Analysing Resource Performance and its Application in Company

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

Vojtěch Mates
Department of Information Systems
BUT, Faculty of Information Technology
Brno, Czech Republic
imates@fit.vutbr.cz

Tomáš Hruška
IT4Innovations Centre of Excellence
BUT, Faculty of Information Technology
Brno, Czech Republic
hruska@fit.vutbr.cz

***Abstract*—Performance analysis of resources is important part of process optimizing. It is useful to analyze performance (execution time) for particular tasks, time deviation, error rate and its improvement during time. It is also possible to use the analyzing of task execution time related to resource for different purposes. For example, it is possible to change process definition according to the results, make predictions using short-term simulation, or use it only as analysis of performance properties of resource. This paper focuses on analyzing resource properties and then makes overview of its applications. Some of these methods will be evaluated on real data in manufacturing company.**

Keywords-resource performance analysis; business process simulation; business process intelligence; data mining; process mining; prediction; optimization; recomendation

I. INTRODUCTION

Current business process management system stores a lot of information about processes, data flow, resources, and execution time. The information is very valuable and it is possible to use it for analysis. Related work deals with process model discovery. There exists also related work dealing with mining decision rules or organizational models of companies. There is also work based on resource perspectives. Nevertheless, resources are an important part of business processes. This paper deals with resource performance analysis and its usage.

It is possible to analyze different performance perspectives. Basic properties can be execution time of task, its deviation, and error rate. However, other important properties could be analyzed – for example ability to raise performance during pressure or ability to quickly adapt to new task (set up time). However, simple look at execution time of task is not enough, because this execution time is not only based on resource productivity but also on another attributes. There is also an influence done by process change, e.g. new, faster machines or cooperation of multiple resources.

The information is useful to improve our process model. This paper describes several approaches related to mentioned topic. The first area of application is analyzing the current resource performance properties and improvement over time. Then manager can look at these data and make some assumptions – better resources will get more money, better resources can show other resources how to perform particular task much better.

Second area of application is about changing process definition according to resource properties. For example, more experienced resource does not need so many checkpoints as less experienced resource does. This can improve performance while technological logic process remains the same.

Third area of application deals with process prediction. This part is based on short-term prediction that uses simulation model built semi-automatically by process mining. These simulation models need more information about resources, because performance could differ significantly between best and slowest worker.

Next possible application is about allocation of resources. This can be static, e.g. manager can decide what resource should be assigned to specific task taking into account properties of resource. This method corresponds with second application – changing process definition. The other approach corresponds to short-term simulation and recommendations. System can simulate multiple scenarios of allocation of resources to task.

This paper is organized as follows: second chapter concludes related work; the third chapter is about analysis of resource properties. Forth chapter is about its application, fifth chapter evaluates some previous methods in real manufactory and last chapter is conclusion and future challenges.

II. RELATED WORK

An interest in process mining raised in last decade. Process mining is based on several perspectives. First is process discovery [3, 12, 13, 14]. Process discovery is able to analyze process event log. Event log contains events that were executed during run of process. Every event corresponds to some task and some case. Events have also start and end execution time (they must be ordered at least). Using this information, process discovery is able to find process model of the task sequence (see figure 1).

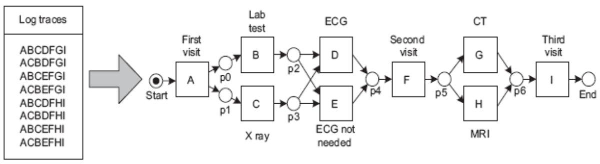

Figure 1. Process discovery [5]. It is possible to discover a process model from logs. The discovered process model must be able to replay most of log traces.

There is also research dealing with mining decision in routing points (OR split) [5, 7, 14, 15]. Using case attributes, decision rules could be discovered as classification problem.

ASA	Diagnosis	Age	class
3	cervix carcinoma	42	D
2	ovarium carcinoma	23	E
4	vulva carcinoma	65	D
1	vulva carcinoma	46	E
4	ovarium carcinoma	60	D
2	ovarium carcinoma	55	E
...	...	...	...

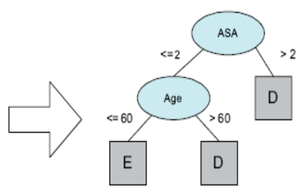

Figure 2. Decision rule discovery [5]. It is possible to discover decision rules from log. Target attribute is class, which corresponds to next task in process model.

Using previous methods, simulation model can be built [5, 7, 10, 15]. This simulation model (figure 3) can be used for either analysis, or short-term prediction and operational decision support [15].

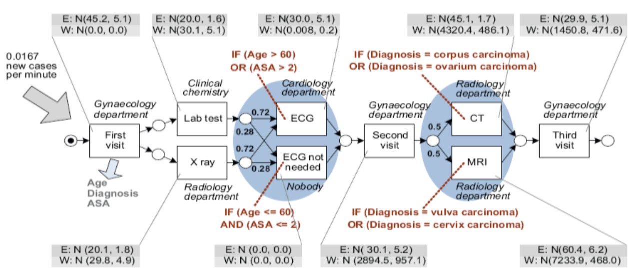

Figure 3. Simulation model [5]. Classic simulation model is enhanced by decision rules. Decision rules can make our routing probabilities more precise, because they depend on case attributes.

Resources are also point of interest. Mates [1] described Resource Dynamic Profiles, which he used for modification of process taking account resource attributes from their Dynamic Profiles. Process mining group also had some research in resources. For example, Song [6] discovered organizational model from process log. Nakatumba [8, 9] examined some resource properties as time availability and ability to increase performance when in pressure.

Another work in Business Process Prediction was from Grigori [2, 16]. She used classification over all case attributes for business process prediction. However, that work did not focus much on resource. Wetzstein [4] did related work where bottlenecks were identified using similar methods (classification). That work also covered resources (resources could be bottlenecks too).

Aalst [11] discovered simulation model using different methods (transition system), but he did not take into account the resource and case attributes at all.

Related work was also done in business process simulations for operational decisions [5, 7, 10], or another methods [2, 11, 16] but only our work [15] deals directly with execution time of tasks ([5, 7, 10, 16] did not examined it at all). However, execution time of task is also dependent on resource performance and thus it is the goal of this paper.

III. Analysing of Resource Performance

It is possible to analyze multiple resource properties:

A. *Execution Time Length of Task*

This property cannot be computed only by looking into the task execution time, because the execution time can be dependent on another attributes. The task with one attribute combination could be much easier than the same task with another attribute combination. For example, assume repair process. “Repair” is one single task, but there is difference between repairing of computer mouse and the notebook. Repairing mouse is easier task than repairing notebook in most cases. In classic workflow system, resource that repaired more mice than notebooks could be considered faster than resource that repaired more notebooks. However, this is false assumption, because second resource is maybe better. There can be the reason why he takes harder repairs.

One solution could be to divide one task into several task that are more similar, but this may not be necessary. The algorithm that computes real worker productivity looks like that:

- For every worker.
- For every worker record.
- Take worker time and predicted time (taken from classifier based on given attributes
- of record).
- Worker productivity = record time length / classifier result.
- Compute average productivity from these records.

The idea of the algorithm is quite simple. Every worker record of executed task is compared to predictor, that is able to predict execution time of task based on provided record attributes (these attributes must not contain resource id). Predictor should be some classifier like Decision Tree, Neural Network, or K-Nearest-Neighbour (that predictor has good results, but it is very slow). Predictor is able to predict execution time of task independently of resource, so resource id must not be part of its input attributes. That means the classifier learns its predictions from all workers and our worker time is compared to all workers times (for only similar task parameters). The productivity ratio is computed as worker time compared to all workers times that worked on task with similar attributes.

This algorithm needs some data cleaning. First, execution times that deviate from average too much are not considered and second, if prediction of classifier is based only on records, that belong to the worker himself, the result is not taken account, because that would mean comparison of worker to himself. Of course, this information should be taken from decision tree, because decision tree should return final leaf, which is based on particular records, whereas neural network does not provide this type of information.

B. Execution Time Variation of Task

It is important to analyze not only the resource performance, but also time variation. Time variation could be the same important information to manager as performance itself. Whole planning is based mainly on variation. Slow resource with low variation and error rate could be good for planning and stable process. Fast resource with high performance could be used in situation, where time is more important than stability.

Variation could be computed using the same method as performance. We will take every worker record and compare it to classifier that predicts variance of all workers according to the task attributes.

C. Error Rate

Error rate could be computed by looking for task execution that were marked as incorrect and then computed by the same method as two previous parameters, because some case attributes could lead to more errors. Error rate have similar usage as variance of time. It brings uncertainty into process that is not desired.

D. Ability to Raise Performance in Pressure

Nakatumba [9] described method how to compute this property using linear regression. We propose different solution. It could by usable to know performance, variability and error rate, when worker is in stress. Therefore, we will compute these parameters for records that were marked as high pressure. How to detect urgent records? It highly depends on context of business process. Usually, it is possible to check work queue, deadlines and available resources. Nevertheless, the question, if the task was in high pressure is beyond that paper.

E. Resource Set up Time

Set up time is important factor that have to be taken into account. When resource is changing task, its performance could be lower than the situation where the task is repeated. This can be computed in the same way as ability to raise performance in pressure. We have to compute performance, variability and error rate the same way but only for tasks, which were changed. If the performance (or variability and error rate) is significantly worse than average properties, we know that we have computed set up time that has to be taken into account in resource allocation. This could be useful information for resource allocation planning, because we can choose resources that have good set up time ability.

F. Present, Historic and Actual properties

Present parameters (performance, variability, error rate – average, in pressure, or set-up time) can be seen also in time plane. We consider present parameters as those, which are two months long. Historic parameters could be analyzed for several months', long periods of time. Using this, we can see if the resource productivity is growing or falling. Actual properties are those, which are valid for example one last week. This reflects actual performance of resource and this could be useful information for prediction – see chapter IV.

G. Triage

Triage is term from business process reengineering. It means dividing one task into more special tasks. Our analysis could be more precise for tasks, which are more similar (in performance). This could be done semi-automatically (maybe fully automatically) by analysis techniques and manager. It is possible to use clustering methods to identify clusters (by execution time). If we found several clusters that are quite different, we can divide one task into that clusters (of course, only for analysis purpose). If those clusters are different (e.g. high distances between clusters) and have low variation (e.g. distances between items in cluster) then we could simplify the operation of analysis of performance (and variance), because we can then compute simple average of times.

IV. Getting Usage Information From Analysis

A. Rewards and Experience

Rewards are most obvious things when analyzing resource performance. It is sometimes difficult for managers to distinguish fast and slow resources only from simple analysis of execution times. Experience is also another valuable property of analysis, because when we found, that one resource has excellent performance in some combination of task attributes comparing to others, it could mean, he has a special approach that could improve performance to others. Similar method was described in [4].

B. Modification of Process Model

In some cases, it is useful to change process definition in runtime in order to adapt particular worker. This can be done by adding special rules using resource dynamic properties or switching variants of process based on results of previous observations. Adaptation of the process according to the user properties is mentioned in [16].Monitoring and analyzing behavior of resources and products is one the most important source of process improvement.

C. Prediction Based on Simulation

Our previous work [15] proposed a method for operational predictions. These predictions were based on simulation model enhanced of process mining. For example, decision rules were discovered and execution time of tasks was predicted based on task attributes and particular resource. Using that (and process model, of course), the method is able to predict business process using simulation. Quality of this depends on predictability of process itself, complexity (complex process model full of communication with web services can be barely predicted) and quality of data. Nevertheless, our experience on data has shown, that this type of prediction could be usable is our manufacture.

What is the influence of resources and their attributes in this method? Significant, some tasks are heavily depended on resource that executes it. Predicted execution time could vary between two different resources (one slow, one fast). There are two basic approaches for this problem:

- Integration with classifier (predictor),
- post modification after prediction,
- prediction with resource attribute,
- prediction without resource attribute.

1) *Integration with Classifier*

First method is based on integration with instance classifier. Instance classifier is classifier that does not return final decision (in our case – time or variation) but rather set of examples that are near to current record we are predicting. This means we need to return set of records with similar attributes. It could be accomplished by several algorithms, for example K-Nearest-Neighbour, or Regression Tree (or Regression Tree Forest). Note that these classifiers must not contain resource as input attribute. In normal situation, result will be computed by simple average of results. However, this is not our case. There are several reasons for that.

First, we want to include both results from our resource and other resources. Second, people tend to change their performance over time, so later records are more important and third is almost the same as second, task performance could change over time due to some another reasons (better machines, ..) , so later records are twice more important. Also, if we are predicting task that is changed for resource (set up time), or resource is in stress (long working queue), we could give more weight to records, that are also changed or in hurry (and opposite – lower, if this is not the case). Result average and variance could be computed by weighting records.

How to compute these weights? It strongly depends on data. Sometimes, newer records are not such important as older records (resource improvement is not so important – for example some easy monotone work). We do not know how to set those weights, our experiments shows that those numbers ranges from 1.0 to 3.0 (another resource vs. this resource, old record vs. new record, etc.). These weights must be set manually in experiments. Of course, we can use some automatic optimization like evolutionary algorithm. Nevertheless, this is beyond this paper.

2) *Post modification After Prediction*

Previous method will work well for instance based classifier. However, there are situation, where instance based classification is not available – a lot of historic data that cannot be stored or another reason like performance. Some classifiers can learn from stream of data that are then forgotten. In that situation, we need to use post modification after prediction. Predictions have to be made without resource attribute. If it is, we do not have to modify its result.

Post modification is made by applying resource attributes (performance, variation) to classification result by multiplying it. Performance and variation of resource attributes is number about 1.0 that says how much better (or worse) resource is compare to other resources. Performance 0.5 means, that resource is two times faster than average resource. We deal with variation in the same way. Low variation means, that resource performance is stable, while high variation means unstable performance results.

3) *Prediction With Resource Attribute*

We could build predictor using resource attribute. There are two types of predictor: One returns set of instance records, second do not. The result of first one could be accomplished by similar method by weighting results of returned records. Second, one is final prediction and does not need any post modification. This approach has several limitations. These limitations depend on classifier we are using. In Regression tree, the result is based only on results by one resource (there is a way from root to leaf using resource attribute, so leaf will cover only records belongs to this resource), which could be not enough. There can be lot of experience in another record especially when there are many combinations of attributes and we do not have so many records for the same attributes. Note that in our case study, execution time varies even for the same attributes, if we want most probable result and variability, we should need much more than ten records (30-50 could be enough).

4) *Prediction Without Resource Attribute*

If work does not depend on resource, we can use simply prediction without resource attribute. This could be case for some workplaces with machines that are little dependent on resource. For this purposes, it could be easier to omit resource attribute at all.

D. Allocation of Resources

Based on simulation described in previous chapter, we can predict future state of process. However, not only predict, we can also recommend some allocation rules. System knows (by analyzing resource performance) who is suitable for what work. Thus, there is space for system recommendation of resources allocation. For example, there are two work queues, one long queue with many same tasks and another with different tasks. We could choose resource that performs well in set-up time attribute. System can simulate those situations by using previous method and compare results, than recommend several good decisions. Static (long-term) allocation is also available.

V. Evaluation

We have tested some methods in real manufacturing company. First, we have tried to make deeper analysis of workers performance and we have discovered that our analysis was closer (by opinion of manager) to real performance than simple average of execution times for particular tasks. Unfortunately, validation of this method cannot be made, because no one knows the real right result, but we believe that it is more precise than simple average of execution times of task.

Second test was more interesting. We have tested four methods from previous chapter about simulation:

- Integration with classifier (predictor),
- post modification after prediction,
- prediction with resource attribute (classifier returned only rows that belong to one resource),
- prediction without resource attributes.

We have made some tests (figure 4) on workplaces divided into two sets – machine workplaces that are not so dependent on resource performance and hand workplaces, which are more dependent on performance. Result was computed as follows: prediction was compared to most simple predictor that supposes always-average value of task execution time for all records. So:

$$\text{Mean diff} = \sum \left| \text{mean} - \text{real value} \right|$$
$$\text{Predictor diff} = \sum \left| \text{predicted value} - \text{real value} \right|$$
$$\text{Ratio} = \text{Predictor diff} / \text{Mean diff}$$
$$\text{Final score} = 1 - \text{Ratio}$$

Mean and Predictor difference is computed as sum of differences over all tested examples. Mean difference is absolute value of mean and real value and Predictor difference is computed from predicted value and real value. Ratio equals ratio of predictor difference and mean difference. We turned over the Ratio, because it is more natural to see the better results as higher.

We can see (figure 4) that first method (integration with classifier) was best, as it was supposed. It has to be better than second method (post modification) because it takes account more deeper dependences (performance of resource is not so precise, because it is overall performance for task and resource could handle some task attributes better than other). Prediction without resource attribute worked well for machine workplace, because this workplace is not so much dependent on resource, but it did not work sufficient for hand workplace. Post modification after prediction is still best choice for resource-dependent workplaces if there is no historic data available and we have only predictor (neural network, regression tree with no leaf data, but only mean and variance information).

Triage (division of task into several tasks by clustering) was not tested, because there were about 18 attributes (high space dimension), high variance of execution time and one big cluster with hundreds thousands of overlapping records.

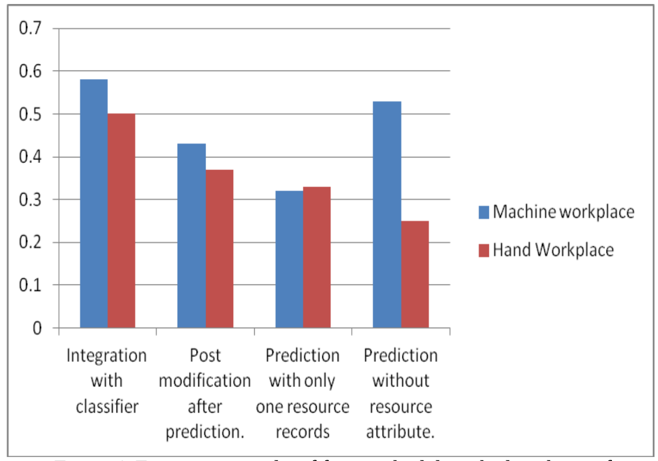

Figure 4. Experiment results of four method described in chapter four. We tested two types of workplaces, machine, that are not so dependent on resource and hand workplace, which is heavily dependent of resource.

VI. Conclusion

The paper has been focused on internal context analysis, especially on resource performance analysis. Many different approaches were discussed. The analysis of internal context, especially combination task and resources, is very important for improving performance of process and it can be used for purposes presented in paper. Evaluation of methods were tested on data covered several million records.

Our future goal is adapting planning algorithms in such way they could increase the planning flexibility and precision, which is very important for logistics purposes, and current planning algorithms do not take so much individuality of product and resource into account. Other goal that has already almost been accomplished is to provide benchmarking methods for comparing performance of workers in company, because using only standards do not reflect individuality of particular tasks. Application of results of the benchmarking can improve overall performance of the process. The paper also describes suitability of particular data mining algorithm for area of research and evaluating it on real data created by manufactory.

Acknowledgment

This research was supported by the grants of MPO Czech Republic TIP FR-TI3 039, the grant FIT-S-10-2, the research plan no. MSM0021630528 and the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] V. Mates: “Using Workflow Management System for Analysis Based on Properties of Resources”, In: DATAKON 2010 Proceedings (Ed. Petr Šaloun), Mikulov, CZ, Ostravská univerzita v Ostravě, 2010, s. 161-168, ISBN 978-80-7368-424-2

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

[2] 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.

[3] 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

[4] A. 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

[5] M. 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.

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

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

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

[9] A. 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] Pospisil, M., Hruška, T., "Business Process Simulation for Predictions" In: BUSTECH 2012 : The Second International Conference on Business Intelligence and Technology, Nice, FR, IARIA, 2012, s. 14-18, ISBN 978-1-61208-223-3

[15] 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.

[16] V. Mates, "Modification of Workflow Management System Focused on Human", Information Systems Development, , New York, US, Springer, 2011, ISBN 978-1-4419-9645-9, pp. 479-490.