Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 110-122
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Performance Analysis and Activity Deviation Discovery in Event Log Using Process Mining Tool for Hospital System

Shanmuga Sundari M[1], Rudra Kalyan Nayak[2*], Vijaya Chandra Jadala[3], Sai Kiran Pasupuleti[4]

[1]. Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation
[2]. School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114
[3]. Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation
[4]. Department of CSE, Prasad V Potluri Siddhartha Institute of Technology

Received: 31 Aug 2021/ Revised: 01 May 2022 / Accepted: 28 May 2022

Abstract
All service and manufacturing businesses are resilient and strive for a more efficient and better end in today's world. Data mining is data-driven and necessitates significant data to analyze the pattern and train the model. Assume the data is incorrect and was not collected from reliable sources, causing the analysis to be skewed. We introduce a procedure in which the dataset is split into test and training datasets with a specific ratio to overcome this challenge. Process mining will find the traces of actions to streamline the process and aid data mining in producing a more efficient result. The most responsible domain is the healthcare industry. In this study, we used the activity data from the hospital and applied process mining algorithms such as alpha miner and fuzzy miner. Process mining is used to check for conformity in the event log and do performance analysis, and a pattern of accuracy is exhibited. Finally, we used process mining techniques to show the deviation flow and fix the process flow. This study showed that there was a variation in the flow by employing alpha and fuzzy miners in the hospital.

Keywords: Alpha Miner; Event log; Fuzzy Miner; Hospital Process; Process Mining.

1- Introduction

In hospital systems, the activities are more volatile and dynamic and affect the efficiency and productivity that are the keys to survival and flourishing in the industry. Business Process Management (BPM) [1] is the basic foundation that efficiently produces the process, which will lead to more minor failures. BPM is an effective method to solve the emotional problems much early in the process activities. The root cause analysis [2] is the best technique to overcome the problem and improve process performance and efficiency. It gives visibility, accountability, and scalability in solving the operational issues and problems in the production or process.

Nowadays, many industries and enterprises are using business process management to construct their technique and improve the operations in process-centric flow, by extension, their efficacy in their field. Process Mining is the central part of the business process management approach. It will extract the input from the event logs collected from the hospital front desk and construct the process flow using this event log. The event log contains all traces of the activities in each flow. The process model is derived from the event log, and further, the model is used to analyze the different categories of the process. Process mining will streamline the related subsequent activities to help the process successfully and reduce the overhead of the process.

Process Mining [3] is the technique that explores all the activities in the process and trains the process model based on the dataset activities executed using the process mining algorithms. It shows the patterns bringing transparent results and delivers the graph with time constraints based on factual evidence and insights from the event log. These are automated process mining algorithms that can produce better and more accurate results when compared with the manual data mining of the process. It is being implemented as how the process ought to train a model.

The majority of the research in the literature is focused on organizational process mining. Medical treatment

✉ **Shanmuga Sundari M**
mshanmugasundari@gmail.com

processes refer to the clinical activities that are responsible for the care of patients. In contrast, administrative functions refer to the total activities carried out within an organization to ensure that successful medical treatment procedures are implemented.

On the other side, there is another set of underlying issues in the healthcare system that is mainly invisible and overlooked. The lack of consistent healthcare systems across the country is mostly to blame for these intangible issues. There is no defined technique to track and control the patient journey process, to put it another way. A patient's journey refers to the entire treatment process, which begins with a consultation with a doctor and ends when the patient has received all necessary therapy, and the case is considered finished. Because many patients are treated in specialized units creating a different process model for each care unit.

The use of process mining in a typical patient treatment procedure in a hospital is demonstrated in this study. We provide a viable solution to tackle most of the visible problems in the healthcare sector by solving the invisible problems utilizing process mining tools and techniques in this work. We used different algorithms to check the conformance of the process. The algorithm shows the more deviation that will find the time taking activity in the entire process.

2- Literature Review

Process mining is the best technique to find patterns [4] from the automatic extraction of the process models, which the process can implement to analyze the process. It is also used to define the conformance checking [5] with the business flow and identify the next level improvements. Business process analysis follows the below perspectives:

1. Data flow: Discover the data flow in the process and produce the pattern based on the data analysis.
2. Process flow: Discover the activities in the process and discover the model for the process; based on this, the process model is enhanced.

We are implementing the process flow in the hospital domain for the above insights with ProM Tool [6]. The information gained from the hospital is changed to the required format of process mining and then derives the conformance checking in emergency activities [7].

Van Der Aalst et al. [8] used the event log activities in the workflow process in water and road maintenance in the Netherlands to find the accuracy and measure the performance of the process from the organizational and case perspectives. The author significantly impacted the performance in terms of all views. He analyzed the outcome of the process in specific benchmarks of the domain variance. By using these activities, the author found abnormal behaviors.

R. Tripathy et al. [9] used a fuzzy C-Mean algorithm to identify the process's prediction. The root cause analysis is demonstrated and incorporated with the practical level. Process mining is used to enhance the process and rerun the flow to correct the process flow. The survey [10] showed that unplanned disruptions and uncertain changes would affect the performance of the process. Uncertainty of the changes will always affect the resource and operational issues in the industry.

All traces of activities in the process must be validated and identified based on the aspects of data evaluation [11]. The results are interpreted, and an assessment of the outcomes is deployed to furnish the following enhancement and future investigations.

The management lacks visibility in a few areas, such as process accuracy, quality, and resource utilization. The mining process will visualize the patterns with deviations which shows the data visualization as well as aiding exploration and automation [12]. The result provides the visualization of the process, and enhancement [13] will be carried out using these deviations to improve the process to the next level.

Petri net [14] is the graphical description of the activity flow and process flow structures, showing the dependencies between the activities [15]. The alpha algorithm proposed extends the process mining to the invisible tasks using the classification algorithm [16] in the hospital and completes the process.

Md Junayed Hasan et al. [17] explained the time-frequency imaging technique using machine learning classifications. It focused on the image's time frame and did the analysis. By evaluating the electroencephalogram (EEG) [18], data of polysomnography (PSG) recorded for three regions of the human brain, namely the prefrontal, central, and occipital lobes, a classification framework for automatic sleep stage recognition in both male and female human subjects was developed. The residual neural network (ResNet) architecture is used to automatically learn the characteristic features of different sleep stages from the raw EEG data's power spectral density (PSD) without using any artifact removal techniques. ResNet's residual block uses EEG data to learn the fundamental properties of various sleep stages while avoiding the vanishing gradient problem.

To obtain more accurate results, an optimum multilevel thresholding hybrid method is combined with Genetic Algorithm (GA) [19] and Particle Swarm Optimization (PSO), named HGAPSO, with an optimization strategy for classification based on grey level range. Process mining is another technique which works before all of the above machine learning methods.

3- Proposed System

The proposed system is a process mining tool to identify the event log from the hospital data, and change the event log into a structured format. The process mining tool will create the model based on the automated algorithms, and the base model will iterate with the event log and produce the different patterns, which will guide us to find the deviation in the process. Fig.1 shows the framework of our proposed model in process mining. The conformity checking is reviewed using the algorithms, and the process enhancement is concluded with the development of discovery model.

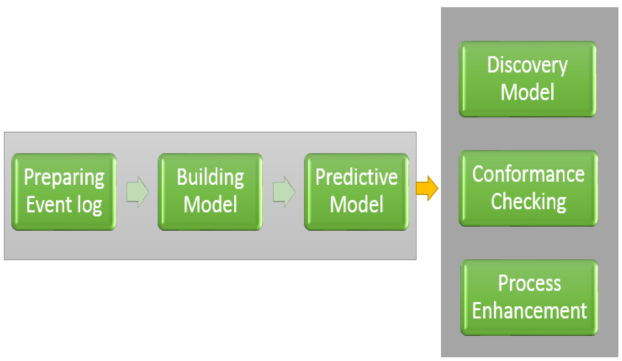


Fig. 1 Framework of proposed process mining

4- Process Mining Algorithms

4-1- Alpha Miner

An event log [20] consists of multiple traces. A trace [21] is a sequence of activity names abstracted from other attributes in the order of events.
The sample event log is:

$$T1 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^2, \langle a,e,d\rangle] \quad (1)$$

Eq. (1) gives the sample event log that explains the sequence of activities in the single trace $T1$. It is the order of series and number of times executed in the activity trace. The activities are *a*, *b*, *c*, and *d* in different patterns means the order of *a*, *b*, *c*, and *d* are executed three times, and *a*, *c*, *b*, and *d* are executed two times, and *a*, *e* and *d* executed one time. The primary purpose of the process mining is to retrieve good possible paths, for instance, that produce Petri net [22]. The Petri net starts the Activity with *a*, followed by *b*, *c*, and ends with activity *d*. Fig. 2 shows the Petri net of the above activity trace *T1*. Similarly, our hospital data set will evaluate the entire traces and analyze the pattern. The system is enhanced using the model-based result to refine the new model.

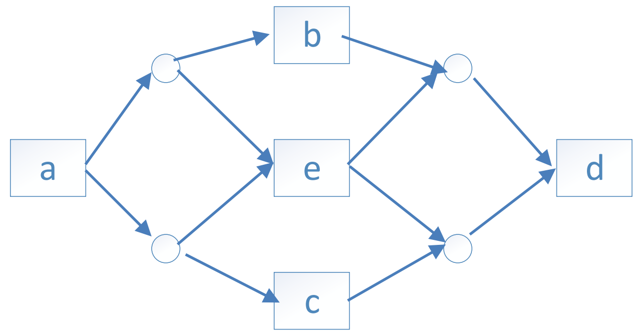


Fig. 2 Petri net for Trace T1

Fig. 2 explains the detailed Petri net of the trace *T1*. The graph starts with Activity a and ends with *d* activity. Based on the traces, the dependency [23] matrix is evaluated, and the order of the relation is calculated based on the dependency matrix.
Rule 1 (Predicting the followed Activity): In *the T1* trace, the flow of activities has to be traced based on the sequence order.
Rule 2 (Predicting entire Traces): In *T1*, the whole activities are followed until the end of the process, and the model is predicted based on all the activities.
Once the transition structure is obtained, the Concurrent Transition Sequence (CTS) is completed. Then the sub-trace log is analyzed for the entire flow, which helps find the trace's significance in the whole hospital dataset—the Petri net result in the reduced structure of input and output functions in the concurrent system.
In the below example, Trace *T2* has a few activity flows ranges *t1* to *t7*. From this, the sub trace is identified using the conditions below.
For example, *T2 = {< t1, t2, t3, t4, t6, t7 >, < t1, t2, t4, t3, t6, t7 >, < t1, t2, t5, t6, t7 >}*, and *N2* is a Petri net model with a concurrent structure. We have CTS= *{t3, t4, t5}*, *σ1|CTS =< t3, t4 >*, *σ2|CTS =< t4, t3 >*, *σ3|CTS = < t5 >*, and the sub-trace log is T_{sub} *= {< t3, t4 >, < t4, t3 >, < t5 >}.*

4-2- Fuzzy Miner

The Fuzzy Miner [24] is different from other algorithms since it will not produce a graph like the Petri net and causal net [25]. The Fuzzy Miner can work on the dynamic structure. The Fuzzy Miner is applicable for semi-structured data in the extensive database. The problem with process graphs is that we cannot differentiate between choice and parallelism. The complex and uncertain activities will happen in the actual time process. The unstructured process [26] is complicated to find the flow. The traditional process cannot support the unstructured process in outpatient discovery [27]. The model discovered in a fuzzy miner is a highly complex net,

and it is called a spaghetti diagram [28]. A fuzzy algorithm is a road map of metaphor.
When less structured activities are available in the dataset, it describes the view in high-level undesired details.
Activities based on the hospital department involved in the treatment and timestamps [29] connected to entering and exiting the event.
Activity in the event log has precedence [30] connection to the department and a timestamp.

Precedence Diagram: Let $W = (E, I, A, C, t, i, a, c)$ be an event log and $S = (N, L)$ and SPD. We say that $Sc = (W, S, la, ln)$ is a *connected SPD*, where $la: A \rightarrow P(N)\backslash\emptyset$ and $ln: NP(A)\backslash\emptyset$, such that for all $a \in A$ and $n \in N$ holds that $n \in la(a) \equiv a \in ln(n)$.

These precedence relations will provide the bonding between the activities identified as a significant role in the traces. The precedence matrix illustrates the value based on the connectivity of the activities.
The main goal of the fuzzy miner algorithm is to cluster the activities that are in the same sequence. These activities are further divided into multiple observations to sense the trace. The main idea behind fuzzy clustering is to find the similarity metrics of the activities. We can choose the number of clusters in the fuzzy miner algorithm to maximize the similarity and minimize the complexity.

5- Experimental Setup

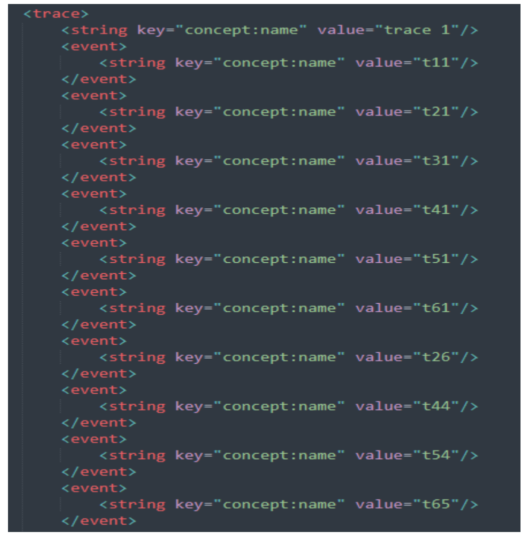
5-1- Dataset

We collected the 1000 Traces dataset [31] from the hospital. In the treatment of patients, each trace will have a variety of actions. Within the traces, the activities are denoted by the letters tXX. The XX values are assigned to each hospital department. We displayed a few department numbers in Table 1.

Table 1. Example Activities from the Event log

Sl. No.	Trace number	Department name
1	*t21*	X-Ray
2	*t26*	Scan
3	*t31*	Blood test
4	*t41*	Consultation
5	*t51*	Nurse care

The related departments are considered based on the patient's treatment. The activities flow will be different, and the trace is illustrated according to their treatment.



```
<trace>
    <string key="concept:name" value="trace 1"/>
    <event>
        <string key="concept:name" value="t11"/>
    </event>
    <event>
        <string key="concept:name" value="t21"/>
    </event>
    <event>
        <string key="concept:name" value="t31"/>
    </event>
    <event>
        <string key="concept:name" value="t41"/>
    </event>
    <event>
        <string key="concept:name" value="t51"/>
    </event>
    <event>
        <string key="concept:name" value="t61"/>
    </event>
    <event>
        <string key="concept:name" value="t26"/>
    </event>
    <event>
        <string key="concept:name" value="t44"/>
    </event>
    <event>
        <string key="concept:name" value="t54"/>
    </event>
    <event>
        <string key="concept:name" value="t65"/>
    </event>
```

Fig. 3 Example trace in the Event log

XES (eXtensible Event Stream) [28], applicable in the ProM tool. The event log's XML (eXentisible Markup Language) format is displayed in Fig.3, including all the traces of events. There are 1000 traces in the event log converted in the XML format. Then this format is changed [29-32].

Fig. 4 Event Dashboard

The dashboard of the activities is shown in Fig. 4. The dataset has 1000 cases and 15995 actions. After noise removal from the dataset, this is the XES file that will be ready to send as input to the process mining. The events are consolidated as a trace in the next step in the mining. The dataset has split 70:30 ratio for train and test the model. The dataset is divided into categories based on how the actions occur. The data is organized into traces, thought of as activity splits. With deviating traces and typical traces, the trace split has occurred.

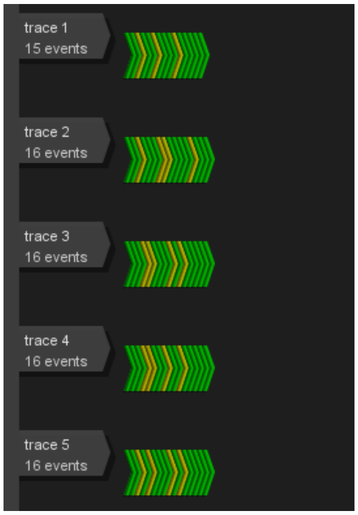


Fig. 5 Activity traces

Fig. 5 displays the activity traces. XML file is the input of the ProM tool then the tool changes the input file to XES format. This stream of activities is called traces which is shown in Fig.4. Hereafter, these traces are the input of the process mining algorithms. All 1000 traces are implemented and ready to apply as input in the algorithms. The green color lines are high-frequency events, and the yellow color lines are low-frequency events. All these events are connected and considered a single trace for a patent treatment flow. For example, in trace 1, 15 events have three low-frequency events. These low-frequency events have approximately 45 to 55 percent. The remaining other activities are 85 to 99 percentage frequency values.

Trace alignment can be represented as a matrix *T*= {activity *i*, *j*} Minimum number of traces <=*i*, *j* >= maximum number of traces. The possible traces can be satisfied in the above relation to form a trace alignment. If there are any gaps between the activities, they can be identified using $\Sigma' \;\; a \cup \{-\}$. The symbol represents the gaps between the activities that do not connect with other activities.

In this paper, the trace alignment is taken, combining all cases in the hospital data using a sequence trace approach. We can use the multiple activities in the event log to diagnose the sequence alignment of all activities. Numerous sequence traces are combined and produce a final trace diagram, as shown in Fig. 6. There are 1000 traces formed according to the sequences. We found one series with four traces of similar flow of lines in activities. Likewise, there are two sequences of 3 traces and three of 2 traces found in the sequence flow. The remaining traces occurred in one sequence flow. There are significant categories also mentioned in the traces. It is proposed to use this approach to find the deviations and understand the traces.

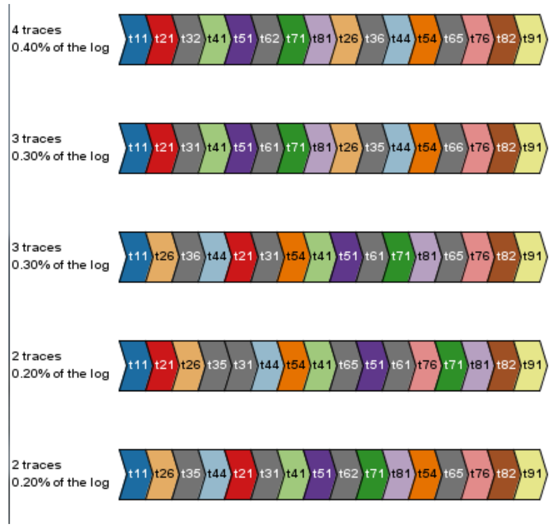


Fig. 6 Traces of Hospital data

In Fig. 7, we drilled down the first sequence flow with four traces. The same activity flow happened four times in the hospital sequence. Fig.7 shows the four activity sequences in the consequent order of activities. Likewise, the traces are designed based on a series of activities. In Fig. 7, the second trace has three similar activity flows. It shows the movements in a trace diagram that can help quickly understand all the flows in the hospital system.

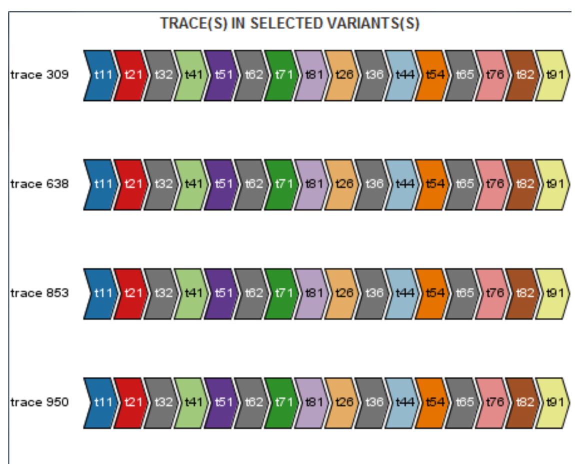


Fig. 7 Selected first trace

5-2- Pattern Evaluation

Fuzzy miner process mining: Conformance checking techniques can quantify the different deviations. For each deviation, the "process model may be wrong," or the "event log (i.e., real data) may be wrong."In the context of compliance and auditing, deviations are often considered undesirable. The term "normative model" refers to negative deviations. However, there are many possibilities of non-conforming activities in the traces, i.e., not affecting the performance of the process, called positive deviants or successful exceptions.

The term "positive deviance" refers to the uncommon process, but successful behaviors can be considered in the process diagram. Irregularities in processes are sometimes needed (like breaking glass to pull a fire alarm). Also, in hospitals, doctors save lives daily by deviating from the medical guidelines. However, flexibility does not imply that it is not valuable to investigate deviations and learn from them, e.g., to change procedures or enforce controls.

Fuzzy miner in hospital dataset:

In Fig. 8, event log traces are applied to the fuzzy miner algorithm that shows the flow of each Activity in the traces. The dark and bold line shows massive activity flow in the entire system, and the lighter color line shows less activity flow in the hospital. The more chances of bottleneck will be possible in the high activity movement path in the process.

This simplified process model shows the darker nodes representing the cluster of less significant activities. Every node in the diagram is rendered with a dependency matrix value. For example, node *t51* shows a value of 0.503, representing this trace comprises 50% of the total traces connected with this node *t51*. When the flow comes to *t54*, the value of *t54* is 0.523. Then the flow is split into two different traces like *t66 and t65*. The total value of 0.523 is divided into two different values that are 0.268 for *t66* and 0.254 for *t65*.

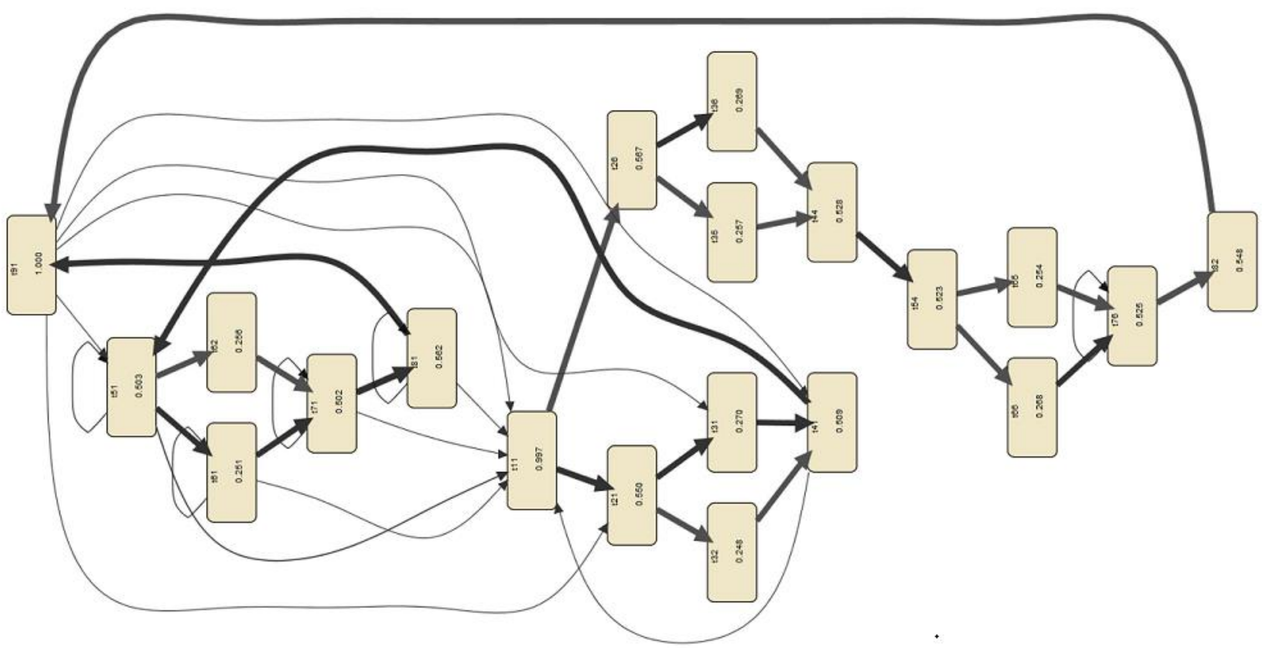


Fig. 8 Fuzzy miner for Hospital dataset

Fig. 8 shows a causal net flow diagram in the simplified process model. Bright square nodes represent significant activities, and the darker nodes are the primary node where all others are connecting through from this node. It is identified as a high significance node in the causal net diagram. All the nodes are denoted clearly with the name and probability of occurrences in the flow. The cluster node is more responsible for connecting all other nodes marked in a darker line. The edges are also labeled with proper correlation values and differentiated with color notation. When we click on the nodes, it will show the importance of the nodes, and we can visualize the connectivity between nodes.

Fuzzy miner is one of the algorithms to animate the process using the event log with various representations. The event log is the primary key point to generating an animation. If this is the case, the spirit will produce the results connecting with the other attribute relations. The features are more dynamic in the hospital dataset, which relies on the various changes and challenges in the resultant diagram.

Cluster in Fuzzy Miner:

Fig. 9 displays the most significant clusters in the hospital dataset. Node t11 is the starting node, and t91 is the ending node. In between, there are two clusters formed due to the lack of resources in the hospital. Cluster_27 was created for five element traces, and cluster_28 was developed for 13 element traces. It is an essential cluster to consider in critical resources in the hospital dataset. The process mining displays the timestamp deviation across the complete sequence of actions. Many patients have gathered around nodes 27 and 28, and they must wait for their acts. It could be due to a shortage of resources to attend the activity or a poor resource working culture. That is why the cluster forms between those two nodes. These two nodes require extra attention from management.

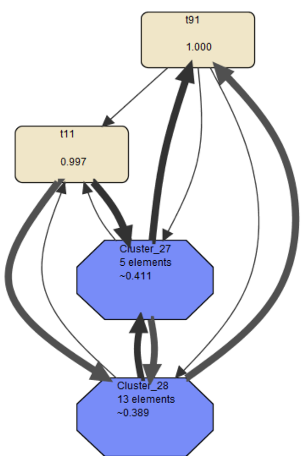


Fig. 9 Cluster of Fuzzy miner for Hospital dataset

Alpha Miner Process Mining:

Finally, Petri net diagram is drawn using the alpha miner algorithm shown in Fig.7. In Alpha miner, the event traces are connected using XOR and AND connectors. The possible flow of activities connects according to the flow of patient's treatment. Trace starting node with the patient registration at the front desk in the hospital and ending activity may differ based on the treatment.

In Fig.10, the basic structures are created with XOR and AND that use split and join conditions to generate a loop structure called a spaghetti graph in the Petri net. The connectivity with the activities is synchronized with non-exclusive conditions connected. The workflow of Petri net represents the business flow with the alpha algorithm. Petri net is a network that was initially considered static. Then the changes happened dynamically during the transition. But in the Petri net, all activities are considered tokens. So the reply tries to make the place changes in the node structure. We will try to replay the transition to eliminate deviation in the process. Circles in the diagram represent the places, and the squares represent the transitions. In this diagram, Activity is the base attribute. All features are designed according to the time stamp based on the base attribute remaining.

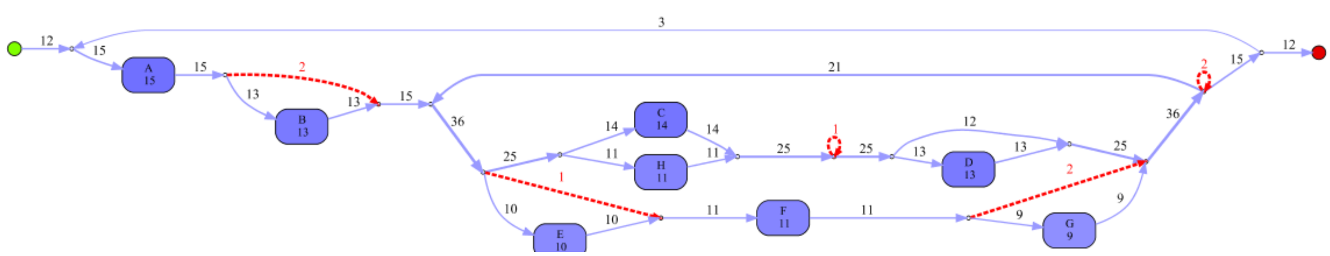


Fig. 10 Petri net with transition

The flow of state determines the waiting stage of activities. If the node is waiting for a particular resource long, that will create a low performance. First, transition results in a state with the red color mentioned as the conjunction of evolution is shown in Fig.10. If the network is static, then the movement is free to flow. The form in a Petri net is called the marking. From the initial level to the inactive status of transition has to be identified to make the node flow in a better chance. It is called reachable marking. That means it can access all activities without missing any action in the flow.

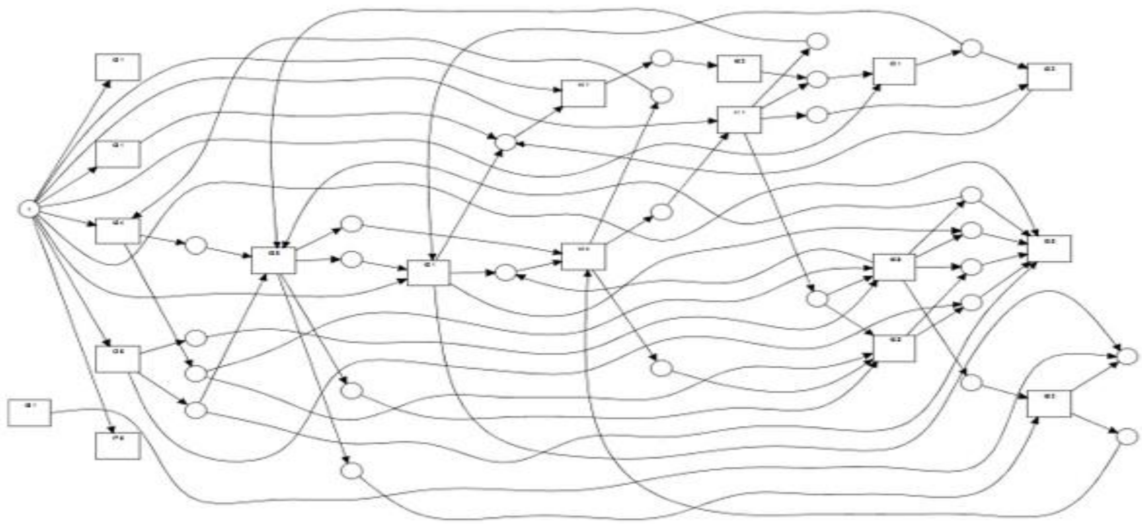


Fig. 11 Petri net diagram for hospital data

If t11 starts the trace from that, multiple other traces can connect based on patients' travel to the hospital. So by this trace, all activities illustrate that the connectivity of tokens may change. In this case, Petri net generates a complete behavior pattern that will express the dependency between nodes or activities. And that is what we need to travel the entire model in the Petri net and find the required abstraction of the Petri net. Now we can look a little deeper into the Petri net and analyze the complexity of the design.

The ProM plug-ins are used to analyze different abstract levels of the result in the hospital dataset. So, if we give the examples that will judge the attributes suitable for the tool and provide the suggestions.

Based on that, we can choose the visualization report, and then it will show the result in graphical representation. Some plug-ins are used to find the entire flow structure and replay activities that will repair the system in terms of enhancement. Sometimes the activities are divided into multiple activities and found whether the sequence flow is correct or not. Moreover, label splitting is also used to find the extensions and refinement approach. ProM uses these extensions to guarantee accurate results.

We displayed the nodes that indicate frequent activities flow, as shown in Fig.11. The previous Petri net offers the entire traces flow called a spaghetti diagram. Process mining can reduce this complicated flow further with a more significant activity sequence. ProM tool has the slider to move the frequency level from low to high. It can use this to identify the essential activities.

Petri net has a complete functionality diagram initially. Later, Petri net has refined and applied the abstract transitions with lower-level models to retain the integrated model of the entire workflow system. The model has been designed after obtaining all possibilities of activities flow, either top-level or bottom-level, from distributed running logs. The event log of an abstract procedure always has a few details: case id, time duration of the event, and the Activity that occurred in the particular time frame. The attributes are different in the event log, reflecting the result based on the features.

Inductive Alpha Miner:

The OR and XOR logic is used to construct the reason for Petri net in the hospital system. The case id and Activity are considered the key points based on which the Petri net is drawn. Fig. 11. And Fig.12 show the Petri net with the attributes of case id and activities followed by the patients in the hospital. Similarly, we add the post-set activities and can build a successor set. The relation between the activities is constructed, and a new model is determined. The casual net shows the transitions corresponding to the successor activities, after which the model and the relation are built.

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 110-122
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

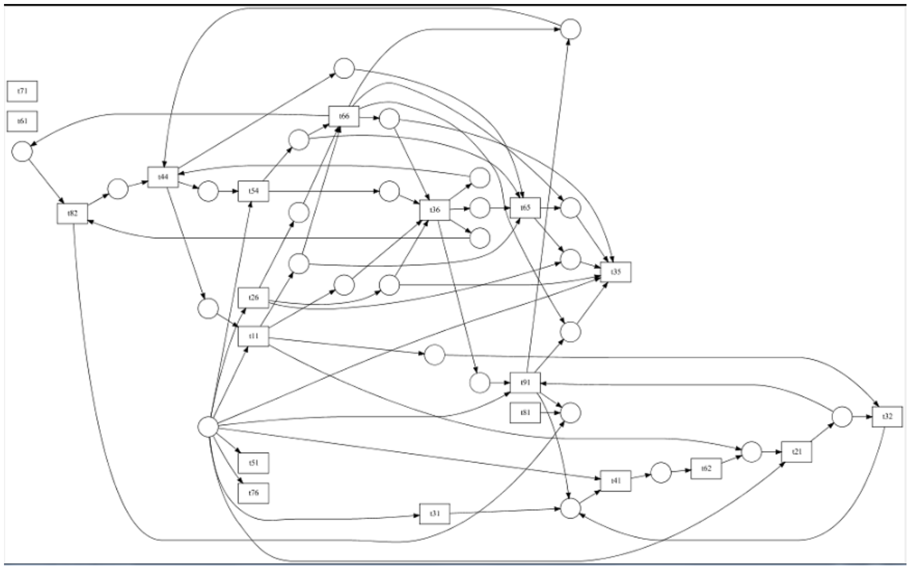


Fig. 12 Inductive Alpha miner for hospital data

The multiple sequences of activities are traced as a single workflow that can be represented in Fig. 13. The number of traces used in the dataset is connected with the activities used in the actual treatment, which will give the sequence graph in dotted format.

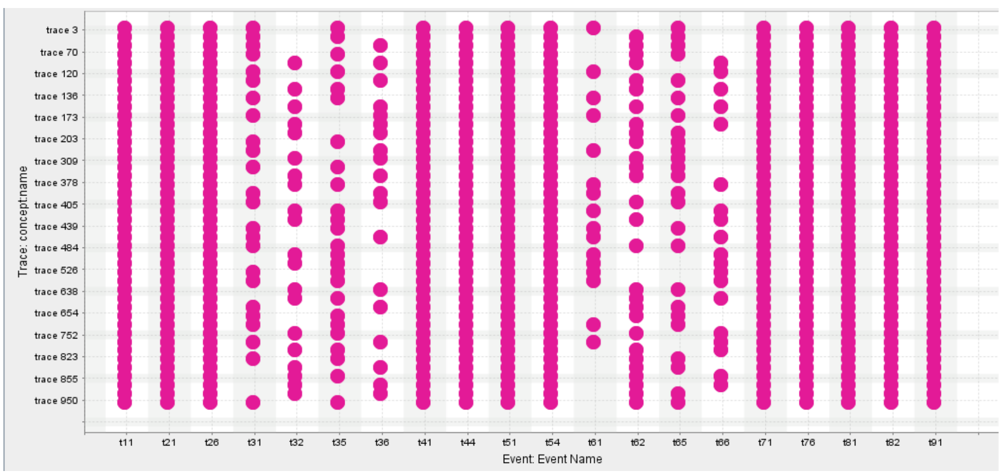


Fig. 13 Event log sequence based on traces

Node Connectivity in Alpha Miner:
The node's connectivity is displayed in Fig.14. Node *t11* is the starting node, and *t91* is the ending node. The complete flow and connectivity of the nodes are shown. The edges of the nodes depict the direction of the node's relation. The array marks provide node guidance and the path from the source node to the destination node. Petri Nets can use process mining to describe the diagram at a high level.

✉ **Shanmuga Sundari M**
mshanmugasundari@gmail.com

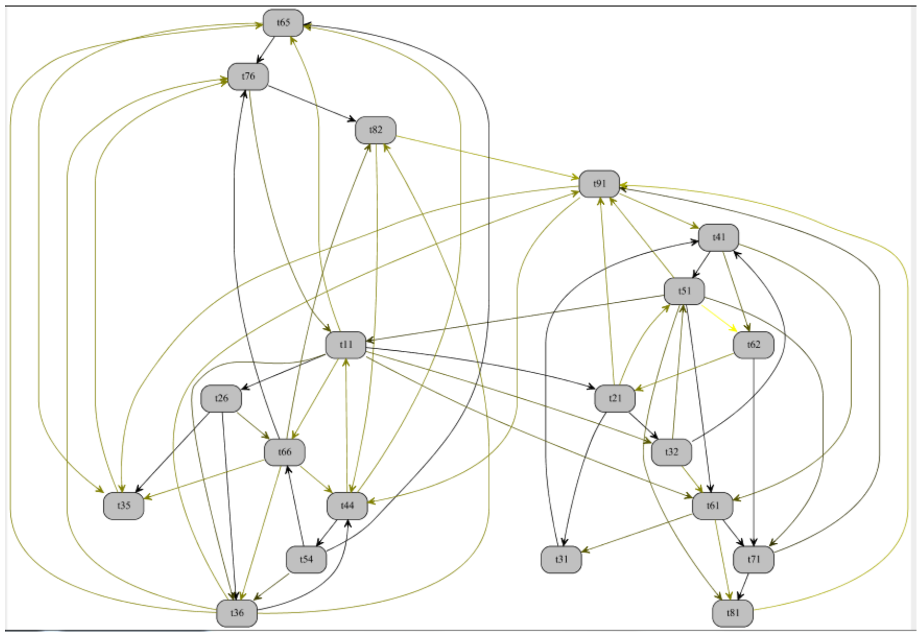


Fig. 14 Node connectivity in Petri net

Deviation will decrease the performance of the process. There are different categories of actions:

(1) high-impact actions: If any changes happen in this Activity, the whole structure will be affected. This node is crucial in timestamp as well.

(2) low-impact actions: If low sequence action can remove this from the flow, it is not affecting the hospital activity.

(3) no-impact actions: This can be removed at any time.

Fig.15 shows the deviation in the hospital database. It shows the deviation nodes. The nodes which are connected in red color edges are deviations. It will be a deviation if the nodes take more time to communicate with the other nodes because of traffic congestion.

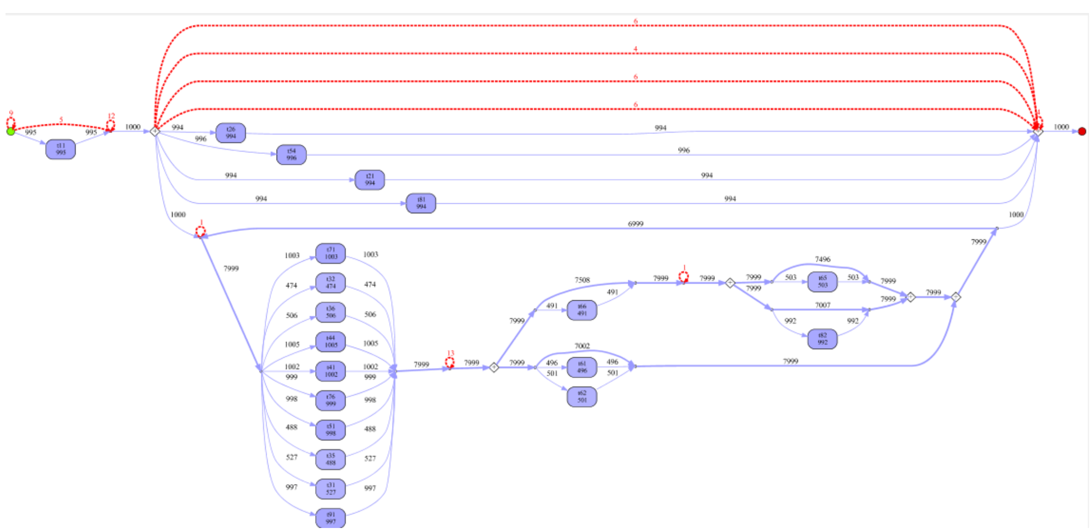


Fig. 15 Deviation for Hospital data

6- Results and Analysis

The result of this research is explained in the performance metric of events. This metric evaluation is presented in process mining, and the assessment is based on the relationship between activity sequences. For this test, we used 1000 patient's traces of activities, and process mining did each Activity in different departments in the hospital. The metrics are dependency and time stamp between the activities. By consideration of these metrics, the fuzzy miner produces the proper matrix for dependency and cluster formation for timestamp. It provides the better analysis than alpha algorithm. Notably, we restrict this test with two algorithms in process mining and find the deviation in the process. Based on this classification the metrics are more accurate using fuzzy miner.

Fuzzy miner interacts with the activities and finds high and low-performance groups. The graphical view can compare the identical core activity in the sequence. The dependency relationship of the traces in the mining process is depicted in Fig. 16. The values are displayed in the matrix based on the interdependence of the activities.

In the fuzzy miner, the blue-colored activities are incredibly bonding. These are considered the bottleneck activities that must take care of in the process. Management can find the bottleneck using this matrix and causal net diagram. Fuzzy miner gives this as a solution to monitor the process flow.

Matrix	t11	t21	t26	t31	t32	t35	t36	t41	t44
t11	-0.68918...	0.356255...	0.186411...	-0.77785...	-0.75558...	-0.72240...	-0.75598...	-0.79517...	-0.79997...
t21	-0.75880...	-0.88277...	-0.44908...	0.382342...	0.062083...	-0.69921...	-0.71202...	-0.87748...	-0.71230...
t26	-0.68661...	-0.43020...	-0.75011...	-0.65463...	-0.66415...	-0.05208...	0.112765...	-0.65128...	-0.78596...
t31	-0.74635...	-0.91595...	-0.48511...	-1.0	-0.91006...	-0.69054...	-0.77455...	0.013182...	-0.69317...
t32	-0.72021...	-0.86347...	-0.52569...	-0.91002...	-0.84426...	-0.67554...	-0.78509...	-0.12782...	-0.70589...
t35	-0.66063...	-0.57802...	-0.72433...	-0.70099...	-0.65483...	-0.69858...	-0.74687...	-0.67172...	-0.16121...
t36	-0.72327...	-0.61182...	-0.76400...	-0.81163...	-0.74939...	-0.74683...	-0.84732...	-0.70277...	-0.06947...
t41	-0.75338...	-0.88014...	-0.58867...	-0.91971...	-0.86364...	-0.68181...	-0.74654...	-0.88299...	-0.63129...
t44	-0.75628...	-0.67414...	-0.79749...	-0.78593...	-0.70905...	-0.78012...	-0.86522...	-0.57162...	-0.88313...
t51	-0.67937...	-0.79939...	-0.67098...	-0.78439...	-0.76097...	-0.67538...	-0.70120...	-0.79503...	-0.57968...
t54	-0.72069...	-0.73215...	-0.76580...	-0.76197...	-0.75452...	-0.75270...	-0.80380...	-0.60591...	-0.83934...
t61	-0.72242...	-0.86458...	-0.73025...	-0.90229...	-0.84537...	-0.72245...	-0.82917...	-0.86472...	-0.81158...
t62	-0.72280...	-0.86164...	-0.74984...	-0.91067...	-0.84556...	-0.73461...	-0.83293...	-0.86491...	-0.77433...
t65	-0.66408...	-0.75011...	-0.72518...	-0.75083...	-0.73846...	-0.69710...	-0.74744...	-0.71778...	-0.78324...
t66	-0.72185...	-0.82659...	-0.76258...	-0.89799...	-0.82356...	-0.74246...	-0.84306...	-0.75457...	-0.86138...
t71	-0.75617...	-0.87743...	-0.77882...	-0.91971...	-0.86364...	-0.75986...	-0.84956...	-0.88028...	-0.79102...
t76	-0.79045...	-0.89721...	-0.82727...	-0.96341...	-0.91030...	-0.81439...	-0.92508...	-0.81784...	-0.93403...
t81	-0.68640...	-0.79463...	-0.70744...	-0.78431...	-0.76086...	-0.71776...	-0.75299...	-0.79723...	-0.67915...
t82	-0.72258...	-0.80373...	-0.77350...	-0.81050...	-0.78924...	-0.74973...	-0.81208...	-0.73680...	-0.83645...
t91	-0.64847...	-0.83158...	-0.69365...	-0.89407...	-0.81237...	-0.67628...	-0.81390...	-0.82782...	-0.82791...

Matrix	t51	t54	t61	t62	t65	t66	t71	t76	t81	t82	t91
t11	-0.75028...	-0.77015...	-0.75784...	-0.76370...	-0.72331...	-0.76000...	-0.79754...	-0.82727...	-0.74651...	-0.77399...	-0.79721...
t21	-0.79721...	-0.78695...	-0.86466...	-0.86485...	-0.76789...	-0.82983...	-0.87751...	-0.90638...	-0.79262...	-0.82647...	-0.88005...
t26	-0.69225...	-0.77001...	-0.74676...	-0.74443...	-0.72526...	-0.75714...	-0.78352...	-0.82718...	-0.74441...	-0.77385...	-0.79477...
t31	-0.78450...	-0.72369...	-0.91061...	-0.91071...	-0.75096...	-0.86487...	-0.91975...	-0.96341...	-0.78454...	-0.83171...	-0.91967...
t32	-0.75541...	-0.71053...	-0.84179...	-0.84556...	-0.72958...	-0.80204...	-0.86364...	-0.88921...	-0.76100...	-0.80409...	-0.86347...
t35	-0.65838...	-0.75270...	-0.71662...	-0.71410...	-0.69948...	-0.74545...	-0.75991...	-0.81143...	-0.71804...	-0.75008...	-0.78223...
t36	-0.66564...	-0.81246...	-0.82208...	-0.81522...	-0.74451...	-0.84661...	-0.84332...	-0.92157...	-0.75313...	-0.80936...	-0.86188...
t41	0.359784...	-0.61835...	-0.85848...	-0.85868...	-0.68019...	-0.77031...	-0.88031...	-0.80859...	-0.79511...	-0.76448...	-0.88285...
t44	-0.59816...	0.457963...	-0.78967...	-0.76496...	-0.78073...	-0.86451...	-0.79373...	-0.93399...	-0.74193...	-0.83911...	-0.88291...
t51	-0.74814...	-0.57962...	-0.04212...	0.011095...	-0.65524...	-0.68016...	-0.79045...	-0.65908...	-0.74630...	-0.69422...	-0.79707...
t54	-0.61730...	-0.80527...	-0.64921...	-0.69903...	-0.08442...	-0.05167...	-0.72760...	-0.86077...	-0.69896...	-0.80501...	-0.83901...
t61	-0.76315...	-0.73049...	-0.84292...	-0.84666...	-0.69084...	-0.74647...	0.090414...	-0.77863...	-0.76047...	-0.72729...	-0.86458...
t62	-0.75806...	-0.69899...	-0.84666...	-0.84685...	-0.70302...	-0.77880...	0.160818...	-0.82758...	-0.76085...	-0.73641...	-0.86477...
t65	-0.62973...	-0.75341...	-0.67618...	-0.70302...	-0.70033...	-0.74602...	-0.70529...	0.083338...	-0.67861...	-0.74850...	-0.77777...
t66	-0.64989...	-0.81139...	-0.78921...	-0.78235...	-0.74602...	-0.84590...	-0.75462...	0.104150...	-0.69394...	-0.80536...	-0.86429...
t71	-0.79969...	-0.69280...	-0.86475...	-0.86493...	-0.69525...	-0.77973...	-0.88031...	-0.72297...	0.802205...	-0.70236...	-0.86929...
t76	-0.69129...	-0.87499...	-0.78223...	-0.83115...	-0.81199...	-0.91776...	-0.69564...	-0.99659...	-0.64605...	1.0	-0.92474...
t81	-0.74990...	-0.62332...	-0.76307...	-0.75523...	-0.63168...	-0.60577...	-0.79961...	-0.53613...	-0.74806...	-0.49699...	0.344011...
t82	-0.57210...	-0.80474...	-0.68338...	-0.65189...	-0.75050...	-0.81102...	-0.55030...	-0.87748...	-0.54288...	-0.80447...	0.379650...
t91	-0.69716...	-0.76232...	-0.81348...	-0.81367...	-0.67689...	-0.81319...	-0.83175...	-0.90408...	-0.69371...	-0.75845...	-0.83158...

Fig. 16 Relation matrix for traces using fuzzy miner

There are more chances of bottlenecks in the entire flow. In Fig. 16, *t44* and *t54* show values of 0.457963 which means that the dependency of these traces is highly significant. Traces *t36* to *t44* value is -0.06947, which shows less significance between the processes. This matrix in fuzzy miner shows the dependency connectivity in between the traces.

According to this matrix, we can find that the pair of (*t81, t71) and (t76, t82)* are the most significant nodes, which means this node is intermediate level to connect other nodes. Node *(t51, t61)* is the many minor nodes that do not affect the flow if it takes more time to process.

The negative values and low-frequency values are not affecting the flow of the process. These are the activities that are not under the bottleneck resources. All positive resources and high-frequency values are considered bottleneck resources, and these are to be monitored carefully in the process flow. This dependency matrix gives the final frequency values for the bonding between each Activity. So, Fuzzy miner is proved to be one of the best algorithms to show the frequency values using the causal net diagram.

7- Conclusions and Future Work

Process mining techniques are a robust business model to generate patterns and analyze the process with the sequence of activity traces. Organizations or any business domain want to improve their production in the initial stage of development. So process mining will help in the beginning to evaluate the process deviations and find the bottleneck in the process sequence to improve their process in the early stage itself. If the process is organized without variation, the management will reach efficient data arrival. This research experimented with process mining in hospital datasets by applying alpha and fuzzy algorithms using the ProM tool.

This research gives the analysis of finding deviation in both the algorithms. Since our data set is unstructured, the fuzzy algorithm produces the dependency relation in an event log with the proper workflow diagram. The fuzzy miner shows the result with a relation matrix that is the exact outcome of the connectivity of process mining. So, our dataset produces an efficient workflow using a fuzzy miner algorithm. We achieved the main objective of this research and found a better algorithm for activity deviation discovery and performance in process mining in the hospital system. In future, our model can be applied over large datasets like ontology, cardiology and it would incorporate process improvements.

References

[1] G. Li and R. M. De Carvalho, "Process Mining in Social Media: Applying Object-Centric Behavioral Constraint Models," IEEE Access, 2019, vol. 7, pp. 84360–84373, doi: 10.1109/ACCESS.2019.2925105.

[2] "Root Cause," 2021. [Online]. Available: https://appian.com/process-mining/root-cause-analysis.html#:~:text=The root cause analysis aims,impact factors such as bottlenecks.

[3] J. Xu and J. Liu, "A Profile Clustering Based Event Logs Repairing Approach for Process Mining," IEEE Access, 2019, vol. 7, pp. 17872–17881, doi: 10.1109/ACCESS.2019.2894905.

[4] G. Akhila, N. Madhubhavana, N. V. Ramareddy, M. Hurshitha, and N. Ravinder, "A survey on health prediction using human activity patterns through smart devices," Int. J. Eng. Technol., 2018, doi: 10.14419/ijet.v7i1.1.9472.

[5] W. Li, Y. Fan, W. Liu, M. Xin, H. Wang, and Q. Jin, "A Self-Adaptive Process Mining Algorithm Based on Information Entropy to Deal with Uncertain Data," IEEE Access, 2019, vol. 7, pp. 131681–131691, doi: 10.1109/ACCESS.2019.2939565.

[6] Q. Zeng, H. Duan, and C. Liu, "Top-Down Process Mining from Multi-Source Running Logs Based on Refinement of Petri Nets," IEEE Access, 2020, vol. 8, pp. 61355–61369, doi: 10.1109/ACCESS.2020.2984057.

[7] Z. Huang et al., "Safety Assessment of Emergency Training for Industrial Accident Scenarios Based on Analytic Hierarchy Process and Gray-Fuzzy Comprehensive Assessment," IEEE Access, 2020, vol. 8, pp. 144767–144777, doi: 10.1109/ACCESS.2020.3013671.

[8] W. Van der Aalst, Process mining: Data science in action. 2016.

[9] R. Tripathy et al., "Spectral Clustering Based Fuzzy C-Means Algorithm for Prediction of Membrane Cholesterol from ATP-Binding Cassette Transporters," in Intelligent and Cloud Computing, Springer, 2021, pp. 439–448.

[10] C. Subbalakshmi, G. Ramakrishna, and S. Krishna Mohan Rao, "Evaluation of data mining strategies using fuzzy clustering in dynamic environment," 2016, doi: 10.1007/978-81-322-2529-4_55.

[11] M. Anila and G. Pradeepini, "Study of prediction algorithms for selecting appropriate classifier in machine learning," J. Adv. Res. Dyn. Control Syst., 2017.

[12] G. Dorgo, K. Varga, and J. Abonyi, "Hierarchical frequent sequence mining algorithm for the analysis of alarm cascades in chemical processes," IEEE Access, 2018, vol. 6, pp. 50197–50216, doi: 10.1109/ACCESS.2018.2868415.

[13] J. Jin, W. Sun, F. Al-Turjman, M. B. Khan, and X. Yang, "Activity pattern mining for healthcare," IEEE Access, 2020, doi: 10.1109/ACCESS.2020.2981670.

[14] T. G. Erdogan and A. Tarhan, "Systematic Mapping of Process Mining Studies in Healthcare," IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2831244.

[15] W. Li, H. Zhu, W. Liu, D. Chen, J. Jiang, and Q. Jin, "An anti-noise process mining algorithm based on minimum spanning tree clustering," IEEE Access, 2018, vol. 6, pp. 48756–48764, doi: 10.1109/ACCESS.2018.2865540.

[16] P. I. C. Kumari, P. Gayathri, N. Rajesh, S. Umar, G. C. Sekhar, and A. M. Abdul, "Designing of medical processor unit for intelligent network-based medical usage," Indones. J. Electr. Eng. Comput. Sci., 2016, doi: 10.11591/ijeecs.v4.i3.pp532-537.

[17] M. J. Hasan, A. Rai, Z. Ahmad, and J. M. Kim, "A Fault Diagnosis Framework for Centrifugal Pumps by Scalogram-Based Imaging and Deep Learning," IEEE Access, 2021, vol. 9, pp. 58052–58066, doi: 10.1109/ACCESS.2021.3072854.

[18] M. J. Hasan, D. Shon, K. Im, H. K. Choi, D. S. Yoo, and J. M. Kim, "Sleep state classification using power spectral density and residual neural network with multichannel EEG signals," Appl. Sci., 2020, vol. 10, no. 21, pp. 1–13, doi: 10.3390/app10217639.

[19] M. J. Hasan, J. Uddin, and S. N. Pinku, "A novel modified SFTA approach for feature extraction," 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2016, 2017, pp. 1–5, doi: 10.1109/CEEICT.2016.7873115.

[20] M. S. Sundari and R. K. Nayak, "Process mining in healthcare systems: A critical review and its future," Int. J. Emerg. Trends Eng. Res., 2020, vol. 8, no. 9, pp. 5197–5208, doi: 10.30534/ijeter/2020/50892020.

[21] V. S. Reddy and B. T. Rao, "A combined clustering and geometric data perturbation approach for enriching privacy preservation of healthcare data in hybrid clouds," Int. J. Intell. Eng. Syst., 2018, doi: 10.22266/ijies2018.0228.21.

[22] K. M. Hanga, Y. Kovalchuk, and M. M. Gaber, "A graph-based approach to interpreting recurrent neural networks in process mining," IEEE Access, 2020, vol. 8, pp. 172923–172938, doi: 10.1109/ACCESS.2020.3025999.

[23] A. E. Marquez-Chamorro, K. Revoredo, M. Resinas, A. Del-Rio-Ortega, F. M. Santoro, and A. Ruiz-Cortes, "Context-Aware Process Performance Indicator Prediction," IEEE Access, 2020, vol. 8, pp. 222050–222063, doi: 10.1109/ACCESS.2020.3044670.

[24] R. Tripathy, R. K. Nayak, P. Das, and D. Mishra, "Cellular cholesterol prediction of mammalian ATP-binding cassette (ABC) proteins based on fuzzy c-means with support vector machine algorithms," J. Intell. Fuzzy Syst., 2020, vol. 39, no. 2, doi: 10.3233/JIFS-179934.

[25] A. Massmann, P. Gentine, and J. Runge, "Causal inference for process understanding in Earth sciences," 2021, pp. 1–24, [Online]. Available: http://arxiv.org/abs/2105.00912.

[26] A. K. A. De Medeiros, A. J. M. M. Weijters, and W. M. P. Van Der Aalst, "Genetic process mining: An experimental evaluation," Data Min. Knowl. Discov., 2007, doi: 10.1007/s10618-006-0061-7.

[27] E. Kim et al., "Discovery of outpatient care process of a tertiary university hospital using process mining," Healthc.

Inform. Res., 2013, vol. 19, no. 1, pp. 42–49, doi: 10.4258/hir.2013.19.1.42.

[28] P. Nets, A. Networks, and R. C. Language, "Modeling of Resource Allocation Mechanisms in Distributed Computing Systems using Petri Nets and Stochastic Activity Networks (SAN): a Review and Reo-based Suggestion."

[29] R. K. Nayak. Sundari, M. Shanmuga, Efficient Tracing and Detection of Activity Deviation in Event Log Using ProM in Health Care Industry, 2021, 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud (ISMAC 2021).

[30] S. C. Sekaran, V. Saravanan, R. RudraKalyanNayak, and S. S. Shankar, "Human Health and Velocity Aware Network Selection Scheme for WLAN/WiMAX Integrated Networks with QoS," Int. J. Innov. Technol. Explor. Eng. (IJITEE), ISSN, pp. 2278–3075.

[31] https://data.4tu.nl/articles/dataset/BPI_Challenge_2012/12689204/1.

[32] P. Selvaraj, V. K. Burugari, D. Sumathi, R. K. Nayak, and R. Tripathy, "Ontology based Recommendation System for Domain Specific Seekers," Proc. 3rd Int. Conf. I-SMAC IoT Soc. Mobile, Anal. Cloud, I-SMAC 2019, no. December, 2019, pp. 341–345, doi: 10.1109/I-SMAC47947.2019.9032634.