Framework of Using Data Mining for Business Process Improvment

Mohammad Khanbabaei
PhD Student, Department of Information Technology Management, Science and
Research Branch, Islamic Azad University, Tehran, Iran,
mohammadkhanbabaei@srbiau.ac.ir
Farzad Movahedi Sobhani[*]
Assistant professor, Department of Industrial Engineering, Science and Research Branch,
Islamic Azad University, Tehran, Iran, farzadmovahedi@gmail.com
Mahmood Alborzi
Associate Professor, Department of Information Technology Management, Science and
Research Branch, Islamic Azad University, Tehran, Iran, m.alborzi@srbiau.ac.ir
Reza Radfar
Associate Professor, Department of Technology Management, Science and Research
Branch, Islamic Azad University, Tehran, Iran, r.radfar@srbiau.ac.ir

Abstract: Organizations should consider process improvement for promoting their performance. The main problem is the high volume of process dataset along with the wide variety of their features that leads to increase the complexity in process improvement methods. Previous methods cannot improve processes in a high volume of process dataset. Data mining approach can support improvement methods by identifying valuable hidden patterns in high volume of process dataset. In this paper a framework of using data mining techniques for extracting valuable hidden patterns in high volume of process dataset is developed for presenting the improvement suggestions. In order to evaluate the proposed framework, a real set of processes along with their features were gathered. After that, using classification, clustering, and features selection algorithms, valuable patterns in high volume of process dataset were identified. After evaluating these patterns, improvement suggestions were recommended by the identified patterns. The results showed that identified patterns can support processes improvement activities with recommending the improvement suggestions.

Keywords: Process Improvement, Data Mining, Business Processes.

Introduction: Processes are one of the most important resources in organizations. Improving processes can lead to enhance the organizational performance. Several methodologies are presented for process improvement. However, they did not consider the problem of high volume of process dataset along with their features in organizations. There are many processes in organizations that lead to increase the complex interactions between processes and high dimensionality problem (Jeong et al., 2008). In addition, there is a single view to processes in the improvement actions (Houy et al., 2011). Huang et al. (2012) stated that there is a low attention to the internal aspects of the processes. In this situation, data mining techniques can identify and discover valuable patterns hidden in the large number of processes in the organizations. These patterns can be utilized for recommending the improvement suggestion for enhancing the performance of the processes. Recently, there is a link between data mining and process improvement. A few studies considered the interaction between data mining and process improvement. However, these studies did not consider a real large number of processes in their computations. In addition, they did not have a comprehensive and practical view to the application of data mining for process improvement. This paper presents a framework for using data mining techniques for identifying and extracting valuable and suitable patterns hidden in the large number of processes. These patterns can be employed for recommending the process improvement suggestions.

Materials and Methods: This paper employs three data mining techniques including clustering, classification, and feature selection techniques for extracting valuable patterns hidden in the large number of processes. CRISP-DM (cross industry standard process for data mining) standard is used to implement data mining activities. In the classification technique, C5 decision tree algorithm is employed to classify processes. In clustering, K-means algorithm is applied to segment processes in several clusters. In feature selection technique, the most important process features are selected for process improvement. In the proposed framework, at first, all processes are gathered from a variety of the resources in the organization. After that, process features are identified from the literature and they are defined based on an interaction between data miner and process improvement expert. In the following, a process dataset is provided for using data mining techniques. For this dataset, a variety of data preparation and preprocessing methods are employed for achieving better results of implementing data mining techniques. After that, three classification, clustering, and feature selection data

---

[*] Corresponding author

mining techniques are applied for extracting valuable patterns hidden in the large number of processes. In the classification techniques, C5 decision tree algorithm employs a target features for classifying processes. Cross validation method is applied to train and test the constructed decision tree. The output is several if-then rules for classifying processes based on the target feature. In clustering, K-means algorithm segments processes in several clusters. Processes in a cluster have a similar behavior and they are dissimilar to the processes in the other clusters. Euclidean distance function is used for calculating the distance between processes. The output is the cluster profiling to describe the behavior of processes in each cluster. In the feature selection, the most important features are selected based on a target process feature. These features are more correlated to the target process feature. In addition, they are more important to consider for process improvement purposes. The output is a variety of more important process features so that they can be considered for recommending the process improvement suggestions. After extracting valuable patterns hidden in the large number of processes, the accuracy and quality and these patterns are evaluated by an interaction between data miner, process owner, and process improvement expert. Evaluated patterns are considered for recommending the improvement suggestions. These suggestions must be aligned with the process improvement concepts in the organization. Processes are improved based on these suggestions. In last, the performance of the improved processes is evaluated. The proposed framework is based on an iterative and continuous method for using data mining in process improvement.

Results and Discussion: The proposed framework was evaluated based on a real process dataset including 1318 processes and 80 process features. Several preprocessing methods were employed to prepare dataset. Three classification, clustering, and feature selection techniques were applied to extract valuable patterns hidden in the large number of processes. Using C5 classification algorithm, processes were classified by the if-then rules based on a target process feature. In the proposed model, as an example, key processes were classified using C5 decision tree. The classification accuracy in the test dataset was set to 92.31%. The output was several if-then rules to classify key processes. These rules can identify key processes. In addition, processes features applied to construct the if-then rules can be employed for recommending the improvement suggestions for key processes. In clustering, using K-means algorithm, processes were segmented into 10 clusters. Some features were considered to cluster processes. The output was cluster profiling to describe the behavior of processes based on the selected features. In this paper, for example, the behaviors of processes in cluster 1 were described through the selected features. Several suggestions were recommended by the selected features for improving processes in this cluster. In the feature selection technique, for example, "cost of process" was considered as the target process feature. The output was the selection of 10 more important process features that they were more related to the target process features. These features can describe the cost of processes better than the other features. The improvement suggestions for reducing cost of processes were recommended based on these selected features.

Conclusion: This paper presented a framework of using data mining techniques for identifying valuable patterns hidden in the large number of processes for the process improvement purposes. A real process dataset was employed to evaluate the applicability and effectiveness of the proposed framework. In the proposed framework, classification, clustering, and feature selection data mining techniques were applied to extract valuable patterns hidden in high volume of process dataset. Process improvement methodologies cannot recommend the improvement suggestions in a rapid and accurate method, when there are many processes along with a variety of the process features. These methodologies are restricted to recommend a limited number of the process improvement suggestions. In the other direction, there are few studies on the application of data mining for the process improvement that they include some weaknesses. The proposed framework employed a lot of real processes in the data mining techniques to discover valuable patterns for process improvement. In addition, a variety of the process features were extracted from the literature to describe the behavior of the processes. In last, a wide variety of the several suggestions were recommended for process improvement based on the extracted patterns hidden in the large number of processes. The organizations can utilize the proposed framework for improving their processes. In addition, this framework can help organizations with the large number of processes for employing the process improvement methodologies in a productive manner. Future studies can apply the other data mining techniques for the proposed framework. In addition, the proposed framework can be developed for knowledge-intensive processes. The proposed framework can be integrated with the knowledge management methodologies for improving knowledge-based processes. In last, the proposed framework can be enhanced as a decision support system for the process improvement.

References
Houy, C., Fettke, P., Loos, P., van der Aalst, W.M.P., Krogstie, J. (2011). Business Process Management in the Large. *Business and Information Systems Engineering*, 3, 385–388.
Huang, Z., Lu, X., Duan, H. (2012). Resource behavior measure and application in business process management. *Expert Systems with Applications*, 39 (7), 6458–6468.
Jeong, H., Song, S., Shin, S., Rae Cho, B. (2008). Integrating data mining to a process design using the robust bayesian approach. *International Journal of Reliability, Quality and Safety Engineering*, 15 (05), 441–464.