ENHANCED CONCEPT DRIFT IN PROCESS MINING

Namrata A. Dumasia[1], Prof. Ankur Shah[2]

[1]P. G. Student, Department of Computer Engineering, Parul Institute of Technology,Vadodara,Gujarat.
[2]Assistant Professor, Department of Computer Science, Parul Institute of Technology,Vadodara,Gujarat.

ABSTRACT

Economic success of an organization is dependent on how they react & adapt changes in its operating environment. Research aims to enhance the existing Drift Detection with Change process discovery in complex datasets. Configurable process model describes a family of similar process models. Process variants discovered using concept drift can be merged to derive a configurable process model. Online learning algorithms often have to operate in the presence of concept drifts. Recent study revealed that different diversity levels in an ensemble of learning machines are required to maintain high generalization on both old & new concepts. Based on this study of diversity with different strategies to deal with drifts, we propose new online ensemble learning approach called Diversity for Dealing with Drifts (DDD). DDD maintains ensembles with different diversity levels and is able to achieve better accuracy than other approaches. It is very robust & outperforming in terms of accuracy when there are false positive drift detections.

Keyword : *Concept Drift; Process Mining; Event Logs; Process Change.*

---

1. INTRODUCTION

Business processes are nothing more than logically related tasks that use the resources of an organization to achieve a defined business outcome. Business processes can be analysed from a number of perspectives, like control flow, data, and the resource perspectives. In today's market scenario, it is necessary for enterprises to streamline their processes so as to reduce cost and to improve performance. Also nowadays customers expect organizations to be flexible and adapt the changes. Extreme variations in supply and demand, natural calamities, seasonal effects, disasters and so on, 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 of work in the pipeline. As another example, in a disaster, hospitals, and banks change their operating procedures. 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. Therefore, flexibility and change have been studied in- depth in the context of business process management (BPM). Business Process Management (BPM) plays a major role in the business environment. The term business process management and the business operational resources are covers how we identify, study, monitor and change, business processes to ensure the process of the event logs and classifies the structure improved over time [8].

2. CONCEPT DRIFT

Concept Drifts refers to situation when the relationship between the input data and the target variable, which the model is trying to changes overtime. Drifts are classified changes into both momentary and permanent. Momentary drifts are nothing but the change will be appear at maximum time and after that it resolves the process where as Permanent Drifts are it gradually changes the whole process which disturbs the event logs while running the features[5].

When dealing with concept drifts in process mining, the following three main challenges emerge.

1) Change point detection: The first and most fundamental problem is to detect concept drift in processes, i.e., to detect that a process change has taken place. If so, the next step is to identify the time periods at which changes have taken place. For example, by analysing an event log from an organization (deploying seasonal processes), we should be able to detect that process changes happen and that the changes happen at the onset of a season.

2) Change localization and characterization: Once a point of change has been identified, the next step is to characterize the nature of change, and identify the region(s) of change (localization) in a process. Uncovering the nature of change is a challenging problem that involves both the identification of change perspective. For instance, in the example of a seasonal process, the change could be that more resources are deployed or that special offers are provided during holiday seasons[4].

3) Change process discovery: Having identified, localized, and characterized the changes, it is necessary to put all of these in perspective. There is a need for techniques/tools that exploit and relate these discoveries. Unraveling the evolution of a process should result in the discovery of the change process describing the second-order dynamics. For instance, in the example of a seasonal process, we could identify that the process recurs every season.

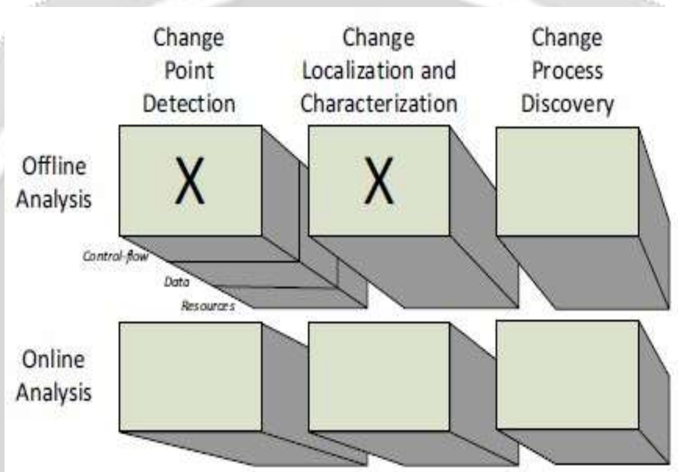


Fig-1 : Different dimensions of concept drift analysis in process mining[8].

We can differentiate between two broad classes of dealing with concept drifts when analysing event logs[7].

1) Offline analysis: This refers to the scenario where the presence of changes or the occurrence of drifts need not be uncovered in a real time. This is appropriate in cases where the detection of changes is mostly used in post mortem analysis, the results of which can be considered when designing/improving processes for later deployment. For example, offline concept drift analysis can be used to better deal with seasonal effects (hiring less staff in summer or skipping checks in the weeks before Christmas).

2) Online analysis: This refers to the scenario where changes need to be discovered in near real time. This is appropriate in cases where an organization would be more interested in knowing a change in the behavior of their customers or a change in demand as and when it is happening.

2.1 Perspectives Of Change :

There are three important perspectives in the context of business processes: Control flow, Data and Resource. One or more of these perspectives may change over time.

1) Control flow/behavioural perspective: This class of changes deals with the behavioural and structural changes in a process model. Just like the design patterns in software engineering, there exist change patterns capturing the common control-flow changes. Control flow changes can be classified into operations such as insertion, deletion, substitution, and reordering of process fragments. For example, an organization which used to collect a fee after processing and acceptance of an application can now change their process to enforce payment of that fee before

processing an application. Here, the reordering change pattern had been applied on the payment and the application processing process fragments. Sometimes, the control-flow structure of a process model can remain intact but the behavioral aspects of a model change. For example, consider an insurance agency that classifies claims as high or low depending on the amount claimed. An insurance claim of $ 1000which would have been classified as high last year is categorized as a low insurance claim this year because of the organization's decision to increase the claim limit. The structure of the process remains intact but the routing of cases changes[3].

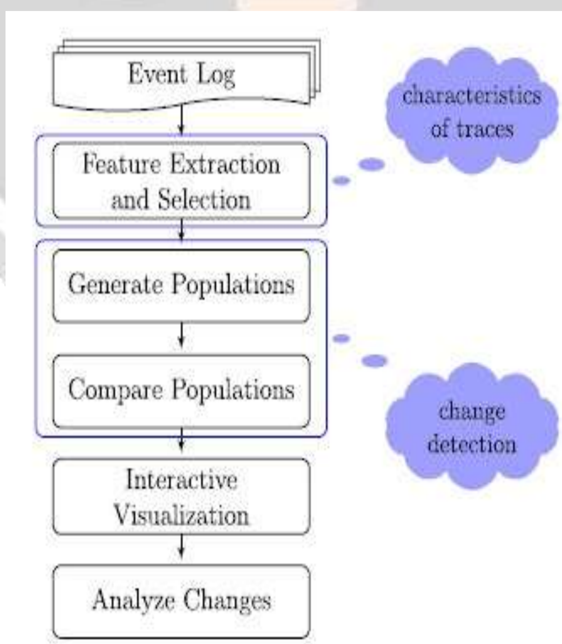
2) Data perspective: This class of changes refer to the changes in the production and consumption of data and the effect of data on the routing of cases. For example, it may no longer be required to have a particular document when approving a claim.

3) Resource perspective: This class deals with the changes in resources, their roles, and organizational structure, and their influence on the execution of a process. For example, there could have been a change pertaining to who executes an activity. Roles may change and people may change roles[7]. As another example, certain execution paths in a process could be enabled (disabled) upon the availability (non-availability) of resources. Furthermore, resources tend to work in a particular manner and such working patterns may change over time, e.g., a resource can have a tendency of executing a set of parallel activities in a specific sequential order. Such working patterns could be more prominent when only few resources are available.

3 EXISTING FRAMEWORK

We propose the framework shown in Fig. 3 for analysing concept drifts in process mining. The framework identifies the following steps[8]:

1) Feature extraction and selection: This step pertains in defining the characteristics of the traces in an event log. There are four features that characterize the control-flow perspective of process instances in an event log. Depending on the focus of analysis, we may define additional features, e.g., if we are interested in analysing changes in organizational/resource perspective, we may consider features derived from social networks as a means of characterizing the event log. In addition to feature extraction, this step also involves feature selection. Feature selection is important when the number of features extracted is large.



Fig–2 : Framework for handling concept drifts in process mining[8].

2) Generate populations: An event log can be transformed into a data stream based on the features selected in the previous step. This step deals with defining the sample populations for studying the changes in the characteristics of traces. Different criteria/scenarios may be considered for generating these populations from the data stream.

3) Compare populations: Once the sample populations are generated, the next step is to analyse these populations for any change in characteristics.

4) Interactive visualization: The results of comparative studies on the populations of trace characteristics can be intuitively presented to an analyst. Troughs in such a drift plot signify a change in the significance probability there by implying a change in the characteristics of traces.

5) Analyse changes: Visualization techniques such as the drift plot can assist in identifying the change points. Having identified that a change had taken place, this step deals with techniques that assist an analyst in characterizing and localizing the change and in discovering the change process.
The framework can be used for designing new change detection approaches.

4 PROPOSED WORK

Initial results show that heterogeneity of cases arising because of process changes can be effectively dealt with by detecting concept drifts. Once change points are identified, the event log can be partitioned and analyzed. An analysis should only be observed as the starting point for a new subfield in the process mining domain. There are lots of challenges that still need to be addressed. And our proposed work aims to deal with some of challenges like Change process discovery, online drift detection and handling concept drift in complex data sets using DDD algorithm.

5 EXPERIMENTAL RESULTS

This section describes the experimental results of proposed system. Here, Dealing with concept drift is done and based on that it detects the drift in online analysis. There are results that shows the concept drift on the basis of different parameters.

Fig-3 : Drift Analysis

Product Name	Year	Sell(in lac.)
cpu	2004	197
cpu	2005	177
cpu	2006	178
cpu	2007	202
cpu	2008	335

Fig-4 : Concept Drift in Product View

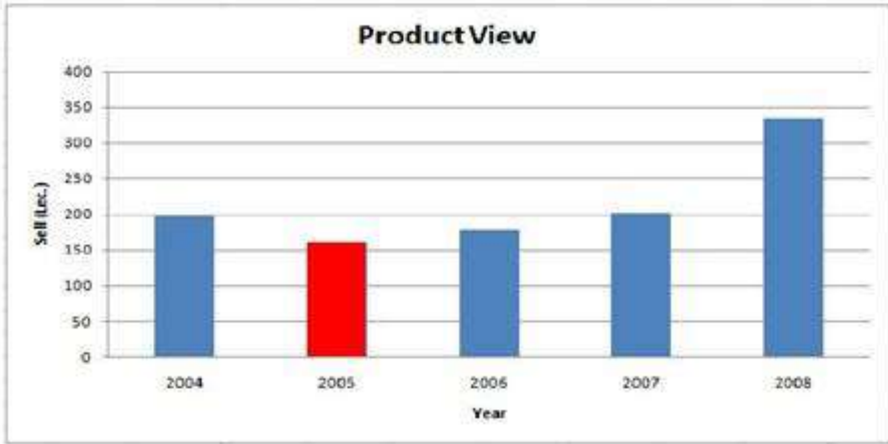


Fig-5 : Comparison of Concept Drift in Product View

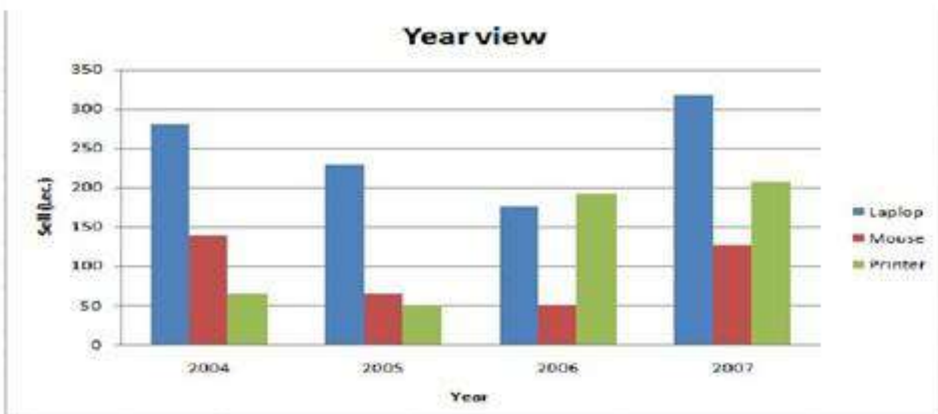


Fig-6 : Comparison of Concept Drift in Year View

Fig-7 : Comparison of Concept Drift in Product & Time View

CONCLUSION

DDD maintains ensembles with different diversity levels and is able to attain better accuracy than other approaches. It is very robust, outperforming other drift handling approaches in terms of accuracy when there are false positive drift detections. Process mining aims at discovering, monitoring and improving the operational process using the traces recorded in log. In this paper, a new algorithm DDD is used for handling drifts in process mining. It performs better than previous results. So, to handle the phenomenon of concept drift, one must be aware of different means from which changes in process may get induced. It presented the three different factors (perspectives, change types and mode of handling) to be considered while designing the solution for the problem of concept drift. Therefore, our analysis should only be observed as the starting point for a new subfield in the process min ing domain and there are few challenges that still need to be addressed.

ACKNOWLEDGMENTS

I express my sincere thanks to my Guide, Prof. Ankur Shah,(Assistant Professor, CSE department, Parul Institute Of Technology, Vadodara), for valuable time and guidance. My special thanks to all the Faculty members of the Department for their valuable advices at every stage.

REFERENCES

[1] Dr. Nick Gehrke, Michael Werner, Dipl.-Wirt.-Inf (2013). "Process Mining".
[2] Du Fei, Zhang Liqun, Ni GuangYun, Xu Xiaolei (2013). "An Algorithm for Detecting Concept Drift Based on Context in Process Mining", IEEE Fourth International Conference on Digital Manufacturing & Automation.
[3] J. Carona and R. Gavald a (2012). "Online techniques for dealing with concept drift in process mining," in Advances in Intelligent Data Analysis XI. Springer.
[4] Josep Carmona and Ricard Gavald (2013)."Online Techniques for Dealing with Concept Drift in Process Mining".
[5] Maddu. Uma Lavanya1, Mr. Sunil Kumar Talluri (2015). "Dealing With Concept Drifts In Process Mining Using Event Logs", International Journal Of Engineering And Computer Science.
[6] Manoj Kumar M V, Likewin Thomas, Annappa B (2014). "Phenomenon of Concept Drift from Process Mining Insight", IEEE Transaction.
[7] R. J. C. Bose, W. M. van der Aalst, I. _Zliobait_e, and M. Pechenizkiy (2011). "Handling concept drift in processmining," in Advanced Information Systems Engineering. Springer.
[8] R. P. Jagadeesh Chandra Bose, Wil M. P. Van Der Aalst, Indr˙E Žliobait˙E, Mykola Pechenizkiy (2014). "Dealing With Concept Drifts In Process Mining", IEEE Transactions On Neural Networks And Learning Systems, Vol. 25, No. 1.
[9] Veena Mittal, Indu Kashyap (2015). "Online Methods of Learning in Occurrence of Concept Drift", International Journal of Computer Applications.