Attribute Selection with Filter and Wrapper: An Application on Incident Management Process

Claudio A. L. do Amaral, Marcelo Fantinato, Sarajane Marques Peres
Graduate Program in Information Systems, Universidade de São Paulo, Brazil
{claudio.amaral, m.fantinato, sarajane}@usp.br

Abstract—Few approaches allow assertive estimates for ticket completion time in incident management. The accuracy level of prediction models depends on how useful the used attributes are. Moreover, to effectively use computational resources, a canonical attribute subset must be used. This paper proposes two automated attribute selection methods to build prediction model. A filter method and two wrapper search techniques were combined with annotated transition systems to automate attribute selectors applied to a real-life incident management process. The results show that the wrapper method surpassed human experts' decision making.

Index Terms—process mining; attribute selection; incident management; ITIL; annotated transition system.

I. Introduction

IN Information Technology (IT), optimization is sought by adopting frameworks such as the Information Technology Infrastructure Library (ITIL) [1]. ITIL covers several IT service management processes, including incident management [2], which is responsible for correcting failures and restoring the normal service operation, as soon as possible, minimizing the impact on business [1]. One of the most relevant monitoring indicators related to this process is the completion time for incident resolution (a.k.a. 'ticket completion time') [2].

Assertive and reliable estimates for completion time is still challenging [3]. A common reason for poor estimates is to conduct predictions based only on a naive and superficial abstraction of the actual process being performed. Fortunately, many companies are using process-aware information systems and recording events about the activities executed. The large amount of data recorded in event logs can be explored in detail through different process mining techniques, which allow to infer a more realistic process model [4]. For example, representing the process as an Annotated Transition System (ATS) allows to estimate the process completion time based on statistics aggregated into the process model [5]. To achieve a proper ATS model representing an incident management process, both the event log and a set of descriptive attributes need to be considered. However, depending on management context, the number of descriptive attributes that may be associated with process instances can be large and complex enough to render unfeasible *(i)* the use of all descriptive attributes, which could generate inefficient ATSs to predict completion time as well as *(ii)* a non-automated decision making about which attributes should be considered to build the ATS. Therefore, two concerns should be considered when building a proper process model: not all attributes are necessarily useful

and much computational resource may be required. Thus, a canonical subset of descriptive attributes must be selected; i.e., an ideal minimum subset of descriptive attributes that minimizes the computational cost and contains the maximum information relevant to build the model.

This paper proposes to apply two classic methods of attribute selection to automatically determine the canonical subset of descriptive attributes. The filter [6] and wrapper [7] methods have been applied to an event log obtained from a real-world enterprise incident management system. For the experiments, ATSs were created using attribute subsets selected by human experts in addition to the two automated methods. The results show that the proposed automated methods surpass human experts in selecting attributes for prediction using the ATS built based on the selected attributes, having wrapper surpassed filter. The remainder of this paper presents: background, related work, proposed solution, experiments and results, and conclusion and future work.

II. Background

The transition systems used in process mining was proposed by Aalst *et al.* [8] and then extended with annotations to describe statistical data that allow predicting the completion time of a process instance [5]. To create the ATS, each state is annotated with data collected from all traces that have visited it [5]. For time analysis, for example, data about the completion time of the instances related to each earlier trace is used. The data are aggregated in each state producing statistics such as average time, standard deviation, median time etc. Two of the proposed strategies are applied here: *maximal horizon* and *representation*, including per *sequence*, *multiset* and *set* [5].

Attribute selection is essential to build a model capable of predicting ticket completion time, by deciding which features to describe the concept to be learned and how to combine them [9]. Methods for selecting attributes are typically classified as filters, wrappers and embedded [6]. In this paper, a filter method based on correlation analysis was applied. Each attribute is individually evaluated based on its correlation with the target attribute (i.e., the ticket completion time). Moreover, two well-known search techniques were applied: hill-climbing and best-first search [7]; having ATSs as the learning model and Mean Absolute Percentage Error (MAPE) [10], [11] as the metric to evaluate the learning model accuracy.

III. Related Work

The proposal presented in this paper is based on the extension of transition systems with annotations, which was originally proposed by Aalst, Schonenberg and Song [5], to predict completion time of running traces. According to them, ATSs include alternatives for state representation, allowing to address overfitting and underfitting in prediction tasks. They concluded that their prediction approach overcomes simple heuristic approaches. Other authors, as Polato *et al.* [12], extended ATS to solve the same task by combining the probability of occurrence of activities with a regression model.

In addition to transition systems, Petri nets have also been used as a technique for prediction work. Rogge-Solti, Vana and Mendling [13] introduced time series Petri net models, making it possible to handle the analysis of temporal aspects of processes. Hinka *et al.* [14], Evermann, Rehse and Fettke [15] and Tax *et al.* [16] presented approaches closer to the study presented in this paper. Tax *et al.* [16] presented a comparison of their approach with ATSs used as predictor and concluded that they obtain more accurate predictions except for instances with a reduced number of events. Approaches were assessed with cross-validation and prediction accuracy metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) [10].

Although with different strategies addressing completion time prediction, there is a lack of concern on choosing the input log configuration for the predictor induction. A preprocessing work for attribute selection, as proposed in this paper, has the potential to improve results of the related work.

IV. Proposed Solution

When an incident occurs, it is identified and reported by a caller. Afterwards, a primary expectation is to know the incident completion time. The usual estimates follow ITIL best practices, which are based on some specific incident attributes such as urgency, category etc. This approach is fairly general and inaccurate since it aggregates a large number of different situations and common target completion times. As the process evolves from early stage to initial support and investigation, some attributes are updated and new ones are added. Depending on the system used, it can usually lead to a number close to 100 attributes. There is an open issue related to providing assertive estimates on incident completion time that is not adequately solved by simple statistical methods. Incident management systems store descriptive information of process instances and audit information about the history of updates of the process in progress. The combination of both types of information allows executing a detailed step-by-step process evaluation and hence deriving estimates for each event.

The problem addressed here lies in such a scenario, where, one needs to discover an attribute subset that allows generating a model capable of minimizing the prediction error of the incident completion time during the process of its resolution. The process starts with a sequence of actions to build the enriched event log used to induce the prediction models. After that, it is possible to apply the three attribute selection methods explored in this paper: *(i)* expert-driven selection, *(ii)* filter with ranking and *(iii)* wrappers with the hill-climbing and best-first search techniques. The search is performed in the context of our attribute selection strategy. In these algorithms, there is a construction function to build an ATS and the evaluation function of the ATS. They use, respectively, a training log excerpt and a testing log excerpt, which represent disjoint subsets of the original event log generated in a cross-validation procedure. After that, the evaluation function is applied and returns the MAPE for the ATS under evaluation. The minimization function applied to the ATS evaluation returns the index of the model that produces the lowest MAPE when applied to the testing log. As final result, it is returned the ATS with lowest MAPE in the set of ATSs under evaluation.

For all the selection methods, ATS is applied as the prediction model responsible for generating the estimates of the incident completion times, including to act as a state evaluator in the wrapper search spaces. For practical purposes, the basic idea is that ATS can be generated from an attribute subset which adequately describes the currently completed incidents. From this point, ATS can be applied to predict the completion time of new incidents while they are running.

V. Experiments and Results

This section presents the used event log, experiments setup and execution details, and results obtained. A cross-validation method with 5 folds was applied to the enriched event log to build the prediction models. The ATS accuracy is given by the average MAPE of test folds in terms of mean and median of incident completion time. In addition, the ATS completeness (or non-fitting) was evaluated as the accounting of how many event records do not have a corresponding state in ATS.

A. Enriched Event Log

An enriched event log of the incident management process was extracted from an instance of the platform used by an IT company (Table I). Information was anonymised for privacy reasons. This enriched event log is composed by information gathered from the audit system and the relational model of the platform. A preprocessing step filtered out noise and sorted audit records in a sequence compatible with event log format.

Some statistical data on the enriched event log is shown in Table II. A well-defined behavior for the incident management process is observed, as most incidents (75%) go up to 7 updates, 50% up to 5 updates and on average 6 updates are needed to close incidents. However, there are outliers, with 58 as the maximum number of updates for one incident. Regarding time, the behavior resembles an exponential distribution.

B. Experiments Setup and Results

For the three experiments conducted, the ATSs parameters were: an enriched event log was randomly sampled by creating two subsets with 8,000 (A) and 24,000 (B) incidents, having $A \subset B$. The maximum horizon parameter values used were: 1 – case with the last event per incident trace; 3, 5, 6 and 7 – most frequent behaviors in this incident process (statistic 'by incident' in Table II); 'infinite' – uses all events in trace.

TABLE I
INCIDENT ENRICHED EVENT LOG EXCERPT

number	incident_state	sys_updated_on	category	assignment_group
	New	3/2/2016 04:57	Internet	Field Service
	Active	3/2/2016 18:13	Internet	Field Service
INC001	Awaiting UI	3/2/2016 19:15	Internet	Field Service
	Active	3/3/2016 12:43	Internet	Field Service
	Resolved	3/4/2016 11:02	Internet	Field Service
	Closed	3/9/2016 12:00	Internet	Field Service

TABLE II
ENRICHED EVENT LOG STATISTICS: DISTRIBUTIONS BY INCIDENT/DAY

	1st Quart.	2nd Quart.	3rd Quart.	Max.	Mean	Std. Dev.
By incident	3	5	7	58	6	3.67
By day	0.01	0.40	5.29	336.21	6.67	21.20

1) Experiment #1 – Expert-Driven Selection: first, attribute selection was driven by data about the domain held by human experts. According to ITIL best practices, in the first stage of incident management process, the caller should provide the initial information, which is complemented by the service desk agent, especially with information related to the incident category and priority. Additional information (i.e, textual descriptions) is also provided to help the support agents; its use is out of the scope of this work. Based on these practices, *incident_state*, *category* and *priority* were considered the most adequate attributes to define the process model in ATS: *incident_state* reports the stage at which the incident is; *category* indicates the type of service the incident belongs; and *priority* determines the focus requested by business. In this scenario, using event log sample with 24,000 incidents and varying the horizon and state representation parameters, 18 ATSs were generated and used as completion time predictor. The best results were obtained with horizon 3, state representation *sequence* and are shown in Table III.

TABLE III
EXPERIMENTS – AVERAGE PREDICTION RESULTS. BEST ATTRIBUTE
SUBSETS SELECTED BY SPECIALIST, FILTER AND WRAPPER. LOG SAMPLE:
24,000 INCIDENTS. METRIC: MAPE. NF = % OF NON-FITTING
INCIDENTS. BOLD: BEST RESULTS ON EACH EXPERIMENT.

Max Horiz	Set			Multiset			Sequence		
	Mean	Median	NF	Mean	Median	NF	Mean	Median	NF
Experiment #1 – Attribute subset: *{incident_state, category, priority}*									
3	106.93	77.46	0.98	91.35	75.87	1.23	**72.36**	**63.66**	**1.38**
Experiment #2 – Attribute subset: *{caller, assigned_to}*									
5	90.73	76.30	33.31	**69.69**	**57.85**	**35.67**	80.97	69.10	35.73
Experiment #3 –Attribute subset: *{incident_state, location}*									
5	**50.45**	**24.49**	**1.11**	41.90	29.35	2.30	35.09	27.28	2.74

2) Experiment #2 – Filter with Ranking: second, attribute selection was driven by filter using a ranking strategy. This approach follows consolidated concepts of specialized literature [6], [7], [9]. In this paper, ranking was applied as preprocessing, as suggested by Kohavi and John [7], to create a baseline for attribute selection, regardless of the prediction model in use. It was created through a variance analysis by correlating the independent variables (i.e., descriptive attributes) and the dependent variable (i.e., attribute 'closed', prediction target attribute). Since most of descriptive attributes are categorical,

the statistic η^2 (Eta squared) was applied, as explained by Richardson [17]. As a design decision, the 15 attributes with the highest correlation were selected to compose the ranking. The variance analysis was carried out on the entire enriched event log. The attributes and correlation scores are listed in Table IV. Using ranking results filter was executed by combining the attributes as follows: *{Caller (1st)}; {Caller (1st), Assigned to (2nd)}; ...; {Caller (1st), Assigned to (2nd), ..., Knowledge (15th)}.* For this scenario, 18 ATSs were generated for each attribute subset. In the second part, the best two result sets obtained in the first part where chosen to generate new sets of ATSs, however, using event log sample with 24,000 incidents (best results in Table III). The prediction results with the ranked attribute subsets were slightly worse than those obtained in the experiment #1. Analyzing the results, it is noticed that, resource-related attributes impair the generation of the prediction model, i.e., they do not reflect the process behavior with the same fidelity as the control attributes do. Particularly regarding to non-fitting, a possible explanation for these poor results could be the frequent changes in the values of the human resources assigned to solve different incidents.

TABLE IV
THE 15 DESCRIPTIVE ATTRIBUTES WITH THE HIGHEST CORRELATION
WITH THE DEPENDENT VARIABLE AND THE RESPECTIVE η VALUES.

Ord	Attribute	η	Ord	Attribute	η	Ord	Attribute	η
1st	Caller	0.54	6th	Incident state	0.32	11th	Created by	0.21
2nd	Assigned to	0.37	7th	Subcategory	0.32	12th	Opened by	0.20
3rd	Assig. group	0.35	8th	Category	0.27	13th	Location	0.14
4th	Symptom	0.33	9th	Active	0.25	14th	Made SLA	0.14
5th	Sys upd. by	0.33	10th	Priority conf.	0.24	15th	Knowledge	0.12

3) Experiment #3 – Wrappers with Hill-Climbing and Best-First: lastly, the attribute selection was driven by wrapper using a forward selection mode[1], with the hill-climbing and best-first search techniques [7]. The search space is composed by all possible combinations of the 15 attributes pre-selected by the filter with ranking strategy, i.e., attributes listed in Table IV. Since each combination represents a state in such a space, whose quality measure is calculated as the predictive power achieved by the ATS generated with the attribute subset associated with this model[2], an exhaustive search procedure is unfeasible and hence the use of a heuristic search procedures is justified. Wrapper was carried out on the enriched event log sample with 8,000 incidents. For the best-first search technique, the maximum number of expansion movements with no improvement was set to 15. Both search techniques selected this same best attribute subset: *{incident_state, location}.* Despite the high agreement between them, this can be highlighted: with hill-climbing, the stopping criterion was reached after the third expansion movement and 42 states of the search space were explored; using best-first, 17 expansion movements were done and 172 states of the search space were

---

[1]In the forward selection, the search start point is a singleton attribute subset to which a new attribute is incorporated at each new step in the search.

[2]The search space had $2^{15} = 32,768$ states, taking the 18 ATSs generated for each state, the range of the horizon and state representation parameters.

explored. The best results were obtained with horizon 7 and state representation *set*; however, the results obtained with the other state representations for the same horizon were very good as well. These results are significantly better than those obtained by the filter and expert-driven selections. Overall, the low non-fitting results are promising. As a second part of experiment #3, a new set of ATSs was generated using as parameters those of best results and enriched event log sample with 24,000 incidents. The best results (Table III) were obtained with maximum horizon set to 5 and overcome those obtained in the previous experiments. The MAPE results are less than half of those measures obtained by the expert-driven selection, keeping non-fitting values at the lowest level.

C. Analysis of Results

When analyzing the results, it is verified that the strategies expert-driven and filter with ranking allow us building models with a similar predictive power. However, when checking the model non-fitting capabilities, differences (1.38% and 35.67%, respectively) are observed between them for the best results. Such differences were caused due to the different process perspectives represented by the attribute subset used in each case. For the former, the ATS generation was driven by incident descriptive attributes recommended by the ITIL best practices suggested by human experts for incident clustering and routing; then, the resulting model was able to accurately represent the process. For the latter, the set of attributes automatically selected to build the ATS represents organizational and resource perspectives of the incident management process. In this case, the ATS captured the way that teams (i.e., people) act to support user requests and became highly specialized and incapable of generalizing the real process behavior. This phenomenon happens because the attributes selected represent information that presumably changes frequently (i.e, 'caller' and 'technical people' in charge of the incident). The MAPE results obtained for the experiment #1 were compared to those obtained for the experiment #2, using the paired Wilcoxon statistical test. This test showed that there is no statistical difference among the distributions of the MAPE values, seeing that, with $p_{value} = 0.3125$, it is not possible to reject the null hypothesis for equal distributions.

The wrapper-based experiment achieved an average MAPE measure (24.49) that is 38.47% of the average MAPE achieved in the expert-driven experiment. The model non-fitting continued in an even lowest level (1.11%) as that obtained in the first one. The paired Wilcoxon statistical test was applied to compare the MAPE results obtained for the experiment #1 with those obtained for the experiment #3. The null hypothesis for equal distributions can be rejected with $p_{value} = 0.0312$. This result allows to affirm that the attribute selection obtained with wrapper is better than the expert's choice.

The attribute subset selected by wrapper is the union of expert knowledge with an organizational perspective, which produced a completion time predictor with high accuracy and low non-fitting rates. Moreover, it was very similar with results obtained with the hill-climbing and best-first search

techniques. This behavior has already been observed in experiments executed by Kohavi and John [7], in which, for different types of datasets, additional search effort did not produce better results.

VI. CONCLUSION AND FUTURE WORK

Wrapper made it possible to select a set of attributes that supported a significant improvement in the accuracy of the ATS to be used as a prediction model when compared to both filter and expert knowledge. Furthermore, such search process pointed out that the maximum horizon and different types of state representations have a high influence on the prediction model results. This approach has the potential to be used as a useful preprocessing step prior to the application of other prediction methods, in addition to the ATS method used here.

As next steps, it is necessary to verify the influence of outliers throughout the process (search and prediction performance), since the results obtained in the experiments presented some variation degree. The use of other search methods such as genetic algorithms or other induction algorithms such as neural networks and the combination of the best models of ATSs with other regression models are points to be explored.

REFERENCES

[1] itSMF, "Global survey on IT service management," The IT Service Management Forum, 2013, http://www.itil.co.il.

[2] M. Marrone, F. Gacenga, A. Cater-Steel, and L. Kolbe, "IT service management: A cross-national study of ITIL adoption," *Communic. of the Association for Inform. Sys.*, vol. 34, pp. 49.1–49.30, 2014.

[3] M. de Leoni, W. M. van der Aalst, and M. Dees, "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs," *Inform. Syst.*, vol. 56, pp. 235–257, 2016.

[4] W. M. P. van der Aalst, *Process Mining – Discovery, Conformance and Enhancement of Business Processes*, 2nd ed. Springer, 2016.

[5] W. van der Aalst, M. Schonenberg, and M. Song, "Time prediction based on process mining," *Inform. Syst.*, vol. 36, no. 2, pp. 450–475, 2011.

[6] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. of Machine Learning Res.*, vol. 3, pp. 1157–1182, 2003.

[7] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intel.*, vol. 97, no. 1, pp. 273–324, 1997.

[8] W. M. P. van der Aalst, V. Rubin, H. M. W. Verbeek, B. F. van Dongen, E. Kindler, and C. W. Günther, "Process mining: A two-step approach to balance between underfitting and overfitting," *Software & Systems Modeling*, vol. 9, no. 1, 2008.

[9] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intel.*, vol. 97, no. 1-2, pp. 245–271, 1997.

[10] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *Int. J. of Forecasting*, vol. 8, no. 1, pp. 69 – 80, 1992.

[11] A. de Myttenaere, B. Golden, B. L. Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, 2016.

[12] M. Polato, A. Sperduti, A. Burattin, and M. de Leoni, "Data-aware remaining time prediction of business process instances," in *Proc. of the 2014 Int. Joint Conf. on Neural Netw.* IEEE, 2014, pp. 816–823.

[13] A. Rogge-Solti, L. Vana, and J. Mendling, "Time series petri net models – enrichment and prediction," in *Proc. of the 5th Int. Symp. on Data-driven Process Discovery and Analysis (SIMPDA)*, 2015, pp. 109–123.

[14] M. Hinkka, T. Lehto, K. Heljanko, and A. Jung, "Structural feature selection for event logs," pp. 20–35, 2017.

[15] J. Evermann, J.-R. Rehse, and P. Fettke, "Predicting process behaviour using deep learning," *Decision Supp. Sys.*, vol. 100, pp. 129 – 140, 2017.

[16] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with lstm neural networks," pp. 477–492, 2017.

[17] J. T. E. Richardson, "Eta squared and partial eta squared as measures of effect size in educational research," *Educational Research Review*, vol. 6, no. 2, pp. 135–147, 2011.