Business process analytics: a dedicated methodology through a case study

Article in EURO Journal on Decision Processes · July 2015

DOI: 10.1007/s40070-015-0050-4

3 authors, including:

Pavlos Delias
International Hellenic University

Nikolaos Matsatsinis
Technical University of Crete

Noname manuscript No.
(will be inserted by the editor)

Business Process Analytics: A Dedicated Methodology through a Case Study

Pavlos Delias · Michael Doumpos · Nikolaos Matsatsinis

the date of receipt and acceptance should be inserted later

Abstract Business Process Analytics is a set of techniques that can be applied to event datasets created by logging the execution of business processes, and emerges as a promising decision aid field. In this work we propose a methodology based on the process mining approach to guide the implementation of process analytics projects. Following a conceptual analysis of existing methodologies, we extract the common methodological steps and present a practical synthesis. The proposed methodology reaches the business need of exposing more than just a static, marginal snapshot of performance by considering a process perspective. We present the methodology in tandem with a case study of a customer service request handling process. We analyze a real dataset containing events from an incident and a problem management information system, and deliver results that eventually can raise the capacity of the company to manage the process.

Keywords Process Analytics, Process Mining, Methodology

Mathematics Subject Classification (2000) 90B50 · 68T99

1 Introduction

Customer service request handling is a reactive business process that is triggered when a customer submits a service request to the help desk of a company. It has been identified as a core function of modern organizations, due to its tight relationship with their marketing function (Wilson et al, 2012). Establishing a service response capability includes a number of actions (Grance et al, 2004), like creating a service response policy,

P. Delias
Department of Accounting and Finance, Eastern Macedonia and Thrace Institute of Technology, Kavala, Greece
Tel.: +30-2510-462368
E-mail: pdelias@teikav.edu.gr

M. Doumpos · N. Matsatsinis
School of Production Engineering and Management, Technical University of Crete, Chania, Greece
E-mail: mdoumpos, nikos@dpem.tuc.gr

setting guidelines for communicating with outside parties regarding customer requests, selecting a team structure and staffing model, establishing relationships between the help desk team and other groups, both internal (e.g., technical support teams) and external, determining what services the incident response team should provide and staffing and training the incident response team.

There are multiple factors that affect the complexity of the process, such as the number of support teams involved, the organizational hierarchy, the number of products / product categories being served, special business rules etc. Due to the complexity of this process, special IT systems are often employed, which can significantly add to the business value (Lin and Kao, 2014). A common practice reference model that introduces standard best practices for IT service management is the Information Technology Infrastructure Library (ITIL) (Hochstein et al, 2005). Nevertheless, the processes described in ITIL are deliberately non-prescriptive, therefore the process flow is not enforced (e.g., by a workflow engine (Tarantilis et al, 2008)). In practice, the actual behavior can significantly vary, not just according to the organizational implementation but because of a plethora of other implementation parameters as well (e.g. the resource performing the activities). Process mining (Van der Aalst et al, 2012) is a promising approach to expose the real behavior of the process from IT systems' logs and conceals business process optimization potentials (Van der Aalst, 1996).

The process mining approach has recently attracted researchers for the service request management process analysis (van Dongen et al, 2013). Since the respective process takes place in a highly flexible environment, multiple techniques are typically combined to deliver a solution. In (De Weerdt et al, 2012), the authors propose a combination of trace clustering and text mining to enhance process discovery techniques with the purpose of retrieving more useful insights from process data, while in (Ferreira and Mira da Silva, 2008) process mining is used to assess whether a business process is implemented according to ITIL guidelines.

In this work we propose a methodology based on the process mining approach to discover coordinated patterns of behavior in a customer service request handling process. Our efforts are not centered in delivering a standard framework, but rather in guiding the implementation of a process analytics application. The goal of this paper is to demonstrate through a real-world case study a roadmap for evidence-based decision making. The case study concerns an IT system used by Volvo IT to support incidents reported by the IT service users. It is a reactive business process, and although there is an Organization structure and some general rules (Steeman, 2013), the company actively looks for inefficiencies. The proposed methodology and its actions deliver effective analytics for such a business quest. Actions are described with a clear reference to the case study, however they are relevant and applicable to any case when event-based data sets are available. We should emphasize that besides guiding an application, additional motivation for delivering such a methodology is to support best practices reporting and sharing, and to endorse evidence-based approaches for decision making.

The next section is a brief presentation to scaffold readers into the proposed approach. Next, we present phases and actions in parallel with their concrete instantiations concerning the case study. Last, a short discussion on the limitations and future work concludes the paper.

2 Outline of the Proposed Approach

The explosion of generated data has lead to several proposals of methodologies for practitioners to extract useful insights from datasets, KDD (Fayyad et al, 1996), SEMMA (Matignon, 2007), and CRISP-DM (Chapman et al, 2000) being the most popular among them. Concentrating on process mining, we regard the L* model (Van der Aalst, 2011, pp. 283-286) that portrays the basic steps to improve a structured (Lasagna) process, context specific approaches (e.g., methodologies for healthcare (Rebuge and Ferreira, 2012; Delias et al, 2015)), and the work of (Bozkaya et al, 2009) to exploit different perspectives of process mining for a specific purpose (to gain an quick overview). Heijden (van der Heijden, 2012) followed a System Engineering Process to identify the requirements and the main activities of a process mining project and to deliver (like CRISP-DM) "an industry-, tool-, and application neutral methodology". There are evident overlaps between the above methodologies. Following a conceptual analysis, we can extract the common methodological steps, necessary to deliver a process analytics project. Thus, the proposed methodology *synthesizes* existing works in order to accelerate project delivery. It focuses and summarizes the bottom line of the cited works, leading to expedited knowledge discovery. The methodology is illustrated in 1. It consists of a set of actions classified into four phases, defined as a higher level of abstraction. Although a sequence is demonstrated (aiming to suggest a consistent and progressive development), it is not rigid. Switching between phases is an expected as well as an essential part of any project implementation. In addition, the dots at the bottom of the list of actions for each phase are used for suspense, to indicate that the actions' lists are not complete checklist but coarse guides.The methodology assumes that a process notion is omnipresent (it exists in problem definition, data format, solutions' intuition, etc.). Therefore, a process mining approach is qualified, since ordinary data analysis or data mining techniques would fail to capture the sequencing of the related events. The basic phases are Business Understanding (figuring out the business context and developping the shape of solutions); Data Collection and Reviewing (acquiring and preparing the raw material); Discovery (extracting bits of knowledge); and Decision Aid (building a rapport between results and business goals). The methodology will be presented in parallel with the case study implementation, while we commit the next sections to the analytical description of the steps.

3 Business Understanding

The rationale of this phase is to help the analyst arrive at a stage of reflection where she has a clear understanding of the business context, and where she can assess how alternative actions can contribute to the business objectives. The case study concerns Volvo IT Belgium. The company's support system comprises of three levels: The first line operates as a common help desk / service desk. The Second line comprises of specialized functional teams within a higher organizational line. The third line is a team of specific product or technical experts and is also within a higher organizational line. The company provided a dataset (Steeman, 2013) from its information system that supports the incidents management for the 2013 edition of the BPI challenge. The suggested actions to reach business understanding are enumerated in the following subsections.

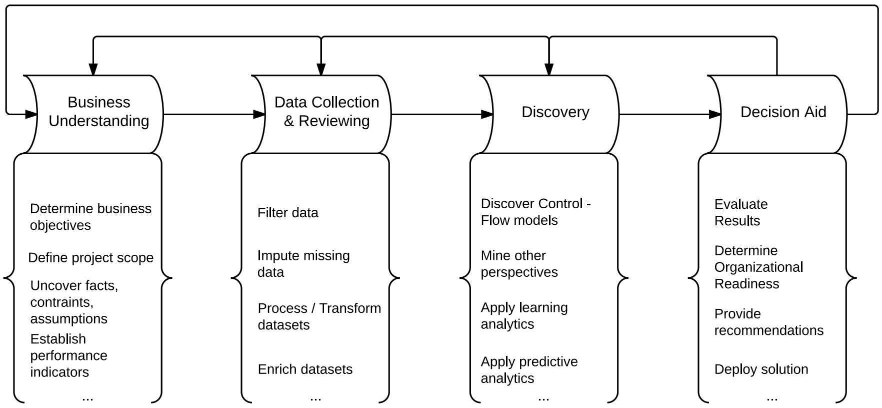


Fig. 1: The proposed methodology: Phases and actions

3.1 Determine Business Objectives and Define Process Scope

Determining business objectives implies informally describing the problem to be solved, specifying all business questions as precisely as possible and pointing out expected benefits in business terms (Chapman et al, 2000). In the case study, the primary goal of the incident management process is restoring a customer's normal service operation as quickly as possible when incidents arise ensuring that the best possible levels of service quality and availability are maintained. So, the pertinent business question is if there are any particular patterns that delay the resolution of issues. Since the quick resolution of the issue is defined within the Service Level Agreement of the company, there are evident gains in settling tactics to avoid those patterns.

Defining process scope deals with acknowledging what parts of the process can be tracked through the logged data and type of information is both available and useful (van der Heijden, 2012). Concerning the case study, each record contains a number of variables such as the unique ticket number of the service request, the impact of the case (a measure of the business criticality of the incident), the case status (queued, accepted, completed or closed) and sub-status (assigned, awaiting assignment, cancelled, closed, in progress, wait or unmatched), the business area of the user reporting the incident, the technology-wise division of the organization, the support team that will try to respond to the service request and the location that takes the ownership of the support team.

The process is roughly the following: A customer submits a service request. The process reactively triggers a "first line" response, in other words, the Service Desk or the Expert Help Desk tries to resolve the issue. When this is not possible, the case should be escalated to Second Line and/or Third Line teams. The dataset contains 65533 timestamped events related to the incident management process.

3.2 Uncover Facts, Constraints and Assumptions

There is an announced policy of the company that most of the incidents need to be resolved by the first line support teams (mainly service desks). This is called “Push to Front” tactic and it is mostly a matter of efficiency. Pushing to front, allows the 2nd and 3rd line support teams to focus on their special, more demanding tasks (usually not related to customer service support). Unless this tactic is consistently applied a lot of 'easy', big volume cases will end up in those lines.

As such, pushing to front is an important coordinated pattern that may arise during the process execution.

Besides pushing work towards the front, any team upon receiving a task can either try to resolve the issue by itself or hand over the task to another team (of the same or of another line). Handover of work is an ordinary action, however if this is excessively used it may have an inadmissible effect on process efficiency. Namely, extensive handover may reveal dodging or deferring behavior. The opposite (extensive takeover) may also reveal some undesired elements like lack of collaboration mentality of lack of knowledge transferring. Therefore, the inter-team handovers may also include coordinated patterns of (social) behavior.

A special case of handover of work is when support teams send the same case to each other again and again. We shall call this undesirable situation “Ping Pong”.

Ping Pong is also an undesirable coordinated behavior that may affect significantly the process performance.

3.3 Establish Performance Indicators

Every indicator should follow some basic requirements, like representativeness, simplicity and easiness of interpretation, feasible data collection, etc. (Franceschini et al, 2007). Generally, each indicator refers to a specific objective, that is to say a sort of reference point used as a basis of comparison. Indicators may originate from a global performance measurement system of the company, but it is recommended to define ad-hoc indicators for the specific process mining project. In this case study, there is a single performance indicator: resolution time. However, since the focus in on discovering patterns that affect the primary indicator, we shall define two secondary indicators: Push to Front and Ping Pong.

Push to Front is measured by a binary variable for each case (a case can either push to front or not) while for the overall performance measuring the percentage of cases that are pushing to front is enough. Since Push to Front is a desired behavior, the greater the percentage, the better for the enterprise. Ping Pong is measured by a numerical variable, because a case may have multiple Ping Pongs, and the amount of Ping Pongs undoubtedly affects the resolution time. We further discuss the need for a numerical scale in section 5.2.

4 Data Collection and Reviewing

This phase consists of data manipulation actions. It is a time-consuming phase that demands for data filtering, dealing with missing values, transforming representations of the variables and adding new information to the existing dataset. Luckily, the case study dataset has been preprocessed by its provider (Steeman, 2013) in a way that very few data manipulation actions were required. In particular: we did not apply any filters to data, and we did not face any missing values problems.

The dataset in its original format contains a list of timestamped events. It is quite hard to elicit patterns of behavior from within this format, since the sequencing of events and their aggregation per case are not exploited. Therefore, the leading step is to reach a process perspective for the dataset. Therefore, wecommitted data to process format following two different perspectives (and thus creating two different datasets)

1. Control flow-wise (trajectories of status / sub-status changes)
2. Social-wise (transactions among support teams or lines)

Finally, to enrich the dataset, we created an additional variable for the service line where every event is performed. This information was embedded within the Support Team variable, so we extracted the pertinent values from the original variable. If no value for the line was logged for a support team (ST), we assumed to be a 1st line ST. In case that a ST spread over more than one line, we used the most front one.

5 Discovery

5.1 Discover Control-flow

The control-flow perspective focuses on the control-flow, i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths expressed in some process notation (Van der Aalst, 2011, p.11), or in other words the goal is to answer the question "what does the actual process look like?" (Bozkaya et al, 2009). The intuition of a common automated process discovery algorithm is to scan the Event Log for sequencing patterns and then to try to aggregate them. However, many different concepts and techniques have been proposed. The interested reader is redirected to (Van der Aalst, 2011, ch. 5-6) for a relevant discussion.

For the specific case study, control-flow refers to how the status / sub-status of a case changes during its lifecycle. There are 13 distinct alternatives for the status / sub-status of a case (presented in Table 1). Although the set of activities (status changes) is small, we noticed that there are 2278 different variants of the same process (for a dataset of 7554 cases). Out these 2278 variants, just 88 have a frequency higher than 100, while the dominant variant represents just a 23% of total cases, a fact that confirms that the process environment is highly flexible.

Since there is no strict sequencing rule, discovering an exact behavior would not reflect the real situation, and would probably be of little importance. In general terms, cases go from some *Accepted* sub-status to either a *Completed* sub-status or to *Queued*. In the latter option, the case returns to an *Accepted* sub-status. A process map is depicted in Fig. 2, where some labels for performance measures are printed. In particular, the heavier the weight of an edge, the worst its performance. The illustration has been created using Disco® (Fluxicon, 2012) and it is a direct way to visualize the process' bottlenecks. The largest delays happen between Completed-Resolve and Completed-Closed (7.2 days), Accepted-Wait User and Completed-Resolve (5.3 days) and Accepted-Wait Implementation and Completed-Resolved (4.7 days). It is also interesting to note that there is a meantime of 4.3 days between the Completed-Closed status and the Accepted-In Progress status, a fact that indicates that some cases are closed only to be re-initiated after 4-5 days.

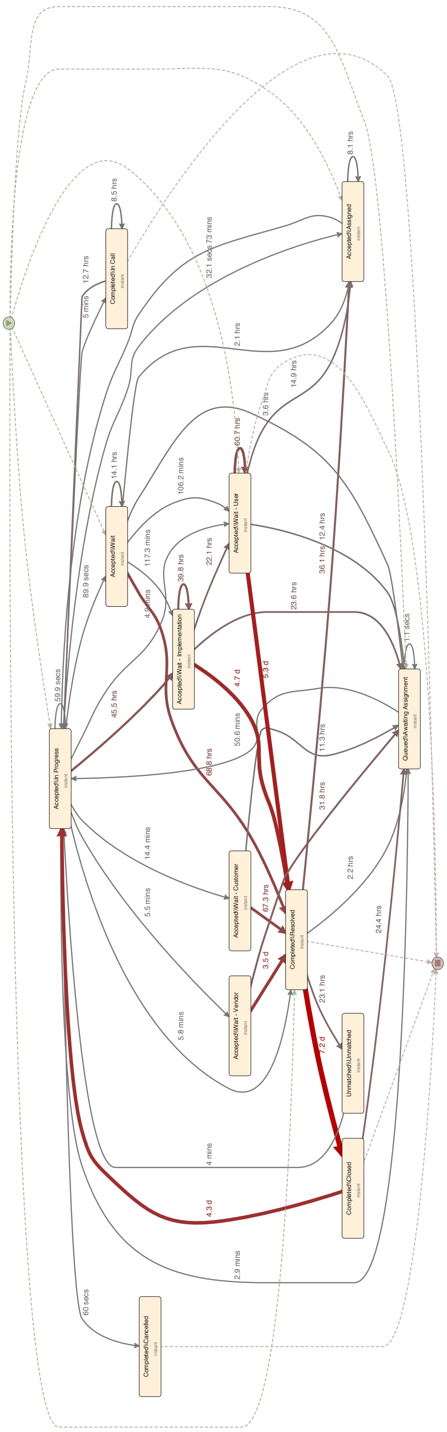


Fig. 2: Process Performance Map

Status	Sub-status
Accepted	Assigned, In Progress, Wait, Wait-User, Wait-Customer, Wait-Implementation, Wait Vendor
Queued	Awaiting Assignment
Completed	In Call, Resolved, Closed, Cancelled
Unmatched	Unmatched

Table 1: Status and Sub-status alternatives

5.2 Mine Other Perspectives

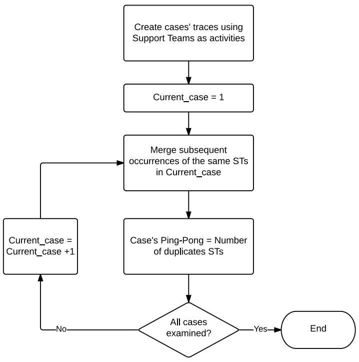
Besides control-flow, other common perspectives are the social or organizational perspective (which focuses on what actors are involved and how they are related) and the case perspective (which focuses on properties of cases). Concerning the case study, the patterns described in section 3.2 are social (organizational) patterns, therefore our focus is on mining that perspective.

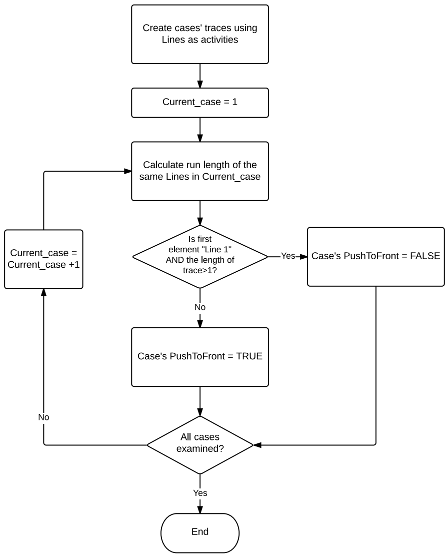
First of all, we need to evaluate the "Ping Pong" and the "Push to Front" patterns for each case, based on the descriptions of section 3.2. To this end, the algorithms illustrated in Fig. 3 were developed. The algorithms follow the definitions of the patterns. We recall that the definition of push to front in this paper refers to the case when the 1st line support teams can resolve the service request without interference of a 2nd or 3rd line support team. The definition of "Ping Pong" is that a Ping Pong occurs when a support team is revisited during the case, after it has passed the work to another team. However, we count a single Ping Pong per support team, even if this is revisited multiple times. This definition allows for a numeric representation of the Ping Pong behavior (a case may have multiple Ping Pongs, yet attributed to different teams).

Figure 4 illustrates these effects for the mainstream cases (outliers, i.e. cases that last more than 50000 minutes are removed). In particular, Fig. 4a depicts what is the difference in duration between cases that Push to Front and cases that do not. The drawn boxes are rectangles with edges defined by the lower and upper quartiles (25% and 75% respectively). The line inside the box is located at the median while values greater than 1.5 times of the upper quartile are presented as dots. It is clear that cases that do Push to Front are resolved quicker than cases that don't. While for *Push to Front* a binary variable is sufficient, for *Ping Pong* a numerical scale is preferred. An illustrative argument for this choice is presented in Fig. 5, where we see that it would not be fair to evaluate Ping Pong with a binary variable, since the number of Ping Pong has a strong effect on the process behavior. In this point we shall remind that a Ping Pong is assigned per team, i.e., even if a pair of teams handover their work multiple times during a case, that will still count for two (one for each team that is revisited). Fig. 4b plots a simple regression line between the duration (in minutes) of cases and the number of Ping Pongs they contain. As expected, both behaviors (Push to Front and Ping Pong) have a negative effect on the case duration.

5.3 Apply Learning Analytics

Learning analytics come in many shapes. Trying to profoundly epitomize we can name as learning analytics techniques that estimate relationships among variables, determine which variables are important in predicting future values, as well as techniques that

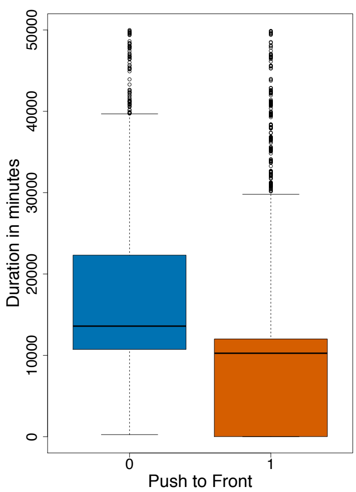


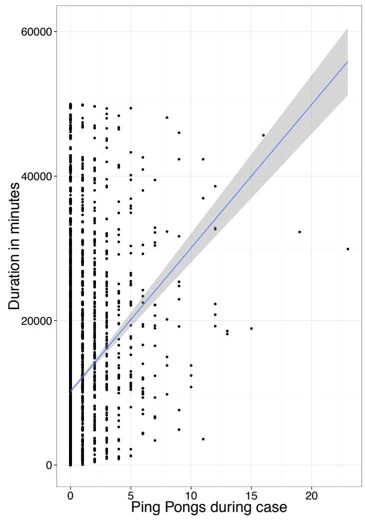


(a) Evaluating the Ping Pong behavior from the Event Log.

(b) Evaluating the Push to Front behavior from the Event Log

Fig. 3: Flow charts to evaluate the social patterns

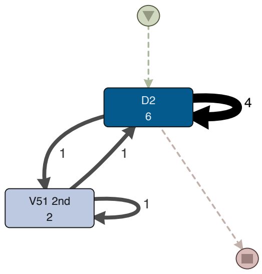




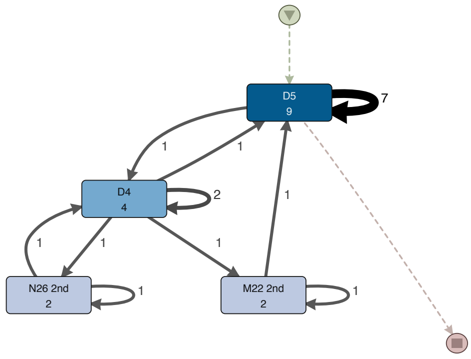
(a) Duration of cases that Push to Front (1) or not (0)

(b) Duration of cases over the number of Ping Pongs they contain

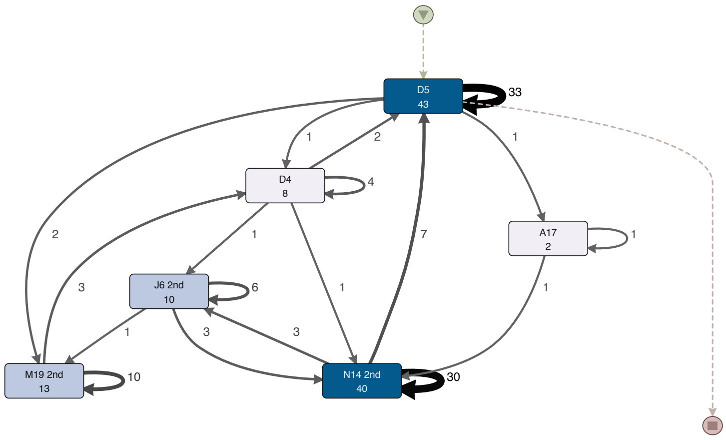
Fig. 4: The effect on case duration



(a) Single Ping Pong



(b) Double Ping Pong



(c) Multiple Ping Pong

Fig. 5: Nodes are Support Teams and arrows show the handovers of work. A numerical scale for the Ping Pong behavior is preferable.

segment the data into homogeneous groups. Casting this description to the case study context, we observe that identifying a set of important factors is a highly relevant question in the customer service field (Tseng and Huang, 2007). In this section, we propose techniques that would reach an answer when the focus is on a process view. The intuition of this task is to discover the features that have a great impact to the process flow and thus facilitate process improvement of reengineering by detecting, listing or classifying best practices (Reijers and Mansar, 2005). In particular, we perform a discrepancy analysis for the observed behavioral variation, as well as we try to assess the importance of factors that cause cases to Push to Front or to Ping Pong.

5.3.1 Discrepancy Analysis

Considering a case as an animating process, discrepancy measures the among-cases variability of the cases' life-cycle trajectories. Therefore, higher discrepancy, would reflect a greater level of uncertainty about the path followed by the cases. In this section, we integrate the sequence discrepancy analysis with the regression tree method introduced in (Studer et al, 2011). The intuition of this regression tree method is the following: Start with all cases grouped in an initial node. Then, recursively partition each node using values of another variable. At each node, the variable and the split are chosen in such a way that the resulting child nodes differ as much as possible from one another or have, more or less equivalently, lowest within-group discrepancy. The process is repeated on each new node until a certain stopping criterion is reached.

An apparent barrier to the application of the above method is that it is not straightforward to calculate the "mean" trace. Therefore, the discrepancy (variance) of the traces will be defined from their pairwise dissimilarities. Perhaps the most popular dissimilarity measure used for sequence analysis is the generalized Levenshtein distance. It is defined as the lowest cost of transforming one sequence into the other by means of state insertions–deletions and state substitutions. However, we still need to find a way to gauge the contribution of each instance to the overall variance. To this end, we exploit the generalization of the Ward criterion (Batagelj, 1988). In particular, Batagelj (1988) introduced the notion of a *gravity center* of a set of sequences and proposes a formula to calculate the distance of any sequence from it. This proposition allows the calculation of metrics like the sum of squares of these distances and the residual within the sum of squares. Based on this fact, and following the ANOVA mindset, Studer et al. (2011) introduced a metric to measure the part of the discrepancy that is explained by differences in group positioning (and they call it pseudo-R^2) and a metric to compare the explained discrepancy to the residual discrepancy (and they call it pseudo-F).

To build the regression tree, we use the pseudo-R^2 as a splitting criterion (we choose to split based on the variable that yields the highest R^2). As a stopping criterion, we trust the pseudo-F significance. In other words, we no-longer split a branch as soon as we get a non-significant F (considering a *p-value* of 0.05) for the selected split. For the implementation of this method, we used the TraMineR (Gabadinho et al, 2011) package of R.

We examined the role of just two predictors (Push to Front and Ping Pong) and as illustrated in Fig. 6, these two social patterns alone explain approximately30% of the total discrepancy. Both of them result in clustered behaviors. In particular, the first split is among cases that Ping Pong or not (0 and greater than 0). Cases of the later category (no Ping Pong) last significantly less and visit a lot less frequently the "*Queued*" status. At the second level, leftmost the split is among cases that Push to Front (>0) and not (0). We regard that cases that Push to Front reach a "*Completed*" status earlier, and that their average duration is smaller. The rightmost split is again based on the Ping Pong behavior, but this time the critical value is two. Cases that Ping Pong more than twice spend an important percentage of their lifetime in a "*Queued*" status, and are naturally prolonged.

5.3.2 Detecting the Factors' Importance

In the previous section we elaborated on learning the role of Ping Pong and Push to Front to the variation of the process. In this paragraph, we try to discover what

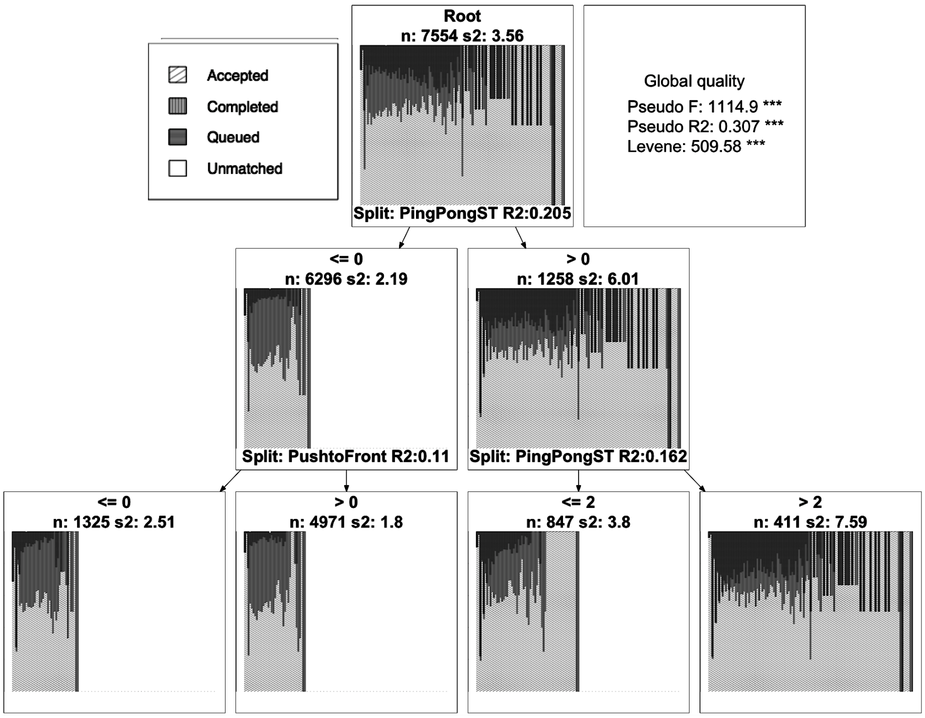


Fig. 6: Discrepancy Analysis for cases lifecycle trajectories

are the factors that affect these coordinated patterns. To this end, we propose to use a tree-based method. The basic reasons that favored this choice is that tree-based methods are more easily interpreted by non-experts. However, a drawback of trees is their accuracy level, since they suffer from high variance. Therefore, in order to create a more powerful and robust model, we propose to use Random Forests (RF) (Breiman, 2001). The basic idea of is that they grow a number of decision trees on bootstrapped training samples. During the creation of every tree, and every time a split is considered, a random set of characteristics (predictors) is used. There are a number of reasons why Random Forests are expected to deliver better results. First of all, since the new dataset is only a subset, it is likely that the number of records it contains is small. RF are more suitable for this kind of problems (small number of records with respect to the number of predictors). Then, by considering different characteristics for every split, RF can deal with high-order interactions and correlated characteristics (Strobl et al, 2008). Moreover, through RF, it is possible to obtain a summary of the importance of each characteristic (how significant it is for the branching decisions) using the Gini index (for classification trees) or the RSS (for regression trees) .

More specifically, since we have a numerical scale for Ping Pong, we grow a regression tree, using seven factors as predictors: the existence of the *Push to Fron*t behavior, the Support Team (*ST*) where the case was initiated (we kept just the top 30 of STs, using “other” for the rest), the *Country* of the ST, the code of the *Product* (again we kept just the top 30 ones), the *Organization Line* where the case was initiated, the *Impact* of the incident, and the *Line* of the ST (1st, 2nd or 3rd). We measure the im-

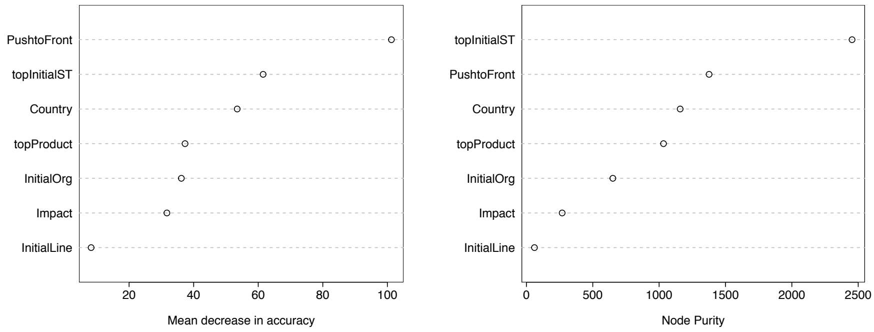


Fig. 7: Variable importance with respect to the Ping Pong behavior

portance of factors with two ways: The first calculates the mean decrease of accuracy in predictions when the corresponding variable is left out of the model (leftmost case of Figure 7) and the second considers the total decrease in node impurity that results from splits over the corresponding variables. As Figure 7 shows, the top three factors that lead to extended Ping Pong is the *Push to Front* tactic, the ST where the incident handling was initiated and its Country.

As far as it concerns the *Push to Front* behavior, we grow a Random Forest of classification trees to assess the variables' importance. A typical way to evaluate the importance is the Gini index, which is actually a measure of total variance across all classes. Nevertheless, the Gini index has been criticized as a tool to asses the importance of characteristics because it is biased in favor of continuous variables and variables with many categories. Therefore, we follow a permutation scheme, as proposed in (Strobl et al, 2007). The basic idea of employing permutation tests, is that if the variable is not important (the null hypothesis), then rearranging the values of that variable will not degrade prediction accuracy. Following this method (which yields an accuracy of 89.54%) *Country* appears to be the most important factor, followed by the *Product* code.

In the previous sections (5.3.1,5.3.2) we applied techniques to fathom the process behavior. The same techniques can be used to predict future outcomes and trends (Predictive Analytics). The most common situations are predicting group membership, predicting a future value, and predict relevant conditions. However, this kind of analysis is not included in this work, due to lack of relevant data.

6 Decision Aid

Decision aid is about "providing decision makers with the most favorable conditions possible for the type of behavior which will increase coherence between the evolution of the process, on the one hand, and the goals and/or systems of values within which these actors operate on the other" (Roy, 1994). There are some basic actions contributing to this mission. We shall dedicate the following paragraphs to briefly describe them.

6.1 Evaluate Results

Evaluating results involves reviewing them to determine whether they are still tied to the original questions, whether they meet the business objectives, and assess the business value they deliver. In case that the business objectives are not met, the analysts should report the reasons of the decline.

Considering the primary objective (resolution time), results are clearly targeted, revealing either paths that slow down the process (Fig. 2), or correlations of resolution time with behavioral patterns (Fig. 4), or even a view of variation of the life-cycles (Fig. 6). Concerning the secondary objectives, with respect to Ping Pong, results suggest an important finding, since now the company can focus on specific factors (Push to Front, Support Team, and Country) and look for assignable causes. As long as for the Push to Front pattern, the knowledge gain is again important, since by focusing on the *Product* factor we can spot the products that display strong Push to Front behavior and the ones that don't. Based on these results, we can provide perceptive recommendations (described in the next paragraph).

6.2 Determine Organizational Readiness and Provide Recommendations

Effective assessment of the organization readiness should result a smooth transition and increased user satisfaction with the proposed changes. Unfortunately, for this case study, we did not have access to relevant information (e.g., are process stakeholders willing to reinforce and reward positive teamwork behaviors? Are they willing to allow time for personnel to attend training?), therefore this step is skipped, while the following recommendations are not aware of any corporate particularities (probably existent, yet not available).

Following the evidence of the previous sections, and to deal with the Ping Pong effect, we shall recommend focusing on STs. This way we can detect that 80% of the total Ping Pongs is due to less than 5% of the STs. STs with extended Ping Pong behavior can thus be identified, enabling the company to take perceptive actions. By focusing on the *Country* factor, we regard that the largest average of Ping-Pongs per case belongs to the Netherlands (4.7 per incident) or that most Ping-Pongs happen within Belgium (with an average of 1.66 per incident). This knowledge facilitates the company to go deeper and look for the reasons that these specific countries are prone to Ping-Pong.

Moreover, to stimulate the Push to Front behavior, a possible response policy could be to assign the latter directly to other lines, or to train the service desk (1st Line) specific for these products, or even to create a knowledge sharing mechanism that will capture solutions specific to those product codes. Focusing on the *Country* factor, we regard for instance that Poland and USA are countries that are Pushing to Front while India has the worst performance. This piece of information should make the company aware and drive it to search concretely for the reasons (e.g., is it a matter of poor training or cultural differences?). Assuming, that we are willing to trade-off accuracy performance for more direct interpretation, it is possible to grow a single classification tree and get a number of 'rules' that can classify / predict Push to Front. Such an output would allow for rule-based process monitoring and support the timely investigation of undesired patterns (Caron et al, 2013).

6.3 Deployment

Deployment for this methodology has the meaning of providing decision aid by participating in the final decision legitimization (Roy and Damart, 2002). In particular, the analyst should be able to enlighten and scientifically accompany decision-making notably (Roy, 1993):

- by making the objective stand out more clearly from the less objective (the entire methodology has an evidence-based mentality)
- by separating robust from fragile conclusions (we applied a robust classification technique to explain the factors affecting the behavior and to deliver a predictive model for undesired behaviors)
- by avoiding the pitfall of illusory reasoning and by emphasizing incontrovertible results (the effect of every variable can be pointedly exhibited)

7 Conclusions

In this work we presented a dedicated approach based on process mining to guide the implementation of process analytics projects. We explored a real case study with the goal to provide insights to this implicit business process and to raise the capability of the company to handle service requests. This work demonstrated that a process perspective generates knowledge gains since ordinary data analysis methods may miss salient information of event based data sets. Our methodology was capable to detect how some social-wise patterns (behaviors) are related with performance and provide insights about the factors that shape these behaviors. Ultimately, the proposed methodology exemplifies how business decisions and process analysis can benefit from the analytical capabilities of a process mining approach.

We avoided to convey the methodology as a standard framework, since the following limitations are acknowledged: The phases include sets of generic actions. Such actions do not necessarily suggest specific techniques. For example, we do not make any particular recommendations about the diagnostic techniques for determining business objectives or defining the project's scope. Likewise, we do not provide recommendations about selecting specific process mining techniques. Yet, that would be a very interesting step for future improvement, since the usefulness of a mining technique decidedly depends on the available data, domain knowledge, expertise, business culture, and the objectives of the project. Moreover, the methodology does not include any mapping technique to match the deliverables with the business objectives, namely to assure that the delivered analytics best support the project's goals. It is out of the scope of this work, but very relevant for future enhancements to add any monitor and control over time functions for the effectiveness and efficiency of the solutions . Finally, the methodology evolved and has been validated through a case study. Future work would target additional validation methods. One way is to collect evidence and to check for realizations of the indicated actions in other applications. A different way is to rely on expert judgement for a more qualitative evaluation to competently meet the challenges of process analytics projects.

Acknowledgement

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

Van der Aalst WMP (1996) Petri net based scheduling. Operations-Research-Spektrum 18(4):219–229

Van der Aalst WMP (2011) Process mining: discovery, conformance and enhancement of business processes, 1st edn. Springer, New York

Van der Aalst WMP, Adriansyah A, Medeiros AKA, Arcieri F, Baier T, Blickle T, Bose JC, Brand P, Brandtjen R, Buijs J, Burattin A, Carmona J, Castellanos M, Claes J, Cook J, Costantini N, Curbera F, Damiani E, Leoni M, Delias P, Dongen BF, Dumas M, Dustdar S, Fahland D, Ferreira DR, Gaaloul W, Geffen F, Goel S, Günther C, Guzzo A, Harmon P, Hofstede A, Hoogland J, Ingvaldsen JE, Kato K, Kuhn R, Kumar A, Rosa M, Maggi F, Malerba D, Mans RS, Manuel A, McCreesh M, Mello P, Mendling J, Montali M, Motahari-Nezhad HR, Muehlen M, Munoz-Gama J, Pontieri L, Ribeiro J, Rozinat A, Seguel Pérez H, Seguel Pérez R, Sepúlveda M, Sinur J, Soffer P, Song M, Sperduti A, Stilo G, Stoel C, Swenson K, Talamo M, Tan W, Turner C, Vanthienen J, Varvaressos G, Verbeek E, Verdonk M, Vigo R, Wang J, Weber B, Weidlich M, Weijters T, Wen L, Westergaard M, Wynn M (2012) Process mining manifesto. In: Van der Aalst WMP, Mylopoulos J, Rosemann M, Shaw MJ, Szyperski C, Daniel F, Barkaoui K, Dustdar S (eds) Business Process Management Workshops, vol 99, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 169–194, URL http://www.springerlink.com/index/10.1007/978-3-642-28108-2_19

Batagelj V (1988) Generalized ward and related clustering problems. Classification and related methods of data analysis pp 67–74

Bozkaya M, Gabriels J, Werf JMvd (2009) Process Diagnostics: A Method Based on Process Mining. IEEE, pp 22–27, DOI 10.1109/eKNOW.2009.29, URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4782560

Breiman L (2001) Random Forests. Machine Learning 45(1):5–32

Caron F, Vanthienen J, Baesens B (2013) Comprehensive rule-based compliance checking and risk management with process mining. Decision Support Systems 54(3):1357 – 1369

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) CRISP-DM 1.0 Step-by-step data mining guide. Tech. rep., The CRISP-DM consortium, URL http://www.crisp-dm.org/CRISPWP-0800.pdf

De Weerdt J, vanden Broucke S, Vanthienen J, Baesens B (2012) Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes. In: Evolutionary Computation (CEC), 2012 IEEE Congress on, pp 1–8, DOI 10.1109/CEC.2012.6256459

Delias P, Doumpos M, Grigoroudis E, Manolitzas P, Matsatsinis N (2015) Supporting healthcare management decisions via robust clustering of event logs. Knowledge-

Based Systems DOI 10.1016/j.knosys.2015.04.012, URL http://linkinghub.elsevier.com/retrieve/pii/S0950705115001501

van Dongen B, Weber B, Ferreira D, De Weerdt J (eds) (2013) Business Process Intelligence Challenge 2013, vol 1052, CEUR

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From Data Mining to Knowledge Discovery in Databases. AI Magazine 17(3):37–54, DOI http://dx.doi.org/10.1609/aimag.v17i3.1230

Ferreira DR, Mira da Silva M (2008) Using process mining for ITIL assessment: a case study with incident management. Bournemouth University

Fluxicon (2012) Disco. Fluxicon, URL http://www.fluxicon.com/disco/

Franceschini F, Galetto M, Maisano D (2007) Management by measurement: designing key indicators and performance measurement systems. Springer, Berlin ; New York

Gabadinho A, Ritschard G, Müller NS, Studer M (2011) Analyzing and visualizing state sequences in r with traminer. Journal of Statistical Software 40(4):1–37

Grance T, Kent K, Kim B (2004) Computer security incident handling guide. NIST Special Publication pp 800–61

van der Heijden T (2012) Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice. Master Thesis, Technische Universiteit Eindhoven, Eindhoven, URL http://alexandria.tue.nl/extra2/afstversl/tm/Van_der_Heijden_2012.pdf

Hochstein A, Zarnekow R, Brenner W (2005) Itil as common practice reference model for it service management: formal assessment and implications for practice. In: e-Technology, e-Commerce and e-Service, 2005. EEE '05. Proceedings. The 2005 IEEE International Conference on, pp 704–710, DOI 10.1109/EEE.2005.86

Lin WT, Kao TWD (2014) The partial adjustment valuation approach with dynamic and variable speeds of adjustment to evaluating and measuring the business value of information technology. European Journal of Operational Research 238(1):208 – 220

Matignon R (2007) Data mining using SAS Enterprise miner. Wiley-Interscience, Hoboken, NJ

Rebuge Á, Ferreira DR (2012) Business process analysis in healthcare environments: A methodology based on process mining. Information Systems 37(2):99–116, DOI 10.1016/j.is.2011.01.003, URL http://linkinghub.elsevier.com/retrieve/pii/S0306437911000044

Reijers H, Mansar SL (2005) Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. Omega 33(4):283 – 306

Roy B (1993) Decision science or decision-aid science? European Journal of Operational Research 66(2):184–203, DOI 10.1016/0377-2217(93)90312-B, URL http://linkinghub.elsevier.com/retrieve/pii/037722179390312B

Roy B (1994) On operational research and decision aid. European Journal of Operational Research 73(1):23–26, DOI 10.1016/0377-2217(94)90136-8, URL http://linkinghub.elsevier.com/retrieve/pii/0377221794901368

Roy B, Damart S (2002) L' analyse coûts-avantages, outil de concertation et de légitimation? Metropolis 108/109:7–16

Steeman W (2013) Bpi challenge 2013, incidents. DOI 10.4121/uuid:500573e6-accc-4b0c-9576-aa5468b10cee, URL http://dx.doi.org/10.4121/uuid:500573e6-accc-4b0c-9576-aa5468b10cee

Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. BMC bioinformatics 8(1):25

Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC bioinformatics 9(1):307

Studer M, Ritschard G, Gabadinho A, Müller NS (2011) Discrepancy analysis of state sequences. Sociological Methods and Research 40(3):471–510, DOI http://dx.doi.org/10.1177/0049124111415372

Tarantilis C, Kiranoudis C, Theodorakopoulos N (2008) A web-based {ERP} system for business services and supply chain management: Application to real-world process scheduling. European Journal of Operational Research 187(3):1310 – 1326

Tseng TLB, Huang CC (2007) Rough set-based approach to feature selection in customer relationship management. Omega 35(4):365–383, DOI 10.1016/j.omega.2005.07.006, URL http://linkinghub.elsevier.com/retrieve/pii/S0305048305000964

Wilson A, Zeithaml VA, Bitner MJ, Gremler DD (2012) Services marketing: Integrating customer focus across the firm. 2nd Eu, McGraw Hill