Advances in Computer Science Research, volume 88
International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)

Task Assignment in Business Processes Based on Completion Rate Evaluation

Tingting Zheng[1, a], Jiexuan Chen[2] and Yingchi Huang[3]
[1]The Open University of Guangdong Guangzhou, China;
[2]School of Data and Computer Science Sun Yat-sen University Guangzhou, China;
[3]Guangdong Digital Research Center China Mobile Guangzhou, China.
[a]ttzheng@gdrtvu.edu.cn

Abstract. In order to ensure the completion of cases and tasks on time in business processes, an approach for task assignment based on completion rate evaluation of resources is proposed. The approach is based on the data analysis of the event logs, in which the related historical temporal information is extracted and analyzed to evaluate the completion rate of the resources for tasks. The capabilities of resources are classifying into different tasks, and the completion rate is evaluated according to the historical completion, proportions of working time and the status of the resources. The results of the evaluation are applied to task assignment including two ways of offering and allocating in business processes to ensure the completion of cases on time. Experiments show that the on time completion rate of cases is increased by the applied of the approach on task assignment.

Keywords: Business process management, probability analysis, resource management, task assignment.

1. Introduction

There are time constraints about cases, tasks and resources in a business process model. The event logs of the processes reflect the running information about resources, cases and tasks, so the performance of a process can be optimized through the data analytics of the event logs.

To achieve more accurate process management at runtime, some detailed temporal information of the running cases and time prediction according to the historical event logs are applied. Resource management is also an important aspect in business processes. The assessment of resource capability and the task assignment strategy is used to ensure the completion of cases by their deadlines and prevent time exceptions caused by improper task assignment and resource allocation..

2. Related Work

Process mining [1], as one of the research hotpots in business process management, has been widely used in process discovery [2], process optimization [3] and so on. The analysis on temporal information which is extracted from event logs is becoming increasingly important in process mining. Aalst et al introduce the time factors to the process mining and construct the transition system based on log mining, and they analyzed the temporal information based on the transition system [4]. The queuing theory, such as random processes [5, 6], Markov chains [7] and so on, were used in the analysis and evaluation of temporal information. Approaches on data analytics and data mining, such as time series analysis [8], pattern matching [9], KNN algorithm [10], and genetic algorithm [11] are used to achieve more accurate prediction on processes. Recently, efforts on deep learning, such as machine learning [12], RNN [13], LSTM [14], are introduced. Some research can provide real-time data analysis [6-7, 12-13], while others are posterior induction.

Some efforts contributed to prevent time exceptions in time prediction, resource management and task assignment. Pika and Aalst et al proposed a series of indications and prediction functions to evaluate the risks of timeout [15-16]. Conforti et al realized a risk prediction system according to event logs [17]. Potential time exceptions can be eliminated by redistribution based on the

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

probability transfer models and GA algorithms [11]. Resource capacity estimate and intelligent task assignment would be used to achieve higher efficiency in process management. But there is less research on how to apply the prediction to process management, such as resource allocation, task assignment except for literature [6]. But the approach in literature [6] is not implemented based on process mining, but on the model of stochastic process. There are some guiding principles for preventing time exceptions in some research [15-17], but no corresponding verifications are proposed.

The resource completion time of task is assessed when the task is assigned to a specific resource. But the existing approach for assessment is mainly the average time [17-19]. However, the time of a resource completing a certain type of task may fluctuate widely. The time of completing a certain type of tasks in different time periods and different conditions (e.g.by different resources) may be quite different. So the consideration of average time in task assignment is deficient.

An approach for task assignment based on the completion rate evaluation of resources is proposed in this paper, in which the capability of resources for different tasks is evaluated and the status of resources is taken into consideration, and the task assignment is implemented according to the result of evaluation.

3. Resource Completion Rate Evaluation

3.1 Resource Historical Capability Assessment.

The cumulative distribution function (CDF) is used to assess the capability of resources here. The CDF is obtained from the event logs, which reflects the probability distribution of the task completion time. But previous works on the application of CDF is mainly for cases and tasks [8] but not for every resource separately. CDF of a separated resource on a specific task reflects the capability of resource for this type of task.

The capability of resource based on the CDFs is defined as follows.

Definition 1(Capability of resource) If the resource R_i is qualified for n different types of tasks $T_1, T_2, \dots, T_n$, then the capability scope of resource R_i is defined as $Skills(R_i)=\{ T_1, T_2, \dots, T_n\}$. Let $F^i_j(x)$be the CDF of completion time of resource R_i for task T_j, then the set describing the capability of resource R_i is defined as follows:

$$Capacity(R_i)=\{ F^i_j(x) \mid T_j \in Skills(R_i), j=1...n\} \tag{1}$$

Then we assess the capability of resource R_i for every type of tasks according to the set $Capacity(R_i)$.For example, when evaluating the probability of R_i completing task T_j, T_j in time units T, the corresponding CDF $F^i_j(x) \in Capacity(Ri)$ is evaluated as the formula (2) below:

$$P(x \leq T)= F^i_j(x) \tag{2}$$

For the tasks to be assigned, the candidate of resources R_i is required to meet the following conditions:

1) For task T_j to be assigned, $T_j \in Skills(R_i)$.

2) If the total completion time of work items at the waiting queue of resource R_i is $WaitingTime_i$ and the task T_j is required to complete in time DLT, the CDF of R_i in required deadline is evaluated as formula(3) below:

$$F^i_j(DLT - WaitingTime) \tag{3}$$

If θ is the minimum required probability of completion rate of the task T_j and can be set according to demands, the completion rate of resource R_i is required to satisfy the formula (4) below:

$$F^i_j(DLT - WaitingTime) \geq \theta \tag{4}$$

For the resource R_i is qualified for n different types of tasks, the total completion rate of resource R_i for T_j is defined as follows.

Definition 2(Historical completion rate of resource) If the task is required to finish in time DLT, the completion rate of resource R_i for the task T_j is

$$F^i_j(DLT - WaitingTime) \times \frac{TotalTime(R_i, T_j)}{TotalTime(R_i)} \tag{5}$$

where the F^i_j(DLT- WaitingTime$_i$) is defined as former, the TotalTime(R_i, T_j) is the proportion of R_i consuming time in completing the type of task T_j, and the TotalTime(R_i) is the total working time of R_i.

In formula (5), F^i_j *(DLT-* WaitingTime$_i$*)* is considered as the probability of completing the task in required time, in which the waiting time is subtracted. If the cumulative distribution probability of the completion time according to formula (3) is considered as the posterior probability of completing the task in required time by resource R_i, the result of $\frac{\text{TotalTime}(R_i, T_j)}{\text{TotalTime}(R_i)}$ can be regarded as the completion proportion of resource R_i in all tasks of the Capacity(R_i). Therefore, the completion rate obtained by formula (5) is actually a priori probability based on the Bayes theorem.

3.2 Resource Completion Rate Evaluation in Current Status.

When a task is assigned to resource R_i, there are many factors that may influence the completion time of R_i. For example, the continual working time, the type of period of working, the length of the work lists, etc. For more accurate completion rate evaluation, these factors should take into consideration in resource evaluation. The influence of these factors on completion time is mostly linear [15, 16]. However, too many parameters in historical event logs may lead to high computation complexity. So some most influential factors should be selected in the computation.

Here the lasso regression can be used for feature selection due to a key feature of the procedure: shrinkage of the vector of regression coefficients toward zero with the possibility of setting some coefficients identically equal to zero, resulting in a simultaneous estimation and variable selection procedure [20].

The lasso regression is an effective dimension reduction method and can realize effectively feature selection with multiple collinearities. The lasso estimator is then defined as the formula (6).

$$\hat{\beta} = \arg\min_{\beta}(\|Y - X\beta\|_2^2 + \lambda\sum_{i=1}^{p}|\beta_i|) \tag{6}$$

To deal with the problem of logistic regression more effectively, some algorithms are proposed in literature [21].

After selecting the features, the linear regression for the completion time evaluation can be conducted, and the logistic regression as the formula (7) is used to evaluate the completion rate of resource R_i with its current status:

$$\hat{F}(x) = \frac{e^{\hat{g}(x)}}{1 + e^{\hat{g}(x)}} \tag{7}$$

The $\hat{g}(x)$ is the linear regression of the completion time estimation constructed by the selected features according to (6). And the result of formula (7) is the evaluation of the completion rate of resource with the current status considering its historical performance by the required deadline, which can be taken into consideration in task assignment besides formula (5).

For resource R_i and the task T_j to be assigned, firstly we can select the most important features using lasso regression. Then the linear regression $\hat{g}^i_j(x)$ on the completion time of resource R_i considering its historical performance and its current status for the task T_j can be constructed. If the total completion time of work items at the waiting queue of resource R_i is WaitingTime$_i$ and the task T_j is required to complete in time DLT, the completion rate can be evaluated as follows:

$$\hat{F}^i_j(DLT - Waitingtime) = \frac{e^{\hat{g}^i_j(DLT-Waitingtime)}}{1 + e^{\hat{g}^i_j(DLT-Waitingtime)}} \tag{8}$$

4. Task Assignment

4.1 The Ways of Task Assignment.

There are two ways of task assignment in process management: offering and allocating [21]. If a resource is offered a task, the resource can choose to accept the task or transfer it to other resources. While in the way of offering, the workflow engine would offer a task to several resources that satisfy the requirements successively. When there is any resource accepting the offering and the response is accepted by the workflow engine, the state of the task turns into assigned. If a resource is allocated a task, the resource is not allowed to reject the task or transfer it to others. The specific differences of the two ways are shown in in Table 1.

Table 1 Difference of offering and allocating

Comparison items	Offer	Allocate
Target	Several resources	One resource
Reject or transfer	Allowed to reject or transfer	Not allowed to reject or transfer
Scope of application	Not urgent cases or tasks.	Urgent cases or tasks

The urgent cases in Table 1 satisfy at least one of the following characteristics:

- Higher priority;
- The sum of the historical average completion time of the remaining tasks is less than the remaining required time.
- The historical completion rate of the case in the remaining required time is less than 50%.The progress of the current case is slower than expected.

The urgent tasks in Table 1 satisfy at least one of the following characteristics:

- Higher priority;
- The probability of the task completion time exceeds the upper limit of the confidence interval of its historical completion time is higher other tasks in the process.
- The total historical completion rate of the task in the remaining required time is less than 50%.
- The progress of the current task is slower than expected.

4.2 Implementation.

The approach for task assignment is designate the way of offering or allocating depending on the priority, the rate of progress of cases, probability of historical completion rate, and etc. The resources for task assignment in both two ways should satisfy the conditions described in IIIA. The main process is shown in Fig. 1. If timeout exceptions exist, the current case would request manual interventions and the remaining tasks would not be assigned automatically.

1) The implementation of allocating: The way of allocating will assign a task to a single resource only every time, which is often used in urgent cases or tasks as described above.

So there are higher requirements for the completion time of resources in the way of allocating. Therefore, the completion rate of resources is the most important consideration. The implementation of allocating is mainly to select a resource which is most likely to complete the task every time, and the factors are described below.

- For all candidates of resources, compute the posterior probability of completing the task according to the formula (3) and the completion rate in the current status according to the formula (7). Average the two results above.
- Select the maximum in the candidates in a given error range.
- If there are more than one resources in the error range, compute the historical completion rate according to formula (4) and select a maximum.

2) The implementation of Offering. When the current case nor the task is not urgent, the task is offered to all resources that satisfy the requirements described in IIIA above.If there are more than one resource accepting the offering, the workflow engine would choose one in the accepting resources in the following two ways:

- Choose the resource that is the fastest one to respond the acceptance;
- Choose the resource by the historical completion rate and the current status similar to the implementation of allocating.

The first way of choosing the fastest one is easier to realize in workflow systems but less efficient. We will adopt the second way as the way of allocating for a more efficient system.

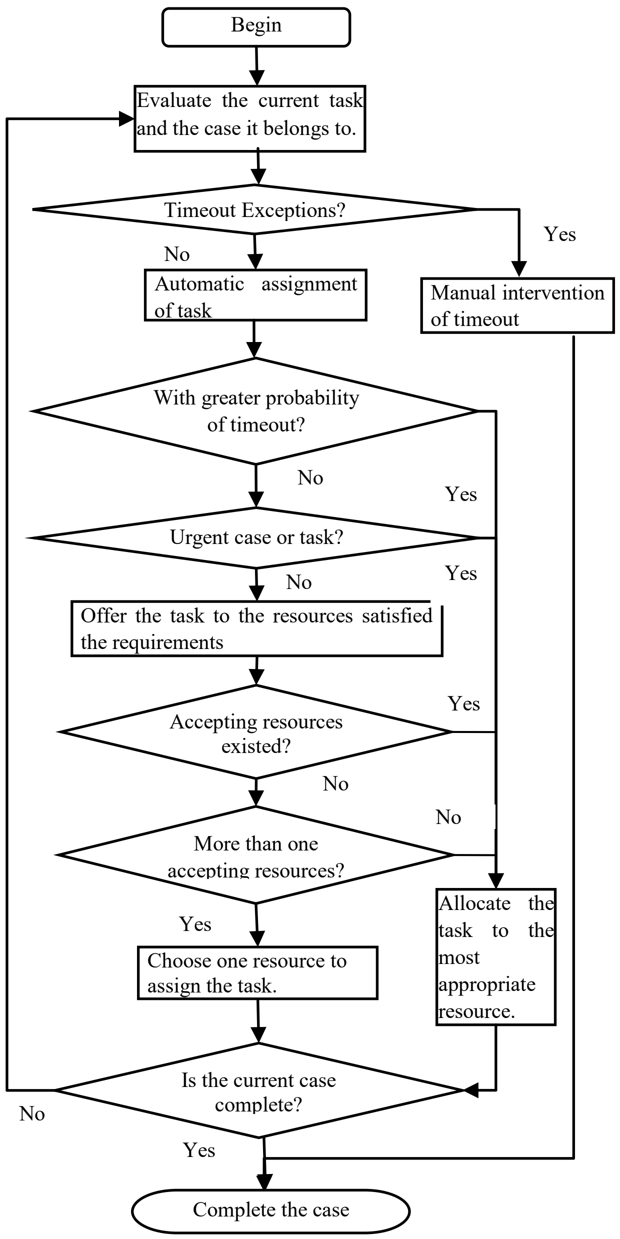

Fig. 1 The main process of task assignment.

5. Experiments

To verify the efficiency of the proposed algorithm, the process derived from a procurement process of a project is simulated in YAWL, which is shown in Fig. 2, and the logs of the process is extracted and analyzed.

Let the number of cases generated per unit time satisfy the Poisson distribution, and each case has its independent deadline. If the case is not completed by the deadline, timeout exception is thrown. The percentage of cases completed by the deadline in a specified time period, which is call on time completion rate, is calculated in the experiments.

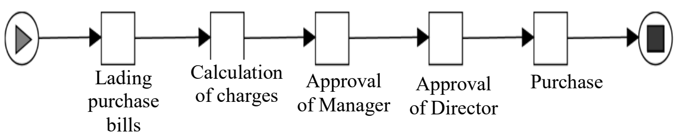

Fig. 2 The procurement process for simulation.

The approach of task assignment proposed in this paper is denoted as DRE (Dispatch on Resource Evaluation), which is compare with the following algorithm in YAWL:

(a) Random Choice Dispatch (RCD): select a resource randomly, which is the algorithm used the most commonly in the process;

(b) Shortest Queue Dispatch (SQD): select a resource with the shortest work list and least acceptance of the task, in which the acceptance of historical and the current workload are considered.

Firstly, in the lighter load (no more than 70% of the maximum load) and with plenty time to complete the cases, the on time completion rate of cases in a given period (several time units) is shown in Fig. 3.

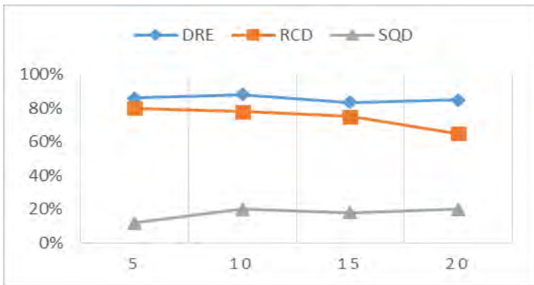

Fig. 3 Comparison of the completion rate with lighter load.

It is shown in Fig. 3 that the algorithm of DRE is the best in performance with the case completion rate always around 85% and the maximum at about 90%, which is also the most stable. The algorithm of RCD starts at a high level, but is declining to the worst of about 65%. The algorithm of SQD has a poor performance with only about 20% at the highest, which is because if only considering the historical proportions of accepting the task but not the capability of the resources, sometimes resources with insufficient capability are selected, so the overall completion rate is seriously lowered.

Then considering higher load and urgent cases, the way of allocating is often designated in DRE in this circumstance. It is shown in Fig. 4 that, the on time completion rate, which is declining significantly with the increase of cases in RCD, is maintained at 78% -80% in DRE. And the completion rate in SQD has poor performance at 0%- 20%. Generally, experiments show that the on time completion rate of DRE can maintain in a high level even in higher load and urgent cases.

Fig. 4 Comparison of the completion rate with higher load.

6. Conclusion

The completion rate evaluation of the resources for the tasks to be assigned is proposed in this paper, in which the capability of resources for every type of tasks is assessed, the historical completion rate of resources is evaluated, and the completion rate considering the status of resources is evaluated by selecting features using the lasso regression and constructing the logistic regression. The results of completion rate evaluation can be used for task assignment at runtime. Experiments are realized to verify the efficiency of the approach.

Future research will be carried out in the optimization of the task assignment strategy. We will try to evaluate the efficiency for the whole system furthermore.

References

[1]. Aalst, Wil Van Der. Process mining. Process Mining. 2012.

[2]. De Leoni, Massimiliano, d. A. W. M. P. Van, and M. Dees. "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs." Information Systems: S0306437915001313, 2015.

[3]. Li, C., M. U. Reichert, and A. Wombacher. "Mining business process variants: Challenges, scenarios, algorithms." Data & Knowledge Engineering vol 70.5, pp. 409-434, 2011.

[4]. Aalst, W. M. P. Van Der, M. H. Schonenberg, and M. Song. "Time prediction based on process mining." Information Systems, vol 36.2, pp.450-475, 2011.

[5]. Guo, Xiaobo, et al. "Dynamically Predicting the Deadlines in Time-Constrained Workflows.". Proc. International Conference on Web Information Systems Engineering, Springer, 2014, pp.120-132.

[6]. Chirkin A M, Belloum A S Z, Kovalchuk S V, et al,"Execution time estimation for workflow scheduling". Future Generation Computer Systems vol. 75, pp. 1-10, 2017.

[7]. Rogge-Solti A, Weske M,"Prediction of business process durations using non-Markovian stochastic Petri nets". Information Systems vol. 54(C), pp.1-14, 2015.

[8]. Lam C Y, Ip W H, Lau C W, "A business process activity model and performance measurement using a time series ARIMA intervention analysis". Expert Systems with Applications vol. 36(3), pp. 6986-6994, 2009.

[9]. Ceci, M., Lanotte, P. F., Fumarola, F., Cavallo, D. P., & Malerba, D, "Completion Time and Next Activity Prediction of Processes Using Sequential Pattern Mining". Discovery Science. Springer International Publishing, 2014.

[10]. Kang B, Kim D, Kang S H, "Real-time business process monitoring method for prediction of abnormal termination using KNNI-based LOF prediction. Expert Systems with Applications, vol. 39(5), pp.6061-6068, 2012.

[11]. Márquez-Chamorro A E, Resinas M, Ruiz-Cortés A, et al, "Run-time prediction of business process indicators using evolutionary decision rules". Expert Systems with Applications, vol.87, pp.1-14, 2017.

[12]. Polato M, Sperduti A, Burattin A, et al,"Time and activity sequence prediction of business process instances", Computing, vol. 10, pp. 1-27, 2016.

[13]. Evermann J, Rehse J R, Fettke P, "A Deep Learning Approach for Predicting Process Behaviour at Runtime In: International Conference on Business Process Management", pp. 327-338. Springer, Cham, 2016.

[14]. Tax N, Verenich I, Rosa M L, et al, "Predictive Business Process Monitoring with LSTM Neural Networks", Advanced Information Systems Engineering, pp. 477-492, 2017.

[15]. Pika A, Aalst W M P V D, Fidge C J, et al, "Profiling Event Logs to Configure Risk Indicators for Process Delays", Advanced Information Systems Engineering, pp. 465-481.Springer Berlin Heidelberg, 2013.

[16]. Pika A, Aalst W M P V D, Fidge C J, et al, "Predicting Deadline Transgressions Using Event Logs". Proc. Business Process Management Workshops, pp. 211-216, Springer Berlin Heidelberg, 2013.

[17]. Conforti R, Leoni M D, Rosa M L, et al," A recommendation system for predicting risks across multiple business process instances", Decision Support Systems, vol. 69, pp. 1-19, 2015.

[18]. Liu Yingbo .Research on Workflow Runtime Intelligent Staff Assignment Technology. CA:Tsinghua University, 2008.

[19]. Byun E K, Kee Y S, Kim J S, et al, "BTS: Resource capacity estimate for time-targeted science workflows", Journal of Parallel & Distributed Computing, vol. 71(6),pp. 848-862, 2011.

[20]. Jones, M. C., and A. Pewsey. "Bayesian lasso regression", Biometrika, vol 96.4, pp. 835-845, 2009.

[21]. Meier, Lukas, V. D. G. Sara, and P. Bühlmann. "The group lasso for logistic regression." Journal of the Royal Statistical Society, vol. 70.1, pp. 53-71, 2008.