Business Process Simulation for Predictions

With focus on decision mining and execution time of tasks

Milan Pospisil
Department of Information Systems
BUT, Faculty of Information Technology
Brno, Czech Republic
ipospisil@fit.vutbr.cz

Tomas Hruska
Department of Information Systems
BUT, Faculty of Information Technology
Brno, Czech Republic
hruska@fit.vutbr.cz

***Abstract*—Simulation can be used for analysis, prediction and optimization of business processes. But models often differ from reality. Data mining techniques can be used for improving these models based on observations of process and resource behavior from detailed event logs. More accurate process models can be used not only for analysis and optimization, but for prediction and recommendation. This paper focuses on decision mining and the duration of tasks in conjunction with personal performance based on case data, workload, and other factors. Some existing ideas are an improvement and others are new. Part of the research was validated on real data.**

Keywords-business process simulation; business process intelligence; data mining; process mining; prediction; optimizatio; recomendation

I. INTRODUCTION

Classic simulation can be used for the analysis of business processes. We can try many scenarios, measure the effects, and then decide on the optimal process settings. For example, we can redesign the process, change resource allocation, and search for the most optimal configuration with respect to requirements (price, effectiveness, customer satisfaction, etc.). Or the current process can be tested for how many cases it can handle.

Nowadays, these models are often built manually, which is error-prone, and time consuming; the main drawback of this approach is that it cannot be used for operational decision support, but only for strategic decisions. This is because classic simulation models have several simplifications – probability the routing and statistical distribution of execution time of tasks. These models are sufficient in long-term simulation (usable for analysis), because simulation parameters are the result of long running processes. But, operational decision support needs short-term simulation. In this situation, we know the running and incoming cases, and the actual resource allocation. Therefore, actual running processes can differ from long measured processes. For example, task A needs to be done and there is a standard execution time of about 30 minutes; but, we have allocated a skilled resource and it is able to execute it under 20 minutes. These, and more problems need to be solved to obtain the simulation model for operational decision support.

Predictions, recommendations, and dynamic optimizations could be accomplished by operational simulation. The system can warn us, that some cases will be probably late. Then some different scenarios can be simulated and evaluated, then the system can recommend us actions and provide dynamic optimization of current running cases – for example; give extra resources from non-critical case to critical, or use a different sub-process – when we have a slower / cheaper version or faster but more expensive.

This work deals with the building of simulation models for operational decision support using data mining, because there is need to find deeper dependencies.

The paper is organized as follows. Related work is described in Section 2. Section 3 reveals the problem with classical simulation and answers the question, why it can not be used for operational decisions. Decision mining in Section 4 advances classic simulation. It is based on current research, but new improving ideas are sketched. Section 5 is about predicting execution time of cases and goes beyond current research in this area. Section 6 compares prediction using simulation and the standard approach by classification (or regression). At the end, Section 7 concludes whole paper.

Our work extends current research of simulation for operational decisions and new ideas are described, some new, some inspired by other works described below. Emphasis is placed on better decision mining and prediction of time execution of task with conjunction of personal performance based on case data, workload, and other factors.

II. RELATED WORK

Data mining techniques can be used in Business Process Management. This new area was called Process Mining [3, 6, 12, 13, 14]. It was based on analysis of information from event logs, that were produced by business processes. Process discovery is one of the methods and it is able to find a process model from an unknown process using many sequence examples of tasks.

 ISBN: 978-1-61208-223-3

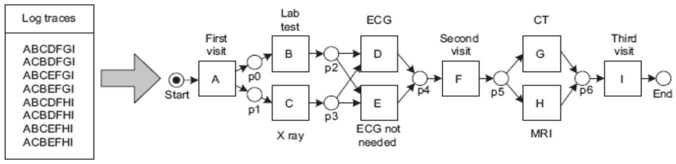

Figure 1. process discovery. We are able to discover a process model from log. The discovered process model must be able to replay most log traces.

A process log (figure 1) contains a sequence of tasks and we are able to discover what process model fits that log. Many algorithms are available for that nowadays and they were successfully used in practice. Discovered process models can be used for simulation, even if the model is not explicitly given or it is not usable (too low level detail, not all paths are described, and so on).

Different techniques are focused on performance analysis [4] (figure 2), where influential factors of Key Performance Indicators (KPI) are investigated.

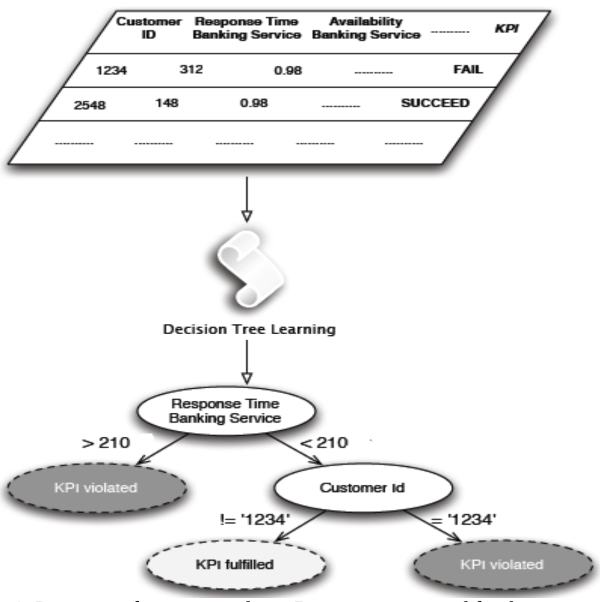

Figure 2. Process performance analysis. Decision tree is used for discovering factors that leads to KPI violation. We can see that KPI is violated when response time of banking service is larger than 210.

The table above (Figure 2) shows data that was collected by running cases (every row is one case). The target value is Key Performance Indicator (KPI) and it reports, if the case was executed right or not (it could be time, quality, or anything else, it depends on process manager). Based on these provided attributes, the target value (KPI) is predicted by a decision tree. Although we used the term 'prediction', this form of decision tree is usable only for analysis of historic data, not for real-time monitoring. We will discus it later. But the tree can tell us what combination of factors lead to KPI violation. For example, we can see, that if the response time of a banking service is higher than 210, the KPI is always violated. And also, cases with customer id 1234 has a problem with KPI too. Performance analysis can be valuable for managers because they can discover and focus on critical factors of processes.

Other work focused on the prediction of execution time [1, 2, 11] using classifiers [1, 2] or process discovery with time information [11]. Note that [1, 2] used similar techniques as [4], not for analysis, but for prediction. While [11] can be used only for time estimation, [1, 2] can predict other things like some events.

Work in papers [1, 2] is based on classifiers. Running cases can produce much usable data. For example, time execution of tasks (start, end) or some data passed from task to task. This information is written into the table - multiple execution of same task in loop is written only twice – first and last occurence. Then some classifier (neural network, decision tree, regression tree…) can be used for the prediction of the target value (total execution time of case or some other quality atribute). Authors test that method in some industrial applications and results were promising. This method will be discussed in Section 6.

Rozinat et al. [5] and Rozinat et al. [10] introduced the idea of building operational simulation models using Process Mining techniques as we described in the introduction. These methods were based mainly on process discovery and decision mining.

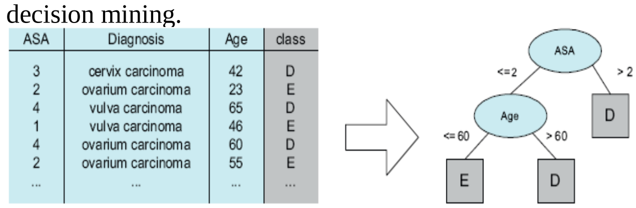

Figure 3. Decision mining. We can discover decision rules in routing points (OR-split nodes). Class attribute represents next task in running process. Decision tree is used for rule discovery.

Decision mining (Figure 3) enhances process discovery with decision rules. It is good to know the process model, but it is useless for simulation, when we do not know what task (in OR node) will be next and routing by percent (which is used in classic simulation models) is not sufficient for operational decisions.

Additional work deals mainly with resource modelling problems [7, 8, 9], which is now a topic of interest, because resources are one of the hardest things to simulate. They try to discover how to simulate resources and what factors influences their productivity – for example it was discovered (from some industrial experiments), that people tend to work faster when there is lot work to do – it is common knowledge, but we need some methods to compute it for every particular person (some people work more at a constant speed, some do not).

III. BUILDING SIMULATION MODELS

Imagine a typical example process (Figure 4) of handling warranties. The process model can be taken from a system, or discovered by Process Mining [5]. The first item is received and then checked for more information and the warranty. Then, a decision is made: the repair process is canceled (warranty not applicable) or send to repair. The repair is either basic or advanced. At last, the item is returned back to customer.

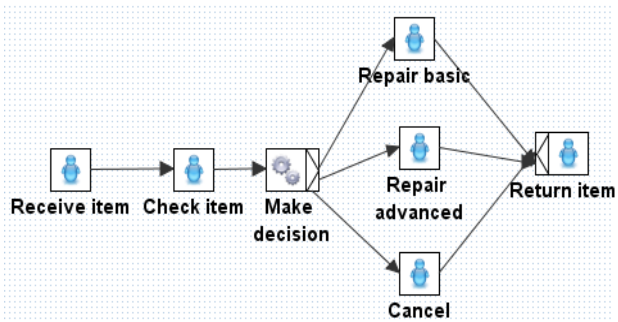

Figure 4. Example process model of handling warranties.

The typical simulation model would provide us a probability of routing (for example 20% of cases are cancelled, 60% sent to basic repair and 20% sent to advanced repair) and an average execution time of tasks with a standard deviation of some distribution (Repair basic takes 1 hour with 0.5 hour deviation, Repair advanced takes 4 hours with 2 hours deviation). We can use this simulation model for long-term simulation over several weeks, but not for short-term simulation over several days. Lets say, we have some running process instance. We are now behind the Check item and we have filled in information about the item (item type, damage type, etc). We want to predict the total execution time. Then routing probabilities are not sufficient, because we now know what type of repair it is and the differences between basic and advanced repair is significant.

IV. Decision Mining

Decision mining can be used to discover what influences the decision of routing. Of course, we can also take that decision from the system (if available). But there is a catch. Decision expression would be probably simple and based on a few attributes known at the time of the decision (filled by a human based on the previous several attributes). In our process, decision rule could look like that – if RepairType is 'Cancelled' – go to Cancel, or if RepairType is 'Basic' – go to Repair basic, if 'Advanced – go to Repair advanced. That rule can be useless if one has not yet filled the attribute RepairType by the time of the decision. But we could have filled some important attributes – for example at the middle of the task Check item. Based on the provided information, one can make a better prediction of the next steps than the basic probability described above.

This can be solved as a classification problem. We have a table of attributes needed for a decision (Figure 5) and we want to predict the next step in the process – if the item will be repaired as basic, advanced, or cancelled.

Item age	Item type	Item type name	Damage	Repair type
3,1	Notebook	Notebook Acer	Broken monitor	Canceled
2,1	Notebook	Notebook HP	Broken HD	Basic
1,2	Mouse	Genius	Unknown	Basic
1,6	Notebook	Notebook Acer	Broken matherboard	Advanced
2,3	Mouse	Logitech	Broken glass	Basic
2,7	Notebook	Notebook Acer	Broken monitor	Canceled
1,3	Mobile	Errickson	Damaged keyboard	Basic
..	..	..	..	..

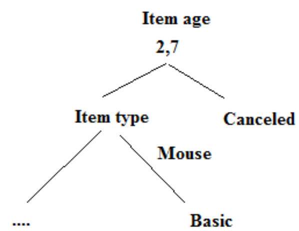

Figure 5. Decision mining. The data in the table is used to predict the next step in the process represented by the attribute Repair Type. The decision tree classifier is used here. Decision tree can be used for prediction and also for analysis.

A. Extension

Similar approach was described by Rozinat et al. [5], but there were several unsolved problems. First, the decision tree with these provided attributes will find this rule – if Repair type is 'Basic' – go to "Basic", etc. (we have discussed it above). It is not a mistake of the algorithm, the decision tree will simply find the attribute that decides everything with 100% precision. Of course, we can delete this attribute from the list. But still, that type of decision mining is good for analysis of historical data, but not for prediction, where not all attributes are provided at runtime. The most important attribute Repair type is filled last, so our prediction will work very poorly.

But, note that the Repair type attribute is filled by a human based on previous provided attributes, so we can still predict the next step even if we do not have all the attributes. The more important attributes we have, the Berger the prediction will work. Most classificators do not work so well on sparse data, so several classificators have to be used at different milestones of the process – a similar problem was solved in [1, 2]. That means we suppose some order of attributes and we build several classifiers using more and more attributes according to that supposed order; from Figure 5, we can see, that if the item type is mouse, it will be probably sent to basic repair. The item type will be one of the first filled attributes.

Another important thing is that many times, we cannot determine one precise decision (mostly when not all attributes are available). In that situation, more decisions have to be provided with some order of probability. A classic decision tree could give us only one final decision, but this can be solved quite easy by providing probabilities of classes in every node (mostly in leaves). Or, we can use a more advanced classifier like a neural network. A neural network can have output neurons corresponding to the next following available task – in our example, we can have three output neurons, first will give us fitness (0..1) of the decision of Repair basic, second about Repair advanced and third about Cancel. A neural network also has some disadvantages. It works as a black box (opposite as a decision tree) and it needs more data to train. Also Bayesian classification could be used, if the attributes are not so dependent. That type of classification will work better than a neural network when not as much data is provided.

V. TIME EXECUTION OF TASKS

In [5], task execution times were modeled classically by a distribution with mean and deviation. This is not sufficient for short-term simulation. Task execution times can depend on several things. We will describe a new and better approach to this problem.

Some people work faster, some slower. Some are good at one task, some at another task. So resource information influences execution time. We have an example in Table I. From this table, we can deduct, that John is faster than Karl (he is also able to repair advanced items, but that information can be also found in process definition). More dependencies could be found (but not at Table I., we do not have so much space), for example mice are repaired faster than notebooks, etc. We can predict task execution time also by classification (Table I.).

These techniques and attributes can be used for the execution time of tasks and for decision mining. In fact, decisions can be influenced by resource, who is responsible for them, weather, time of day, etc. It is the reason, why these approaches have to be done semi-automatically. The process designer has to decide what attributes are needed for what decision. Also, there is an option to automatically find important data. But still, a human has to provide all the important data to system.

Additional works about resource modeling in simulation are in [7, 8, 9]. Human productivity can be influenced by many factors – by weather, day of week, time of day (especially after lunch or dinner). Another important factor is workload [9]. People tend to work faster, when there is full work queue, but not so long. After some time (it depends on the individual) productivity fails. There is a place for future research – resources have several attributes and those attributes can be measured. It could be productivity variance, ability to increase performance, when there is too much work, endurance to illness (people are usually weaker when the weather changes or during a flu epidemic). The question is how to deal with that information, because some tasks can be more influenced by resource productivity, and some tasks not. Mainly stereotyped work or machine operations – These tasks will probably be influenced more by data parameters of case (in our example - item type, damage type, etc.) than personal productivity.

We can put that information into the table and provide the same data mining techniques as in decision mining. It could work, but it requires more data to learn than the classifier. So, there is a space for research how to accomplish that with measuring resource parameters and without the need to put all this information into the classifier.

VI. SIMULATION AND CLASSIFICATION

Grigori et al. [1][2] uses classification to predict total execution time of cases (and potentially some events) using all important case parameters. It works similar to our approach, with distinction, that process model is not used. All process attributes are in one table and the final predicted value is the total execution time of cases.

In our simple example, this method does not differ from the simulation model, because we also used classification for decisions and predictions of execution time. It is because our model is too simple and we are using almost all attributes to predict both decision and time. In a more complicated model, not all attributes will be needed for classification.

What are advantages and disadvantages of these two approaches? Clearly, when there is no predictable process model behavior, or even no model available, classification based on all attributes will be better.

First, the simulation model will be better, when something changes – for instance, a faster machine, a change in the process model will be much worse for the classifier. We do not have to learn the whole classifier again, but we have to deal only with one change – the data needed to predict the execution task of the new machine can be provided from an expert, for example. Second, short-term simulation can give us what-if analysis. We can simulate several situations (with different resource allocation) and the system can choose the best solution.

Third, and maybe the most important advantage, is that this method is contextual. Prediction based on classification of all attributes without a model is bad for resource modeling. If case attributes show that the case will be in time that does not have to be true, because we can be short of workers. It is hard to give information to the classifier about the resource workload. But in the simulation model, we know what resources are available, what tasks can be accomplished by what resources, etc.

VII. CONCLUSION AND FUTURE WORK

We proposed an improvement of the existing simulation model for operational decisions. Improvement was based on better decision mining and mainly on the execution time of tasks. This work is now in progress, so ideas are described and compared with some other approaches.

We did some industry experiments and results were quite good. We were able to predict execution time 40% better than methods that do not také into account parameters of cases and were based only on global mean and deviation of all execution time of cases. So we belive, that methods are able to improve prediction in some industrial companies.

Next research could be focused on industrial experiments and dealing with resources – we need to meause resource productivity at particular tasks.

We believe, this type of simulation will be able to support operational decisions and predict execution times of cases. But more work need to be done, mainly at the field of resource modeling.

ACKNOWLEDGMENT

This research was supported by the grants of MPO Czech Republic MPO TIP FR-TI3 039, by the grant FIT-S-10-2 and by the research plan no. MSM0021630528.

REFERENCES

[1] D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, and M.C. Shan, Business Process Intelligence, Computers in Industry, Volume 53, Issue 3, Process / Workflow Mining, April 2004, Pages 321-343, ISSN 0166-3615, DOI: 10.1016/j.compind.2003.10.007.

[2] D. Grigori, F. Casati, U. Dayal, and M.C. Shan, Improving Business Process Quality through Exception

Understanding, Prediction, and Prevention

Proceedings of the 27th VLDB Conference,Roma, Italy, 2001, 1-55860-804-4

[3] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, and H.M.W. Verbeek, Business process mining: An industrial application, Information Systems, Volume 32, Issue 5, July 2007, Pages 713-732, ISSN 0306-4379, DOI: 10.1016/j.is.2006.05.003.

[4] B. Wetzstein, P. Leitner, F. Rosenberg, I. Brandic, S. Dustdar, F. Leymann, "Monitoring and Analyzing Influential Factors of Business Process Performance," *Enterprise Distributed Object Computing Conference, 2009. EDOC '09. IEEE International, pp. 141-150, 1-4 Sept. 2009* doi: 10.1109/EDOC.2009.18

[5] Anne Rozinat, R.S. Mans, M. Song, and W.M.P. van der Aalst, Discovering simulation models, Information Systems, Volume 34, Issue 3, May 2009, Pages 305-327, ISSN 030

[6] Minseok Song and W.M.P. van der Aalst, Towards comprehensive support for organizational mining, Decision Support Systems, Volume 46, Issue 1, December 2008, Pages 300-317, ISSN 0167-9236, DOI: 10.1016/j.dss.2008.07.002.

[7] W.M.P. Van der Aalst, Business Process Simulation Revisited, 2010, ISSN: 1865-1348

[8] J. Nakatumba, A. Rozinat, and N. Russell, Business Process Simulation: How to get it right, 2010,Springer-Verlag,doi=10.1.1.151.834

[9] J. Nakatumba and W.M.P.V.D. Aalst, "Analyzing Resource Behavior Using Process Mining", in Proc. Business Process Management Workshops, 2009, pp. 69-80.

[10] Anne Rozinat, M.T. Wynn, W.M.P. van der Aalst, A.H.M. ter Hofstede, and C.J. Fidge, Workflow simulation for operational decision support, Data & Knowledge Engineering, Volume 68, Issue 9, Sixth International Conference on Business Process Management (BPM 2008) - Five selected and extended papers, September 2009, Pages 834-850, ISSN 0169-023X, DOI: 10.1016/j.datak.2009.02.014.

[11] W.M.P. van der Aalst, M.H. Schonenberg, and M. Song, Time prediction based on process mining, Information Systems, Volume 36, Issue 2, Special Issue: Semantic Integration of Data, Multimedia, and Services, April 2011, Pages 450-475, ISSN 0306-4379, DOI: 10.1016/j.is.2010.09.001.

[12] W. M. P. van der Aalst, and A. J. M. M. Weijters, Process mining: a research agenda, Computers in Industry, Volume 53, Issue 3, Process / Workflow Mining, April 2004, Pages 231-244, ISSN 0166-3615, DOI: 10.1016/j.compind.2003.10.001.

[13] W. M. P. van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters, Workflow mining: A survey of issues and approaches, Data & Knowledge Engineering, Volume 47, Issue 2, November 2003, Pages 237-267, ISSN 0169-023X, DOI: 10.1016/S0169-023X(03)00066-1.

[14] W. M. P. van der Aalst, Process Mining, Berlin, Heidelberg 2011, ISBN 978-3-642-19344-6

TABLE I.

Item age	Item type	Item type name	Damage	Repair type	Resource	Time
3,1	Notebook	Notebook Acer	Broken monitor	Canceled	Mary	1:00
2,1	Notebook	Notebook HP	Broken HD	Basic	Karl	2:00
1,2	Mouse	Genius	Unknown	Basic	John	1:30
1,6	Notebook	Notebook Acer	Broken matherboard	Advanced	John	10:50
2,3	Mouse	Logitech	Broken glass	Basic	Karl	3:30
2,7	Notebook	Notebook Acer	Broken monitor	Canceled	Mary	1:10
1,3	Mobile	Errickson	Damaged keyboard	Basic	Karl	1:50
..	..	..	..	..		

Execution times of tasks. Six attributes are used to predict execution time.

