Developing an integrated framework for using data mining techniques and ontology concepts for process improvement

Mohammad Khanbabaei , Farzad Movahedi Sobhani ,
Mahmood Alborzi , Reza Radfar

PII: S0164-1212(17)30261-3
DOI: 10.1016/j.jss.2017.11.019
Reference: JSS 10067

To appear in: *The Journal of Systems & Software*

Received date: 17 September 2016
Revised date: 3 September 2017
Accepted date: 6 November 2017

Please cite this article as: Mohammad Khanbabaei , Farzad Movahedi Sobhani , Mahmood Alborzi , Reza Radfar , Developing an integrated framework for using data mining techniques and ontology concepts for process improvement, *The Journal of Systems & Software* (2017), doi: 10.1016/j.jss.2017.11.019

Highlights

- A framework using data mining and ontology concepts for process improvement.
- An integrated three-part, five-stage framework for process improvement.
- Extracting process ontologies using data mining in a high volume of process data.
- Recommending process improvement suggestions based on the process ontology.

Developing an integrated framework for using data mining techniques and ontology concepts for process improvement

Mohammad Khanbabaei
Department of Information Technology Management,
Science and Research Branch,
Islamic Azad University,
Tehran, Iran.
mohammadkhanbabaei@srbiau.ac.ir

Farzad Movahedi Sobhani
Department of Industrial Engineering,
Science and Research Branch, Islamic Azad University,
Tehran, Iran.
*Corresponding author
fmovahedi@iau.ac.ir
Address: Science and Research Branch, Daneshgah Blvd, Simon Bulivar Blvd, Tehran, Iran
Post Office Box: 14515/775
Postal Code: 1477893855
Phone Number: +98021-44865179-82 & +98021-44865154-8

Mahmood Alborzi
Department of Information Technology Management,
Science and Research Branch, Islamic Azad University,
Tehran, Iran.
m.alborzi@srbiau.ac.ir

Reza Radfar
Department of Technology Management,
Science and Research Branch, Islamic Azad University,
Tehran, Iran.
r.radfar@srbiau.ac.ir

Abstract: Process, as an important knowledge resource, must be effectively managed and improved. The main problems are the large number of processes, their specific features, and the complicated relationships between them, which all lead to the increase in complexity and create a high-dimensionality problem. Traditional process management systems are unable to manage and improve processes with a high volume of data. Data mining techniques, however, can be employed to identify valuable patterns. With the aid of these patterns, suggestions for process improvement can be presented. Further, process ontology can be applied to share the process patterns between people, facilitate the process understanding, and develop the reusability of the extracted patterns for process improvement.

This study presents a combined three-part, five-stage framework of data mining, process improvement, and process ontology. To evaluate the applicability and effectiveness of the proposed framework, a real process dataset is applied. Two clustering and classification techniques are used to discover valuable patterns as the process ontology. The output of these two techniques can be considered as the recommendations for improving the processes. The proposed framework can be exploited to support process improvement methodologies in organizations.

Key words: Data mining, Process improvement, Ontology, Classification, Clustering

1. Introduction

Nowadays, the vast majority of large organizations possess hundreds of different business processes (BPs), which are typically poorly documented. Furthermore, the relationships between the different types of processes are not clearly specified (Houy et al., 2011). Rebuge and Ferreira (2012) presented characteristics of processes, including the following: the dynamic and changing nature, complexity, interdisciplinary nature of the processes; interactions between the processes in different departments; and the requirement for obtaining experience, knowledge, and expertise for implementing the processes.

Moreover, they stated that the traditional analysis of processes was time consuming. In addition, creating a common understanding of the processes between employees was difficult. The other problems related to this research work are presented in Table 1.

Table 1. Problems related to the research work

Author (s) (year)	Problem
Lepmets et al. (2012)	Evaluating and improving processes without considering the effect of one process on other processes
Jeong et al. (2008)	1. Employing statistical methods for process improvement (PI) in the past 2. Occurrence of a high-dimensionality problem due to the increasing the number of process features (PFs) and data related to the processes
Huang et al. (2012)	The lack of knowledge regarding the internal aspects of the processes
Darmani and Hanafizadeh (2013)	The problems of PF selection
Houy et al. (2011)	1. Mainly focusing on single isolated processes in past investigations of business process management (BPM) 2. The requirement to address a large set of organizational processes interrelated to each other in the near future 3. Increasing the complexity of the processes 4. Concentrating on BPM on a large scale
Delgado et al. (2014)	1. The lack of an integrated view and complete picture for analyzing BPs based on process information 2. The absence of intelligence in the majority of PI methodologies
Vukšić et al. (2013)	1. There was no overlapping relationship between the data and the process in traditional business intelligence (BI) approaches 2. Combining BPM and BI for improving the performance and decrease excessive costs resulting from the separate implementation of these two approaches

With respect to the above-mentioned problems, data mining (DM) techniques can extract valuable patterns hidden in the high volume of BPs for recommending PI suggestions. In this regard, a process ontology concept can be considered to share the patterns extracted from the application of DM in PI to gain process ontology benefits.

Pivk et al. (2014) stated that applying ontology for implementing DM in BPs has benefits: (1) sharing a common understanding between people, (2) re-using the domain knowledge, (3) making explicit domain assumptions, (4) differentiating between domain and operational knowledge, and (5) analyzing the domain knowledge.

This paper contributes an integrated three-part, five-stage framework for applying the DM approach for PI under a process ontology concept. Figure 1 exhibits the conceptual model of the proposed framework including three parts as follows: 1. process ontology, 2. DM, and 3. PI. Each part has a series of attributes and behaviors. The attributes explain the characters of these parts and describe what they are. The behaviors describe the activities that these parts can implement.

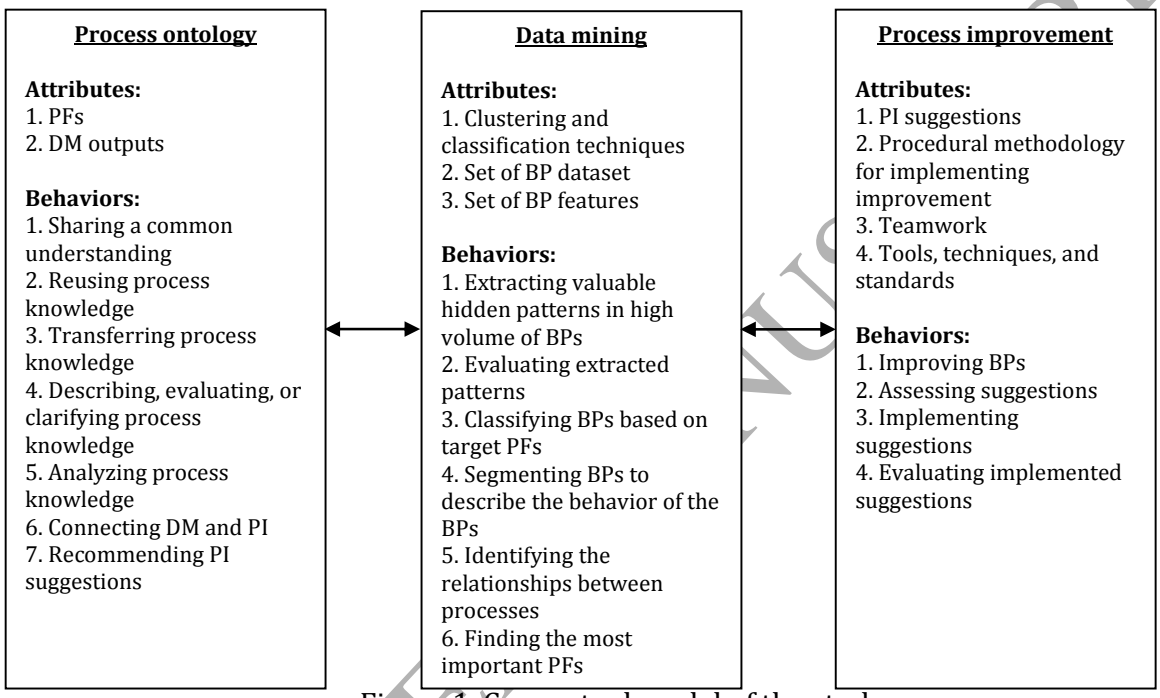

Figure 1. Conceptual model of the study

With respect to the conceptual model displayed in Figure 1, clustering and classification DM extract valuable hidden patterns for PI in the high volume of BPs. These patterns are identified as process ontologies and include two types, PFs and outputs of the DM. The process ontologies are exploited to share a well-understood, explicit, and reusable process knowledge (extracted using DM techniques) for PI in the organization. Using these ontologies, PI suggestions can be recommended.

Moreover, the process ontologies can act as a connection between the DM and PI approaches. This connection is established such that DM extracts patterns from the high volume of BPs. These patterns are considered as process ontologies where they can recommend PI suggestions.

An actual, real BP dataset including a variety of PFs is applied to evaluate the effectiveness of the proposed framework. The operational stages of the proposed framework are explained in additional detail in Section 5.

Based on the conceptual model, in the proposed framework, several mutual relationships are established to connect the components of each of these parts. These

relationships and the operational stages create an integrated framework for using the DM approach for PI under the concept of process ontology.

2. Research bases

It must be indicated that our study was performed with knowledge of four previously director researches; these can be considered as the foundations of this paper. However, they do not address important issues for increasing the applicability and effectiveness of the concept of using DM techniques for improvement in the high volume of BPs.

In the first research, Rupnik and Jaklic (2009) presented a framework for using DM in operational BPs. In the second study, Wegener and Rüping (2010) described the integration between DM techniques and BPs as a critical issue in today's business environment. They stated that integrating DM methods in BPM frameworks is not a straightforward matter. Afterward, they integrated DM and BPs and evaluated this integration in the business process reengineering (BPR) context.

In the third research, Ghanadbashi et al. (2013) employed DM techniques for BPR based on the simultaneous use of the two following approaches. In the first approach, a literature review on the application of DM in the phases of the BPR-consolidated methodology proposed by Muthu et al. (2006) was presented. Then, a DM model based on BPR (DMbBPR model) was proposed to integrate DM techniques in each phase of the BPR.

In the second approach, a new combinational model of the Cross Industry Standard Process - Data Mining (CRISP-DM) standard, knowledge management (KM), and process monitoring architecture was presented. Also, the model was evaluated using a sample SIPOC (supplier-input-process-output-customer) dataset.

In the fourth research, Pivk et al. (2014) proposed an approach employing DM in BPs. This study was a follow-up to the study presented by Rupnik and Jaklic (2009). In this study, Pivk et al. (2014) explained that in the utilization of DM in BPs, ontology is beneficial in the following aspects:

1. BP ontology explains the PFs such as inputs, outputs, effects, constraints, and activities and 2. DM ontology describes the DM process and can be applied in selecting the DM algorithms and searching for the respective methods.

Table 2 presents the four mentioned researches including the director issues derived from these studies to develop the originalities and scientific evolutions of the proposed framework. Further, the weaknesses related to the director researches that can be covered by the proposed framework are described in Table 2.

Table 2. Director issues and weaknesses related to the four director researches considered as the foundations of the current paper

Research	Director issues	Weaknesses points
Rupnik and Jaklic (2009)	1. Defining CRISP-DM standard in their proposed framework	They only highlighted the concept of using DM in BPs and its importance in general; they did not consider the detailed operational activities.
Wegener and Rüping (2010)	1. Defining CRISP-DM standard in their proposed framework 2. Defining the role of business users, and information technology (IT) and DM experts in each stage of their proposal	In this study, only a theoretical description of the relationship between DM and the BP roles was considered; the technical dimensions, output of the DM, and their application in BPM were not addressed.
Ghanadbashi et al. (2013)	1. Defining DM activities in phases of BPR methodology 2. Integrating CRISP-DM standard with process monitoring architecture	1. The study did not employ a real process dataset and a large number of PFs. 2. The proposed model did not recommend PI suggestions resulting from the extracted patterns of DM.

	3. Using a sample process dataset for extracting hidden patterns	
Pivk et al. (2014)	1. Defining CRISP-DM standard in their proposed framework 2. Exploiting the ontology concept for the utilization of DM in BPs	1. The study was more related to the process mining concepts and employed DM in BPs to a lesser extent. 2. Their study did not employ the ontology concept in a more applicable and extensive method than the proposed framework in the current study.

The remaining structure of this study follows. The main concepts of the paper are presented in Section 3. In Section 4, related works and a comparison with the current study is provided. The proposed frameworks along with a case study are explained in the subsequent sections. In Section 7, the discussion and conclusions of the paper are presented.

3. Background

3.1. Process improvement

According to (Gómez-Pérez et al., 2010), process can be defined as a specific concept including pre-assumptions, results, contents, actors, and reasons. In another definition, process is expressed as a set of events implemented in connection with the characteristics of a system or an object. Processes are associated with a set of operations, consisting of events, time, place, expertise, and other resources that ultimately result in producing one/several outputs. Borrego and Barba (2014) stated that a process is a set of activities implemented in a technical and organizational environment for realizing the business objectives.

In this field, PI plans can support processes via different methods, techniques, and software, to design, approve, control, and analyze the operational processes. There are several methodologies to improve processes.

Tonchia (2004) presented the stages of PI as follows: 1. identifying the processes for improvement, 2. defining an intervention team, 3. analyzing the current processes and improvement methods, 4. implementing the improvement actions, and 5. evaluating the results.

Further, Damij and Damij (2014) clarified that PI concentrates on improving the function of the current processes by searching the appropriate methods and solutions for increasing the performance and quality and reducing costs.

These methods include: removing bureaucracy, analyzing value add, removing duplicates, simplifying methods, shortening cycle times, correcting errors, upgrading processes, and simplifying the language of processes, standardization, supplier participation, automation, and IT.

3.2. Data mining

Data mining is the process of selecting, discovering, and modeling high volume of data to find and clarify unknown patterns (Koh and Chan, 2004). Lee and Siau (2001) stated that these patterns are explicit, useful, and potential. Larose (2005) presented CRISP-DM as a standard process for implementing DM. The stages of the CRISP-DM standard are as

follows. 1. Business understanding. 2. Data understanding. 3. Data preparation. 4. Modeling. 5. Evaluation. 6. Deployment.

There are several DM techniques, which include the following: feature selection, clustering, classification, estimation, prediction, association rules, time series analysis, segmentation, trend analysis, deviation detection, and profiling. Next, two clustering and classification techniques applied in the current study are explained in detail.

First, clustering, this is one of the descriptive DM techniques, segments the records of a dataset into several clusters. For this purpose, the distance between two records is calculated so that the records in a cluster are similar and dissimilar to the records in other clusters (Larose, 2005). K-means clustering is the most frequently used clustering algorithm. This algorithm uses the Euclidean distance function for calculating the distance between records in an iterative manner to achieve better performance. In this paper, K-means clustering is used to segment BPs.

Second, the decision tree classification algorithm, which is one of the most applicable DM techniques, classifies the records of a dataset based on a target variable. The output of this algorithm is a set of if-then rules in a tree shape. This algorithm uses the top-down inference process based on the recursive partitioning method (D'heygere et al., 2003).

The decision tree consists of several nodes, branches, and leaves. Each node introduces one variable (feature). The branches divide the dataset into a smaller dataset until leaves can be observed at the end of the tree. For this purpose, the cross-validation technique is employed to create two training and test datasets (D'heygere et al., 2003). In the current work, the C5 decision tree algorithm is applied for classifying the processes.

3.3. Ontology

According to Kharbat and El-Ghalayini (2008), Gruber (1995) defined ontology as a formal and explicit characteristic of a shared concept. Each ontology must possess such specifications. Singh et al. (2010) stated that ontologies are applied in IT for a systematic presentation of knowledge in one domain. Ontologies must be designed in a manner to be generalized in practice. They define a set of classes, relationships, functions, and objects for a domain.

To solve problems, different ontologies have been defined. Different methodologies were provided to design the ontologies. Several researchers have already compared these with each other. One of the main ontologies is the organizational ontology. It is an organizational model to represent processes, information, resources, people, behavior, goals, and constraints (Rao et al., 2012). Figure 2 presents an organizational ontology including its components.

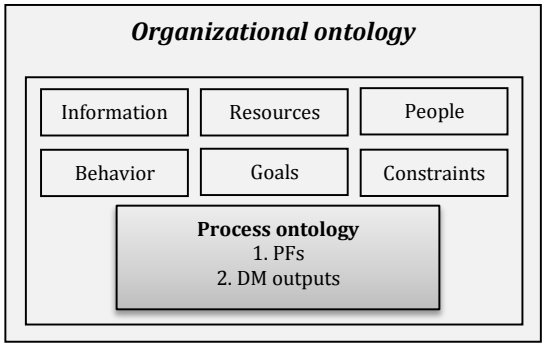

Figure 2. Organizational ontology and two types of the process ontology

With respect to the above-mentioned definition, this paper only considers the process aspect of the organizational ontology (see Figure 2). Gangemi et al. (2004) cited "process ontology is a description of the components and their relationships that make up a process". Key concepts of process ontology can be type of activity, role and authority, agent for implementing the process, resources for executing the process, activity, and the organizational unit.

Through process ontology, the relationships between organizational units and relationships between the components of the processes can be identified. Process ontology can recognize the relationships between PFs and facilitate the understanding of the process-related concepts for employees.

As indicated in Figure 2, in this study, ontology is referred to as process ontology and can be defined by two types.

1. A set of components that represent a process. These components can be nominated as the "process features". These features explain the knowledge and behavior of the processes. Using these types of ontologies, models can be designed such that the process samples can be embedded in the models in a fashion to segment (for clustering objective) and classify (for classification objective) the processes.

2. The output of DM that can be used for recommending improvements for processes. These outputs can transfer knowledge, are explicit and reusable, and are comprised of concepts.

4. Related works

This section presents selected previous studies on DM and related issues with improving BPs. There are three studies presented, which are as follows:

- Previous studies on using DM for PI
- Prevalent PI methodologies
- Proposed framework of the current study using DM and process ontology for PI

First and second studies have some weaknesses and so the proposed framework in the current (third) study addresses these weaknesses. A comparison between the proposed framework and the two studies is presented.

First, a number of researchers have considered the use of DM in PI, BPM, or BPR concepts. They are presented in Table 3.

Table 3. Previous studies on using DM for PI and other related concepts

Author (s) (year)	Contribution
Chen and Wang (1999)	Developing an integrated DM system to analyze process operational data for applications such as pattern detection, trend and deviation analysis, affiliation and link analysis, summarization, and sequence analysis
Grigori et al. (2004)	Presenting the concept of BP intelligence as an application of BI in BPs
Folorunso and Ogunde	Applying DM as a technique to support BPR by extracting hidden knowledge from a high volume of

(2005)	data
Zhonghua and Limei (2008)	Applying DM for BPR to analyze data by identifying the key processes, examining the key success factors, and improving the information flow and process feedback
Rupnik and Jaklic (2009)	Demonstrating a DM framework for operational BPs
Marjanovic (2012)	Developing a relationship between operational BI and operational BPs
Wegener and Rüping (2010)	Integrating DM to BPs in the context of BPR
Sohail and Dhanapal Durai Dominic (2012)	Recognizing a model to diminish the gap between process intelligence and PI by considering the process logs
Mathew and George (2012)	Employing DM to support BPR using the extracted hidden knowledge from a high volume of data
Ghattas et al. (2014)	Demonstrating a semi-automatic approach to improve the performance of processes with respect to the decisions of past processes using DM techniques
Groger et al. (2014)	Establishing a prescriptive analysis for PI based on recommendation systems using DM
Pivk et al. (2014)	Exhibiting an approach based on ontology and service-oriented architecture for implementing a DM process to optimize BPs

The above-listed studies had weaknesses and the present study aims to overcome these, as shown in Table 4.

Table 4. Comparison of previous studies with the current study for using DM to improve processes

Weaknesses of previous studies using DM for improving BPs	**Ways to mitigate the weaknesses using the current study using DM and process ontology for PI**
Lack of adequate understanding of organizational processes and a unified and exhaustive overview of the process information Inability to collect, update, and easily access the information of all processes in the organization	Use of a real BP dataset from all organization departments to obtain the complete overview of BPs
The absence of sufficient consideration of the peripheral issues of processes, such as PFs	Apply a broad variety of PFs to describe the behavior of processes obtained based on a comprehensive literature review of two concepts: 1. BPM and 2. KM for the processes in the organization
Lack of application of a high volume of BP dataset in the computations	Employ a large incorporated dataset of processes to extract valuable patterns for PI
Describe only the theoretical relationship between DM and BPs	Comprehensively consider the technical results of using DM for PI

Second, the proposed framework overcomes the weaknesses of prevalent PI methodologies, as shown in Table 5.

Table 5. Comparison of prevalent PI methodologies and the proposed framework

Weaknesses of prevalent PI methodologies	**Ways to overcome the weaknesses by the proposed framework**
Requirement of intervention and interview of experts for the identification of processes on the interactions between departments and employees Difficulty in team construction in contrast with high quantities of data	Employ the high volume of data and quickly extract valuable patterns, which can be used to identify and analyze the processes
Low access to the high volume of process information	Integrating the high volume of process information
Provide tools and technologies whose output requires employee training	Provide easy-to-understand outputs of the proposed framework for users, which is a benefit of process ontology
A lack of thoroughly considering the relationship between PFs and between processes	Analyze all processes simultaneously and examine PFs correspondingly
Incur high cost and experience a long duration for implementing the PI projects	Reduce the cost, time, and employee involvement and resistance by creating a variety of models to describe the relationships between processes Developing diverse analyses, and simulating several PI scenarios that can be virtually designed and implemented

Incur cost and time in addition to increased employee resistance for obtaining documents and information related to processes in the PI methodologies	Provide and achieve documents with full information and knowledge of the processes by the organization
A lack of consideration for a prescriptive model of PI for the entire organization	Provide the specific obtained process knowledge of implementing DM for the organization and thus reduce the resistance to change for managers
Employ only manual, subjective, and prejudiced computations in current PI methodologies	Considering an automatic extraction of the patterns in the process dataset

With respect to the above-mentioned issues, the objective of this work is to present the main novelty of developing a new integrated framework for combining DM and PI life cycles under an ontology concept. In this framework, the PFs and models extracted by DM are identified as ontologies (see Figure 2). The proposed framework uses the benefits of the process ontology concept for PI. Process ontologies transfer knowledge hidden in the process dataset and are explicit, reusable, and easy to understand for users.

Other studies, (Rao et al., 2012), (Brandt et al., 2008), (Dalmaris et al., 2007), (Papavassiliou et al., 2002), and previous PI methodologies used ontologies. However, they did not use automated methods such as BP exploration for determining diverse patterns. Further, these studies employed a minimal number of PFs in their computations. The proposed framework applies BP exploration for a high volume of process data with a large number of PFs using DM. It constructs process ontology for sharing valuable process patterns extracted through DM. Moreover, process ontology supports PI procedures.

As a noticeable point, we clearly state that the work presented in this paper is not related to the well-known process mining concept. Claes and Poels (2014) stated that process mining discovers, monitors, and improves processes using the extracted knowledge from the event logs. Although several studies have focused on the process mining concept, it is not the intention of this paper to present a replacement for nor is it in contrast to process mining.

The centrality of this paper is in improving BPs based on the extracted patterns of the DM approach hidden in a high volume of process data. The idea of this paper is more related to PI concepts rather than process mining. The proposed framework has several differences compared to process mining. Some of these are presented in Table 6.

Table 6. Distinctions between the proposed framework and the process mining approach

Process mining approach	**The proposed framework**
Lack of usage of literature and concepts of PI methodologies in the process mining procedures	Consider the literature of PI concepts in its computations and promote improved understanding by the organization
Using fewer number of PFs in the process description	Use a high variety of PFs to describe the behavior of BPs
Extract rigid patterns from event logs	Extract valuable patterns from high volume of BPs
Presenting fewer improvement suggestions due to fewer number of PFs	Presenting a high variety of PI suggestions by using a large number of PFs
Focusing more on internal aspects of BPs Little understanding on BPs	Apply more attention to the organizational dimensions of the BPs and the objectives of the processes and business Apply a deeper understanding regarding external aspects of BPs in the organization (as explained in PFs)
A rigid view to BPs and giving lesser importance to viewpoints related to managers and employees	Include additional consideration of the viewpoints of organizational employees in analyzing the behavior of BPs
Low generalizability and flexibility characters of the extracted results from process mining	Effective use of the advantages of process ontology

5. Methodology

In this section, the proposed framework for integrating DM and PI under the organizational ontology (process-typed) concept is demonstrated.

5.1. Relational model for the research elements

Figure 3 portrays the four-phased relational model for the elements of the study including conceptualization, developing the proposed framework, implementing the proposed framework, and comparisons.

As can be observed, Phase 1 identifies the research problems (see Section 1) and contribution statements (see Section 4). The main goal is to solve the problems and present the contribution.

In Phase 2, DM is used for extracting valuable patterns from the high-volume process dataset for improving BPs using process ontologies under the proposed framework (see Figure 4).

In Phase 3, using DM, valuable patterns are extracted as process ontologies. Then, using accuracy measures, the quality of the extracted patterns are evaluated. In this study, the proposed framework employs clustering and classification techniques to construct the patterns.

In clustering, processes are segmented into numerous clusters; the key objective is to segment similar and dissimilar BPs in one and different clusters. A K-means algorithm, an easy and popular technique, is employed to segment the BPs; it can create the patterns with lower complexity and more simple interpretation than other clustering techniques. Other clustering algorithms can be used to segment BPs. The improvement suggestions can be recommended based on the patterns extracted from implementing these algorithms.

In classification, a C5 decision tree algorithm is applied to classify the BPs by extracting a set of if-then rules. This algorithm is user-friendly and can produce simple patterns with easy and interesting interpretations. Other classification algorithms can be employed to classify BPs and based on their extracted patterns, PI suggestions can be recommended. A 10-fold cross-validation technique is employed to divide the process dataset into training and testing datasets.

After extracting the patterns, it is important to evaluate the conformance of the created patterns by considering the PI concepts. In this regard, after the construction of the patterns, they can be promoted through interviewing with the process owners, organizational managers, BP manager, and data miner. It is considerably important to interpret the extracted models correctly. Further, the extracted patterns must have low complexity and easy interpretability.

Then, by considering the extracted patterns, a variety of PI suggestions is recommended. The conformity of the proposed suggestions must be assessed with PI perceptions and real issues related to the organization and its environment by a collaboration between the business and process experts and the data miner.

Finally, the PI suggestions are implemented by considering all issues related to the organizational context. Further, it is verified whether the resulting changes from the implementation of the PI suggestions have conformity with the existing organizational context and the real world.

In Phase 4 of the relational model, the proposed framework is compared with the previous studies of using DM in PI and current PI methodologies (see Tables 4 and 5, respectively).

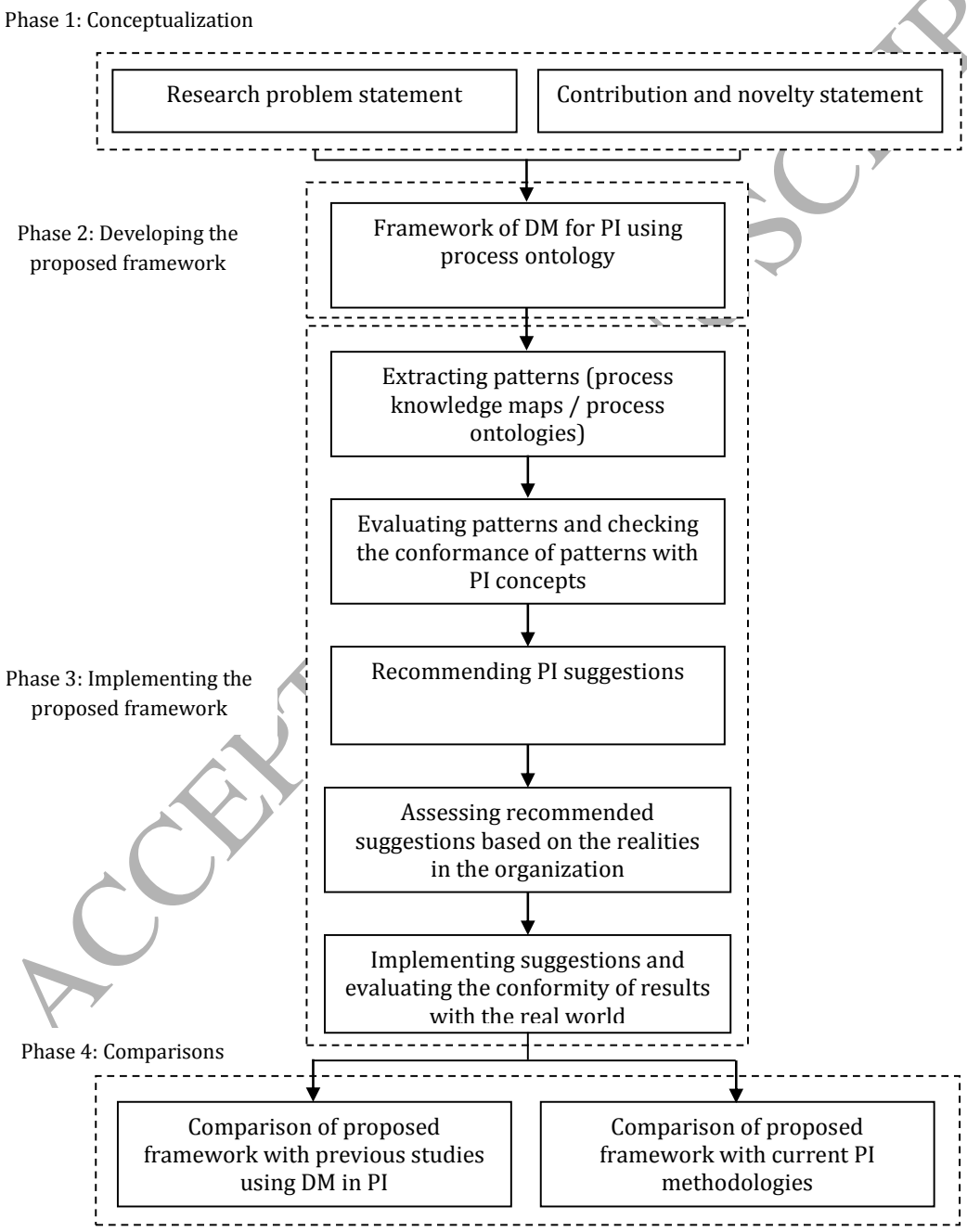

Figure 3. Relational model for the elements of the research

5.2. Proposed integrated framework

Figure 4 displays an integrated framework of DM and PI based on the organizational ontology. As can be observed in Figure 4, the proposed framework is composed of three main parts including organizational ontology (with focus on process ontology), DM, and PI, which were identified in the conceptual model (see Figure 1). Each of these parts includes several sequential phases (indicated by the one-way arrows) and are based on the studies presented in the following. For more understanding, the following studies can be read in detail.

The first part, according to Rao et al. (2012), explains the phases of the organizational ontology as follows: 1. Organizational ontology acceptance; 2. Process identification; 3. Knowledge maps development; 4. Processes modifications; and 5. Ontology updating. Table 7 describes the steps of each phase of the organizational ontology as shown by Rao et al. (2012).

Table 7. Steps of phases related to the organizational ontology

phase	Steps
Organizational ontology acceptance	Accepting/adopting the organizational ontology; understanding the relationships between organizational goals, BPs, resources, and decision makers
Process identification	Identifying and prioritizing BPs for reengineering
Knowledge maps development	Developing knowledge maps (process ontologies)
Processes modifications	Analyzing knowledge maps to identify the inefficiencies of the processes; modifying the processes
Ontology updating	Updating the ontology (process models) for a proper response to the changes

The next part is related to the DM approach, where the phases of the DM are based on the CRISP-DM standard and are as follows (Larose, 2005): 1. Business understanding, 2. Data understanding, 3. Data preparation, 4. Modeling, 5. Evaluation, and 6. Deployment. Table 8 shows the steps of each phase of the CRISP-DM standard (Larose, 2005).

Table 8. Steps of phases related to the CRISP-DM standard

phase	Steps
Business understanding	Defining the requirement of DM project; translating business/ project objectives into DM objectives; presenting a strategy for achieving DM objectives
Data understanding	Collecting data; using exploratory data analysis to familiarize with the data; assessing the quality of data
Data preparation	Preparing final dataset; selecting the cases and variables for analysis; data cleaning and transformation
Modeling	Selecting and implementing suitable modeling techniques; model setting
Evaluation	Evaluating the quality of the models; determining how the model fits with the objectives presented in the "business understanding" phase
Deployment	Using of the created model; generating a report from deployment

The third part of the proposed framework explains a PI procedure based on (Adesola and Baines, 2005) and includes seven phases as follows: 1. understand business requirements, 2. understand process, 3. modeling and analyzing process, 4. redesign process, 5. new process implementation, 6. evaluate new process and methodology, and 7.

review new process. Table 9 explains the steps of each phase of the PI procedure presented by Adesola and Baines (2005).

Table 9. Steps of phases related to the process improvement procedure

phase	Steps
Understand business requirements	Developing and prioritizing strategic objectives and vision; analyzing competitors; developing the organizational model; assessing the existing practices; determining measurable targets; determining the scope of changes; developing the objectives of the BPs and evaluating the readiness; obtaining agreement and initial project resource; benchmarking BPs
Understand process	Identifying BP architecture; defining the BP and its scope; capturing and modeling as-is situation; BP information gathering
Modeling and analyzing process	BP modeling; verifying and validating the model; measuring the performance of the BP
Redesign process	BP benchmarking; identifying the measures for process redesigning; redesign activities; modeling and validating new to-be process model; determining IT requirements; evaluating the performance of redesigned BP
New process implementation	Planning for implementation; approving the implementation; reviewing change management plan; communicating the change; technological development; making the new BP; training staff; roll outing the changes
Evaluate new process and methodology	Conducting BP deployment; reflecting the performance data; revising organizational approach
Review new process	Developing strategic vision of the business; determining BP targets and measures; planning for achieving targets; implementing plan

This procedure is simple and can be better aligned with the DM approach than other procedures in the literature. There exist several PI methodologies. To create a superior alignment, the PI methodology developed by Adesola and Baines (2005) was selected for integrating with the CRISP-DM standard in the proposed framework.

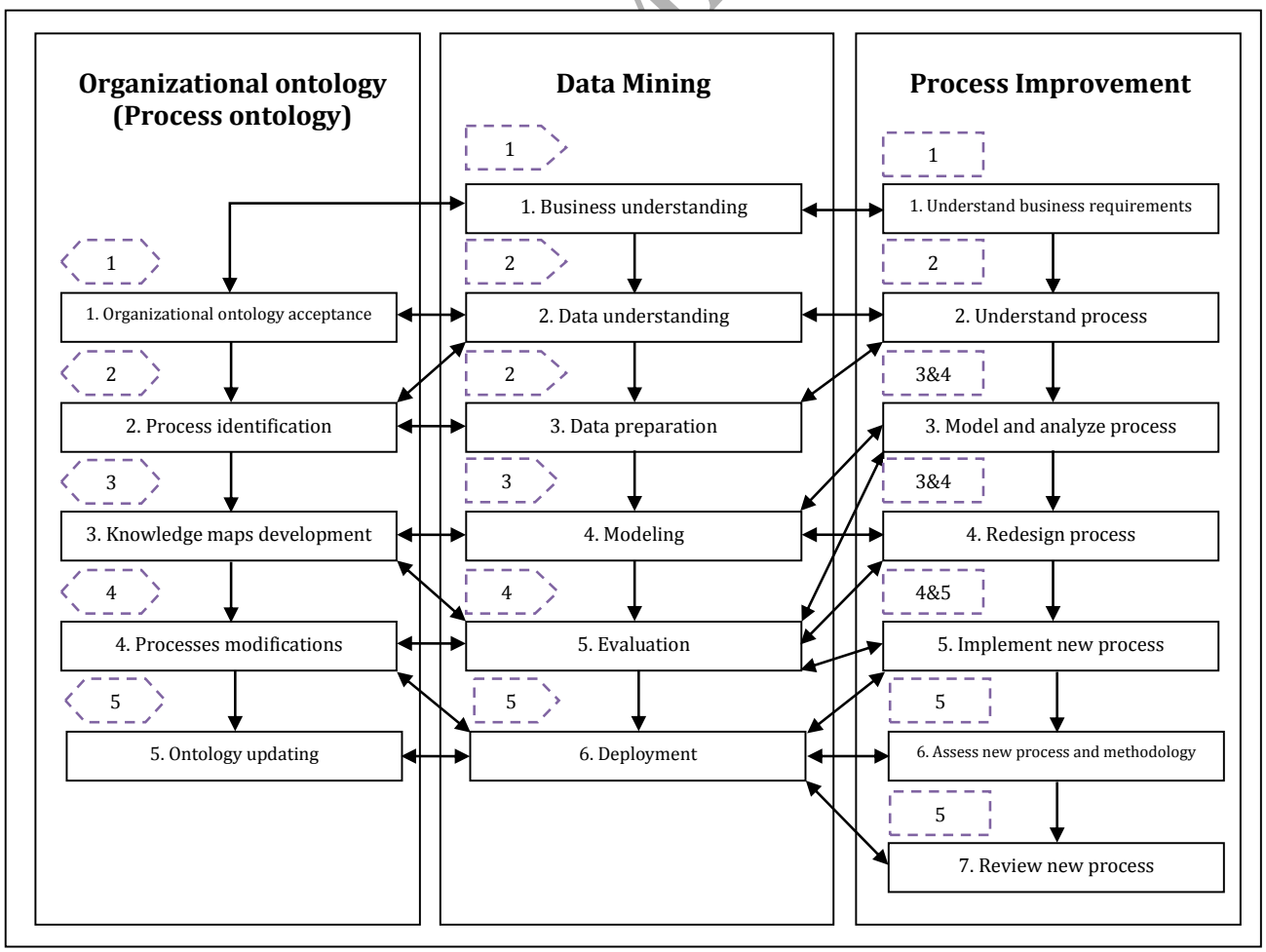

Figure 4. Integrated framework of DM and PI based on organizational ontology

5.2.1. Integration method

To implement the proposed framework, an integration method was clearly essential to achieve the valuable results of PI. The integration method in the proposed framework establishes the relationships between the three parts and their related phases and is based on two schemes as follows: 1. mutual arrows and 2. the numbers inside the dashed figures.

1. First, the mutual arrows explain how the phases of these three parts are related. For example, in the beginning of PI, it is important to understand the business requirements (Phase 1 of PI). For this purpose, some activities regarding the understanding of the business (Phase 1 of DM) and accepting the organizational (process) ontology (Phase 1 of process ontology) must be considered.

As another example, for understanding the processes (Phase 2 of PI), it is important to accept the process ontology and identify all the BPs in the organization (Phases 1 and 2 of process ontology, respectively) and understand and prepare the process data (Phases 2 and 3 of DM). Therefore, it is important to address and understand the mutual relationships between the phases of the parts in implementing the proposed framework.

2. Secondly, the numbers inside of dash present the stages for implementing the proposed framework by simultaneous execution of the related phases in each part. For example, as seen in Figure 4, in the third stage of the proposed framework, processes are modeled, analyzed, and redesigned. In this regard, DM can create and extract some interesting patterns (models) such that they are developed as process knowledge maps (also, process ontology).

5.3. Operational stages of the integrated framework

Table 10 clarifies the operational stages of the proposed framework using an integrated method of the relationships between the phases of the three parts. To implement these stages, the related phases must be executed together, simultaneously. Further, for executing each phase, its relationships with the other phases must be considered with respect to the mutual arrows illustrated in Figure 4.

Table 10. Stages of proposed framework of DM and PI based on organizational ontology

Five stages of the proposed framework	**Three parts of the proposed framework**		
	Part 1: Process ontology	**Part 2: DM**	**Part 3: PI**
Stage 1			
Related phases	Organizational ontology acceptance	Business understanding	Understand business requirements
Stage 2			
Related phases	Process identification	Data understanding Data preparation	Understand process
Stage 3			
Related phases	Knowledge maps development	Modeling	Model and analyze process Redesign process
Stage 4			

Related phases	Processes modifications	Evaluation	Model and analyze process
			Redesign process
			Implement new process
Stage 5			
Related phases	Ontology updating	Deployment	Implement new process
			Assess new process and methodology
			Review new process

The integration method underscores that the simultaneous application of these three parts can enhance the PI procedure using DM. Further, the proposed framework can benefit from the use of the ontology concept in PI.

As indicated in Figure 4 and Table 10, in the first stage of the proposed framework, a business analysis is performed to align both DM and PI perspectives. In this field, the main task is to identify the organizational problems and requirements for PI. For this purpose, the organizational goals are determined, which are associated with the DM and PI objectives.

In this stage, the current situation of the organization is studied. Another core task is creating a general acceptance for building the process ontologies between the main employees in the organization, such as the executive managers, PI expert, and data miner. Finally, in this stage, the proposed framework underlines the alignment between both process and business goals.

In the second stage, a high volume of the processes in the organization are identified and documented using an insightful collaboration between process analyzer, data miner, PI expert, and process owners. A process dataset is designed, in a tabular format. It is containing a large number of PFs placed in the columns of the table and a large number of process names located in the rows of the table. The processes include their values for each feature positioned in each cell in the table. In this table, each process is described by the values assigned to the PFs. In this stage, various process data preparation methods are applied to prepare the processes for analysis.

In the third stage, with the aid of DM techniques including clustering and classification, a diversity of valuable and knowledge-based patterns hidden in the high volume of process data are discovered and developed. They become the process knowledge maps that express and analyze the behavior and the characteristics of the BPs based on the PFs. Further, they can be described as the process ontologies in the proposed framework. With respect to the extracted patterns, processes can be analyzed and redesigned. This stage requires team collaboration including a data miner and PI expert.

In the fourth stage, the extracted knowledge-based maps are evaluated using tools such as follows: interviews with experts including the PI expert and data miner, cross-validation techniques in the DM classification algorithm, and similarity measures between processes in each cluster for the DM clustering algorithm. Moreover, in this stage, the PI suggestions are recommended according to the extracted process knowledge maps (process ontologies); the processes will be modified and enhanced using these suggestions.

In clustering, using cluster profiling (see Table 14), each cluster has processes such that they have similar behavior based on the PFs. Important suggestions are recommended for each process cluster based on the values assigned to the PFs. In classification, improvement

suggestions are recommended based on the target PF with the aid of the ten most selected PFs and the if-then rules extracted from the classification algorithm.

In both clustering and classification techniques, improvement suggestions are recommended in a subjective and judgmental method with cooperation between a PI expert and data miner. This stage applies an iterative method to analyze, model, and modify the processes on one side and employ the mentioned evaluation methods on the other side. Finally, after evaluating the results and process modification and redesign, the initiation of the process implementation is the main task in the fourth stage.

In the fifth stage, new designed and modified processes are implemented and then the performance of the newly implemented processes is evaluated. Moreover, the new process knowledge map (process ontology) can be updated and employed in an incremental and continuous improvement plan. Further, the BP dataset and PFs can be updated for application in a renewed implementation of the proposed framework. Finally, all the previous stages are iterated after utilizing the PI suggestions.

6. Case study

The goal of this section is to describe the applicability of the proposed framework. In other words, this case study illustrates how the proposed framework extracts valuable patterns from a high volume of BP dataset while employing the process ontology concept for PI. In the current study, the efficiency and effectiveness of the proposed framework was evaluated using a real BP dataset containing a large number of BPs along with their interrelated PFs.

The references of the BPs in the organization are as follows: related standards; methods and instruments; mission statement; job and process descriptions; duty description; other related documents of the organization; and information related to interviews with the managers, BP experts, and PI manager.

The PFs were identified using the above-mentioned references and other documents including advanced product quality planning (APQP) classification framework, Porter's value chain, past studies on BPM, BPR, and PI, their tools and standards, and the researches related to aligning BP and KM approaches.

There were 1,318 BPs along with 80 PFs with the following four types: Continuous (c), Nominal (Binary) (b), Nominal (Multiple values) (m), and Ordinal (o). The PFs applied in the case study are presented in Table 11.

Table 11. PFs applied in the case study

PFs applied in case study
The process name (m), documented process (as-is) (b), documented process (to-be) (b), most important supplier (here, process owner) (m), most important input (m), most important output (m), most important customer (m), most important mechanism (m), most important control (m), importance of process technology (o), type of process technology (m), process complexity 1 (simple, complex) (m), process complexity 2 (simple, step-by-step, very complex, knowledge-intensive) (m), process scope (number of engaged departments during the process implementation) (c), contingency to environmental factors (o), addressing processes outside organization (b), influence on other related organizations (o), influence on the business (o), direct relationship of process with projects of the organization (b), considerable impact on other processes (b), considerable influence from other processes (b), supporting knowledge by the process (PI through cooperative, incremental, and continuous methods) (b), capability of knowledge transfer between people (does it include a knowledge object/asset?) (b), imitability of the process (ease of

understanding process) (o), substitutability of the process (o), rarity of the process (o), type of waste (muda) for the process (customer perspective) (m), strategic importance of the process (o), key process (b), competitive excellence (o), capability of outsourcing the process (m), repeatability of the process (c), number of improvements (o), monetary value of the process (o), cost of the process (o), type of the process based on the value (customer perspective) (m), place of implementing the process (m), main employee related to the process (m), customer-oriented process (b), discourse process (b), process formality (o), degree of structuredness (m), degree of automation (m), level of abstraction (m), required inspection and measurement for implementing process (m), type of frequent changes occurred in process (m), ease of process implementation (implementability) (o), risk of process (o), risk of process failure (crisis of the organization in event of process failure) (o), type of process resources (m), resource accessibility (m), process implementation speed (o), time required for process (per working day) (c), requirement for managerial expertise (o), requirement for skill in process implementation (o), requirement for human judgments and experiences in most parts of the process (b), requirement for studying task and process for implementation of the process (o), requirement for education (o), requirement for innovation in the process (o), requirement for attention to quality during the process implementation (o), relationship with KM activities (b), type of process expert (m), requirement for IT in process (o), capability of enabling process by IT (o), requirement for information gathering during process implementation (o), requirement for security in the process (o), type of process security (m), requirement for case management (b), direct relationship of process with objectives and missions of organization (b), requirement for process to comply with organizational objectives and missions (o), purpose of process (m), PI methods (m), atomic process (b), type of process (m), type of process based on Brown (2008) (m), type of process based on Linden et al. (2011) (m), type of process based on Amaravadi and Lee (2005) (m), field of study related to process (m), process predictability (o), uncertainty and ambiguity level in process (o)

It should be noted that for ordinal features, the values from one to five refer to the linguistic words very low, low, moderate, high, and very high, respectively. In nominal features (binary), value of one (zero) reflects that the process includes (excludes) the feature. Appendix 1 demonstrates the values for each nominal PF (multiple values).

Next, the implementation stages of applying DM in PI are presented. In this regard, the clustering and classification DM techniques are applied to construct valuable patterns for a high-volume BP dataset. These patterns are considered as process ontologies to support the recommendation of PI suggestions.

6.1. Implementation stages of applying data mining for process improvement

Figure 5 shows the implementation stages of applying DM in PI based on the proposed framework. These are as follows: 1. Gathering real BP dataset, 2. Preparing and pre-processing BP dataset, 3. Selecting PFs, 4. BP clustering with the K-means algorithm, 5. Creating cluster profiling for BPs, 6. Determining and analyzing the behavior of BPs in each cluster, 7. Identifying the extracted patterns from clustering for PI, 8. BP classification with the C5 decision tree algorithm, 9. Training and testing of the classification models, 10. Comparing classification models and selecting the best, 11. Identifying the extracted patterns from the classification for PI.

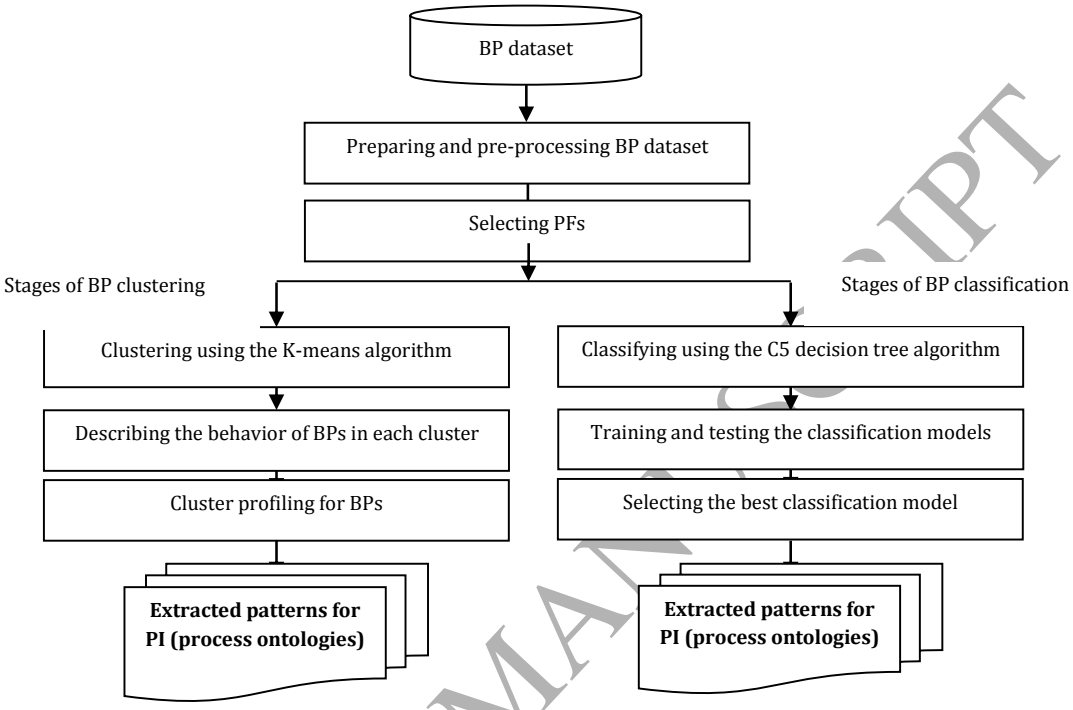

Figure 5. Implementation stages for applying DM in PI

6.2. Clustering

In this section, a K-means clustering algorithm is presented to segment the BPs and describe their behavior in each cluster. Figure 6 indicates the stages of the clustering model to describe BPs and recommend improvement suggestions.

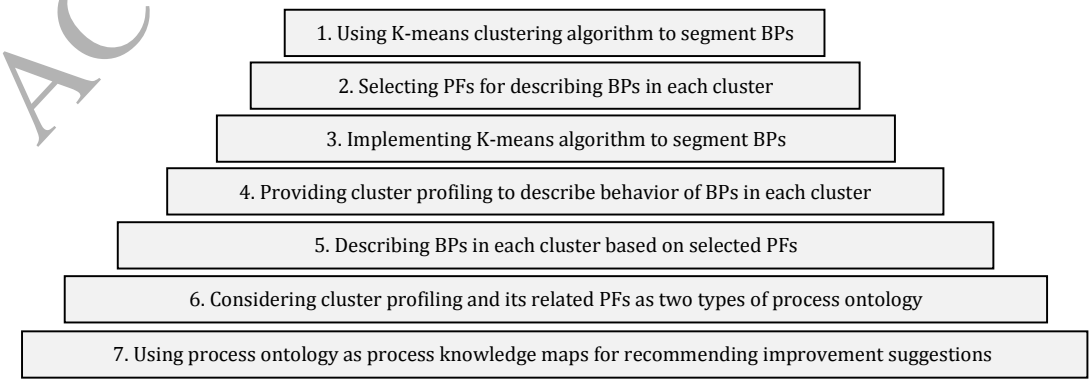

Figure 6. Stages of the clustering model in proposed framework

In the clustering model, at first, a variety of PFs are selected. They are related to the organization and can describe the relationships of BPs with the organization and the environment. These PFs are displayed in Table 12. Then, using the K-means algorithm, BPs are clustered into a pre-defined number of clusters. For this purpose, the distance between BPs is calculated using the Euclidean distance measure. BPs with similar values in the PFs are segmented in one cluster and vice versa.

Table 12. PFs to describe relationships of BPs with organization and the environment

PFs applied in the clustering
Contingency to environmental factors; addressing processes outside organization; influence on other related organizations; influence on the business; direct relationship of process with projects of the organization; competitive excellence; capability of outsourcing the process; risk of process failure (crisis of the organization in event of process failure); direct relationship of process with objectives and missions of organization; requirement for process to comply with organizational objectives and missions; purpose of process

After using the K-means clustering algorithm with five clusters, the numbers of BPs and sample processes in each cluster are determined, as shown in Table 13. Moreover, using the clustering algorithm, Table 14 shows cluster profiling to describe the behavior of BPs in each cluster based on the PFs. In the other words, Table 14 shows the characteristics of BPs such as the sample BPs presented in Table 13.

Table 13. Number of BPs and sample processes in each cluster

Cluster number	Number of BPs	Sample processes
1	360	Determining the information security risk; paying the salaries of employees
2	160	Designing the system; reviewing the conceptual design
3	255	Financial appraisal of the design department; compatibility of the financial strategy with the organization
4	252	Designing the cause and effect diagram of the project; receiving the documents related to knowledge items in each phase of the project
5	291	Issuing the identification card of employees; procuring general items

Table 14. Cluster profiling based on the PFs related to the relationships of the BPs with the organization and environment

	PF					
Cluster number	Contingency to environmental factors	Addressing processes outside organization	Influence on other related organizations	Influence on the business	Direct relationship of process with projects of the organization	Competitive excellence
	Values of each PF for each cluster					
Cluster 1	3 (37.50%)*	0 (69.44%)	1 (48.06%)	3 (41.39%)	0 (73.33%)	1 (43.61%)
Cluster 2	3 and 4 (30.63%)	0 (57.50%)	3 (36.88%)	3 (46.88%)	1 (97.50%)	4 (36.25%)

Cluster 3	4 (38.04%)	0 (65.88%)	3 (35.29%)	4 (41.57%)	1 (53.33%)	2 (53.33%)
Cluster 4	3 (38.49%)	0 (81.35%)	2 (42.86%)	3 (61.90%)	1 (72.22%)	2 (45.24%)
Cluster 5	1 (75.26%)	0 (78.69%)	1 (90.72%)	1 (88.66%)	0 (94.85%)	1 (68.73%)

Table 14. Cluster profiling based on the PFs related to the relationships of the BPs with the organization and environment (continue)

	PF				
Cluster number	Capability of outsourcing the process	Risk of process failure	Direct relationship of process with objectives and missions of organization	Requirement for process to comply with organizational objectives and missions	Purpose of process
	Values of each PF for each cluster				
Cluster 1	1 (69.44%)	3 (43.33%)	0 (92.78%)	3 (38.06%)	S (50.00%)
Cluster 2	1 (72.50%)	4 (46.25%)	0 (51.25%)	1 (26.88%)	D (88.75%)
Cluster 3	1 (85.88%)	4 (52.55%)	1 (81.57%)	4 (41.57%)	O (49.41%)
Cluster 4	1 (76.98%)	3 (44.05%)	0 (82.54%)	3 (36.51%)	O (45.24%)
Cluster 5	1 (47.08%)	1 (55.67%)	0 (100%)	1 (91.75%)	E (49.83%)

Notice: For assistance reading Table 14, please see Section 6 and Appendix 1.

Legend 1: "3 (37.50%)"*: the number outside of the parenthesis is the dominant value of the PF in the cluster. The number inside of the parenthesis is the percent of the processes in the cluster including the dominant value.

Legend 2: For reading the letters in the column "purpose of process", please see legend 3 or Appendix 1.

Legend 3: Product and process support, and improvement (S); Product development (D); Organizational improvement and problem solving (O); Services to employees (E).

6.2.1. Inference mechanism of cluster profiling

As shown in Table 14, there are five clusters in the rows and several PFs in the columns. The inference mechanism of Table 14 is explained as follows. The numbers in Table 14 display the dominant values of the PFs in each cluster, which can be distinguished using the numbers inside the parentheses. For higher values of the numbers inside the parentheses, the similarity measure between the processes in the cluster for the related feature is higher than the other features. This means that this PF can cluster BPs better than other features.

For example, as can be observed in Table 14, in the PF "influence on other related organizations", cluster 5 was set to one for almost all (90.72%) of the processes (as indicated in red). These processes were more similar to each other in cluster 5, compared to the other clusters.

As another example, as displayed in Table 14, in cluster 1, the PF "direct relationship of process with objectives and missions of organization" was set to zero for almost all (92.78%) of the processes (as indicated in blue). These processes are more similar to each other than to the processes in the other clusters based on the mentioned feature.

With this cross-sectional analysis, the best PFs for each cluster were determined to describe the BPs in clusters. Further, for each PF, the similarity measure between the processes was compared for each cluster.

6.2.2. Describing business processes using cluster profiling

Table 14 shows cluster profiling, which describes the behavior of the processes in the clusters based on their relationships with the organization and environment. That is, the

PFs displayed in Table 14 can explain the characteristics of BPs based on their relationships with the organization and the environment (also, see Table 12).

As demonstrated in Table 14, in clusters 2 and 3, the BPs were primarily influenced by environmental factors. A large number of BPs in cluster 2 had more interactions with the BPs outside the organization than the other clusters. The level of influence on the other related organizations was set to moderate and low for the majority of the BPs. Some of the BPs in cluster 3 were influential on the business.

The majority of the BPs in clusters 2 and 4 had a direct relationship with the projects of the organization. The highest level of competitive excellence was for the BPs in cluster 2, which were more able to achieve the competitive excellence. A significant majority of the BPs in clusters 1, 2, 3, and 4 had no outsourcing capability and based on various reasons, must be implemented in the organization. The failure risk level for the majority of the BPs in clusters 2 and 3 was high and a crisis could be created in the organization by the failure of these BPs.

A considerable number of the BPs in cluster 3 had a direct relationship with the missions and objectives of the organization. The BPs in cluster 3 were required to comply, to a greater extent, with the organizational missions and objectives.

It is worth mentioning that the main purpose of the majority of the BPs in cluster 2 was development. Whereas the target of the majority of the BPs in clusters 3 and 4 was organizational improvement and problem solving. Further, the most considerable purpose of the BPs in cluster 5 was providing services to employees. Finally, half of the BPs in the first cluster involved the purpose of supporting and improving products and BPs.

6.2.3. Providing improvement suggestions

In clustering, there are two types of process ontology that can be considered as a process knowledge map. Using these process ontologies, suggestions are recommended for PI. These process ontologies are depicted in Figure 7.

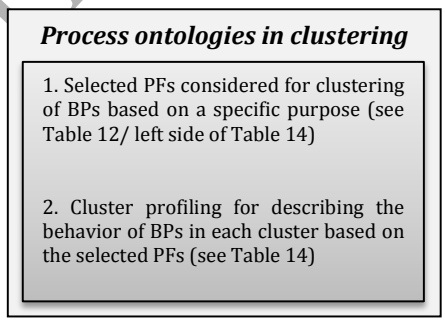

Figure 7. Two types of process ontology in clustering

After describing the BPs by cluster profiling, some PI suggestions can be recommended based on the cluster profiling presented in Table 14. These PI suggestions are based on the PFs that describe the relationships of the BPs with the organization and the environment, which are listed in Table 15.

These PI suggestions are inferred based on the subject of PF for the process cluster that has more distinguished BPs than other clusters with the aid of collaboration between a PI expert and data miner. For example, as seen in Table 14, BPs in clusters 2 and 3 have more risk of failure than BPs in the other clusters. Then, a PI suggestion is recommended for the BPs in clusters 2 and 3, as indicated in Table 15 (see Row 5). It can be observed that PI suggestions can be related to more than one PF.

Table 15. PI suggestions based on results of clustering

Row	PFs	PI suggestion related to PFs
1	Contingency to environmental factors; addressing processes outside organization; influence on other related organizations; influence on the business	Using a suitable environmental scanning system for the BPs in clusters 2 and 3
2	Direct relationship of process with projects of the organization; direct relationship of process with objectives and missions of organization	Aligning the purpose of the processes with the strategies of the organization and the system of the projects for the processes in clusters 2, 3, and 4
3	Competitive excellence	Designing business excellence plans based on the processes in cluster 2
4	Capability of outsourcing the process	Identifying and determining the processes that can be outsourced in all clusters (particularly cluster 5)
5	Risk of process failure (crisis of the organization in event of process failure)	Developing risk management systems and employing failure mode and effect analysis (FMEA) for the majority of the processes in clusters 2 and 3
6	Purpose of process	Designing organizational improvement and problem-solving plans for the majority of the processes in clusters 3 and 4
7	Purpose of process	Applying the principles and systems of innovation and technology management to design a technology and product roadmap for the processes in cluster 2 in the direction of product development
8	Purpose of process	Providing specific plans to increase the satisfaction level of employees and assess the satisfaction related to the processes in cluster 5
9	Purpose of process	Planning for PI and designing instruments for the processes in cluster 1

6.3. Classification

In this section, the C5 decision tree algorithm is applied to classify BPs based on a target PF. After implementing the classification algorithm, the ten most important PFs are selected such that they describe the target PF. Note that different target PFs can be employed to classify BPs and therefore, diverse PI suggestions can be recommended.

In this paper, the target feature "requirement for innovation in the process" was applied as an example to classify the BPs and verify the feasibility of the proposed framework in this section. Figure 8 displays the stages of the classification model to construct a decision tree for recommending improvement suggestions.

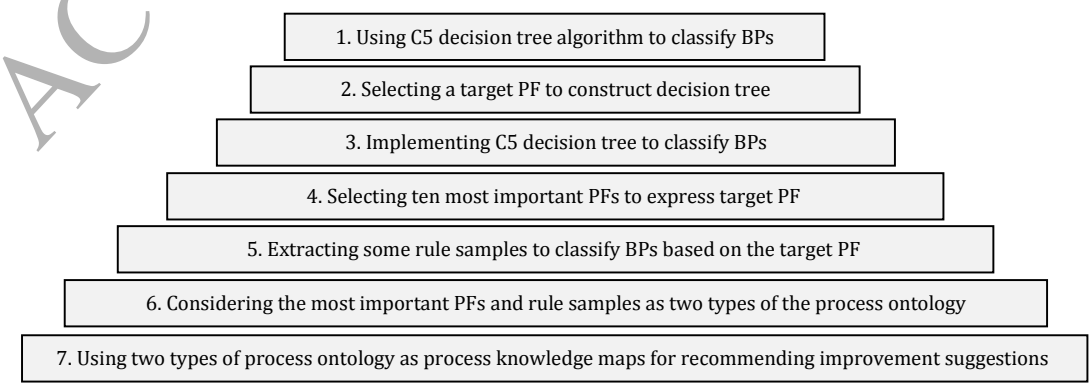

Figure 8. Stages of the classification model in the proposed framework

In the classification model, first, a target PF is selected. Then, using the C5 decision tree algorithm, BPs are classified. For this purpose, a 10-fold cross-validation technique is used to divide the BP dataset into training and testing BP datasets. The C5 decision tree algorithm is implemented by the recursive partitioning method until higher classification accuracy is achieved.

Table 16 presents the results of the C5 decision tree for classifying the BPs. This includes the ten most important selected features to express the target feature "requirement for innovation in the process", two rule samples, and two related BP samples.

Table 16. Results of applying the C5 decision tree classification algorithm

Target PF: Requirement for innovation in the process **Classification accuracy:** 91.54%	
Most important selected PFs	Rarity of the process (more considerable important than other selected features); PI methods; requirement for education; ease of process implementation; supporting knowledge by the process; uncertainty and ambiguity level in process; type of frequent changes occurred in process; addressing processes outside organization; most important control; time required for process
Rule sample 1	If the rarity of the process (3, 4, and 5), number of improvements (3), requirement for studying task and process for the implementation of the process (4 and 5), requirement for education (4 and 5), ease of process implementation (1), time required for process (more than 12), requirement for innovation in the process (5).
BP sample 1	Technical design; sub-system integration; engineering analysis
Rule sample 2	If the rarity of the process (1 and 2), requirement for human judgments and experiences in most parts of the process (1), uncertainty and ambiguity level in the process (1), requirement for innovation in the process (1).
BP sample 2	User education; selecting ethical employees

Notice: For assistance reading Table 16, please see Section 6 and Appendix 1.

6.3.1. Providing improvement suggestions

As can be observed in Figure 9, the most important selected features and the two rule samples are actually the two types of process knowledge map and process ontology. They can be applied for recommending PI suggestions with support and cooperation between the PI expert and data miner.

Process ontologies in classification

1. The most important selected PFs extracted from implementing C5 that can better explain target PF to classify BPs (see Table 16)

2. A collection of if-then rules extracted from the implementation of C5 that classify BPs based on target PFs (see Table 16)

Figure 9. Two types of process ontology in classification

That is, PI suggestions are concluded based on the most important selected PFs that are effective for determining the amount of requirement for innovation in the processes. Further, using if-then rules, BPs can be classified based on their degree of required innovation.

For example, rule sample 1 (see Table 16) declares the requirement for education of extremely high innovative processes (the value is set to 5) is equal to high and very high (the values are set to 4 and 5, respectively). Then, the organization must consider the three recommended samples of suggestions for improving the situation of "innovation measure" in the processes.

They are as follows and as presented in Table 17: (1) employing suitable educational plans for the experts in innovative processes (see Row 5), (2) applying improvement methods including knowledge-based methods and enhancing teamwork and creative education for less-innovative processes (see Row 9), and (3) decreasing the uncertainty of innovative processes by educating experts (see Row 11).

Table 17 presents the PI suggestions related to the most important selected features and two rule samples. It is clear that some PI suggestions are related to more than one PF.

Table 17. Sample PI suggestions based on the results of classification

Row	PFs	PI suggestion related to the PFs and two rule samples
1	the need for innovation in the process	Identifying and classifying the processes that require innovation for their activities more than others
2	the rarity of the process	Planning for innovation management for the rare processes with a significant requirement for innovation
3	PI methods; type of frequent changes occurred in the process; the most important control; supporting the knowledge by the process	Using knowledge-based, incremental, cooperative, and continuous improvement methods for the processes that require a high level of innovation
4	supporting the knowledge by the process; rarity of the process	Developing a specific KM strategy to advance the knowledge embedded in rare and innovative processes
5	the need for education; easiness of the process implementation	1. Using the competency model for employing valuable experts for innovative processes 2. Employing suitable educational plans for the experts in innovative processes
6	the rarity of the process; the need for studying the task and process for the implementation of the process; the need for education; the time required for the process	Studying the activities and tasks related to the rare processes, especially for developing the activities and experts of the rare processes
7	dealing with the processes outside the organization; the rarity of the process	Interacting with technological and innovative companies / universities / research centers to promote innovation in rare processes
8	the rarity of the process; easiness of the process implementation; uncertainty and ambiguity level in the process; the time required for the process	Facilitating the difficulties and constraints related to rare processes to increase the feasibility and ease of implementation
9	PI methods; the need for education; number of improvements	Applying improvement methods including knowledge-based methods and enhancing teamwork and creative education for less-innovative processes
10	easiness of the process implementation; studying the task and process for the implementation of the process; number of improvements	Studying the job and performing evaluations to enrich the tasks related to less-innovative processes
11	uncertainty and ambiguity level in the	Decreasing the uncertainty of innovative processes by educating experts

	process; the need for education	
12	the need for human judgments and experiences in most parts of the process; number of improvements	Developing creativity and innovation circles

7. Discussion and conclusions

In organizations, there are typically many BPs with specific features, leading to an increase in the dimensionality, complexity, uncertainty, time, cost, resistance of employees, and misunderstanding of the processes. In this situation, DM techniques can support PI procedures by extracting valuable patterns hidden in the high volume of BPs for the purpose of recommending improvements.

The contribution of this research work is in four main areas as follows:

- First, this paper presents a broad variety of PFs in order to identify the behavior of BPs. These PFs were provided based on the vast literature related to the concepts of BPM and KM. In addition, a large real dataset including the information related to these PFs for the entire BPs in the organization was prepared. This large BP dataset along with the wide variety of PFs were considered as an input to the DM techniques in the framework developed in the current study.
- Second, this paper developed a three-part, five-stage framework implementing DM techniques and a process ontology concept for PI. An actual high-volume BP dataset was employed to evaluate the applicability of the proposed framework. The proposed framework integrated three life cycles (including organizational ontology (process ontology), DM, and PI) to identify the behavior of processes. This framework can simultaneously benefit from these life cycles with a unified approach. It consists of five stages, where, in each stage, the activities of these life cycles are implemented. Further, there are mutual relationships between the activities of these life cycles.

The proposed framework automatically extracts valuable patterns and recommends PI suggestions using clustering and classification DM techniques. The proposed framework employs the process ontology concept to achieve the benefits of using ontologies in PI. The process ontology in the proposed framework is of two types as follows:

1. The PFs that explain the characteristics of BPs in the organization.
2. The valuable patterns extracted using the clustering and classification DM algorithms. These patterns can portray the process knowledge maps to describe the behavior of processes and provide PI suggestions (see Tables 14 (clustering) and 16 (classification)).

- Third, clustering and classification DM techniques were employed for the PFs in order to find valuable patterns hidden in the large number of BPs.

In clustering, the K-means algorithm segmented the BPs into five clusters based on specified PFs. These PFs describe the relationships between BPs on one side and the organization and the environment on the other side. Cluster profiling extracted by implementing K-means and specified PFs are two types of process ontologies discussed in Section 6.2 of the proposed framework (see Table 14).

In classification, a C5 decision tree algorithm classified the BPs based on the target PF “requirement for innovation in the process”. The most important selected PFs and the

collection of if-then rules extracted by implementing the C5 algorithm are two types of process ontologies discussed in Section 6.3 of the proposed framework (see Table 16).

- Fourth, the present study was inspired by the four following studies as explained in Section 2. The main subjects in these studies included: (1) using the CRISP-DM standard for operational BPs (Rupnik and Jaklic, 2009); (2) explaining the roles of business users, IT, and DM experts in the integration of DM stages and BPs (Wegener and Rüping, 2010); (3) presenting a literature review of the DM application in the BPR methodology for developing an integrated framework for the simultaneous implementation of the two approaches; and developing a combinational model of DM, KM, and process monitoring architecture (Ghanadbashi et al., 2013); and (4) proposing a framework for using DM in BPs with ontology under process mining concepts (Pivk et al., 2014).

Previous PI methodologies had certain problems as elaborated in Section 1. Although they employed DM for BPs, they presented only a theoretical and conceptual framework for the BPs. They did not include a large number of BPs with many PFs in their computations. Moreover, whereas several studies on BPs adopted the concept of ontology, none applied process ontology to present DM patterns for PI.

The proposed framework attempts to overcome the weaknesses of past studies. These weaknesses are relative and may vary considerably in severity from one study to another. However, they are related to the previous studies evaluated.

Five types of studies were considered in the current work to evaluate their relative weaknesses in comparison with the proposed framework. A method for comparing these weaknesses is presented in this paper. The weaknesses were overcome by the proposed framework, as explained in the respective sections. The five types of studies considered were as follows:

1. Four director researches as the foundations of the proposed framework (See Table 2).
2. Previous studies regarding using DM approach for PI (See Table 4).
3. Current PI methodologies (See Table 5).
4. Studies related to process mining approach (See Table 6).
5. Studies (Rao et al., 2012), (Brandt et al., 2008), (Dalmaris et al., 2007), and (Papavassiliou et al., 2002) that applied ontology concept (See Section 4).

Table 18 explains how the proposed framework can overcome the existing weaknesses of the five types of the previous related studies.

Table 18. Characteristics of proposed framework to overcome weaknesses of previous related studies

Row	Type of previous studies	Characteristics of proposed framework to overcome weaknesses
1	Four director researches as the foundations of the proposed framework	Considering details and applicable, operational, and technical dimensions of using DM in recommending PI suggestions by application and extensive use of process ontology with an actual, high volume of BPs with a large variety of PFs
2	Previous studies regarding using DM approach for PI	Using a real BP dataset with an extensive variety of PFs for DM approach to extract practical results for recommending PI suggestions
3	Current PI methodologies	Applying a high volume of information regarding the processes for automatic extraction of easy-to-use process ontologies to analyze the processes behavior for recommending a variety of PI scenarios with improved performance in cost, time, and other indices
4	Studies related to process mining approach	Using the literature of PI issues to determine a variety of PFs for describing a large BP dataset

5	Studies (Rao et al., 2012), (Brandt et al., 2008), (Dalmaris et al., 2007), and (Papavassiliou et al., 2002) that applied ontology concept	Applying an automatic method to discover process knowledge maps as process ontologies for appropriate sharing PI concepts using an extensive variety of PFs and high volume of process data

The proposed framework can support PI methodologies by presenting valuable process ontologies for sharing an effective process understanding between employees. Further, the extracted process ontologies can be re-used in all parts of the organization for sharing the process knowledge. Process ontologies can also be updated automatically and rapidly, in accordance with the changes accrued in the organizational processes. Moreover, process ontologies can clarify the hidden and embedded knowledge in processes and their relationships with each other for PI.

In future research, the proposed framework can be developed for knowledge-intensive BPs. Furthermore, a KM approach can be combined with the proposed framework to integrate the PI and KM methodologies using DM techniques and the process ontology concept. Finally, other clustering and classification algorithms can be applied to extract valuable patterns for recommending PI suggestions.

References

Adesola, S., Baines, T., 2005. Developing and evaluating a methodology for business process improvement. Bus. Process Manag. J. 11, 37–46. doi:10.1108/14637150510578719.

Amaravadi, C.S., Lee, I., 2005. The dimensions of process knowledge. Knowl. Process Manag. 12, 65–76. doi:10.1002/kpm.218.

Borrego, D., Barba, I., 2014. Conformance checking and diagnosis for declarative business process models in data-aware scenarios. Expert Syst. Appl. 41, 5340–5352. doi:10.1016/j.eswa.2014.03.010.

Brandt, S.C., Morbach, J., Miatidis, M., Theißen, M., Jarke, M., Marquardt, W., 2008. An ontology-based approach to knowledge management in design processes. Comput. Chem. Eng. 32, 320–342. doi:10.1016/j.compchemeng.2007.04.013.

Brown, S., 2008. Business Processes and Business Functions: a new way of looking at employment. Mon. labor Rev. 131, 51–70.

Chen, F.Z., Wang, X.Z., 1999. An integrated data mining system and its application to process operational data analysis. Comput. Chem. Eng. 23, S787–S790. doi:10.1016/S0098-1354(99)80193-8.

Claes, J., Poels, G., 2014. Merging event logs for process mining: A rule based merging method and rule suggestion algorithm. Expert Syst. Appl. 41, 7291–7306. doi:10.1016/j.eswa.2014.06.012.

Dalmaris, P., Tsui, E., Hall, B., Smith, B., 2007. A framework for the improvement of knowledge-intensive business processes. Bus. Process Manag. J. 13, 279–305. doi:10.1108/14637150710740509.

Damij, N., Damij, T., 2014. Business Process Approaches, in: Process Management: A Multi-Disciplinary Guide to Theory, Modeling, and Methodology. Springer Berlin Heidelberg, pp. 45–60. doi:10.1007/978-3-642-36639-0_4.

Darmani, A., Hanafizadeh, P., 2013. Business process portfolio selection in re-engineering projects. Bus. Process Manag. J. 19, 892–916. doi:10.1108/BPMJ-08-2011-0052.

Delgado, A., Weber, B., Ruiz, F., Garcia-Rodríguez de Guzmán, I., Piattini, M., 2014. An integrated approach based on execution measures for the continuous improvement of business processes realized by services. Inf. Softw. Technol. 56, 134–162. doi:10.1016/j.infsof.2013.08.003.

D'heygere, T., Goethals, P.L.M., De Pauw, N., 2003. Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. Ecol. Modell. 160, 291–300. doi:10.1016/S0304-3800(02)00260-0.

Folorunso, O., Ogunde, A.O., 2005. Data mining as a technique for knowledge management in business process redesign. Inf. Manag. Comput. Secur. 13, 274–280. doi:10.1108/09685220510614407.

Gangemi, A., Borgo, S., Catenacci, C., Lehman, J., 2004. Task taxonomies for knowledge content. Trento.

Ghanadbashi, S., Khanbabaei, M., Abadeh, M.S., 2013. Applying data mining techniques to business process reengineering based on simultaneous use of two novel proposed approaches. Int. J. Bus. Process Integr. Manag. 6, 247–267. doi:10.1504/IJBPIM.2013.056963.

Ghattas, J., Soffer, P., Peleg, M., 2014. Improving business process decision making based on past experience. Decis. Support Syst. 59, 93–107. doi:10.1016/j.dss.2013.10.009.

Gómez-Pérez, J.M., Erdmann, M., Greaves, M., Corcho, O., Benjamins, R., 2010. A framework and computer system for knowledge-level acquisition, representation, and reasoning with process knowledge. Int. J. Hum. Comput. Stud. 68, 641–668. doi:10.1016/j.ijhcs.2010.05.004.

Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., Shan, M.-C., 2004. Business Process Intelligence. Comput. Ind. 53, 321–343. doi:10.1016/j.compind.2003.10.007.

Groger, C., Schwarz, H., Mitschang, B., 2014. Business Information Systems, Lecture Notes in Business Information Processing, Lecture Notes in Business Information Processing. Springer International Publishing, Cham. doi:10.1007/978-3-319-06695-0.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing? Int. J. Hum. Comput. Stud. 43, 907–928. doi:10.1006/ijhc.1995.1081.

Houy, C., Fettke, P., Loos, P., van der Aalst, W.M.P., Krogstie, J., 2011. Business Process Management in the Large. Bus. Inf. Syst. Eng. 3, 385–388. doi:10.1007/s12599-011-0181-5.

Huang, Z., Lu, X., Duan, H., 2012. Resource behavior measure and application in business process management. Expert Syst. Appl. 39, 6458–6468. doi:10.1016/j.eswa.2011.12.061.

Jeong, H., Song, S., Shin, S., Rae Cho, B., 2008. Integrating data mining to a process design using the robust bayesian approach. Int. J. Reliab. Qual. Saf. Eng. 15, 441–464. doi:10.1142/S0218539308003155.

Kharbat, F., El-Ghalayini, H., 2008. Building Ontology from knowledge base systems, in: Data Mining in Medical and Biological Research. I-Tech Education and Publishing, Vienna, pp. 55–68.

Koh, H.C., Low, C.K., 2004. Going concern prediction using data mining techniques. Manag. Audit. J. 19, 462–476. doi:10.1108/02686900410524436.

Larose, D.T., 2005. Discovering Knowledge in Data, an Introduction to Data Mining, 1st ed. John Wiley & Sons, New Jersey. doi:10.1002/0471687545.

Lee, S.J., Siau, K., 2001. A review of data mining techniques. Ind. Manag. Data Syst. 101, 41–46. doi:10.1108/02635570110365989.

Lepmets, M., McBride, T., Ras, E., 2012. Goal alignment in process improvement. J. Syst. Softw. 85, 1440–1452. doi:10.1016/j.jss.2012.01.038.

Linden, M., Felden, C., Chamoni, P., 2011. Dimensions of Business Process Intelligence, in: Lecture Notes in Business Information Processing. pp. 208–213. doi:10.1007/978-3-642-20511-8_19.

Marjanovic, O., 2012. The Importance of Process Thinking in Business Intelligence, in: Organizational Applications of Business Intelligence Management. IGI Global, pp. 76–94. doi:10.4018/978-1-4666-0279-3.ch006.

Mathew, S., George, S., 2012. Implementation of data mining technique for BPR, in: International Conference on Electrical Engineering and Computer Science. Trivandrum, pp. 262–266.

Muthu, S., Whitman, L., Cheraghi, S.H., 2006. Business process reengineering: a consolidated methodology, in: 4 Th Annual International Conference on Industrial Engineering Theory, Applications, and Practice, 1999 U.S. Department of the Interior - Enterprise Architecture. San Antonio, Texas, pp. 8–13.

Papavassiliou, G., Ntioudis, S., Mentzas, G., Abecker, A., 2002. Business process knowledge modelling: method and tool, in: Proceedings. 13th International Workshop on Database and Expert Systems Applications. IEEE Comput. Soc, pp. 138–142. doi:10.1109/DEXA.2002.1045889.

Pivk, A., Vasilecas, O., Kalibatiene, D., Rupnik, R., 2014. Ontology and SOA Based Data Mining to Business Process Optimization, in: Information System Development. Springer International Publishing, Cham, pp. 255–268. doi:10.1007/978-3-319-07215-9_21.

Rao, L., Mansingh, G., Osei-Bryson, K.-M., 2012. Building ontology based knowledge maps to assist business process re-engineering. Decis. Support Syst. 52, 577–589. doi:10.1016/j.dss.2011.10.014.

Rebuge, Á., Ferreira, D.R., 2012. Business process analysis in healthcare environments: A methodology based on process mining. Inf. Syst. 37, 99–116. doi:10.1016/j.is.2011.01.003.

Rupnik, R., Jaklic, J., 2009. The deployment of data mining into operational business processes, in: Data Mining and Knowledge Discovery in Real Life Applications. I-Tech Education and Publishing, Vienna, pp. 373–388.

Singh, R., Gernaey, K. V, Gani, R., 2010. An ontological knowledge-based system for the selection of process monitoring and analysis tools. Comput. Chem. Eng. 34, 1137–1154. doi:10.1016/j.compchemeng.2010.04.011.

Sohail, A., Dhanapal Durai Dominic, P., 2012. A gap between Business Process Intelligence and redesign process, in: 2012 International Conference on Computer & Information Science (ICCIS). IEEE, pp. 136–142. doi:10.1109/ICCISci.2012.6297227.

Tonchia, S., 2004. Fundamentals of Process Management and Business Process Reengineering, in: Process Management for the Extended Enterprise. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11–27. doi:10.1007/978-3-642-17051-5_2.

Vukšić, V.B., Bach, M.P., Popovič, A., 2013. Supporting performance management with business process management and business intelligence: A case analysis of integration and orchestration. Int. J. Inf. Manage. 33, 613–619. doi:10.1016/j.ijinfomgt.2013.03.008.

Wegener, D., Rüping, S., 2010. On Integrating Data Mining into Business Processes. pp. 183–194. doi:10.1007/978-3-642-12814-1_16.

Zhonghua, D., Limei, S., 2008. Business processes reengineering based on data mining, in: International Conference on Logistics Engineering and Supply Chain. China, pp. 164–169.

Appendix 1. Values for nominal PF (binary and multiple values)

Feature names	Values for features
The process name	Process name
Most important supplier Note: values are departments of the organization.	Commerce 1, designer/ researcher 2, education 3, employee 4, finance 5, future study 6, human resources 7, IT 8, inspection 9, KM 10, manager director 11, planning and control 12, project manager 13, quality 14, support and logistics 15, test 16, top manager 17
Most important input	Documents 1, human resources 2, information 3, money 4, notifications and superiors 5, request 6, system 7, work piece 8
Most important output	Decision 1, documents 2, knowledge 3, money 4, service 5, system 6, work piece 7
Most important customer Note: values are departments of the organization.	Commerce 1, design office 2, education 3, employee 4, finance 5, future study 6, human resources 7, IT 8, KM 9, manager director 10, planning and control 11, project manager 12, quality management 13, support and provisions 14, test 15, top manager 16
Most important mechanism	IT 1, human resource (expert) 2, management 3, money 4, technology 5, no important mechanism 6
Most important control	Knowledge 1, management 2, rules and methods 3
Type of process technology	Uncomplicated 0, complex 1
Process complexity 1	Simple 0, complex 1
Process complexity 2	Simple procedural 1, step-by-step procedure 2, very complex 3, knowledge intensive 4
Type of waste (muda) for the process (customer perspective)	Waste type 1 (1), waste type 2 (2), non-waste (3)
Capability of outsourcing the process	Incapable 1, parts of the process are capable 2, all part of the process are capable 3
Type of the process based on the value (customer perspective)	Non-value added 0, necessary to add value 1, value-added 2
Place of implementing the process	CEO (chief executive officer: management team) 1, human resources management 2, IT 3, commerce 4, design office 5, education 6, finance 7, future study 8, inspection 9, KM 10, planning and control 11, quality management 12, support and logistics 13, test 14
Main employee related to the process	Designer 1, employee 2, engineer 3, general employee 4, manager 5, project manager 6, researcher 7, top manager 8
Degree of structuredness	Low 0, high 1
Degree of automation	Manual 1, semi-automated 2, automated 3
Level of abstraction	Very detailed 0, only high level 1
Required inspection and measurement for implementing the process	Low 0, high 1
Type of frequent changes occurred in process	Low change 1, incremental improvement 2, fundamental changes 3
Type of process resources	Data 1, hardware 2, human resources 3, information 4, machines 5, money 6, software 7, system 8
Resource accessibility	Direct 1, indirect 0
Type of process expert	Expert 1, knowledge type (knowledge worker) 2, usual

	(ordinary worker) 3
Type of process security	Secret 1, top-secret 2, unclassified 3
Purpose of process	Services to employees (E), product development (D), product and process support and improvement (S), organizational improvement and problem solving (O), finance (F)
PI methods	Traditional 0, knowledge-based 1
Type of process	Design-based 1, managerial 2, research (operation) 3, research (support) 4, staff-related 5, technical 6
Type of process based on Brown (2008)	Managerial 1, operational 2, supportive 3
Type of process based on Linden et al. (2011)	Technical 0, business 1
Type of process based on Amaravadi and Lee (2005)	Behavioral process 1, change process 2, managerial process 3, work process 4
Field of study related to process	Business and management 1, commerce management 2, computer engineering 3, engineering 4, financial management 5, human resources management 6, industrial engineering 7, IT management 8, technology management 9, no academic field 10

Mohammad Khanbabaei is currently PhD in Information Technology Management from Islamic Azad University, Science and Research Branch, Tehran, Iran. His research focuses in data mining, business intelligence, artificial intelligence, neural networks, fuzzy inference systems, optimization techniques, business process management, systems analysis, decision making, knowledge management, and technology management.

Farzad Movahedi Sobhani received his PhD in Industrial Engineering from Department of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran. He is currently a faculty member at the Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran. His interests include business process management, knowledge management, and data mining.

Mahmood Alborzi received his PhD in Neural Networks from Department of Computer Science & Information Systems, Brunel University, Uxbridge, London, England. He is currently a faculty member at the Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran. His research has focused on the application of neural networks for real-time log interpretation in oil well drilling. His interests include neural networks, genetic algorithm, artificial intelligence and supply chain management.

Reza Radfar received his PhD in Technology Management from Department of Technology Management, Islamic Azad University, Science and Research Branch, Tehran, Iran. He is currently a faculty member at the Department of Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran. His interests include technology management, knowledge management, data mining, and systems modeling.