



DATA MINING AND PROCESS MINING IN PRINTING HOUSE

Chekanov I.O., Projects manager and Software developer, WPS Hong Kong
Grigoriev A.V., Ph.D., docent, MST Department KhNURE
Zhu Tai Hai, China branch director and Project manager, WPS Hong Kong

Data Mining – is a set of methods for detecting previously unknown, non-trivial, practically useful and accessible interpretation of knowledge, which necessary for making decisions in various fields of activity.

Process Mining – is a set of methods and approaches for analyzing and improving of processes in information systems or business processes that based on the study of event logs.

These two tools, which listed above, are widely used in the analysis of data and processes in large enterprises, regardless of the scope of their activities.

The foundation of the instrumental and theoretical basis for analyzing data and processes consists of various methods of classification, modeling and forecasting. These methods based on the use of decision trees, artificial neural networks, genetic algorithms, evolutionary programming and fuzzy logic. These methods are often referred to statistical methods such as descriptive analyze, correlation analyze, regression analyze, etc. However, it is worth noting that such analysis methods suggest some a priori representations of the analyzed data, which slightly contradicts to the original view of Data Mining.

Data Mining and Process Mining have great importance in monitoring the activities of the enterprise and its software and hardware. This applies to both production divisions and other departments, whose activities are far from production. Therefore, any printing company, regardless of the specifics of its activities, can use methods of analyzing data and processes. Whether it's a printing house or the organization for distributing printing products and services.

In this paper, Data Mining and Process Mining are considered within the polygraphic company Worldwide Printing Solutions, which deals with the manufacturing of printing products. Furthermore, this company deals with delivery wares to the customer and working with publishers around the world. This enterprise includes many offices and printing houses, which are territorially located in the different countries and on the different continents. That's why experimental software for organizing interaction between offices and printing houses among themselves was implemented on this enterprise. The program stores information about customers, products and some processes that occur at printing houses and in the company. The stored information has a practical interest for studying and processing by methods of Data Mining and Process Mining for extracting useful information, analysis and subsequent visualization.

Examples of realization for methods described above, can serve as sociograms. Sociograms show the interaction of people at the enterprise, employees in offices and printing houses among themselves and with people outside the enterprise namely customers and suppliers. Sociograms allow identifying bottlenecks in the structure of



communications and after reorganize communications and units. In case of using sociograms together with diagrams about the employment of workers and the data on the company's profits, become possible to reveal data about the usefulness of departments or individual employees for the company, data about the time of the greatest activity of employees, and so on. The obtained information can be used for improving the efficiency of the company, both in the short-term planning of activities and long-term one.

Another example of using data mining is the creation of algorithms with statistics of the execution of operations, based on information obtained as the result of processing event logs. These diagrams allow identifying bottlenecks in the algorithms of the enterprise, job descriptions, schedules of employees and equipment, as well as schedules of the supplying of materials. The information, which obtained by analyzing algorithms, can be used for correcting these graphs and instructions.

One of the most useful tool, which can be obtained from the analysis of data and processes, is Petri net. This net is a mathematical apparatus for modeling dynamic discrete systems. Petri net is a bipartite oriented multi graph that consisting of two types of vertices - positions and transitions that are connected by arcs. Methods for analyzing of its properties are based on the use of graphs of achievable (covering) markings, solving the equation of network states and calculating linear invariants of positions and transitions. In addition, the application of reduction methods makes possibility of reducing the size of the Petri net with preservation of its properties. Reducing the Petri net allows simplification of the algorithm that is used for enterprise control.

For the introduction of Data Mining and Process Mining methods in the enterprise it was decided to introduce additional program modules for the collection of data sets and for the generation of event logs into the experimental software, that used in the company. XML files and relational databases are used for the storing data sets. The event logs are stored in an XES format file.

The XES format is the XML standard for event logs. It has become a universally accepted format for exchanging data from event logs between different tools and areas of activity. The compatibility of generated reports with third-party applications for the analysis of data and processes served as a justification for choosing this format. For example, the event log in XES format can be analyzed by using the application ProM 6, which provides all basic tools for extracting data, processing. Furthermore, this program can build matrices, algorithms, Petri nets, etc.

To improve the quality of data analysis, it was decided to introduce a system of special flags for denoting semantic blocks and dependencies between structural information units. These structural units are located within the information space of the enterprise's experimental software that was mentioned earlier.

The implementation of Data Mining and Process Mining methods at the enterprise were considered in this paper. These methods will allow to receive more important information in accessible form. It will be useful not only for any levels of planning and control at the enterprise, but also for all spheres of company activity.