

# Dealing with Concept Drifts in Process Mining

N. Abinaya, S. Subathra

Department of Computer Science and Engineering,  
Sri Vidya College of Engineering & Technology,  
Virudhunagar- 626 005.

Mrs. R. Anitha,

AP/CSE,  
Sri Vidya College of Engineering & Technology,  
Virudhunagar-626 005.

**Abstract:-** The concept of drift is used in the transaction. Concept drift is an important concern for any data analysis scenario. The drift may be periodic (e.g. due to seasonal influences) or one-of-a-kind (e.g., the effects of new legislation). For process management it is crucial to discover and understand such concept drifts in processes.

**Keywords:** *Concept drift, flexibility, hypothesis tests, Process changes, process mining.*

## 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 viewed from a number of perspectives, including the control flow, data, and the resource perspectives. In today's dynamic marketplace, it is increasingly necessary for enterprises to streamline their processes so as to reduce cost and to improve performance. New legislations such as the WABO [1] and the Sarbanes-Oxley [2], extreme variations in supply and demand, seasonal effects, natural calamities and disasters, deadline escalations 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. Therefore, flexibility and change have been studied in-depth in the context of business process management (BPM). For example, process-aware information systems (PAISS) [4] have been extended to be able to flexibly adapt to changes in the process. State-of-the-art workflow management (WFM) and BPM systems provide such flexibility, e.g., we can easily release a new version of a process. Many of today's information systems are recording an abundance of event logs. Process mining is a relatively young research discipline aimed at discovering, monitoring, and improving real processes by extracting knowledge from event log. Although flexibility and change have been studied in-depth in the context of WFM and BPM

systems[5], contemporary process mining techniques assume the processes to be in a steady state. For example, when discovering a process model from event logs, it is assumed that the process at the beginning of the recorded period is the same as the process at the end of the recorded period. These practical experiences show that it is very unrealistic to assume that the process being studied is in a steady state. Concept drift refers to the situation in which the process is changing while being analyzed. There is a need for techniques that deal with such second-order dynamics. Analyzing such changes is of most importance when supporting or improving operational processes and to obtain an accurate insight on process executions at any instant of time.

## 2. RELATED WORKS:

In [1], The Act has created one overarching procedure for granting permission for projects like the construction, alteration or use of a house or building. There is now one permit, one procedure and one set of submittal requirements, followed by one legal remedies procedure and enforcement by one authority. Applications may be submitted electronically to the Online Portal 24 hours a day. They will be processed electronically as far as possible. This modernisation of the permit system will not establish any new or amended criteria for examining applications. Nor will the Act alter the level of protection afforded by current legislation. The same goes for the policy latitude of a competent authority to attach conditions to a permit. But the Act does regulate avoidance of conflicting conditions. The Act gives applicants considerable freedom in arranging the process of requesting a permit. In principle they may decide whether to apply in one go for a permit that covers all their activities, or first to apply for a permit for one activity or a few activities and later for the other activities. An example is an application for permission to cut down trees, followed by an application to demolish some structures and then an application to ready the sites for construction work. Successive landowners can then request planning permits to build one or more houses, a

retail building, company building, school, etc., for example. Besides the integrated permit procedure (i.e. one application for several activities), the Act regulates coordination. Government authorities involved in the application are required to cooperate with each other to take one harmonized decision, issued by one competent authority.

**In [2]**, Decision making in process-aware information systems involves build-time and run-time decisions. At build-time, idealized process models are designed based on the organization's objectives, infrastructure, context, constraints, etc. At run-time, this idealized view is often broken. In particular, process models generally assume that planned activities happen within a certain period. When such assumptions are not fulfilled, users must make decisions regarding alternative arrangements to achieve the goal of completing the process within its expected timeframe or to minimize tardiness. We refer to the required decisions as escalations. This paper proposes a framework for escalations that draws on established principles from the workflow management field. The paper identifies and classifies a number of escalation mechanisms such as changing the routing work, changing the work distribution, or changing. The requirements with respect to available data. A case study and a simulation experiment are used to illustrate and evaluate these mechanisms.

**In [3]**, Operational processes need to change to adapt to changing circumstances, e.g., new legislation, extreme variations in supply and demand, seasonal effects, etc. While the topic of flexibility is well-researched in the BPM domain, contemporary process mining approaches assume the process to be in steady state. When discovering a process model from event logs, it is assumed that the process at the beginning of the recorded period is the same as the process at the end of the recorded period. Obviously, this is often not the case due to the phenomenon known as concept drift. While cases are being handled, the process itself may be changing. This paper presents an approach to analyze such second-order dynamics. The approach has been implemented in Prom and evaluated by analyzing an evolving process.

**In [4]**, Variants of the same process may be encountered in different organizations, e.g., any municipality will have a process to handle building permits. New paradigms such as Software-as-a-Service (SaaS) and Cloud Computing stimulate organizations to share a BPM infrastructure. The shared infrastructure has to support many processes and their variants. Dealing with such large collections of similar process models for multiple organizations is challenging. However, a shared BPM infrastructure also enables cross-organizational process mining. Since events are recorded in a unified way, it is possible to cross-correlate process models and the actual observed behavior in different organizations.

**In [5]**, The Software as a Service (SaaS) paradigm is particularly interesting in situations where many organizations need to support similar processes. For example, municipalities, courts, rental agencies, etc. all need to support highly similar processes. However, despite these similarities, there is also the need to allow for local variations in a controlled manner. Therefore, cloud infrastructures should provide configurable services such that products and processes can be customized while sharing commonalities. Configurable and executable process models are essential for realizing such infrastructures. This will finally transform reference models from "paper tigers" (reference modeling a SAP, ARIS, etc.) into an "executable reality". Moreover, "configurable services in the cloud" enable cross-organizational process mining. This way, organizations can learn from each other and improve their processes.

### 3. IMPLEMENTATION:

#### 3.1 Heuristic Guilt Agent Analysis:

Once the agent is tried to leak the file, his information will be stored in log file. There are two changes in this model with respect to the previous one. The first one is related to the checking for leaking of the registered documents. The second one is that someone is tried to change the documents. Illegal Agent could not open the leaked file.

**3.2 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 postmortem 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).

**3.3 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. Such real-time triggers (alarms) will enable organizations to take quick remedial actions and avoid any repercussions.

#### 3.4 Modules:

4.4.1 Data transfer

4.4.2 Guilt model

4.4.3 Change point detection module

4.4.4 Agent-guilt model

3.3.1 Data Transfer:

- This module is mainly designed to transfer data from distributor to agents.
- The same module can also be used for illegal data transfer from authorized to agents to other agents.

3.3.2 Guilt Model:

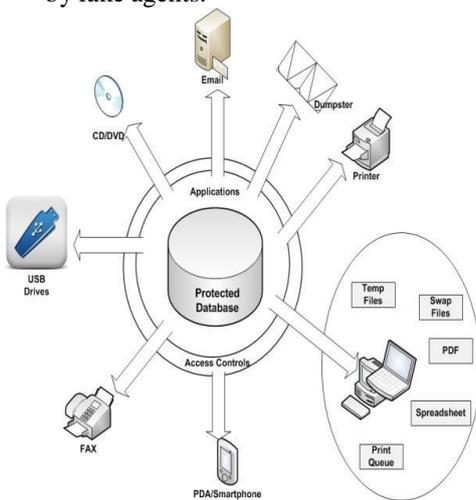
- This module is designed using the agent – guilt model.
- Here a count value(also called as fake objects) are incremented for any transfer of data occurrence when agent transfers data.
- Fake objects are stored in database.

3.3.3 Change Point Detection Module:

- 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 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.

3.3.4 Agent-Guilt Model:

- This module is mainly designed for determining fake agents.
- This module uses fake objects (which is stored in database from guilt model module) and determines the guilt agent along with the probability.
- A Statistical report is used to plot the probability distribution of data which is leaked by fake agents.



4.CONCLUSION:

We analyzed event logs of three processes from a large Dutch municipality for the presence of concept drifts (i.e., process changes). The detection of such change points can help us put the results of process mining in a right perspective and enables an organization to take appropriate measures when a change in behavior is perceived. Using the framework proposed in\ for dealing with concept drifts in process mining, we are able to detect changes in real-life event logs even with a small number of cases.

5. FUTURE ENHANCEMENT:

The proposed four features characterizing the control flow dependencies between activities. These features are shown to be effective in detecting process changes. An event log can be transformed into a data set D, which can be considered as a time series by these features. In future, six features characterizing the control flow dependencies between activities. These features are shown to be effective in detecting process changes. An event log can be transformed into a data set D, which can be considered as a time series by these features. Change detection is done by considering a series of successive populations1 of feature values and investigating if there is a significant difference between two successive populations. The premise is that differences are expected to be perceived at change points provided appropriate characteristics of the change are captured as features.

6. REFERENCES:

- [1] (2010) All-in-one Permit for Physical Aspects:(Omgevingsvergunning)in a Nutshell[Online].Available:http://www.answersforbusiness.nl/regulation/all-in-one-permit-physical- aspect.
- [2] W. M. P. van der Aalst, M. Rosemann, and M.Dumas, "Deadline-basedescalation in process- aware information systems," Decision SupportSyst., vol. 43, no. 2, pp. 492–511, 2011
- [3] M. Dumas, W. M. P. van der Aalst, and A. H. M. TerHofstede,Process- Aware Information Systems: Bridging People and Software ThroughProcess Technology. New York, NY, USA: Wiley, 2005.
- [4] R. P. J. C. Bose, W. M. P. van der Aalst, I. Zliobaite, and M. Pechenizkiy,"Dealing with concept drifts in process mining: A case study in a Dutch municipality," BPM Center, Univ. Technol., Singapore, Tech. Rep. BPM
- [5] W. M. P. van der Aalst, "Configurable services in the cloud: Supportingvariability while enabling cross-organizational process mining," in Onthe Move to Meaningful Internet Systems (OTM 2010), LNCS 6426. New York, NY, USA: Springer-Verlag, Jan. 2010, pp. 8–25.