International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Mining Event Log Framework Implementation

Senin MS[1], Albert Feisal@Muhd Feisal Ismail[2], Mohd Norazmi Nordin[3]

[1]*Independent Researcher, Malaysia*
[2]*Faculty of Technology Management and Technopreneurship, Universiti Teknikal Malaysia Melaka*
[3]*Open University Malaysia*

Abstract: In this study, a hierarchical temporal memory modeling is implemented. The study is based on the classifying processes in multi-set event logs. The Hierarchical Temporal Memory (HTM) learning algorithms that handle sparse representation of data and online learning were used to build the model. The framework exhibits potency in process discovery. Every organization has set goals to achieve which could span from delivery of services to end users, transformation of unrefined materials into a desired product, rendering of support service to guarantee clients satisfaction, recording of financial transactions for the purpose of budgeting and management amongst others [1]. The accomplishments and achievements of these objectives and aim require the presence of an activity or a task, or perhaps, multiple activities and tasks. These set of tasks and activities are logically related and flow in a logical manner and are termed business processes. Business process is all about the daily operations of organizations and businesses irrespective of the business domain or industry they operate in. It can further be defined as a way of specifying the approach in which the resources of an enterprise are used [2].

1. INTRODUCTION

The financial standing of a business depends on the precision and value of the business process being used. Some processes in the business are so broad that they house many other sub processes, for instance, manufacturing itself can be considered as a business process, consisting of sub processes such as, product assembly, product distribution, product billing and quality assurance. Taking a closer look at the nature of business processes, brings to mind that they need to be managed, particularly because process can be complex and it involves a lot of people [2]. Some processes are straightforward, other are complex consisting of multiple steps and involving multiple users. In the bid to manage business processes, a discipline in management called Business Process Management (BPM) was adopted. The business process management can be defined as a management discipline that describes a systematic approach to identify, measure, document, execute and control both computerized and non-computerized business processes to

achieve reliable and directed results in line with the organization’s goals [3]. In other words, we refer to BPM as a systematic approach in improving processes in business to run effectively and efficiently so also to improve its ability to be adaptive in an ever-changing business environment and all of these processes involve data.

2. METHODOLOGY

Most process miners that have been developed in literature suffer a setback of misclassifying novel processes, as a result of the models being only a replica of the existing event log, rather than being an intelligent software. The framework proposed in this work is set to improve on the

intelligence of existing process miners by integrating the spatial pooler algorithm and the temporal memory algorithm from the HTM theory. The spatial pooler has a huge advantage of representing events and data in a sparse distributed manner, while the temporal memory has the capability of performing on-learning of sequence data. The HTM is a biological plausible theory that attempts to replicate the operations of the neocortex in computer software algorithm. As a machine learning framework, it primarily employs a spatio-temporal learning algorithm to form invariant features of the input world.

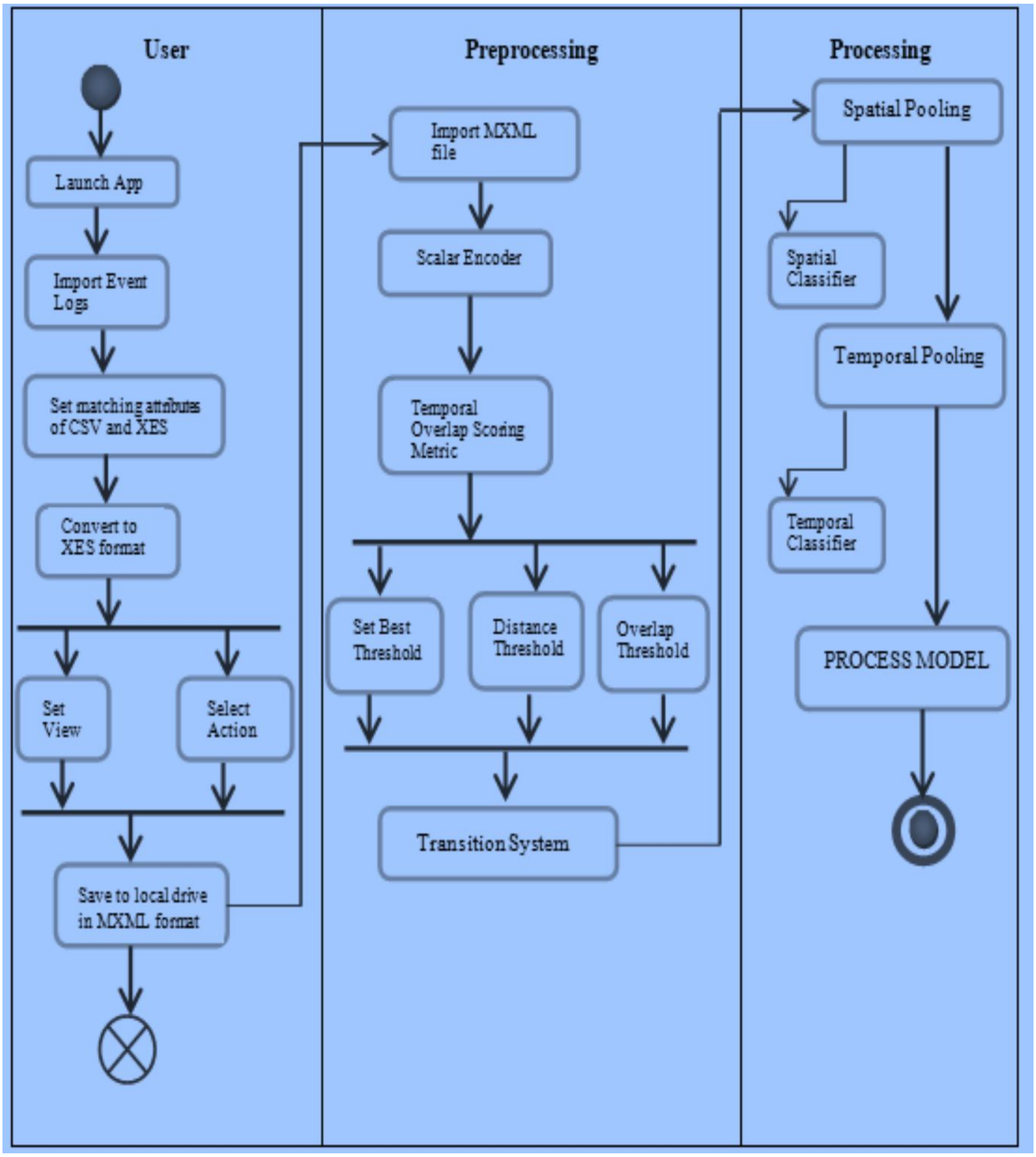
3. ESULTS AND DISCUSSION

Analysis of the Spatial Pooler Algorithm (SPA)

The SPA handles the actual learning of the system. It is responsible for the relationship between the columns of the region and input bits. The SPA acts on the input bits and returns a list of active columns, thereby representing information in a sparse manner. The SPA is executed in three stages.

Analysis of Temporal Memory Algorithm (TMA)

The TMA continues from where the SPA ended by carrying out two main functions which are; firstly, learning the sequences of the SDR formed by the spatial pooler and secondly, making predictions. Both algorithms share similar concepts like; Permanence, binary weights, dendrite segment, synapses and learning.



Three (3) swim-lanes can be seen from the figure above can be explained thus:

a. User: This swim-lane accounts for the actions that are being triggered by the user.

- The user launches the app by clicking the icon on the desktop
- Clicks the import icon to import event logs from existing path.
- Sets the matching attributes of case id, activity and timestamp of both CSV and XES format.
- Flows into two parallel actions of either set view or select action.
- Specify the MXML format at the point of saving on the system drive.

b. Preprocessing: This second swim-lane reveals the activities that concern the preprocessing of the even logs.

- The MXML file is imported from where it was previously stored.

- The scalar encoder converts it to binary sparse distributed representation (SDR).
- The temporal overlap scoring metric (TOSM) acts on it to generate a best start.
- The various threshold parameters are set
- A threshold system is generated as the intermediate representation of the log file.

c. Processing: the actual processing and generation of the process model occurs in this swim-lane.

- The three stages of the spatial pooler which are the overlap, inhibition and learning algorithms are implemented in the module.
- A trained spatial pooler is generated.
- The two stages of temporal learning and prediction have their algorithms implement

Processes are mined to discover, control and enhance the genuine process running in a system by extracting information from event logs that are usually available in information systems [5]. Gracia et al., (2019) reported a growing number of process mining publications across the years, with the peak of publication in the year 2014 as seen in figure 1 [6]. The Netherlands, Germany, China, USA and Italy have been recognized as the countries with the most contributions in

the process mining research area. Figure 2 describes a chart showing the number of publications per country, having The Netherlands as the country with the most produced significant contribution. With process mining, business strive better because they can analyze every process

running within the business in reality and reconstruct these processes, such that they can be carried out optimally [7]. Most businesses and institutions, if not all, are process oriented and in the coming years, there will be an overwhelming amount of processes on information systems, so, the ability to effectively mine processes today, will be a proactive step in the right direction. Processes are mined by a process miner to extract the true reality of events occurring within an organization in form of a petrinet. A petrinet otherwise known as place/transition net is a formal

way of representing process model. The information received from a petrinet influences organizational decisions in a positive way. Process miners can be referred to as the end products built from a process mining research. They are the software that mine huge number of processes to discover novel information and trends that lies beneath. Without process mining, industries will struggle in making realistic decisions or changes in their processes. Process mining makes use of event logs to draw out quality and hidden information on the processes that take place within an organization [1, 4]. Event logs are hard evidence of real events that have taken place, they also refer to data about business processes occurring during the system’s performance and they are collected and stored in the process aware information systems (PAIS) which are advantageously used as input information for building and retrieving business process model [5]. In event logs, each event refers to a case, an activity and a point in time. They are also said to be data that reveals the real events that has taken place rather than how it is supposed to be or how it is perceived by its authors. Event logs come from a wide variety of sources such as; patient data in a hospital, customer service unit in an institution, financial data on spreadsheets, transaction logs from trading systems, message log from middleware and a host of others [4].

When a cyber-criminal uses a suspicious or false identity to access a system, a log will be created. These suspicious activities can either occur in a minimum of 0 in a day i.e. a minimum of 0 - zero day attack and a maximum of N attack in a day. When such suspicious activities are observed once in a system or network, The developed system will record such activities as mistakes. Data recorded as mistakes are saved on the system which can be further used to process and analyze the rate of accuracy of the user. When a cyber-criminal tries to access a system or network with a suspicious activities for the second time using the same identity, the developed system will record and see it as machine error. Machine errors could be as a result hardware failure or computational error. These data can be further used to process the efficiency of hardware or used to measure the performance of the existing component of the developed system. Finally, when fraudulent or suspicious activities are detected on the developed system for the third time from the same identity, the system will no longer see it as either a mistake nor machine error. It will randomize it and save it using the bidirectional recurrent neural network as an attack. The volume of the attacked data can be used to model the prediction. Although it is possible for the first and second attempt (whether successful or not) to be an attempted attack, the developed system chooses to record the first and second foiled attempt as mistakes and errors. This is because our developed model may consider human computer error which may range from hardware failure and computational error. However, the model tries to evaluate the captured attack for prediction. The data collected was imported into weka. Weka is software designed in Java which is used in data mining specifically for prediction.

The result that was gathered during the experiment while predicting the rate of Zero-day attack for a specific domain during the research provides the information below. After a period of fifteen days, data was collected and recorded, the study was able to model the prediction after implementing the Bi- directional recurrent neural network algorithm. The table 1 below shows the data that was captured from domain A.

The result classified in the model for Domain A was 92.2% correctly classified, instances with a precision of 92% and an Fmeasure of 96%. In a similar result in Domain B, the correctly classified instances of 70%, with a precision of 70% and Fmeasure of 82%. A further result during the experiment in Domain C shows the correctly classified instances as 65% with a precision of 65% and a F-measure of 75%. The F-measure for predicting an attack in the developed Model using BRRN are as follows: Domain A is 0.960, Domain B is 0.824, and Domain C is 0.789.

The developed model gives a higher accuracy of about 92% from the dataset. The prediction of incoming attacks is achieved in a timely manner which enables security professionals to install defense systems in order to reduce the possibility of such attacks. Finally, the model performs better than the gray box prediction and black box prediction because a small sample of data was used. The mode of data collection was real time which makes data to be trained properly when modeling the prediction as against publicly available data and social data.

4. CONCLUSION

A framework based on the hierarchical temporal memory theory has been developed, to mine large dataset of event logs in the bid to generate a process model. The framework serves are a platform to aid process model developers in designing robust process models that can represent novel process intelligently. Leveraging on the sparse distributed learning nature of the hierarchical temporal memory, data scientist can design models that will not suffer the short coming of overfitting. The proposed framework also highlights the learning efficiency of the HTM learning algorithms as a result of its overlap and inhibition features. This capability will save process time and overall system time and at same time, produce a robust system. The framework proposed in the work will enhance process mining capabilities in the process mining domain.

5. REFERENCES

[1] Djenouri, Y., Belhadi, A., Fournier-Viger, P.: Extracting useful knowledge from event logs: A frequent itemset mining approach. Knowledge Based Systems 139(1), 2018, pp. 132-148.

[2] Alizadeh M., Lu, X., Fahland, D., Zannone,N., van der Aalst, W.M.P.: Linking data and process perspectives for conformance analysis. Computer Security 73(1), 2018, pp. 172-193.

[3] Pham, B., Bagheri, E., Rios, P., Pourmasoumi, A., Robson, R.C., Hwee, J., Isaranuwatchai, W., Darvesh, N, Page, M.J., Tricco, A.C.: Improving the conduct of systematic reviews: a process mining Perspective. Journal of Clinical Epidemiology 103(1), 2018, pp. 101-111.

[4] Ayo, F.E., Folorunso, O., Ibharalu, F.T.: A probabilistic approach to event log completeness. Expert Systems With Applications 80(1), 2018, pp. 263–272.

[5] Fahland, D., van der Aalst, W.M.P.: Simplifying discovered process models in a controlled manner. Information Systems 38(5), 2012, pp. 585-605.

[6] Huang, Z., Kumar, A.: A study of Quality and Accuracy Tradeoffs in Process Mining. Informs Journal on Computing 24(2), 2012, pp. 311-327.

[7] Dongen, B.F., van der Aalst, W.M.P.: A Meta model for process mining data. EMOI-INTEROP 130(1), 2005, pp. 309-320