IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

ASSIMILATION AND DATA FILTRATION OF CORRUGATED INDUSTRY

Aruna Devi .T*, Dr.Kumudavalli M.V, Dr.Sudhamani
* Research Scholar, Rayalaseema University, Kurnool, Andhra Pradesh - India
Dept. of Computer Applications, Dayananda Sagar College, Bangalore - India
Principal, Vivekananda College for Women, Bangalore - India

DOI: 10.5281/zenodo.290141

ABSTRACT

The increased availability of event data in the industry has growing interest in process mining. Industries and organizations are using process mining to know the operational processes which are executed. Process mining can be used to systematically drive innovation in a digitalized world. Process discovery approaches have problems in dealing with fine-grained event logs and less structured processes. Huge amount of data is being collected on a day-today basis pertaining to the manufacturing industry which in turn is updated in the information system. This unprocessed data is the motivation for the present study. The data is collected from the manufacturing industry to create and investigate the event logs. The data is maintained in the industry in various forms. The collection of data and pre-processing of it is a necessary step for building a tailor made event log; which is of a great concern in manufacturing industry. Data pre-processing converts the data into a format which will be simpler and effective process for the purpose of the user in order to help improve the quality of the data for further study. If no quality data, then no quality mining results. Process mining techniques facilitates the industries to produce accurate insights regarding their processes while showing only the desired traits and removing all irrelevant details. It is not easy to reproduce the best practices followed in data creation and collection in many settings due to the quality of raw data and the nature of processes. In this paper we are focusing on available data of a corrugated industry and various filtering techniques that can be used to filter the data and produce industry friendly event logs for further use by the industry.

KEYWORDS: Process Mining, Data, Event logs, Filtering

INTRODUCTION

The amount of data available in industry has increased in recent years. The huge growth of event data is rapidly changing the Business Process Management (BPM) discipline. Without using the valuable information which is hidden in the information system does not make sense to focus on model based analysis. Industries are competing on analytics and only industries that intelligently use the vast amounts of data available will survive in the market. Most real-life logs tend to be incomplete, so we need to perform filtering technique on the event logs. Process mining techniques are used in the industry to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's information systems [1]. Normally, "flat" event logs serve as the starting point for process mining ([1], [2]). These logs are created with a particular process and a set of questions in mind. An event log will be observed as a multiset of traces. For each trace the life-cycle of a particular case is defined in terms of the activities which are executed. Frequently event logs store extra information about events. The event logs have extra information like initiating the activity or resource executing, the timestamp of the event or data elements recorded with the event. Most event logs are incomplete, so we need to perform filtering technique on the event logs. The growth of data in industry is from a wide variety of sources and systems across many domains and applications. The quality of the event logs are poor, i.e., recorded events may not resemble to reality and events may be missing. The manufacturing industry manufactures corrugated boxes in different size and shape with different process involved in it. The data related to the manufacturing industry is stored in the information systems, where there may be incomplete data which should be filtered.

RELATED WORK

Many organizations and industries apply process mining techniques to analyses the process to discover, examine and progress real processes by extracting knowledge from event logs. It is understood that enormous data in data sources is often unclean and thus needs to be cleaned [3]. Data quality problems exist in industry so there is a need for event log pre-processing. For example, the healthcare domain is a prime example of a domain where event logs suffer from various data quality problems. In ([4], [5]) the gynecological oncology healthcare process within an university hospital has been analyzed; in [6] several processes within an emergency department have been investigated; in [7] all Computer Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound, and X-ray appointments within a radiology workflow have been analyzed; in [8]. In total, the indicated problem such as "missing values", "missing events", "duplicate events", "evolutionary change", "noisy data", etc.,

DATA

Data can be described as primary data and secondary data. In the corrugated manufacturing industry the primary data was collected through questionnaires. Broad survey has been conducted, interviews and free discussion with selected respondents was done. Through all this methods the data was collected. The operation process was discussed. Then the secondary data was built from the event logs stored in the information system. The data which has been collected from the industry for specific purpose. The data recorded in information systems are event logs. An event log captures the display of events concerning to the instances of a single process. A process instance is also denoted as a case. Each event in the log relates to a single case and can be associated to an activity or a task. An event log may also convey further information like time, transaction type, resource, costs, etc. Timing information such as date and timestamp of when an event occurred is required to analyses the performance associated to the process. Resource information is analyzed with concern to person executing the activity in the manufacturing industry. We refer to these additional properties as attributes. The event logs of corrugated industry which captures the execution of a process. An event log consist of cases or process instances where each case contains an ordered list of events, a case can have attributes such as case id/job id, etc., each event can be associated exactly to a single case/job, events can have attributes such as activity (i.e., the process of work), time (i.e., timestamp), resource (i.e., person assigned the job), etc.

FILTERING FOR OBTAINING A QUALITY EVENT LOG

The data recorded in information systems are event logs. It is necessary to present these data in an understandable way for manufacturing industry, who cannot deal with large amounts of log data. Event logs [9] is often referred to as sample data or table. The rows in the tables are called instances. Another terms are: individuals, entities, cases, objects, and records. Instances can be related to students, customers, orders, order lines, messages, etc. The columns in the tables are called variables. Variables are frequently denoted to as attributes, features, or data elements. Normally variables have no logical ordering. Before applying any process mining technique the data is typically pre-processed (i.e., filtering the data), e.g., rows and columns may be removed for various reasons. Most event logs are incomplete, so we need to perform filtering technique on the event logs. For instance, columns with less relevant information should be removed before hand to reduce the dimensionality of the problem. Instances that are corrupted should be removed. Also, the value of a variable for a certain instance may be omitted or have the wrong type. This may be due to an error while recording the data, but it may also have a particular reason. Data pre-processing describes any type of processing implemented on raw data to organize it for another processing procedure. Commonly used as a preliminary process mining practice, data pre-processing converts the data into a format that will be more easily and effectively processed for the purpose of the user. Data in the real world is unclean. It can be in incomplete, noisy and inconsistent form. These data needs to be pre-processed in order to help improve the quality of the data, and quality of the mining results. If no quality data, then no quality mining results. The quality result is always based on the quality data. If there is much irrelevant and redundant information present or noisy and unreliable data, then analysis is difficult. Incomplete data can be not applicable data value when collected, different considerations between the time when the data was collected and when it is analyzed, Human/hardware/software problems. Noisy data (incorrect values) can be faulty data collected by human, instruments or computer error at data entry, errors in data transmission. Inconsistent data can be from different data sources.

With regard to the quality of an event log there can be various issues. We distinguish four broad categories of problems.

Missing Data: This corresponds to the scenario where different kinds of information can be missing in a log although it is mandatory. For example, a certain entity of a log such as an event, event attribute/value, and relation is missing. Missing data mostly rejects a problem in the logging framework/process.

Incorrect Data: This corresponds to the scenario where although data may be provided in a log, it may turn out that, based on context information, the data is logged incorrectly. For example, an entity, relation, or value provided in a log is incorrect.

Imprecise Data: This corresponds to the scenario where the logged entries are too coarse leading to a loss of precision. Such imprecise data prohibits in performing certain kinds of analysis where a more precise value is needed and can also lead to unreliable results. For example, the timestamps logged can be too coarse (e.g., in the order of a day) thereby making the order of entries unreliable.

Irrelevant Data: This corresponds to the scenario where the logged entries may be irrelevant as it is for analysis but another relevant entity may have to be derived/obtained from the logged entities. However, in many scenarios such transformation of irrelevant entries is far from trivial and this poses as a challenge for process mining analysis.

During the data collection of the corrugated manufacturing industry some of the above said filtering techniques are used to get quality event log.

METHODOLOGY

Methodology used to collect, create and filter data for corrugated manufacturing industry is as follows:

Stage 1: Survey about the corrugated industry is completed.

Stage 2: Study on the existing system is performed. The behavior of the corrugated industry is analyzed.

Stage 3: Existing manual data which is in records and information system are examined.

Stage 4: For the current study the data from various sources is collected and then pre-processed by using the filtering techniques.

Stage 5: Finally we create quality data as event logs in various formats like CSV, XES (Extensible Event Stream) etc., for further study.

EXPERIMENTAL OUTCOMES

Step 1: Raw data set from the corrugated industry.

Step 2: Analyses is performed and .xls format data sheet is prepared.

Step 3: Filtering is performed on the data to get a quality event logs.

Table 1. XLS Data Sheet

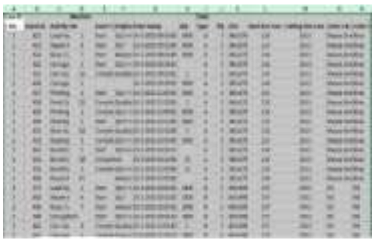
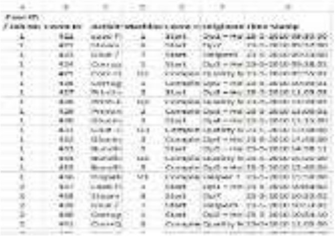


Table 2. Filtered Data



Step 4: Then the pre-processed data is converted to .CSV file.

Table 3. CSV File

Step 5: The .CSV file is imported into the ProM Tool using the button Import.

Figure:

Importing the .csv file in to the ProM tool

Using the ProM Tool the process mining techniques are used to create process model to analyze the process in the corrugated manufacturing industry. The event logs are used in the ProM tool for further analysis. The process followed in the corrugated manufacturing industry is analyzed by using the different techniques which are available for further study.

CONCLUSION

Abundant data is available in industry, it is difficult to extract the desired knowledge from raw event data. Real-life applications of process mining tend to be demanding due to process related issues expressed in event logs and data quality issues. In this paper we have highlighted the techniques to filter the data to get quality event logs. The steps involved to collect, pre-process and import the file into the tool is described pertaining to corrugated manufacturing industry which is a stepping stone for further study of the process so as to improve the production.

ACKNOWLEDGEMENTS

One of the authors Mrs. Aruna Devi T., acknowledges Dayananda Sagar Institutions, Bangalore, Karnataka and Rayalaseema University, Andhra Pradesh, India for providing the facilities for carrying out the research work.

REFERENCES

[1] W.M.P. van der Aalst, “Process Mining: Discovery, Conformance and Enhancement of Business Processes”, Springer, 2011.

[2] Van der Aalst, Wil MP. "Extracting event data from databases to unleash process mining", BPM- Driving innovation in a digital world, Springer International Publishing, 2015, 105-128.

[3] Kim, W., Choi, B.J., Hong, E.K., Kim, S.K., Lee, “A Taxonomy of Dirty Data - Data Mining and Knowledge Discovery”, 2003, 81 – 99.

[4] Mans, R.S., Schonenberg, M.H., Song, M.S., van der Aalst, W.M.P., Bakker, P.J.M, “Application of Process Mining in Healthcare : a Case Study in a Dutch Hospital”, In Fred, A., Filipe, J., Gamboa, H., eds.: Biomedical engineering systems and technologies (International Joint Conference, BIOSTEC 2008, Funchal, Madeira, Portugal, January 28-31, 2008, Revised Selected Papers). Volume 25 of Communications in Computer and Information Science, Springer-Verlag, Berlin (2009) 425 – 438.

[5] Mans, R, “Workflow Support for the Healthcare Domain”, PhD thesis, Eindhoven University of Technology (June 2011), http://www.processmining.org/blogs/pub2011/workflow support for the healthcare domain.

[6] A., Ferreira, D, “Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining”, Information Systems, 37(2), 2012, 99 -116.

[7] Lang, M., Burkle, T., Laumann, S., Prokosch, H.U, “Process Mining for Clinical Workflows: Challenges and Current Limitations”, In: Proceedings of MIE 2008, Volume 136 of Studies in Health Technology and Informatics, IOS Press, 2008, 229 – 234.

[8] Poelmans, J., Dedene, G., Verheyden, G., van der Mussele, H., Viaene, S., Peters. E, “Combining Business Process and Data Discovery Techniques for Analyzing and Improving Integrated Care Pathways”, In: Proceedings of ICDM'10, Volume 6, 171 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2010, 505 – 517.

[9] Aruna Devi. T, Dr.Kumudavalli M.V, Dr.Sudhamani, “2-D Method of Assessment for Corrugated Industry Using Process Mining Approach”, To be presented in ICACET 2017 organized by WARSE, Bangalore, 2017.