In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 8, No.2, April-June 2020, Serial Number 30

Research Institute for Information and Communication Technology
Iranian Association of Information and Communication Technology
Affiliated to: Academic Center for Education, Culture and Research (ACECR)

Print ISSN: 2322-1437
Online ISSN: 2345-2773
Publication License: 91/13216
Editorial Office Address: No.5, Saeedi Alley, Kalej Intersection., Enghelab Ave., Tehran, Iran,
P.O.Box: 13145-799 Tel: (+9821) 88930150 Fax: (+9821) 88930157
E-mail: info@jist.ir , infojist@gmail.com
URL: www.jist.ir

Indexed by:
- SCOPUS www. Scopus.com
- Index Copernicus International www.indexcopernicus.com
- Islamic World Science Citation Center (ISC) www.isc.gov.ir
- Directory of open Access Journals www.Doaj.org
- Scientific Information Database (SID) www.sid.ir
- Regional Information Center for Science and Technology (RICeST) www.ricest.ac.ir
- Iranian Magazines Databases www.magiran.com

Publisher:
Iranian Academic Center for Education, Culture and Research **(ACECR)**

This Journal is published under scientific support of
Advanced Information Systems (AIS) Research Group and
Digital & Signal Processing Research Group, ICTRC

Acknowledgement

JIST Editorial-Board would like to gratefully appreciate the following distinguished referees for spending their valuable time and expertise in reviewing the manuscripts and their constructive suggestions, which had a great impact on the enhancement of this issue of the JIST Journal.

(A-Z)

- Ahmadi, Parvin, Telecommunication Research Center, Tehran, Iran
- Ahmadizad, Arman, University of Kurdistan, Kurdistan, Iran
- Asgari Tabatabaee, Mohammad Javad, University Of Torbat Heydarieh, Razavi Khorasan, Iran
- Ebadati, Omid Mahdi, Kharazmi University,Tehran, Iran
- Ebrahimi, Seyed Babak, K. N. Toosi University of Technology, Tehran, Iran
- Eshaghi, Mohadeseh, Islamic Azad University of Tehran, Iran
- Fadaeieslam, Mohammad Javad, Semnan University, Iran
- Fakhar, Fatemeh, Payame Noor University, Ahvaz, Iran
- Farsijani, Hassan, Shahid Beheshti University, Tehran, Iran
- Farsi, Hassan, University of Birjand, South Khorasan, Iran
- Ghasemzadeh, Ardalan, Urmia University of Technology, West Azerbaijan, Iran
- Ghayoomi, Masood, Institute for Humanities and Cultural Studies, Tehran, Iran
- Ghanbar Tehrani, Nasim, Kharazmi University,Tehran, Iran
- Haghzad Klidbary, Sajad, University of Zanjan, Iran
- Kasaei, Shohreh, Sharif University, Tehran, Iran
- Khalili, Karam, Islamic Azad Ilam University, Ilam, Iran
- Mavadati, Samira, Mazandaran University, Mazandaran, Iran
- Masoomi, Mohsen, Islamic Azad University, Tehran, Iran
- Mirzaei, Abbas, Islamic Azad University, Ardabil, Iran
- Mohammadi, Esfandiar, Ilam University, Ilam, Iran
- Moslehi,Mohammadreza, Institute of Higher Education ACECR, Isfahan, Iran
- Mohammadpour, Davoud, Zanjan University, Zanjan, Iran
- Mahmoudzadeh, Mahmood, Islamic Azad University, Tehran, Iran
- Rafighi, Masoud, Malek Ashtar,Tehran, Iran
- Reshadat, Vahid, Malek Ashtar, Tehran, Iran
- Soleimanian Gharehchopogh, Farhad, Islamic Azad University Urmia, Iran
- Soryani, Mohsen, University of science and technology, Tehran, Iran
- Tanhaei, Mohammad, Ilam University, Ilam , Iran
- Yaghoobi, Kaebeh, ManavRachna International University, India

Table of Contents

Body Field: Structured Mean Field with Human Body Skeleton Model and Shifted Gaussian Edge Potentials

Sara Ershadi-Nasab
Faculty of Electrical Engineering, Sharif University of Technology, Tehran, Iran
ershadinasab@sharif.edu
Shohreh Kasaei*
Faculty of Computer Engineering, Sharif University of Technology, Tehran, Iran
kasaei@sharif.edu
Esmaeil Sanaei
Faculty of Electrical Engineering, Sharif University of Technology, Tehran, Iran
sanaei@sharif.edu
Erfan Noury
Faculty of Computer Engineering, Sharif University of Technology, Tehran, Iran
erfan.noury@gmail.com
Hassan Hafez-Kolahi
Faculty of Computer Engineering, Sharif University of Technology, Tehran, Iran
hafez@ce.sharif.edu

Received: 28/Aug /2019 Revised: 2/Aug/2020 Accepted: 15/Aug/2020

Abstract

An efficient method for simultaneous human body part segmentation and pose estimation is introduced. A conditional random field with a fully-connected graphical model is used. Possible node (image pixel) labels comprise of the human body parts and the background. In the human body skeleton model, the spatial dependencies among body parts are encoded in the definition of pairwise energy functions according to the conditional random fields. Proper pairwise edge potentials between image pixels are defined according to the presence or absence of human body parts that are near to each other. Various Gaussian kernels in position, color, and histogram of oriented gradients spaces are used for defining the pairwise energy terms. Shifted Gaussian kernels are defined between each two body parts that are connected to each other according to the human body skeleton model. As shifted Gaussian kernels impose a high computational cost to the inference, an efficient inference process is proposed by a mean field approximation method that uses high dimensional shifted Gaussian filtering. The experimental results evaluated on the challenging KTH Football, Leeds Sports Pose, HumanEva, and Penn-Fudan datasets show that the proposed method increases the per-pixel accuracy measure for human body part segmentation and also improves the probability of correct parts metric of human body joint locations.

Keywords: Human Body Parts; Skeleton Model; Mean Field Approximation; Pose Estimation; Segmentation; Shifted Gaussian kernel**.**

1- Introduction

Human body part segmentation is the problem of segmenting a given image to human body (HB) parts and the background. The main difference between this process and the general object segmentation is that the HB has an articulated structure. Human pose estimation is defined as the problem of localization of human body joints in the 2D image or 3D space. Human body part segmentation and pose estimation are challenging tasks in computer vision. Their wide applications include surveillance, motion analysis, human-computer interaction, image understanding, augmented reality, and action recognition. As HB has an articulated structure, pose estimation methods aim to find that configuration in a given image. The articulation in HB is often realized by a skeleton model with 14 body joints as well as the corresponding connections among them [1], [2], [3], [4], [5]. The main challenges involved in HB part segmentation and pose estimation are the occluded body parts, the foreshortening effect on the length of some body parts (caused by projection from the 3D space to the 2D image plane), and the ambiguity in defective body parts (due to motion blur or self-occlusion).

In this paper, a new and efficient method for simultaneous HB part segmentation and pose estimation is introduced. The block diagram of the proposed method is shown in Figure 1. The method is based on a *conditional random field* (CRF) graphical model.

* Corresponding Author

The graphical model is a fully connected graph (shown in Figure 2). The graphical model for human skeleton in the proposed dual pose and segmentation method is shown in Figure 3. The label of each image pixel (graph node) is a random variable of this CRF, taking values from the set $p_0, \ldots, p_{14}$, where labels $p_1, \ldots, p_{14}$ are body part labels and p_0 is the background label (see Figure 4). In this work, HB joints are modeled in a graph with 14 nodes and the corresponding connections among graph nodes are determined according to the HB skeleton, as it is shown in Figure 3. In the proposed method, the HB skeleton is not restricted to tree; it can also have cycles. Only the unary and pairwise relations are considered in defining the energy function, and higher order relations (e.g. ternary, quadratic, etc.) are neglected.

The spatial dependency of HB joints in the skeleton model, the length of limbs, and the difference between the features of two joints are encoded in the pairwise terms of the CRF energy function. The main contributions of this paper are summarized as following.

- The semantic human body part segmentation and pose estimation problems are modeled, simultaneously, in a single graphical model. Then, an efficient inference method is proposed to minimize the energy function defined by the model.
- The body length constraint is modeled in the proposed fully connected graphical model by the shifted Gaussian kernels considered in the definition of pairwise energy terms.
- It is demonstrated that although the proposed graphical model is fully connected and Gaussian kernels are shifted, the message passing operation in the inner part of the mean field inference can be computed using the fast bilateral filtering approach. Therefore, the inference algorithm remains tractable.
- Experimental results on the popular and challenging pedestrian parsing benchmark Penn-Fudan dataset [6] for semantic human segmentation, and also on the HumanEva I [7], Extended Leeds Sports Pose [8], and KTH Football I [9] datasets show that the proposed method outperforms the method of Xia [10] that is the state-of-the-art in HB segmentation in terms of per-pixel accuracy measure. It also achieves substantial improvement in finding the locations of corresponding joints according to the *probability of correct pose* (PCP) and *probability of correct key points* (PCK) measures in comparison with Chu *et al.[2]* that is state-of-the-art in 2D pose estimation.

The rest of this paper is organized as follows. In Section 2-, related literature and previous research is reviewed. In Section 3-, the method of Kraehenbuehl *et al.*[11] is reviewed that is necessary for explaining the proposed method. In Section 4-, the proposed method is explained. Next, in Section 5-, experimental results are given. Finally, Section 6- concludes the paper.

2- Related Work

The problem of HB part segmentation and pose estimation can be approached simultaneously. The best graphical model for solving this problem would have to take into account the relations among all image pixels. However, when considering image pixels as the nodes of a fully connected graphical model, the computational cost of the inference step will be very high. Kraehenbuehl *et al*. [11] showed that the inference in dense CRF can successfully be performed by mean field approximation using efficient high dimensional Gaussian filtering operations [12]. The method is specifically designed for the general segmentation problem without any constraint on articulation of HB part.

The kernels are Gaussian functions on the position or color space. No other image features, such as *histogram of oriented gradients* (HOG) [13] are used. Other researchers tried to use this efficient inference and filtering in pose estimation tasks. Vineet *et al*.[14] used this efficient inference in the joint HB pose estimation, segmentation, and depth estimation in a method called *PoseField*.

However, the energy function defined by them is not specialized for HB and does not reflect the HB skeleton model. Kiefel *et al*.[15] tried to extend the inference method introduced in [11] to pose estimation problem. They introduced the *field of parts* method to detect HB joints in 2D images. In their method, the local appearance and joint spatial configuration of HB are modeled. Recently, models based on *deep convolutional neural networks* (DCNN) have been studied extensively in 2D human pose estimation [1], [2], [3], [16].

The *convolutional pose machines* (CPM) architecture proposed by Wei *et al*.[16] is a sequential convolutional neural network that enforces intermediate supervision at the end of each stage to prevent vanishing gradients. DeeperCut [1] is a multi-person pose estimation approach that adapts the deep residual network for human body part detection and uses integer linear programming to jointly detect multiple persons and estimate their body part configurations. Chu *et al*.[2] incorporated the DCNN with a multi-context attention mechanism into an end-to-end framework for human pose estimation. They adapt stacked hourglass networks to generate attention maps from features at multiple resolutions with various semantics. Bulat *et al*.[3] designed a DCNN cascaded architecture specifically for learning part relationships and spatial context. The first part of their cascade outputs part detection heat maps and the second part performs regression on these heat maps to estimate the 2D body pose. Kazemi *et al*.[9] tried to learn the body shape in a discriminative approach using *random forest* (RF) classifier to capture the variations in appearances of HB parts in 2D images. Semantic segmentation and human parsing based on shape-based methods has been studied in [17]. They generate region proposals, rank them using shape and appearance features, and assemble the proposals with simple geometric constraints. A Bayesian framework for jointly estimating articulated body pose and pixel-level segmentation of each body part is proposed in [18].

Wang *et al*.[19] proposed a joint solution that tackles the semantic object and part segmentation, simultaneously. In that method [19], the higher object-level context is provided to guide the part segmentation process. Also, more detailed part-level localization is utilized to refine the object segmentation process.

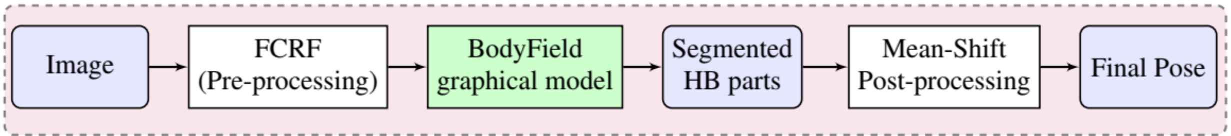


Figure 1: Schematic view of proposed method.
Input: Image block. Outputs: Segmented HB Parts and Final Pose.

A *deep decompositional network* (DDN) for parsing pedestrian images into semantic regions is proposed in [20]. This method tries to directly map low-level visual features to the label maps of body parts. Top-down pose cues as well as deep-learned features are used in an *and-or graph* (AOG) for semantic part assembling [10]. This method tries to refine the semantic parts of objects by using the pose cues. DeepLab framework [21] augments fully convolutional network with dilated convolutions, atrous spatial pyramid pooling, and CRF. DeepLab obtains state-of-the-art performance in general problem of semantic segmentation. Guler *et al*. [22] proposed a surfaced based framework for dense human pose estimation and body part segmentation. It is based on finding dense correspondence between image and a surface of human body. Since there is not a large-scale dataset containing correspondence between image and human body surface, this method has some challenges with general and natural images. An occlusion aware framework for human pose estimation is proposed in [23]. It is based on adversarial training of a *Convolutional Neural Network* (CNN). They designed discriminators to distinguish the real poses from the fake ones (such as biologically implausible ones) to avoid fake estimated poses. Peng *et al*. [24] used data augmentation method in training phase of an adversarial learning framework. They proposed to optimize data augmentation and network training jointly to avoid overfitting for the task of human pose estimation. *Yang et al.* [25] tried to learn 3D human pose structure from a dataset with only 2D pose annotation as the ground-truth. Their method is based on an adversarial learning framework using multi-source discriminators to distinguish the predicted 3D poses from the ground-truth one. In fact, they tried to enforce the pose estimator to generate anthropometrically valid poses even with images from natural scenes. Chen *et al.* [26] proposed a method for multi-person pose estimation in challenging scenes that contain occluded or invisible keypoints and complex backgrounds. They used cascaded networks of GlobalNet and RefineNet. Simple key points like eyes and hands are localized with the GlobalNet. Hard keypoints such as occluded or invisible key points are addressed with the RefineNet network. Also, this method handles only the pose estimation problem and does not handle the body part segmentation problem. PoseTrack is a large-scale benchmark for video-based human pose estimation and articulated tracking [27]. It is a more suitable dataset for multiple human tracking task in video sequences rather than body part segmentation since it does not have any ground-truth information for human body segmented regions. It is worth mentioning that the proposed method is different from the Kraehenbuehl *et al.*'s work[11], in that in the proposed method, the CRF formulation is specifically defined according to the HB configuration such that HB segments are naturally considered to appear in a set of constrained positions relative to each other.

Also, the definitions of pairwise energy terms are different from that work. Since the graphical model used in the proposed method is a fully-connected graph constrained to image pixels, it is similar to the work of Kiefel *et al.* [15], albeit that method does not produce the HB part segmentation and they only report the PCP values on the Leeds Sports Pose [8] dataset.

3- Efficient Mean Field in Object Segmentation

Kraehenbuehl *et al.* [11] proposed an efficient inference in mean field approximation for general segmentation problem. Their method is not designed for articulated objects such as human body and is only evaluated in PASCAL dataset for general object segmentation problem.

In this Section a brief description of Kraehenbuehl *et al.*'s [11] method is reviewed that is needed for introducing the proposed method in the next Section. They defined a conditional random field over a set of random variables $X = \{x_1, \dots, x_N\}$, where N is the total number of pixels in image I. Each variable has a set of possible labels $P = \{p_0, \dots, p_k\}$, where p_0 corresponds to the background and $p_1, \dots, p_k$ are possible pixel labeling.

The conditional random field is characterized by the Gibbs energy function defined on this graph by

$$E(x) = \sum_{i} \psi_{unary}(x_i = p) + \sum_{i<j} \psi^{(1)}_{pairwise}(x_i = p, x_j = p') \quad (1)$$

where i, j range from 1 to N.

The Gibbs energy function is a summation of pairwise and unary terms. The $\psi_{unary}(x_i = p)$ is the cost of assigning label p to random variable x_i . The second term, $\psi^{(1)}_{pairwise}(x_i = p, x_j = p')$, measures the cost of assigning label p and p' to two neighboring pixels i and j , respectively. The pairwise term is the cost of assigning two different labels to two arbitrary pixels, given by

$$\psi^{(1)}_{pairwise}(x_i = p, x_j = p') = \mu_1(p, p') \sum_{m=1}^{M} w_1^{(m)}\, k^{(m)}(f_i, f_j) \quad (2)$$

where $k^{(m)}(f_i, f_j)$ is a Gaussian kernel and is defined as

$$k^{(m)}(f_i, f_j) = \exp\left\{-\frac{1}{2}\left((f_i - f_j)^T (\Sigma^{(m)})^{-1} (f_i - f_j)\right)\right\} \quad (3)$$

in which vectors f_i and f_j are feature vectors of pixels i and j in an arbitrary feature space, respectively, $w_1^{(m)}$ is the weight of the kernel, m is the index of the kernel, and M is the number of kernels. $\Sigma^{(m)}$ is matrix of variances between f_i and f_j of $m-$th Gaussian kernel. The energy function defined in the CRF formulation is minimized during the inference phase. The mean field approximation is an iterative process that instead of computing the exact distribution P, computes the approximated $Q(x)$ such that minimizes the KL -divergence $D(Q||P)$ among all distributions, where Q can be expressed as the product of independent marginal $Q(x) = \prod_i Q_i(x_i))$. According to the energy function defined in Equation(1), the closed-form solution of the mean field approximation can be written as

$$Q(x_i = p) = \frac{1}{Z_i} \exp\{-\psi_{unary}(x_i = p) - \hat{Q}_1(x_i = p)\} \quad (4)$$

where $Q(x_i = p)$ is the belief of pixel i about having the label p and is updated in iterative steps. Z_i is defined as $Z_i = \sum_{p=1}^{P} Q(x_i = p)$ and is the normalization term. Also, $\psi_{unary}(x_i = p)$ is the initial belief about pixel i having the label p. The belief of all other pixels about pixel i having the part label p is defined as

$$\hat{Q}_1(x_i = p) = \sum_{p' \in P} \mu_1(p, p') \sum_{m=1}^{M} \omega_1^{(m)} \tilde{Q}_1^{(m)}(x_i = p') \quad (5)$$

in which, $\mu_1(p, p')$ is the label compatibility function between two possible labels p and p' for each pixel.

A simple label compatibility function is the *Potts model*, in which

$$\mu_1(p, p') = 1(p \neq p') \quad (6)$$

where $1(p \neq p')$ denotes the indicator function.

$\omega_1^{(m)}$ is the weight of $m-$th Gaussian kernel, M is the total number of kernels, and

$$\tilde{Q}_1^{(m)}(x_i = p') = \sum_{i \neq j} k^{(m)}(f_i, f_j) Q(x_j = p') \quad (7)$$

in which $k^{(m)}(f_i, f_j)$ is a Gaussian kernel as is defined in Equation (3). It is worth mentioning that Equation (7) is performed once for all pixels by using the Permutohedral lattice filtering. Every channel p' of matrix Q is blurred by Gaussian kernel of $k^{(m)}(f_i, f_j)$ as in Equation (7) that are applied on all image pixels. By substituting Equations (5), (6), and (7) in Equation (4) the message passing is performed as

$$Q(x_i = p) = \frac{1}{Z_i} \times \exp\{-\psi_{unary}(x_i = p) - \sum_{p' \in P} \mu(p, p') \sum_{m=1}^{M} \omega_1^{(m)} \sum_{j \neq i} k^{(m)}(f_i, f_j) Q(x_j = p')\}. \quad (8)$$

Since the graphical model is a fully-connected graph, the message passing step is the bottleneck of the mean field approximation. Its run-time is quadratic in the number of pixels N.

4- Proposed Method

The block diagram of the proposed method is illustrated in Figure 1. The Image block is input to the method, the FRCF block is a pre-processing step that computes the initial pose that is needed in the next block. The details of this pre-processing step are explained in Subsection 4-1-. The BodyField Graphical model is the proposed method that is explained in detail in Subsection 4-2-. The Segmented HB parts are the output of the method. The Mean-Shift block is a post-processing step that is applied to the distribution of the segmented body parts for computing the final estimated pose. The Final Pose block is the final estimated pose and output of the method.

4-1- Pre-processing: Computing the Initial Pose by a Fully Connected Pairwise CRF

A fully connected pairwise CRF is proposed for computing the initial pose that is needed in the proposed dual pose and segmentation method. The graphical model of human body according to this CRF is shown in Figure

2. The nodes of this graph are human body joints that all of them are connected to each other.

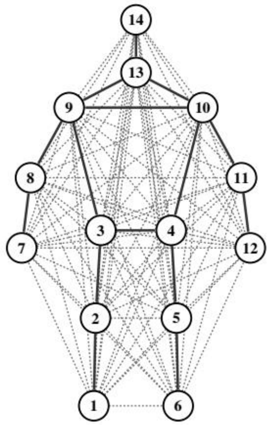


Figure 2. Proposed fully connected model of human body. This model is used in pre-processing step to find the initialy estimated pose.

Images are initially processed with the DeeperCut 2D part detector [1] and the score map of body joints in the images are obtained. The score map, S, is an array of size $W \times H \times 15$ where W and H are the width and height of the image, respectively, and 15 is the number of body joints (14) plus a special class for the background. The unary term of the energy function is computed by the first output of body part detector, S, as

$$\psi_{unary}(X_{(i,j)} = p) = S(x^{(i,j)}, y^{(i,j)}, p). \quad (9)$$

Another output of 2D part detector is R, that is an array of size $W \times H \times 14 \times 13 \times 2$, where 14×13 indicates the number of permutations of length two of 14 distinct variables, and 2 is for two dimensions x and y.
According to the output R of the part detector [1]

$$\left(\delta x^{(i,j)}, \delta y^{(i,j)}\right) = R\left(x^{(i,j)}, y^{(i,j)}, \upsilon(p,p')\right) \quad (10)$$

which implies that if a pixel in location (i,j) has the joint label p, it is expected that the joint p' will occur with an offset $\left(\delta x^{(i,j)}, \delta y^{(i,j)}\right)$ from it. Also, $\upsilon(p,p')$ is an index between 1 and 182 which indicates one of the possible permutations 14×13 belonging to joints p and p', according to [1]. Therefore, if joint i is in location $(x^{(i,j)}, y^{(i,j)})$, then the model expects that joint j to be in location

$$\left(\tilde{x}^{(i,j)}, \tilde{y}^{(i,j)}\right) = \left(x^{(i,j)} + \delta x^{(i,j)}, y^{(i,j)} + \delta y^{(i,j)}\right). \quad (11)$$

In the same way, if a pixel in location (i',j') has the joint label p', acording to the output of the part detector[1], it expects that the joint p be in the offset

$$\left(\delta x^{(i',j')}, \delta y^{(i',j')}\right) = R\left(x^{(i',j')}, y^{(i',j')}, \upsilon(p,p')\right), \quad (12)$$

from it. Therefore the expected location of joint p from the point of view of pixel (i',j') that has joint label p' is

$$\left(\tilde{x}^{(i',j')}, \tilde{y}^{(i',j')}\right) = \left(x^{(i',j')} + \delta x^{(i',j')}, y^{(i',j')} + \delta y^{(i',j')}\right). \quad (13)$$

The difference vector between the expected location of joint p' from the point of view of pixel (i,j) that has joint label p and pixel (i',j') that has joint label p' is

$$\Delta^{(1)} = \| (\tilde{x}^{(i',j')}, \tilde{y}^{(i',j')}) - (x^{(i',j')}, y^{(i',j')}) \|. \quad (14)$$

Also, the difference vector between the expected location of joint p from the point of view of pixel (i',j') that has joint label p' and pixel (i,j) that has joint label p is

$$\Delta^{(2)} = \| (\tilde{x}^{(i,j)}, \tilde{y}^{(i,j)}) - (x^{(i,j)}, y^{(i,j)}) \|. \quad (15)$$

The pairwise term as the cost of assigning label p to pixel (i,j) and label p' to pixel (i',j') is defined as

$$\psi_{pairwise}(X_{(i,j)} = p, X_{(i',j')} = p') = exp\left\{-\frac{1}{2} \| \Delta^{(1)} + \Delta^{(2)} \|_2^2\right\}. \quad (16)$$

The inference in the proposed fully connected CRF is computed by the loopy belief propagation method [28]. Using this pre-processing step improves the estimated pose of the DeeperCut method. The comparison between the estimated pose in this pre-processing step, (FCRF), and DeeperCut method is provided in experimental results Section 5-. The initial pose obtained by the pre-processing, (FCRF), is used in computation of the amount of needed shift values in the proposed method in the next Section.

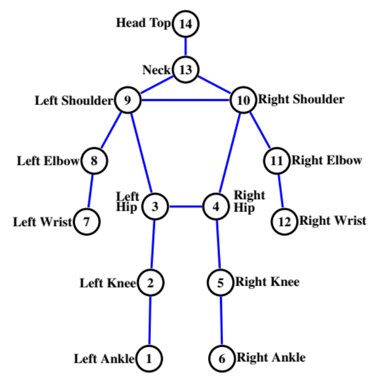


Figure 3. Proposed graph for human skeleton model in proposed dual pose and segmentation method.

4-2- BodyField Graphical Model Definition

According to Figure 3, human skeleton model is considered to contain 14 joints and their connections are set according to the HB configuration. Furthermore, as the graph is not restricted to be a tree, the model can easily be extended to arbitrary number of HB parts and there is no hard constraint on the number of joints in the model. Figure 4 illustrates the proposed fully-connected graphical model. The nodes in the proposed graphical model are image pixels, and pairwise terms are weights of any connection between two arbitrary pixels. A pixel i is shown to be connected to all other pixels with labels in $p_0, \ldots, p_{14}$. It is also true for all other pixels (due to visualization restrictions, other connections are not shown). Also, it is important to note that there are no connections, and thus pairwise terms, between a pixel and itself. Since the pairwise terms between two pixels are constrained to the label compatibility, for visualization purposes, image labels are separated to $L = 15$ channels. These 15 channels should be added to create a fully connected graphical model. Therefore, there are $W \times H$ nodes in the graph, in which W and H are the width and height of the image, respectively. Also, $Q(X = p)$ is the probability of assigning label p to a set of image pixels X. Energy function should be defined such that a true configuration of HB corresponds to the minimum value of the energy function, otherwise, finding the minimum value of the energy function will not lead to a good configuration. Note that the sum of probability values of parts for each pixel is one. The label assigned to each pixel is the HB part with the highest probability value among all HB parts and the background. The pairwise energy terms are defined such that the pairwise terms have lower values when the two paired pixels have corrected HB part labels. In the general segmentation problem, it is assumed that pixels that are close to each other (in the feature space) lie in the same segment. It can be met in general segmentation problems, but it does not always hold in HB part segmentation.

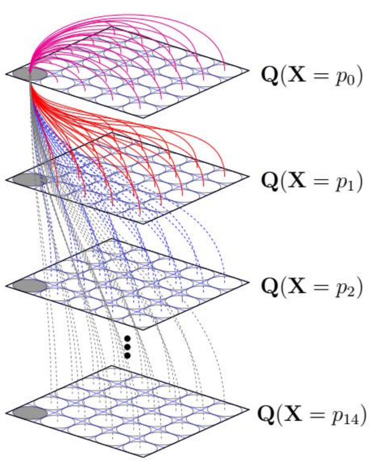


Figure 4. Fully connected graph for human skeleton model.

Some body parts should occur in pre-defined distances to each other in accordance to the existence of a connection among related joints in the HB skeleton model. The inference process tries to find the minimum of the energy function; in the final solution, all nearby (generally, in the feature space) pixels will have similar labels. But, in pose estimation problems, this is not always true. The reason is simply that there might be nearby and similar pixels in the image of HB that do not belong to the same part. In images of HB, there are three common categories of relationships among pixels.

- Pixels that are close in the feature space and belong to the same HB part.
- Pixels that are close in the feature space but do not belong to the same HB part, however their corresponding parts are connected in the HB skeleton model.
- Pixels that may or may not be close in the feature space and do not belong to the same HB part, but their corresponding parts are not connected in the HB skeleton model.

Pixels belonging to the third type can move and eventually appear close to each other; e.g. the wrist can appear near the other parts of HB. When defining the energy function and pairwise terms, all of these situations should be considered and the suitable kernel and compatibility functions should be assigned to any two labels. For any two arbitrary pixels, according to their labels, two different pairwise terms are defined. One for resolving the first and the third type and the other for resolving the second type. In the proposed method the energy function is defined as

$$E(X|I,\theta) = \sum_i \psi_{unary}(x_i = p|I,\theta) + \sum_{i<j} \psi^{(2)}_{pairwise}(x_i = p,\ x_j = p'|I,\theta) + \sum_{i<j} \psi^{(2)}_{pairwise}(x_i = p,\ x_j = p'|I,\theta) \quad (17)$$

where i, j range from 1 to N. Variable θ denotes the set of parameters of HB. It is computed by the initial pose that is estimated in the Subsection 4-1-.

For the sake of conciseness, in the remainder of the paper, I and θ are omitted in equations. If these two pixels are close to each other in the feature space, the energy cost for assigning different labels to these two pixels is high. When minimizing the energy function during the inference process, this configuration of labeling (two nearby pixels with two different labels) will be avoided. Therefore, in the best configuration, neighboring pixels approach towards getting identical labels. This is generally true in articulated HB shapes and therefore these pairwise terms are defined between any two arbitrary pixels by using a simple Potts model. The second type of pairwise energy terms is specifically defined to encode HB joints' constraints in the proposed CRF formulation, given by

$$\psi^{(2)}_{pairwise}(x_i = p, x_j = p') = \mu_2(p,p') \sum_{m=1}^{M} \omega_2^{(m)} \kappa^{(m)}_{p,p'}(f_i, f_j) \quad (18)$$

where $\omega_2^{(m)}$ is the weight of the shifted kernel function. The label compatibility function $\mu_2(p, p')$ is defined as

$$\mu_2(p,p') = -1(p\ is\ connected\ to\ p') \quad (19)$$

according to the existence of a connection between body part p and body part p' in the HB skeleton model as it is shown in Figure 3. The value of $\kappa_{p,p'}^{(m)}(f_i,f_j)$ is defined as

$$\kappa_{p,p'}^{(m)}(f_i,f_j) = \exp\left\{-\frac{1}{2}\left((f_i-f_j-\eta_{p,p'}^m)^T(\Sigma_{p,p'}^{(m)})^{-1}(f_i-f_j-\eta_{p,p'}^m)\right)\right\} \quad (20)$$

in which $(\Sigma_{p,p'}^{(m)})^{-1}$ is the variance matrix between feature vector of joint p and p' of m −th shifted Gaussian kernel. $\eta_{p,p'}^m$ is the mean expected difference vector between the features f_i and f_j. When the features are simply the positions of points, the value of $\eta_{p,p'}^m$ is a difference vector that is computed from the initial pose that is estimated by the preprocessing step of Subsection 4-1-.

Let us consider two arbitrary pixels which have two different labels and are connected according to their labels in the HB skeleton model. The pairwise term that is defined for these two pixels, takes the minimum value when these pixels are placed at a predefined distance from each other. By this definition, the Gaussian term is shifted by $\eta_{p,p'}^m$, such that the mean of Gaussian lies on pixels for which the difference between their features and the feature of pixel i is $\eta_{p,p'}^m$. The pairwise energy is the weight of edges in the fully connected model between pixels and it is constrained on labels of pixels. There will be 15×14 pairwise energy terms between any two pixels. There are some constraints on HB skeleton model according to the skeleton graph. The goal of the proposed method is enforcing all constraints presented in the HB skeleton model in the mean field approximation process. Note that, up to here, body part lengths and nearby joints that are connected in the skeleton graph are successfully encoded in the energy function definition of the fully connected conditional random field model that is defined on image pixels. In defining the pairwise energy terms between two arbitrary pixels, $XYRGBHOG$ kernel is used that is a 36-D vector $f_i = (x_i, y_i, r_i, g_i, b_i, f_i^{(1)}, f_i^{(2)}, \ldots, f_i^{(31)})$ where x_i and y_i are coordinates of pixel i, r_i, g_i, b_i, the RGB values of the pixel, and $f_i^{(l)}$ is the l^{th} element of HOG feature vector of the cell containing that pixel. f_i and f_j in Equation (20) are defined as above.

4-3- Efficient Inference Via High Dimensional Gaussian Filtering

According to the energy function defined in Equation (17), the closed-form solution of the mean field approximation can be written as

$$Q(x_i=p) = \frac{1}{Z_i}\exp\{-\psi_{unary}(x_i=p) - \hat{Q}_1(x_i=p) - \hat{Q}_2(x_i=p)\} \quad (21)$$

where $Q(x_i = p)$ is the belief of pixel i about having the label p and is updated in iterative steps. Z_i is defined as $Z_i = \sum_{p=1}^{P} Q(x_i = p)$ and is the normalization term.

Also, $\psi_{unary}(x_i = p)$ is the initial belief about pixel i having the label p. $\hat{Q}_1(x_i = p)$ and $\hat{Q}_2(x_i = p)$ are the belief of all other pixels about pixel i having the part label p.

The value of $\hat{Q}_2(x_i = p)$ in Equation (21) is defined as

$$\hat{Q}_2(x_i=p) = \sum_{p'\in P}\mu_2(p,p')\sum_{m=1}^{M}\omega_2^{(m)}\tilde{Q}_2^{(m)}(x_i=p') \quad (22)$$

in which $\mu_2(p,p')$ is the label compatibility function and is defined in Equation (19), $\omega_2^{(m)}$ is the weight of shifted Gaussian kernel, and

$$\tilde{Q}_2^{(m)}(x_i=p') = \sum_{j\neq i}\kappa_{p,p'}^{(m)}(f_i,f_j)Q(x_j=p') \quad (23)$$

in which $\kappa_{p,p'}^{(m)}(f_i,f_j)$ is a shifted Gaussian kernel, is defined in Equation (20).

It is worth mentioning that Equations (7) and (23) are performed once for all pixels by using the Permutohedral lattice filtering.

Every channel p' of matrix Q is blurred by Gaussian kernel of $k^{(m)}(f_i,f_j)$ as in Equation (7) and by shifted Gaussian kernel of $\kappa_{p,p'}^{(m)}(f_i,f_j)$ as in Equation (23) that are applied on all image pixels. By substituting Equations (5), (7), (22), and (23) in Equation (21) the message passing is performed as

$$Q(x_i=p) = \frac{1}{Z_i}\times\exp\{-\psi_{unary}(x_i=p) - \sum_{p'\in P}\mu_1(p,p')\sum_{m=1}^{M}\omega_1^{(m)}\sum_{j\neq i}k^{(m)}(f_i,f_j)Q(x_j=p') - \sum_{p'\in P}\mu_2(p,p')\sum_{m=1}^{M}\omega_2^{(m)}\sum_{j\neq i}\kappa_{p,p'}^{(m)}(f_i,f_j)Q(x_j=p')\}. \quad (24)$$

Since the graphical model is a fully-connected graph, the message passing step is the bottleneck of the mean field approximation. Its run-time is quadratic in the number of pixels N. As another contribution of the proposed method, shifted Gaussian kernels are used in the pairwise terms in addition to the non-shifted Gaussian kernels, while keeping the inference step computationally tractable.

4-4- Implementation Details of Shifted Gaussian Kernels

Permutohedral lattice high dimensional Gaussian filtering, performs the filtering task in three steps [12]:

- Splatting the points to the lattice space.
- Performing the blurring process in lattice space.
- Slicing the lattice to find the final values of blurred points.

Splatting is the initial phase of lattice construction in high dimensional space according to the definition in [12]. Since we want to blur the value of $Q(x_j = p')$ with position vectors that are shifted by $\eta_{p,p'}^{(m)}$, it implies that at first the position vector shifts by $\eta_{p,p'}^{(m)}$ before performing the blurring task. But, in lattice space the operation of shifting and then blurring is equivalent to blurring and then slicing at the shifted positions. Substituting Equation (20) in Equation (23) will result in

$$\tilde{Q}_2^{(m)}(x_i = p') = \sum_{i \neq j} exp\left\{-\frac{1}{2}\left(f_i - f_j - \eta_{p,p'}^{(m)}\right)^T \left(\Sigma_{p,p'}^{(m)}\right)^{-1} \left(f_i - f_j - \eta_{p,p'}^{(m)}\right)\right\} \times Q(x_j = p'). \quad (25)$$

The *permutohedral lattice filter* [12] is implemented in the ImageStack library [29], which is a toolbox for high dimensional Gaussian filtering. It is used for performing the high dimensional blurring in the inference step of the proposed method. In implementation process, according to Equation (25), Q is a matrix of size $(W \times H \times L)$, given that the input image is of size $(W \times H)$, where L is the total number of body parts and background labels ($L = 15$). $Q(x_i = p)$ is the probability of part p for each arbitrary x_i pixel of the image. It is necessary that $\sum_{p \in P} Q(x_i = p) = 1$, in which $i \in N$ and N is the set of all image pixels. Taking Equations (7) and (23) into account, it is apparent that both equations are similar, except that in the former, Gaussian weights are shifted. Baek *et al*.[30] proved that to use shifted Gaussian kernels, it is sufficient to slice the lattice at shifted positions. Using the ImageStack library, the lattice points are ordinary position vectors without shifting and the values of $Q(x_j = p')$ are blurred in the lattice space by using position vectors in Gaussian weights. Afterwards, the lattice should be sliced to find the final values of $Q(x_j = p')$ in the initial space. In Equation (7), for all channels of matrix Q, these operations are performed once using only a single lattice. Permutohedral lattice filter reduces the time complexity of Gaussian operation to $O(nd^2)$, where n is the number of points to be blurred and d is the dimension of the position space (despite the fact that its three required steps of splatting, blurring, and slicing are time consuming; specifically in high dimensional spaces like HOG feature space). In the proposed method, the shifted Gaussian filtering is performed for several times, which is time consuming. It is worth mentioning that to further speed-up the process, one can force all $\Sigma_{p,p'}^{(m)}$ to be the same for all p and p', and therefore some steps need only be performed once for updating the belief about each label, as done in Equation (24). Note that for all shifted Gaussian kernels that use the same feature space and covariance matrix, constructing and blurring the lattice in the feature space is only performed once. On the contrary, due to different values of $\eta_{p,p'}^{(m)}$, the lattice is sliced in different shifted positions.

5- Experimental Results

The proposed method is evaluated on (i) KTH Football I dataset with 3900 training images and 2007 test images, (ii) Extended Leeds Sports Pose dataset with 11000 training and 1000 test images, and (iii) Sequences 1 and 2 of the HumanEva I dataset for jogging, walking, and balance actions. Since the proposed method has the 2D pose output in addition to the body part segmentation output, it is also evaluated on those datasets that have 2D pose annotations. The obtained results are then compared with that of the 2D pose estimation methods. It should be noted that as the KTH Football I, Extended Leeds Sports Pose, and HumanEva I datasets do not have any ground-truth data annotations for HB part segmentation, the results can only be evaluated qualitatively. For evaluating the proposed method in human body segmentation, the Penn-Fudan dataset is used. It contains 170 test images and the ground-truth of the body part segmentation. Some typical results of the proposed HB part segmentation are shown in Figure 5, Figure 6, Figure 7 and Figure 8 in which the first column (a) shows the original images. BodyField segmentation result and estimated pose are shown in column (b) of these Figures. In column (c), the HB part segments of the BodyField method is visualized by similar colors as in the ground-truth of Penn-Fudan dataset using the estimated pose of the BodyField method. As it can be seen from Figure 5, Figure 6, Figure 7, and Figure 8, the locations of joints have been estimated accurately, due to the refined HB part segmentation obtained by the proposed method. In KTH Football the PCP metric is used to evaluate the accuracy of pose estimation methods. According to the definition in [31], a part is considered correctly localized if the average distance between its endpoints (joints) and the ground-truth data is less than α times of the length of annotated endpoints in the ground-truth data.

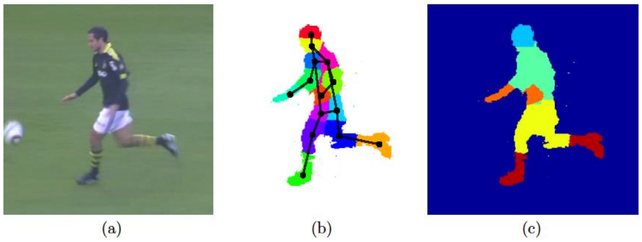


Figure 5. (a) Original image from KTH Football I dataset. (b) Pose and segmentation result of proposed method. (c) Different visualization of proposed method.

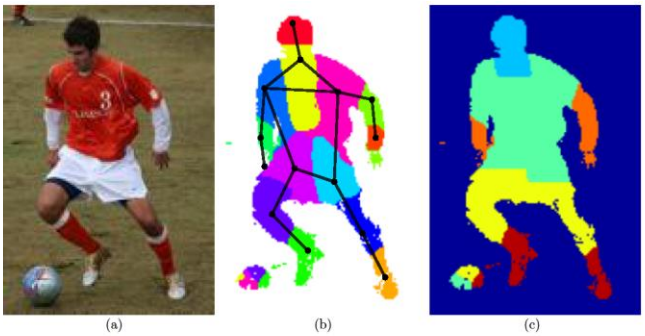


Figure 6. (a) Original image from Leeds Sports Pose dataset. (b) Pose and segmentation result of proposed method. (c) Different visualization of proposed method.

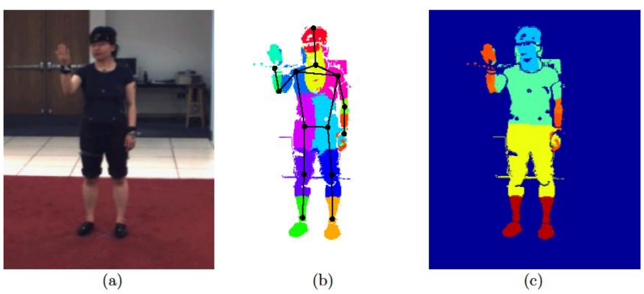


Figure 7. (a) Original image from HumanEva I dataset. (b) Pose and segmentation result of proposed method. (c) Different visualization of proposed method.

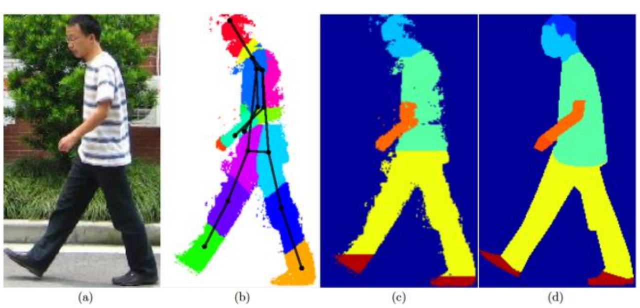


Figure 8. (a) Original image from Penn-Fudan dataset. (b) Pose and segmentation result of proposed method. (c) Different visualization of the proposed method. (d) Ground-truth.

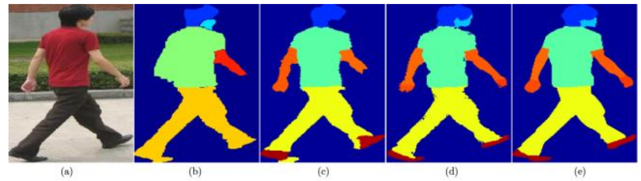


Figure 9. (a) Original image from Penn-Fudan dataset. (b) Output of Bo et al. [17] method. (c) Output of Xia [10] method. (d) Segmentation result of proposed method. (e) Ground-truth.

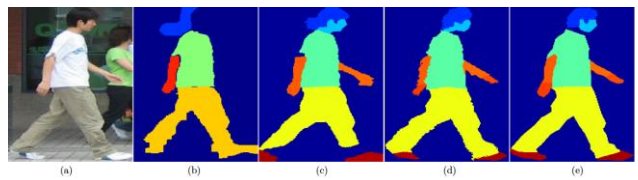


Figure 10. a) Original image from Penn-Fudan dataset. (b) Output of Bo [17] et al. method. (c) Output of Xia et al. [10] method. (d) Segmentation result of proposed method. (e) Ground-truth.

For the Penn-Fudan dataset there is a ground-truth segmentation for body part segmentation. Since the face and hair are segmented in two different classes in this dataset we used these data in training phase of the Body Field method. In fact the training phase is performed with one more extra class for this dataset. It shows the generalizability of the proposed method to the datasets that have more fine body parts segmented regions. We could define extra classes for each of the fine segmented regions and compute the mean expected difference vector between fine regions and other classes to use in training phase of the BodyField method. Quantitative results of the proposed method on the challenging KTH Football I datasets are summarized in Table 1. The pre-processing step, (FCRF), improves the results obtained by the DeeperCut method by up to 6%. Also DeeperCut [1] method is evaluated on this dataset and it has 85% total PCP. DeeperCut is a powerful body part detector. It uses integer linear programming for estimating the pose from the probability map. However, it sometimes fails to estimate the correct pose of player because of high degree of motion blur in images of KTH Football I dataset. The proposed BodyField method has 99% PCP and improves the results of the original DeeperCut method by 14%. Also the proposed BodyField method improves the results obtained by Kazemi *et al*.[9] in terms of PCP measure by up to 10% due to its better and refined HB part segments. In the Extended Leeds Sports Pose dataset, the standard *probability of correct key points* (PCK) evaluation metric is used [1], [2]. According to the definition in [31], a candidate key point is considered to be correct if it falls within $\beta \times max\,(h, w)$ pixels of the ground-truth key point, where h and w are the height and width of the bounding box of human respectively, and β controls the relative threshold for considering correctness. Results in Table 2 is based on Person-Centric ground-truth with $\beta = 0.2$. In Table 2, comparison results of the proposed method with the method of Chu *et al.*[2], Bulat *et al.*[3], Wei *et al.*[16], and Insafutdinov *et al.*[1] are presented. As it can be seen from Table 2, the original method of Insafutdinov *et al.*[1] has 90.1% PCK. The pre-processing step, FCRF, improves the PCK to 91.8%. The proposed BodyField method has 96.2% efficiency in terms of PCK measure. It outperforms the original method of Insafutdinov *et al.*[1] by 6.1%, and also the method of Chu *et al.*[2] by 3.6% in terms of PCK measure. For the HumanEva dataset, the standard method for computing the accuracy of pose estimation methods is the average 2D error [7]. The proposed method is evaluated on sequences 1 and 2 in walking, jogging and balance actions. As it can be seen in Table 3, the 2D error between the estimated pose and ground-truth location of joints is decreased by using the proposed BodyField method.

The overall average 2D error of Sigal *et al.*[7] is $10.7 \pm 1\,pix$, while it decreases to $5.4 \pm 1.2\,pix$ in pre-processing step, FCRF, and to $2.9 \pm 1\,pix$ in the proposed BodyField methods. Since the official evaluation server of the HumanEva dataset, http://humaneva.is.tue.mpg.de/, is currently out of service, we used the validation set for reporting the average values of 2D error.

Table 1. PCP values for KTH Football dataset (Observer-Centric,$\alpha = 0.5$)

Method	Torso	Upper Leg	Lowe Leg	Upper Arm	ForeArm	Head	Total
BodyField	**100**	**98.6**	**98.7**	**98.9**	**98.8**	**100**	**99**
FCRF(Pre-processing)	95	98	90	92	82	91	91
Kazemi *et al*. (RF) [9]	96	94	84	90	69	94	87
Kazemi *et al*.(RF+PosePrior) [9]	98	97	88	93	71	96	89
DeeperCut [1]	91	91	87	89	72	82	85
Belagiannis *et al*. [32]	98	92	80	88	57	86	84
Yang & Ramanan [31]	98	89	73	86	55	84	80

Table 2. PCK values for Extended Leeds Sports Pose dataset (Person-Centric,$\beta = 0.2$)

Method	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
BodyField	98.2	96.8	97.5	93.7	85.9	96.3	95.6	96.2
Chu *et al*. [2]	98.1	93.7	89.3	86.9	93.4	94.0	92.5	92.6
FCRF (Pre-processing)	90.9	90.8	92.2	92.9	90.7	91.9	90.8	91.8
Bulat *et al*. [3]	97.2	92.1	88.1	85.2	92.2	91.4	88.7	90.7
Wei *et al*. [16]	97.8	92.5	87.0	83.9	91.5	90.8	89.9	90.5
DeeperCut [1]	97.4	92.7	87.5	84.4	91.5	89.9	87.2	90.1

Table 3.Average 2D error for HumanEva I dataset

	Sequence 1			Sequence 2			Overall
Method	Walk	Jog	Balance	Walk	Jog	Balance	
BodyField	3.9± 0.3 pix	2.9±0.6	2.2±1.4	2.8±1 pix	2.4±1.7pix	3.2±0.9pix	2.9±1pix
FCRF(Pre-processing)	4.1±0.4pix	5.6±0.8pix	5.4±1.6pix	3.9±1.2pix	7.2±2.1pix	5.9±1.2pix	5.4±1.2pix
Sigal[7]	10.1±0.9pix	11.3±0.7pix	11.3±2.3pix	7.9±0.6pix	12.4±2.3pix	10.9±2.8pix	10.7±1pix

Table 4. Comparison of our approach with other state-of-the-art methods on the Penn-Fudan benchmark dataset in terms of per-pixel accuracy (%). The Avg$*$ means the average without shoes class since it was not reported in other methods.

Method	hair	face	u-cloth	arms	l-cloth	legs	shoes	Avg*
BodyField	63.4	61.7	79.8	58.4	82.3	65.0	47.2	65.4
AOG [10]	63.2	56.2	78.1	40.1	80.0	45.5	35.0	60.5
DDN [20]	43.2	57.1	77.5	27.4	75.3	52.3	------	56.2
SBP [17]	44.9	60.8	74.8	26.2	71.2	42.0	------	53.3
P&S [18]	40.0	42.8	75.2	24.7	73.0	46.6	------	50.4
Wang *et al*.[19]	48.7	49.1	70.2	33.9	69.6	29.9	36.1	50.2

The proposed method is evaluated on the popular Penn-Fudan benchmark [6], which consists of pedestrians in outdoor scenes with much pose variations. Labels of the dataset include 7 body parts namely hair, face, upper-clothes, lower-clothes, arms (arm skin), legs (leg skin), and shoes.

Also, since the proposed method segments the human body parts into 14 classes, we used the mapping process to convert the corresponding classes to those used in the Penn-Fudan dataset. For this conversion, the estimated pose is used as auxiliary information. The typical part segmentation results in these datasets are illustrated in Figure 8, Figure 9 and Figure 10. This dataset does not have the ground-truth of joints and therefore the results in pose estimation cannot be compared in this dataset in terms of PCP or PCK measures. But since for this dataset the ground truth for segmentation is available in pixel by pixel, the standard evaluation metric is used as per-pixel accuracy [10]. In Figure 9 and Figure 10 the comparison between the proposed method and the method of Xia *et al.* [10] and Bo *et al*. [17] are provided. For the method of Xia *et al*. [10] and Bo *et al*. [17] we used the source image provided by the authors. As shown in Table 4, the proposed method is compared with state-of-the-art methods, namely, AOG [10], DDN[20], P&S[18], SBP [17], and Wang *et al.* [19] on the Penn-Fudan dataset. The proposed BodyField method outperforms the DDN [20] method by over 9% and it has 4.9% improvement in comparison with the state-of-the-art method of Xia *et al*.[10] (AOG method). The improvement in the proposed method is due to the fact that estimated pose and corresponding body part segments are refined simultaneously. In other words, use of pose information in semantic human body part segmentation has increased the per-pixel accuracy. More visual output results of the inner steps of the BodyField method are available in http://ipl.ce.sharif.edu/bodyfield.html.

6- Conclusion

A new and efficient method for simultaneous single-view human body part segmentation and pose estimation is introduced that opens a new approach to the problem of structured semantic segmentation. A new energy function is introduced that encodes the spatial dependency between human body parts, in addition to the available segmentation constraints. In the proposed method, despite the fact that shifted Gaussian kernels are used, it is shown that finding the minimum of the proposed energy function is possible by applying an efficient mean field approximation process. Due to challenges such as occlusion and self-occlusion effects that occur frequently in human body pose data, the previous learning methods that only use the appearance model cannot converge to a proper pose estimation. That is because there are not enough evidences about the occluded and self-occluded parts available. The proposed BodyField method uses the probability map of the DeeperCut method to define a proper energy function with shifted Gaussian kernels between connected body parts. During the inference step, the evidence for occluded parts is refined by using the information of other parts that are connected in the human body skeleton model. Although shifted Gaussian kernels (in pairwise terms of the proposed energy function) add huge computational cost to the inference process, the problem is solved by proposing an efficient mean field approximation algorithm that speeds up message-passing steps, despite the fact that kernels are shifted.

For demonstrating the effectiveness of the proposed fully connected model in comparison with the state-of-the-art pose estimation methods, the probability maps of the DeeperCut method are used in the training phase and it is shown that results improve significantly in KTH Football I, LSP, HumanEva I in comparison with the original DeeperCut method. Also it is shown that the BodyField method has substantial improvement in HB segmentation in Penn-Fudan dataset in per pixel segmentation measure.

The experimental results on the challenging KTH Football I, Extended Leeds Sports Pose, HumanEva I, and Penn-Fudan datasets show the superiority of the proposed method over other existing methods in terms of PCP, PCK and per pixel segmentation accuracy.

References

[1] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in European Conference on Computer Vision, 2016: Springer, pp. 34-50.

[2] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang, "Multi-context attention for human pose estimation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1831-1840.

[3] A. Bulat and G. Tzimiropoulos, "Human pose estimation via convolutional part heatmap regression," in European Conference on Computer Vision, 2016: Springer, pp. 717-732.

[4] S. Ershadi-Nasab, S. Kasaei, and E. Sanaei, "Regression-based convolutional 3D pose estimation from single image," Electronics Letters, vol. 54, no. 5, pp. 292-293, 2018.

[5] S. E. Nasab, S. Kasaei, E. Sanaei, A. Ossia, and M. Mobini, "Multiview 3D reconstruction and human point cloud classification," in 2014 22nd Iranian Conference on Electrical Engineering (ICEE), 2014: IEEE, pp. 1119-1124.

[6] L. Wang, J. Shi, G. Song, and I.-f. Shen, "Object detection combining recognition and segmentation," in Asian conference on computer vision, 2007: Springer, pp. 189-199.

[7] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," International journal of computer vision, vol. 87, no. 1-2, p. 4, 2010.

[8] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in CVPR 2011, 2011: IEEE, pp. 1465-1472.

[9] V. Kazemi, M. Burenius, H. Azizpour, and J. Sullivan, "Multi-view body part recognition with random forests," in 2013 24th British Machine Vision Conference, BMVC 2013; Bristol; United Kingdom; 9 September 2013 through 13 September 2013, 2013: British Machine Vision Association.

[10] F. Xia, J. Zhu, P. Wang, and A. L. Yuille, "Pose-guided human parsing by an and/or graph using pose-context features," in Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[11] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in Advances in neural information processing systems, 2011, pp. 109-117.

[12] A. Adams, J. Baek, and M. A. Davis, "Fast high - dimensional filtering using the permutohedral lattice," in Computer Graphics Forum, 2010, vol. 29, no. 2: Wiley Online Library, pp. 753-762.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005.

[14] V. Vineet, G. Sheasby, J. Warrell, and P. H. Torr, "Posefield: An efficient mean-field based method for joint estimation of human pose, segmentation, and depth," in International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, 2013: Springer, pp. 180-194.

[15] M. Kiefel and P. V. Gehler, "Human pose estimation with fields of parts," in European Conference on Computer Vision, 2014: Springer, pp. 331-346.

[16] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4724-4732.

[17] Y. Bo and C. C. Fowlkes, "Shape-based pedestrian parsing," in CVPR 2011, 2011: IEEE, pp. 2265-2272.

[18] I. Rauschert and R. T. Collins, "A generative model for simultaneous estimation of human body shape and pixel-level segmentation," in European Conference on Computer Vision, 2012: Springer, pp. 704-717.

[19] P. Wang, X. Shen, Z. Lin, S. Cohen, B. Price, and A. L. Yuille, "Joint object and part segmentation using deep learned potentials," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1573-1581.

[20] P. Luo, X. Wang, and X. Tang, "Pedestrian parsing via deep decompositional network," in Proceedings of the IEEE international conference on computer vision, 2013, pp. 2648-2655.

[21] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834-848, 2017.

[22] R. Alp Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7297-7306.

[23] Y. Chen, C. Shen, X.-S. Wei, L. Liu, and J. Yang, "Adversarial posenet: A structure-aware convolutional network for human pose estimation," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1212-1221.

[24] X. Peng, Z. Tang, F. Yang, R. S. Feris, and D. Metaxas, "Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2226-2234.

[25] W. Yang, W. Ouyang, X. Wang, J. Ren, H. Li, and X. Wang, "3d human pose estimation in the wild by adversarial learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5255-5264.

[26] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded pyramid network for multi-person pose estimation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7103-7112.

[27] M. Andriluka et al., "Posetrack: A benchmark for human pose estimation and tracking," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5167-5176.

[28] J. M. Mooij, "libDAI: A free and open source C++ library for discrete approximate inference in graphical models," Journal of Machine Learning Research, vol. 11, no. Aug, pp. 2169-2173, 2010.

[29] A. Adams and J. Dolson, "ImageStack Library," https://github.com/abadams/ImageStack.

[30] J. Baek, A. Adams, and J. Dolson, "Lattice-based high-dimensional gaussian filtering and the permutohedral lattice," Journal of mathematical imaging and vision, vol. 46, no. 2, pp. 211-237, 2013.

[31] Y. Yang and D. Ramanan, "Articulated human detection with flexible mixtures of parts," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 12, pp. 2878-2890, 2012.

[32] V. Belagiannis, C. Amann, N. Navab, and S. Ilic, "Holistic human pose estimation with regression forests," in International Conference on Articulated Motion and Deformable Objects, 2014: Springer, pp. 20-30.

Sara Ershadi-Nasab was born in Mashhad, Iran, in 1986. She received the B.Sc. degree in Electrical Engineering from Ferdowsi University of Mashhad in 2008, her M.Sc. degree in 2010 from Amir Kabir University of Technology and her Ph.D. degree in 2017 at Sharif University of Technology. Her research interests are in 3D pose estimation, computer vision, and human activity recognition.

Shohreh Kasaei received the B.Sc. degree from the Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran, in 1986, the M.Sc. degree from the Department of Electrical and Electronics Engineering, University of the Ryukyus, Japan, in 1994, and the Ph.D. degree from Signal Processing Research Centre, School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia, in 1998. She joined Sharif University of Technology since 1999, where she is currently a full professor and the director of image processing laboratory (IPL). Her research interests include image and video processing as well as 3D computer vision with primary emphasis on 4D reconstruction and graphical element addition in dynamic sports scenes, human activity recognition, pose estimation, 4D object tracking, virtual reality, semantic scene understanding, 3D SLAM, image/video mosaicing, multi-resolution texture analysis, scalable video coding, image retrieval, video indexing, face recognition, hyperspectral change detection, video restoration, fingerprint authentication, and watermarking.

Esmaeil Sanaei received the B.Sc. degree in Electronics Engineering from the Department of Electrical Engineering, Amir Kabir University of Technology (Tehran Polytechnic), Iran, in 1979, the M.Sc. degree in Control Systems from the University of Technology of Compi`egne (Universit`e de Technologie de Compi`egne), Paris, France, in 1981, the M.Sc. degree in Information Technology Systems from the ` Ecole sup`erieure d'`electricit`e (Sup`elec), Paris, France, in 1982, and the Ph.D. degree in Information Technology Systems from University of Paris, Paris, France in 1984. He joined Sharif University of Technology since 1986, where he is currently an assistant professor. His research interests include computer vision and machine learning.

Erfan Noury was born in Urmia, Iran, in 1993. He received a B.Sc. degree at Computer department at Sharif University of Technology in 2017. His research interests include deep learning and computer vision.

Hassan Hafez-Kolahi was born in Mashhad, Iran, in 1989. He received the B.Sc. degree in Computer Engineering from Ferdowsi University of Mashhad in 2011, her M.Sc. degree in 2013 from Sharif University of Technology and he is currently a Ph.D. candidate at Sharif University of Technology. His research interests are in deep learning, computer vision.

Recognizing Transliterated English Words in Persian Texts

Ali Hoseinmardy
Computer Engineering Department, Amirkabir University of Technology, Iran
ali.hoseinmardy@aut.ac.ir
Saeedeh Momtazi*
Computer Engineering Department, Amirkabir University of Technology, Iran
momtazi@aut.ac.ir

Received: 28/Jan/2020 Revised: 12/Feb/2020 Accepted: 19/Apr/2020

Abstract

One of the most important problems of text processing systems is the word mismatch problem. This results in limited access to the required information in information retrieval. This problem occurs in analyzing textual data such as news, or low accuracy in text classification and clustering. In this case, if the text-processing engine does not use similar/related words in the same sense, it may not be able to guide you to the appropriate result.

Various statistical techniques have been proposed to bridge the vocabulary gap problem; e.g., if two words are used in similar contexts frequently, they have similar/related meanings. Synonym and similar words, however, are only one of the categories of related words that are expected to be captured by statistical approaches. Another category of related words is the pair of an original word in one language and its transliteration from another language. This kind of related words is common in non-English languages. In non-English texts, instead of using the original word from the target language, the writer may borrow the English word and only transliterate it to the target language. Since this kind of writing style is used in limited texts, the frequency of transliterated words is not as high as original words. As a result, available corpus-based techniques are not able to capture their concept. In this article, we propose two different approaches to overcome this problem: (1) using neural network-based transliteration, (2) using available tools that are used for machine translation/transliteration, such as Google Translate and Behnevis. Our experiments on a dataset, which is provided for this purpose, shows that the combination of the two approaches can detect English words with 89.39% accuracy.

Keywords: Transliteration; Text processing; Words Relation; Neural Network-Based Sequence2Sequence Model; Google Translate; Behnevis

1- Introduction

Searching textual information on the Web has become one of the main usages of the Internet. People use the Internet to find the information they need. For this reason, the intelligence of language processing tools can be much helpful in interacting with computers.

One of the challenges encountered in text processing systems is recognizing related words in a language. Different models have been proposed to address this issue, most notable methods are based on dictionary and statistical co-occurrence of words.

The available methods, however, have not been able to produce good results for new words entered into a language. Transliterated words from dominated languages such as English to other languages are examples of new words in the target languages. Transliterated words are words that are entered in the target language with their vocals. These words are written in the target language as they are pronounced in the source language.

The words "دیپ" (the transliteration of the word "deep" in Persian), "تیبل" (the transliteration of the word "table" in Persian) are examples of this phenomenon in non-English languages which should be captured as related words to "عمیق" and "میز", respectively, which are the correct translations.

The goal of this paper is detecting this type of words in a text and finding a relation between a word in the target language and a transliterated word from another language; e.g., the words "دیپ" and "عمیق" in the above example in Persian. To the best knowledge of the authors, this issue has not been studied in Persian and this is the first attempt in this direction.

The structure of this paper is as follows: In Section 2, we describe word relation, Section 3 describes the transliteration in the Persian language and its challenges. In Section 4, we show the proposed model and the solutions which include 2 different techniques: the seq2seq

* Corresponding Author

model and the tool-based technique. Section 5 represents the evaluation dataset and the results. Finally, Section 6 concludes the paper.

2- Word Relationship

Various methods have been proposed to overcome the word miss-match problem. Dictionary-based and distributional methods are the main techniques used for this goal.

The dictionary-based methods are very simple and its implementation is very easy. The main problem of this approach is the fact that the dictionary vocabulary is constant; i.e., new words in a language, do not exist dictionaries, since updating dictionaries is very costly and time-consuming. Another problem with this method is that it does not consider foreign words. Words transliterated from another language to Persian, which is the target of this research, are not included in the Persian dictionaries.

In the distributional methods, the contexts of the words are used to identify their concept. The advantage of this method compared to the former one is the possibility of adding new vocabulary items at any time by using texts published daily on the Internet. For example, in the Persian language, if we search for the phrase "نرخ ماشین" ("the price of the car"), we expect to receive quite similar results compare to the query "قیمت اتومبیل", which has the same meaning.

One of the problems of this approach is requiring a large amount of data and context for each word. Therefore, although these methods have achieved desirable results in different languages including Persian [7], their success is limited to the words used frequently in those languages.

By the rapid growth of the WEB2 content and the huge amount of information that ordinary people enter into the Internet via social networks, blogs, etc., transliterated foreign words are added daily in non-English texts.

In some cases of transliterated words, as they are used very common in the target language, they can be detected by distributional methods due to their high frequency; e.g., the word "تست" (transliteration of the word "test" from English), which is used by everyone, became a formal translation of the word as well. Therefore, its concept and its similarity to the word "آزمون" can be detected with the statistical methods. For most of the transliterated words, however, it is almost impossible to identify them using these two methods, due to the lower frequency they have.

For example, the word "دیپ" (the transliteration of the word "deep"), which is used in Persian text, should be recognized as a related word to "عمیق", but since this word is not available in the Persian dictionary and its frequency is not high in Persian, we could not find such relation. As a result, we cannot expect similar results when searching for the words "دیپ" and "عمیق". Such a problem motivated us to propose a new solution to overcome this problem.

3- Transliteration

Transliteration means a word with fixed vocals (sound and pronunciation) transfers from one language to another, and each letter is transmitted in the same way as another language. For example, the word "کامپیوتر" is the transliterated word of "computer" in Persian [8].

Transliteration is almost obvious for people who know the both languages. This makes transliteration a personalized task; therefore, it is different from transcription. Because transcription maps the sound of one language to another. There are lots of difficulties in transliteration, like "ق" in Arabic and Persian. Because in English it's not "k" or "g" and maybe something between these two [8].

In the transliteration task between two languages, reconstructing the original form of writing a word in the source language from the transliterated form with no ambiguity is an important issue. Therefore, a character/letter does not need to be always translatable. Another thing is conceivability when several different characters are spelled alike. In this case, there is usually a rule in the original language that decides which characters should be converted when the word is spelled. For instance, if we consider "ک" and "ک" in the Persian language as two different characters, the spelled-out form of the word "کک" in English is "kk". Then, to reconstruct the word to the original language from "kk", it should be known that "k" is written "ک" in the Persian language if it comes first, and if it comes at the end of a character that adheres to the next letter, "k" is written "ک". Therefore, with the basic knowledge of Persian script, we can still reconstruct the original form of the word.

One important issue in transliteration is that in some cases, people use different types of transliteration for a foreign word; e.g., the English word "balcony", is transliterated to "بالکن" and also "بالاکن" in the Persian language.[1]

To summarize, the main limitations of transliteration are as follows:

- There may be more than one transliteration for a word. People may pronounce or write an English word in different ways in Persian, Like "Mouse" in English which can be written like "ماوس" and "موس"
- There might be some words that are written in the same way in another language. Like "Test" & "Toast" in English which are written "تست" in Persian.

[1] according to the transliteration style sheet, the second word is wrong, and the first word should be used.

Table 1: Constants of Persian language according to Transliteration style sheet presented by Dr. Mo'in

Constants																						
English Alphabets	B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	š	T	V	X	Z	ž	'
Persian Alphabets	ب	چ	د	ف	گ	ح،ه	ج	ک	ل	م	ن	پ	غ،ق	ر	ث،س،ص	ش	ت،ط	و	خ	ذ،ز،ض،ظ	ژ	ع،ء

3-1- Transliteration style Sheet

This transliteration style sheet was first used by Dr. Mohammad Mo'in in its famous Encyclopedia in 1971[14] to represent the correct pronunciation of words. Later in 2012, the same model, entitled "General Transcription Style Sheet for Geographical Names of Iran", was sent to all ministries and state. This standard is accepted by the United Nations. [21]

Table 2: Vowels of Persian language according to Transliteration style sheet presented by Dr. Mo'in

Vowels								
Persian Alphabets	اَ	اِ	اُ	آ	ئی	ئو	ع،ء	اَوْ
English Alphabets	a	e	o	ā	i	u	ë'/	ow
Persian Example	ابر	ارم	اردک	آسمان	ایران	بوستان	مسئول	فردوسی
Equivalents	Abr	Eram	Ordak	Āsemān	Irān	Bustān	Mas'ul	Ferdowsi

As can be seen in Tables 1 and 2, the transliteration of Persian to English and the transliteration of English to Persian is based on phonemes and letters. Transliteration is usually used in proper nouns and named entities, such as the particular name of persons, organizations, locations, or events; e.g., "Jonathan", "Apple", or "Nowruz". Ghayoomi et al. [6] and Bijankhan et. al. [2] provided a comprehensive study on challenges related to the Persian transliteration style.

3-2- Transliteration Challenges

Transliteration has different challenges that we describe in this section. Different dialects in languages: there are several dialects in the languages; e.g., the way American people speak is different from the Canadian people, but both of them speak English.
Different transliteration for a word: Due to phonemes and pronunciations in different languages, a word in one language (for example, the word "اسلام" in Persian) may have several English transliterations ("Islam" or "Eslam"). In this state, we must refer to the original word chosen for the meaning of "اسلام" in English ("Islam"), and count it as the correct word.
Words with letters that are not pronounced: some words have letters that we do not pronounce them; e.g., "خواهر" and "خورشید" in Persian or "knife" and "knee" in English.
According to the method used in this project, they will be implicitly solved when processing them in our proposed methods.

4- The Proposed Model

This research aims to provide a tool which receives a Persian text as an input and detects foreign (transliterated) words in the text, Furthermore, it finds the corresponding Persian meaning. For this purpose, the project can be implemented in two steps: (1) detecting if a word is an original word or a transliterated word, and (2) finding the equivalent of the foreign word in Persian; i.e., the Persian meaning of that word.
To this aim, we propose two different approaches which will be described in the next sections:

- Using neural network-based transliteration
- Using available tools, such as Google Translate, and Behnevis.

Figure 1 shows the overall architecture of our proposed model. The detailed description of the components is presented in the rest of this section.

4-1- Transliteration with Deep Sequence2sequence Model

A deep neural network is an artificial neural network with multiple layers between the input and output layers. Deep neural networks are powerful models with excellent performance in learning tasks [18]. Sequence2Sequence models are one of the well-known neural network-based models that are used for processing sequential data. In this model, a sequence is given to an autoencoder/decoder as an input and the system returns a sequence as an output. Both encoder and decoder parts of the model consist of Long Short Term Memory (LSTM) units, such that each unit supports one term in the input/out text.
The structure of a sequence2sequence model for machine translation is represented in Figure 2. The

* Corresponding Author

sequence2sequnce model in transliteration is similar to the translation model, but the units are characters instead of the words.

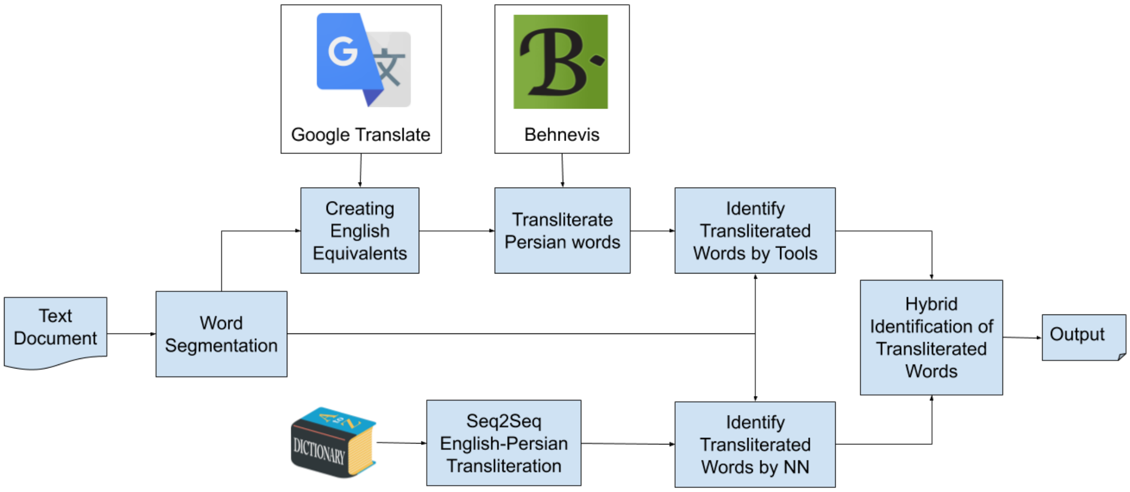


Fig. 1 Overall architecture of the proposed model

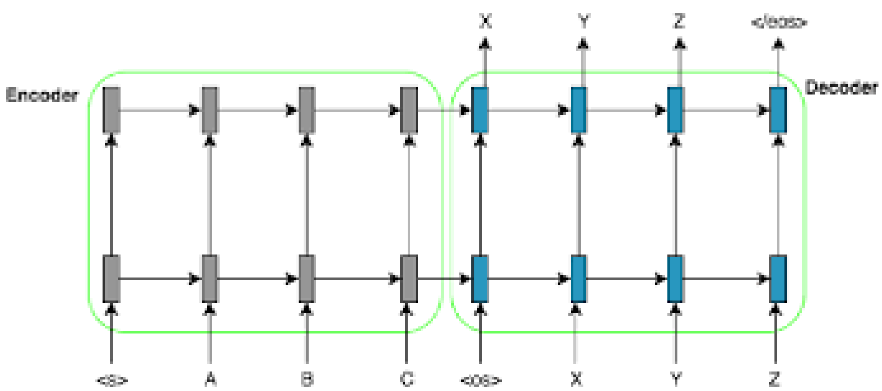


Fig. 2 A sample architecture for sequnce2sequence neural transliteration [9]

This method has been widely used in different natural language processing tasks, such as machine translation [3,1], question answering [11,15], dialog systems [23], and speech recognition [4]. In machine translation, an input sentence, as a sequence of words, is given to the system, and a sentence, again as a sequence of words, is generated in the output. The transliteration task has also the same behavior. It receives an input word, as a sequence of characters, and returns an output word, again as a sequence of characters [20].

To implement a sequence-to-sequence transliteration model, we used a 3-layer LSTM network as encoder, hidden, and decoder layers with Adam optimizer and learning rate 0.001. Each letter is represented by a 200-dimension embedding vector. The vectors are generated randomly and the representations are learned during the training phase. The implementation is done within the Tensorflow framework in Python. The model is trained on the English-Persian (EnPe) part of the transliteration shared task at Named Entities Workshop (NEWS) [24]. The dataset contains 14,758 training samples. Each sample consists of a pair of English and transliterated Persian words.

The trained model is then used to transliterate all words in the Persian input text into English. The transliteration is then compared to the English vocabulary. In case of any matching between the transliterated words and the English words in the dictionary, the word is detected as foreign word and the translation of the corresponding word in English-Persian bilingual dictionary is returns as the related word. The English vocabulary is selected based on the top 35k high frequent English terms from the Wall Street Journal Penn Treebank [12], and the Brown Corpus [19].

To be more precise, the proposed model captures the relationship between English and Persian words in a triangle as represented in Fig 3. The double line means that the main purpose of this model is to find the transliterated English word and its equivalent in Persian.

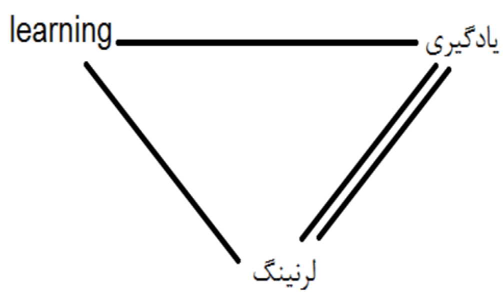


Fig. 3 a desired sample of the relationship between an original Englsih word (learning), its transliteration (لرنینگ), and the equivalent Persian word (یادگیری), which indicates the final goal of the research.: We want to find the transliterated English words in the Persian Language and its Persian equivalent. For example, "learning" is an English word and "لرنینگ" is its transliterated word in Persian. "یادگیری" is the Persian equivalent of "learning".

4-2- Tool-Based Transliteration

The tool-based approach tries to capture transliterated words by utilizing available toolkits that are compatible with this task. To this end, the tool-based approach is designed as follows such that the output of the first step is the input to the second step:

The first step is to find the English equivalent of the input words using *Google translate*. Google Translate is a free multilingual translation program developed by Google. It offers an API that helps developers build software applications. The tool is widely used in different researchers for question retrieval [17] as well as other text applications [22] and it is one of the best translators that can be used by programmers and researchers.

In this section, we translate each Persian word to English using the Google translate API. As it is known, Persian words are translated and foreign words are transliterated because these words come from foreign languages. Therefore, it is not possible to distinguish foreign words from original words; i.e., we cannot automatically find out if the output of the Google translate API is the translation of the Persian word to English, or it is the original English word that was transliterated to Persian. This question will be solved in the next steps.

The second step is to find the Persian equivalent of the English words using *Behnevis*. Behnevis is a website for transliterating English words into Persian. Most of the usage of this site is in the transliteration of Persian texts written in English. Here we use it to distinguish foreign words. In this way, if the word is given as an input to the Behnevis, the word is transliterated in Persian. For example, the word "بهینه", turns into "Optimum" in the first step and the second step becomes "اپتیموم". As another example, the foreign word "آمبولانس" becomes "Ambulance" in the first step and the second step again becomes "ambulance". As can be seen in these examples, the detection of transliterated words is based on the output word of Behnevis. In this way, if the word is an original Persian word, after translating it with Google Translate, the English translation will be returned. Google Translate passes the translated word to Behnevis, the word is spelled out and varies with the original input word. But if the word is foreign, Google Translate essentially transliterates the word to English and then Behnevis transliterates it back to Persian in the next step and returns the original word. Therefore, we need to compare the input word with the final word. If they are the same, the word is foreign and if they differ, it is a Persian word. Also, in this step, we can consider Google translate output as the equivalent of foreign words in Persian and English and reach the related word to Persian.

Although this method captures a large number of foreign words in texts, it still suffers from a shortcoming. Since some words have different pronunciations, they are transliterated to different forms when they come from foreign languages to Persian. Therefore, it is difficult to capture them with either of the proposed approaches. For example, by applying the above steps to the word "اورجینال" (the transliteration of the word "Original"), it becomes "اوریجینال" in which one character is added to it. We may also have such changes in consonants, such as "هندزفری" (the transliteration of the word "Hands-free"), which becomes "هندسفری" which replaces one letter. In such cases, although the transliteration is correct, the system cannot detect it, due to the difference between the final output and the input word. To solve this issue, we need to relax the exact matching between the terms and accept the matching between terms that differ in one character. To this aim, we use Levenshtein distance [10] algorithm to compare two strings and find the possibility of skipping one character when matching two words.

Levenshtein distance is used to calculate the difference between two strings. The output of the comparison of the two strings is the minimum number of changes needed to convert a string to another string. We have 3 actions, namely replacing a letter with another one, removing a letter, and adding a letter to the word. This algorithm works better than the Hamming algorithm because it calculates the distance between two words regardless of their lengths, whereas in the Hamming distance algorithm, the size of the 2 strings must be equal, and it only considers the replacement of characters [13].
Levenshtein distance uses dynamic programming, and its time complexity is O (m*n), where n and m are the lengths of two words.

5- Results

5-1- Dataset

To evaluate the performance of the proposed foreign word detection model, we provided a list of 979 words. The list consists of 491 Persian words and 488 foreign words. Persian words are those that have origin in the Persian language[1]. Foreign words are English words that are transliterated from English to Persian; i.e., all words are written with Persian alphabets.
It should be mentioned that in the input texts, there is no complete sentence and there is no need to word context information because the processing of each word is regardless of the previous and next word. Therefore, the proposed model can also be applied to single terms.

5-2- Evaluation Metrics

To evaluate our proposed model, we used 3 evaluation metrics, namely precision, recall, and F-measure. These metrics are calculated based on the confusion matrix of the results, as presented in Table 3.

Table 3: Confusion matrix

		Actual Values: English	Actual Values: Persian
Predicted Values	English	True Positive	False Positive
	Persian	False Negative	True Negative

[1]There are Arabic words entered in Persian, however, since the characters are not identical with Persian, we can consider them

In the confusion matrix, true positive means the word is originally an English word and the system correctly identified as a transliterated word (English), true negative means the word is Persian and identified as Persian, false positive means the word is Persian but identified as English and false negative means the word is English but identified as Persian.
Considering the above descriptions, the metrics are calculated as follows:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{True negative}} \quad (1)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad (2)$$

$$\text{F-Measure} = \frac{\text{Precision . Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

5-3- Results of the Sequence2sequence Model

The first experiment was performed using sequence2sequence transliteration on the input text. The confusion matrix and the results of the model can be seen in Tables 3 and 4, respectively.

Table 4: Confusion matrix of output produced by the neural network method: There are 488 English words and 491 Persian words for testing.

	English	Persian
English	109 (22.33%)	14 (2.85%)
Persian	379 (77.67%)	477 (97.15%)

Table 5: Results of the foreign word prediction method using the neural network method

	English	Persian	Average
Precision	0.8862	0.5572	0.7217
Recall	0.2234	0.9715	0.5974
F-Measure	0.3568	0.7082	0.6537

As you can be seen in Table 3, the system has intention toward detecting words as Persian words rather than transliterated words. This results in high Precision in detecting English words and high recall in detecting Persian words. The overall f-measure of the system is 65.37%, which shows that the system has a correct prediction on about two-thirds of the words in the input list.
The neural network-based model is relatively fast in transliterating one word, but it needs to compare the transliterated word with all English vocabulary words

5-4- Results of the Tool-Based Model

A similar experiment has been done on the tool-based approach and the results are presented in Tables 5 and 6.

Table 6: Confusion matrix of output produced by the tool-based method: There are 488 English words and 491 Persian words for testing.

	English	Persian
English	275 (56.35%)	3 (0.61%)
Persian	213 (43.65%)	488 (99.39%)

Table 7: Results of the foreign word prediction method using the tool-based approach

	English	Persian	Average
Precision	0.9892	0.6961	0.8427
Recall	0.5635	0.9939	0.7787
F-Measure	0.7180	0.8188	0.8094

As can be seen in the tabulated results, although this method is slower than the neural network-based approach, its accuracy is very high and it detects about 80% of foreign words. Using this approach, we can see the following results for a sample English and Persian word respectively:

- "اسنک" ------> "snack" ------> "اسنک"
  prediction: English word
- "زبان" ------> "language" ------> "لنگویج"
  prediction: Persian word

The output of the tool-based approach shows that the second step of the model can clearly distinguish the words transliterated by Google from those that are translated.
Moreover, among not detected foreign words, we can see items with only 1 difference, so the recognition of foreign words even with minor differences can help to increase the accuracy of the system. The word "باینری" is an example of such words.

- "باینری" ------> "Binary" ------>" بینری"
  prediction: Persian word

Although these words are borrowed from English, due to differences between the source and target word, it cannot be detected as a foreign word with the proposed tool-based approach. Therefore, relaxing the exact match assumption and accepting one-character difference will help us to detect these words as well.
To this aim, in the next step of our experiment, we performed the tool-based approach while accepting 1 distance based on Levenshtein. The results of these experiments are presented in Tables 7 and 8.

Table 8: Confusion matrix of output produced by the tool-based method with accepting 1 Levenshtein distance: There are 488 English words and 491 Persian words for testing.

	English	Persian
English	364 (74.59%)	4 (0.81%)
Persian	124 (25.41%)	487 (99.19%)

Table 9: Results of the foreign word prediction method using the tool-based approach with accepting 1 Levenshtein distance

	English	Persian	Average
Precision	0.9891	0.7971	0.8931
Recall	0.7459	0.9919	0.8689
F-Measure	0.8505	0.8838	0.8808

5-5- Using the Joint Neural Network and Tool-Based Method

We can see in the results that the tool-based method outperforms the neural network-based method. But there is still room to improve the system by combining both models. To this end, we try to use the neural network-based transliteration besides the tool-based method. The results of the combined model that is applied to the same input text are presented in Tables 9 and 10.

Table 10: Confusion matrix of output produced by the joint method: There are 488 English words and 491 Persian words for testing.

	English	Persian
English	398 (81.55%)	18 (3.66%)
Persian	90 (18.45%)	473 (96.34%)

Table 11: Results of the foreign word prediction method using the joint approach

	English	Persian	Average
Precision	0.9567	0.8401	0.8984
Recall	0.8156	0.9633	0.8895
F-Measure	0.8805	0.8975	0.8939

As expected, comparing to the tool-based approach, the output improved and the F-measure of detecting English and Persian words increases 3% and 1%, respectively.
Given the fact that very few Persian words have been detected as foreign words; we hope that this method is working and we have greatly remedied the concern.

5-6- Error Analysis

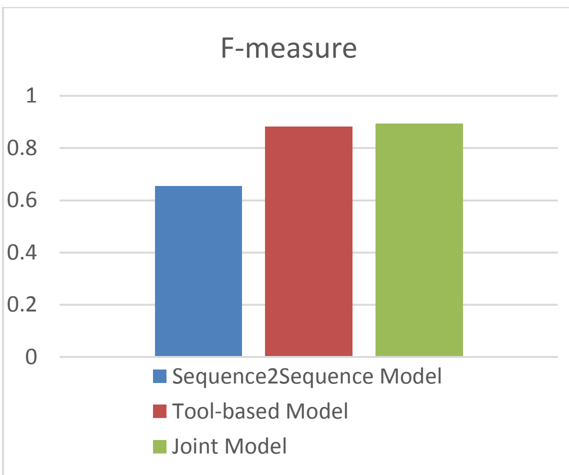
The error analysis of the words that cannot be recognized yet shows difficult items, such as, the word "دیلیت" (Delete). Although the proposed model achieved promising results in the task, there are some words that cannot be detected by our approach. To have a better understanding about our model. We had an error analysis and found the following issues the main reasons of the 11% error:

- For some words, the Google translate provides wrong translation instead of a correct transliteration. For example: the word "پرشین" is translated to "Jump" by Google translate. The NN output for this word is "perchin" which is wrong, too.
- The variation of letters in transliteration is more than one word and cannot be captured by our levenshtein distance approach. For example, the

word “آداپتور” which is “adapter” in English is transliterated to “آدپتر” by Google translate, which is not completely wrong. Since in both languages the “a” vowel and the “ā̲” vowel are similar to each other, there can be more than one transliteration for a word and this makes the task difficult.

Overall, by using the combined model, our proposed system can correctly identify 89.39% of the words. For a better overall comparison, the average F-measure of all the models, including the hybrid model are presented in Figure 4.

Fig. 4 F-measures of Models respectively: Seq2Seq model, Tool-based model, Joint model



6- Conclusion and Future Work

Using the transliteration of foreign words instead of original Persian words becomes a common issue in recent years by the advent of Web2. This results in different problems when capturing the semantic meaning of a text. In this paper, we addressed this problem by proposing two different approaches and their combination. We showed that our proposed model, which benefits from a neural sequence2sequence model and a tool-based approach using Google translate and Behnevis APIs can detect foreign words and their corresponding Persian words with the F-measure of 89.39%.

As mentioned, although distributional approaches show great success in various tasks, their power comes from the data and the frequency of words in training corpora. The transliterated words, however, suffer from low frequency in training corpora and this is their main weakness to be used for the present work. In the future, we plan to use these approaches and try to adapt the models to be used for low-frequency words as well. We first try to discover the usage of ontology for the task [16] and then study more advanced contextualized techniques, such as BERT [5].

References

[1] P. Bahar, C. Brix, H. Ney. “Towards Two-Dimensional Sequence to Sequence Model in Neural Machine Translation.” Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, p. 3009–3015.

[2] M. Bijankhan, J. Sheykhzadegan, M. Bahrani, M. Ghayoomi. “Lessons from building a Persian written corpus: Peykare”, Language Resources and Evaluation, Vol. 45, No. 2, 2011, pp. 143-164.

[3] M.X. Chen, O. Firat., A. Bapna, M. Johnson, W. Macherey, G. Foster, L. Jones, N. Parmar, M. Schuster, Z. Chen, Y. Wu, M. Hughes. “The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation.” arXiv:1804.09849, 2018.

[4] C.C. Chiu, T.N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R.J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, M. Bacchiani, “State-of-the-Art Speech Recognition with Sequence-to-Sequence Models.”, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 4774-4778.

[5] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, In Proceedings of the NAACL-HLT Conference, Minneapolis, Minnesota, 2019, pp. 4171 - 4186.

[6] M. Ghayoomi, S. Momtazi, M. Bijankhan. “A study of corpus development for Persian.” International Journal on Asian Language Processing, 2010.

[7] A. Hadifar, S. Momtazi, “The impact of corpus domain on word representation: a study on Persian word embeddings”, Journal of Language Resources and Evaluation, Vol. 52, No. 4, 2018, pp. 997-1019.

[8] N. S. Kharusi and A. Salman, “The English Transliteration of Place Names in Oman”. Journal of Academic and Applied Studies Vol. 1, No. 3, 2011, pp. 1–27.

[9] N.T. Le, F. Sadat, L. Menard, and D. Dinh, “Low-Resource Machine Transliteration Using Recurrent Neural Networks”. ACM Trans. Asian Low-Resour. Lang. Inf. Process. Vol. 18, No. 2, 2019.

[10] *V.I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals”. Soviet Physics Doklady. Vol. 10, No. 8, 1966, pp. 707–710.*

[11] B. Li, "A Question Answering System Using Encoder-Decoder Sequence-to-Sequence Recurrent Neural Networks.", Master Thesis, The Faculty of the Department of Computer Science, San José’s State University, 2018.

[12] M. Marcus, B. Santorini, M.A. Marcinkiewicz. “Building a Large Annotated Corpus of English: The Penn Treebank”, Journal of Computational Linguistics, Vo. 19, N. 2, 1993, pp. 313-330.

[13] F.P. Miller, A.F. Vandome, and J. McBrewster, Levenshtein Distance: Information Theory, Computer Science, String (Computer Science), String Metric, Damerau? Levenshtein Distance, Spell Checker, Hamming Distance, 2009, Alpha Press.

[14] M. Mo'in, Mo'in Encyclopedic Dictionary, Amirkabir Publisher, 1972.
[15] A. Otsuka, K. Nishida, K. Bessho, H. Asano, and J. Tomita. "Query Expansion with Neural Question-to-Answer Translation for FAQ-based Question Answering". In Proceedings of the Web Conference (WWW), 2018. pp. 1063-1068.
[16] V. Reshadat, M.R.F. Derakhshi, "Studying of Semantic Similarity Methods in Ontology". Research Journal of Applied Sciences, Engineering and Technology, Vol. 4, No. 12, 2012, pp. 1815-1821.
[17] A. Rücklé, K. Swarnkar, I. Gurevych. "Improved Cross-Lingual Question Retrieval for Community Question Answering.", In Proceedings of the Web Conference (WWW), 2019, pp. 3179-3186.
[18] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview". Neural Networks. Vol. 61, 2015, 85–117.
[19] L. Schubert and M. Tong. "Extracting and evaluating general world knowledge from the Brown corpus". In Proceedings of the HLT-NAACL workshop on Text meaning, Association for Computational Linguistics, Vol. 9, 2003, pp. 7-13.
[20] I. Sutskever, O.Vinyals, Q.V. Le. "Sequence to Sequence Learning with Neural Networks." In Proceedings Neural Information Processing Systems (NIPS), 2014.
[21] Terminology Department, "A collection of terms approved by the Academy of Persian Language and Literature". Vol. 6, 2009, Academy of Persian Language and Literature. Tehran. (ISBN 978-964-7531-85-6)
[22] E.D. Vries, M. Schoonvelde, and G. Schumacher, "No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications". Political Analysis, Vol. 26, No. 4, 2018, pp. 417-430.
[23] B. Wei, S. Lu, L. Mou, H. Zhou, P. Poupart, G. Li, Z. Jin, "Why Do Neural Dialog Systems Generate Short and Meaningless Replies? a Comparison between Dialog and Translation", In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019. pp. 7290-7294.
[24] M. Zhang, H. Li, A. Kumaran, M. Liu, "Report of NEWS 2011 Machine Transliteration Shared Task". In Proceedings of the 5th international Joint Conference on Natural Language Processing (lJCNLP), 2011.

Saeedeh Momtazi is currently an assistant professor at the Amirkabir University of Technology, Iran. She completed her BSc and MSc education at the Sharif University of Technology, Iran. She received a Ph.D. degree in Artificial Intelligence from Saarland University, Germany. As part of her Ph.D., she was a visiting researcher at the Center of Language and Speech Processing at Johns Hopkins University, US. After finishing the Ph.D., she worked at the Hasso-Plattner Institute (HPI) at Potsdam University, Germany and the German Institute for International Educational Research (DIPF), Germany as a postdoctoral researcher. Natural language processing is her main research focus. She has worked in this area of research for more than 15 years.

Ali Hoseinmardy received the B.S. degree in Computer Engineering from Amirkabir University of Technology, Tehran, Iran in 2018. Currently he Is a MBA graduate student in Sharif University of Technology, Tehran, Iran. His research interests Natural Language Processing, Data mining, Organizational Leadership and Gamification.

A Two-Stage Multi-Objective Enhancement for Fused Magnetic Resonance Image and Computed Tomography Brain Images

Leena Chandrashekar*
Research Scholar, R V College of Engineering, Visvesvaraya Technological University, India
Leenamaheshnikam10@gmail.com
Sreedevi A
Associate Professor, R V College of Engineering, Visvesvaraya Technological University, India
sreedevia@rvce.edu.in

Received: 04/Oct/2019 Revised: 24/Jan/2020 Accepted: 08/Feb/2020

Abstract

Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are the imaging techniques for detection of Glioblastoma. However, a single imaging modality is never adequate to validate the presence of the tumor. Moreover, each of the imaging techniques represents a different characteristic of the brain. Therefore, experts have to analyze each of the images independently. This requires more expertise by doctors and delays the detection and diagnosis time. Multimodal Image Fusion is a process of generating image of high visual quality, by fusing different images. However, it introduces blocking effect, noise and artifacts in the fused image. Most of the enhancement techniques deal with contrast enhancement, however enhancing the image quality in terms of edges, entropy, peak signal to noise ratio is also significant. Contrast Limited Adaptive Histogram Equalization (CLAHE) is a widely used enhancement technique. The major drawback of the technique is that it only enhances the pixel intensities and also requires selection of operational parameters like clip limit, block size and distribution function. Particle Swarm Optimization (PSO) is an optimization technique used to choose the CLAHE parameters, based on a multi objective fitness function representing entropy and edge information of the image. The proposed technique provides improvement in visual quality of the Laplacian Pyramid fused MRI and CT images.

Keywords - Glioblastoma; Laplacian Pyramid; Image Fusion; Image Enhancement; Contrast Limited Adaptive Histogram Equalization; Particle Swarm Optimization

1- Introduction

Glioblastoma is the fastest growing Grade IV malignant tumors found in the brain with a survival time of less than a year after their detection [1]. The detection of these tumors has always been a challenge to doctors. Some of the reasons for delay in detection are due to failure to understand the early symptoms, lack of awareness, inadequate healthcare facilities like imaging, preliminary screening for the patients and expertise with doctors. An inability to distinguish tumors by doctors and experts is mostly due to incorrect imaging procedures, patient's condition at the time of image acquisition and noise may lead to delayed prognosis. For this reason, medical imaging is paramount in detection, identification, grading and diagnosis of the Glioblastoma. Doctors recommend many imaging techniques for detection of Glioblastoma like Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and its variants, Fluid Attenuated Inverse Recovery (FLAIR) and Positron Emission Tomography (PET) [2]. These images are acquired sequentially through different scanning machines at different times. Each of the modalities provide different information of the brain. For example, the CT image provides the structural information of the brain like bone structure, tissue symmetries, changes in tissue density and space occupying lesions [3]. It also shows changes made in the nearby skull region due to tumor extension and calcification of tumors. Conversely, CT images fail to indicate tumor borders and infiltration in the nearby regions. These can be easily visualized with MR (Magnetic Resonance) images, which provide structural and functional information of the brain along with high contrast and resolution for soft tissues like tumors or lesions [3].

FLAIR images are a special type of MRI, which is sensitive to changes at the periphery of the cerebral hemispheres. PET images assess the tumor growth and spread [4]. These multimodal images are the noninvasive ways to detect Glioblastoma. Based on the multimodal images, surgical resection or complete removal of tumor is made followed by radio chemotherapy [4]. Thus, a single imaging technique is never sufficient to confirm the presence or extent of the tumor [5]. The latest development in medical imaging is generation of a complete image, obtained by fusing the multimodal images to a single

* Corresponding Author

image [6]. This helps in early detection of tumors and requires lesser digital storage [7-8]. Some of the techniques for fusing the multimodal images are Discrete Wavelet Transform (DWT), Laplace Transform (LT), Contourlet Transform (CT) and Non-sub Sampled Contourlet transform (NSCT) [9]. The process of fusion begins with decomposing the multimodal images in to approximation coefficients (low frequency components of the image) and detailed coefficients (high frequency components) using the above techniques. These components are combined using different rules like averaging, summing, weighted summing, max-min or max-max fusion rules [10]. The fused coefficients are recomposed using inverse transforms to generate a fused image. Fig.1 shows the methodology of the fusion process using DWT [10]

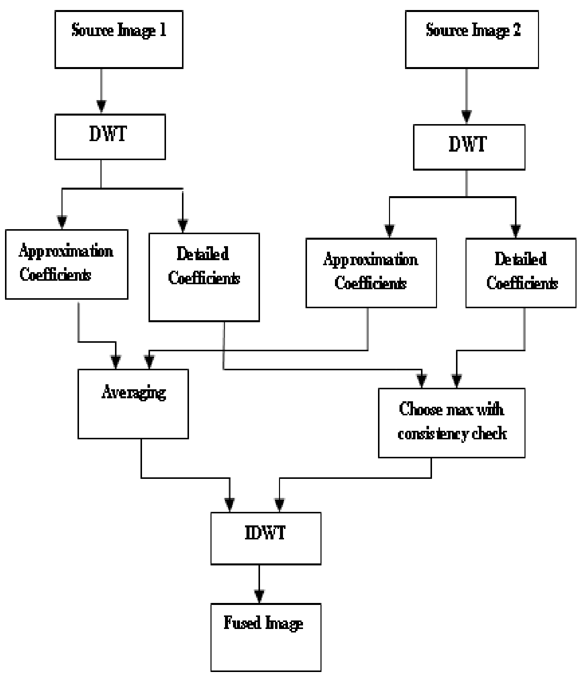
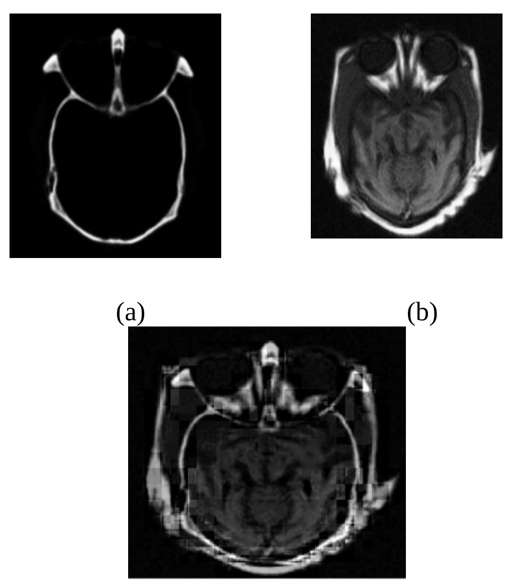


Fig.1.Methodology for Multimodal Image Fusion

The visual quality of the fused image is evaluated based on contrast, edge information and peak signal to noise ratio. Contrast is measured in terms of standard deviation of the image and is significant in fused images, because it helps to distinguish the healthy and tumorous cells. The edge information in the image represents the tumor boundary and also assists in locating the tumor. Noise exists at the time of image acquisition due to various reasons like patient position, scanner machines or in processing the MRI or CT machine. Peak signal to noise ratio indicates the signal to noise ratio in the fused image. The fusion process introduces blocking effect, noise and artifacts that greatly reduce the visual quality of the fused image. Fig.2 shows a fused CT and MR image obtained by DWT indicating blocking effect at the edges and corners [10]. The fused images with poor visual quality make it difficult for experts to interpret the tumor presence or its spread. Hence, there is a need for an enhancement technique to improve the visual quality and reduce blocking effect, noise and artifacts. Most of the enhancement techniques deal with improving image contrast, as it helps in differentiating the Glioblastoma and the normal cells. Nevertheless, the drawback with these techniques is that it only increases the dynamic range of the image, which is a function of pixel intensity alone. Tumors or any abnormality in the brain appears distinct, bright or light intensity in MR images. Thus, by varying the contrast, the normal and the tumorous cells can be differentiated easily. On the contrary, the high contrast images or low contrast MR image makes it challenging to differentiate the Glioblastoma [6]. The contrast enhancement is mostly a 2-fold process, consisting of contrast stretch and tonal enhancement. The contrast stretch improves the brightness differences uniformly across the dynamic range of the image and tonal enhancement improve the brightness differences in different areas like dark, gray or bright regions in the image [6]. The paper deals with enhancement techniques for fused images – contrast enhancement, improvement in entropy and structural information.

2- Enhancement Techniques

Generally, the contrast enhancement is achieved by the following techniques – Non-Linear Transfer function, Histogram based and Frequency Domain [11]. Among them, Histogram Equalization (HE) is the most popular



(a) (b)

(c)

Fig.2 (a) CT image (b) MR image (c) Fused image technique for contrast enhancement performed in a spatial domain. This deals with remapping the gray scale values of the original input image to a new level of gray scale values. The transformation is made by a simple scaling of values or by use of linear or non-linear functions. Consequently, HE attains the contrast enhancement by flattening and stretching the dynamic range of the image's histogram. The enhanced image shows improved characteristics like contrast, entropy, peak signal to noise ratio and edge information with respect to the original image. Although HE is a popular enhancement technique, it suffers from visual artifacts like intensity saturation and amplification due to large number of homogenous pixels. Moreover, the equalization is accomplished uniformly for all the pixels of the image, leading to enhanced global contrast. However, the lowest intensity pixels become less significant, thereby reducing the local contrast [12]. In order to overcome the drawbacks of HE, newer enhancement techniques are proposed for improved visual quality with the use of median filters, adaptive gamma correction and homomorphic filtering [13-14].

An Adaptive Histogram Equalization (AHE) technique is a block based adaptive method, which is capable of reducing the drawbacks of HE. It deals with the local contrast rather than global contrast. The local contrast is more significant in the detection of Glioblastoma. In this technique, histogram equalization is performed on sub-images or small and equal sized blocks obtained by splitting the image. The equalization is executed on every block independently and mapped to new intensity levels based on a transformation function. The new pixel values are solely based on the neighboring pixel characteristics. Then, bilinear interpolation is used to combine the blocks after equalization [14]. The major challenge with AHE is the selection of the block size and the transformation function, which are significant and greatly affect the quality of the enhanced image. However, AHE suffers from blocking effect, at the time of combining the blocks. In addition to blocking effect, over-amplification is also seen due to large homogenous regions of the image. Youlian Zhu et al have proposed an adaptive histogram equalization technique for CT images. A user defined parameter β, is suggested based on the gray level of the image. The entropy is used as an objective function to select the β adaptively [15].

A variant of AHE is the Contrast Limited Adaptive Histogram Equalization (CLAHE), proposed by K Zuierveld, is also a block-based contrast enhancement technique with focus on local contrast. Unlike AHE, CLAHE provides uniform equalization with clipping the excess portion of large peaks found after the histogram equalization, thereby avoiding over-amplification. The excess portion removed depends on a parameter called clip limit, which is a function of the dynamic range of the image and block size. CLAHE involves setting of three operational parameters – clip limit, block size and distribution function, which must be selected before performing the image enhancement to achieve good contrast images, free from noise and artifacts [16]. Various histogram based enhancement techniques are compared and analyzed, CLAHE is observed to perform better for MRI brain Images [17].

The simplest technique of setting the operational parameters for CLAHE is by trial and error. However, this is time consuming, may deviate from the actual values and changes for every image. Some of the techniques used to set the parameters are based on textureness of the image, maximum curvature of entropy, Least Mean Square (LMS) algorithm, multi-objective optimization technique and fuzzy rules. Despite various techniques for contrast enhancement, CLAHE seems to provide good local contrast, however it largely fails to enhance the pixels with low gray level intensity. Moreover, there is no standard for finding the optimal clip limit for a specific region of interest in medical images. Generally, clip limit is proportional to the multiple of average height of the histogram, where the multiplication factor is user-defined and varies for different images, which is the major drawback of CLAHE. Therefore, there is a need to choose clip limit adaptively for every block of image without any user intervention [23].

The clip limit is also a function of dynamic range of the grayscale image, block size and slope of transformation function. Initially, the clip limit and block size is chosen empirically and then obtain the optimal values based on statistical parameters like entropy, peak signal to noise ratio or edge information. Yet clip limit may change depending on the type of images. This makes the enhancement process very extensive and time consuming [23]. Moreover, inaccurate selection of clip limit can cause over-amplification in CLAHE. Bilateral Filter and Median filters are used to overcome this drawback [24-25].

Optimization techniques provide a convenient way of determining the CLAHE parameters without any heuristics and compute them adaptively for every image block. Particle Swarm Optimization (PSO) - a population-based optimization technique proposed by Eberhart and Kennedy [27]. The motivation for PSO is from the biological social groupings of animals, which interact with each other to find food or save each other from predators**.** It uses swarm intelligence to solve any optimization problem. A swarm is a group of possible solutions also called particles, which provide a solution to the optimization problem. The performance of the optimization is evaluated based on a fitness function. Each particle of the swarm is identified with its velocity and position, which are updated through iterations. The search for best solution terminates at the end of the iterations or when the solution generates the highest fitness value [27]**.** Malik Braik and Alaa Sheta have implemented the PSO algorithm for enhancing general images [28].

The objective of this paper is to enhance the contrast of Multimodal and Multiresolution fused images, obtained by a fusing MRI and CT brain images. The fusion process induces blocking effect, noise and artifacts that greatly reduce the visual quality of the fused image. CLAHE is used to improve the contrast of the fused image. The structural information is of great importance in the fused MRI and CT images, as they indicate the periphery of tumor region. The PSO algorithm is used to select of the operational parameters of CLAHE automatically as well as enhance the image quality based on entropy and edge information. The operational parameters for them are set manually or empirically, however they change with images, making it tedious to set them manually every time.

3- Enhancement Technique for Fused MRI and CT Image

The research work includes preprocessing of MRI and CT images, image registration, and image fusion followed by image enhancement. More than 200 MRI and CT images containing Grade IV tumors - Glioblastoma is taken from www.Radiopedia.org for this work. There are quite a few databases available publicly for MRI images, but the challenge in our research is to get multimodal images for the same patient. Since these are acquired at different times and from different machines, they must be registered and preprocessed before fusing them. The Fig.3 shows the block diagram for the enhancement process. The preprocessing stage deals with resizing the image to size 256x256 and converting them to gray scale. The images contain Gaussian and Rician noises, which need to be eliminated; Non-Local Means Filter is used in filtering the CT and MRI images [29].

The preprocessing is followed by image registration - mandatory step, where both the images are matched for size, orientation and scaling. Subsequently, the images are fused using Laplacian Pyramid [31], since it provides excellent contrast for fused images. The fusion process introduces blocking effect and noise in the fused image, thereby reducing the image quality as discussed in section 1 [30]. Consequently, the enhancement technique is required to enhance the fused image with minimum loss of original information. Since the focus is the tumor region, adaptive block-based enhancement technique like CLAHE is chosen.

The operational parameters for CLAHE are block size, clip limit and distribution function. The proposed technique is executed for different block sizes like 2x2, 3x3, 4x4, 5x5, 7x7, 8x8 and 10x10. The 8x8 provided superior results in terms of contrast and structural information, hence chosen to be constant for the enhancement process. There is a trade-off for high contrast large block size is considered and for high edge information low block size is chosen. Similarly, the enhancement process was performed with the various distribution functions like Uniform, Rayleigh and Exponential distribution. The Uniform distribution performed better than the other distribution functions. Hence, uniform distribution is considered in the proposed enhancement process. The clip limit is initialized randomly in the range of 0 to 0.01, the PSO algorithm adaptively chooses the clip limit based on a fitness function. The proposed enhancement process is two-stage; firstly, the contrast of the MRI and CT fused image is enhanced by CLAHE algorithm, which increases the dynamic range of the image. Secondly, the multi-objective fitness function assists in choosing the clip limit that maximizes the entropy and edge information of the image. The Particle Swarm Optimization (PSO) algorithm is used to find the optimal clip limit. It begins with initialization of a group of 50 particles called as a swarm. In this work, the particles in the swarm represent the values for clip limit randomly chosen in the range of 0 to 0.01. The fitness function given by Eq. (1) helps to determine the best clip limit.

$$F(I_e) = \log(\log(E(I_s))) \times \frac{n\_edges(I_s)}{M \times N} \times H(I_s) \quad (1)$$

The cost function/fitness function is a product of entropy, sum of edge intensities and number of edge pixels. Since, multiple parameters are considered to measure the degree of enhancement; this function is called multi-objective function. Every particle i in the swarm is represented by two parameters velocity and position. For any particle ‘i’, the position and velocity indicate its location in the swarm and fitness value respectively. These two parameters are initialized with some random values and is updated in every iteration using Eq. (2) and Eq. (3).

$$v_i(t+1) = wv_i(t) + c_1 r_1 (p_i(t) - x_i(t)) + c_2 r_2 (g(t) - x_i(t)) \quad (2)$$

$$x(t+1) = x_i(t) + v_i(t+1) \quad (3)$$

where $v_i(t)$ and $x_i(t)$ represent the velocity and position for an particle i and iteration t. Eq.(2) comprises of three components – first component representing the initial velocity of the particle, the second component represents the particle’s decision based on its own experience and the third component indicates the particle’s decision based on swarm’s experience. In every iteration, the image is enhanced using CLAHE with the

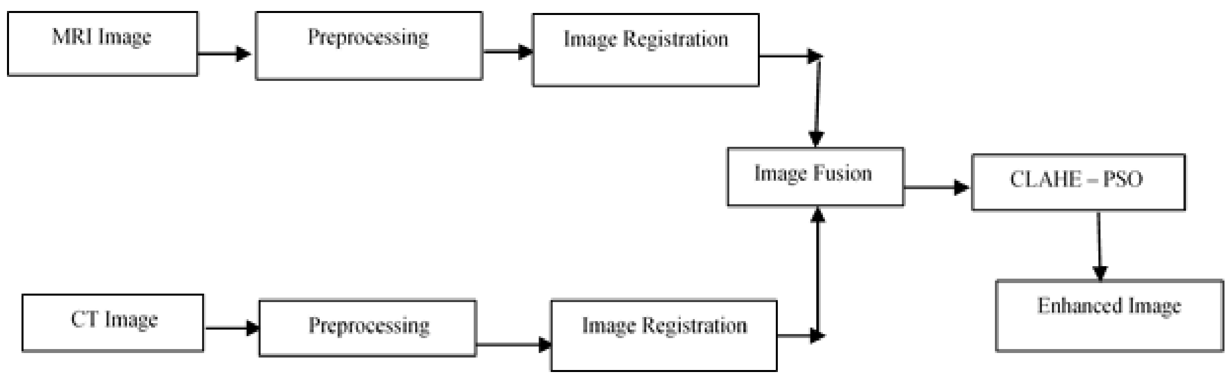


Fig.3 Block diagram for the Enhancement Process

selected clip limit (each particle). The fitness function is computed using Eq. (1). This process is repeated for all the particles to get **the best fit (clip limit). The clip limit that maximizes the fitness function can be accessed from the swarm based on its position and velocity and is represented as 'pbest' or $p_i(t)$**. This denotes the best local solution for that iteration. The enhancement process is repeated for all the iterations to get 'pbest' or $p_i(t)$ for each iteration. In case the 'pbest' value in the current iteration is greater than the previous one, then the 'pbest' is updated with a new 'pbest' and 'gbest', otherwise the 'pbest' from previous iteration is retained as 'pbest' and 'gbest'. The 'gbest' or g(t) in Eq.(2) is the global solution for the enhancement process obtained at the end of all the iterations. When "pbest' appears equal 'gbest' over a predefined number of iterations the enhancement process terminates.

Initialize the particle swarm
For each iteration
For each particle
Enhance the image using CLAHE
Compute the fitness value for the enhanced image as per Eq. (1) for every particle
If the fitness value is greater than the previous fitness value (pbest)
Set current value as the new pbest (gbest)
End
Choose the particle with the best fitness value among all the pbest (gbest)
For each particle
Calculate particle velocity as per Eq. (2)
Calculate the particle position as per Eq. (3)
End
Continue while maximum iterations are attained.
End
Report the gbest and pbest

Fig.4 Pseudo code for Proposed Technique (CPSO)

A balance between 'pbest' and 'gbest' is achieved by inertia weight represented as w, c1 and c2 - the positive acceleration constants and r1 and r2 are random values in the range of [0,1] . **The, an optimal value for the clip limit is obtained by PSO algorithm based on maximum value of the fitness function [32].** Fig. 2 shows the pseudo code for proposed CLAHE-Particle Swarm Optimization (CPSO) algorithm.

4- Experiment and Results

The CT and MRI images prior to enhancement are preprocessed, registered and fused using **Laplacian Pyramid** [31]. The dataset includes many variants for CT and MRI images – (i) images with contrast agent (ii) images without contrast agent and (iii) delayed images with contrast agent. The combination of CT and MRI T1 image, CT and MRI T2 and CT and MRI FLAIR images are used to generate the fused image. The proposed enhancement technique is tested on more than 200 MRI and CT fused images, containing the different types of Glioblastoma - Multicentric Glioblastoma, Multifocal Glioblastoma, Cystic Glioblastoma and Giant Cell Glioblastoma.

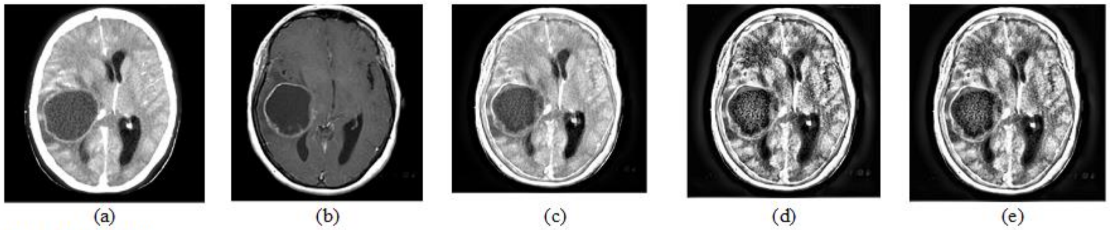


Fig. 5 Dataset D20

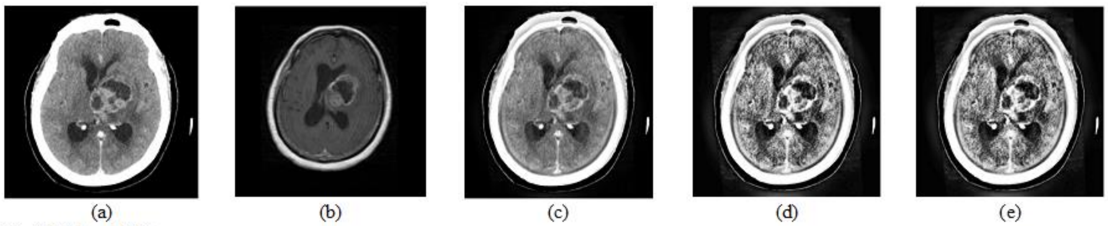


Fig. 6 Dataset D53

Fig. 7 Dataset D65

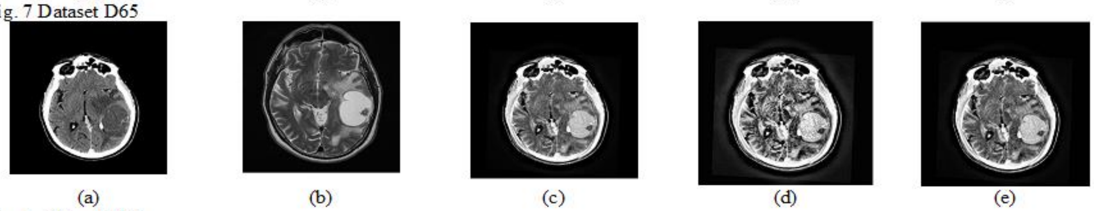


Fig. 8 Dataset D65

Table 1: Characteristics for Dataset D20 – D50

Dataset	CT	MRI	Size of Tumor	Group	Type of tumor
D20	High Contrast	Low contrast	Large	I	Cystic High Grade Glioblastoma
D53	Large size	Small size	Large	I	Glioblastoma (Grade IV)
D60	Large Size	Large Size	Large	I	Giant cell Glioblastoma
D67	Large size	Small size	Large	I	Primary CNS Lymphoma
D65	Small size	Large size, Poor contrast	Large	I	differential diagnosis – metastasis and glioblastoma
D17	Misaligned	Aligned, Poor contrast	Small	II	Glioblastoma (Grade IV)
D16	Aligned	Misaligned, Poor contrast	Small	II	Multicentric Glioblastoma
D15	Aligned	Misaligned, Poor contrast	Small	II	Multifocal glioblastoma
D6	Large	Small, Poor contrast	Small	II	Glioblastoma (Grade IV)
D50	Small	Large	Small	II	Glioblastoma (Grade IV)

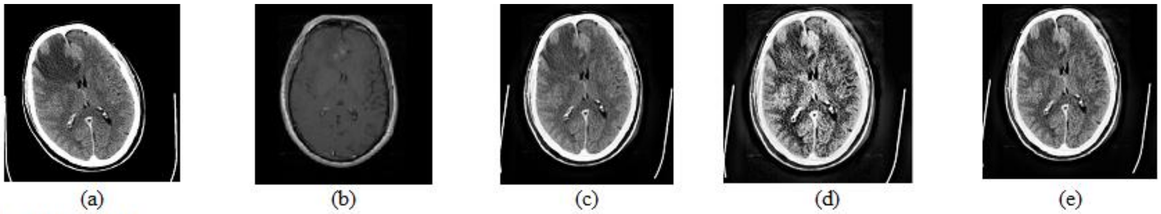


Fig. 9 Dataset D17

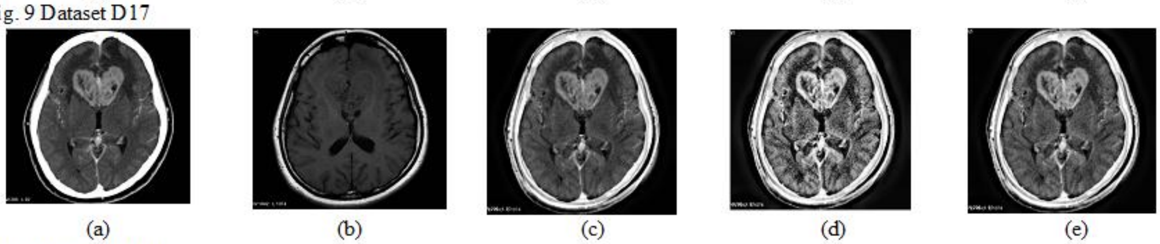


Fig. 10 Dataset D60

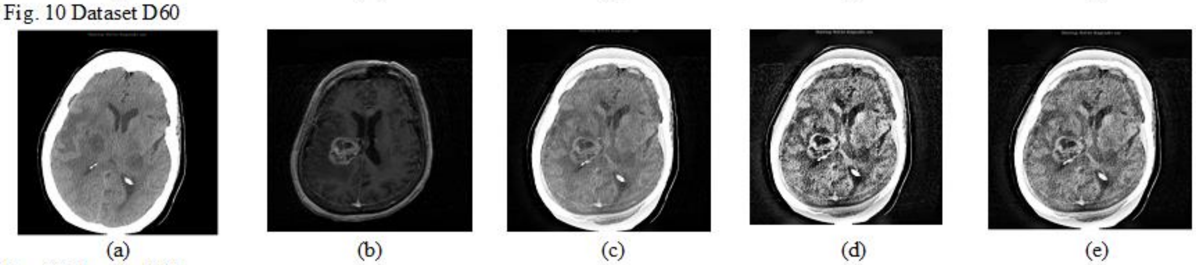


Fig. 11 Dataset D16

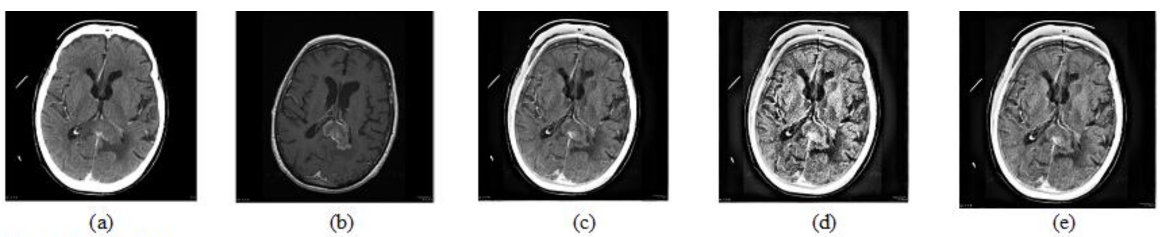


Fig. 12 Dataset D15

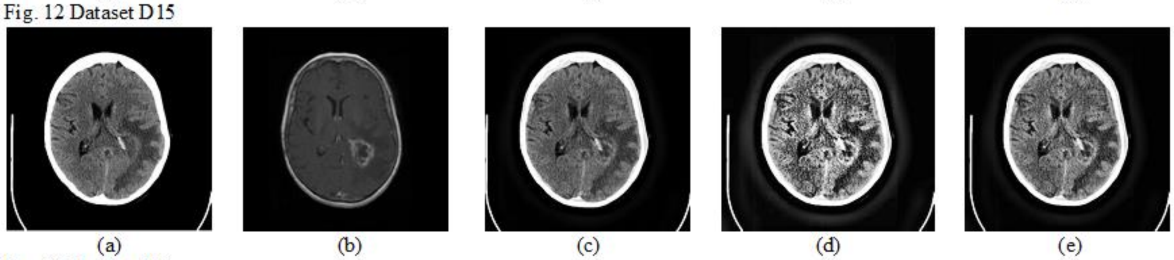


Fig. 13 Dataset D6

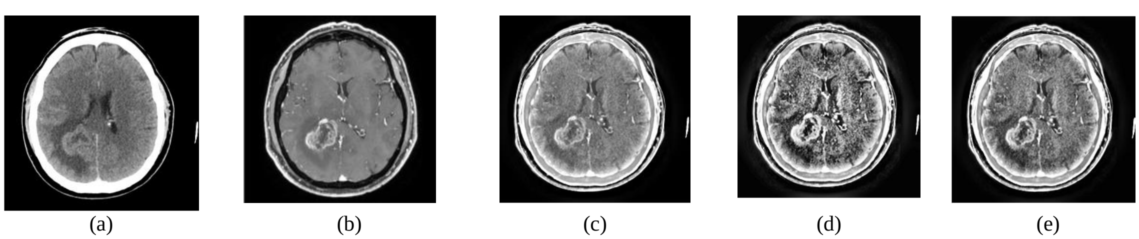


Fig. 14 Dataset D50

Fig.5-14 Dataset D20 – D50 (a) CT Image (b) MRI Image (c) Fused Image (d) CLAHE (e) Proposed

Table 2: The Performance Parameters for Evaluation of the Enhanced image

S.no	Parameter	Equation
1.	Standard Deviation (SD)	$\sqrt{\frac{1}{HxW}\sum_{x=1}^{H}\sum_{y=1}^{W}(F(x,y)-\mu)^2}$
2.	Entropy (En)	$\sum_{l=0}^{L-1} p(l)\log_2 p(l)$
3.	Mean Square Error (MSE)	$\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I_1(i,j)-E(i,j))^2$
4.	Peak Signal to Noise Ratio (PSNR)	$20\log_{10}\left(\frac{255}{\sqrt{MSE}}\right)$
5.	Structural Similarity Index Metric (SSIM)	$\frac{(2\mu_e\mu_s+C_1)(2\sigma_e+C_2)}{(\mu_e^2+\mu_s^2+C_1)(\sigma_e^2+\sigma_s^2+C_2)}$
6.	Universal Image Quality Index (UIQI)	$\frac{4\sigma_e\sigma_s(\mu_e+\mu_s)}{(\sigma_e^2+\sigma_s^2)(\mu_e^2+\mu_s^2)}$

Table 3: Experimental Results for Dataset D20 - D50

	STD			En			SSIM		
Dataset	**Fused**	**CLAHE**	**Proposed**	**Fused**	**CLAHE**	**Proposed**	**Fused**	**CLAHE**	**Proposed**
D20	91. 32	85.24	**85.41**	6.51	6.81	**6.78**	0.5	0.70	**0.73**
D53	84.77	86.03	**85.77**	6.24	6.65	**6.49**	0.74	0.65	**0.82**
D67	70.13	78.01	**71.84**	6.57	7.15	**6.75**	0.75	0.65	**0.92**
D65	83.34	85.03	**83.79**	6.32	6.79	**6.32**	0.64	0.69	**0.96**
D60	72.81	75.03	**73.02**	4.92	5.30	**4.96**	0.73	0.59	**0.93**
D17	79.54	80.34	**80.25**	6.23	6.79	**6.78**	0.55	0.62	**0.64**
D16	78.29	80.19	**79.18**	5.60	6.19	**6.11**	0.64	0.59	**0.66**
D15	84.24	81.39	**81.85**	6.28	7.02	**6.69**	0.55	0.60	**0.80**
D6	74.42	80.23	**77.42**	5.30	5.72	**5.54**	0.54	0.57	**0.76**
D50	73.59	80.09	**75.71**	6.22	6.60	**6.34**	0.56	0.64	**0.92**

	UIQI			PSNR			MSE	
Dataset	**Fused**	**CLAHE**	**Proposed**	**Fused**	**CLAHE**	**Proposed**	**CLAHE**	**Proposed**
D20	0.51	0.73	**0.72**	16.22	19.77	**20.86**	685.3	**532.63**
D53	0.75	0.63	**0.70**	24.62	21.43	**25.49**	467.07	**183.57**
D67	0.72	0.70	**0.95**	16.80	18.17	**27.50**	989.00	**115.60**
D65	0.64	0.73	**0.93**	13.41	21.50	**34.87**	459.57	**21.16**
D60	0.72	0.33	**0.66**	20.13	23.65	**36.39**	280.46	**14.92**
D17	0.55	0.62	**0.64**	22.37	21.59	**22.20**	450.82	**391.61**
D16	0.62	0.47	**0.59**	25.80	21.45	**23.77**	465.56	**272.78**
D15	0.50	0.55	**0.92**	26.47	21.08	**26.85**	506.09	**134.15**
D6	0.49	0.47	**0.43**	30.32	20.92	**25.99**	517.74	**163.66**
D50	0.48	0.58	**0.95**	21.61	19.13	**27.34**	793.91	**119.88**

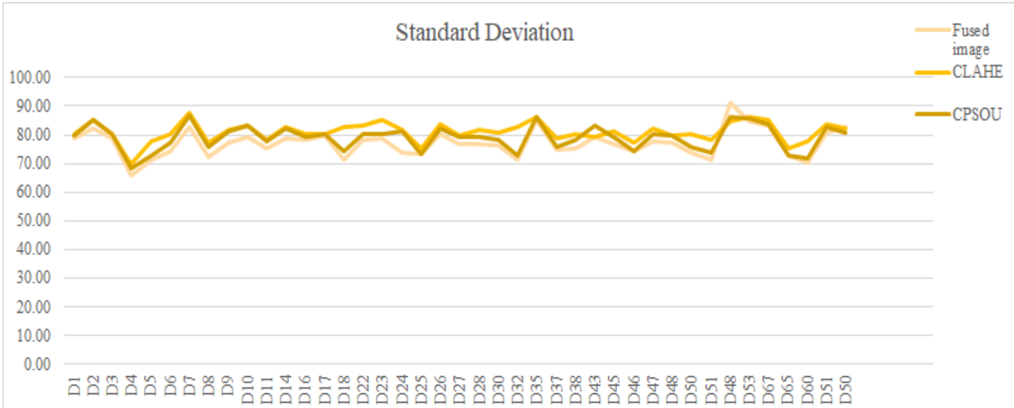


Fig.15 Standard Deviation

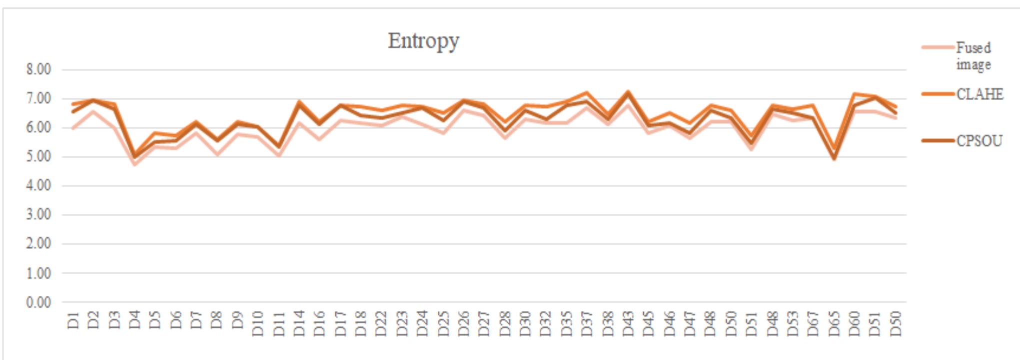


Fig.16 Entropy

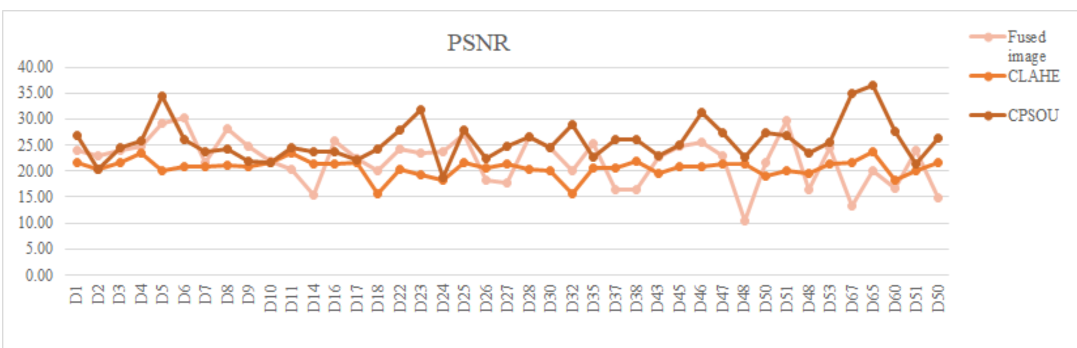


Fig. 17 PSNR

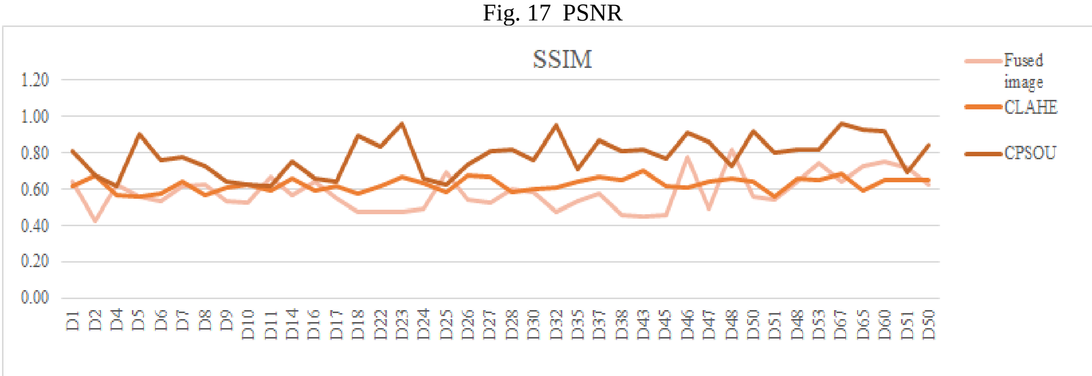


Fig.18 SSIM

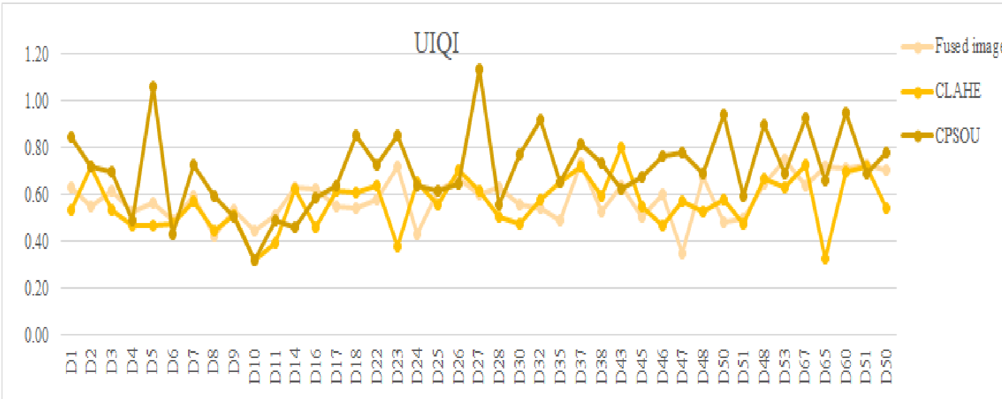


Fig.19 UIQI

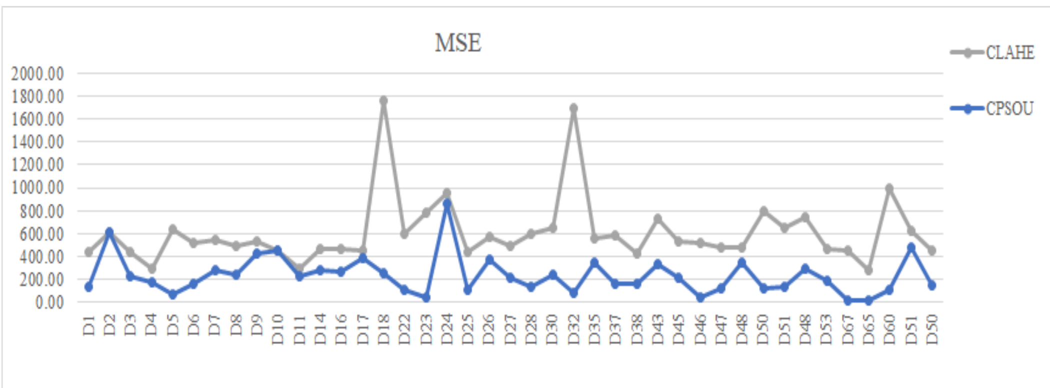


Fig.20 MSE

Fig. 5-14 shows some sample CT image, MRI image, fused image, CLAHE enhanced image and the CPSO (proposed) enhanced image for 10 datasets D20-50. Dataset belonging to Group I represent large tumors and Group II contain images with small tumors. In addition, high contrast CT and poor contrast MRI images are also tested with the proposed technique. For example, the CT image in dataset 20 has very high contrast and MRI T1 in dataset 16 and 17 has poor contrast. The proposed enhancement technique is implemened for fused CT and MRI image generated in [30-31], thus the performance parameters are the same. Standard Deviation (SD), Entropy (En), Structural Similarity Index (SSIM), Peak Signal to Noise Ratio (PSNR), Unique Image Quality Index (UIQI) and Mean Square Error (MSE) are considered to measure the quality of the enhanced image and compared with the fused image quality. The Table 2 shows the expressions for various performance parameters, the detailed definitions are in [30].

The Standard Deviation (SD) represents the contrast of the enhanced image. It helps in differentiating the soft tissues and tumors. The information content of the fused image is measured by Entropy. The ratio of the maximum pixel intensity to Mean Square Error(MSE) in the enhanced image is indicated by Peak Signal To Noise Ratio (PSNR) MSE is a measure of similiarity between the orginial image and the enhanced image. Structural Similarity Index Metric (SSIM) is a measure of the structural similiarity between the source image and enhanced image. The value of SSIM lies in the range of 0 and 1. The value 1 or close to 1 indicates high structural similiarity between the source image and enhanced image. Universal Image Quality Index (UIQI) represents coefficient correlation, illumination and contrast of enhanced image and source image. Fig.15-20 indicate the results for all that datasets and performance parameters. An efficient enhancement technique can provide high STD, En, PSNR, SSIM, UIQI and a minimum MSE. It can be observed that images enhanced by CLAHE seem enhanced, however it is only in terms of contrast or variaion of brightness but the structural information is poor.

Although, the standard deviation and entropy is high for enhanced image obtained by CLAHE, the PSNR, SSIM, UIQI is lesser than those obtained with the proposed enhancement technique.This is because, the CLAHE deals with enhancing the dynamic range and increasing the high intensity pixels and further decreasing the low intensity pixels of the image, however ignores the entropy or structural information of the image. Moreover, the MSE value is minimum for enhanced image in the proposed technique, indicating more similiarity to the orginal image with minimum loss of information. The clip limit in CLAHE is kept constant at 0.01, but with the proposed technique the clip limit is chosen adaptively in the range of 0 to 0.01 for each block of size 8x8, and it is found to present good results. Table 3 shows the performance paramters for the 10 image Datasets.

Table 4: Average values for 50 images

Parameter	**Fused**	**CLAHE**	**CPSO(Proposed)**
STD	77.12	81.02	**79.09**
En	5.98	6.48	**6.29**
PSNR	22.11	20.62	**25.72**
SSIM	0.59	0.63	**0.78**
UIQI	0.59	0.57	**0.71**
MSE	--	608.51	**234.67**
CL	--	0.01	**0.0 - 0.01**

5- Conclusion

A novel technique for enhancement of fused CT and MRI T1 weighted images is proposed. CLAHE is used to enhance the contrast of the fused image and the selection of operational parameters like clip limit, block size and distribution function is automated. The block size is chosen for the experiment as 8x8. The uniform distribution provides better results for the fused images. The clip limit is in range of 0-0.01, adaptively chosen by PSO algorithm. The proposed technique improves the visual quality. Further, tumors can be segmented and classified efficiently. Experiments show superior results for PSNR, SSIM, UIQI and MSE. This technique can be applied for enhancement of CT-PET images and MRI-PET images. There is an improvement in contrast and entropy for Laplacian Pyramid fused image and CPSO enhanced image by 2.55 % and 5.18% respectively. CPSO enhanced image indicates an huge increase in PSNR, SSIM and UIQI by 16.32%, 32% and 20.3% respectively. Also, the MSE has greatly reduced. The proposed technique performs better than CLAHE in terms of PSNR, SSIM, UIQI and MSE.

References

[1] A.Prof Frank Gaillard et al. “Glioblastoma”. Article. https://radiopaedia.org/articles/glioblastoma

[2] Javier E Villanueva, Marc C Mabray, Soonmee Cha, “Current Clinical Brain Tumor Imaging”, Neurosurgery 81:397-415, 2017.

[3] Olivier Keunen, Torfinn Taxt, et al., “Multimodal imaging of gliomas in context of evolving cellular and molecular therapies,” Adv. Drug Delivery Reviews, Vol.76, 2014, pp. 98-115.

[4] Wolf-Dieter Heiss, Peter Raab, Heinrich Lanferman, “Multimodality Assessment of Brain Tumors and Tumor Recurrence”, J Nucl Med 2011, Vol.52, pp.1585-1600.##

[5] Maikel Verduin, Inge Compter, Danny Seijvers et al. “Noninvasive Glioblastoma Testing: Multimodal Approach to Monitoring and Predicting Treatment Response”. Hindawi, Disease Markers, Vol. 2018, Article ID 2908609.

[6] Yin Fei, Gao Wei, Son Zongxi. “Medical Image Fusion Based on Feature Extraction and Sparse Representation”. Hindawi, International Journal of Biomedical Imaging, Vol. 2017.

[7] Hui Huang, Xi’an Feng, Jionghul Jiang. “Medical Image Fusion Algorithm Based on Nonlinear Approximation of Contourlet Transform and Regional Features”. Hindwai Journal of Electrical and Computer Engineering, Vol. 2017.

[8] Hiba Mzoughi, Ines Njeh, et al. “Histogram Equalization-Based Techniques for Contrast Enhancement of MRI Brain Glioma Tumor Images: Comparative Study,” in Proc. 4th

[9] International Conference on Advanced Technologies for Signal and Image Processing, ATSIP, March 21-24, 2018.

[10] Bin Yang, Shutao Li, “Multifocus Image Fusion and Restoration with Sparse Restoration,” IEEE Transactions on Instrumentation and Measurement, Vol. 59, No.4, 2010.

[11] Robert D Fiete. “Modelling the Imaging Chain of Digital Camera”, Image Enhancement Processing, Chap. 9, [Online], SPIE Digital Library.

[12] Yakun Chang, Cheolkon Jung et al. “Automatic Contrast-Limited Adaptive Histogram Equalization with Dual Gamma Correction”, IEEE Access, Vol. 6, 2010. ##

[13] Byong Seok Min, Dong Kyun Lim et al. “A Novel Method of Determining Parameters of CLAHE Based on Image Entropy”. International Journal of Software Engineering and Its Applications, Volume 7, No.5, pp. 113-120, 2013.

[14] Monika Agarwal and Rashima Mahajan. “Medical Image Contrast Enhancement using Range Limited Weighted Histogram Equalization”, in Proc.6th International Conference on Smart Computing and Communication, ICSCC December 2017.

[15] Monika Agarwal and Rashima Mahajan. “Medical Image Contrast Enhancement using Quad Weighted Histogram Equalization with Adaptive Gama Correction and Homomorphic Filtering”, in Proc. 7th International Conference on Advances in Computing & Communications, ICACC-2017.

[16] Youlian Zhu, Cheng Huang. “An Adaptive Histogram Equalization Algorithm on the Image Gray Level Mapping”, in Proc. 2012 International Conference on Solid State Devices and Materials Science, Physics Procedia 25, pp.621-628.

[17] K Zuiderveld. "Contrast Limited Adaptive Histogram Equalization", in Graphics gems IV, pp474-485, San Diego, CA, USA, Academic Press Professional, Inc. ##

[18] Hardeep Kaur, Jyothi Rani. "MRI brain image enhancement using Histogram equalization techniques", in Proc. IEEE WISPNET, 2016.

[19] Kitti Koonsanit, Saowapak Thongvigitmanee, et al. "Image Enhancement on Digital X-Ray Images using N-CLAHE", in Proc. IEEE 10th Biomedical Engineering International Conference (BMEiCON), Aug 2017.

[20] Shelda Mohan and T R Mahesh. "Particle Swarm Optimization Based Contrast Limited Enhancement for Mammogram Images", in Proc. 7th International Conference on Intelligent Systems and Control, 2013.

[21] Madhukar Bhat, Tarun Patil M S. "Adaptive Clip Limit for Contrast Limited Adaptive Histogram Equalization (CLAHE) of Medical Images using Least Mean Square Algorithm", in Proc. International Conference on Advance Communication Control and Computing Technologies (ICACCCT), IEEE, 2014.

[22] Luis G More, Macrcos A Brizuela et al. "Parameter Tuning of CLAHE-based on Multi-Objective Optimization to Achieve Different Contrast Levels in Medical Images", in Proc. International Conference on Image Processing (ICIP), IEEE 2015.

[23] Sheeba Jennifer, S Parasuraman, Amudha Kadirvelu. "Contrast Enhancement and Brightness preserving of digital mammograms using fuzzy clipped contrast-limited adaptive histogram equalization algorithm", Applied Soft Computing Vol. 42, pp.167-177, May 2016.

[24] Justin Joseph, J Sivaraman, R Periyasamy, V R Simi. "An Objective method to identify optimum clip-limit and histogram specification of contrast limited adaptive histogram equalization for MR Images," Nalecz Institute of Biocybernetics and Biomedical Engineering 37, pp.489-497, 2017.

[25] Gurshan Singh, Anand Kumar Mittal. "Controlled Bilateral Filter and CLAHE Based Approach for Image Enhancement," International Journal of Engineering and Computer Science, Volume 3, Issue 11, Nov 2014.

[26] Aziz Makandar, Bhagirathi Halalli. "Breast Cancer Image Enhancement using Median Filter and CLAHE," in Proc. International Journal of Scientific & Engineering Research, Volume 6, Issue 4, 2015.

[27] Mehemet Zeki, Sarp Erturk. "Enhancement of Ultrasound Images with Bilateral Filter and Rayleigh CLAHE", in Proc. 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, Turkey, 2015.

[28] James Kennedy and Russell Eberhart. "Particle Swarm Optimization," in Proc. International Conference on Neural Networks, 4, pp.1942-1948, IEEE 1995.

[29] Malik Braik, Alaa Sheta, Aladdin Ayesh. "Image Enhancement Using Particle Swarm Optimization," Proceedings of World Congress on Engineering, Vol. 1, London, UK, 2007.

[30] Leena Chandrashekar, Sreedevi A. "Assessment of Non-Linear Filters for MRI Images," in Proc. Second IEEE International Conference on Electrical, Computer and Communication Technologies, Feb 22-24, Coimbatore, India, 2017.

[31] Leena Chandrashekar, Sreedevi A. "A Hybrid Multimodal Medical Image Fusion Technique for CT and MRI brain images," IGI Global International Journal of Computer Vision and Image Processing (IJCVIP), Vol. 8, Issue 3, Sept 2018.

[32] Leena Chandrashekar, Sreedevi A. "A Novel technique for fusing Multimodal and Multiresolution Brain Images," in Proc. 7th International Conference on Advances in Computing and Communications, ICACC 2017, Aug 22-24, Cochin, India, 2017.

[33] Leena Chandrashekar, Sreedevi A, "Leena Chandrashekar, A Sreedevi, "A Multi Objective Enhancement Technique for Poor Contrast Magnetic Resonance Images of Brain Glioma blastoma", Third International Conference on Computing and Network Communication, CoCoNet 2019.

Leena Chandrashekar pursing her PhD in Department of Electrical and Electronics, RVCE, VTU under the guidance of Prof. Sreedevi A. Her research interests include Biomedical Image Fusion, Biomedical image processing, Neural Networks and Deep Learning. She has published papers in international conferences and journals.

Sreedevi A, is working as Associate Professor in Department of Electrical and Electronics Engineering, RVCE, Bangalore, obtained her PhD in the area of Signal Processing. She has published many papers in international conferences and journals. Her research interests include Digital Signal Processing, Digital Image Processing and ControlSystems.

Model of Technological, Managerial and Marketing Infrastructure for Intelligent Technology Efficiency in Telecommunication Industry - Case Study: Telecommunication Infrastructure Company of Ilam Province[1]

Heshmat Moradhaseli
Department of management, central Tehran branch Islamic azad university, Tehran, iran
h.haseli123@gmail.com
Jalal Haghighat Monfared*
Department of management, central Tehran branch Islamic azad university, Tehran, iran
jhm1847@gmail.com

Received: 01/Jan/2020 Revised: 04/Feb/2020 Accepted: 28/ Apr /2020

Abstract

Today's, intelligent agent system (IAS) are considered as an important part of people's lives. Therefore, many of organizations try to implement IAS in their mechanism. One of these organizations in Iran is telecommunication Infrastructure Company. Because any implementation need a model which clarify the structural and contextual components, therefore, the current research is conducted to provide a model for developing the necessary infrastructure for implementation of intelligent technologies in the communication and telecommunication mechanisms of Ilam Province. To achieve the goal, a qualitative approach and thematic analysis method were used. The research population consisted of all experts in the field of ICT in Ilam province Infrastructure Communications Company that using purposeful sampling method and relying on theoretical data saturation, 10 of them were selected as sample. Semi-structured interviews were used to collect the data. The data were analyzed through theme analysis. Based on the method, 4 themes, 10 main categories and 153 open codes were extracted. The findings of the study showed that to transform communication mechanisms into intelligent technologies, there must be technological, management, marketing and cultural infrastructure. Technological infrastructure consisted of intelligent software and hardware; management infrastructure consisted of knowledge and belief; marketing infrastructure included attracting intelligent technology to audiences, encouraging ideas, physical and virtual channels; and finally, cultural infrastructure, it was staff training and public awareness.

Keywords: Intelligent Agent System; Intelligent Technology; Intelligent Communication; Technological Infrastructure; Management Infrastructure; Marketing Infrastructure.

1- Introduction

The industrial revolution experience and the age of enlightenment in many countries have led to the ever-expanding and accelerating movement of societies towards growth and development, and the required tools and technologies were produced and offered to facilitate this process. However, technologies have played a significant role in these societies, and new generations of these technologies are increasingly being introduced today, but intelligent technologies has been conducted in the different area. Therefore, the need for the intelligent technology is very urgent.

In the past, communication was thought to be merely a one-way process in which the information holder transmits information to its recipients [1] as there were only one or more suppliers presenting their information. But with the advent of technologies, communication mechanisms have changed dramatically. On the other hand, not only macro-technology policies have been based on the equipment of existing technologies with intelligent technologies, but there has also been a widespread social wave in society that calls for intelligent acting of their social and organizational relationships [2], [3]. Now, if this tendency is to take an organized form and to apply such an approach to the country's telecommunications system, there needs to be a coherent conceptual model to build the necessary infrastructure on it.

Today, telecommunications systems have complex tasks. It is not only expected that traditional communication tools are truly established in every country, but new generations of information technologies, especially intelligent technologies, are also expected to grow. Meanwhile,

* Corresponding Author
1. Extracted article of Phd.Thesis

connectionism has emerged as a significant paradigm in the telecommunications industry. Under this paradigm, telecommunications companies strive to serve their customers through its connections. In total, there are two orientations in connectivism theory. One is the orientation of Stephen Downes, who works in the field of online communication content composition, and the other is George Siemens's approach [4, 5]. Downs focuses more on connective knowledge and communication and considers it an interactive knowledge that takes place on a social network. But Siemens sees connectivism as more of a communication paradigm for the digital age [5].

From a connectionism standpoint, we acquire our knowledge, interactions, relationships, and learnings through the connections we make or the connections we belong to. Downes defines connectivism as a communication theory that considers knowledge as the creation of a network of interactions [4]. In fact, networks are created from a set of interactions between professionals, students, professors, general public, databases, blogs, colleges, telecommunications companies and websites, and learning in such an environment requires the ability to create and develop networks [6]. However, the basic theory of this research is based on connectionism. This theory is based on the following dimensions and components:

1. Connectionism interactions are based on diversity of opinions and behaviors [7].
2. Learning and communication is the process of connecting to nodes and information resources [8].
3. Communications may be created through non-human (machine) applications [9].
4. Communication capacity and ability in how to communicate is much more important than communication itself [10].
5. Maintaining and developing continuous communication is essential [11].
6. The ability to understand the connections between branches, thoughts and concepts is an essential skill [12].
7. Up-to-date and accurate knowledge is the main goal of all connectionism communication activities [13].
8. Decision making itself is a kind of learning. Choosing what to learn and the meaning of the input information is itself a kind of learning [14].

On the other hand, by looking at these dimensions, it can be seen that in order to establish a connectionism intelligent system, it is necessary to provide the appropriate infrastructure. Daft [16] believes that these infrastructures which are needed to establish information technology includes three infrastructures: technological infrastructure, management infrastructure, and marketing infrastructure. Therefore the present study considered these foundations and infrastructures as the research foundations.

The research questions are:

1. What kind of technological infrastructures are necessary for the development of intelligent technologies in the Infrastructure Communications of Ilam Province?

1. What kind of managerial infrastructures are necessary for the development of intelligent technologies in the Infrastructure Communications of Ilam Province?

1. What kind of marketing infrastructures are necessary for the development of intelligent technologies in the Infrastructure Communications of Ilam Province?

The present study aimed to provide a clear answer to these questions. The purpose of this study is to identify the areas and infrastructure needed for the development of intelligent technologies in Telecommunication Infrastructure Company of Ilam. In addition, the research was seeking to address the contexts, infrastructures, and backgrounds should be provided to transform the Telecommunication Infrastructure Company's mechanisms into intelligent communications.

Ignoring this model to develop intelligent technologies can lead to high costs for the organization. Because the use of traditional generations of information technology always requires the existence of significant agents and human resources to exploit it. But intelligent technologies conducted their job intelligently without human intervention and without the need for constant monitoring. On the other hand, ignoring intelligent technologies deprives the infrastructure communications company of the ability to respond quickly and efficiently. Because today, one of the indicators that is considered by the audience and customers of telecommunication and communication companies is the ability to provide easy, fast and quality service, all of which can be achieved by intelligent technologies.

2- Theoretical Background

Intelligent systems or intelligent agent systems have different types. However, the common feature of all these systems is that they can carry out their activities without human intervention. These systems also include very simple and basic systems such as thermostats they also include more sophisticated systems, such as simulation systems or learning management systems. Some intelligent systems have the ability to learn in new situations and others can only adapt to different situations. Intelligent systems must have the capacity to recognize the status quo and make appropriate decisions and actions for different situations.

There are many different types of intelligent systems, but all of them have four features in common which are:

1. Independence and self-control
2. Activity
3. Accountability
4. Adaptability [5].

One of the features of intelligent systems is their flexibility and adaptability. In the context of communication, such a feature means identifying the needs of users and providing the necessary feedback and information resources for them. Existing systems focus more on a particular aspect of communication. In the world of communication and education, part of intelligent systems are intelligent cognitive systems that are responsible for recognizing users' cognitive level and responding appropriately to their cognitive needs. Part of it focuses on the emotional aspects of users and provides appropriate responses based on the emotional state of users. In some researches, showed that using several strategies, intelligent systems can be built based on Web 2 and the linking system. These strategies are providing credible knowledge, management integration management guide, data recovery on time, provide data in a homogeneous manner [5]. It should be noted that the effect of increased agent transparency on operator workload depends on the amount of information provided and whether the information is required for the operator’s task performance [6]. This is while it has been stated before that intelligent systems perform their activities without the need for human intervention. In its absence of a user manual, an intelligent entity must learn to align its values with that of humans [7]. It is in such a complex world that intelligent systems are introduced [8]. However, some believe that the deployment of intelligent systems requires different infrastructures [9]. Among these, they have shown that technological, cultural, marketing and management infrastructures are among the most important infrastructures [10].

3- Related Works

Mohammadi 2017 designed a conceptual model of intelligent learning management system responsive to cultural diversity of Mehr Alborz virtual university students. This research was conducted using a mixed methods and its findings showed that to govern a multicultural approach in our higher education system, its infrastructure that include multicultural education policy and multicultural human resources are needed to be provided first to respond to the diverse needs of cultural minority students by relying on macro education strategies (at university and ministry level) and micro strategies (at classroom level). Based on the findings of the first part of the research, the conceptual model of the intelligent system was also designed and validated, whereby the present systems, curriculum content, educational activities and specific cultural programs were presented to each student. Finally, suggestions were made to deploy the system in question [3].
In another study, Rocher, Janse, Portolan and Streitz (2014) examined the demands of users of intelligent home environment systems. Ambient Intelligence encompasses the concept of future life, with much of the day-to-day home activities being carried out by intelligent systems. The study sought to identify and categorize the demands that different users of various cultures from the system. To achieve this goal, mixed research method was used. The findings of the study showed that user expectations and demands for the intelligent home environment system are as follows: 1. The system can detect all user behaviors and never go astray; 2. The system must be highly secure and protected to be able to provide proper privacy for users; 3. The system should be able to gain added value compared to other existing systems; 4. The system should never replace actual interactions between people in unnecessary situations; 5. Convenience and comfort should always be preserved at home and the system should not disturb it [15].
Mazadi, Qasim Aghaei and Oren, (2008) in another study merely sought to model culture and its dimensions for simulation in culturally-intelligent systems. The findings of this study showed that the design of this system required the following information to be collected: 1. Identifying the nationality of each user, since Schwartz's research provides comprehensive information about the cultural characteristics of each nation and can be used to identify the cultural characteristics of each user. 2. Learning from the behaviors and performance of the student (user) according to which the cultural characteristics of the user can be classified based on the indicators specified in different cultural categories. 3. Assessment of student's cultural characteristics through PVQ questionnaire [16].

4- Research Methodology

To achieve the purpose of the present study, qualitative approach and theme analysis were used. The qualitative approach was used because the pattern of intelligent technologies had to be consistent with the cultural, political, economic, and social context of any society, and identifying such a context would probably require qualitative research. On the other hand, identifying the dimensions and components needed to transform existing communication into intelligent communication requires an indigenous model appropriate to the cultural, social and economic context of the community in question. For this reason, the theme analysis research method can be instrumental in achieving this goal. Theme analysis attempts to extract themes and content from various discourses, interviews, and texts, and to identify the focal ideas of those documents [15].
The research population consisted of all experts in the field of ICT in Telecommunication Infrastructure Company of Ilam. Of these, 10 were selected using purposive sampling. In addition, sample size estimates were based on

theoretical data saturation. It was purposeful because the researcher consciously, not based on randomize selection, selects people who have specialized knowledge and attitudes about the phenomenon and can provide in-depth information to the researcher. The reason for the selection of these people was that they all had educational backgrounds and academic knowledge in the field of information technology, especially intelligent technologies. Secondly, all of them were operating and managing the information systems operating and operating in the Telecommunication Infrastructure Company of Ilam. Thus, all interviewees had adequate theoretical and conceptual knowledge of intelligent technologies.

The data collection tool was semi-structured interview. The reason for using this type of interview was to allow the researcher to analyze and evaluate their perspectives and experiences without limiting the participants' views to a particular context or basis.

1. In your opinion, what kind of technologies features should be able to run and implement intelligent systems in order to make intelligent communication technology in Ilam?
2. In your opinion, what kind of management and leadership approaches should be able to run and implement intelligent systems in order to make intelligent communication technology in Ilam?
3. what kind of marketing procedure should be able to run and implement intelligent systems in order to make intelligent communication technology in Ilam?

Finally, the researcher analyzed and interpreted the research findings using a thematic analysis method based on two types of open coding and axial coding. In this type of analysis, information is first collected from the participants and then the data is coded, without any specific theoretical basis (open coding). Next, duplicate and redundant codes are removed, and the process of code reductions continues until we reach broader categories that can include more minor codes called axial coding [17]. During the open coding procedure, 153 open codes were extracted. After the open coding procedure, a number of duplicate and redundant codes were removed and 10 main categories and 4 general themes were extracted by constant reference to interviews, analysis, comparison and review of data. What is presented in the following table is a report of extracted sample codes, subcategories, main categories and main research theme.

Two methods were used to validate the research data. First, Triangulation which is the process of corroborating evidence from different individuals (e.g., a principal and a student), types of data (e.g., observational field notes and interviews), or methods of data collection (e.g., documents and interviews) in descriptions and themes in qualitative research. Second, Member checking is a process that the researcher asks one or more participants in the study to check the accuracy of the account.

5- Data and Findings

In view of the tangible relationship between the experts of the Telecommunication Infrastructure Company with the research problem, they were asked to participate in the present study and to provide their views on the research questions. What follows are the findings from the interviews with these individuals, which are summarized and analyzed in Table 1.

Table 1. Themes and categories extracted from the interviews

Themes	Main categories	extracted sample codes
Technological infrastructure in the telecommunications industry	- Intelligent hardware - Intelligent software	Purchasing the right equipment; Purchasing the right software; Sufficient bandwidth; IT based communication channels; Providing telecommunication data such as fiber, FTTH, ADS; high speed Internet.
Managerial infrastructure in the telecommunications industry	- Knowledge - Belief and faith	Targeted management; Committed to adherence to goals and methods; Scientific evaluation of programs and goals; Defining intelligent strategic perspectives; Defining medium-term and long-term goals, mechanisms and structures; The firm conviction that inteligent ing is performed; Manager awareness of intelligent features; manager awareness of intelligent services and benefits
Marketing infrastructure in the telecommunications industry	- Attracting intelligent technologies - Encouraging people with ideas	Proving intelligent Technology Attraction for Investor Companies, Providing Reliability to Investor Companies for Investing, Providing Return on Investment for Investor Companies
Cultural infrastructure in the telecommunications industry	-Staff training - Informing people - Virtual channels	Raising public awareness, use of visual media for awareness, use of print media for awareness, use of national media for awareness, use of virtual spaces for awareness, full awareness of intelligent features to Client, Human Resources Training

The findings of this study revealed that all of these infrastructures can be summarized and categorized into four infrastructures. What follows is a report of findings from experts in this field.

A) Technological Infrastructure in the Telecommunication Industry

Findings from the analysis of the interviews showed that to become an intelligent organization and to make Ilam's telecommunications mechanisms intelligent, it is necessary to develop and strengthen existing technologies. This development can be examined in two dimensions.

1. Hardware Development

Interviewees believe that in order to integrate intelligent technologies into communication mechanisms, existing hardware needs to be further developed. They believe that in order to become more sophisticated, communication infrastructure needs to have specific hardware in place. For example, interviewee No. 4 has raised important points in this regard.
"To become intelligent, you have to acquire good hardware infrastructure. We need wider bandwidth. Otherwise, many intelligent programs won't run at all."

2. Software Development

Interviewees not only believe that the company's existing hardware needs to be upgraded, but some interviewees also believe that the existing software infrastructure is inadequate and needs to be strengthened. In this regard, the viewpoint of interviewee No. 3 can be considered.
"In order to be able to implement intelligence, we need to strengthen the software of the Telecommunication Infrastructure Company. We need to have the telecommunications data such as fiber and ADSL to be optimized so that it can be used for intelligent purposes."

B) Management Infrastructure in the Telecommunications Industry

In order to find intelligent technologies in any organization, it is not only necessary to provide the right software and hardware infrastructures, but also to efficiently manage the organization. Based on the data obtained from the interviews, the management infrastructure can be examined in the following two dimensions.

1. Knowledge of Managers

Interviewees strongly believe that a manager who does not have the appropriate knowledge and information required by intelligent technologies cannot provide the appropriate context for its advancement. To this end, interviewee No. 4 has put forward such a view.
"In my opinion, a manager who wants to implement intelligent technologies must first have sufficient knowledge and expertise about these technologies. A specialized manager can provide the necessary background for the success of intelligent technologies."

2. Managers' Beliefs

Data from interviews with IT professionals suggest that there must be a firm belief that these technologies are effective and efficient. Some of the participants such as interviewee # 1 are serious critics of existing sloganistic approaches and believe that only by relying on the true belief of managers can these technologies be successfully implemented.
"Organizers must have a firm belief in intelligent technologies, not in slogans or in luxury, performing no appropriate action." "

C) Marketing Infrastructure in the Telecommunications Industry

Today's world is accompanied with increasing competition for organizations and companies to achieve sustainable development. In this regard, outsourcing and privatization is one of the most important steps that organizations take to succeed in their intelligent mechanisms. The marketing infrastructure can be discussed in the following dimensions.

1. Make Intelligent Technologies Attractive

Data from interviews with experts suggest that in order to attract customers for intelligent technologies, they must be offered as an attractive technology. Some interviewees have explicitly stated that in order to be successful in privatizing and outsourcing intelligent technologies, the necessities and benefits of inteligent ening must be clearly stated for outsider companies. For example, interviewee No. 5 has commented on this.
"To be successful in this (intelligent marketing) business, one has to realize the benefits of using intelligent technologies for investor companies."

2. Encourage the Proponents

Many interviewees believe that in order to be successful in marketing intelligent technologies, it is important to encourage those who have ideas to invest in intelligent technologies or implement their technological ideas. Various approaches have been suggested by the interviewees, but overall, many believe that telecommunications companies should be able to provide the opportunity to attract creative people. Interviewee No. 5 expressed his view in this regard.
"Proponents should be encouraged to invest in intelligent technologies or implement their ideas."

3. Advertise Via Virtual Channels

With the advent of virtual communications, today, it is expected that intelligent technology mechanisms will also be propagated through virtual channels. Some interviewees believe that these technologies can be advertised using the infrastructure systems and automation. Interviewee No. 8 has identified the Enterprise Portal of the Telecommunication Infrastructure Company as an important channel for advertising.
"The Telecommunication Infrastructure Company’s Portal itself is an important channel for outreach. Because anyone who visits an organization's website will see the most important news and announcements and can take action in investment."

D) Cultural Infrastructure of the Telecommunications Industry

Intelligent human activities also require cultural infrastructure to enable users of both technologies to learn how to use those technologies and to minimize the potential for misuse of these technologies. Interviewees believe that cultural infrastructure is a key pillar for inteligent ening, and without it we cannot hope for the successful implementation of intelligent technologies. In sum, cultural infrastructure can be studied in two dimensions, each of which is discussed below.

1. Staff Training

Interviewees believe that effective and efficient cultures for intelligent technologies are possible only if the providers of those technologies have the appropriate knowledge, recognition, and attitude of those technologies. Interviewees have abundantly stated that providers of these technologies must have a precise and accurate understanding of intelligent technologies in order to effectively train and cultivate.
For example, interviewee No. 6 reflected this belief in his statement.
"Adequate training should be given to staff on how to use intelligent technologies."

2. Informing People

The end users of intelligent technologies include the people who are always the clients of these technologies. In order to realize all the positive benefits of intelligent technologies, people need to be aware of the mechanisms and how to use them. Interviewees believe that informing people is an essential step for cultural contexts. Interviewee # 7 identified informing people as a key step in cultivation.
"People need to be informed on how to use intelligent technologies. As long as information is not available, the benefits of intelligent technologies cannot be effectively exploited."

5-1- System Model

Nevertheless, a review of the findings of this study revealed that in order for intelligent technologies to be well developed, all four infrastructures needed to be interoperable. On the other hand, these findings suggest that all four infrastructures play an essential and effective role in the implementation of intelligent technologies. The final model of the research is shown in Figure 1.

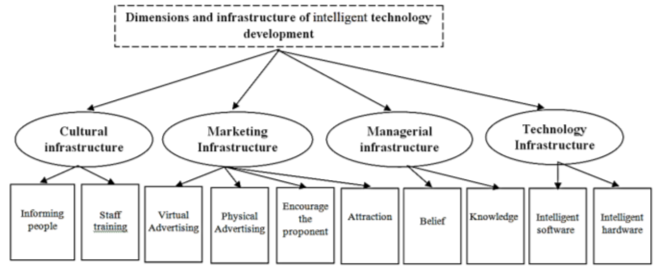


Fig. 1. Final research model

6- Results and Discussion

The findings of this study showed that in the first place, to transform existing communications into intelligent communications, there must be adequate hardware and software infrastructure in the company. Purchasing the right equipment; Purchasing the right software; Sufficient bandwidth; IT-based communication channels; Providing telecommunication data such as fiber, ADSL; FTTH; all were suggestions made by the participants. This part of the research findings is consistent with the findings of Mohammadi 2017, Rocher and et al 2014 and Mazadi and et al 2008 researches. Because they have also shown that there must be technological infrastructure for intelligent technologies.
Second, the company's management needs to be flexible and, in this respect, the most central emphasis is on executives who are responsible for implementing intelligent systems. Interviewees believed that relevant managers should have both an appropriate knowledge and information about intelligent technologies and a strong belief in the effectiveness of these technologies. As such, the firm conviction about intelligent making; manager awareness of the intelligent making features; and manager awareness of the services and benefits of intelligence were among the suggestions the interviewees made. This part of the findings is new. Because previous studies have not been addressed this issue.
Third, marketing infrastructure is also one of the most important infrastructures needed for intelligent communication and telecommunications activities. In the meantime, the findings of the study showed that to provide such an infrastructure, intelligent technologies must be attractive to investor companies, encourage proponents to invest in the field, and eventually take place through appropriate and effective virtual advertising channels. This part of the findings is new. Because previous studies have not been addressed this issue.
Fourth, there is a need for cultural infrastructure to transform communication mechanisms into intelligent mechanisms, and to provide such infrastructure, firstly, the staff of the communication infrastructure company must receive the necessary training and secondly, to provide

appropriate information to the general public to implement the necessary cultivation.

This part of the research findings is consistent with the findings of Mohammadi 2017, Rocher and et al 2014 and Mazadi and et al 2008 researches. They all dealt with culture as a fundamental element

The present study includes innovation in several ways.

First, this research contributes to the development of literature because no comprehensive and serious research has been done on this issue in Iran so far, and the model presents new views to the intelligent world.

Second, the present research can help the planners and policy makers of the Infrastructure Communications Company to make coherent and codified plans to improve the state of intelligent technologies by relying on the data and evidence obtained from this research.

Third, this research has shown that in order to establish an intelligent system, not only technological infrastructure is needed, but also other infrastructures must be provided.

Based on the results, the following suggestions are made.

1. Since the present study showed that there must be appropriate technological infrastructure for the establishment of an intelligent system, it is recommended that the communications company provide and deploy the appropriate software and hardware infrastructure.
2. Since the findings showed that leadership and management skills are essential elements for intelligent technologies, it is recommended that appropriate training courses be held for current managers on the necessities of intelligent systems in organization.
3. Since the findings show that proper marketing should be done for intelligent technologies, it is recommended that the infrastructure communications company use virtual and physical methods to advertise its intelligent technologies.

References

[1]. Constantinides, E and Fountain, S. Web 2: conceptual foundation and marketing issues. Journal of Direct, Data and Digital Marketing Practice. Vol. 9, No. 3, 2008, 231-244.

[2]. Mohammed, Ph, Mohan, P. The design and implementation of an encultured web- based intelligent tutoring system for computer science education. The IEEE international Conference on advanced learning technology. 2011, 501-505.

[3]. Mohammadi. Sh. designing the pattern of intelligent learning responding to Cultural diverse of students in Mehr University of alborz. Thesis of educational Management of Tehran University. 2017.

[4]. Ahrens, A; Wen, M, L; Huang, Y; Zascerinska, J and Bassus, O. A comparative study of the relationship between social dimension of web 2.0 technologies and E- learning: students view in German and Taiwan. Paper presented at the 3rd International Conference on Innovation and Entrepreneurship and its implications in Engineering and Business Education, The La Consolacion College Manila, Manila, Philippines, 2010. 16 - 18 November.

[5]. Trannsue, B, M. Connectivism and information literacy: moving from learning theory to pedagogical practice. Public Services Quarterly, 9:3, 2014, 185-195.

[6]. Mercado, J, M; Rupp, M, A; Barbet, D & Procci, K. Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management. Vol. 58, No. 3, May 2016, pp. 401–415.

[7]. Kumar, V; Dixit, A; Rajshekar, R & Javalgi, G. Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing. J. of the Acad. Mark. Sci. DOI 10.1007/s11747-015-0426-9. 2017.

[8]. Rosenfeld, A; Agmon, N; Maksimov, O; Azaria, A & Kraus, S. Intelligent Agent Supporting Human-Multi-Robot Team Collaboration. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[9]. Lemoine, P, M, P; Trentesaux, D; Rey, G, Z & Millot, P. Designing Intelligent Manufacturing Systems through Human-Machine Cooperation Principles: A Human-Centered Approach. Computers & Industrial Engineering, Volume 111, September 2017, Pages 581-595.

[10]. Dermeval, D; Paiva, R; Borges, D; Bittencourt, I, I & Vassileva, J. Authoring Tools for Designing Intelligent Tutoring Systems: A Systematic Review of the Literature. International Journal of Artificial Intelligence in Education volume 28, pages336–384, 2018.

[11]. Lin, Ch, F; Yeh, Y, Ch; Hung, Y, H and Chang, R, I. Developing an intelligent agent for managing web 2.0 services. International conference on future computer supported education. IEPRI, 2, 2012, 738-741.

[12]. Alexander, S, Th, V. An affect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor. A thesis presented at Massey University, Albany, New Zealand. 2007.

[13]. Anderson, T and Dron, J. Three generations of distance education pedagogy. International review of research in open and distance learning. 2011, Vol. 12.3.

[14]. Blanchard, E and Frasson, C. Making intelligent tutoring systems culturally aware: the use of Hofstede's cultural dimensions. HERON Laboratory, Computer Science Department. 2010.

[15]. Brown, S. (2010). From VLEs to learning webs: the implications of Web 2.0 for learning and teaching. Interactive Learning Environments. Vol. 18, No. 1, 1–10.

[16]. Hao. Y & Borich, G. A Practical Guide to Evaluate Quality of Online Courses. In Holim Song & Terry Kidd. Handbook of Research on Human Performance and Instructional Technology. New York: Information science reference. 2010.

[17]. Hixon, E., Barczyk, C., Buckenmeyer, J., & Feldman, L. Mentoring university faculty to become high quality online educators: A program evaluation. Online Journal of Distance Learning Administration, 2011, 14 (5).

[18]. Kaplan, A, M and Haenlein, M. Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. Business horizon, 2016, 59, 441- 450.

[19]. Kazemifard, M; Ghasem- Aghaee, N and Oren, T. Emotive and cognitive simulation by agents: roles of three levels of information processing. Cognitive systems research, 2011, 13, 24-38.

[20]. Rocker, C; Janse, M; Portolm, N and Streitz, N. User requirements for intelligent home environments: A scenario-driven approach and empirical cross-cultural study. Published in: proceeding of the 2014 joint conference on intelligent objects and ambient intelligence, 2014, 111-116.

[21]. Mazadi, Z; Ghasem Aghaee, N and Oren, T. Prelude to cultural software agents: cultural backgrounds in agent simulation. Conference: Proceedings of the Spring Simulation Multi conference. 2008, 135-142.

[22]. Creswell J. w. Educational research: planning, conducting and evaluating quantitative and qualitative research. (Fourth edition). Pearson. Boston. 2012.

Heshmat Moradhaseli received the B.S. Degree in Computer Engineering from Payame Noor University, Ilam Center, Iran in 2011 and Master of Information Technology Management Degree from Azad University, electronici Tehran Branch, Iran in 2015. He is currently a PhD student at Azad University, Central Tehran Branch, Tehran, Iran. His research interest is in the development of intelligent technologies.

Jalal Haghighat Monfared is an assistant professor of Industrial Management at Azad University, Central Tehran Branch, Iran. His research is interest is in the development of intelligent technologies.

Complexity Reduction in Massive-MIMO-NOMA SIC Receiver in Presence of Imperfect CSI

Nilufar Tutunchi
Faculty of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University
nilufar.totonchi@gmail.com
Afrooz Haghbin*
Faculty of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University
a.haghbin@srbiau.ac.ir
Behrad Mahboobi
Faculty of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University
b.mahboobi@srbiau.ac.ir

Received: 28/Aug/2019 Revised: 01/Mar/2020 Accepted: 08/Apr/2020

Abstract

One of the main reasons for switching to the next generation of communication systems is the demand of increasing capacity and network connections. This goal can be achieved using massive multiple input - multiple output (massive-MIMO) systems in combination with Non-orthogonal multiple access (NOMA) technique. NOMA technology uses the successive interference cancellation (SIC) receiver to detect user's signals which imposes an additional complexity on the system. In this paper, we proposed two methods to reduce the system complexity. The proposed method despite imperfect channel state information (CSI) in the receiver, there is not significantly reduction in the system performance. Since the computation of matrices inverse has a high computational complexity, we used the Neumann series approximation method and the Gauss-Seidel decomposition method to compute matrices inverse in the SIC receiver. Simulation results are provided at the end of the paper in terms of bit error rate (BER) at the receiver which show, these methods have lower computational complexity in comparison with the traditional methods while they cause a slight performance reduction in the SIC receiver. Also, we examined the increasing and decreasing value of imperfect channel state information in the system performance which shows the increasing value of imperfect channel state information, cause a slight performance reduction in SIC receiver.

Keywords: Massive MIMO; NOMA; Complexity; SIC; Neumann; Gauss - Seidel.

1- Introduction

Reviewing and estimating the wireless networks traffic shows an increase in the networks traffic. Researchers are providing an overview for 5G systems that can increase the capacity of telecommunication cells, spectral efficiency and performance efficiency as well as decrease PAPR [1]. Multiple access techniques use the orthogonality principle in order to reduce the inter channel interference. A new non- orthogonal multiple access method has been introduced that in this method, with accepting a certain level of interference, we can remove the orthogonal principle and increase the number of sub-channels.
NOMA is one of the key techniques of multiple access methods, which is very promising for increasing the performance of the fifth generation, and in comparison, with FDMA, offer more favorable advantages that can be used to increase the spectral efficiency. This method uses a new field of work that called 'power'. For a certain level of interference to be accepted, all users should have a power limit. So, the main point in this method is the use of signals that have a different power level.
NOMA can support more connections than other multiple access techniques, which is important in the 5G systems. On the other hand, the use of NOMA enables users to access all available sub-carriers. Therefore, it can be said that NOMA can allocate channel capacity and bandwidth equally to all users. The combination of NOMA with other existing technologies, such as MIMO, effectively increases network bandwidth. So, more users can be used in the communications system. Reducing complexity in these systems is also an important point especially in downlink. Since the SIC receiver in the NOMA method imposes additional complexity on the system, solving this problem is particularly useful to reduce user equipment complexity. In this paper, we examine the massive MIMO transmission system combined with NOMA technique and use algebraic methods in the SIC receiver to reduce the receiver complexity. We have examined the performance of system

* Corresponding Author

with these methods and compared the results with the exact SIC receiver.
Many studies on NOMA technique and its combination with MIMO systems have been presented. An article[2] explains the challenges, theories and research that have been done for NOMA technique. In [3], by assuming that users are randomly deployed in a cell, the effect of path loss on the performance of NOMA in 5G systems has been characterized. Simulation results shows that NOMA achieves higher rates and better performance than the OMA technique. In [4] an optimal method for allocating power and maximizing ergodic capacity with low complexity has been proposed. Then a simulation has been carried out to examine the proposed method performance. By analyzing the numerical results, we can see significant increase in the interest of the NOMA technique in comparison with the OMA technique. Also, an article [5] is devoted to a new design of the detection matrix for MIMO-NOMA and its performance. In [6], the NOMA system was initially introduced, both of uplink and downlink. Simulations have also been performed to evaluate the rate of NOMA and OMA cells, which show that increasing the number of users, will increase the NOMA rate, but the OMA rate will go to saturation. In [7] a design for massive MIMO systems combined with NOMA has been developed that makes designing easier. To increase the efficiency of NOMA, the effect of the SIC receiver in ferroelectric mode is investigated in system [8]. In article[9] the user clustering scheme and optimization of precoding matrix for a NOMA system are investigated. The user selection plan and power plan for the NOMA system are based on ZF radiation to improve the capacity of the NOMA system in [10]. In [11], researchers have proposed a low complexity sub-optimal user-clustering method for Rayleigh fading channels by considering the effects of intra-cluster pilot contamination, inter-cluster interference, imperfect successive interference cancellation (SIC) and statistical downlink channel state information (CSI) at secondary users.

In this paper, we focus on the complexity reduction in massive-MIMO-NOMA SIC receiver in presence of imperfect channel state information. The concept of massive-MIMO-NOMA has been validated by using systematic implementation in [7]. Compared to these existing works, the contributions of this paper are as follows: we consider a general Massive-MIMO-NOMA system in which all users have a fixed set of power allocation coefficients. The interference between clusters has been removed by a precoder and the interference between the users in a cluster has been removed by the SIC receiver. We have used the Neumann series approximation method and the Gauss-Seidel decomposition method to compute matrices inverse in SIC receiver and a comparison between traditional LU decomposition method and Neumann series approximation and Gauss-Seidel decomposition methods have been investigated.
The rest of this paper is organized as follows: In Section.2, we present the system model of massive MIMO-NOMA. Section.3 examine complexity and discusses on algebraic methods to reduce computational complexity. Section.4 discuses on simulation results. Finally, Section.5 concludes the paper and offers suggestions to further works.

2- System Model

Consider a cellular downlink transmission scenario, in which the base station and receiver is equipped with M and N antennas, respectively. Channel between each transmit and receive antenna is flat fading with Rayleigh distribution. The modulation scheme is BPSK. In this system, user clustering is used, so that users' channels in a cluster have the same correlation matrix, and the channel matrix of the users in the different clusters is almost orthogonal. We divide the users into K clusters and each cluster has L user. The channel state information is imperfect in the receiver and the channel coefficients are summed with a random Gaussian variable Δ with the variance $\delta^2 = 10^{-4}$. The scheme used in this simulation is using a precoder at the BS and the SIC receiver at the user side.
Transmitted signal from the BS for k'th cluster is as follows

$$\mathbf{t_k} = \mathbf{w}_k \sum_{l=1}^{L} \sqrt{\alpha_{k.l}} \mathbf{S}_{k.l} \quad (1)$$

$\mathbf{S}_{k.l}$ is a signal that contains information of the l-th user in the k-th cluster. $\alpha_{k.l}$ is the power allocation coefficients for the l-th user in the k-th cluster. $\mathbf{w_k}$ is $k \times 1$ precoder vector. The received signal is as follows

$$\mathbf{Y}_{kl} = \mathbf{h}_{kl}\mathbf{t}_k + \mathbf{h}_{kl} \sum_{i=1;i\neq k}^{K} \mathbf{t}_i + \mathbf{N}_{kl} \quad (2)$$

for l=1, 2…, L.

$\mathbf{N}_{kl}$ is the receiver noise which is complex Gaussian random variable with zero mean and variance N_0.
As seen in Eq. (2), the first term of the equation is the useful signal for the l-th user in the k-th cluster. The second term shows the interference of the other clusters. In addition to the interference between the clusters in Eq. (2), in a cluster, each user interferes with other users, which is received by the SIC receiver. In this paper we have considered the SIC receiver with the use of the MMSE detector.

2-1- The Process of Removing Interference between Users with the SIC receiver

For eliminating interference between users with the nonlinear SIC receiver and the MMSE detector, we consider the allocation of power to users based on the SINR received by the users as follows

$$|\mathbf{h}_{11}|^2/N_{01} < |\mathbf{h}_{12}|^2/N_{02} < \cdots < |\mathbf{h}_{1L}|^2/N_{0L} \quad (3)$$

$$\alpha_{11} > \alpha_{12} > \cdots > \alpha_{1L}\,; \quad \sum_{l=1}^{L} \alpha_{1l}^2 = 1\ . \quad (4)$$

According to the NOMA principles [6], the first user that has the best channel condition detects the signal of the other users with more allocated power and then removes their effects from the received signal. Then the first user can detect its corresponding signal with a slight amount of interference from the other users. Similarly, this process is will be done for other users to eliminate the interference between all users. Now we have to eliminate the inter clusters interference. We will do this with the use of precoder at BS.

In the case of using a precoder at the BS, it is necessary for the channel matrix to be perfect at the BS. We have also assumed that the channel matrix **H** is perfect at BS, and it is used to construct the MMSE precoder vector, **w**, to eliminate the inter clusters interference. The vector **w** is constructed based on the stronger user-channel vector that eliminates inter clusters interference. To remove the inter clusters interference, the following condition should be met

$$\mathbf{H}^{\mathrm{H}}\mathbf{P} = 0 \quad (5)$$

$$\mathbf{h}_{il}\mathbf{w}_k = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases}\ . \quad (6)$$

As seen in Eq. (6), the inter clusters interference can be removed.

3- Computational Complexity

Computational complexity is one of the important criteria for implementing communications system. This important criterion examines the amount of resources needed to implement the algorithms. The most commonly used resources are the calculation time and computational complexity. In this section, the computational complexity of the system is examined. Since the purpose of this paper is to reduce the computational complexity, two methods used to reduce matrix inverse calculations will be introduced.

One of the algebraic methods to calculate the matrix inverse is the Neumann series approximation method. In this method, with the order of n, the matrix inverse is calculated in the following order[12]

$$\mathbf{X^{-1}}_L = \left(\sum_{n=1}^{L}(\mathbf{I} - \mathbf{AX})^{n-1}\right)\mathbf{A} \quad , (\mathbf{A}=\mathbf{X}^{-1})\ . \quad (7)$$

For this approximation method, the complexity is as follows

If **A** is a matrix with dimensions $N \times N$, then the complexity of calculation the matrix $\mathbf{A}^{-1}$ is as shown in Table 1[13].

Table 1: Complexity of Neumann method with different orders for calculating the inverse of matrix $\mathbf{A}_{N\times N}$ [13]

Neumann-Series Method	***Complexity***
Order of one	N
Order of two	$3N^2$-$2N$
Order of three	$16N^3$+$3N^2$-$4N$

It is noteworthy that we use 'order of two' for the above approximation method, because the first order with the lowest complexity does not provide a good performance for the system. Also, the third order with best performance, provide higher complexity, so we use the second order to ensure that, along with an appropriate complexity, the system performance dose not reduce significantly.

Another method to iteratively compute matrix inverse is 'gauss-seidel' decomposition method. In this method. we decompose matrix **A** as follows

$$\mathbf{A} = \mathbf{L}_* + \mathbf{U} \quad (8)$$

where $\mathbf{L}_*$ is lower triangular component and **U** is strictly upper triangular component given by[14];

$$\mathbf{L}_* = \begin{bmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad \mathbf{U} = \begin{bmatrix} 0 & a_{12} \cdots & \cdots a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \quad (9)$$

If $\mathbf{A}^{-1}$= **X**, and

$$\mathbf{AX} = \mathbf{I} \quad (10)$$

According to the multiplication properties of matrices, we know

$$\mathbf{A}x_i = e_i \quad ; i = 1.2. \ldots .n\ . \quad (11)$$

In the above relation x_i is the i th column of matrix **X** and e_i is i th column of matrix **I**.

So according to Eq. (8) and (10), and with m iterations, we have.

$$(\mathbf{L}_* + \mathbf{U})\mathbf{X} = \mathbf{I} \quad (12)$$
$$\mathbf{L}_*\mathbf{X} = \mathbf{I} - \mathbf{U}\mathbf{X} \quad (13)$$
$$\mathbf{X}^{(k+1)} = \mathbf{L}_*^{-1}\left(\mathbf{I} - \mathbf{U}\mathbf{X}^{(k)}\right) \quad (14)$$
$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(e_i - \sum_{j>i} a_{ij}\, x_j^{(k)} - \sum_{j<i} a_{ij}\, x_j^{(k+1)}\right)$$
$$i = 1,2,3,\dots,n \ \ and \ \ k = 1,2,\dots,m \quad (15)$$

Table 2: Complexity of Gauss-Seidel decomposition method with m iterations for calculating of matrix inverse $\mathbf{A}_{N\times N}$ [14]

Method	***Complexity***
Gauss-Seidel	$m\times O(N^2)$

In Table 3 we compare the complexity of SIC receiver using three methods of matrix inverse calculation.

Table 3: Comparison complexity of three methods of inverse matrix $\mathbf{A}_{N\times N}$ for entire system with L users

Methods	***Complexity***
LU	$\frac{L(L+1)}{2}O(N^3)$
Neumann	$\frac{L(L+1)}{2}O(N^2)$
Gauss-Seidel	$\frac{L(L+1)}{2}mO(N^2)$

4- Simulation Results and Discussions

In this section, computer simulations are used to study the performance of the proposed Massive-MIMO-NOMA scheme. The BS is equipped with M = 50 antennas and the receiver is equipped with one antenna. In Fig. 1 the BER performance of the Massive-MIMO-NOMA scheme, in the case of traditional LU decomposition method for calculating the matrices inverse is studied. In Fig. 2 and Fig. 3 the impact of using of Neumann series method and gauss-seidel decomposition method for calculating the matrices inverse are illustrated, respectively. Fig. 4 shows a comparison between three methods. As can be seen from these figures, the use of the Neumann series method and gauss-seidel decomposition method for calculating matrices inverse in Massive-MIMO-NOMA scheme, reduce the system performance slightly.

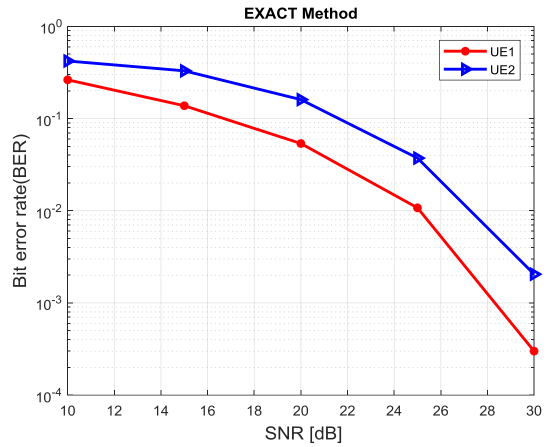


Fig. 1 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters (a_1^2 =6/7, a_2^2 =1/7) for exact method.

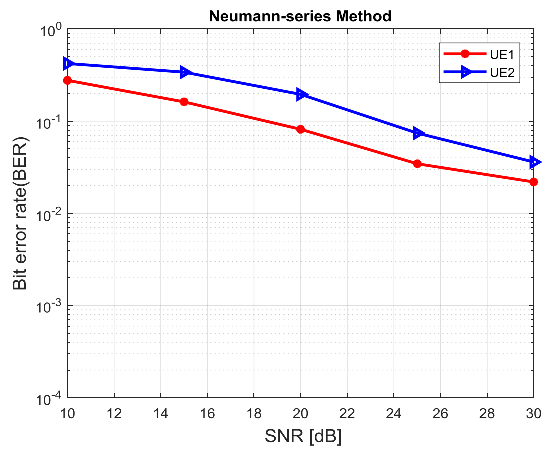


Fig. 2 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters (a_1^2 =6/7, a_2^2 =1/7) for Neumann series method.

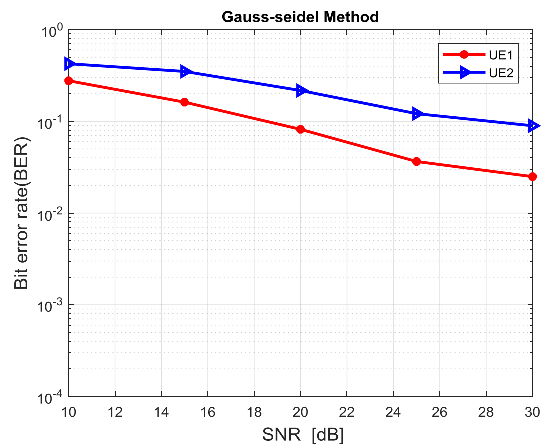


Fig.3 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters (a_1^2 =6/7, a_2^2 =1/7) for Gauss-Seidel decomposition method.

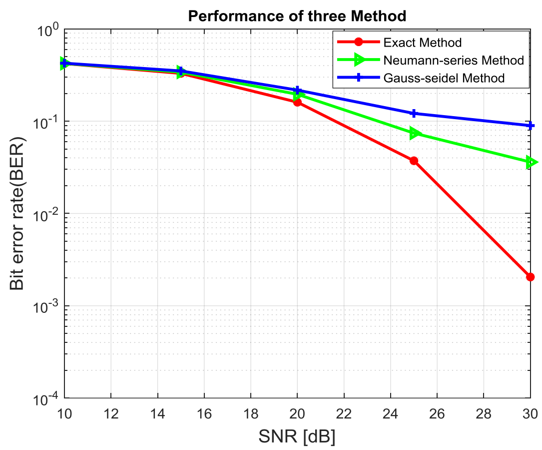


Fig. 4 Comparison the BER curve of the scheme in the matrix inverse calculation between three methods for 2 users and 2 clusters ($a_1{}^2$ =6/7, $a_2{}^2$ =1/7)

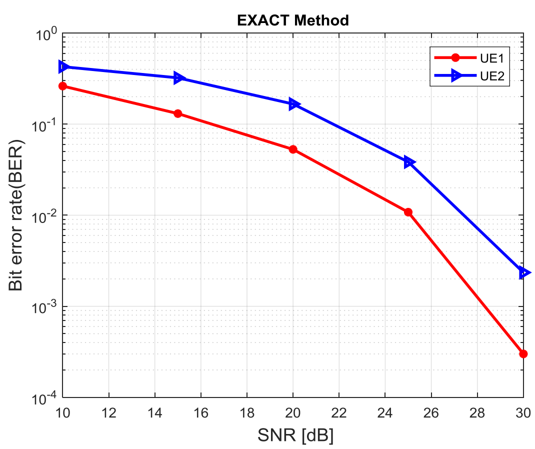


Fig. 5 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters ($a_1{}^2$ =4/5, $a_2{}^2$ =1/5) exact method

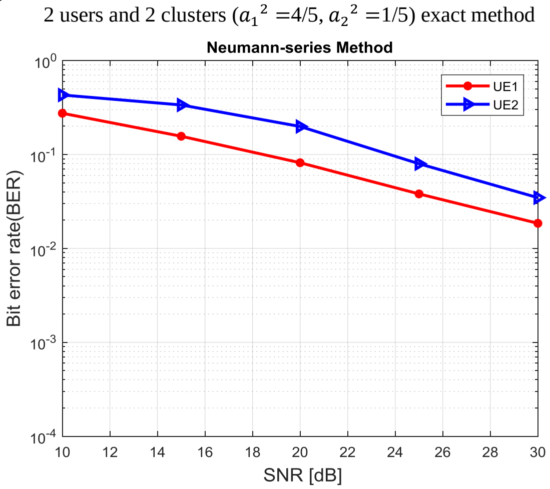


Fig. 6 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters ($a_1{}^2$ =4/5, $a_2{}^2$ =1/5) for Neumann series.

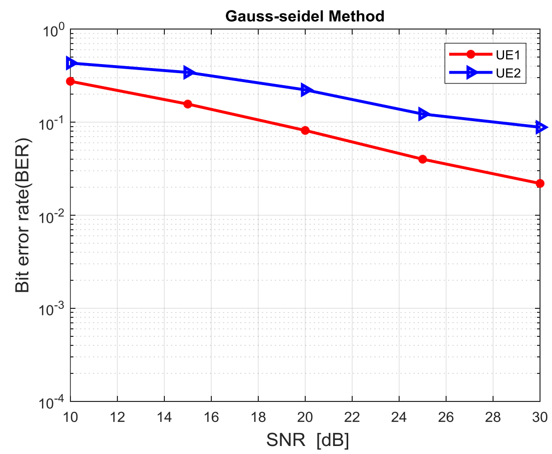


Fig. 7 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters ($a_1{}^2$ =4/5, $a_2{}^2$ =1/5) for Gauss-Seidel decomposition method

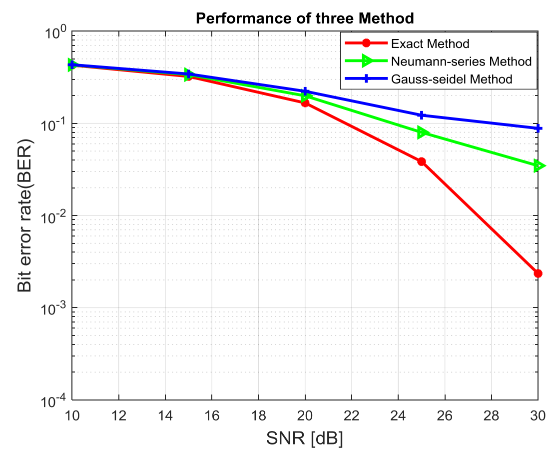


Fig. 8 Comparison the BER curve of the scheme in the matrix inverse calculation between three methods for 2 users and 2 clusters($a_1{}^2$ =4/5, $a_2{}^2$ =1/5)

Now we change the power allocation coefficients and repeat the simulation with $a_1{}^2$ =4/5, $a_2{}^2$ =1/5 values. The effect of increasing power allocation coefficients can be seen in the Fig. 5 to Fig. 8.

Then we increase the amount of variable power that indicates imperfect channel state information in the receiver and set it to $\delta^2=10^{-3}$ and observe its effect on system performance.

As it can be seen in Fig. 9 to 12, increasing the amount of δ^2 for channel state information in the receiver, reduce the system performance slightly. So, we can conclude that these methods are not so sensible to imperfect channel state information.

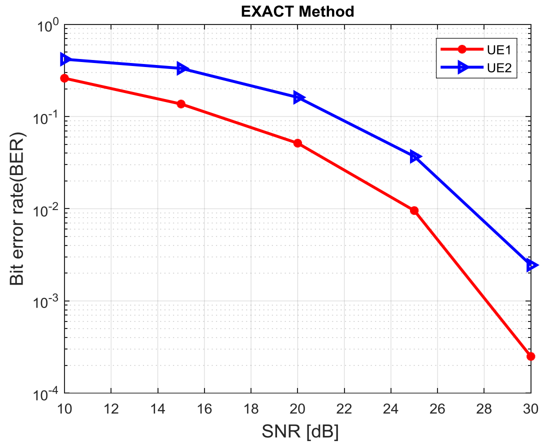


Fig. 9 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters ($a_1{}^2 = 6/7$, $a_2{}^2 = 1/7$) for exact method

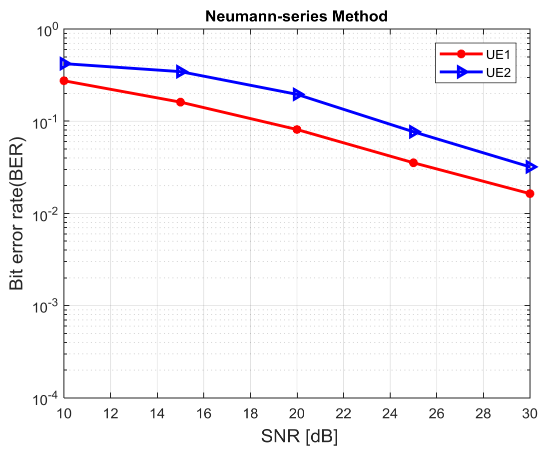


Fig. 10 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters($a_1{}^2 = 6/7$, $a_2{}^2 = 1/7$) for Neumann series method

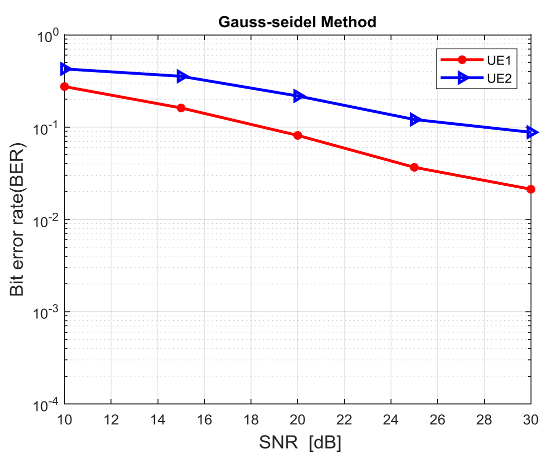


Fig. 11 The BER curve of the scheme in the matrix inverse calculation for 2 users and 2 clusters($a_1{}^2 = 6/7$, $a_2{}^2 = 1/7$) for gauss-Seidel decomposition method

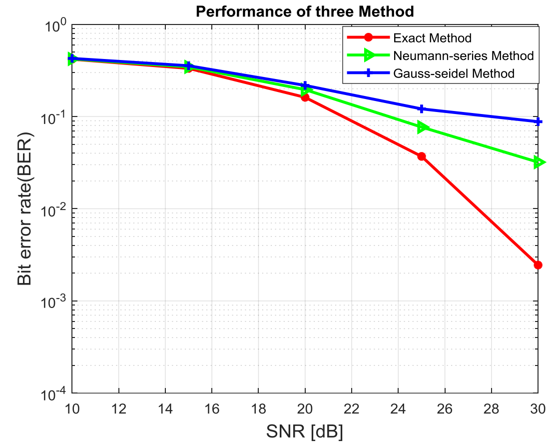


Fig. 12 Comparison the BER curve of the scheme in the matrix inverse calculation between three methods for 2 users and 2 clusters($a_1{}^2 = 6/7$, $a_2{}^2 = 1/7$)

As we know, performance of the SIC receiver is dependent on the matrix inverse accuracy. Better accuracy of matrix inverse results better performance of SIC receiver. In the Gauss-seidel method, there is a factor *m* that indicates the number of iterations in the algorithm and affects the accuracy of matrix inverse. The larger the factor *m*, the better approximation of matrix inverse be. In this work we have chosen *m*=10 to have the complexity order of Gauss-Seidel method not so larger than that of Neumann series method. So, on the other hand we have a lower accuracy of matrix inverse than that of Neumann series method because of low convergence rate of Gauss-Seidel method [14]. So, using Gauss-Seidel method to approximate matrix inverse, causes a lower performance of SIC receiver in comparison to that of using Neumann series method. The LU decomposition method performs better than two other methods in SIC receiver, because it exactly calculates the matrix inverse. To better illustration of the results comparisons, in Table 4, we have examined the three methods mentioned in terms of complexity and performance.

Table 4: A numerical example for comparison the complexity and performance of three methods of inverse matrix $\mathbf{A}_{N\times N}$ for entire system with *L* users in 25dB SNR

Methods	***Complexity***	***Approximate BER***
LU	$\frac{\boldsymbol{L(L+1)}}{\mathbf{2}}\boldsymbol{O(N^3)}$	4×10^{-2}
Neumann	$\frac{\boldsymbol{L(L+1)}}{\mathbf{2}}\boldsymbol{O(N^2)}$	7×10^{-2}
Gauss-Seidel	$\frac{\boldsymbol{L(L+1)}}{\mathbf{2}}\boldsymbol{mO(N^2)}$	12×10^{-2}

5- Conclusions

NOMA technology uses the SIC receiver to detect user’s signals which imposes an additional complexity on the system. In this paper, we proposed two methods to reduce the system complexity. The goal was complexity reduction in massive-MIMO-NOMA SIC receiver in presence of imperfect channel state information. we considered a general Massive-MIMO-NOMA system. The interference between clusters removed by a precoder and the interference between the users in a cluster removed by the SIC receiver. we used the Neumann series approximation method and the Gauss-Seidel decomposition method to compute matrices inverse in SIC receiver and a comparison between traditional LU decomposition method and Neumann series approximation and Gauss-Seidel decomposition methods investigated. Due to the SIC receiver method where users with the worth channel only detect their signal and users with the best channel, first have to detect the signal with more allocated power and then detect their signal, the BER for the user with more power is better than the user with less power. Also, with using matrix inverse calculation methods to reduce system complexity, there is a slight reduction in system performance. So, we can say that the Neumann series method and the gauss-seidel decomposition method for calculating matrices inverse, are two methods that we can use them in communication systems with less complexity. Also, we examined the effect of increase the δ^2 for channel state information in the SIC receiver and it was observed that increasing the amount of δ^2 for channel state information in the receiver, reduce the system performance slightly.

In general, since the goal of the paper was to reduce the receiver complexity, this was achieved, but the following suggestions can be used to make a tradeoff between the complexity and the performance of the system and improve the performance of the system

- Designing an optimal algorithm for clustering users with the least interference.
- Design and optimization a precoder to eliminate interference between clusters in order to provide the best performance for the system.
- Optimization of power allocation coefficients to reduce error probability.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, *et al.*, "What will 5G be?," *IEEE Journal on selected areas in communications,* vol. 32, pp. 1065-1082, 2014.

[2] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," IEEE Communications Magazine, vol. 53, pp. 74-81, 2015.

[3] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," IEEE signal processing letters, vol. 21, pp. 1501-1505, 2014.

[4] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," IEEE Wireless Communications Letters, vol. 4, pp. 405-408, 2015.

[5] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," IEEE Transactions on Wireless Communications, vol. 15, pp. 537-552, 2015.

[6] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Non-orthogonal multiple access (NOMA): Concept and design," Signal Processing for 5G, 2016.

[7] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," IEEE Signal Processing Letters, vol. 23, pp. 629-633, 2016.

[8] X. Chen, A. Beiijebbour, A. Li, H. Jiang, and H. Kayama, "Consideration on successive interference canceller (SIC) receiver at cell-edge users for non-orthogonal multiple access (NOMA) with SU-MIMO," in 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015, pp. 522-526.

[9] J. Kim, J. Koh, J. Kang, K. Lee, and J. Kang, "Design of user clustering and precoding for downlink non-orthogonal multiple access (NOMA)," in MILCOM 2015-2015 IEEE Military Communications Conference, 2015, pp. 1170-1175.

[10] S. Liu, C. Zhang, and G. Lyu, "User selection and power schedule for downlink non-orthogonal multiple access (NOMA) system," in 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 2561-2565.

[11] F. Rezaei, A. R. Heidarpour, C. Tellambura, and A. Tadaion, "Underlaid Spectrum Sharing for Cell-Free Massive MIMO-NOMA," IEEE Communications Letters, 2020.

[12] O. Gustafsson, E. Bertilsson, J. Klasson, and C. Ingemarsson, "Approximate Neumann series or exact matrix inversion for massive MIMO?," in 2017 IEEE 24th Symposium on Computer Arithmetic (ARITH), 2017, pp. 62-63.

[13] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," IEEE Journal of Selected Topics in Signal Processing, vol. 8, pp. 916-929, 2014.

[14] C. H. Rycroft, "Iterative methods for linear systems," School of Engineering and Applied Sciences, Harvard University, Cambridge, 2007.

Nilufar Tutunchi received the B.S. degree in Electrical Engineering from Azad University, Saveh Branch, Iran in 2014, and M.S. degree in Electrical Engineering from Azad University, Science and Research Branch, Tehran, Iran, in 2019. Her research interests include wireless communications and optimization.

Afrooz Haghbin received her Bs.C. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2001. She received her Ms.C. degree from Tehran University and her PhD degree from Tarbiat Modares University, Tehran, Iran, all in electrical engineering in 2004 and 2009, respectively. She is currently with the Faculty of Electrical and Computer Engineering at Science and Research Branch in Azad University, Tehran, Iran, as assistant professor. Her research interests include MIMO wireless communications, channel coding, precoding, multicarrier modulation, information theory and detection and estimation theory.

Behrad Mahboobi received BSC. Degrees in electrical engineering from K.N.Toosi University of technology, Tehran, Iran, in 2003 and his Master Degree from K. N. Toosi university of technology, in 2012. Since 2012, he has been assistant professor at Science and Research Branch of Azad University with major interest in convex optimization, resource allocation, game theory, information theory and queuing networks.

An Autonomic Software Defined Network (SDN) Architecture With Performance Improvement Considering

Alireza Shirmarz*
Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.
A.shirmarz@iau-tnb.ac.ir
Ali Ghaffari
Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
A.Ghaffari@iaut.ac.ir

Received: 04/Feb/2020 Revised: 09/Mar/2020 Accepted: 16/Apr/2020

Abstract

SDN makes the network programmable, agile, and flexible with data and control traffic separating. This architecture consists of three layers which are application, control and data. The aim of our research is concentrated on the control layer to improve the performance of the network in an autonomic manner. In the first step, we have categorized the performance improvement researches based on network performance improvement solutions proposed in the recent papers. This performance improvement solution clustering is one of our contributions to our paper. The significant contribution in this paper is a novel autonomic SDN-based architecture to ameliorate the performance metrics including blocking probability (BP), delay, jitter, packet loss rate (PLR), and path utilization. Our SDN-based autonomic system consists of three layers (data, autonomic control, and Route learning) to separate the traffics based on deep neural networks (DNN) and to route the flows with the greedy algorithm. The autonomic SDN-based architecture which has proposed in this paper makes better network performance metrics dynamically. Our proposed autonomic architecture will be developed in the POX controller which has developed by python. Mininet is used for simulation and the results are compared with the commonly used SDN named pure SDN in this article. The simulation results show that our structure works better in a full-mesh topology and improves the performance metrics simultaneously. The average performance is improved by about %2.5 in comparison with pure SDN architecture based on the Area Under Curve (AUC) of network performance.

Keywords: SDN; Performance; Autonomic System; Resource Management; QoS.

1- Introduction

The number of internet users is growing rapidly, so the network resources will be restricted for Internet service providing in the future. According to this Internet growth, different architectures have been proposed in various scrutinies [1]. The resource shortage can give rise to the service provisioning problem in the networks and causes network equipment configuration complexity as mentioned in [2]. Based on the future network requirements, SDN has been proposed as a programmable architecture [3]. SDN is an architecture that has been investigated in many papers [4][5][6][7]. SDN architecture consists of three layers (APP, Control, and Data) and three APIs (Northbound, Southbound, and East-West) which are shown in Fig.1, briefly.

Data plane is composed of FEs which is simple forwarding elements, control plane ha the role of decision making in SDN and all FEs send their flows' first packet of flows in the packet-in message to the controllers, the controllers impose the flow entries with suitable action to the Fes flow tables. The controller can be single centralized or conceptual centeralized. The Conceptual centralized controller is composed of some controller which are related together with east-west APIs. The data plane and control plane are connected by southbound API. The application layer is based on network application and is connected with the controller with northbound APIs.

* Corresponding Author

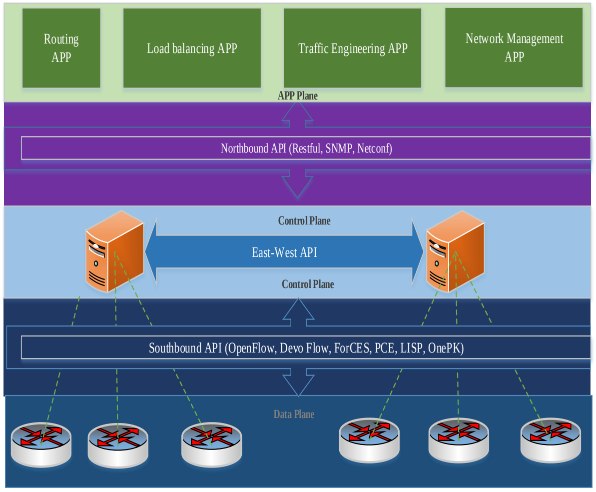


Fig.1 SDN Architecture [8]

We have surveyed the performance improvement in SDN and published our research in [8]. According to our published survey, the solutions for SDN performance improvement have been categorized into three groups that are DC & Cloud, Wireless, and WAN. The SDN-based network performance metrics which have been used are delay, throughput, delay, jitter, packet loss rate (PLR), blocking probability (BP), flow rate (FR), bit error rate (BER), and energy. The main question in this paper is how to improve software defined network's performance in an autonomic manner? Which we solve this problem with network performance improvement approaches categorization in SDN and a novel autonomic SDN-based architecture to cover our problem and improve the network performance metrics simultaneously.
In this paper, we investigate the solutions and propose an SDN-based architecture to make dedicated performance metrics improved. Finally, this our proposed architecture's advantages and drawbacks will be discussed in comparison with other ones

2- Related Works

The performance improvement in SDN is a problem which has been paid attention in [8]. The paper has extracted performance metrics and solutions in three dedicated SDN-based networks. This paper is the continuing work of that survey which was published. The other solutions for Quality of Service (QoS) in SDN have been investigated in [9]. The authors have proposed a new SDN-based architecture to provide QoS for application with the use of SDN capabilities. Omar Aldhaibani et al have suggested an SDN-based architecture for the Handover (HO) decision in Wi-Fi environments in [10]. They have used SDN to make hand-over smart in IEEE 802.11 wireless LAN and proved that the QoE (Quality of Experience) has been improved. Feng et al have analyzed the applications of SDN to different types of wireless networks. They have discussed the performance improvement in SDN-based wireless networks and presented future direction in SDWN (Software Defined Wireless Network) [11]. T. Shozi et al have worked on an SDN-based overlay solution for the existing traditional Wide Area Networks (WAN) environment, targeting provisioning flexibility and control in case of network failures, through a distributed SDN network overlay and edge SDN devices [12]. The goal was to facilitate the transition to SDN in developing economies through the adoption of SDN while retaining legacy infrastructure. We present an SDN overlay on top of the existing WAN using the South African National Research Network (SANReN) network as a use case. Kleinrouweler et al have worked on Dynamic adaptive streaming over HTTP (DASH) which is a simple, but effective, technology for video streaming over the Internet in [13]. They proposed an SDN-based architecture for DASH-aware networking that also enables internet service providers, network administrators, and end-users to configure their networks to their requirements. In paper [14], Autonomic QoS Management Mechanism in Software Defined Network has been proposed to improve QoS with open flow protocol modification. G. Poulios et al have proposed an SDN-based architecture for Self-Organization networking (SON) in LTE [15]. Pedro Neves et al have worked on 5G and suggested an architecture with SDN capabilities in [16]. These papers have proposed SDN-based architecture in a dedicated use case, so we analyze the performance improvement solutions in SDN to propose an innovative architecture to improve network performance with the SDN concept.

3- SDN Performance Improvement Solutions Analysis

In this paper, in continuing our survey in [8], we analyze the researches to extract performance improvement in SDN and according to the recent researches, we can divide the solutions in a tree which is shown in Fig 2 in summary. The main resource which is used to control the network performance is bandwidth. These solutions are divided into two significant taxonomies which are Link Capacity Expansion, and bandwidth allocation management which are the goals of different researches to ameliorate network performance metrics.

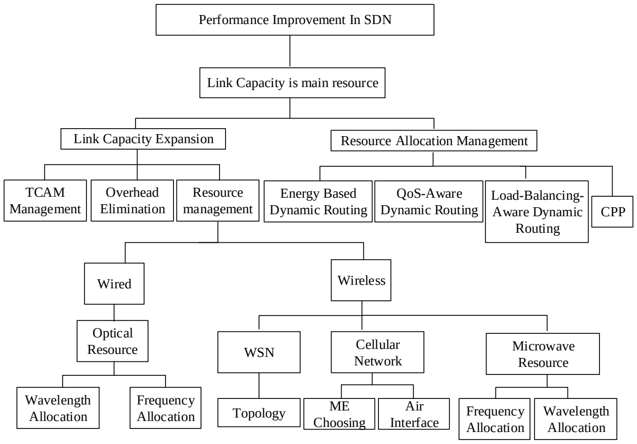


Fig. 2 Performance improvement solutions taxonomy

3-1- Link Capacity Expansion

The expansion of link capacity is the aim of some researches that seek to develop bandwidth capacity with TCAM management, overhead elimination, and link resource management.

- TCAM management: This solution concentrates on memory space management which causes performance improvement. Switches, routers, and Forwarding Elements (FEs) use the ternary content-addressable memory (TCAM) which is fast in reading/writing speed, but this memory type is expensive.
- Overhead Elimination: This solution is used to eliminate the overhead of reading/writing protocols in southbound API. The most southbound protocols which have been used in the researches are open-source which is open flow.
- Resource Management: The links in the network can be wired or wireless. The resources which have been used in wired media are optical one and the researchers increase bandwidth with frequency and wavelength management. In wireless media, the resources are different based on network applications. In WSN, topology is managed, in the cellular network, air interface and ME selection are managed, and in microwave-based resources like frequency and wavelength are used to control.

3-2- Bandwidth Allocation Management

In this solution, the researchers work to manage resources with awareness. The awareness types discriminate the solutions.

- Energy-based Dynamic Routing: this solution tries to route the traffic with energy consumption consideration, so the energy will be improved.
- QoS-aware Dynamic Routing: this solution routes the flows based on QoS metrics consideration. This method betters QoS.
- Load-balancing Aware Dynamic Routing: this method routes the flows based on the load in FEs and Servers. This one decreases congestion and improve performance measures.
- Controller Placement Problem (CPP): this way works on improvement in performance measures with controlling of the number of controllers and controllers place.

This section analyzes the solutions which have been proposed for performance improvement in SDN to design the architecture for performance improvement in SDN. In the following, our proposed architecture will be suggested.

4- Proposed Architecture for Autonomic Performance Improvement in SDN

This architecture is inspired by the IBM architecture which is presented in [17]. This architecture makes the network autonomic and consists of four main elements as shown in Fig 3. This model is used as a reference model in our proposed architecture.

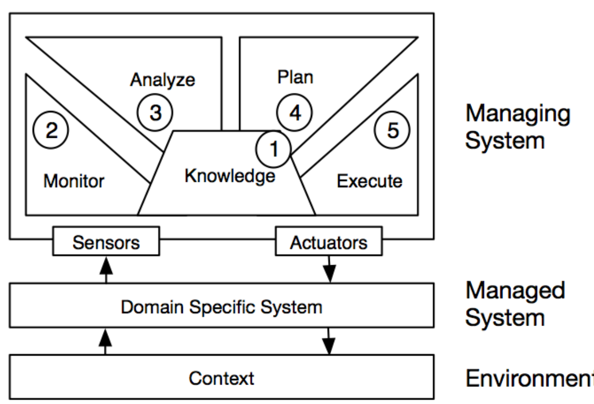


Fig. 3 Reference Model for Autonomic System [17]

- Knowledge: this is a database that is collected from the system and makes the required knowledge to make the autonomic system.
- Monitor: this is composed of the monitoring process
- Analyze: this section is the module for analyzing the system to make a plan for the system
- Plan: this module makes the plan for execution

- Execute: this module imposes the plan to the system

SDN makes the network programmable and agile. It also causes the network flexibility. We propose an architecture that is composed of SDN and MAPE-K for performance improvement. Our proposed architecture should be an autonomic one to support the performance metrics in ISPs and network service providers.

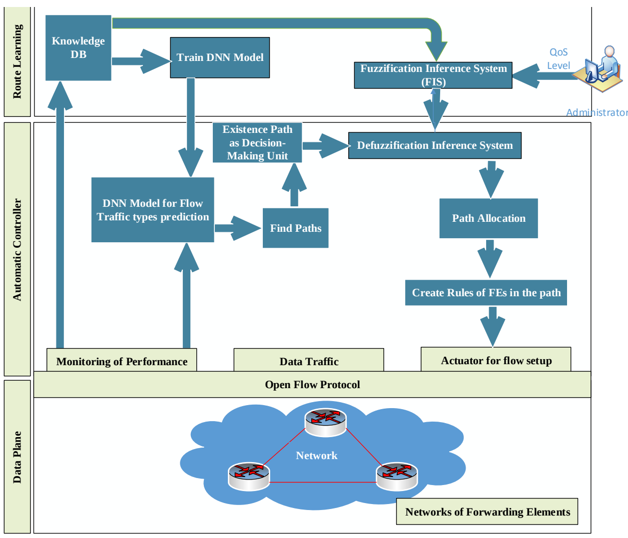


Fig. 4 Proposed autonomic architecture for performance improvement in SDN

This autonomic system provides the performance requirements automatically and involuntarily. The performance metrics include blocking probability (BP), delay, jitter, packet loss rate (PLR), and Path utilization which are monitored and analyzed in our scheme.

- Blocking Probability (BP): this metric shows the percentage of blocked flow requests in the controller.
- Delay: this metric indicates the delay of the path which contains propagation, processing, queuing, and transmission delay
- Jitter: it is the variance of delay
- Packet loss rate (PLR): this is the percentage of packets lost in one second.
- Path Utilization: this metric shows the percentage of path capacity which has been used in the network.

QoS separates the traffics in the link layer, network layer, and transport layer to make sure that the expected quality of applications will be provided by the network with a variety of required resources. On the contrary, network performance should be set out independently and without awareness of applications and services, to a dependable network.

Our proposed architecture has three layers which are data, Autonomic control, and route learning layers.

- Data layer: this layer indicates the data which is passing through the network
- Autonomic control layer: this layer contains the controller with some modules to make the controller as an autonomic one.
- Route learning layer: this layer gathers the information and trains the model periodically till makes the machine-learning flow discriminator more accurate.

In the next sub-section, the modules of each layer will be expressed.

4-1- Data Layer

In this layer, packets send the first packets of the flows and the non-dedicated rule packets are sent to the controller as a packet-in message format of open flow ver1.3 as mentioned in Fig 5.

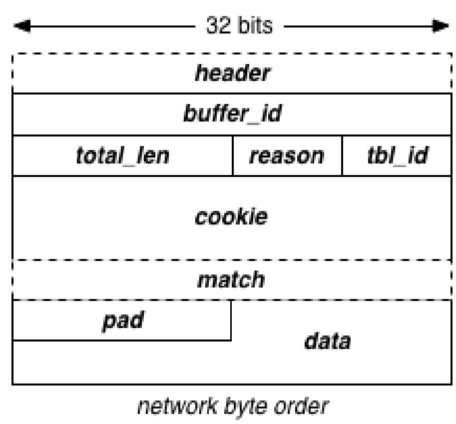


Fig. 5 Packet-in format in Open flow ver1.3 [18]

The flow is a sequence of packets that have the same source and destination. The unique source and destination are specified by the unified source and destination MAC, IP, and destination port number. The controller, based on the flow information routes the flow path that will be discussed in the controller subsection. The other information which is needed to collect will be sent to the route learning layer for knowledge database collection to increase the training samples.

4-2- Autonomic Control Layer

This layer has 4 main modules and one common module which has discussed here.

- DNN Model for traffic types' discrimination: this module used a deep neural network (DNN) to classify the traffic types automatically. This module makes our architecture independent from the applications. This model needs the training set which had been gathered in the route learning layer and trained there before. This Module is implemented using the paper [19].

- Find Path Module: a module is needed to find the path with source and destination information that had been extracted from the packet-in message. This module constructs the existing paths in which there are between source and destination.
- Path Allocation Module: This module allocates the proper path to the flow requests with the performance considering. All existing paths are sorted based on the performance metrics and select the path with the highest performance.
- Create Rules for the FEs: the selection path needs the configuration of all FEs which exist in the path, so this module makes ready all flow entity for specified FE. Finally, all configurations will be imposed on all FEs.

The autonomic controller has been implemented in our work in [20]. In the next subsection, the Route learning layer will be discussed. The tasks which have been don is shown in Algorithm 1.

Algorithm.1: Steps of routing based on the greedy algorithm in [20]

Algorithm 1: The greedy adaptive flow routing algorithm

1. Input→ F as the requested flows set
2. Sort F based on B^f descending
3. For each flow in F
   Begin for F
4. Find all existing paths between current flow source and destination with Findpaths function
5. Sort the paths of the current flow ascending
6. For each path in paths
   Begin for paths
7. **If** the current path supports the $B^f_{src,dst}$ and $D^f_{src,dst}$ and $J^f_{src,dst}$ and $L^f_{src,dst}$ as the current flow performance thresholds
   Begin **if**
   Allocate the current flow to the current path and return the current path
   Goto the next flow if exists
   End if
   Else
   Goto another path if exists, otherwise return the null value for the current flow
   End Else

   End for paths
   End for F
8. Output→ a path or null for each flow in F

4-3- Autonomic Control Layer

This layer has two significant roles in our proposed architecture, at first, it is used to information collection and model training, and the second is a channel for communication with network admins.

- Training Role: the required information is gathered from the data layer and stored in a PostgreSQL database and the data is analyzed and store the number of packets in each flow, the number of bytes in each flow, the number of blocked flow, and so on. The suitable action is tagged and trained with Moore traffics which is a reference dataset in ten days [21].
- Network Admins Interface: this module communicates with the human being, so we use fuzzy logic to show the network performance situation in 5 degrees which are very bad, bad, average, good, and very good. These states show the state of the performance of the networks linguistically. The other option in the admin interface is the prioritization between BP and QoE (Quality of Experience) that should be defined by the network admins. This option can prioritize that if there is bandwidth capacity resource more than required and the network situation is good or very good the network can improve the quality of experience or decrease the BP. It depends on the policy of networks.

This algorithm is shown in Algorithm 2.

Algorithm 2 Training algorithm [19]

1. **Input**→X_L: The labelled training set
   X_U: The unlabeled training set
2. Randomly initialize the stacked autoencoder
3. **For** $X_{U1} \in X_U$, do
   Calculate the output of the first hidden layer $y(X_{U1})$
   Calculate $z(X_{U1})$ by formula
   End for
4. Compute the weight and bias of the first hidden layer by minimizing $LS(\theta_1)$
5. **For** i=2 to M, do
   Use the output of the (i-1)th hidden layer as the input of the ith
   Hidden layer to train the ith hidden layer by minimizing $LS(\theta_1)$
   End for
6. Randomly initialize the softmax regression layer
7. Train softmax regression layer to achieve the minimization of $LS(\theta_2)$ by using the output of the final hidden layer as the input
8. Initialize the hybrid deep learning network with obtained Parameters through the pre-training process
9. Fine-tune the hybrid deep network to minimize $LSS(\theta)$ by BP algorithm
10. Output→ θ: The weights and bias of all layers

This architecture is developed in the POX controller to show its behaviour in comparison with the pure SDN one. Three scenarios are defined to show that our architecture has better performance in comparison with pure SDN.

The architecture of a deep learning-based application classifier is shown in Fig 6.

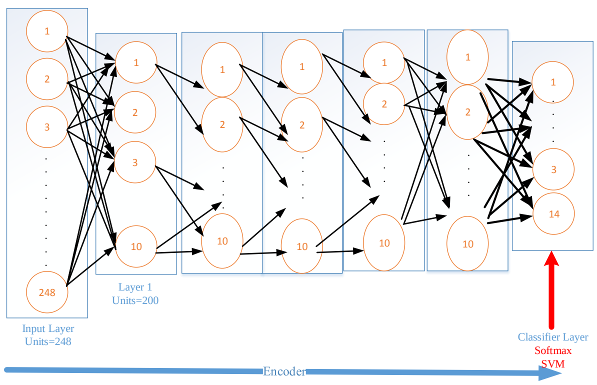


Fig. 6 Application Classifier based on Deep Learning Architecture

5- Experiments and Results

Our proposed architecture is implemented in the POX controller and Mininet. The requirements for simulation are presented in Table 1.

Table 1: Simulation Requirements

Hardware/ Software	Framework	specs
Hardware	CPU	Intel(R) Core(TM)-i7-2.4GH
	Ram	12 GB
	Hard	2 TB
Software	OS	Windows 10-64bit
	Programming language	Python 2.7
	Programming IDE	Spyder 3.3.3
	Deep learning software	Tensorflow/python
	Deep learning library	Keras
	Machine Learning library	scikit-learn v0.21.3
	Controller	Pox 0.2.0

The simulation runs for 15 min with 10 times iteration. The average BP, delay, jitter, PLR, and utilization will be considered and normalize with the Eq.(1).

$$Normalize(x) = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Three scenarios are defined with a different number of nodes and links which will be defined in the following subsection.

5-1- Simulation Scenarios

Three scenarios are defined to show the application of our proposed architecture with performance improvement considering.

- Scenario-1: This scenario is full-mesh one which has specs and the limitation value which is selected randomly as mentioned in Table 2.

Table 2: Full-mesh scenario

Specs	Limitation of value
The number of nodes	9
The number of links	36
The number of flow requests	40
Link speed	10-100 Mbps
Link delay	10-120 ms
Link jitter	10-20ms
Link PLR	1%-7%

- Scenario-2: This scenario is partial-mesh one which has specs and the limitation value which is selected randomly as mentioned in Table 3.

Table 3: Full-mesh scenario

Specs	Limitation of value
The number of nodes	9
The number of links	24
The number of flow requests	40
Link speed	10-100 Mbps
Link delay	10-120 ms
Link jitter	10-20ms
Link PLR	1%-7%

- Scenario-3: This scenario is sparse one which has specs and the limitation value which is selected randomly as mentioned in Table 4.

Table 4: Full-mesh scenario

Specs	Limitation of value
The number of nodes	9
The number of links	9
The number of flow requests	40
Link speed	10-100 Mbps
Link delay	10-120 ms
Link jitter	10-20ms
Link PLR	1%-7%

All values are assigned randomly in each scenario between min and max value which have been determined in Tables 2, 3, and 4.

5-2- Evaluation

For evaluation, we do the algorithm which is shown in Algorithm 2.

Algorithm 2: The evaluation of two architectures

1. Input→Pure SDN | Proposed Architecture
2. Run simulation for 15 minutes
3. Sampling every 5 seconds the performance metrics
4. average of BP, delay, jitter, PLR, utilization for 15 minutes
5. Normalize the performance metrics
6. complementary of utilization is calculated
7. Draw the AUC for 5 performance metrics
8. Calculate the area of the curve for each scenario
9. Output→ comparison of Proposed architecture and pure SDN

To evaluate the proposed autonomic SDN-based architecture which has been developed in POX as the controller and Mininet as an emulator the AUC is used. AUC considers all performance metrics that should be minimized and utilization which should be maximized. We use complementary utilization that can minimize. The AUC is used to show performance improvement in our proposed autonomic SDN-based architecture in collation with pure SDN with OpenFlow spanning tree in POX to prevent a loop.
The average results of simulation in five performance metrics are shown in Fig 7, Fig 8, and Fig 9 as the AUC chart for scenario-1, scenario-2, and scenario-3 respectively. All metrics should be minimized, but the utilization should maximize, so the complementary of utilization is considered as 1-U.

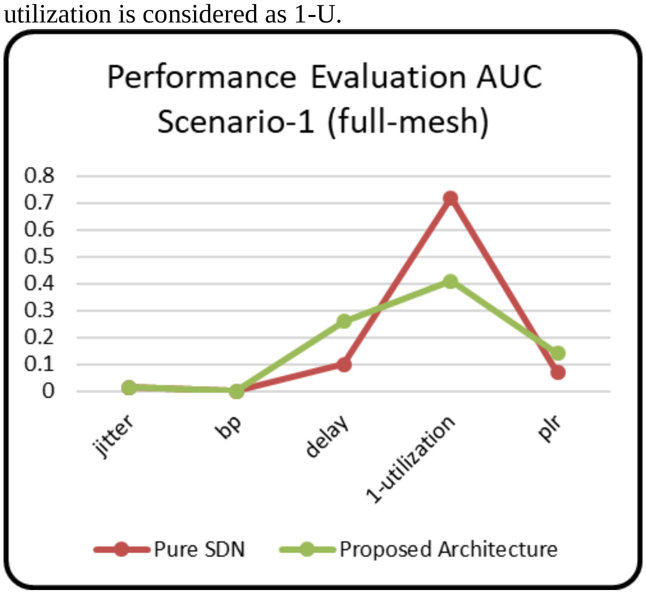


Fig. 7 Average of performance metrics in scenario-1 (full-mesh)

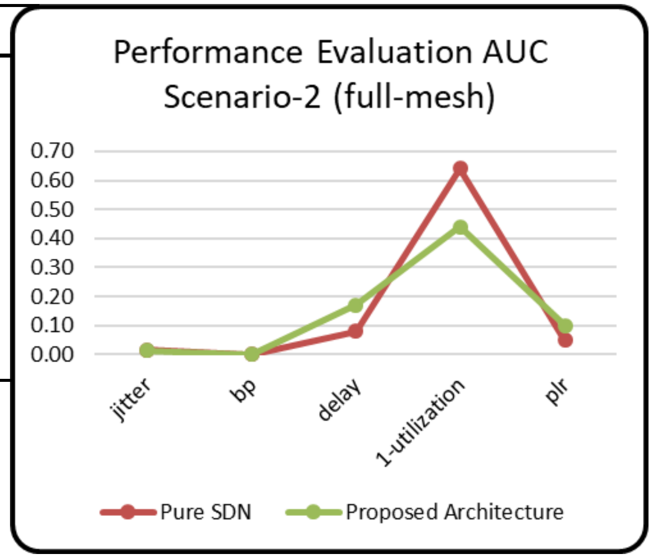


Fig. 8 Average of performance metrics in scenario-2 (partial-mesh)

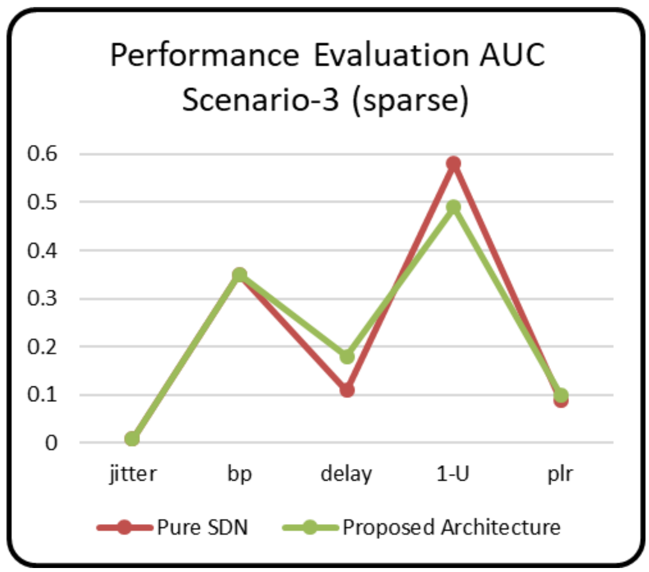


Fig. 9 Average of performance metrics in scenario-3 (sparse)

The area in each scenario is measured and presented in Table 5.

Table 5: Area under the curve for scenario1, 2, and 3

	Scenario-1	Scenario-2	Scenario-3
Autonomic SDN	0.412	0.362	0.564
Pure SDN	0.452	0.392	0.569

The area under curve shows that the proposed architecture has better performance in comparison to the pure SDN.

6- Experiments and Results

SDN is an architecture that makes the network programmable and agile. The autonomic model is a reference model that has been proposed by IBM to make the systems autonomic. In this paper, we designed a new architecture that is composed of SDN and IBM autonomic model. The momentous aim of our research is an SDN-based architecture to control and improve network performance metrics. The performance metrics which have been used in this paper are blocking probability, delay, jitter, packet loss rate and utilization. This problem is an optimization one that should be optimized. Delay, jitter, and packet loss rate should be minimized, but the utilization should be maximized.

Our proposed autonomic SDN architecture has three layers, including the data layer, autonomic control layer, and Learning route learning layer. The data layer is the same as in pure SDN, but the other two layers have been developed in the POX controller with version 0.2.0. Our autonomic controller routes the flows based on the greedy algorithm. The proposed routing algorithm maximizes utilization and minimizes the other performance metrics. The other layer which is route learning collects data in PostgreSQL and analyzes the number of bytes, packets and time to transmission. This layer trains a deep neural network model for flow discrimination. The model updates periodically to make the model more accurate. Our proposed architecture is used in comparison with POX OpenFlow.spanning_tree module to tackle the loop in the network. To evaluate our autonomic controller which makes the performance of the networks improved, we use the area under the curve to show that all 5 metrics are optimized. Three scenarios with 9 nodes are assumed in this research with different numbers of links which are 36, 24, and 9 called full-mesh, partial-mesh, and sparse respectively. The results of simulation show that our proposed SDN architecture has more (minimized) optimized performance in comparison with pure SDN. According to the simulation results, the area under curve difference with pure SDN is increasing with moving from the full-mesh scenario towards the sparse scenario. Our proposed architecture works better in the network with more links, and with the decrease of links the effect of our architecture is declining.

References

[1] J. Pan, S. Paul, and R. Jain, "A survey of the research on future internet architectures," Commun. Mag. IEEE, vol. 49, no. 7, pp. 26–36, 2011.
[2] H. Qi and K. Li, Software Defined Networking Applications in Distributed Datacenters. Dalian, China: Engineering, SpringerBriefs in Electrical and Computer, 2016.
[3] N. Feamster, J. Rexford, and E. Zegura, "The Road to SDN: An Intellectual History of Programmable Networks," ACM Sigcomm Comput. Commun., vol. 44, no. 2, pp. 87–98, 2014.
[4] R. Masoudi and A. Ghaffari, "Software defined networks: A survey," J. Netw. Comput. Appl., vol. 67, no. May, pp. 1–25, 2016.
[5] A. K. Singh, "A survey and classification of controller placement problem in SDN," Int. J. Netw. Manag., no. December 2017.
[6] S. Bera, S. Misra, and A. V. Vasilakos, "Software-Defined Networking for Internet of Things: A Survey," IEEE Internet Things J., vol. 4662, pp. 1–1, 2017.
[7] Y. E. Oktian, S. Lee, H. Lee, and J. Lam, "Distributed SDN Controller System: A Survey on Design Choice," Comput. Networks, vol. 121, pp. 100–111, 2017.
[8] A. Shirmarz and A. Ghaffari, "Performance issues and solutions in SDN-based data center: a survey," J. Supercomput., 2020.
[9] A. T. Oliveira, B. Jos, C. A. Martins, M. F. Moreno, and A. B. Vieira, "SDN-Based Architecture for Providing QoS to High-Performance Distributed Applications," in IEEE Symposium on Computers and Communications (ISCC), 2018.
[10] O. Aldhaibani, F. Bouhafs, M. Makay, and A. Raschellà, "An SDN-based Architecture for Smart Handover to Improve QoE in IEEE 802. 11 WLANs," in 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2014, pp. 287–292.
[11] I. Engineering, "Enhancing the performance of future wireless networks with software-defined networking ∗," Front. Inf. Technol. Electron. Eng., vol. 17, no. 7, pp. 606–619, 2016.
[12] T. Shozi, P. Mudali, and O. Matthew, "An SDN Solution for Performance Improvement in Dedicated Wide- Area Networks," in 2019 Conference on Information Communications Technology and Society (ICTAS), pp. 1–6.
[13] J. W. Kleinrouweler, S. Cabrero, and P. Cesar, "Delivering Stable High-Quality Video : An SDN Architecture with DASH Assisting Network Elements Categories and Subject Descriptors," in Proceedings of the 7th International Conference on Multimedia Systems, 2016.
[14] W. Wendong, Q. I. Qinglei, G. Xiangyang, H. U. Yannan, and Q. U. E. Xirong, "Autonomic QoS Management Mechanism in Software Defined Network," China Commun., no. July, pp. 13–23, 2014.
[15] G. Poulios, K. Tsagkaris, P. Demestichas, A. Tall, Z. Altman, and C. Destré, "Autonomics and SDN for self-organizing networks," 11th Int. Symp. Wirel. Commun. Syst. ISWCS 2014 - Proc., pp. 830–835, 2014.
[16] P. Neves et al., "The SELFNET Approach for Autonomic Management in an NFV/SDN Networking Paradigm," Int. J. Distrib. Sens. Networks, vol. 2016, no. 4, 2016.
[17] IBM, "Autonomic Computing White Paper: An Architectural Blueprint for Autonomic Computing," IBM White Pap., no. June, p. 34, 2005.
[18] V. W. Protocol, "OpenFlow Switch Specification," 2012.
[19] C. Zhang and X. Wang, "Deep learning-based network application classification for SDN," Trans. Emerg. Telecommun. Technol. Wiley Online Libr. J., no. February 2018.
[20] A. Shirmarz and A. Ghaffari, "An adaptive greedy flow routing algorithm for performance improvement in a software-defined network," Int. Numer. Model. Electron. networks, Devices, Fields-Wiley online Libr., no. March, pp. 1–21, 2019.
[21] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian Neural Networks for Internet Traffic Classification," IEEE Trans. Neural Networks, vol. 18, no. 1, pp. 223–239, 2007.

Alireza Shirmarz received his BS, MS, and P.hd degree in computer engineering, IT, and Computer system architecture from Tehran Shahed, Tehran Polytechnic, university respectively. He has worked on software defined networking and performance. His research interests include Network, Recommendation systems, AI, Data mining and Data Science.

Ali Ghaffari received his BSc, MSc. and Ph.D. degrees in computer engineering from the University of Tehran and IAU (Islamic Azad University), TEHRAN, IRAN in 1994, 2002 and 2011 respectively. As an associate professor of computer engineering at Islamic Azad University, Tabriz Branch, IRAN, his research interests are mainly in the field of software defined network(SDN), Wireless Sensor Networks (WSNs), Mobile Ad Hoc Networks(MANETs), Vehicular Ad Hoc Networks(VANETs), networks security and Quality of Service (QoS). He has published more than 60 international conference and reviewed journal papers.

Context-Based Expert Finding in Online Communities Using Ant Colony Algorithm

Mojtaba Sharifian
Department of Electrical, Computer and IT Engineering, Qazvin Islamic Azad University, Qazvin, Iran
sharifian68@gmail.com
Neda Abdolvand*
Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran
n.abdolvand@alzahra.ac.ir
Saeedeh Rajaee Harandi
Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran
Sa.Rajaeeharandi@gmail.com

Received: 04/Dec/2019 Revised: 24/Feb/2020 Accepted: 08/Apr/2020

Abstract

Online communities are the most popular interactive environments on the Internet, which provide users with a platform to share their knowledge and expertise. The most important use of online communities in cyberspace is sharing knowledge. These communities are a great place to ask questions and find answers. The important challenges of these communities are the large volume of information and the lack of a method to determine their validity as well as expert finding which attracted a lot of attention in both industry and academia in. Therefore, identifying persons with relevant knowledge on a given topic and ranking them according to their expertise score can help to calculate the accuracy of the comments submitted on the internet. In this research, a model for finding experts and determining their domain expertise level by the aid of statistical calculations and the ant colony algorithm in the MetaFilter online community was presented. The WordNet Dictionary was used to determine the relevance of the user's questions with the intended domain. The proposed algorithm determines the level of people's expertise in the intended field by using the pheromone section of the Ant colony algorithm, which is based on the similarity of the questions sent by the users and the shared knowledge of the users from their interactions in the online community.

Keywords: Online Communities; Experts Finding; Ant Colony Algorithm; Word Net.

1- Introduction

Online communities are among the most important achievements of Web 2.0 technologies that have been noticed by researchers and business organizations due to their large volume of valuable raw data [1]. The most important use of online communities in cyberspace is sharing knowledge. These communities are a great place to ask questions and find answers [2] the unique features of these communities such as ease of access and lack of time and space constraints for accessing them make them as one of the most important sources of problem solving [3]. Since, there is no distinction between the levels of the users' knowledge in these communities; there is uncertainty about the value of the answers and comments submitted in these communities. In fact, the questioner does not know how much he/she can trust the answers sent by other users [4]. Considering the growth of information generated by users in question-and-answer forums (QA forums) and the need for speed and precision in finding the true answer, finding expert people in is necessary. Therefore, the important issue is finding the level of expertise of people in different fields as well as the level of users' trust in experts to use their answers in various issues '[2]. Given that only a small fraction of users are responsible for answering a significant number of questions, it is difficult to find potential experts [4]. Therefore, by using expert advisory methods, especially the recommender systems questions can be provided to knowledgeable people [5].

Expert finding is a challenging problem that has been considered in both industry and academia in the past few years. Identifying people with relevant knowledge about the issue and ranking them according to their expertise is a challenging task
[6-7].

In general, there are two main approaches to expert finding: the content analysis of user messages and analysis of social networks between people. Each has some flaws. For example, in the social network analysis, the content of messages sent by users is not considered. In the content analysis approach, communication between people is not

* Corresponding Author

considered and there is no difference between the answers given to people with different levels of expertise, while the knowledge of a person responding to an expert user is different from that of a user responding to a beginner [8].
Given that there are many queries in QA forums, and there is no way to determine the accuracy of the answers, users cannot trust the answers [4]. Therefore, the most important issue is how to identify and evaluate the experts as well as their level of expertise in professional communities with high precision and trust.. Hence, this study aims to rank the experts in different fields based on their answers as well as to determine their performance in QA forums. Therefore, for the first time in this field, using the Ant colony algorithm and considering the quality of communications and users' trust and reputation, a new model is proposed, which is a good solution to solve problems and challenges of online communities with high accuracy
The research literature is reviewed first; then, the research method and the proposed algorithm are presented. Finally, the conclusion and recommendations for future studies are expressed.

2- Literature Review

2-1- Expert Finding Systems

Expert finding systems are part of Computer Supported Cooperative Work (CSCW) systems, and one of the most important classes of recommender systems [9]. Given the importance of expert finding in both industry and academia, many approaches came into existence [10]. In general, two main approaches are used for expert finding, including content analysis and social network analysis. The former focuses on analyzing the content of user messages to find experts in online communities. In this approach, text-mining techniques and data retrieval algorithms are used to analyze messages sent by users. Then, based on the information extracted from the message, a user knowledge model is created using user's knowledge modeling techniques or a probabilistic model of the communication between the user and the messages is created. Expert people can be identified by using the knowledge model or probabilistic model. For example, Liu et al. [11] have identified experts using the data retrieval algorithms.
In latter, the features of the social network structure are used to estimate the people expertise. In this method, based on the relationships between people in a network, a graph is made [12], and the diffusion method [13] and the similarity of users [5] are used to weigh the edges and people, and sometimes the PageRank algorithm is used to score nodes and edges [2-13,14].
The study of Yuan et al. [15] the recent solutions for expert finding were classified into four different categories of matrix factorization based models (MF-based models), gradient boosting tree based models (GBT-based models), deep learning based models (DL-based models) and ranking based models (R-based models). The results of the study indicated that MF-based models outperform other categories of models in the crowd-sourcing situation. However, their results indicated that different algorithms can perform well on different data sets and different problems.
The study of Dadkhaha et al. [16] used business intelligence tools to find potential experts as well as the analytic hierarchy process for assigning weights to both selection criteria and potential experts.
The study of Zhang [17] has tested a set of network-based ranking algorithms including Indegree, AnswerNum, Z-Score, PageRank, and HITS. These simulations have been used to find a few simple simulation rules governing dynamic network queries and answers. Kardan et al. [2] have also proposed a SNPageRank method for the accurate and comprehensive evaluation of the experts' knowledge. Their proposed method was specifically implemented and tested on FriendFeed's social network. However, this method was unable to identify experts in different areas. Zhao et al. [18] have proposed a method to improve the performance of expert finding in CQA systems using users' social networks. They also developed two iterative procedures, GRMC-EGM and GRMC-AGM, to solve the optimization problem. Their proposed algorithm had better performance in comparison to the state-of-the-art expert finding algorithms.
Besides, Cifariello et al. [19] proposed an unsupervised search engine for expert finding in academia. This system combines classical language modeling techniques, based on text evidences, with the Wikipedia Knowledge Graph, via entity linking. In another study, Rampisela et al. [20] used the thesis abstract and fasilkom UI students' metadata to find experts in the Faculty of Computer Science in Universities Indonesia. They used combination of word2vec and doc2vec to model semantic information, and indicated the comparable performance of the embedding models in retrieving experts using expertise queries in both Indonesian and English languages.
Given that people's posts are in different fields, they may have different levels of knowledge. So, Stephens-Martinez et al. [14] proposed the CEF method by changing the PageRank algorithm to run online forums in the CEF. The CEF algorithm can determine the level of users' knowledge in each domain. The relationship between topics was also considered in this study. Omidvar et al. [21] have proposed a novel algorithm to determine people's expertise level in AskMe forum using WordNet dictionary and social network analysis. Their proposed method outperforms other expert finding algorithms. Fu et al. [22] proposed a model that considers the relationship between users as well as the link between topics to

determine the expertise of each user. In addition, Brusilovsky and Millán [13] identified the people who had the highest rank in the list of experts. Then, they compiled a list of experts by calculating the level of communication of other people on the web pages or their emails with these experts. The algorithm used in this study was the diffusion algorithm, an important feature of which is the selection of some sample class of people. The problem with this research is it ignores the dynamic change of users who have been selected as experts. The two mentioned approaches for finding experts are incomplete. For example, in the social network analysis, malicious users can expose their communications network by sending numerous unrelated or empty messages, and the algorithm automatically detects these users as experts. Moreover, in the content analysis approach, communication between people is not considered, and there is no difference between answers given to people with a different level of expertise. Therefore, in some studies, a combination approach has been used to find the experts. For example, Karimzadegan et al. [10] by using the hierarchical relationships within the organization have tried to identify the similarities of people in the organizational environment and improve their ranks. Their method was more helpful to new users who have not had much work so far. But they did not considered the topic similarity between users. In order to solve this problem, Ziaimatin et al. [23] have proposed a model that uses the PageRank algorithm. They indicated that the relevance and proximity of people who interact in different fields is of paramount importance. Zhou et al. [24] have proposed a different method that, in addition to counting the number of user answers in the expert network, also specifies the number of answers to the topics. Then, using the PageRank algorithm, calculates the similarity between the questioner and the respondent. The evaluations carried out in this study indicated that the proposed model works better than the content based models introduced in earlier studies. Rafiei and Kardan [25] have also proposed a hybrid method for expert finding in java online communities, which was based on content analysis and social network analysis. They indicated the better performance of their proposed method on expert finding. In the study [26] the LDA label model was proposed to determine the distribution of the user subject. In this study, features like post content, answer votes, proportion of the best answers and user relationships were considered to find the correct answer to the new question. Moreover, EL-Korany [3] proposed a method that combines two content-based approaches and social network analysis and uses ratings and the reputation of users. In this model, user profiles are first created using content analysis methods, then user rating posts are used to compute user reputation and to combine this rating with the content analysis step. In addition, using network methods, user ratings are computed in the QA network, and the output of the two steps is combined. Their proposed method improved the accuracy of the expert finding. Wang et al. [27] have also proposed a novel expert finding algorithm, ExpertRank, which evaluates expertise based on both document-based relevance and one's authority in his or her knowledge community. They used PageRank algorithm to evaluate one's authority, and explored three different expert ranking strategies of linear combination, cascade ranking, and multiplication scaling that combine document-based relevance and authority. Their results indicated the better performance of the proposed algorithm than others. Gui et al. [6] proposed a new method for expert finding in heterogeneous bibliographical networks based on two aspects of textual content analysis and authority ranking. They indicated that the proposed framework outperforms existing methods.

2-2- Ant Colony Optimization Algorithms

Ant Colony Optimization (ACO) algorithms are constructive metaheuristics which inspired by the natural behavior of ants in finding the shortest paths between their nest and food sources by depositing pheromone on the ground in an iterative process. ACO uses the pheromone model to provide a set of solutions. The pheromone model is a set of values obtained by a reinforcement type of learning mechanism. Therefore, the pheromone model can be used to increase the probability of constructing good solutions from the components [28-29].

Different ACO algorithms have been used in the studies to solve the NP-hard problems. For example Anuradha et al. [30] used AntRank algorithm for ranking the web pages. In the same study, Setayesh et al [31] proposed a PageRank algorithm based on ACO. The results of simulation indicated that in their proposed algorithm, ratings are closer to real data, and more distinct ratings are generated.

As the research studies indicate, in earlier studies the Ant colony algorithm has not been used to find experts. In addition, all the basic algorithms for finding experts that are usually used for evaluation are not based on queries and just rank experts. The importance of using the Ant colony algorithm is that the existence of a possible component allows the ants to make various answers, and so they will answer more questions than greedy heuristics. In addition, the use of heuristic information, which, in many cases, helps the ants move towards more promising answers. Most importantly, the experience of ants during search by pheromone affects the answer construction in the next iterations of the algorithm and creates a dynamic learning process. In addition, the use of an ant colony gives more stability to the algorithm and the interaction of the population of agents (ants) together helps to solve the problem more effectively. Therefore, for the first time in this field, using the Ant colony algorithm and considering the quality of communication and users' trust and reputation, a new method is proposed, which similar to [32] methodology uses some features of users such as ranking of posts to improve the algorithm's accuracy.

3- Proposed Algorithm

In this research, a new algorithm called the Expertise AntRank (EAR) is presented to find experts in each domain using the MetaFilter dataset, an online forum for discussing various issues [33]. In this algorithm, a new approach is proposed to solve the existing challenges in finding experts using statistical calculations and ant colony algorithm in online communities. In addition to expert finding, this algorithm indicates the level of expertise in different fields. The proposed algorithm consists of four phases: I. data preparation, II calculating the share of users' knowledge, III calculating the level of attraction between people based on the field, IV implementing the ant colony algorithm.

After data preparation, the Bayes' theorem is used to obtain the share of each user's knowledge who participated in the question and answer process, which its application is changed based on the conditions of this research. WordNet's dictionary, a processing program based on the psychology of language rules and encodes concepts as sets of synonymous words, is also used to determine the relevance of user-submitted questions to the domain of interest, [34]. In addition, changes were made to the PageRank algorithm in order to calculate the degree of attraction between people by considering the activity records and the similarity of these records with the users requested field. Finally, inspired by the pheromone part of the ant colony algorithm, users are treated as a route, questions are considered as ants, the level of attraction between people in the specialized network is treated as innovative information, and knowledge share of users in each question is considered as pheromone. . At the end of the ants' tour, the experts are identified and ranked by the amount of pheromone placed on each path.

Phase I: Data Preparation

The data preparation is done through the extract, transform and load (ETL) process. First, the required information is extracted from the HTML files pertaining to the MetaFilter website (Metafilter.com). Next, the extracted data is examined for any incorrect entries, and all users whose total number of questions and answers is less than 20 are deleted from the database. In addition, all related information, such as the answered questions and corresponding answers, all unanswered questions, and all questions that have not been assigned a keyword are deleted from their records. Then, the cleaned data is uploaded to the MetaFilter online database. Table 1 shows statistical information about the data in the designed database.

Table 1. Statistical data of the MetaFilter online community

Number of Questions	Number of Answers	Number of Intended Answers	Number of Tags	Number of Users
229401	543699	2412653	951118	43676

Phase II: Calculation of User Knowledge

In online communities, each user can ask a new question, which will be answered by different users. Each question may receive more than 10 answers, all of which are somehow correct. But, the best answer is always chosen, which indicates that the proposed solution is more efficient and better. Once the best answer is given to a question, and the process of knowledge transfer ends, and the distribution of knowledge or the calculation of the share of each person's knowledge in response to the particular question is completed. To calculate the share of each user's knowledge in the QA process, the likelihood probability relationship based on the Bayes theory is adopted, which is modified based on the research conditions (equation 1):

$$K_{U_i P_j} = \frac{Votes(\mathrm{U}_i.\mathrm{P}_j)+1}{\sum_{k=1}^{N} Vote(P_{jk})+N} \quad (1)$$

Where $K_{U_i P_j}$ is the knowledge gained by each user i in the question or post j, and $Votes(\mathrm{U}_i, \mathrm{P}_j)$ is the votes of the user i in question j. In addition, $\sum_{k=1}^{N} Vote(P_{jk})$ is the total number of votes earned in all (N) questions and answers.

For example, as table 2 indicates a question that asked by a user received 5 positive votes. Moreover, there are also 4 answers with different scores., The knowledge share of each user who contributed to this QA process is calculated using equation 1 (see Table. 2).

Table 2. An example of user knowledge calculation

Type of Post	Number of Votes	Best Answers	Users' share of knowledge
Question	5	-	0.24
Answer	4	Yes	0.2
Answer	1	No	0.08
Answer	0	No	0.04
Answer	10	No	0.44

At the end of this phase, a user-post table is created, in which the knowledge share of each person is specified from each post. The information of this table is used as a pheromone in the ant colony algorithm.

Phase III: Calculation of the Attractiveness

In the PageRank algorithm, nodes are web pages that are connected by page links [35]. In this study, nodes are memers of the online community instead of web pages, and and links are based on the responses sent between them. Then according to the questions and links among people

, the Chief Executive Network (CEN) is drawn [36]. At CEN, each user is linked to those who answered his/her questions, and each link is given a certain weight depending on the number of answers and the field of expertise. User who answers another user's question is thus deemed to have more information on that subject. Therefore, a link is drawn from the question to the respondent, by the questioner validates the respondent. To determine the relevance of a question to a subject, the distance between all the labels of that question to the intended field is calculated using the OSS distance function. The OSS function takes two concepts as inputs, and returns a numeric range of zero to one using the WordNet ontology. The closer the output number is to one, two concepts are more similar to each other [15]. Then, the average distance of the labels is considered as the relevance of the question to the subject and the weight of each edge is calculated using equation 2:

$$W_{AB} = \sum_{p=1}^{N_{AB}} \left(\frac{\sum_{t=1}^{N_p} (\mathrm{Distance}(T_t.C))}{N_p}\right) \quad (2)$$

Where N_{AB} is the number of answers of user B to user A, N_p is the number of the labels of question P, C denotes the domain of the desired expertise, T is its label, and distance is the OSS distance function that is used to calculate the distance between field C and label T.

The graph proximity table is used to construct a possible transfer table. If there is a non-zero row in the graph proximity matrix, all the entries of that row are placed in the probabilistic transfer matrix as $1/_N$, where N is the number of members in the online community. Then, for the other rows, the value of each cell is calculated by using equation 3:

$$\frac{W_{ij}}{\sum_{z=0}^{z=N} W_{iz}} \quad (3)$$

Where, W_{ij} is the weight of the link from person i to person j. using Equation 3, the weighted average of each relationship is calculated with respect to other relationships of that person. After making changes to the proximity table, all table cells are multiplied by 1-a, where the value of a is the probability of a mutation. In the mutation, the searcher can jump to any node in the graph. The destination of the mutation is also randomly selected. If the number of nodes in the graph is N, the mutation may move the searcher to any node (even to the current node) in the graph with probability of 1/N.

Phase IV: Implementing the Ant Colony Algorithm

At this phase, inspired by the pheromone part of the ant colony algorithm, users are treated as a route, questions are considered as ants, the level of attraction between people in the specialized network is treated as the information, and knowledge share of users in each question is considered as pheromone. First, a weighted graph G = (N, A) is considered, in which N is a set of nodes or members of an online community with n = |N| members, and A is the set of edges. The dij is assigned to each (i, j) ϵ A edge, which indicates the level of attraction between members i and j. Ants can be considered as a possible construction procedure that produces an answer by moving on the graph G = (N, A). The ants do not move arbitrarily on the graphs, but rather use a construction policy that is subsidiary of the Ω constraints. The effect of pheromone T_i is considered for node i, which has a long-term memory of the search process and is updated by ants. A heuristic value n is also considered, which includes pre-problem information or information obtained at the time of problem solving from a source other than ants. Due to this feature, the algorithm, in addition to having an order based on ants' observation, also uses other desirable information that is outside of their observations to make decisions in finding more appropriate paths. The pseudocode of the metaheuristic behavior of the algorithm is given below:

```
Start the EAR metaheuristic process
Adjust the activity of the ants
Place each ant on a graph point
Place the initial value of the pheromone and calculate the heuristic values
Calculate the probability of an ant moving to the neighbouring points
Keep moving the ants to complete a tour for each ant
End setting activities
Update the pheromone
Pheromone placement based on the quality of the answer quality function
Pheromone Evaporation
End of pheromone update
End of the EAR's metaheuristic process
```

First, the variables and parameters of the algorithm, the number of program execution (termination condition), the initial pheromone value, α and β values, and the evaporation coefficient of the input are set. Then the probability of transition to neighboring nodes is calculated and the ants are randomly placed on the nodes by marking all the nodes as unappointed nodes. Then, each ant must be assigned to a primary city. To do this, we assign each ant to a random primary node. Finally, each ant makes a complete tour. At each step of answer making, the ants select the next node to navigate based on based on the constraint and the probability of transition. The only constraint on this issue is that all nodes should be met at most once. This constraint forces the ants to choose the next node through the unselected nodes during the answer construction process (that is, the possible neighborhood N^k_i of the ant k in node *i* and includes all nodes that have not yet met). When the ants are placed in randomly selected nodes, at each step of answer construction, an answer, the ant K uses a probabilistic selection law, called the relative random law, to select the next node. The probability of selecting node j by ant k which is located on the node i is calculated by using equation 4:

$$p_{ij}^{k}=\frac{[T_i]^{\alpha}[n_{ij}]^{\beta}}{\sum_{i\epsilon N_I^K}[T_i]^{\alpha}[n_{ij}]^{\beta}};\ if\ j\epsilon\ N_i^k \quad (4)$$

Where n is a heuristic value that is the inverse of users' desire (n=1/d_{ij}) (which was calculated in the third phase. T_j is the amount of pheromone on the node j at any moment. Besides, α and β are two parameters that determine the effect of the dependent pheromone route and the heuristic information. N_i^k is the possible neighborhood for the ant k, (the probability of selecting a node out of N_i^k is zero).

After calculating the transition probability of all neighboring nodes, the rotary method is used to select the next node to navigate. The rotary method generates a random number between zero and one to select each node, and the mentioned number corresponding node of each interval that the given number is located in will be selected.

When all the ants have made their tours, the pheromone routes are updated. In this way, first, the amount of pheromone is reduced from all nodes by a fixed factor (evaporation), and then the amount of pheromone that is calculated in the second phase is added on the nodes through which the ants passed in their tours. Pheromone evaporation is calculated by using equation 5:

$$T_i \leftarrow (1-p)T_i \quad (5)$$

Where the evaporation rate of pheromone is 0 > p <= 1, parameter p is used to prevent indefinite accumulation of pheromone pathways and enables the algorithm to forget the wrong decisions that have already made. If a node is not selected by ants, the amount of pheromone assigned to it decreases exponentially in repetitions. After evaporation, all the ants leave the pheromones on the nodes they have passed through their net. This means that the level of expertise decreases over time. The amount of pheromone that the ant k leaves on the nodes is calculated by using equation 6:

$$\Delta T_i = \begin{cases} \frac{K_i}{C^k}\ if\ node\ (i) belongs\ to\ T^K \\ 0 \qquad Otherwisw \end{cases} \quad (6)$$

Where, C^k, the fitness of the tour T^k that is made by the ant k, is calculated as the sum of the edge sizes belonging to T^k, and K_i is the share of user knowledge. The pheromone update is calculated by using equation 7:

$$T_i(t+1) = (1-p).T_i(\mathrm{t}) + \Delta T_i\ (t) \quad (7)$$

The condition for terminating the algorithm is the number of iterations that are taken from the input. The appropriate answer can be reached by repeating the algorithm several times and keeping the best answers. That is, at the end of the last iteration, a user-pheromone table is generated, which is sorted by pheromone columns to determine expert users and their level of expertise.

4- The Proposed Algorithm Evaluation

The proposed EAR method, along with other methods such as Indegree, Z-degree, Z-number, AnswerNum, ExpertiseRank and SNPageRank was run through MetaFilter's online community database.

The Indegree method is used to count the number of users that have helped others; Z-degree is the Z-score of the number of people whom the user has asked or answered; Z-number is the Z-score of the number of questions and answers that a user submitted; AnswerNum is the simplest way to identify experts by counting the number of their answers to the questions; ExpertiseRank provides the ranking of users in terms of decreasing the level of expertise in online communities. It counts the number of people who have helped the user as well as the number of users who have answered to each of these people [37]; SNPageRank is based on PageRank algorithm and is used for the accurate and comprehensive evaluation of the experts' knowledge in a social network [2].

The results are then compared under three different tests in three domains: travel, music, and internet. Fifty questions are selected for each field, ensuring that each question had an answer. The best answer for each question is specified by the person who originally asked the question. The selection of set of questions is done three times for each field, with the next steps being performed each time.

- First, questions with fewer than five answers are selected. This is done in such a way that expert-oriented approaches have a better chance of being selected by the appropriate respondent. In the next step, questions with 5 to 10 answers are selected. In the last step, questions with 10 or more answers are selected.
- The set of selected questions are then deleted from the database.
- All algorithms are run on a data set in order to rank the users.
- The best responder is selected for each question in the test based on the ranking of the algorithms. Finally, it is confirmed whether or not the selected participant gives the best answer.

The accuracy of each algorithm is calculated according to equation 8:

$$Accuracy = \frac{N_1}{N_1 + N_2} \quad (8)$$

Where N_1 is the number questions in which the best respondent is correctly selected, N_2 is the number of questions in which the best respondent is not selected correctly. According to the results, the proposed method has better results than other methods in finding the best respondent for each question. Figure 3 indicates the average accuracy of the proposed expert finding in three different test areas. As figure 1 demonstrates, the proposed algorithm with the average accuracy of 93.33% has better results in finding the best respondent than other algorithms in all three domains. Subsequently, two SNPageRank and ExpertRank algorithms yielded better results than other methods. Of course, the ExpertRank algorithm is able to detect the best music answer better than the SNPagerank algorithm. It is also able to identify the best responder for 35 of the 50 survey questions.

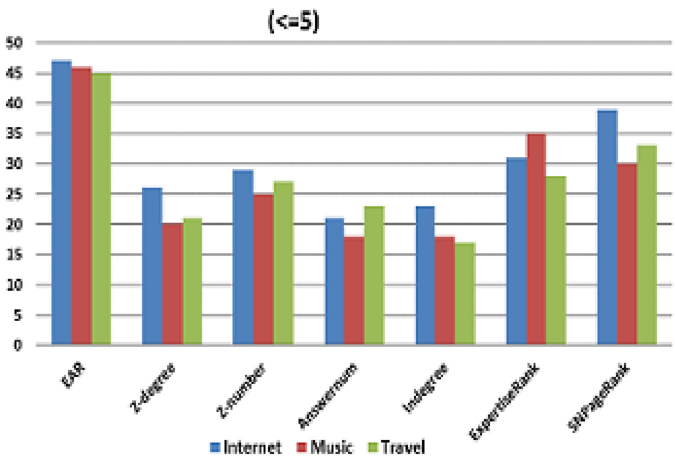


Figure 1. The results of the test for questions with 5 answers

Figure 2 indicates the accuracy of algorithms to find the best answer for questions with 5 to 10 answers. According to the figure, as in the previous test, the proposed algorithm is able to find the best respondent better than other algorithms in all three domains with the average accuracy of 74%. Of course, the results of the SNPageRank algorithm are close to the proposed algorithm. In addition, the results of Z-number method are better than those of ExpertiseRank algorithm and place in the third rank among the methods.

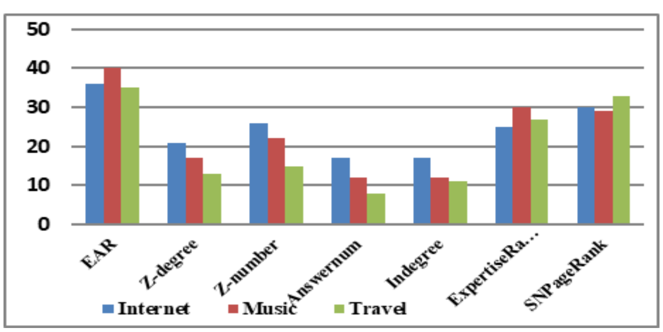


Figure 2. The results of test for questions with 5 to 10 answers

Figure 3 indicates the accuracy of the algorithms to find the best answer for the questions with more than10 answers. As with previous tests, the proposed method with the average accuracy of 50% has obtained better results than other methods.

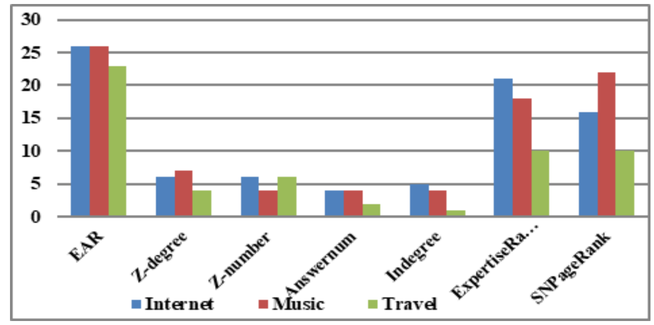


Figure 3. The results of the test for questions with more than 10 answers

Based on the results, the expert-oriented methods have better results in the first test, and the results of the second test are better than the third test. This is due to h increased number of respondents in this test compared to the first and third test. This reduces the probability of finding the best respondent in the expert-oriented approaches.

In the next step, the Spearman correlation method is used to compare the performance of proposed method with basic methods. To calculate the Spearman correlation value, n which is the raw score that people may have obtained in each subject, are presented by two symbols of X_i ($X_1 + X_2 + \ldots + X_n$) and Y_i ($Y_1 + Y_2 + \ldots + Y_n$). The value of d_i which is the difference between the ranks of each observation is then calculated as ($d_i = x_i - y_i$). In the absence of equal rankings, equation 9 is used [38]:

$$p = 1 - \frac{6 \sum i d_i^2}{n(n^2 - 1)} \tag{9}$$

Where X_i and Y_i are two points that user i received in two different rankings, and x_i and y_i are the rankings that person i received as a result of getting X_i and Y_i points respectively. As such, d_i is the difference in the ranks a user received in two subjects, and n is the total number of people.

In this evaluation, first all users who not answered any questions or their in-degree is equal to zero, are removed from the dataset. Then the first 1,000 users with a highest number of best answers are separated from the remaining users. The reason for this separation is that after observing the points earned by others, users with a rank of 1,000 onwards have similarly low scores. Thus, it can be said that the competition in expertise is among the first 1,000 users. Of these 1,000 users, 50 are randomly selected and Spearman correlation is calculated on them based on the total number of their best answers and output scores. The results indicate a correlation of 75.34%, which is a better result than other methods (Figure 4).

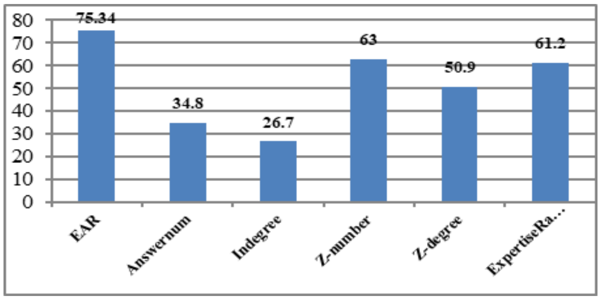


Figure 4. The accuracy of the proposed EAR method compared with basic methods of experts finding

One of the problems of the proposed method is its high computational complexity and its high temporal complexity compared to the basic methods. The reason for this complexity is that the proposed algorithm is context-based, meaning that all the basic expert finding algorithms are not query-based, and only rank experts. However, this advantage of the proposed method added the computational load of calculating the similarity between words to the algorithm.

5- Discussion and Conclusion

Today, people from all over the world can freely share their questions or opinions in online communities and social networks. Given that the number of answers to most questions is high and it is not possible to determine the accuracy of the received answers, online community users cannot trust on the answers of other users. Moreover, in some online communities such as MetaFilter, the expertise level of members is unclear. This requires the proper participation of all users, as well as the provision of accurate comments. In this study, the role and importance of online communities in knowledge sharing was discussed, and for the first time in this field, a model for finding experts and determining their domain expertise level by the aid of statistical calculations and the ant colony algorithm in the MetaFilter online community was presented. This algorithm uses the pheromone part of the ant colony algorithm to determine the level of people expertise in the intended field based on the similarity of users' questions with the studied field and the share of the users' knowledge of their interactions in the online community. The MetaFilter online community data was used to evaluate the proposed method.

The results of the proposed EAR method were compared with other methods like Indegree, Z-degree, Z-number, AnswerNum, ExpertiseRank, and SNPageRank in three different domains of internet, travel and music. Based on the results, the accuracy of the proposed method in finding the best respondent for each question was much better than other methods proposed by [2, 15-19]. In addition, the results of Spearman correlation indicated the correlation of 75.34% which was a better result than other methods. Therefore, the proposed method can be used in recommender systems, in such a way that people can submit their questions to the online community. Then, instead of exposing the question to everyone, it is suggested only to those who can provide appropriate answers.

The novelty of this research is the introduction of a new context-based expert finding method for and determining their level of expertise in each domain using the Ant colony algorithm in the MetaFilter online community. Since the WordNet dictionary has ontology, it was used to select the distance between concepts and to identify the background of members' knowledge. Moreover, the pheromone part of the ant colony algorithm was used to determine the level of users' expertise in each domain.

None of the methods used for expert finding can determine the level of individuals' knowledge in different fields. To achieve this goal, users' opinions should first be categorized and then the previous methods should be applied to different categories. Therefore, in order to change, improve and complete the identification model of experts provided in this research, a recommender system can be designed for online communities to answer any questions asked by users who have enough knowledge to answer them. Users can now see all the questions asked by other. Because the number of questions is too high, the large number of specialized questions that only a few users can respond, may remain unanswered, or take the long time to respond. Therefore, a recommender system could be designed and developed for QA forums to employ EAR algorithm for expert finding in order to show each question intelligently to users who are more knowledgeable for answering the question.

The biggest problem of a user in using online communities is the time it takes to review the entire content of a conversation, some of which reach more than hundreds of massages. Therefore, it seems desirable to summarize conversations in order to provide sufficient knowledge for user without losing important information. So far, various summarization methods have been proposed for multiple documents, but none of them consider the main feature of online communities, which is multiple writers. Therefore, the most important issue in generating a desire summary will be to identify experts and use their opinions in the summary form. In fact, the proposed method should be used for expert finding first, and then the summarization algorithm should be designed, which uses the writings of the experts in the process of summarizing.

Using concept map, a graph-like tool for knowledge representation, in each scientific domain, the semantic similarity between each term can be determined. These maps represent a meaningful relationship between the concepts of a subject, and can be used to determine the similarity of concepts. In fact, for specialized concepts such as Java programming language terms that are not found in the WordNet dictionary, a concept map created by experts for this programming language can be used. [28].

Moreover, some distance-definition functions, such as the method presented in [39] uses the meanings of the words in WordNet to determine the distance between concepts. Such methods can be used in the proposed EAR method to investigate their effects on the accuracy of the proposed algorithm.

In the proposed EAR method, the keywords of the questions are used to determine the relevance of the questions to the field of knowledge. Questions in the MetaFilter online community are tagged, but in many online communities such as YA, this does not matter. Therefore, keyword extraction methods can be used to tag questions and answers in online communities.

To improve the accuracy and efficiency of the proposed algorithm in finding experts, future studies might determine the difficulty of question by using factors such as time elapsed after asking a question and the number of answers, and add it as a coefficient to the equation of knowledge-level computing.

In addition, future studies can use other methods of ensemble intelligence and compare their results with the proposed algorithm. Besides, future studies can apply the proposed model to other online communities and compare the results.

References

[1] T., Lappas, K., Liu, & E, Terzi. "A survey of algorithms and systems for expert location in social networks." In Social Network Data Analytics (pp. 215-241). Springer, Boston, MA. 2011.

[2] A., Kardan, A., Omidvar and F., Farahmandnia "Expert finding on social network with link analysis approach." In *2011 19th Iranian Conference on Electrical Engineering*, pp. 1-6. IEEE, 2011.

[3] A., El-Korany. "Integrated Expert Recommendation Model For Online Communities." Arxiv preprint arxiv: 1311.3394. 2013.

[4] A., Daud, M., Ahmad, M. S. I., Malik & D., Che. "Using machine learning techniques for rising star prediction in co-author network." Scientometrics, Vol. 102, No. 2. 2015. pp. 1687-1711.

[5] L. A., Adamic, J., Zhang, E., Bakshy, & M. S., Ackerman. "Knowledge sharing and yahoo answers: everyone knows something." In Proceedings of the 17th international conference on World Wide Web (pp. 665-674). ACM. 2008, April.

[6] H., Gui, Q., Zhu, L., Liu, A., Zhang, & J., Han. "Expert Finding in Heterogeneous Bibliographic Networks with Locally-trained Embeddings." arXiv preprint arXiv: 1803.03370. 2018.

[7] M., Neshati, Z., Fallahnejad, & H., Beigy. "On dynamicity of expert finding in community question answering." Information Processing & Management, Vol. 53, No. 5. 2017. pp. 1026-1042.

[8] X., Ding, B., Liu & P. S., Yu "A holistic lexicon-based approach to opinion mining." In Proceedings of the 2008 international conference on web search and data mining (pp. 231-240). ACM. 2008.

[9] M., Karimzadehgan, R. W., White & M., Richardson. "Enhancing expert finding using organizational hierarchies." In European Conference on Information Retrieval (pp. 177-188). Springer, Berlin, Heidelberg. 2009.

[10] S. B., Sriramoju. "Heat Diffusion Based Search for Experts on World Wide Web." International Journal of Science and Research (IJSR), https://www. Ijsr. Net/archive/v6i11/v6i11. Php, Vol. 6, No. 11. 2017. pp. 632-635.

[11] X., Liu, W. B., Croft & M., Koll. "Finding experts in community-based question-answering services." In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 315-316). ACM. 2005, October.

[12] S., Ravi, B., Pang, V., Rastogi & R., Kumar. "Great Question! Question Quality in Community Q&A." ICWSM, Vol.14, No. 2014. pp. 426-435.

[13] P., Brusilovsky, & E., Millán. "User models for adaptive hypermedia and adaptive educational systems." In The adaptive web (pp. 3-53). Springer, Berlin, Heidelberg. 2007.

[14] K., Stephens-Martinez, M. A., Hearst & A., Fox. "Monitoring moocs: which information sources do instructors value?" In Proceedings of the first ACM conference on Learning@ scale conference, (pp. 79-88). ACM. 2014.

[15] S., Yuan, Y., Zhang, J., Tang, W., Hall, & J. B., Cabotà. "Expert finding in community question answering: a review." *Artificial Intelligence Review*, Vol. 53, No. 2, 2020, pp. 843-874.

[16] M., Dadkhaha, M., Lagziana, F., Rahim-niaa, & K., Kimiafar. "The potential of business intelligence tools for expert finding." *Journal of Intelligence Studies in Business*, Vol. 9, No. 2. 2019. pp. 82-95.

[17] J., Zhang, J., Tang & J., Li. "Expert finding in a social network." In International Conference on Database Systems for Advanced Applications, (pp. 1066-1069). Springer, Berlin, Heidelberg. 2007.

[18] Z., Zhao, L., Zhang, X., He & W., Ng. "Expert finding for question answering via graph regularized matrix completion." IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 4. 2015. pp. 993-1004.

[19] P., Cifariello, P., Ferragina, & M., Ponza. "Wiser: A semantic approach for expert finding in academia based on entity linking." *Information Systems*, Vol. 82 No.2019, 2019. pp. 1-16.

[20] T. V., Rampisela, & E., Yulianti. "Academic Expert Finding in Indonesia using Word Embedding and Document Embedding: A Case Study of Fasilkom UI. In *2020 8th International Conference on Information and Communication Technology (ICoICT)* pp. 1-6. IEEE. (2020, June).

[21] A., Omidvar, M., Garakani, & H. R., Safarpour. "Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis." Information Technology and Management, Vol. 15, No. 1. 2014. pp. 51-63.

[22] F., Yupeng, R., Xiang, Y., Liu, M., Zhang, and Sh., Ma. "Finding experts using social network analysis." In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 77-80. IEEE Computer Society, 2007.

[23] H., Ziaimatin, T., Groza, G., Bordea, P., Buitelaar, & J., Hunter. “Expertise Profiling in Evolving Knowledge curation Platforms.” GSTF Journal on Computing (JoC), Vol. 2, No. 3. 2018. pp. 118-127.

[24] G., Zhou, S., Lai, K., Liu & J., Zhao. “Topic-sensitive probabilistic model for expert finding in question answer communities.” In Proceedings of the 21st ACM international conference on Information and knowledge management (pp. 1662-1666). ACM. 2012.

[25] M., Rafiei & A. A., Kardan. “A novel method for expert finding in online communities based on concept map and PageRank.” Human-centric computing and information sciences, Vol.5, No. 1. 2015. pp. 10-28.

[26] H., Li, S., Jin & L. I., Shudong. “A hybrid model for experts finding in community question answering.” In Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2015 International Conference on (pp. 176-185). IEEE. 2015.

[27] G. A., Wang, J., Jiao, A. S., Abrahams, W., Fan & Z., Zhang. “Expert Rank: A topic-aware expert finding algorithm for online knowledge communities.” Decision Support Systems, Vol. 54, No. 3. 2013. pp. 1442-1451.

[28] Z., Dorrani, & M. S., Mahmoodi “Noisy images edge detection: Ant colony optimization algorithm.” Journal of AI and Data Mining, Vol 4, No 1. 2016. pp. 77-83.

[29] M. B., Dowlatshahi & V., Derhami. “Winner Determination in Combinatorial Auctions using Hybrid Ant Colony Optimization and Multi-Neighborhood Local Search.” Journal of AI and Data Mining, Vol. 5, No 2. 2017. pp. 169-18.

[30] G., Anuradha, G. L., Devi & M. P., Babu. “Antrank: An Ant Colony Algorithm for Ranking Web Pages.” International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol. 3, No. 2. 2014. pp. 208-212.

[31] S., Setayesh, A., Harounabadi & A. M., Rahmani. “Presentation of an Extended Version of the PageRank Algorithm to Rank Web Pages Inspired by Ant Colony Algorithm.” International Journal of Computer Applications, Vol. 85, No. 17. 2014. pp. 7-10.

[32] D. R., Liu, Y. H., Chen, W. C., Kao & H. W., Wang. “Integrating expert profile, reputation and link analysis for expert finding in question-answering websites.” Information processing & management, Vol. 49, No. 1. 2013. pp. 312-329.

[33] L., Silva, L., Goel & E., Mousavidin. “Exploring the dynamics of blog communities: the case of MetaFilter.” Information Systems Journal, Vol. 19, No.1. 2009. pp. 55-81.

[34] S. H., Hashemi, M., Neshati, & H., Beigy. “Expertise retrieval in bibliographic network: a topic dominance learning approach.” In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (2013, October), pp. 1117-1126.

[35] R., Navigli. “Word sense disambiguation: A survey.” ACM Computing Surveys (CSUR), Vol. 41, No.2. 2009. pp. 10-79.

[36] M., Rowe. “Mining User Development Signals for Online Community Churner Detection.” ACM Transactions on Knowledge Discovery from Data (TKDD), Vol. 10, No. 3, 2016. pp. 1-29.

[37] M., Neshati, D., Hiemstra, E., Asgari, & H., Beigy. „Integration of scientific and social networks.” World Wide Web, Vol. 17, No. 5, 2014. pp. 1051-1079.

[38] A. D., Well & J. L., Myers. *Research design & statistical analysis*. Psychology Press. USA. 2003.

[39] D., Petkova & W. B., Croft “Hierarchical language models for expert finding in enterprise corpora.” International Journal on Artificial Intelligence Tools, Vol. 17, No.1. 2008. pp. 5-18.

Mojtaba Sharifian holds a Master’s in Information Technology Engineering from the Islamic Azad University of Qazvin, Qazvin, Iran. He is interested in research in the field of information systems and technology, including innovation in ICT, business intelligence, electronic commerce, and big data analytics and cloud computing.

Neda Abdolvand is an Associate Professor in the Alzahra University and was a Postdoctoral Research Fellow at Tarbiat Modares University. She holds a PhD and MS in Information Technology from the Tarbiat Modares University and a Postgraduate Certificate in Information Systems from Melbourne University. She is interested in research in the field of information systems and technology, including innovation in ICT, business intelligence, electronic commerce, big data analytics and cloud computing. She has been a member of the Association for Information Systems since 2009.

Saeedeh Rajaee Harandi holds a Master’s in Information Technology Management from the Alzahra University, Tehran, Iran. She is interested in research in the field of information systems and technology, including innovation in ICT, business intelligence, electronic commerce, big data analytics and cloud computing.