# Potentials of Image Mining for Business Process Management

**Rainer Schmidt, Michael Möhring, Alfred Zimmermann,
Ralf-Christian Härting and Barbara Keller**

**Abstract** An enormous amount of data in the context of business processes is stored as images. They contain valuable information for business process management. Up to now this data had to be integrated manually into the business process. By advances of capturing it is possible to extract information from an increasing number of images. Therefore, we systematically investigate the potentials of Image Mining for business process management by a literature research and an in-depth analysis of the business process lifecycle. As a first step to evaluate our research, we developed a prototype for recovering process model information from drawings using Rapidminer.

## 1 Introduction

Business Process Management is an important means for adapting enterprises to changing requirements [1]. Most information used for business process management is either represented as formal process models using approaches such as ARIS [2] or BPMN [3] or laid down in written documentation. Although this approach is conceptually sound, the use of formal methods excludes many stakeholders at least partially [4], because they are not familiar with these tools. In general, the need to

R. Schmidt (✉) · M. Möhring
Munich University of Applied Sciences, Munich, Germany
e-mail: Rainer.Schmidt@hm.edu

R.-C. Härting · B. Keller
Aalen University of Applied Sciences, Aalen, Germany

A. Zimmermann
Reutlingen University, Reutlingen, Germany
e-mail: alfred.zimmermann@reutlingen-university.de

429

integrate more information from heterogeneous information sources has been identified as a precondition for improving process quality [5].

At the same, the amount of digitized information has increased in many enterprises significantly [6]. Not only by the advance of smartphones, but also through scanners, cameras etc. images created in huge numbers. Capturing and storing images has become an everyday activity in many enterprises [7]. Images are taken to capture handwritten notes, whiteboards, comments on printed documentation etc. Apps such as OfficeLens [8] allow to capture images very easily and use them within office software. Software for extracting textual information from images has become a standard. Detection of geometric figures in hand-written drawings is available [9], the detection of semantically deep drawings is in close reach. Often information relevant to business process management is contained in the images.

However, up to now this affluence of images is not used as an input for business process management. Thus, the potentials using these images for business process management are not exploited. Therefore, this paper address the following research question: *What are the potentials of Image Mining for Business Process Management?* To explore this research question, we made a systematic literature review and described the basics of Image Mining and Business Process Management. Furthermore, a prototype was designed to show the potentials of one scenario of Image Mining for Business Process Management. The paper is structured as follows: after this introduction, we define basics of Image Mining as well as Business Process Management. In Sect. 3, we define potentials of Image Mining for Business Process Management. A prototype as an example of the potentials is designed and tested in Sect. 4. In Sect. 5, related work is described and the paper concludes with a discussion of the results.

## 2 Background

### 2.1 Image Mining

Image Mining extract implicit knowledge, relationships through image data and other implicit patterns from images or image databases according to Zhang et al. [10]. Therefore, Image Mining integrates different research streams and results from Data Mining, Machine Learning, Database Management etc. [10, 11]. Zhang et al. [10] argues that Image Mining is not just an extension to the traditional Data Mining. Image Mining can be interpreted as a unique research field and uses as well as integrates different methods from different research fields [10].

The typical Image Mining process can be divided into different steps [12]. The first step is preprocessing of image data, like loading the image and special segmentations [12]. The next step is feature extraction and transformation [12], where common image attributes (e.g. color, edge, shape, texture) are extracted from the images. The third step are Image Mining techniques [12]. There are different Image

Mining techniques according to Zhang et al. [10] (can may be used in the field of BPM):

- Object Recognition
- Image Retrieval
- Image Indexing
- Image Clustering and Classification
- Association Rule Mining
- Neural network

Object Recognition tries to find known as well as similar objects in different images [10]. Through Image Retrieval users as well as information systems can easily find images e.g. based on different patterns [10, 13]. Further, image retrieval is a process of processing limited information to support users retrieval goals at a short time [12]. Furthermore, approaches to indexing images for implementing a information system to retrieve images as well as image data are necessary [10]. To find out similarities of different images as well as cluster different images according to their individualities an image classification and image clustering is needed [10]. Through Association Rule Mining [10] interesting trends, patterns and (pattern) rules of different images can be extracted, Rule Mining can be applied based on a large database of images or e.g. a combined collection of images [10, 14]. Artificial neuronal networks can be used to mine a large amount of image data for feature extraction [10, 15, 16].

Finally, based on the results of these techniques an evaluation and knowledge creation are the last steps of the typical Image Mining process [12]. Therefore, decisions related to BPM are possible (e.g. analyze of graphical process documentation). General aspects of BPM are described in the next section.

## *2.2 Business Process Management*

Business process management [17] is the method-based application of methods, techniques, and tools to business processes during their lifecycle [1]. There are different definition of the business process lifecycle [1, 17]. The most frequently used phases are design, deployment, operation and optimization. We will use the definition developed in [18].

**Process Identification**

Starting from a business perspective, the processes are identified that contribute to achieving a business goal [18]. They are also delimited and related to each other. Often, the processes found are integrated into an enterprise architecture. The process identification phase consists of two sub-phases. In the designation phase [18], an understanding of the processes in an organization and their interrelationship shall be achieved. Depending on the abstraction level, different numbers of processes

may emerge. In the following evaluation phase, the processes found in the designation phase are prioritized according to their need for modeling, redesign etc.

**Process Discovery/Design**

The goal of process discovery [18] is to collect information about an existing process and create a documentation of the present state of the process. Using modeling approaches such as BPMN [3] or ARIS [2] the current or strived for state of the process is depicted in one or several models. Process discovery has to cope with three challenges [18]. First, the knowledge about the process is fragmented. Frequently, no single domain expert has a complete picture of the process but only parts of it. Second, the knowledge of the domain experts is often organized from a case-oriented way, but not a process-oriented view. Third, the business domain experts are not familiar with process modeling methods. Instead, they are using ad hoc defined approaches for depicting their knowledge. Therefore, it is important to assure that these informal descriptions are in sync with the formal ones. The prototype presented in section four supports this use case.

**Process Analysis**

The processes are analyzed using both qualitative and quantitative means [18]. Value-Added Analysis [19], Root Cause Analysis [20] and Issue Documentation are important steps of the qualitative process analysis.

Goal of the value-added analysis is to identify unnecessary process steps and to remove unnecessary steps. To do so, the process steps are classified into value-adding, business value-adding and non-value adding tasks. Then, the non-value adding tasks are eliminated either completely or automated as far as possible.

In the root-cause analysis [20], the relationship between adverse effects on one hand and causal and contributing effects on the other side shall be identified. Cause-effect diagrams and why-why diagrams are frequently used by means of root-cause analysis.

As a result, a register of issues is created. Whenever possible the issues are not only described qualitatively, but also their impact is quantified. Furthermore, the list of issues should be prioritized, pareto charts are important means of doing this.

Quantitative Process Analysis [21] start with capturing data covering the process performance dimensions' time, cost, quality and flexibility. In flow analysis, important performance indicators such as cycle time are determined. Other analysis objects are queues and queue lengths.

**Process Redesign/Improvement**

Using the register of issues during process analysis changes shall be identified to resolve these issues [22]. If there are multiple ways to resolve an issue, they should be compared. As a result a set of changes is proposed that addresses seven elements [18]: internal and external customers, business operations, business behavior, organization, information, technology and the external environment. However, the

possible goals time, cost, quality and flexibility cannot be achieved to the same extent at the same time.

### Process Implementation

This phase transforms the as-is process into the to-be process [18]. Process implementation embraces both organization and information systems. Often it is started by selecting a process automation platform such a dedicated business process management system [23], workflow management systems or enterprise resource planning software. In this way manual tasks can be replaced by automated one. These information systems furthermore provide execution transparency and the enforcement of regulatory rules and laws.

### Process Monitoring and Controlling

After deployment, the business process the operational phase of the business process starts and process instances are created [18, 24]. They represent the execution of the business process such as business transactions. During the operation phase, data representing is collected for later analysis. This data is used in the optimization phase, in order to find possible improvements. Themes for performance analysis are time, cost, quality and flexibility.

## 3 Potentials of Image Mining for Business Process Management

Based on a systematic literature review according to Kitchenham [16] in databases like SpringerLink, IeeeXplore, AISel, Sciencedirect ACM digital library with keywords like "Image Mining" AND "BPM" or "Business Process Management" for the last decade, we cannot find research papers which address an overview or broad insights of the use of Image Mining for Business Process Management. Therefore, we define in the following some core aspects and potentials of Image Mining for Business Process Management according to the fundamentals of BPM according to Sect. 2.2 of the paper.

Images can be differentiated into documents, drawings and pictures. Documents contain textual information, that can be recovered using optical character recognition. Drawings contain graph-based information. Approaches for recovering graphical information are just starting, such as graph detection in OneNote Drawings. Pictorial data contains not directly recoverable information, but often allows to detect metadata, such as the types of products depicted etc. During the business process lifecycle, image data are created on many occasions, such as workshops, meetings, documentation etc. There are a number of sources of image data. In many enterprises, paper documents do not travel within the organization, but are scanned on arrival. Other significant sources are mobile phones and tablets with cameras.

**Process Identification**

A lot of Image data is created in the designation phase, especially during workshops and meetings. Drawings depicting the anticipated process architecture are created. The image data originate from scans and cameras. Object recognition and retrieval can be used to identify and find processes.

**Process Discovery/Design**

Although powerful process modeling tools are available, plenty of image-based data is created during process discovery and design, because the domain experts are not familiar with process modeling methods and the tools supporting them. Often ad hoc defined approaches are used for depicting processes. A particular challenge is the fragmentation of process knowledge leading to multiple separate images that cover the same process. Furthermore, the case-oriented perspective of the domain experts has to be transformed to a process-oriented view. In addition, checking if the modeled processes are correct (e.g. with a comparison of textual descriptions) can be made through object recognition.

**Process Analysis**

Image data containing drawings are a major source for value-added analysis in order to identify unnecessary process steps and to remove unnecessary steps. In root-cause analysis, cause-effect diagrams and why-why diagrams are frequently used means. Pareto charts are important means for prioritizing issues found during process analysis. Image-based data supports quantitative process analysis e.g. to capture customer queues and queue lengths. Object as well as image recognition can be used to detect differences in business process models. Furthermore, image clustering and classification can help to e.g. better understand similarities in processes.

**Process Redesign/Improvement**

During process redesign and improvement, image-based data is created as a result of workshops, meetings etc. They contain suggestions to redesign and improve areas such as internal and external customers, business operations, business behavior, organization, information, technology and the external environment. In this step, e.g. object recognition can be used to check if the improved business process is modeled correct (e.g. in comparison to the specification).

**Process Implementation**

Image-based data in this phase often contains specifications as well as external information such as of regulatory rules and laws. Image mining can support the process implementation phase by using image mining techniques.

**Process Monitoring and Controlling**

During process operation, a lot of image data is created in many enterprises. Paper documents do not travel within the organization, but are scanned on arrival.

Another important source are mobile phones and tablets with their built cameras. In production environments, images are used for documenting productions quality. All this data are collected for later analysis and to find possible improvements. Image indexing and retrieval are an important in the means to do this analysis.

## 4 Prototype: Object Recognition in Business Process Models Through Image Mining

To show the potentials of one area of Image Mining for Business Process Management, we implement a Prototype for object recognition of Business Process Models. The goal of this prototype is to detect business process modeling elements like gateways, activities etc. from images. The detected modeling elements and their order can be further used to interpret the model and to pre-check with other models (e.g. textual descriptions [25]). The prototype was designed based on general prototyping principles according to [26] and implemented through the software Rapid Miner [27] and the Image Mining package [28, 29] (BurgSys). As a modeling notation, we used the EPC notation [2]. The EPC is well known and used in practice and has e.g. not so much modeling elements like BPMN [2, 3].

In the following, we describe the implementation of our prototype for the modeling element "XOR" (exclusive disjunction) of the EPC. For other modeling elements (like activities, AND, OR) the implementations are similar. First two types of images were load into the mining software. One type are images with different representations of XOR, like shown below (Fig. 1).

The other types are images without XOR representations. The next step is the core detector algorithm, which uses these two types of images to learn the differences and to build a model for correct detection of the business process modeling element XOR. As a detector algorithm, we used the fast haar detector, which is very common for detecting objects [29, 30]. The generated model was used to detect the modeling element XOR in the modeled business processes. Therefore, the modeled business processes were load and then discovered based on the model. Finally, the detected XOR were extracted.

To check the possibilities of our prototype, we used the modeled business process of [25] of the business case "important credit application processing" for evaluation (Fig. 2).

The results of the prototype for detecting "XOR" in this sample case is shown in the following figure (Fig. 4). Furthermore, other modeling notation elements like AND, OR, activities etc. can be detected similarly to the described case. The

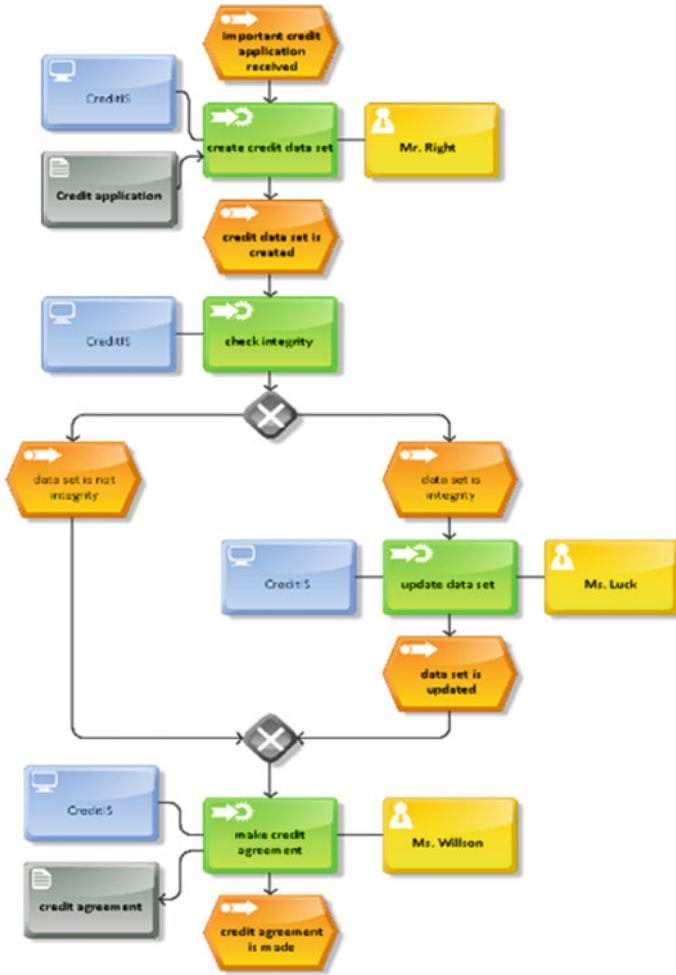**Fig. 1** Two examples of XOR representations in images



etc.

**Fig. 2** Sample business process "important credit application processing" [25]

following figure shows an short excerpt of the implementations through Rapid Miner 5.4 [27, 28] (Fig. 3).

As seen in Fig. 4 the two "XOR" are detected correctly for this sample process. Further evaluation experiments show similar results. The results of the prototype can be used e.g. for checking the modeled business process with textual descriptions. Therefore, a comparison of the occurrences of notations elements and their order can be a quality check if the business process was modeled correctly. Furthermore, textual analysis of the process model can be made through Text Mining according to [31, 25] and compared with the results of the prototype.
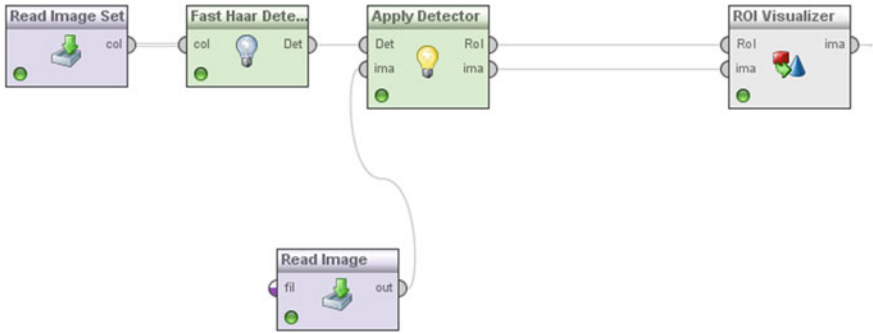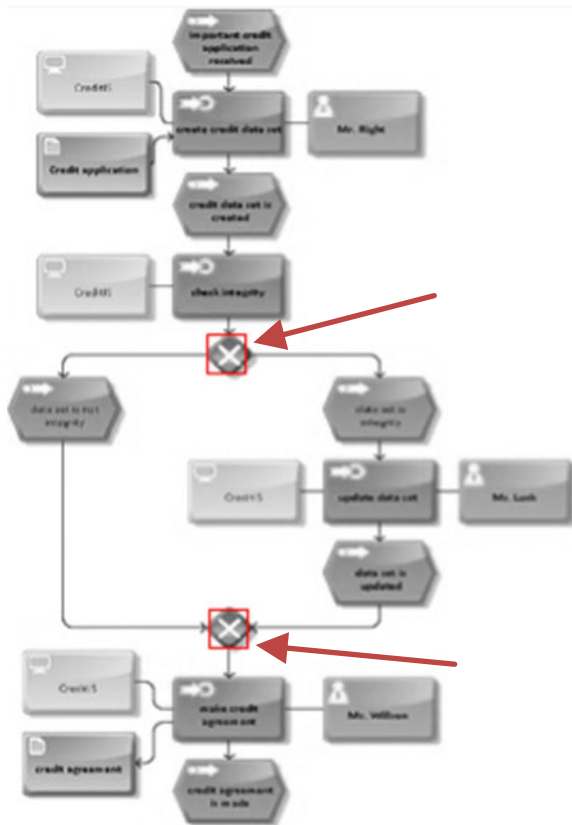
**Fig. 3** Excerpt of implementation in Rapid Miner 5

**Fig. 4** Correct detected XOR of the sample case (*Sample case* [25])

## 5   Related Work

According to Sect. 3, there is no specific research on the potentials of Image Mining for Business Process Management. General research on Image Mining can be found e.g. in [10, 13]. Further concepts of the use of Image Mining are defined in [12]. On an abstract level, there is a relationship to process mining [32]. Both approaches try to extract process related from digitized data. Contrary to process mining, our approach starts from a business expert view and not from low-level system events. Fundamental aspects of Business Process Management can be found in [1, 17]. Modeling notations like EPC and BPMN are defined in [2, 3]. The use of Text Mining for pre-check business process models is discussed in [25]. The need to capture a broader part of the reality in order to improve business process management has already been identified in [33]. Also in [4, 5] the need for extending the input to business process management has been described. The concepts developed here fit very well with social extensions of business process management [5].

## 6   Conclusion and Outlook

In this paper, we introduced basic ideas of the use of Image Mining for Business Process Management. According to the business process lifecycle and the fundamentals of Image Mining, we define core potentials to combine both concepts. Furthermore, we test image recognition from process models.

Our research contributes to the current literature by generating a new view of the combination of Image Mining and Business Process Management. Researchers can use our results to adapt current approaches and to improve business process modeling behavior as well as techniques. Industry managers can use our approach to implement software tools for (pre-) checking modeled business process models and specifications etc. Therefore, the quality of business process projects can be improved. Our prototype addresses the synchronization of image and textual representations.

There are some limitations according to our work. We cannot address to all possibilities of Image Mining and Business Process Management. Furthermore, our prototype for process element recognition is at an early stage and should be improved (e.g. for more notations).

There are great possibilities for future research based on our work. Empirical validation as well as sector specific adoption of the results and the prototype should be done in the future. Furthermore, the functionality of the prototype can be increased e.g. to different notations like UML, BPMN, etc. as well as further specific adoptions.

# References

1. Weske, M.: Business Process Management: Concepts, Languages, Architectures. Springer, Berlin, Heidelberg (2007)
2. Scheer, A.-W., Nüttgens, M.: ARIS architecture and reference models for business process management. In: van der Aalst, W., Desel, J., Oberweis, A. (eds.) Business Process Management, pp. 376–389. Springer, Berlin, Heidelberg (2000)
3. White, S.A.: Introduction to BPMN. IBM Coop. 2008–029 (2004)
4. Bruno, G., Dengler, F., Jennings, B., Khalaf, R., Nurcan, S., Prilla, M., Sarini, M., Schmidt, R., Silva, R.: Key challenges for enabling agile BPM with social software. J. Softw. Maint. Evol. Res. Pract. **23**, 297–326 (2011)
5. Schmidt, R., Nurcan, S.: BPM and social software. In: Ardagna, D., Mecella, M., Yang, J., Aalst, W., Mylopoulos, J., Rosemann, M., Shaw, M.J., Szyperski, C. (eds.) Business Process Management Workshops, pp. 649–658. Springer, Berlin, Heidelberg (2009)
6. Brynjolfsson, E.: Understanding the Digital Economy: Data, Tools, and Research: Data, Tools and Research. The MIT Press (2000)
7. Vincent, A., Varghese, G.: Towards A Robust and Stand-Alone System for Binarization and OCR of Document Images (2015)
8. Office Lens—Windows-Apps im Microsoft Store. https://www.microsoft.com/de-de/store/apps/office-lens/9wzdncrfj3t8
9. Team, O.: OneNote—what's new in January 2016. https://blogs.office.com/2016/01/29/onenote-whats-new-in-january-2016/ (2016)
10. Zhang, J., Hsu, W., Lee, M.L.: Image mining: Issues, frameworks and techniques. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Multimedia Data Mining (MDM/KDD'01). University of Alberta (2001)
11. Burl, M.C., Fowlkes, C., Roden, J.: Mining for image content. Syst. Cybern. Inf. Inf. Syst. Anal. Synth. (1999)
12. Mishra, N., Silakari, D.S.: Image mining in the context of content based image retrieval: a perspective. IJCSI Int. J. Comput. Sci. Issues **9**, 98–107 (2012)
13. Stanchev, P., Flint, M.: Using image mining for image retrieval. In: IASTED Conference "Computer Science and Technology," Cancun, Mexico, pp. 214–218 (2003)
14. Ordonez, C., Omiecinski, E.: Discovering association rules based on image content. In: IEEE Forum on Research and Technology Advances in Digital Libraries, 1999. Proceedings, pp. 38–49. IEEE (1999)
15. Carpenter, G., Grossberg, S., Markuzon, N., Reynolds, J.H., Rosen, A.B., et al.: Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Trans. Neural Netw. **3**, 698–713 (1992)
16. Kitchenham, B.: Procedures for performing systematic reviews. Keele UK Keele Univ. **33**, 1–26 (2004)
17. Van der Aalst, W., ter Hofstede, A., Weske, M.: Business process management: a survey. Bus. Process. Manage. 1019–1019 (2003)
18. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer, Berlin, Heidelberg (2013)
19. Conger, S.: Six sigma and business process management. In: Handbook on Business Process Management, vol. 1, pp. 127–146. Springer (2015)
20. Wilson, P.F.: Root Cause Analysis: A Tool for Total Quality Management. ASQ Quality Press (1993)
21. Zhu, J.: Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets. Springer (2014)
22. Bach, V., Brecht, L., Hess, T., Österle, H.: Enabling Systematic Business Change: Integrated Methods and Software Tools for Business Process Redesign. Springer (2013)
23. Hajo, A.: Reijers: Implementing BPM systems: the role of process orientation. Bus. Process Manag. J. **12**, 389–409 (2006)

24. Metzger, A., Leitner, P., Ivanovic, D., Schmieders, E., Franklin, R., Carro, M., Dustdar, S., Pohl, K.: Comparing and combining predictive business process monitoring techniques. IEEE Trans. Syst. Man Cybern. Syst. 45, 276–290 (2015)
25. Schmidt, R., Möhring, M., Härting, R.-C., Zimmermann, A., Heitmann, J., Blum, F.: Leveraging textual information for improving decision-making in the business process lifecycle. In: Neves-Silva, R., Jain, L.C., and Howlett, R.J. (eds.) Intelligent Decision Technologies. Sorrent (2015)
26. Naumann, J.D., Jenkins, A.M.: Prototyping: the new paradigm for systems development. Mis. Q. 29–44 (1982)
27. Akthar, F., Hahne, C.: RapidMiner 5 Operator Reference. Rapid-GmbH (2012)
28. Burget, R., Karasek, J., Smékal, Z., Uher, V., Dostal, O.: Rapidminer image processing extension: a platform for collaborative research. In: Proceedings of the 33rd International Conference on Telecommunication and Signal Processing, pp. 114–118 (2010)
29. Masek, J., Burget, R., Karasek, J., Uher, V., Guney, S.: Evolutionary improved object detector for ultrasound images. In: 2013 36th International Conference on Telecommunications and Signal Processing (TSP), pp. 586–590. IEEE (2013)
30. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001, pp. I–511. IEEE (2001)
31. Tan, A.-H., et al.: Text mining: The state of the art and the challenges. In: Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, p. 65 (1999)
32. Van Der Aalst, W.: Process mining. Commun. ACM 55, 76–83 (2012)
33. Wang, M., Wang, H.: From process logic to business logic—a cognitive approach to business process management. Inf. Manage. 43, 179–193 (2006)