

Data-driven and Context-aware Process Provisioning

(Extended Abstract)

Renuka Sindhgatta

Queensland University of Technology, Australia

1 Introduction

Business process provisioning deals with the execution of a process by providing sufficient resources (people, technology, and information), to realize its goals. To ensure effective execution of the business process, process execution logs or event logs are analyzed considering multiple perspectives: i) ordering of activities (process), ii) completion time of tasks and process (timing), iii) variances based on case attributes (case), and iv) capabilities, roles of resources and their allocation (organization). A large part of existing work has focused on these perspectives in an isolated manner. This thesis is mainly concerned with analyzing variance in resource efficiency based on case attributes and its impact on process performance, thus considering organizational, case, and timing perspective together. Further, an important dimension of operational context is considered as event log data suggests that human workers or resources with the same organizational role and capabilities can have heterogeneous efficiencies based on their operational context. In practice, experienced managers use this knowledge and consider these heterogeneous resource ¹ efficiency, when allocating tasks to meet the required process performance. The thesis is mainly concerned with answering operational questions such as, “How does resource efficiency vary with case attributes, resource behavior (that manifests as context), further impacting task allocation?”, “How do we support task allocation based on process context, for pull-based and push-based dispatching scenarios?”, and “How do we identify useful contextual information from process data?” Hence, we refer to this work as *context-aware* process provisioning.

The importance of context has been recognized by many researchers in disciplines such as mobile computing, ubiquitous computing, information retrieval, as well as business process management [1], [2]. In BPM, context has been considered as “the initial state of the process and the set of external events that affect the process instance at run time” [3]. The context can also contain important information about resource behavior (which, are external to the process execution) and yet could influence the resource and process. Early work on dynamic resource allocation, alludes to the importance of considering certain resource and case attributes such as suitability, availability, conformance and urgency [4].

¹ this work focuses on human resources only

The influence of individual contextual parameters such as workload and cooperation on resource performance have been studied [5], [6]. As several contextual parameters can influence resources and their efficiency, we learn the influence of operational context and case on resource efficiency from event logs. The key argument of this thesis is that, for an effective resource allocation, it is necessary to consider the influence of several factors related to process instance and operational context of resources in conjunction. To this end, we present a set of data-driven methods (machine learning models) relevant to different dispatching scenarios.

2 Contributions

The overall contribution of this thesis is to validate the influence of multiple factors related to case attributes and the context on resource performance, develop solutions that learn the influence of the factors from event data using supervised machine learning algorithms, and explore the possibility of mining contextual factors from additional process data sources such as textual information to support process analysis. In this thesis, in line with other definitions of context, we consider context as *exogenous knowledge potentially relevant to the execution of the process that is available at the start of the execution of the process, and that is not impacted/modified via the execution of the process*

The key contributions to context-aware resource performance analysis and allocation are as follows:

- Evaluate the influence of context and case attributes on resource performance using event logs which leads to a context model comprising resource behavior indicators such as experience, cooperation, preference [7] and other task related contextual factors.
- A recommender system that uses the context model in addition to task and resource attributes, for pull-based dispatching scenarios.
- A supervised machine learning-based method, to derive resource allocation policies, in push-based dispatching scenarios.
- A method to explore and discover contextual information that impacts performance outcome, from unstructured textual data available in communication and message logs of business processes tasks.

2.1 Heterogeneous Resource Performance

In a business process where contractual service levels have to be maintained, the time taken (completion time) to complete a task is a critical performance metric. The completion time comprises of: (a) queue waiting time in the process, and (b) the service time of the resource (time required to complete a single unit of work). The queue waiting time depends on the amount of work that exists in the system and the resources available for doing that work. The service time of a worker (or resource) is influenced by several factors. In this work multiple factors are

considered: (a) complexity of work (b) importance or priority of the work, and (c) contextual factor such as expertise level of the worker required for a work. It is observed that, *while experts have lower service time than novices for complex work and important work, they tend to have the same efficiency as novices for less important work*. The study illustrates that the efficiency of human resources in any process, depends on multiple factors that go beyond the role or availability of the workers. A simulation model is built to account for the behavior of experts and novices for varying work complexity and priority. A search based optimizer uses the simulation model to arrive at an optimal staffing. The simulation model results in different staffing solutions based on the contextual factors considered. Hence, the results presented serve as the basis for guidelines on using data-driven analysis taking into consideration the operational context when allocating task to resources.

2.2 Context-aware Resource Allocation

In a process, where tasks allocation follows a pull-based dispatching policy, the ownership of selecting the right task to work on, lies with the resource. The proposed approach involves recommendation of tasks to resources taking into consideration the context of the resource and the task. Such a decision support is useful to novice service workers. To this end, we build a context-aware recommender system (CARS). The event logs are used to infer context and the outcome of the execution. The outcome is the performance indicator defined for the task. The recommender predicts the suitability of a task for a resource, by providing a rating. Prediction is based on the assumption that resources who have similar ratings on certain tasks are likely to have similar ratings towards other tasks. The rating of a task is predicted, by identifying resources who have had similar ratings on tasks under similar context, hence considering the notion of context). A context model based on resource characteristics or behavior and task context such as ‘time of the day’ is defined. In order to conduct the evaluation, two Business Process Intelligence Challenge (BPIC) event logs are used (BPIC 2012 and BPIC 2013). Based on the identified performance outcome (completion time, quality), ratings are computed for each resource-task pair. The event logs are enriched by inferring contextual factors such as resource workload, experience and preference. The rating of a task for a resource is predicted and compared with the actual rating. Prediction with and without contextual information are compared to evaluate the importance of context in deriving a suitable rating of a resource for a task. Results indicate that the mean absolute error (MAE) considering contextual factors is lower than MAE without considering context². Further, the improvement in MAE is studied on individual contextual factors. The factors that influence rating vary for each business process.

² https://github.com/renuka98/context_aware_allocation

2.3 Learning Influence of Context on Resource Allocation

In push-based dispatching, a system or a person, is responsible for allocating the tasks of the process to relevant human resources. To study the influence of context on resource allocation, two common explainable machine learning algorithms are used: (1) Decision tree learning, and (2) the k-nearest neighbor (k-NN). With the former, context inferred from event logs are used as input features and the process outcome is used as a target variable to derive a decision tree, predicting the performance of a process instance. The decision tree thus obtained can also be used to extract rules correlating contextual knowledge with process data when the intent is to guarantee a certain set of outcomes (in other words, a certain performance profile). With the k-NN approach, k-NN regression is used to determine from the nearest neighbors of a process instance, those values of the process context (and particularly those that characterize resources) that would likely lead to the desired outcomes. The evaluation of the approach is presented using both a real-world dataset (BPIC 2013) and a synthetic dataset. Evaluation of the synthetic dataset aims to verify the ability of using the approach to discover context dependent task allocation rules, that are synthetically introduced in the log. The real-world data is used to validate influence of context when allocating tasks. The decision tree model helps explain the influence of contextual factors on process outcome (and hence resource efficiency). The model predicts process outcome with an F1-score of 67.13% on the hold-out test data.

2.4 Context Mining

Observing and analyzing impact of the context of a process or the environmental factors, on its execution outcome helps adapting and improving the process. There can be two views of context [8]. First, explicit context which is information that is identified by domain experts and can be defined a priori. Second, implicit context where some situations arise as a part of performing a task or an activity, and may not be known prior to task or process execution. These *implicit contextual dimensions* need to be discovered from various sources of information [9]. In this work, the problem of exploiting unstructured textual data to discover implicit context is studied. In the proposed approach, phrases of textual data are extracted from relevant textual logs of process instances. The phrases are used instead of the entire user comments, as comments would include several aspects of task execution. The extracted phrases or segments of information are clustered. Since the phrases are very short documents, different clustering approaches are evaluated to identify the most suitable method for short document clustering. The clusters are semi-automatically pruned by applying filtering rules such as size and process outcome, to arrive at subset of textual clusters that are likely to relate to implicit contextual information and impact process outcome. The final decision of the information being a contextual dimension or not, is made by domain experts. Different clustering algorithms and textual similarity measures are evaluated on an annotated dataset containing cluster labels. This

enables identifying a suitable clustering algorithm and similarity metric. Next, the approach using clustering and similarity metric is used to identify phrases on a business process log. Clusters of phrases that have statistically significant variance in the performance outcome, are identified. The phrases mined from the approach present interesting insights on possibility of identifying implicit context through unstructured textual data.

However, a key challenge in the evaluation of this study is the use of unsupervised learning algorithm on business process logs that do not contain a gold-standard (or ground truth), that other natural language processing tasks utilize for evaluation. Lack of process event logs containing textual information limits the ability to evaluate on multiple process logs and assess the repeatability of results. Hence, we consider this work as a first step in mining contextual data from process logs that could be further improved with the use of supervised learning on an expert annotated or labeled dataset.

In conclusion, this thesis compiles a set of methods that are data-driven and context-aware to help managers and resources allocate suitable tasks and help improve the overall process outcome.

References

- [1] O. Saidani, C. Rolland, and S. Nurcan, "Towards a generic context model for BPM," in *48th HICSS 2015, Hawaii, USA*, 2015, pp. 4120–4129.
- [2] M. Rosemann and J. Recker, "Context-aware process design exploring the extrinsic drivers for process flexibility," in *Proc. of the CAISE*06 Workshop on BPMDS '06, Luxemburg*, 2006.
- [3] J. Ghattas, P. Soffer, and M. Peleg, "A formal model for process context learning," in *BPM 2009 International Workshops, Ulm, Germany, September 7, 2009. Revised Papers*, 2009, pp. 140–157.
- [4] A. Kumar, W. M. P. Van Der Aalst, and E. M. W. Verbeek, "Dynamic work distribution in workflow management systems: How to balance quality and performance," *J. Manage. Inf. Syst.*, vol. 18, no. 3, pp. 157–193, Jan. 2002.
- [5] J. Nakatumba and W. M. P. van der Aalst, "Analyzing resource behavior using process mining," in *BPM Intl. Workshops, Germany*, 2009, pp. 69–80.
- [6] A. Kumar, R. M. Dijkman, and M. Song, "Optimal resource assignment in workflows for maximizing cooperation," in *BPM 2013, China*, 2013, pp. 235–250.
- [7] A. Pika, M. T. Wynn, C. J. Fidge, A. H. M. ter Hofstede, M. Leyer, and W. M. P. van der Aalst, "An extensible framework for analysing resource behaviour using event logs," in *CAiSE, Greece, June*, 2014, pp. 564–579.
- [8] P. Dourish, "What we talk about when we talk about context," *Personal Ubiquitous Comput.*, vol. 8, no. 1, pp. 19–30, Feb. 2004, ISSN: 1617-4909.
- [9] J. Kiseleva, "Context mining and integration into predictive web analytics," in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, 2013, pp. 383–388.