Applying process mining techniques and neural networks to creating and assessment of business process models

K. Grigorova[1], K. Mironov[1], E. Y. Malysheva[2]

[1]University of Ruse, Studentska street 8, Ruse, Bulgaria, 7017
[2]Volga Region State University of Services, Gagarina street 4, Togliatti, Russia, 445677

Abstract. The article presents an approach for automated generation of business process models by applying process mining techniques to event logs created during the operation of information systems used in an organization. Existing algorithms for process mining are discussed. Criteria for performing a comparative analysis of these algorithms are specified. A framework is proposed in which to build and analyze business process models. The framework includes tools for initial analysis of the event log file, extracting elements of a business process model, and composing a new model by applying a trained neural network.

Keywords: algorithm, process mining, neural networks, business process model.

1. Introduction

The rapid development of technology coupled with the proven benefits of implementing a process-oriented management approach are the causes of the growing use of business process models in organizational decision-making. Enhanced efficiency, better resource monitoring, and enhanced competitiveness are one of the reasons why business process management is increasingly involved into medium and small business organizations. Modern marketing methods, combined with software solutions that support and facilitate them, convey the understanding of business beyond the organization.

The need for businesses to get more information on how their processes are performed in a real environment is one of the main reasons for the spreading and use of process mining [1]. Modern business management systems and large volumes of data they generate create a favorable environment for the application of various process mining methods.

In addition to applying the most common algorithms and generating new business process models, analyzing system information allows to validate the integrity of the overall process, as well as to provide valuable metrics about its performance [2]. Another important feature of process mining is the ability to be used as a part of methodology for automating life cycle stages of business processes.

This article aims to present an experimental framework that uses process mining methods and enables automated generating and performance evaluating of identical business process models. A multilayer neural network is used for decision making during the business process generation as well as process evaluation according to the specific use case. In the previous related works the use of artificial neural networks is discussed as a tool for extracting business process models from event logs.

The framework described in this paper involves neural network that is used on the stage of decomposing and evaluation of generated process model.

2. Process mining algorithms

The creation of a methodology for retrieving information from system data requires the knowledge of modification and application of the basic approaches of process mining [3]. Following is a review of the most commonly used algorithms for retrieving process models from system data, as well as analysis and classification based on their core characteristics and summarizing the most appropriate algorithms for specific use cases.

2.1. Alpha algorithm

The Alpha algorithm is one of the main in process mining [4]. It is characterized by creating a complete business process model by analyzing and processing event logs. The algorithm works by detecting Workflow nets and merging them into a complete logical structure. Typical for the algorithm is its fast performance, but it is also characterized by some limitations. The presence of a complete log file is required for the correctness of the generated business process model through Alpha miner. Incompleteness or lack of events, as well as describing of some exceptions in the event logs, may lead to incorrect final results. The Alpha algorithm is suitable for use in conjunction with systems that generate complete and well-structured event log files.

2.2. Heuristics miner

This algorithm is considered to be an extended version of the Alpha algorithm, but provides additional options for analyzing system event logs. The application of Heuristics miner is characterized by analyzing log data, sequencing events and determining the occurrence frequency. Based on these data, a network is created that consists of linked sequences and reflects the basic behavior of the workflow. Unlike the Alpha algorithm, Heuristics miner can be used for processing event logs with missing, incomplete and unassigned data, and even there are records of "emergency" situations.

2.3. Genetic miner

The genetic algorithm uses an evolutionary selection approach that mimics the process of natural evolution. This algorithm is significantly slower but offers optimal results based on system event log files [6]. It is characterized by analyzing and identifying initial "process candidates" in the form of related events, most often represented by Petri nets. Genetic miner has a cyclical character, and each of the potential processes is compared, and the one that comes closest to the general scenario in the event log is chosen. The genetic algorithm is suitable for use in system logs with missing or incomplete data, as well as with repeating events. It is not suitable for real-time application, and the final result is a detailed and very accurate business process model.

2.4. Fuzzy miner

Very often, business process models that are generated by event log analysis are large, unstructured and chaotic, without the ability to accentuate important events. The fuzzy algorithm is configurable and allows compact representation of detected patterns from different viewpoints (user or system) [7]. This is done through a set of techniques for identifying and removing isolated nodes, as well as merging several linked nodes into one. As a result of Fuzzy miner performance a logically true and identical copy of the original unstructured process is obtained, which is easy readable [5]. It is suitable for application on complete system log files without missing or incomplete information.

2.5. Selection of algorithm comparison criteria

When comparing algorithms for Process mining in the context of automated creation and assessment of business process models, consideration should be given to criteria that affect the program implementation as well as the number of characteristics of the input and output data. For this purpose, event logs and generated processes are classified according to the following criteria:

- Integrity - this characteristic indicates the ability of the algorithm to work in the case of missing event data.
- Uniqueness – indicates the ability of the algorithm to handle repeating events within a single process.
- Exceptions handling - indicates whether the algorithm can work normally when there are some exceptions.
- Real-time - indicates the ability of the algorithm to be used in real-time on a running system.
- Model integrity - indicates whether the generated model is a complete business process model or only a part of it (a series of related tasks).
- Abstraction level - indicates whether the obtained business process model is generalized and simplified.
- Structured model - indicates whether the obtained business process model is structured and suitable for reading by a human.

2.6. Comparative analysis of process mining algorithms

Automated generation and evaluation of business process models requires the processing event logs that differ by type and structure and the creation of a unified method for working –reading and analyzing the event log files. In order to determine the most appropriate algorithms for different use cases, process mining algorithms are compared and analyzed on criteria affecting their input and output data.

Using the specified criteria and taking into account the characteristics of process mining algorithms discussed in the previous sections, the results of the comparative analysis are summarized and presented in Figure 1.

	Alpha Algorithm	Heuristics Miner	Genetic Miner	Fuzzy Miner
Integrity		✓	✓	
Uniqueness	✓	✓	✓	✓
Exceptions		✓	✓	
Real-Time	✓			
Model Integrity	✓	✓	✓	✓
Abstraction level				✓
Structured model				✓

Figure 1. Results from comparative analysis.

In the context of the framework created, when generating a business process model from an event log file, a preliminary analysis of the input data should be performed in order to determine which of the algorithms considered is the most appropriate for the specific case.

3. Framework for automated creation of business process model and evaluation of its performance

The automated creation of business process models combines the optimization and reengineering of existing models. Based on the fact that an activity can be implemented in several different ways, the purpose of the framework is to assemble different business process models that perform a particular business task. Components of each business process are already available business tasks extracted from system event log files. Each business process model will be evaluated for performance in order to identify the most effective for the considered business task.

Figure 2 shows a general architecture of the proposed framework. Each stage is considered and its main features are presented.

3.1. Pre-analysis

At this stage, a preliminary analysis of the system event log file to be processed is performed. The purpose is to determine its integrity and type. Based on the results of the preliminary analysis as well

as those from the comparative analysis, an optimal process mining algorithm is selected to be applied in the specific case.

3.2. Extracting business process model

The selected process mining algorithm is applied to the event log file and a business process model is generated. In order to avoid future errors, the model is validated and its integrity is checked. In case that a correct process is retrieved, it is simulated and evaluated for performance. These data will be used in the future to determine the most efficient process.

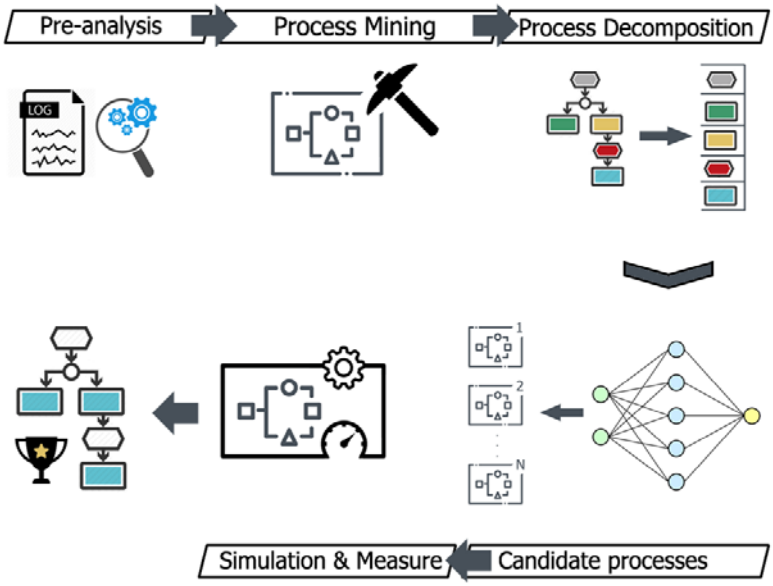

Figure 2. Framework's architecture.

3.3. Business process decomposition

The decomposition consists of segmenting the separate components of the model and storing them as independent elements. Intra-system storage of the elements allows the use of non relational model and, depending on the amount of business components, the application of Big data methods is recommended.

3.4. Generation of "candidate" processes

Creating processes that will be simulated and evaluated is performed using the decomposed business elements from the previous step and with the help of a neural network. Depending on the elements found in the decomposition of the model, the number of layers and neurons in the neural network may vary. The neural network training is based on the dependencies found in the log file, the relationships between the individual components of the process as well as the initial design and the subsequent redesign of the processes.

As a result of this phase, there are several logically identical "candidate" processes that perform the same functionality as the primary process generated by applying process mining algorithm.

3.5. Simulation and assessment

At this stage, it is intended to identify the most appropriate business process model for the particular business task. For this purpose, simulation is performed on each business process in a controlled and preconfigured work environment. During the execution of each process, its behavior is monitored and data is collected that will help to generate baseline assessment of performance.

3.6. Framework operation

After selecting a system event log file, which is the input data of the developed framework, a preliminary analysis is performed on it that aims to determine its structure, integrity, and availability

of information about recurring events. After the analysis and on the basis of its results, a process mining algorithm is chosen which is most appropriate for the specific case. The selected algorithm is applied to the input log data and a business process model is derived. It is presented in XML-based format that allows easy processing. Additionally, metadata about the process logic is created used for initial neural network training.

The next step is decomposing the resulting business process model of components and storing them in a data repository, created as Redis database. Next the separate components are used to create "candidate" processes by using a neural network. The number of layers and neurons in the network is determined dynamically depending on the case and the business process model under consideration. As an environment Java neural network framework Neuroph is used.

The generated "candidate" processes are simulated, taking into account their productivity metrics during the simulation. After comparing the results obtained, the most effective business process model for the particular business task can be determined.

4. Conclusion

The development of technology and increasing competitiveness in business environments contributes to the partial or complete automation of business tasks. The variety of business process modeling systems and the growing tendency for their use by business creates favorable conditions for the application of process mining techniques. This article offers a solution for automated generating and evaluating the performance of business process models.

The presented framework uses process mining and artificial intelligence techniques and aims to automate part of the monitoring and optimization activities. As a final product of the examined framework, a business process model is generated that is the most effective for a particular business task. This increases the organization's competitiveness, and the automation reduces the resources used.

5. Acknowledgements

This work is supported by the National Scientific Research Fund under the contract DFNI - I02/13.

6. References

[1] Wang, J. Efficient selection of process mining algorithms/ J. Wang, R.K. Wong, J. Ding, Q. Guo, L. Wen // IEEE Transactions on Services Computing. – 2013. – Vol. 6(4). – P. 484-496.

[2] Van der Aalst, W.M. Business process mining: An industrial application / W.M. van der Aalst, H.A. Reijers, A.J. Weijters, B.F. van Dongen, A.A. De Medeiros, M. Song, H.M.W. Verbeek // Information Systems. – 2007. – Vol. 32(5). – P. 713-732.

[3] Tiwari, A. A review of business process mining: state-of-the-art and future trends / A. Tiwari, C.J. Turner, B. Majeed // Business Process Management Journal. – 2008. – Vol. 14(1). – P. 5-22.

[4] Van der Aalst, W.M. Workflow mining: Discovering process models from event logs / W.M. van der Aalst, T. Weijters, L. Maruster // IEEE Transactions on Knowledge and Data Engineering. – 2004. – Vol. 16(9). – P. 1128-1142.

[5] Weijters, A.J.M.M. Process mining with the heuristics miner-algorithm / A.J.M.M. Weijters, W.M. van Der Aalst, A.A. De Medeiros // Technische Universiteit Eindhoven, Tech. Rep. – 2006. – Vol. WP 166. – P. 1-34.

[6] De Medeiros, A.A. Genetic process mining. Applications and Theory of Petri Nets / A.A. De Medeiros, A.J.M.M. Weijters // Lecture Notes in Computer Science. – 2005. – Vol. 3536.

[7] van Dongen, B.F. Mining: Fuzzy Clustering and Performance Visualization / B.F. van Dongen, A. Adriansyah // Business Process Management Workshops. – 2009. – Vol. 43. – P. 158-169.