Application of Bayesian Networks to Recommendations in Business Process Modeling*

Szymon Bobek, Mateusz Baran, Krzysztof Kluza, Grzegorz J. Nalepa

AGH University of Science and Technology,
al. A. Mickiewicza 30, 30-059 Krakow, Poland
{kluza,matb,s.bobek,gjn}@agh.edu.pl

Abstract Formalized process models help to handle, design and store processes in a form understandable for the designers and users. Modeling of business processes is a complex task, which can be supported by recommendations. It is important, as designers prefer to receive and use suggestions during the modeling process. Recommendations make modeling faster and less error-prone because a set of good models is automatically used to help the designer. In this paper, we propose a method that uses Bayesian Networks for recommendation purposes in process modeling. To create such a network, we use configurable business processes that combine a set of reference models.

1 Introduction

Business Processes (BP) in organizations are designed using visual representations and stored as process models. This helps to manage process complexity, especially by using human-understandable notations like Business Process Model and Notation (BPMN). As process modeling by inexperienced users can be a difficult task due to complexity of the problem as well as the richness of the BPMN language, suggestions during the modeling would facilitate it significantly. Based on current progress or additional pieces of information, a designer can be supported during the design of the model. Such an assistance can provide autocompletion mechanisms with capabilities of choosing next process fragments from suggested ones.

There are number of challenges related to the delivery of such assistive mechanisms, including: 1) a suitable repository of existing models that would allow for the training of recommender module, 2) an appropriate machine learning model and learning method for the recommendation, and 3) a feasible integration with the BP modeling on the level of the BPMN representation.

In our research we assume availability of a model repository that provides consistency of models and detects their similarities for further reuse. In fact, such a repository is being developed in the Prosecco project[1]. The objective of the project is to provide tools supporting the management of Small and Medium Enterprises (SMEs) by the introduction of methods for declarative specification of business process models and their

* The paper is supported from the *Prosecco* project funded by NCBR.

[1] See http://prosecco.agh.edu.pl

semantics. We are using the formalization of BP models with BPMN as well as ontological modelling with RDF and OWL to capture the semantics of the models.

In this paper we propose the application of Bayesian Networks to recommend process fragments in BPs modeling. Such a method can help in speeding up modeling process and producing models that are less error prone compared to these designed from scratch. Finally, we support our approach on the use of a configurable process model built on similarities of models of the same group. We propose a simple algorithm that is based on the configurable process model. This supports the reuse of existing process models, especially when a process repository is provided.

The rest of this paper is organized as follows: In Section 2 we present a short overview of the recommendation types and in Section 3 we describe the current state of the art in this research area. Section 4 presents configurable process model that can be used as an input for recommendation algorithm, while Section 5 describes application of Bayesian Networks for such recommendation in process modeling. The paper is summarized in Section 6.

2 Recommendations Types in BP Modeling

Recommendation methods in Business Process modeling can be classified as one of two types: subject-based and position-based classification [1], which are complementary as they are suited for different purposes.

In the first case, the suggestion itself is not directly dependent on the context it is placed in. However, the recommendation algorithms may actually inspect the context to deliver more accurate results. This classification focuses on the subject which is actually suggested, such as:

- attachments to a process model (e.g. decision tables, links, service tasks, subprocesses or call subprocesses),
- textual pieces of information (e.g. names of elements, guard conditions, etc.),
- structural fragments (single elements or structures of elements).

Position-based classification focuses on the part of the model where the suggested artifact is to be placed in the model; thus, we distinguish:

- forward completion – a part of the process is known and the further fragment of the process is to be suggested,
- backward completion – a part of the process is known and the previous fragment of the process is to be suggested,
- autocomplete – any part of the process is known and the rest of the process is to be suggested (a number of items with no outgoing or incoming flows is selected – missing flows will lead to or from the suggested structure).

In this paper, we consider structural fragment recommendation approach that can be used as completion in any of the abovementioned positions. In the following section, we present related works in the BP recommendation area.

3 Related Works

As empirical studies have proven that users prefer to receive and use suggestions during modeling processes [2], several approaches to recommendations in BP modeling have been developed. They are based on different factors such as labels of elements, current progress of modeling process, or additional pieces of information like process descriptions or annotations.

Among attachment recommendations, support with finding appropriate services was proposed by Born et al. [3] and Nguyen et al. [4]. Such a recommendation mechanism can take advantage of context specified by the process fragment [4] or historical data [5]. Approaches that recommend textual pieces of information, such as names of tasks, were proposed by Leopold et al. [6] and extended in [7].

In the case of structural recommendations, Kopp et al. [8] showed how to autocomplete BPMN fragments in order to enable its verification. Although this approach does not require any additional information, it is very limited in the case of recommendations. The more useful existing algorithms are based on graph grammars for process models [9,10], process descriptions [11], automatic tagging mechanism [12,2], annotations of process tasks [13] or context matching [14]. Case-based reasoning for workflow adaptation was discussed in [15]. It allows for structural adaptations of workflow instances at build time or at run time, and supports the designer in performing such adaptations by an automated method based on the adaptation episodes from the past.

In our research, we use Bayesian Networks for recommendation purposes. As such a network is created based on a configurable Business Process, in the following section we present a short overview of BP configuration.

4 Configurable Processes

BP configuration is a method allowing for expressing similarities between two or more BP models. There are mechanisms for comparing processes and managing them in large repositories [16,17], refactorization of such repositories [18], as well as automatic methods for extracting cloned fragments in repositories of models [18,19].

There are several methods of extraction of configurable processes. They are focused on different goals. Analyzing a configurable Business Process reveals a high-level workflow that might not be apparent when analyzing particular models. The structure is partially lost in the process, so this gives no benefit for process recommendation. The method of interest in this paper are models merged into configurable model [20]. They allow the designer to see several processes as special cases of one configurable model. The model emphasizes similarities preserving all the important details. There is an active research field in the area of configurable Business Processes. In [21], Rosemann et al. described an approach focused on hand-made diagrams for the purpose of reference modeling. La Rosa et al. [22] extended it with objects and roles.

Figure 1 shows an example of configuration of parts of BP models. First, four similar processes are presented. They are composed of certain elements that are either present or not. In Figure 1 there is a configurable process that generalizes all of them. The numbers in square brackets indicate in which process a given element appears. If an element does not appear, the control flow goes on to the next element. This is a simplified approach in comparison to La Rosa's [22] version. Nevertheless it is sufficiently expressive in given case and more appropriate for later usage with Bayesian Network that we will use. There is also a technique that enables process designer to specify more than one variant of a task. The so-called variant-rich process models were explored in PESOA project [23,24].

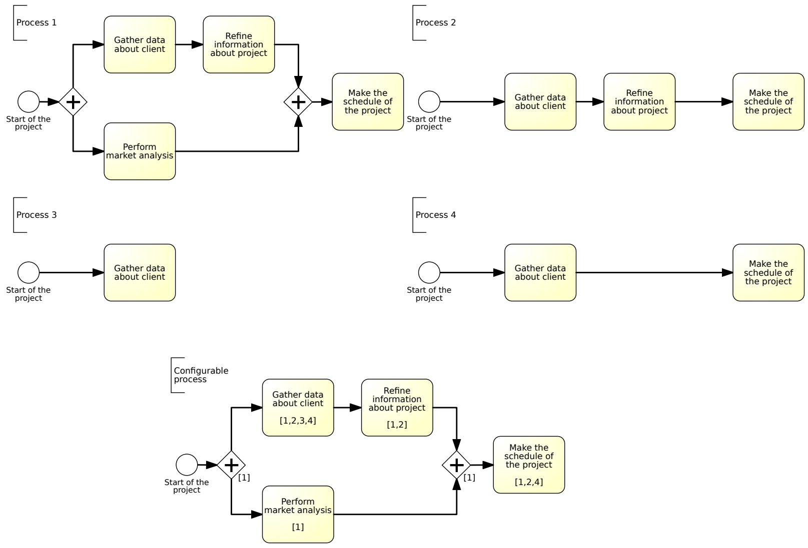

Figure 1. An example of Business Process configuration

For the purpose of evaluation of the method we use a case study of four BP (see Figure 2). The processes describe the implementation of a development project in four different SME. The project starts when appropriate decision is made in the company. Initial preparations are made first, then the project is created and several end-of-the-process tasks are performed. The "Make settlements" task appears twice – in some companies it appears later and in some earlier in the process but always exactly once. In case of one company the "Perform tasks" task triggers an event at the end of each milestone. Exactly the same steps are performed then as are after the project is completed, so the flows actually merge.

5 Bayesian Networks for Recommendation

In this section we present our method that applies Bayesian Networks (BN) for recommendation purposes in process modeling. Thus, the following subsections describe BN representation, modeling and training issues. Then, the recommendation scenario is presented with open issues discussion.

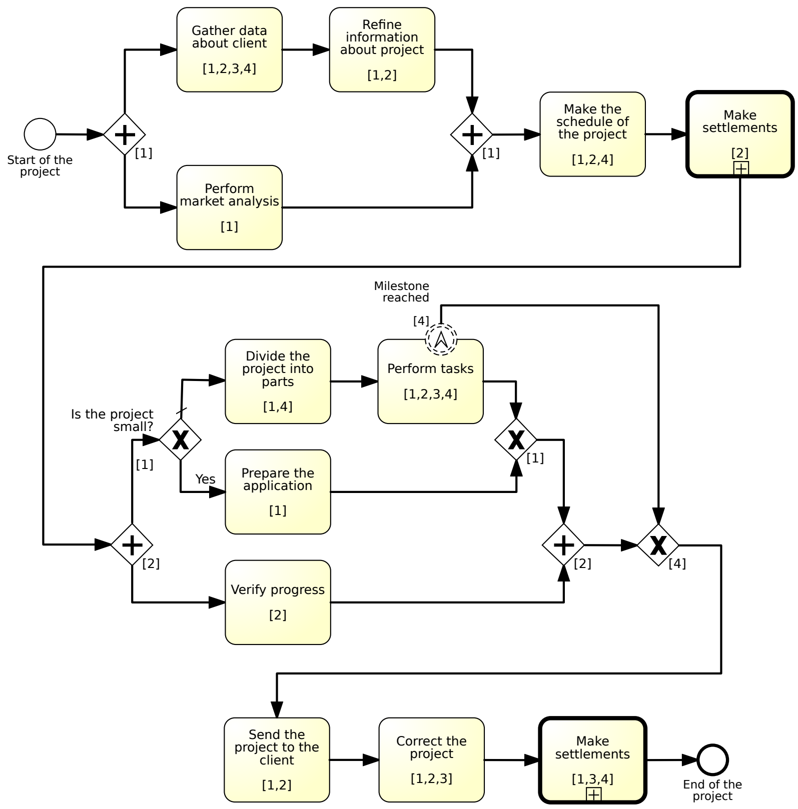

Figure 2. Configurable Business Process

5.1 Bayesian Network Representation

Bayesian Network [25] is an acyclic graph that represents dependencies between random variables and provide graphical representation of the probabilistic model. This representation serves as the basis for compactly encoding a complex probability distribution over a high-dimensional space [26]. The most important advantage of Bayesian Network models is that it is possible to directly exploit the graphical representation of BP diagrams, which can be easily translated into such model. Another advantage is that the output of a recommendation is a set of probabilities, which allows for ranking the suggestion from the most probable to the least probable.

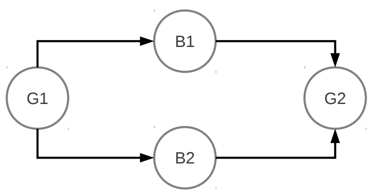

Figure 3. A simple Bayesian Network example

In Figure 3 a simple example of a BN is presented. In order to calculate the probability of the value of the random variable *B1*, the equation 1 can be used. The G1 and G2 can be denoted as BPMN gateways, and B1 and B2 as other BPMN blocks, like Tasks or Events. Thus, having any of these blocks given, we can calculate a probability of a particular block being a missing part.

$$P(B1) = \sum_{G1} \sum_{G2} \sum_{B2} P(G1)P(B1|G1)P(B2|G1)P(G2|B1, B2) \tag{1}$$

The following subsections provides the details how this can be achieved.

5.2 Bayesian Network Modeling

The transformation from a configurable model to a BN model is straightforward. Each node in a configurable process has to be modeled as a random variable in BN. Therefore, each node in a configurable process is translated into a node in the network. The flow that is modeled by configurable process represent dependencies between nodes. These dependencies also can be translated directly to the BN model. For instance, a BN that defines dependencies between particular nodes of configurable process presented in Figure 2 can be modeled using the BN representation presented in Figure 4.

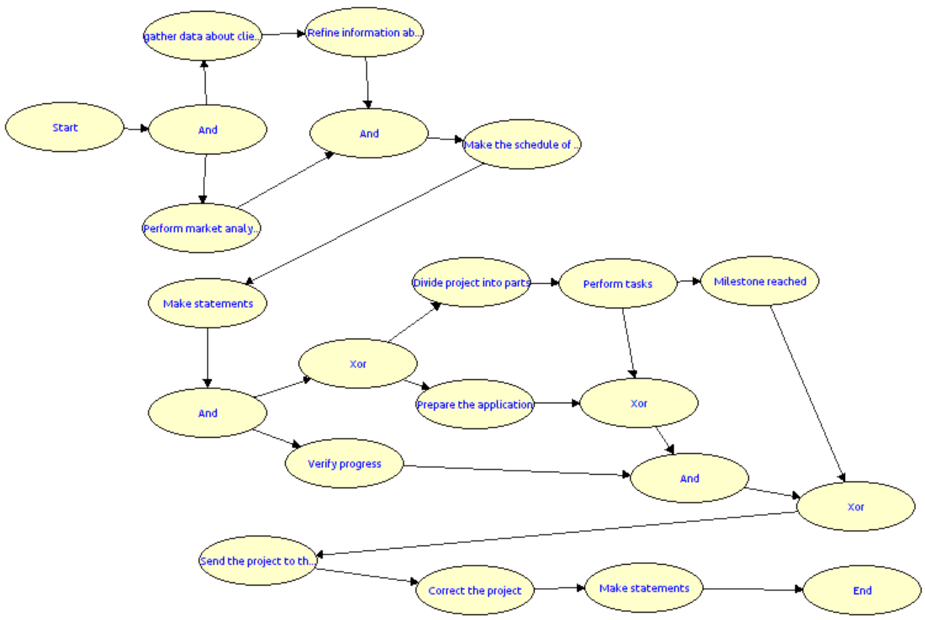

Figure 4. Bayesian Network representing the configurable process from Figure 2

The Bayesian Network presented in Figure 4 captures only dependencies that are a direct consequence of the control flow in the process model. However, we can add more semantics to the Bayesian Network, what would allow us to capture some indirect dependencies that arise from the company characteristics that are not included in configurable process. Such dependencies encoded into BN would allow for better recommendation accuracy, preserving the BPMN grammar at the same time. An example of such a dependency is presented in Figure 5. It encodes that the project size, type and company size may influence other components presence or absence in the Bayesian Network model.

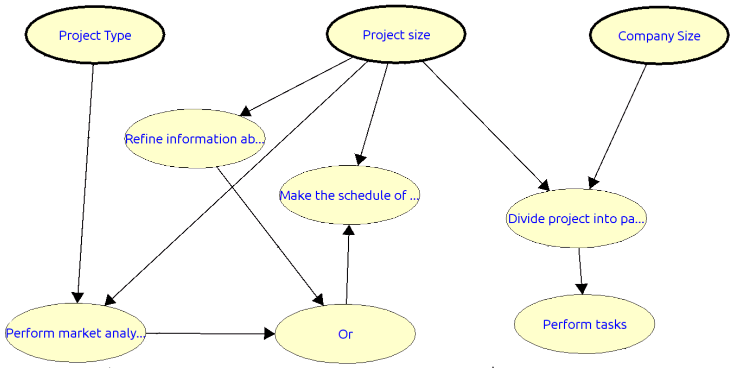

Figure 5. Bayesian Network capturing dependencies between BPMN and additional background knowledge about project and company characteristics.

5.3 Bayesian Network Training

There are several methods for training Bayesian Networks. The comprehensive list and comparison of them can be found in [27]. For the purpose of this paper, we use the Expectation Maximization algorithm to perform bayesian network training. The software we used to model and train our network is called Samiam[2]. The important part of the learning process is providing training data. In our case, the training data was a configurable process serialized to a CSV file. Each column in the file represents a node in configurable process, whereas each row represents a separate process model that was used to create the former.

5.4 Recommendation Scenario

The trained Bayesian Network can be used to recommend to a user which elements of the BPMN diagram should be included in the currently modeled process. Such suggestions can be done in two ways, either by recommending a next possible element, or by suggesting a group of elements presence of which is highly probable.

In the case of a single element structural recommendation, we recommend the next element in the currently modeled process. This situation is presented in Figure 6. The red circled elements are the elements that a user has already included or excluded from his or her process. The green blocks represent the confidence that the particular node should be included into the diagram. As it can be noticed, it is not only possible to recommend the further nodes, but also to *autocomplete* omitted elements (like the *Refine information about the project* node from the Figure 6).

[2] See: `http://reasoning.cs.ucla.edu/samiam`.

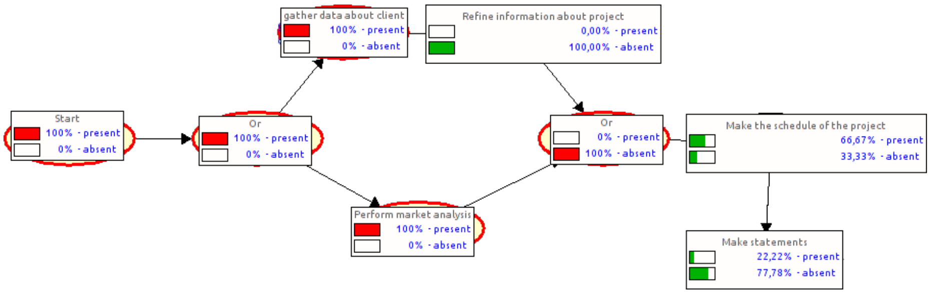

Figure 6. Recommending BPMN nodes using the Bayesian Network learned from the configurable process (presented in Figure 2)

Another, extended approach can focus not only on the structure of a process model but rather on dependencies between its components. In such the approach, only the elements that are highly probable to be present in the process, based on the already added elements, would be recommended. This, however, requires from a designer of the BN to capture all dependencies between BPMN nodes.

5.5 Open issues

The proposed solution has still some unresolved challenges to face. The structural recommendation assumes that a user builds a process model step by step, choosing a next probable element of the learned model. The conformance checking issue arises at this stage, since the BN model has to be compared to the BP model. This is easily achievable when a user picks only the nodes from the learned model, however when he or she enters some elements unseen in the learning phase, the conformance checking is not a trivial task any more. Such a case is related with other issues. Once the Bayesian Network model observes unseen configuration of random variables, it produces probabilities very close or equal to zero. As a result, the recommendation process is stopped. Currently, a Bayesian Network model is built manually based on a configurable process. However, this task can be automated and it is considered as a future work.

6 Conclusion and future work

This paper proposes an application of Bayesian Networks to recommendation purposes in process modeling. Such a method can help in speeding up modeling process and producing less error prone models than modeling from scratch. The algorithm is based on the configurable process models.

The proposed suggestion mechanism was tested on a configurable model, prepared from four real-world processes. The processes were adapted to preserve anonymity and highlight the similarities. In fact more detailed processes differ so much that typical configuration mechanism would produce almost worthless results as most of the tasks would not be matched between processes. This issue can be solved by using hierarchical BP configuration [28].

Our future work will focus on specifying recommendation approach for company management systems in order to enhance modeling process and evaluation of the selected recommendation methods. We plan to carry out a set of experiments aimed at testing recommendation approaches on various model sets. We also plan to extend the presented approach in order to use it in collaborative modeling environment [29].

References

1. Kluza, K., Baran, M., Bobek, S., Nalepa, G.J.: Recommendation methods in business process modeling. In Nalepa, G.J., Baumeister, J., eds.: 9th Workshop on Knowledge Engineering and Software Engineering (KESE9) at the KI 2013, the 36th Annual Conference on Artificial Intelligence, in Koblenz, September 16-20, 2013. (2013)
2. Koschmider, A., Hornung, T., Oberweis, A.: Recommendation-based editor for business process modeling. Data & Knowledge Engineering **70** (2011) 483 – 503
3. Born, M., Brelage, C., Markovic, I., Pfeiffer, D., Weber, I.: Auto-completion for executable business process models. In Ardagna, D., Mecella, M., Yang, J., eds.: Business Process Management Workshops. Volume 17 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg (2009) 510–515
4. Chan, N., Gaaloul, W., Tata, S.: Context-based service recommendation for assisting business process design. In Huemer, C., Setzer, T., eds.: E-Commerce and Web Technologies. Volume 85 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg (2011) 39–51
5. Chan, N., Gaaloul, W., Tata, S.: A recommender system based on historical usage data for web service discovery. Service Oriented Computing and Applications **6** (2012) 51–63
6. Leopold, H., Mendling, J., Reijers, H.A.: On the automatic labeling of process models. In Mouratidis, H., Rolland, C., eds.: Advanced Information Systems Engineering. Volume 6741 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2011) 512–520
7. Leopold, H., Smirnov, S., Mendling, J.: On the refactoring of activity labels in business process models. Information Systems **37** (2012) 443–459
8. Kopp, O., Leymann, F., Schumm, D., Unger, T.: On bpmn process fragment auto-completion. In Eichhorn, D., Koschmider, A., Zhang, H., eds.: Services und ihre Komposition. Proceedings of the 3rd Central-European Workshop on Services and their Composition, ZEUS 2011, Karlsruhe, Germany, February 21/22. Volume 705 of CEUR Workshop Proceedings., CEUR (2011) 58–64
9. Mazanek, S., Minas, M.: Business process models as a showcase for syntax-based assistance in diagram editors. In: Proceedings of the 12th International Conference on Model Driven Engineering Languages and Systems. MODELS '09, Berlin, Heidelberg, Springer-Verlag (2009) 322–336
10. Mazanek, S., Rutetzki, C., Minas, M.: Sketch-based diagram editors with user assistance based on graph transformation and graph drawing techniques. In de Lara, J., Varro, D., eds.: Proceedings of the Fourth International Workshop on Graph-Based Tools (GraBaTs 2010), University of Twente, Enschede, The Netherlands, September 28, 2010. Satellite event of ICGT'10. Volume 32 of Electronic Communications of the EASST. (2010)
11. Hornung, T., Koschmider, A., Lausen, G.: Recommendation based process modeling support: Method and user experience. In: Proceedings of the 27th International Conference on Conceptual Modeling. ER '08, Berlin, Heidelberg, Springer-Verlag (2008) 265–278
12. Koschmider, A., Oberweis, A.: Designing business processes with a recommendation-based editor. In Brocke, J., Rosemann, M., eds.: Handbook on Business Process Management 1. International Handbooks on Information Systems. Springer Berlin Heidelberg (2010) 299–312

13. Wieloch, K., Filipowska, A., Kaczmarek, M.: Autocompletion for business process modelling. In Abramowicz, W., Maciaszek, L., Węcel, K., eds.: Business Information Systems Workshops. Volume 97 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg (2011) 30–40
14. Chan, N., Gaaloul, W., Tata, S.: Assisting business process design by activity neighborhood context matching. In Liu, C., Ludwig, H., Toumani, F., Yu, Q., eds.: Service-Oriented Computing. Volume 7636 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 541–549
15. Minor, M., Bergmann, R., Görg, S., Walter, K.: Towards case-based adaptation of workflows. In Bichindaritz, I., Montani, S., eds.: ICCBR. Volume 6176 of Lecture Notes in Computer Science., Springer (2010) 421–435
16. Kunze, M., Weske, M.: Metric trees for efficient similarity search in large process model repositories. In Muehlen, M., Su, J., eds.: Business Process Management Workshops. Volume 66 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg (2011) 535–546
17. Dijkman, R., Rosa, M.L., Reijers, H.A.: Managing large collections of business process models – current techniques and challenges. Computers in Industry **63** (2012) 91–97
18. Weber, B., Reichert, M., Mendling, J., Reijers, H.A.: Refactoring large process model repositories. Computers in Industry **62** (2011) 467–486
19. Uba, R., Dumas, M., García-Bañuelos, L., Rosa, M.: Clone detection in repositories of business process models. In Rinderle-Ma, S., Toumani, F., Wolf, K., eds.: Business Process Management. Volume 6896 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2011) 248–264
20. van der Aalst, W.M.: Business process management: A comprehensive survey. ISRN Software Engineering **2013** (2013)
21. Rosemann, M., van der Aalst, W.M.P.: A configurable reference modelling language. Inf. Syst. **32** (2007) 1–23
22. Rosa, M.L., Dumas, M., ter Hofstede, A.H., Mendling, J.: Configurable multi-perspective business process models. Information Systems **36** (2011) 313 – 340
23. Puhlmann, F., Schnieders, A., Weiland, J., Weske, M.: Variability Mechanisms for Process Models. PESOA-Report TR 17/2005, Process Family Engineering in Service-Oriented Applications (pesoa). BMBF-Project. Report, Hasso Plattner Institut, Postdam, Germany (2005)
24. Schnieders, A., Puhlmann, F.: Variability mechanisms in e-business process families. In: Proc. International Conference on Business Information Systems (BIS 2006). (2006) 583–601
25. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning **29** (1997) 131–163
26. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. MIT Press (2009)
27. Neapolitan, R.E.: Learning Bayesian Networks. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (2003)
28. Baran, M., Kluza, K., Nalepa, G.J., Ligęza, A.: A hierarchical approach for configuring business processes. In Ganzha, M., Maciaszek, L.A., Paprzycki, M., eds.: Proceedings of the Federated Conference on Computer Science and Information Systems – FedCSIS 2013, Krakow, Poland, 8-11 September 2013, IEEE (2013) 931–937
29. Nalepa, G.J., Kluza, K., Ciaputa, U.: Proposal of automation of the collaborative modeling and evaluation of business processes using a semantic wiki. In: Proceedings of the 17th IEEE International Conference on Emerging Technologies and Factory Automation ETFA 2012, Kraków, Poland, 28 September 2012. (2012)