Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2011 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-2-2011

Enhancing Decision Patterns Dicovered By Process Mining With Semantic Related Data

Wirat Jareevongpiboon

Paul Janecek

ENHANCING DECISION PATTERNS DICOVERED BY PROCESS MINING WITH SEMANTIC RELATED DATA

Wirat Jareevongpiboon, Asian Institute of Technology, wirat@cs.ait.ac.th
Paul Janecek, Asian Institute of Technology, paul.janecek@gmail.com

ABSTRACT

Business processes can be automatic, semi-automatic or manual processes. Semi-automatic and manual processes are involved in some parts by people. Understanding how people work or make judgments in processes can help management to evaluate their performance and suggest essential information to enhance their decision making. This paper describes a case study of using process mining to discover decision patterns of a worker in a semi-automatic business process. It was found that the discovered rules could be improved by enhancing the business execution log file with semantic related data. The experimental results before and after improvements were compared.

Keywords: Process Mining, Decision pattern, Decision making, Semantic related attribute, Log enhancing

INTRODUCTION

Business process mining is a recent technology used in Business Process Analysis, which aims at monitoring, diagnosing, simulating and mining enacted processes in order to support analysis and enhancement of process model [4]. It starts by gathering process data logged by the IT systems of a company, such as Enterprise Resource Planning (ERP), Supply Chain Management (SCM) and Customer Relationship Management (CRM) [5]. Through mining and analyzing these data, we can extract knowledge about the actual process execution, uncover patterns in process data as well as predict potential problems in current processes. ProM, a project by the Process Mining Group, Eindhoven Technical University-Netherlands, is an open source and extensible framework that supports a wide variety of process mining techniques and analysis tools in the form of plug-ins. Researchers have applied ProM to analyze many areas of business, such as process discovery (automatically construct a process model from the event logs), conformance checking (compare the recorded behaviour with some already existing process model to detect possible deviations) or extension (e.g. how the value of data attributes may affect the routing of a case) [2].

Analysis of various perspectives is also possible in ProM depending on the data recorded in log files. The basic attributes of each process required for mining include a process instance ID, activity name, event type, originator or performer and timestamp. These attributes allow ProM to mine in *control-flow* and *organizational* perspectives. The *control flow* mining displays the abstraction of process model as a flow of activities where a decision point (if any) indicates an alternative flow and the route taken is determined at run time (FIGURE 1 shows an example). The *organizational* perspective discovers social networks of people handing over tasks or an organizational hierarchy of people involved in a process. For other perspectives, such as *performance* and *quality*, additional attributes need to be recorded in a log file.

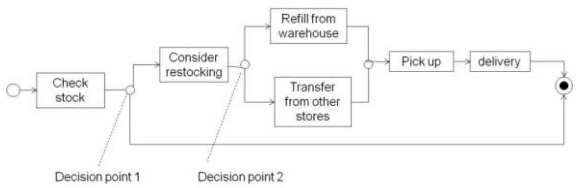

FIGURE 1 : Abstract Representation of Business Process.

The perspective of our interest is decision pattern of activities' performers. Along a process flow, performers can be people, the system or both depending on whether the process is manual, automatic, or semi-automatic. In a manual process, the process model can represent the way people work or decide on something, especially when the human decision involves routing the direction of activities. By mining the manual or human part of a semi-automatic process, we can discover not only the process model, but also the human decision model along the process flow. At each decision point accomplished by a human, we can analyze the attributes that influence the human decision. The analysis results may reveal "business rules" or "individual rules" that humans use in process execution, which can be used for guiding decision process improvement and human performance evaluation.

The main contribution of this work is to find the decision patterns of real case process which is semi-automatic. It compared decision rules derived from system and human and argued that data in the event log is not enough for discovering decision rules especially human decision. Humans are able to use not only explicit, but also implicit knowledge in their judgement. This work suggested the way to improve the results of discovery by enhancing the log file with semantic related data.

In this paper, a semi-automatic business process in a retail warehouse was studied. The process execution data in the past was extracted from the database and transformed into a log file. This file was mined and analyzed by ProM with the purpose of decision rules discovery. The remainder of the paper is organized as follows. Section 2 reviews related research in business process mining and work that makes use of business process mining technology. Section 3 introduces the background of this case study, and the details of the studied process and tested data. Section 4 shows the results from process mining. Section 5 discusses the analysis of results and enhancement. Section 6 talks about future work and section 7 presents the conclusion.

RELATED WORK

The application of process mining in the context of workflow management was first introduced in [10]. In 2005, the ProM framework for mining business processes [2] was developed, and has since been applied to many application domains to discover processes and analyze decision making, from multiple perspectives.

An early example of the practical application of business process mining is found in [13], where process mining was used to analyze the processing of handling invoices for the Dutch National Public Work Department. In this paper, three perspectives of process mining (i.e., control-flow, organizational and case perspectives) were analyzed and the results were combined to reach a richer understanding of the process. The work of [9] presented a process mining approach in the gas industry, converting the process model discovered by ProM into a Petri Net that was editable and redesigned by CPN Tools. In the Health Care domain [8] presented an evaluation and comparison of process models derived from real life clinical data using seven different process mining algorithms. Process mining has also been used as a tool to detect exceptions in a process, such as in the case of fraud. This was studied in [7] using ProM's HeuristicsMiner to detect flaws in the procurement process of a company.

The input to process mining is often extracted from information systems in the form of log files. When sufficient data to define process instances is not available in the log files of enterprise systems, cases together with related events and data are often extracted from the relevant tables in databases to create process instances. In the case of [11], the log files to support three perspectives of analysis were derived from queries over a hospital information system, and then mined to discover an understandable process model.

For the analysis of decision patterns, [1] introduced an approach to analyze how data attributes influence the choices made in a process. This approach identifies decision points in a Petri Net process model discovered by an algorithm in ProM to turn a decision point into a learning problem. The authors in [1] developed the Decision Miner plug-in to the ProM framework and tested it with the liability claim process in an insurance company.

Our work is similar to the work mentioned above, but slightly extends its purpose by applying the process mining and decision point analysis to the semi-automatic process called restocking process. This studied process involves tasks accomplished by both human and machine. The rules discovered by Decision Miner are not only business rules coding in program logics, but also decision rules of human where their judgments are really important to the success of business process. Like the most other real cases and the work of [11], the input data is not perfect. Cases related info is derived from associating tables in relational database where data from applications used by human are recorded. Our work enhances the result of Decision Miner and increases the possibility to discover rules by adding semantic related data to log file. It shows that transactional data found in most log files are not sufficient for decision analysis especially human decision.

CASE STUDY

The case study of this paper is a company that sells clothes for ladies and gentlemen, which we will refer to as the "Acme Company". Acme owns several major brands of women ready-made dresses and leatherwear, and distributes products to department stores, as well as its own retail stores. Distributing and transferring products to/from these stores are critically key process of the company. Usually, new products are launched every two weeks and these products are distributed to each store in only a small amount, two pieces per stock keeping unit (SKU) for the beginning. SKU identifies product type, size and color. The undistributed products are stored in the central warehouse, and are sent daily to stores that have sold products (the "refill process") to maintain inventory level. When the stock of new products in the central warehouse becomes empty, products are redistributed among retail stores and department stores as needed, which is called the "transfer process". This process is controlled and decided by "merchandisers", the person or group of persons who are responsible for promoting their individual brands. The success of this process strongly depends on the decision making of merchandisers as the results directly affect the sale amount of products being transferred. Deciding to move products to the right places at the right time may increase the opportunity of sales and generate more revenue to the company. From now on, we refer to both refill and transfer processes as a restocking process.

Details of the Studied Process

Even though the restocking process involves many units in the company such as commercial, logistics and points of sale, we focused only on the decision parts in the commercial department where merchandisers are the key actors. The restocking process is semi-automatic, using the Restocking Information System (RIS) in the commercial department, but paper-based system in others. All activities are done in sequence and on a daily basis. Starting from the commercial department, RIS detects out-of-stock Point of Sales (POS) where products are sold out. It begins refill process automatically to restock products if available from central warehouses to those POS. POS that are not refilled because the stock of central warehouse is empty will be moved on transfer process. In this process, RIS generates two main reports called NoStk report (out of stock report) and StkAvail report (stock available report). Grouped by SKU, the NoStk report lists all stores or POS, where products are sold out and not refilled. The StkAvail report, on the other hand, lists all POS that have products available in stock. For each SKU, merchandisers have to decide to move products logically from stores in the StkAvail report, where stock is available, to stores in the NoStk report, where stock is not available, via the interface of the system (FIGURE 2).

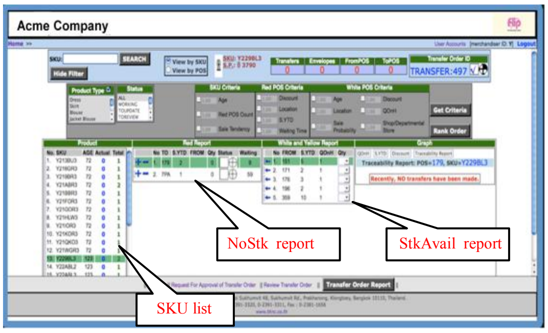

FIGURE 2: Restocking Information System's Interface Used by Merchandisers for Making Transfer Decision.

Typically, in each SKU, a merchandiser selects a POS from the NoStk report and then selects the quantity to transfer from one or more POS from the StkAvail report. Before completing each POS-POS transfer, the merchandiser has an option to see additional information in pop up windows or in paper-based reports to support his/her decision. The examples of additional information for each POS include sales performance, size of POS, current percent discount and waiting days for restocking etc. A merchandiser usually finishes the logical transfer of products among stores in the morning and all selections made from RIS including refill are recorded in the database. Related documents (e.g. transfer notes, moving notes) are generated and printed. The results of selection are then passed to brand managers who are responsible for approving the refill and transfer lists for their own brands. After approval, those printed documents are signed and passed to the logistics department, who physically relocate products among POS based on the approved results. The complete circle of a refill (warehouse to POS) and transfer case (POS to POS) takes at least 2 days beginning from logical transfer decision by merchandisers and ending with physical transfer by logistics.

Details of Data

The data used for this study is historic data in May 2009. Since this company owns many major brands, each of which has many new and aging products restocked in that period, we only considered refill and transfer cases for products launching from 2007-01-01 of a brand in order to analyze the decision pattern of one merchandiser. Within this period, there were about 11544 instances waiting for restocking while only about 207 cases or 1.8% was accomplished. The company had recently replaced a paper-based system with a new restocking information system (RIS).

The company does not keep data in event log form, but rather in relational database. In order to analyze decision patterns in process, a log file was generated from this data source. The stock data is recorded in SKU unit, which identifies the product type, color and size. We defined stock of a SKU in a POS as a process instance in order to see its changes while passing along the process flow. The steps done in RIS lists activities the system and merchandiser did when making logical refill or POS-POS transfer and were used to complete process instances. To get completed data for a transfer case, a document named transfer note recorded as a table, generated automatically after completing logical refill or transfer, is used to link data from all related tables. This document table records date of restocking, SKU, from-POS and to-POS. In case of refill, from-POS will be either 777 or 7BP which are identifications of central warehouses. These data were extracted and organized in the form of a log file containing the date of logical transfer, person who made transfer, activities, and the data related attributes. Apart from physical product movement by logistics, timestamp data were recorded in date as all activities done by merchandiser were completed in a day for just several clicks from the RIS.

PROCESS MINING WITH CASE STUDY

As shown in FIGURE 3, there are three main steps using ProM to mine and analyze this process. The first step is data preparation. The historic data from the relational database was loaded, arranged in the form of an event log and was transformed into MXML (Mining XML).

MXML [2] is a format required by ProM, using a framework called ProMImport [3], which is able to load various types of log file from several type of information systems and convert from original format to MXML [13]. The second step is process mining, where the MXML file is mined with the desired perspective (process discovery in this case) to create a process model. The third and final step is analysis, where the resulting model is analyzed.

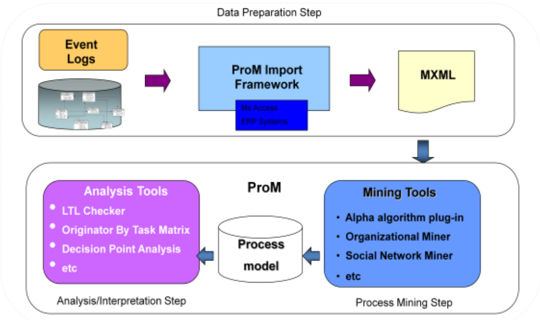

FIGURE 3: Process Mining Steps with ProM.

Data Preparation Step

TABLE 1: The Extract of Data in Log in Table Format.

Instance#	Task/Activity	Actor	EventType	Timestamp	Data
1	Notify Stock Empty	RIS	Complete	06/05/09	POS=187, prd_code= 2006040288895
1	Check from warehouses	RIS	Complete	06/05/09	WH status = NOT EMPTY
1	Refill from warehouses	RIS	Complete	06/05/09	
1	Refill old products	RIS	Complete	06/05/09	from-POS=777 , quantity = 1
1	Create Moving Notes	RIS	Complete	06/05/09	bill_no = 7651
1	Approve transfer & refill list	Mrs Praiya	Complete	06/05/09	status = OK
2	Notify Stock Empty	RIS	Complete	06/05/09	POS=117, prd_code= 2006040332321
2	Check from warehouses	RIS	Complete	06/05/09	WH status = NOT EMPTY
2	Refill from warehouses	RIS	Complete	06/05/09	
2	Refill new products	RIS	Complete	06/05/09	from-POS=7BP , quantity = 2
2	Create Moving Notes	RIS	Complete	06/05/09	bill_no = 7652
2	Approve transfer & refill list	Mrs Praiya	Complete	06/05/09	status = OK
3	Notify Stock Empty	RIS	Complete	06/05/09	POS=142, prd_code= 2006040332116
3	Check from warehouses	RIS	Complete	06/05/09	WH status = EMPTY
3	Make transfer decision	Ms. Hatairat	Complete	06/05/09	SKU= G33ABL4 to-POS = 142
3	Confirm transfer	Ms. Hatairat	Complete	06/05/09	from-POS=185, quantity = 1
3	Create moving Notes	RIS	Complete	06/05/09	bill_no = 7501
3	Approve transfer & refill list	Mrs Praiya	Complete	06/05/09	status = OK
4	Stock Empty Notification	RIS	Complete	06/05/09	POS=116, prd_code= 2006040332109
4	Check from distribution WH	RIS	Complete	06/05/09	WH status = EMPTY
4	Make transfer decision	Ms. Hatairat	Complete	06/05/09	SKU= G33ABL3, to-POS=116
4	Suspend transfer	Ms. Hatairat	Complete	06/05/09	
4	Create moving Notes	RIS	Complete	06/05/09	
4	Approve transfer & refill list	Mrs Praiya	Complete	06/05/09	status = OK

As previously mentioned, the restocking information system itself does not record event in form of log file, but the system records data as transactional tables in MySQL database. Although ProMImport can convert files from many types of information system, it has no plug-in to work directly with MySQL, but MSAccess. We therefore used MSAccess as a mediation format. We linked transactional data about this process in form of four MSAccess tables which were instance, instance data, activity, and activity data tables and used the available plug-in in ProMImport to convert them into an MXML file containing process instances describing process instance id, activity/task name, event type (start, suspend, resume, complete etc), originator (person who initiates task), timestamp and other data attributes. The extract of four process instances represented in one table format is shown in TABLE 1.

Process Mining Step

At this step, the MXML log file was loaded into ProM. As mention earlier, ProM provides various plug-in algorithms for the purposes of mining. In our case, we used the Alpha algorithm plug-in [14], to derive a process model in Petri Net, a format required by Decision Miner later used in the analysis section) showing the real execution of this merchandiser using RIS. Based on the data in the log file, a process model with 10 activities was discovered. These activities are:

1. Notify stock empty
2. Check from warehouses
3. Make transfer decision
4. Refill from warehouses
5. Suspend transfer
6. Confirm transfer
7. Refill old products
8. Refill new products
9. Create moving notes
10. Approve transfer and refill lists

The abstract model of the derived process is shown in FIGURE 4 Focused only on activities in the commercial department using RIS, we considered activities 1 and 10 as starting and ending activities. Activities 1, 2, 4, 7, 8, 9 were done by the system while activities 3, 5, and 6 were performed by a merchandiser and activity 10 by a brand manager.

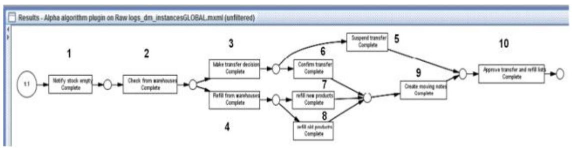

FIGURE 4: A Process Model Discovered by Alpha Algorithm.

Process Analysis Step

The derived model showed several alternative paths (FIGURE 5) where the decision involve routing the direction of activities. Decision point A and B are done by RIS to determine whether a stock should be refilled or transferred and the refill moves products from old product warehouse or new product warehouse respectively. Decision point C is done by merchandiser to determine whether a stock should be restocked from a transfer process. We would like to analyze the factors that influenced the decision points to discover decision rules and check whether all decisions rules follow business rules. Especially merchandiser's decision in transfer process because the results of transferring are directly affects sales opportunity. ProM supports this kind of analysis with Decision Miner, a decision point analysis plug-in. This algorithm determines all decision points (if any) in the discovered model to find which properties of a case might lead to taking a certain route/path in the process model.

Using the decision tree algorithm C4.5 provided by the Weka library, Decision Miner converts decision points into classification problems where the classes are the different decisions that can be made in a process [1].

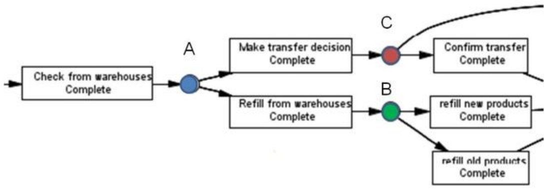

FIGURE 5: All Decision Points Discovered by Decision Miner.

The following summarizes the results from the Decision Miner at each decision point where data attributes in log file were considered.

Decision point A shows the system decision on selecting refill or transfer

Statistics

Correctly Classified Instances	99.9827%
Kappa statistic	0.9374
Mean absolute error	0.0003
Root mean squared error	0.0132
Relative absolute error	11.4449%
Root relative squared error	34.3219%

Rule discovered

WH status = EMPTY: Transfer
WH status = NOT EMPTY: Refill

FIGURE 6: Decision Point A - Result from Decision Miner.

The decision point A was done by the system with two alternatives: "*Refill from warehouses*" and "*Make transfer decision*. With almost 100% correctly classified instances, it was revealed that the status of warehouses for requested products (*WH status*) is the attribute that routes the case. If the warehouse of selected SKU is not empty, the system refills products to *to-POS*, otherwise forward it to transfer process. This discovered factor is straightforward as the decision point A is decided by the system which works according to business rules encoded in the programs.

Decision point B shows the system decision on selecting type of refill.

The result of decision point B, which is the step of selecting warehouse to move product from, is shown in FIGURE 7. For Acme Company, the stock of new products, launched in latest season, are stored in warehouse 7BP, while the stock of dated products, and were moved and stored in warehouse 777. This decision point, therefore, determines the warehouse (*from-POS*). It is discovered that requested POS for products having code > 2006040325088 will be refilled from warehouse 7BP; otherwise it will be refilled from warehouse 777.

Statistics

Correctly Classified Instances	88.2353 %
Incorrectly Classified Instances	11.7647 %
Kappa statistic	0.7463
Mean absolute error	0.1961
Root mean squared error	0.3131
Relative absolute error	40.3397 %
Root relative squared error	63.6096 %

Rule discovered

If prd_code <= 2006040325088 : refill old products
If prd_code > 2006040325088 : refill new products

FIGURE 7: Decision Point B - Result from Decision Miner.

Decision point C shows the merchandiser's decision on selecting POS from the NoStk report. Some POS were selected to be restocked while others were not.

Statistics

Correctly Classified Instances	98.3517 %
Incorrectly Classified Instances	1.6483 %
Kappa statistic	0
Mean absolute error	0.0324
Root mean squared error	0.1273
Relative absolute error	99.7505%
Root relative squared error	100%

Rule discovered

suspend transfer decision

FIGURE 8: Decision Point C - Result from Decision Miner

The result of decision point C, which is the step of selecting POS from NoStk report to be restocked, is highlighted (FIGURE 8). The selected POS continued to the "*Create moving notes*" activity while the unselected POS were passed to "*Approve transfer and refill lists*" activity. Merchandiser, named "Ms Hatairat" of studied brand plays a key role in making decision. With the information recorded before decision point C in the log such as POS id (to-POS) and product code (*prd_id*) no rule was discovered.

ANALYSIS OF RESULT AND ENHANCEMENT

The decision point analysis works well to discover business rules, but not human rule from data in the log. As clear business rules are fixed and encoded in the program, which can direct the system performing tasks automatically in automatic process such as *refill* process in our case. The RIS system first checks the status of warehouse to decide whether a stock should be refilled or transferred. It also checks from the barcode of products to determine whether products are latest or dated. However, in semi-auto process which deals with people, business rules sometimes cannot direct people in deciding, but instead as guideline. Each person may use his/her own personal judgement in doing something. Like our case, merchandisers are group of people who are experts in the fashion market of the company. They are familiar with their brands, market segments, fashion trends, target groups and target areas. So, they use their expertises and experiences in their work to fulfil the global objective of company which is increasing sales. As their individual decision in restocking products may affect the company in overall sales, the patterns of them usually interests the management to study.

It is noticed that data supplied to Decision Miner are limited to data found in event log. These logged data are input, output or updated data which are usually entered or updated by the programs or users. For example, in our TABLE 1, *POS* and *prd_code* are the beginning input to the RIS. *SKU*, *to-POS* (from NoStk report), *from-POS* (POS from StkAvail report) and *qty transferred* are directly selected from RIS main interface and recorded as activity's data of *Make transfer decision* and *Confirm transfer* activities. These recorded data are enough to discover business rules for process performing by machine, but not enough for human. Human decision is not quite straightforward as machine decision. Human can use not only explicit knowledge but also implicit knowledge about data for judgement. It is observed from the case study that merchandisers may use semantic related data about POS from the pop-up screen on the system interface or from the paper reports, or from their own heads (implicit knowledge) while selecting POS from NoStk report. The examples of these semantic related data about POS are area, size, sales performance and current discount etc (the description of these properties is listed in TABLE 2). These semantic related attributes were not found in the original log file (TABLE 1).

TABLE 2: Semantic Related Data of POS.

Properties	Description
area	the location of POS
size	the size of POS which is measured by the amount of products in pieces (no actual size recorded in database)
sales for SKU	the amount of sales (in pieces) for a given SKU since its lauching date in the POS
overall sales	the overall amount of sales (in pieces) for any products in the POS
discount	the current %discount in POS
waiting	the number of waiting days for out of stock SKU (only for POS in OoS report)
type of POS	the owner of POS e.g. Acme, department stores etc

With the assumption that these related data should influence the decision of merchandisers in some way, we added these POS related attributes to the log and rerun Decision Miner. The new results of decision point C (decided by merchandiser) asserted our assumption. The discovered rules is quite long, but the major factors influenced this point are uncovered. They are POS size, waiting days for restocking, current discount at POS and type of POS respectively.

size => if greater than 426, *confirm transfer*
else consider **waiting**
waiting => if greater than 2 weeks, *suspend transfer*
else consider **discount**, **sales for SKU**
discount => if greater than 0% then *confirm tran*sfer
else consider **sales for SKU**, **POS type**
POS type =>if POS belongs to Acme than *confirm transfer* else *suspend transfer*

It was showed that merchandiser for this brand selected POS to be restocked from its size (larger stores tend to be located in target areas), waiting (POSs waits too long may be ignored as they are not prioritized or no products available in any stores), discount (products tend to be moved to POS having current discount promotion as they are likely to be sold) and POS type (POSs owned by Acme itself are given high priority).

Similar to POS, we added product semantic related data to see their effects to decision point 1 and 2 as product code is the major attribute for these two decision points. The added attributes and their descriptions can be found on TABLE 3.

TABLE 3: Semantic Related Data of Product

Properties	Description
SKU	stock keeping unit identified by product code
brand	brand of product
type of product	type of product e.g. shirt, skirt, blouse, pant etc.
season	the season that product was launched e.g. 2, 3, B etc
size	size of product e.g. 2, 3,4, 5 etc.
color	color of product e.g. white, black, brown etc
birthdate	the launching date of this product

After rerun Decision Miner with additional data attributes about product, the discovered rule in decision point1 was not changed. However, rules of decision point2 was changed to

birthdate <= 1217782800000: refill old products
birthdate > 1217782800000 : refill new products

Remark : *birthdate* is changed to numeric format by the C4.5 algorithm

The attribute in the rule is changed from *prd_code* to *bithdate.* We found that *prd_code*, *bithdate* and season are correlated attributes. Without *prd_code* and *birthdate*, the rule is relied on *season* attribute.

season = 3: refill new products
season = B: refill old products
season = 2: refill old products

It was shown that products that have *prd_code* greater than 2006040325088 are products that have *birthdate* > 1217782800000 (numeric format of date) and are products in *season* 3 which is the latest season, so they are executed by *Refill new products* or refill from warehouse 7BP. Adding related data attributes of product to log did not make the algorithm to discover new more rules, but add semantics to the previous discovered rule. In case of decision point 2, it is more readable to analyst to determine product by its *season* or *birthdate* rather than its *prd_code*.

DISCUSSION AND FUTURE WORK

The experiment with the case study above demonstrates the enhancement of results from Decision miner. It was found that semantic related data which are part of domain knowledge should be deserved to be considered in data mining phase. In most data mining research, data miner needs some domain knowledge for data preparation where data set is filtered and organized [12]. This pre-processing step is required so that the results are likely to exhibit real patterns of interest. In our case, Decision Miner simply applies data mining algorithm to a process model discovering from process mining. It does not involve data preparation. Thus, semantic related attributes deriving from domain knowledge, could be added to the input log at the step before which is data preparation step.

However, specifying semantic related attributes in program codes is not suggested as the code cannot be reused with other data sets. It is time-consuming to modify program every time the data set changes. The algorithm to enhance a log file should only provide the mean to add attributes, but not to include the attributes to be added. The practical way to model these semantic related attributes is using ontology. Ontology provides method to model domain knowledge via concepts and relationships. In our case, we can model domain knowledge in form of apparel ontology including concepts like POS, SKU and their relationships. Semantic related attributes of SKU and POS which are part of domain knowledge can be derived from this ontology and they will become the input for enhancing log file. The design of domain ontology is on progress. Once successfully designed, domain ontology can be reused in many semantic analysis tools [4].

Although useful to improve the results of Decision Miner, including semantic related data can make the system overload if there are many attributes and many process instances. The enhanced log file becomes large and it may take much more time to finish process mining. To overcome this problem, we can filter irrelevant attributes and add only most relevant ones before decision mining step. Several algorithms are available in Weka [6] for selecting attributes. These algorithms evaluate, rank or choose the most relevant attributes to supervised class for classification problem. They provide the means to filter out unrelated attributes from the analysis. Including attribute filtering feature is being implemented in our work.

Since data about restocking process are quite a lot, we selected only process data for one brand of products involving only one merchandiser. We also interest to see the decision patterns of other merchandises working for other brands and test how semantic related attributes affect their decision rules. The result of this investigation will be found in our future work.

CONCLUSION

This paper describes the use of process mining to analyze business processes with a real case study in the fashionable products industry. The case study focuses on the “restocking process”, where products are refilled or transferred among points of sale to optimize the opportunity for sale. This process is semi-automatic; some activities are done automatically by the systems and some are done by “merchandisers” who decide to move products from one location to another. Since the decisions of merchandisers strongly affect the opportunity to sell products, studying of their strategies is quite beneficial to the company. The experiments were conducted in 3 steps. Firstly, historic data related to the activities of this process were collected to create a log file. Secondly, this log file was mined with the Alpha algorithm plug-in to discover a process model that represents the real execution flow of the process. Thirdly, the discovered process model was analyzed with the Decision Point Analysis plug-in to locate all decision points that influence the paths of execution. At each decision point, the patterns of attributes and factors that influence the decision were uncovered and analyzed. The experiments showed that no rule was discovered in some decision points and some rules are not easy to understand. Our work proved that the data from the log file are not enough for decision point analysis especially human decision. We introduced the way to improve the analysis results by enhancing log file with semantic related data. It was showed that new rules and more meaningful rules could be disclosed.

Our current and future work will focus on designing domain ontology to derive semantic related attributes and looking for methods in selecting only related attributes to enhance the log file, so that mining algorithm will not be overloaded. The more data sets varying on actors and brands will also be tested.

REFERENCES (BIBLIOGRAPHY)

[1] A. Rozinat and W. M. P. van der Aalst, "Decision Mining in ProM," in *Business Process Management*, 4102, 2006, pp. 420-425.

[2] B. F. van Dongen and W. M. P. van der Aalst, "A Meta Model for Process Mining Data," presented at the Conference on Advanced Information Systems Engineering, 2005.

[3] C. Günter and W. M. P. van der Aalst, "A Generic Import Framework for Process Event Logs," in *Business Process Management Workshops*. vol. 4103, J. Eder and S. Dustdar, Eds., Springer Berlin / Heidelberg, 2006, pp. 81-92.

[4] C. Pedrinaci and J. Domingue, "Towards an Ontology for Process Monitoring and Mining," in *In Proceedings of Semantic Business Process and Product Lifecycle Management in conjunction with the 3rd European Semantic Web Conference*, ed, 2007.

[5] D. Grigori*, et al.*, "Business Process Intelligence," *Computers in Industry,* vol. 53, pp. 321-343, 2004.

[6] E. F. Ian H. Witten, Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques* Third Edition ed.: Morgan Kaufmann, January 2011.

[7] M. Jans*, et al.*, "Business Process Mining for Internal Fraud Risk Reduction: Results of a Case Study," ed, 2008.

[8] M. Lang*, et al.*, "Process Mining for Clinical Workflows: Challenges and Current Limitations," in *MIE*, 2008, pp. 229-234.

[9] N. R. T. van Beest and L. Maruster, "A Process Mining Approach to Redesign Business Processes - A Case Study in Gas Industry," in *Symbolic and Numeric Algorithms for Scientific Computing, 2007. SYNASC. International Symposium on*, 2007, pp. 541-548.

[10] R. Agrawal*, et al.*, "Mining process models from workflow logs," in *In Proc. of Sixth International Conference on Extending Database Technology, pp 469-483*, ed, 1998.

[11] R. S. Mans*, et al.*, "Application of Process Mining in Healthcare - A Case Study in a Dutch Hospital," in *BIOSTEC (Selected Papers)*, 2008, pp. 425-438.

[12] V. Svátek*, et al.*, "Ontology-Enhanced Association Mining," in *Semantics, Web and Mining*. vol. 4289, M. Ackermann*, et al.*, Eds., ed: Springer Berlin / Heidelberg, 2006, pp. 163-179.

[13] W. M. P. van der Aalst*, et al.*, "Business process mining: An industrial application," *Inf. Syst.*, vol. 32, pp. 713-732, 2007.

[14] W. M. P. Van der Aalst*, et al.*, "Workflow mining: discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering,* vol. 16, pp. 1128-1142, 2004.