International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

A Novel Approach for Handling the Concept Drifts in Process Mining by Using Events

J.Vasantharao [1]& **Sk.Jahangeer Basha** [2]
[1]PG Scholar,Dept of CSE, Rao & Naidu Engineering College, Ongole,Prakasam Dist,Andhra Pradesh
[2]Associate Professor,Dept of CSE, Rao & Naidu Engineering College, Ongole,Prakasam Dist,Andhra Pradesh

Abstract:
Now a days in this e-world most of the business totally related to the process mining trends, it is been generated and checks the process flow in terms of changes in the whole system. Here a few process mining techniques are used to analyze the changes in the drifts, this drifts may be changes according to the process and it can be sudden change, gradually change or the random change .This changes can be visualized by using the Rapid Miner tool, in that a ProM framework is used to get the better Results. Not only it tracks the change it also finds the localization in the business Process. Here Feature Extraction and Generate Population method is used to find the relationship among the activities. Event logs are used to establish the process flow

Keywords: - Process Mining; Drifts; Event Logs; Feature Extraction; Rapid Miner

I. Introduction

Authors Process mining techniques have got matured recently. Provided that the process is stable and enough example traces have been recorded and logged in the event, and is possible discovering a high-quality process model that can be used for performance analysis, compliance checking, and prediction [1]. Unfortunately, most processes will be in every dynamic process that increases necessity for enterprises that provide analysis towards their processes so as to reduce costs and to improve performance. Moreover, today's customers expect organizations to be flexible and adapt to changing circumstances [2]. New legislations created as many acts that have extreme variations in supply and demand that affects seasonally with natural calamities leads to disasters that are also forcing organizations to change their processes. For example, governmental and insurance organizations reduce the fraction of cases being checked when there is too much work in the network of process. In case of any calamities or problems many government institutions change their operating procedures etc. It is evident that the economic success of an organization is more and more dependent on its ability to react and adapt to changes in its operating environment [3]. Concept drift refers to the situation in which the process is changing while being analyzed. The need for techniques that deal with such second order dynamics analyzing such changes is of utmost importance when supporting or improving operational processes and to get an accurate insight on process executions at any instant of time. Processes can change in with respect to the three main process perspectives viz., Control flow, data, and resource. There are different changes that pursues drift process concept which gives behaviour with activity process. Such changes are perceived to induce a drift in the concept (process behaviour), e.g., in the way which activities are executed when, how, and by whom. There are three topics when dealing with concept drifts in process mining [4]. Whenever the data is transmitted from agent IP address, the heuristic algorithm is triggered and enabled in distributor side. Data transferred means that is indicated by the count value 1. There are many values that triggers display agents for unauthorized IP for different kinds of process points. Graph is displayed by agent IP address and unauthorized IP address with the notification of which data is transferred.

II. Related Work

Process mining serves a bridge between data mining and business process modeling. Business processes leave trails in a variety of data sources

(e.g., audit trails, databases, and transaction logs). Process mining aims at discovering, monitoring, and improving real processes by extracting knowledge from event logs recorded by a variety of systems (ranging from sensor networks to enterprise information systems). The starting point for process mining is an event log, which is a collection of events. We assume that events can be related to process instances (often called cases) and are described by some activity name. The events within a process instance are ordered [8]. Therefore, a process instance is often represented as a trace over a set of activities. In addition, events can have attributes such as timestamps, associated resources (e.g., the person executing the activity), transactional information (e.g., start, complete, suspend, and so on), and data attributes (e.g., amount or type of customer). For a more formal definition of event logs used in process mining, the reader is referred to shows a fragment of an example log.

Process mining determines how the process models are created. A broad variety of mining algorithms does exist. The following three categories will be discussed in more detail.

A. Heuristic mining algorithms Determinism means that an algorithm only produces defined and reproducible results. It always delivers the same result for the same input [14]. Heuristic mining also uses deterministic algorithms but they incorporate frequencies of events and traces for reconstructing a process model.

B. Fuzzy Miner Fuzzy Miner mines behavior of less structured processes [14]. It applies a variety of techniques, such as removing unimportant edges, clustering highly correlated nodes in to a single node, and removing isolated node clusters.

C. Genetic mining algorithms Genetics Miner uses a genetic algorithm to mine a Petri net representation of the process model from execution traces. The algorithm employs a search technique that mimics the evolution of biological systems. Although the algorithm can mine process models that might contain all the common structural constructs and can handle noise, it can take a large amount of computational time [14].

Event logs are completely standard in the process mining community and event log formats. The emergence of semi-structured processes, combined with improvements in computing power and the speed of data transmission, have fueled the need for mining algorithms to address new challenges. In addition to discovering process models from logs and examining the level of conformance of an actual process to its modeled counterpart, process mining should find, merge, and clean event data; handle changes in a process that occur while its being mined; and provide operational support to process users in an online manner [13].

Challenges faced in the process mining are:

1. The first challenge is that numerous uncorrelated events (with possible noise) are gathered from disparate heterogeneous and distributed systems.
2. The second challenge is the Dramatic variation in execution behavior.
3. The third challenge is parallel and repeated task execution.

In the real world concepts are often not stable but change with time. The underlying data distribution may change as well. The model built on old data will be necessarily updated. This problem is known as concept drift. The change in the distribution or the concept-drift may be – Sudden, Gradual or Recurring [12].

Concept drift is a relatively young research topic that has gained popularity in data mining and machine learning communities in the last 10 years. Concept drift research primarily has been focusing on two directions: 1) how to detect drifts (changes)

online and 2) how to keep predictive models up to date Concept drift has been shown to be important in many applications. The basis for drift detection could be a raw data stream, a stream of prediction errors, and, more rarely, a stream of predictions or a stream of updated model parameters. Two types of concept drift detection approaches have been used: monitoring evolution of a stream or comparing data distributions in two time windows. The cumulative sum (CUSUM) approach is a representative sequential analysis technique for change detection, different extensions to which have been proposed [6]. One notable example is computational intelligence-based CUSUM or CI-CUSUM that aims to detect a non stationary condition by monitoring a multidimensional vector, i.e., multiple features. Adaptive windowing is a representative approach for online change detection using an adaptive size sliding detection window. In this paper, we consider offline change detection and its localization and therefore focus on studying what features to monitor and how to identify when these characteristics change [8].

III. Change of Process Points

3.1 Change Point Detection

The first and most fundamental problem is to detect concept drift in processes that takes place for detection of process [5]. If so, the next step is to identifying the time periods at which changes occur. And also by analyzing an event log from an organization, one should be able to detect that process changes happen and that the changes happen at the onset of a season.

3.2 Change Localization and Characterization

Once a point of change has been charged, then it is applied to characterize the nature of change, and identify the region(s) of change (localization) in a process [6]. Uncovering the nature of change is a challenging problem that involves both the identification of change perspective (e.g., control-flow, resources which is sudden and gradual) identifies exact and change itself. In the seasonal process, there could be a change that more resources are deployed or that special offers are provided during holiday seasons

3.3 Change Process Discovery

Having identified, localized, and characterized the changes, it is necessary to put all of these in perspective. The main need for techniques is that which exploit and relate these discoveries [7]. Unravelling the evolution of a process should result in the discovery of the change process describing the second dynamics. A seasonal process, one could identify that the process recurs every season. A process evolved over a period of time with annotations showing several perspectives such as the performance metrics of process at different instances of time.

One can consider an event log L as a time series of traces (traces ordered based on the timestamp of the first event). The basic premise in handling concept drifts is that the characteristics of the

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

traces before the change point differ from the characteristics of the traces after the change point. The problem of change (point) detection is then to identify the points in time when the process has changed, if any. Change point detection involves two primary steps:

(i) Capturing the characteristics traces, and

(ii) Identifying when these characteristics change.

The process is stable and enough example traces have been recovering and is possible to discover a high quality process model that can be used for analyzing the performance, checking, and predicting. Fairly most processes are not in steady-state [8]. In today's dynamic place which is increasingly necessary for enterprises to streamline their processes so as to reduce costs and to improve performance. We consider applications where the original sensitive data cannot be perturbed and analyzed. A completely specified workflow design is required in order to enact a given workflow process. The design for creation of workflow which is very complicate process that consumes more time for typical process that have more discrepancies between flow of workflow processes that follows the management. Thus the proposed techniques that rediscover the workflow models for distributed design. This technique uses workflow logs to discover the workflow process as it is actually being executed.. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified [9]. Watermarks can be very useful involve some modification of the original data. Other watermarks cause malicious code.

IV. Proposed System

The proposed model aims to remove the problems related to inefficiency in accurately classifying the data examples in the presence on concept drift as the base classifier was not able to learn the old concept well and thus a need of completely new classifier is needed to train on new concept in order to make negative instances positive and detect drifted data along with classification based on completely new concept. The model aims not only to accurately classify the data but also detect drifted data correctly. The steps that would be followed to obtain accurately classified data along with drifted data are:

1) The system evaluates the dataset properties.

2) The redundant and irrelevant attributes are eliminated.

3) Minority class is discovered and iteratively the boundaries are refined to obtain accurately classified instances.

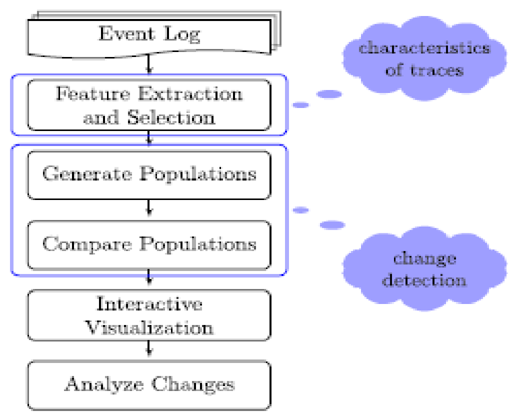

Fig :1 Framework for handling concept drifts in process mining.

The framework identifies the whole process how we are getting the result. First of all we have to import the particular event logs in an data base server then we will follow the frame work to detect the changes in the environment to recognize the event logs in an appropriate visualization and drift changing in an Prom 6 network to regain the very it is being change in the process. This framework identifies the 4 modules to detect the drift changes in process mining.

1) Feature extraction and selection: In This feature extraction and selection step we have to identify what type of process perspectives is suitable like organizational or resource then we have to select any one of them to extract the changes in the environment system.

2) Generating Population and Compare Population: In this generating populations step we will collect the sample event logs from particular log data and we will generate the populations after that we can see the changes in the process then compare the previous event process.

3) Interactive Visualization: The results of comparative studies on the populations of the events. And we can visualization Prom the changes what type of drift occurred, by using the ProM6 software.

4) Analyze Changes: Here we can analyze and identify the exact where the change has been occurred in the form of bar chart we can notice the changes

IV. Implementation

ProM framework works very well in finding out the visualization of the process event logs in an appropriate manner. Process ProM has emerged to be the facto standard for process mining. First we have to download the prom6 extension from the rapid miner tool .Then automatically all the packages will be uploaded from that prom framework we have to read the log data and it should be connected to the multiplier, in that all the operations are involved and it performs to generate the results. And now we have to import two more packages they are Execute the read file and also the alpha miner operation step to get the relevant features. This step involves many actions to get the source related process mining drifts from the event logs, and also it follows where the change has been occurring and it tracks the recourses whether it is data perspective or the organizational perspective. We will notice all the changes in the form of a bar chart and visualizes very clearly with the help of a x-axis and the y-axis in the prom framework according to the detect in the changes in process mining.

V. Conclusion

In this paper, the main purpose of the project is to describe the concept drifts in process mining using event logs and this process depends on the event logs. A process may change suddenly or gradually for that purpose it is important to track the process where there is a change. A platform is developed in order to find out the drifts in the event logs using the feature technique and the rapid miner tool. By this we can visualize the changes in the events .Mainly it is used to effectively identify the changes in the events.

VI. REFERENCES

[1] Jianmin Wang, Raymond K. Wong, Jianwei Ding, Qinlong Guo, And Lijie Wen,“Efficient Selection Of Process Mining Algorithms**”**, *Ieee Transactions On Services Computing, Vol. 6, No. 4,* October-December 2013.

[2] R.P. Jagadeesh Chandra Bose, Wil M.P. van der Aalst, Indr_Zliobait, and Mykola Pechenizkiy, “Handling Concept Drift in Process Mining”, *Department of Mathematics and Computer Science, University of Technology, The Netherlands*.

[3] Van der Aalst, W. M. P. “Process Mining: Discovery, Conformance and Enhancement of Business Processe*”,* (*1st Edition) Heidelberg: Springer.*

[4] R. P. J. C. Bose, W. M. P. van der Aalst, I. Žliobait˙e, and M. Pechenizkiy,“Handling concept drift in process mining,” in *Proc. Int. CAiSE*, 2011,pp. 391–405

[5] J. Carmona and R. Gavaldà, “Online techniques for dealing with concept drift in

process mining," in *Proc. Int. Conf. IDA*, 2012, pp. 90–102

[6] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. New York, NY, USA: Springer- Verlag, 2011.

[7] B. F. van Dongen and W. M. P. van der Aalst, "A meta model for process mining data," in *Proc. CAiSE Workshops (EMOI-INTEROP Workshop)*, vol. 2. 2005, pp. 309–320.

[8] C. W. Günther, (2009). *XES Standard Definition* [Onlilne]. Available: http://www.xes-standard.org

[9] F. Daniel, S. Dustdar, and K. Barkaoui, "Process mining manifesto," in *BPM 2011 Workshops*, vol. 99. New York, NY, USA: Springer-Verlag, 2011, pp. 169–194.

[10] R. P. J. C. Bose, W. M. P. van der Aalst, I. Žliobait˙e, and M. Pechenizkiy, "Handling concept drift in process mining," in *Proc. Int. CAiSE*, 2011,
pp. 391–405.

[11] J. Carmona and R. Gavaldà, "Online techniques for dealing with concept drift in process mining," in *Proc. Int. Conf. IDA*, 2012, pp. 90–102.

[12] J. Schlimmer and R. Granger, "Beyond incremental processing: Tracking concept drift," in *Proc. 15th Nat. Conf. Artif. Intell.*, vol. 1. 1986, pp. 502–507.

[13] A. Bifet and R. Kirkby. (2011). *Data Stream Mining: A Practical Approach*, University of Waikato, Waikato, New Zealand [Online]. Available: http://www.cs.waikato.ac.nz/~abifet/MOA/StreamMining.pdf

[14] I. Žliobait˙e, "Learning under concept drift: An Overview," *CoRR*, vol. abs/1010.4784, 2010 [Online]. Available: http://arxiv.org/abs/1010.4784

[15] Van Der Aalst, Wil MP, Arthur HM Ter Hofstede, and Mathias Weske. "Business process management: A survey." *Business process management*. Springer Berlin Heidelberg, 2003. 1-12.