O. Chala

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

LOGICAL - PROBABILISTIC REPRESENTATION OF CAUSAL DEPENDENCIES BETWEEN EVENTS IN BUSINESS PROCESS MANAGEMENT

The **subject matter** of the article are the processes of identifying knowledge in the form of causal relationships based on the analysis of the log of the business process. The **goal** is to develop a logical-probabilistic model of cause-effect relationships between pairs of log events that describes the implementation of the business process's action to support the solution of the task of automating the construction of the knowledge base of the process management system. **Tasks**: Select context constraints and limitations on the execution of business process actions that can be obtained as a result of log analysis; develop an approach to extract the probabilistic and logical components of cause-effect dependencies; to develop a logical-probabilistic model of causal relationships. The **methods** used are: methods for constructing predicate models; Bayesian methods of constructing probabilistic models. The following **results** are obtained. Formalized class of causal dependencies for knowledge-intensive business processes. Such dependencies can take into account informal knowledge of the business process. Within this class there are: a predicate description of the state of the context based on information about values of attributes of log events; contextual constraints on doing business process actions; probabilistic conditions for implementing the business process. **Conclusions.** The scientific novelty of the results obtained is as follows: a logical-probabilistic model of cause-effect relationships between pairs of log events describing the performance of the business process is proposed. The model binds a logical description of the state of the context before and after the completion of each activity of the business process, as well as a logical description of the constraints on the actions of the process and a probabilistic description of the conditions for the execution of these actions. In practical terms, the model provides an opportunity to solve problems of extracting, expanding and integrating knowledge based on the analysis of logs of business processes.

Keywords: causality; knowledge; knowledge base; dependencies; business process; workflow; event log.

Introduction

Business process management systems are designed to manage the enterprise through the management of business processes (BP). The process control cycle includes the stages of building business process models, configuring these models in the management system, using BP models for managing and improving process models [1].

Business process models specify the sequence of actions that are required to obtain the product or service needed by the consumer. Improvement of process models allows to eliminate bottlenecks, and also to adapt the model to new conditions of business process execution. Such conditions arise as a result of changes in the activities of the enterprise, new market contracts, etc.

The problem of adaptation of process models is especially important for knowledge-intensive business processes. Such processes differ in that knowledge workers can change their structure. They use their personal knowledge and experience to change activities when executing a business process. As a result, the a priori model of the process becomes inadequate to the current instance of the business process [2].

To continuously improve the BP, it is necessary to supplement the process model with a knowledge base. This knowledge base can be used to assess the effectiveness of possible changes in the process model, taking into account constraints on the performance of business process actions.

When building a knowledge base, it is necessary to take into account the limited, inaccurate personal knowledge of the knowledge workers, who change the sequence of actions of the business process [3].

Related Work

At present, approaches to building knowledge bases to support process management are focused primarily on knowledge engineering. The specialists in knowledge engineering provide for human-machine interaction in the allocation and formalization of knowledge. The Semantic Web paradigm is used, as well as micro-blogging ideas [4].

At the same time, the methods of automated construction of knowledge bases [5-7] and approaches to the probability representation of knowledge are intensively developed [7-9]. These approaches use probabilistic causalities [10, 11], probabilistic relations [12], probabilistic graphical models [13], probabilistic logical reasoning [14, 15].

However, these methods and models are intended primarily for use in information retrieval and information referral systems. They are also used to support speech and image recognition.

Further research is required to solve the problems of supporting decisions on process management, taking into account the inexact, probabilistic nature of knowledge of people who perform knowledge-intensive business processes.

As a result, information regarding the further behavior of the business process becomes uncertain. This indicates the importance of constructing a logical-probabilistic representation of knowledge.

The purpose of this article is to develop a probabilistic-logical representation of cause-effect dependencies in a knowledge base of the process management system in order to perform such tasks:

– using personal knowledge of employees to improve the business process model;

– supporting decisions on the assessment of the actions of knowledge workers, who changes the business process.

Knowledge-intensive business process: causality

Knowledge-intensive business process has a number of properties that affect the representation of causal dependencies.

The process has three structural levels: a workflow; contextual knowledge about the actions of the process; context.

Workers can change the sequence of actions based on their knowledge and the current state of the context. This means that there are causal relationships between the context and the process activities (Fig. 1).
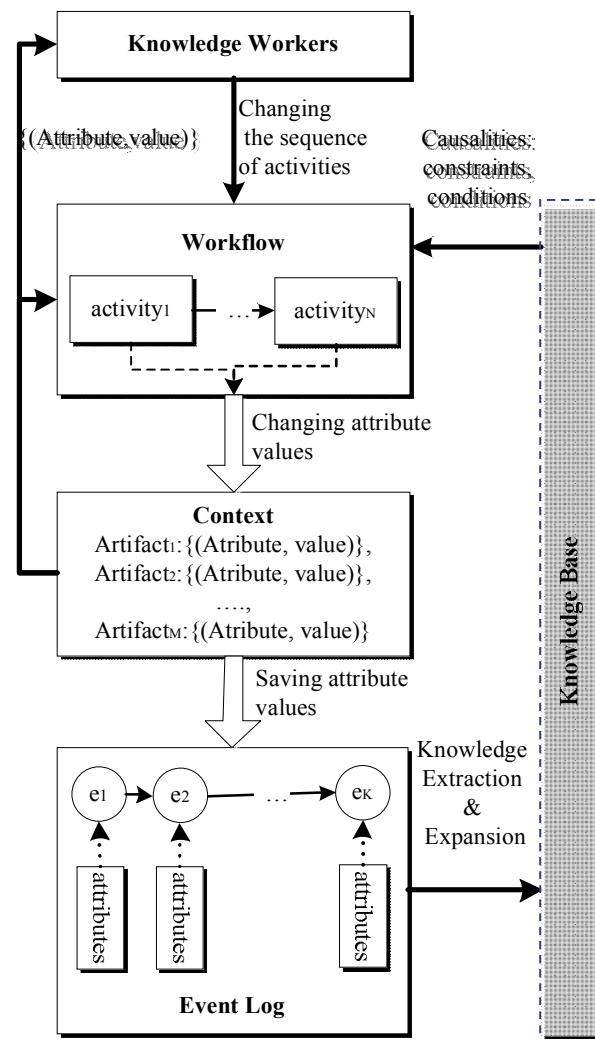

Fig. 1. Causal dependencies
of knowledge-intensive business process

Context affects the execution of actions in the domain and includes a set of artifacts.

Artifacts are objects that the process uses when executing. Artifacts are characterized by a set of properties. Properties are represented as attribute - value pairs.

The current state of the business process context depends on such variables:

- a set of artifacts that this process uses;
- a set of properties of artifacts;
- the values of the properties of these artifacts.

Sets of artifacts and immutable attributes set the static aspect of the context. In other words, artifacts with predefined attribute sets are used by the process all the time.

The absence of such artifacts does not allow performing the corresponding actions of the process. Therefore, the static aspect of the context sets limits on the execution of business process actions. In this case, for each constraint of the business process, the rule is triggered: as long as the corresponding activity does not satisfy the constraint, it cannot be included in the workflow:

$$\forall c \in C \, \exists D_l \subseteq D, D_l \neq \varnothing : D_l \bigcap wf = \varnothing | \neg c, \qquad (1)$$

where c is the current constraint; D_l – a subset of activities that must satisfy the constraint; wf – the workflow of the current instance of the business process..

The set of all possible sequences of actions forms the general flow of work of the business process, Wf i.e. $Wf = \{wf\}$.

Limitations for a particular action must be satisfied in all instances of the process. A process instance is an implementation of the BP model, similar to an instance of a class in object-oriented programming.

The set of pairs "attribute-changing value" characterizes the dynamic aspect of the context. Attribute values changes as a result of process actions or environmental influences.

In order to execute a process action, a subset of attributes must have certain values.

Therefore, the dynamic aspect is characterized by the conditions for performing BP actions at the time when the desired attribute values are obtained:

$$wf, t \models d \mid l \wedge c, \; l \in L, \; c \in C, d \in D, \qquad (2)$$

where l is the condition to perform activity d; t – the moment of time when the condition l and constraint c are satisfied.

Unlike constraints, the conditions for starting actions cannot be satisfied for all instances of the BP:

$$\exists wf \in Wf : wf, t \nvDash d \mid l \wedge c. \qquad (3)$$

This difference is due to the fact that employees can change the process and perform the action under other input conditions. In the general case, there is a nonzero probability that the action will be performed with the true expression l.

In order to identify the causal dependence of the "condition-activity" kind and then to predict possible sequences of knowledge-intensive business process work, we set the threshold value ε of the probability for this condition. The business process is characterized by the implementation of actions in time. Therefore, this causal dependence is represented using the temporal modal logic operators:

$$X^+ p(l) > \varepsilon \Rightarrow Xd \,|\, c, \qquad (4)$$

where $p(l)$ is the probability of the condition l being met; X – the temporal modal logic operator $neXt$, it specifies the execution of the action d at the next discrete point in time; X^+ operator sets the condition being met from the current moment to the next discrete point in time.

Thus, causal dependencies for a knowledge-intensive business process have logical limitations and probabilistic conditions.

Approach to extracting cause-effect dependencies between business process events

The sequence of automated knowledge base construction includes the tasks of extracting knowledge, expanding the knowledge base and integrating knowledge with existing models (for example, business process models).

In order to solve the first task, the extraction of knowledge in the process management system, it is necessary to identify the context of the actions of the process. Then, based on the context analysis, we can set constraints and conditions for the execution of actions.

We can identify the context in two ways:

– By using a priori knowledge of the subject area;

– By analyzing the event logs that the process management system forms when executing a business process.

In the first case, the expert in the domain sets the permissible sequences of actions for each artifact. Usually such dependencies are determined by the life cycle of these objects. Therefore, we can specify a restriction on the sequence of the process actions in the form of permissible links between the operations for processing the artifact in accordance with its life cycle.

The conditions for the execution of actions are determined by comparing the permissible variants of the life cycle of artifacts. Therefore, it is necessary to compare several alternative sets of conditions when performing one action.

The disadvantages of this method are as follows.

First, the approach requires the involvement of a qualified expert who knows the specific business process well.

Secondly, in addition to the business process model, it is necessary to construct life cycle models for artifacts, and then establish constraints by comparing the sequence of process actions and the sequence of life cycle operations.

In the second case, a log is used to identify conditions and limitations for the execution of business process actions. It contains a description of the events that occurred while the process was running.

The log is formed by the process management system. Log events are characterized by a set of attributes and their values. Attributes of events are also attributes of artifacts. For example, they can contain: the name and division of the performer; the name of the product that the business process creates; name of equipment, etc.

The disadvantage of the second method is that the log contains incomplete information about the business process. However, the log is constantly updated with new events as the business process proceeds. This allows you to perform both the task of extracting knowledge, and the task of expanding the knowledge base.

With the constant expansion of the knowledge base, the task of integrating knowledge is also performed in order to ensure the adequacy of existing models.

In general, with the automated building of the knowledge base of the business process management system, it is advisable to combine both methods. General dependencies are formed by an expert, and more detailed ones are identified on the basis of an analysis of the event log.

However, if the log contains detailed information about the activities of the process, you can limit yourself to identifying the context based on the analysis of the log.

In Fig. 1 the cycle of knowledge extracting, expanding and integrating is presented.

In this cycle, the log is viewed as the current "photo" of the workflow of the process. This sequence changes according to the current context using causal dependencies.

Such dependencies can be changed by knowledge workers. However, then, after analyzing the logs, the found altered dependencies are included in the knowledge base.

Then these cause-effect relationships can be included in the business process model.

In order to implement this cycle, we propose a logical-probabilistic approach to the extraction of cause-effect relationships in the field of process management. The proposed approach uses the properties of causal dependencies for the knowledge-intensive business process (4), as well as the properties of the BP log.

We can choose one of two possible categories of uncertainty when we construct a logical-probabilistic description of the causality of a knowledge-intensive business process.

– uncertainty about empirical facts;

– uncertainty about logical facts.

Empirical facts are characterized primarily by information and cognitive components. First, they contain information about the state of objects and phenomena of the domain. Secondly, the cognitive component determines the way in which these facts are interpreted.

Empirical facts in the logs are represented by a set of attribute-property pairs for each event, and therefore we will interpret these facts using the mathematical apparatus of predicate logic.

The predicate R_i interprets each event in this way:

$$R_i (a_1,...,a_j,...a_J) = \bigwedge_j (a_j = v_{i,k}), \qquad (5)$$

where a_j – attribute of i-event; $v_{i,k}$ – is the value of the attribute a_j.

Logical facts determine the relationship between empirical facts. Logical facts are represented in the log in the form of sequences of events corresponding to the sequence of actions of the business process. The information system of management of business processes records in a log a trace of events separately for each copy of the business process. Therefore, the logical facts for the process are of a probabilistic nature.

From expression (4) we can see that causal dependencies for a knowledge-intensive business process contain deterministic and probabilistic components.

For each pair of events $i, i+1$, we will also distinguish two components: the deterministic constraint and the probabilistic condition. The constraint $c_{i,i+1}$ is represented by a predicate that links the unchanged attributes $a_{i,j}, a_{i+1,j}$ of a pair of successive events. The first event describes the state of the context in which the action is started, and the second describes the state of the context after the action is completed:

$$R^{*}_{i,i+1}(a_{i,1},....a_{i,J},a_{i+1,1},...a_{i+1,J}) = \\ \bigwedge_j (a_{i,j} = v_{ik} \mid \forall i = \overline{1, I-1} \; a_{i,j} = a_{i+1,j}). \tag{6}$$

Using the representation of constraints in the form of a predicate (6), we can find subsets of attributes that do not change for the selected action on all log paths. This means that the selected action is performed with unchanged values of the artifact attributes in all instances of the business process:

$$a_{i,j} = v_{ik}.$$

If we order such subsets of attributes by run-time, we can obtain a set of sequences of unchanged attribute pairs for all possible variants of executing a business process. This sequence describes an invariant part of the context that does not depend on possible actions of the BP. Therefore, we can treat such sequences as static context constraints.

Each action is represented in the log by a pair of events: $i, i+1$.

Therefore, we calculate the probability of a condition l_{i+1} for performing an action using:

– the probability $p(c_{i,i+1})$ of restriction for these events;

– the probability $p(r_{i,i+1})$ that an event is represented by a set of attributes that ensure the truth of the predicate $R^{**}_{i,i+1}$ (Fig. 2):

$$p(l_{i+1}, c_{i,i+1}, r_{i,i+1}) = \\ = p(c_{i,i+1})p(r_{i,i+1})p(l_{i+1} \mid c_{i,i+1}, r_{i,i+1}). \tag{7}$$
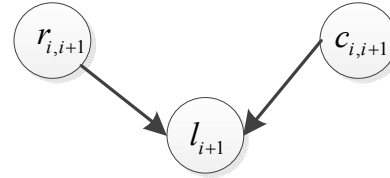

Fig. 2. Probabilistic dependencies

The predicate $R^{**}_{i,i+1}$ will be true only for pairs of event attributes a_i, a_{i+1}, which change their value after the execution of the action $i+1$ of the business process:

$$R^{**}_{i,i+1}(a_{i,1},....a_{i,J},a_{i+1,1},...a_{i+1,J}) = \\ \bigwedge_j (a_{i,j} = v_{ik} \mid \forall i = \overline{1, I-1} \; a_{i,j} \neq a_{i+1,j}). \tag{8}$$

Conclusions

The article considers the problem of representation of causal dependencies in modeling and management of business processes.

Our contribution is as follows. A logical-probabilistic model of cause-effect relationships is proposed. Each such dependency links a pair of log events that describe the execution of a single action of the business process. The model connects the logical description of the context state before and after the completion of each action of the business process, as well as a logical description of the constraints on the process actions and a probabilistic description of the conditions for performing these actions.

The proposed model can be used to solve problems of extracting, expanding and integrating knowledge based on the analysis of business process logs.

REFERENCES

1. Vom Brocke, J., Rosemann, M. (2015), *Handbook on Business Process Management 1. Introduction, Methods, and Information Systems*, Springer-Verlag Berlin Heidelberg, 709 p.
2. Gronau, N. (2012). *Modeling and Analyzing knowledge intensive business processes with KMDL: Comprehensive insights into theory and practice (English)*, Gito, 522 p.
3. El-Den J.A. (2009), *Tacit knowledge externalization among geographically distributed small groups*, PhD Thesis's, University of Technology, Sydney, Australia, 323 p.
4. Warren, P., Kings, N., Thurlow, I., Davies, J., Brger, T., Simperl, E., Ruiz, C., G´omez-P´erez, J., Ermolayev, V., Ghani, R., Tilly, M., Bsser, T. and Imtiaz, A (2009), "Improving knowledge worker productivity the active approach", *BT Technology Journal*, No. 26, pp.165-176.
5. Dong, X. L., Gabrilovich, E., Heitz, G., Horn, W., Murphy, K., Sun, S. and Zhang, W. (2014), "From data fusion to knowledge fusion", *International Conference on Very Large Data Bases (VLDB)*.

6. B. Min, R. Grishman, L. Wan, C. Wang, and D. Gondek. (2013). Distant supervision for relation extraction with an incomplete knowledge base. *In Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL).*

7. Murphy, K. (2013), "From big data to big knowledge", *ACM International Conference on Conference on Information and Knowledge Management (CIKM),* New York, NY, USA.

8. Chen, Y. and Wang, D.Z. (2014), "Knowledge expansion over probabilistic knowledge bases", *ACM SIGMOD International Conference on Management of Data (SIGMOD).*

9. Niu, F., Zhang, C., R´e, C. and Shavlik, J.W. (2012), "Elementary: Large-scale knowledge-base construction via machine learning and statistical inference", *Int. J. Semantic Web Inf. Syst.*

10. Pearl, J. (2000), *Causality: models, reasoning, and inference*, Cambridge University Press, Cambridge.

11. Alon N. and Spencer J.H. (2008), *The Probabilistic Method,* Wiley-Interscience, third edition.

12. Friedman, N., Getoor, L., Koller, D. and Pfefier, A. (1999), "Learning probabilistic relational models", *International Joint Conference on Artificial Intelligence (IJCAI).*

13. Koller, D., Friedman, N. (2009), *Probabilistic Graphical Models: Principles and Techniques*, MIT Press.

14. Ritter, J., Li, A., Jurafsky, D. (2014), "Inferring user preferences by probabilistic logical reasoning over social networks", *arXiv preprint arXiv:1411.2679*, available at: https://arxiv.org/abs/1411.2679 (last accessed March 01, 2018).

15. Bröcheler, M., Mihalkova, L. and Getoor, L. (2012), "Probabilistic similarity logic. Computing Research Repository", *abs/1203.3469,* available at: http://arxiv.org/abs/1203.3469 (last accessed March 01, 2018).

Логіко-ймовірнісна модель причинно-слідчих зв'язків між подіями журналу системи процесного управління

О.В. Чала

Предметом вивчення в статті є процеси виявлення знань у формі причинно-наслідкових залежностей на основі аналізу логу подій бізнес-процесу. **Метою** статті є розробка логіко - ймовірнісної моделі причинно-слідчих зв'язків між парами подій журналу, що описують виконання дій бізнес-процесу для підтримки рішення задач автоматизованого побудови бази знань системи процесного управління. **Завдання:** Виділити контекстні умови та обмеження на виконання дій бізнес-процесу шляхом аналізу лога подій цього процесу; розробити підхід до виділення ймовірнісної та логічної складової причинно-слідчих залежностей; розробити логіко-ймовірнісну модель каузальних залежностей Використовуваними **методами** є: методи побудови предикатних моделей; байєсовські підходи до побудови ймовірнісних моделей. Отримані наступні **результати**. Формалізовано клас причинно-наслідкових залежностей для знаннямних бізнес-процесів. Такі залежності можуть враховувати неформалізоване знання про бізнес-процес. В межах даного класу виділені: прогнозний опис стану контексту на основі інформації про значення атрибутів подій журналу; контекстні обмеження на виконання дій бізнес-процесу; ймовірнісні умови реалізації дії бізнес-процесу. **Висновки.** Наукова новизна отриманих результатів полягає в наступному: запропонована логіко-ймовірнісна модель причинно-слідчих залежностей між парами подій журналу, що описують виконання дій бізнес-процесу. Модель зв'язує логічне опис стану контексту до і після завершення кожного дії бізнес-процесу, а також логічний опис обмежень на дії процесу та ймовірнісний опис умов для виконання цих дій. У практичному аспекті модель забезпечує можливість вирішення завдань вилучення, поповнення та інтеграції знань на основі аналізу логів бізнес-процесів.

Ключові слова: причинно-наслідкові залежності; знання; база знань; бізнес-процес; потік робіт; журнал подій.

Логико-вероятностная модель причинно-следственных зависимостей между событиями журнала системы процессного управления

О.В. Чалая

Предметом изучения в статье являются процессы выявления знаний в форме причинно-следственных зависимостей на основе анализа лога бизнес-процесса. **Целью** является разработка логико-вероятностной модели причинно-следственных зависимостей между парами событий лога, описывающими выполнение действия бизнес-процесса для поддержки решения задачи автоматизированного построения базы знаний системы процессного управления. **Задачи:** Выделить контекстные условия и ограничения на выполнение действий бизнес-процесса, которые могут быть получены в результате анализа лога; разработать подход к выделению вероятностной и логической составляющей причинно-следственных зависимостей; разработать вероятностно-логическую модель каузальных зависимостей Используемыми **методами** являются: методы построения предикатных моделей; байесовские методы построения вероятностных моделей. Получены следующие **результаты.** Формализован класс причинно-следственных зависимостей для знаниеемких бизнес-процессов. Такие зависимости могут учитывать неформализованное знание о бизнес-процессе. В рамках данного класса выделены: предикатное описание состояния контекста на основе информации о значениях атрибутов событий лога; контекстные ограничения на выполнение действий бизнес-процесса; вероятностные условия реализации действий бизнес-процесса. **Выводы.** Научная новизна полученных результатов состоит в следующем: предложена логико-вероятностная модель причинно-следственных зависимостей между парами событий лога, описывающими выполнение действия бизнес-процесса. Модель связывает логическое описание состояния контекста до и после завершения каждого действия бизнес-процесса, а также логическое описание ограничений на действия процесса и вероятностное описание условий для выполнения этих действий. В практическом аспекте модель обеспечивает возможность решения задач извлечения и пополнения и интеграции знаний на основе анализа логов бизнес-процессов.

Ключевые слова: причинно-следственные зависимости; знания; база знаний; бизнес-процесс; поток работ; журнал событий.