A METHOD FOR DISCOVERING THE RELEVANCE OF EXTERNAL CONTEXT VARIABLES TO BUSINESS PROCESSES

Eduardo Costa Ramos, Flavia Maria Santoro and Fernanda Baião
NP2Tec, Department of Applied Informatics
Federal University of the State of Rio de Janeiro (UNIRIO), Rio de Janeiro, Brazil

Keywords: Business Process, External Context, Knowledge Management, Competitive Intelligence, KDD.

Abstract: Organizations have been demanded to efficiently detect and respond to changes in their environment, which depends on its ability to adapt their business processes. Taking internal and external environment variables into account enables to address issues, such as, how a business process was executed last time the country experienced a similar economic scenario; whether that process execution brought positive results or not; which were the external environmental reasons that provoked changes in previous process executions. These environmental variables are typically referred in the literature as the context of the process. In this paper, we propose a method to identify and prioritize external variables that impact the execution of specific activities of a process. The proposed method applies competitive intelligence concepts and data mining techniques, and was evaluated in a case study.

1 INTRODUCTION

Organizations are pressured to quickly detect and respond to changes in their environment, which may include issues about social, political, economical or technological areas. This fast adaptation depends on its ability to use both internal and external information about the environment and adapt itself to changes and other contingencies imposed. Such disruptions in the routine should be reflected in business processes (Recker and Rosemann, 2006). Knowledge Management and Competitive Intelligence approaches can be used in this direction (Jung et al., 2006).

Both Knowledge Management (KM) and Competitive Intelligence (CI) focus on the strategic organization goals. While CI focuses on the outside, monitoring and internalizing information from the external environment, KM encodes, shares and uses knowledge generated and stored internally in the organization. Taking internal and external environment variables into account enables the organization to address important questions such as how a business process was executed last time the country experienced a similar economic scenario; whether that process execution brought positive results or not; which were the external environmental reasons that posed changes in previous process executions. Those environmental variables are typically referred in the literature as the context of the process.

Context is defined as any information that can be used to characterize the situation of an entity (Dey, 2001). In a business process scenario, context is the minimum set of variables containing all relevant information impacting the design and implementation of a business process. Context information could be associated to any process element, such as activities, events, or actors. Furthermore, its analysis should provide insights to identify problems and learn with the past, besides helping to make decisions.

However, manipulating all stored organizational knowledge, as well as environmental external information, requires the application of knowledge discovery techniques so as to automatically handle and extract patterns from it. In this regard, Liebowitz (2003) proposed a set of frameworks to help a project manager in conceptualizing and implementing knowledge management initiatives, and poses some important questions that need to be addressed: (i) how knowledge discovery techniques can be applied for mining Knowledge bases; (ii) how is Knowledge originating from outside a unit evaluated for internal use?; (iii) does lack of a shared context inhibit the adoption of knowledge

Costa Ramos E., Maria Santoro F. and Baião F..
A METHOD FOR DISCOVERING THE RELEVANCE OF EXTERNAL CONTEXT VARIABLES TO BUSINESS PROCESSES.
DOI: 10.5220/0003668603990408
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (RDBPM-2011), pages 399-408
ISBN: 978-989-8425-81-2

originating from outside a unit?; (iv) How much context needs to be included in knowledge storing to ensure effective interpretation and application?

Although there are a few proposals that deal with context associated to business process (Nunes et al., 2009); (Rosemann et al., 2008); (Saidani and Nurcan, 2007), defining the relevance of external information for the execution of a process in an organization is still a challenge.

We propose a method to identify and prioritize external variables that impact the execution of specific activities of a process. The proposed method applies Competitive Intelligence concepts and data mining techniques (feature selection and decision trees). We have evaluated the method in a case study, which showed how the discovered variables influenced specific activities of the process.

This paper is structured as follows: Section 2 defines context and KM concepts, and presents related work. Section 3 details the proposed method, which was applied to a case study explained in Section 4. Section 5 concludes this work and points to promising evolutions of it.

2 RELATED WORK ON CONTEXT-AWARE PROCESS

The concept of context has recently revealed its relevance in business process management area. Identifying, documenting and analyzing contextual issues might help to make clear how changes in the environmental setting of an organization should lead to adaptations in processes. Literature points to the importance of considering contextual information, both in the design of business processes; and also, throughout process instances execution. As a result, an important issue should be identifying contextual elements that impact the process.

A taxonomy for context, described by Saidani and Nurcan (2007), which is composed of the most usual contextual information (location, time, resource and organization) aims at supporting context elicitation. Nunes et al. (2009) also presented a model for context to support knowledge management within the scenario of a business process. The model developed by these authors is an ontology that establishes a representation for context elements associated with process activities. Based on this model, process instances and their context are stored and further could be re-used. The types of context elements presented are: (i) information that exist during the execution of an activity (time, artifacts), (ii) information about individuals or groups that perform an activity, (iii) information to spell out the interaction between individuals within the activity performed. Both proposals do not provide explicit methods for context elicitation and neither consider external environment context.

Rosemann et al. (2008) integrate context in process modeling and define a meta-model concerned to the structure of a process, its goals, and context. They also describe a context framework where diverse context levels are depicted in layers, and a procedure to use it: (i) identify process goals; (ii) decompose process, (iii) determine relevance of context, (iv) identify contextual elements, (v) type context. Our research is directly related to the detailing of step 4 as an evidence-based task.

Another approach for bringing out context is stated by Soffer et al (2010) with the goal of learning and gradually improving business processes considering three elements: process paths, context and goals. Similar to our work, they argue that the success of a process instance can be affected not only by the actual path performed, but also by environmental conditions, not controlled by the process. Their work is based on an experience base, including data of past process instances: actual path, achieved outcome, and context information.

We propose context identification to be handled at the activity level, thus enabling process stakeholders to dynamically interfere into a specific activity result by applying previously acquired knowledge during the execution of a process. The circumstances are defined according to the external environment. External contingencies can be considered as opportunities or constraints that influence the structure and internal processes of organizations, according to Competitive Intelligence initiatives (Jung et al., 2006). The CI implementation cycles generally include steps to identify information that should be collected. Therefore, based on (Jung et al., 2006); (Kimball and Ross, 2002); (Cook and Cook, 2000); (Herring, 1999); (Ramos et al., 2010) described the CI process cycle steps to support a Context-based KM Model

The first step is to identify process, therefore key business processes are chosen from goals and organization strategy. Then, external variables should be identified and represented and associated to the process model through a Bus Matrix (Kimball and Ross, 2002). After that step, it is possible to start collecting and keeping these information through properly sources (databases, sensors, etc.). All information is stored in a repository called Organizational Memory, and a number of techniques (KDD, inferences) are applied in order to search for evidences of their impact in process instances. This might result in scenarios and recommendations, which might improve the process, either at the instance or at the model level. The process manager is able to make decisions based on that outcomes; it could possibly cause process adaptations. Then, the cycle starts in on again.

The problem addressed in this paper is specifically related to steps 2 and 3 from this cycle. Next section describes a method to identify the external context, or the kind of information that generally cannot be captured in transactional systems, but from outside of the organization.

3 A METHOD FOR DISCOVERING EXTERNAL CONTEXT

In order to capture and use context information, it is first necessary to specify which context information has to be handled by the organization (Nunes et al., 2009). We propose a method to discover external context variables (Figure 1) that may not be part of the organizational memory elements, but can be very relevant to the organization in achieving its process goals. This method also identifies which specific activities and process outcomes are impacted by the external context variables. Once discovered, the intelligence analyst may retrieve and analyze external context variables to define scenarios and recommend actions for decision-makers. The decision-makers evaluate the previous decisions and make new decisions that can reflect on improving, creating or removing processes.

There are several methods related to the definition of information needs, e.g., questionnaire, interview and observation that are widely used in different contexts (Vuori and Pirttimäki, 2005). However, the most suitable methods for the definition of information at the strategic level used by competitive intelligence are Key Intelligence Topics (KIT) (Herring, 1999) and Critical Success Factors (CSF). The use of a systematized or formal “management-needs identification process” is a proven way to accomplish this task (Herring, 1999). Key Intelligence Topic (KIT) support specification, definition and prioritization of information needs at the strategic level of the organization. KITs are items that must be constantly monitored to guarantee business success. They should be more detailed in the form of KIQs (Key Information Questions), which are items that specify the contents of each KIT. For example, the KIT “Strategic Investment Decisions” may consist of the following KIQs: "What is the involvement of other investors in competitors?" and "What are the critical investments from competitors?" (Vuori and Pirttimäki, 2005).

The KITs are identified through interviews with managers, asking open questions. They fall into three categories: (i) strategic decisions and actions; (ii) topics for early warning, considering threats and issues on which decision makers do not want to be surprised, and (iii) major players in the market, such as customers, competitors, suppliers and partners (Herring 1999). The technique also proposes the concept of surveillance areas, which are macroeconomic variables that impact the business sector, and that should be monitored.

The method steps are described as follows.

Step 1 – Identify Process Goal(s). Identify the goal related to a given process and their appropriate measures (Rosemann et al, 2008). Repeat this step to identify others goals after concluding the last step.

Step 2 – Select KIT Category. Herring (1999) has divided KITs into three categories: 1) Strategic Decisions and Issues, 2) Early-warning KITs, considering threats and issues on which decision makers do not want to be surprised and 3) Key player KITs (such as customers, competitors,

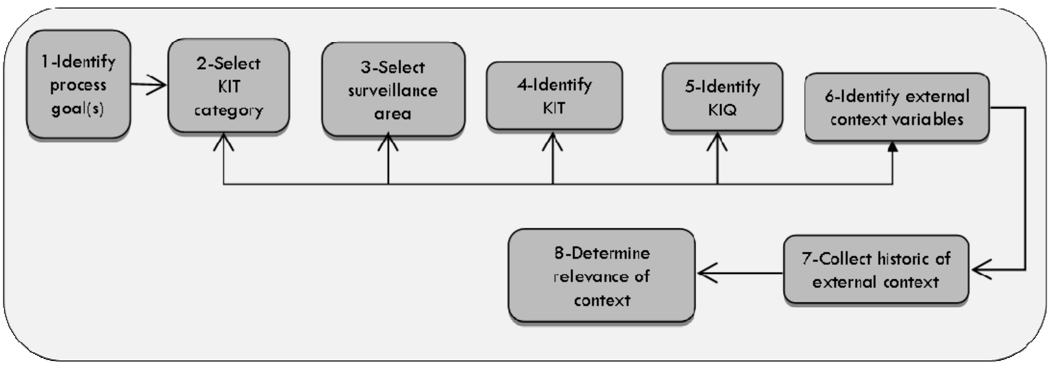

Figure 1: Method for external context variables identification.

suppliers and partners).

Step 3 – Select Surveillance Area. To define the external context variables, the steps 3 to 6 are part of a top-down approach. Top level areas must be considered to give support to the next step. A model to categorize context information would help to select those areas. The areas can be selected from any framework or a combination of them, such as Five Forces model (Porter, 1979), or SLEPT or STEEP Analysis (The Times, 2010). In general, they are: social, technology, economic, ecology, political, legal and competitors, due to all industries are influenced by them. These forces are continually in a state of change and then should be scanned. Most research about context in business process deal with internal context, i.e. process attributes inherent to the way process is performed, to the organization of activities and internal rules. Few context categories are proposed, such as location, time, and organization environment. Our work focuses on the events that occur externally to the process, or ultimately to the organization where it runs, but somehow interfere within this process, provoking good or bad effects. There are not many proposals to categorize this kind of context information. Rosemann et al. (2008) propose that the external layer of their model is composed of the following types of context: suppliers, capital providers, workforce, partners, customers, lobbies, states, competitors. Repeat this step for each of the three KIT categories.

Step 4 – Identify KIT. Key Intelligence Topics (KITs) are identified by interviewing the key decision-makers and asking them open-ended, non-directive questions (Herring and Francis, 1999). An interview protocol can be very useful to ensure the consistency of results (Herring, 1999). Repeat this step for each of the surveillance area selected.

Step 5 – Identify KIQ. Key Intelligence Questions (KIQs) should be identified for each KIT. KIQs represent the information needs listed in the KIT, i.e. what the manager needs to know to be able to make the decisions. It is possible to have the same KIQ for more than one KIT. Repeat this step for each KIT selected.

Step 6 – Identify External Context Variables. Each KIQ may reference one or more external variables. These are the external context variables and are identified in this step. It is possible to have the same variable for more than one KIQ. Repeat this step for each KIQ identified in the previous step. For each process goal, the result of all the executions of steps 2 to 6 will be the final Intelligence Tree with the following columns: Process Goal, KIT category, Surveillance Area, KIT, KIQ and External Context Variable.

Step 7 – Collect Past Information of the External Context. In this step, the historic of the external context is collected and stored in the organizational memory.

Step 8 – Determine Relevance of the External Context to the Process outcomes and to the Process Activities Outcomes. It is not feasible to store all context information that could form part of the Organization Memory. That's is why, this step helps prioritizing which context to capture and store, by classifying the variables by relevance using data mining. This step follows the KDD process of Fayyad et al (1996) that is interactive and iterative, involving numerous steps with many decisions made by the user. The term Knowledge Discovery in Databases (KDD) is generally used to refer to the overall process of discovering useful knowledge from data, where data mining is a particular step in this process (Fayyad, et al., 1996)

Several data mining problem types or analysis tasks are typically encountered during a data mining project. Depending on the desired outcome, several data analysis techniques with different goals may be applied successively to achieve a desired result (Jackson, 2002). Before applying the KDD process, it is necessary to develop an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the customer's viewpoint (Fayyad et al., 1996). Our method uses KDD for the following KDD goal: predict the process goal and determine the relevance of the external context to the process outcomes and to the process activities outcomes to achieve the process goal defined in step 1. The KDD process steps (Fayyad et al., 1996) are:

Step 8.1 (Selection) - this step consists on creating a target data set, or focusing on a subset of variables or data samples, on which discovery is to be performed. In this step, the historic of the external context is associated to the process activities outcomes and to the process execution results, for the same period.

Step 8.2 (Pre-processing) - this step consists on the target data cleaning and pre processing in order to obtain consistent data;

Step 8.3 (Transformation) - this step consists on data reduction and projection: finding useful features to represent the data depending on the goal of the task. With dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found (Fayyad et al., 1996).

Step 8.4 (Data Mining - DM) - this step consists on the searching for patterns of interest in a particular representational form, depending on the DM objective (usually, prediction). Many models can be created to allow comparing which one has the best accuracy for predicting the target attribute, in the case of prediction. The chosen model must easily show the relevant variables that must be scanned and what specific values may trigger some decisions.

Step 8.5 (Interpretation/Evaluation) - this step consists on the interpretation and evaluation of the mined patterns.

4 A CASE STUDY USING DATA FROM OPEN SOURCE PROJECTS

An explanatory case study was made in order to evaluate the method proposed. A case study was used in this research because it does not require control of behaviours events and because it focus on contemporaneous events (Yin, 2009). This research question is: “how to determine the relevance of variables of the external context to a business process?”.

4.1 Source Forge Software Development Process Model

We applied the approach in a scenario on the domain of Open Source Software Development. Figure 2 presents a process of Source Forge software development projects modeled with the Bizagi Process Modeler (Bizagi, 2011) using BPMN 1.2 notation (OMG 2010). In this software development process, the organizations may be interested in the information if new projects or existing ones will be concluded under the production or mature status, i.e., the organizations must make decisions such as: authorize or no the start of a software development project?; what to do to maximize the chances of an on going project to be concluded in the production or mature status?; when is it better to deactivate a project than continuing with it?

The software development process of Source Forge (SF) is not published formally by Source Forge, thus, we made some considerations in creating the process model in Figure 2, as for example, we considered only the projects that started in the Specify Requirements activity, despite there were others projects getting started in others activities.

Each project can be classified into one of six different levels, from the earliest stage of production to a fully developed software: planning, pre-alpha, alpha, beta, production stable and mature (Comino *et al.*, 2007). The process in Figure 2 was based on these status and on literature (PMI, 2008); (Madey, 2011). In the Authorize the start of the project activity, the decision maker, that can be a project manager for example, creates the project in SF; in

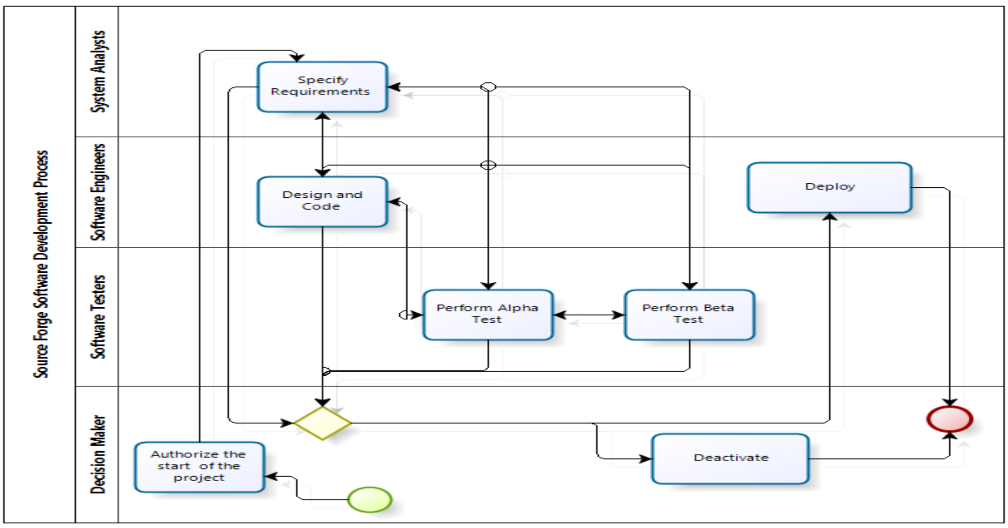

Figure 2: Source Forge Software Development Process Model.

the Specify Requirements activity the requirements are specified; in the Design and Code activity the software engineers design and develop the software, and perform the unit tests; in the Perform Alpha Test activity and in the Perform Beta Test activity, the software is tested; in the Deployment activity the official software is published to the users in the production or mature status, that is why we consider just one status: “production/mature”; in the Deactivate activity, the project is canceled temporarily or definitely by the decision maker.

4.2 The Data Set

The proposed method was applied to the Open Source (OS) projects from Source Forge projects database (Madey, 2011). SourceForge (SF) thrives on community collaboration to help creating the leading resource for open source software development and distribution. With the tools it provides, 2.7 million developers create software in over 260,000 projects. SF connects more than 46 million consumers with these open source projects and serves more than 2,000,000 downloads a day (Madey, 2011). SourceForge.net is the largest existing online platform providing OS developers with useful tools to control and manage software development. Project administrators register their software project on SF and provide the required information which is then available on-line (Comino et al. 2007).

The dataset we employed in our analysis consists of 1,087 OS projects that were hosted on SF and that had an English version and that got started after January 2005 at the “Specify Requirements” activity, and that achieved firstly one of the following activities before January 2011: “Deploy” or “Deactivate”. All the 1,087 projects are aligned with the process of Figure 2. This dataset has 1 dependent variable and 10 predictors pertaining to projects. These predictors consist of 1 process outcome and 9 process activities outcomes.

For each project, the binary outcome (dependent) variable “final status” is available and indicates whether the project achieved firstly the status of “production/mature” (good projects) or “inactive” (bad projects). This dataset contains 295 bad projects and 792 good projects. It means that 27% of the 1087 projects achieved the “final status” as inactive, and 73% of them, as production/mature. In addition, this dataset has also 9 process activities outcomes available for each project, describing the total duration of the project in each process activity and the percentage it represents of the project duration. The project duration is one process outcome and represents the duration of the project from the Specify Requirements activity to the first month of one of the following activities: Deploy or Deactivate. The duration is measured in quantity of months.

In our work, we introduce new variables of the external context and relate it to the process activities and to the process execution results to support these decisions.

4.3 Application of the Method

In this explanatory case study, we applied all the 8 steps of the proposed method to define relevant external variables that influenced the project conclusion of SF projects using the dataset detailed in section 4.2 and considering the software development process defined in section 4.1. The result after applying the steps 1 to 6 of the proposed method 1 is a list of possible relevant external variables. The result applying the steps 7 to 8 is a list showing just the relevant variables among the external contexts, the activities outcomes and the process outcomes; and a decision tree showing the relation among these relevant variables.

Step 1 – The goal “Conclude the software development in the Deploy activity” was considered for the process of Figure 2. This goal is achieved

Table 1: Part of the Final Intelligence Tree after all the executions of steps 2 to 6.

KIT category	Surveillance Area	KIT	KIQ	External Context Variable
Strategic decisions and actions	Economic	Economic recession	What are the predictions for IT investments of public and private organizations for the next years?	**IT Investment Prediction;**
			What are the predictions for the unemployment rate for next years?	**Unemployment Rate prediction; Unemployment Rate;**
			What are the predictions for the inflation rate for next years?	**Inflation Rate prediction; Inflation Rate;**
Strategic decisions and actions/ Early-warning	Politic	IT goals of the Govern	What are the Open Source Software patterns adopted by the Govern?	**Open Source Software patterns;**

when the dependent variable “final status” is production/mature.

Step 2 to 6 – For the defined process goal, the result of all the executions of steps 2 to 6 was a table similar to the Table 1. This table contains possible relevant external variables that can impact the process goal.

Step 7 – In this step the focus is on collecting the past information of the external variables defined previously. As the projects could be developed by people that lives in different countries anywhere in the planet, it was necessary to make a simplification assuming that the USA was the original country of every one involved in the 1,087 projects of the dataset. The USA was chosen because it is one of the most influential countries in the global economy, as we could see in the global economy crises of 2008 that got initiated in the USA. Another aspect to consider in this step is that sometimes it is not possible to collect the past information of all the external variables because, for example, it may not exist. In our research, we have collected the historic of 2 external variables defined previously: the USA unemployment rate and the USA inflation rate (IndexMundi, 2010).

Step 8 – In this step we followed the KDD process (Fayyad et al. 1996) and we applied the Feature Selection technique to show the variables relevance, and we used Decision Tree C&RT (Standard Classification Trees with Deployment) to show explicitly the rules of the relation between the relevant external contexts, the relevant process outcomes and the relevant process activities outcomes for predicting the dependent variable “final status”. This was the KDD goal.

Below, we explain how the data mining technique determined that Unemployment Rate was a relevant external context variable to the defined process goal and to one of its activity outcome. We used the STATISTICA Data Miner software (StatSoft, 2010) that uses the CRISP-DM process (CRoss-Industry Standard Process for Data Mining). According to Azevedo and Santos (2008) CRISP-DM can be viewed as an implementation of the KDD process of Fayyad et al (1996). KDD process steps:

Steps 8.1 (Selection) and **8.2 (Pre-processing)** - These 2 steps were some of the most time consuming steps, as Mack *et al.* (2005) already experienced. The data requirements for what is necessary as well as the data acquisition itself have been taken care of already with the data dump from SourceForge (SF). The output of the step 8.1 is the process log, the dataset that was detailed in section 4, and the output of the step 8.2 is a new dataset with the historic of the collected external contexts (step 7) associated to the process activities outcomes and process outcomes (step 8.1).

Step 8.3 (Transformation) – In this case study, we run the Feature Selection of STATISTICA Data Miner (StatSoft, 2010) to automatically find and rank important predictor variables for predicting the dependent variable “final status” that discriminates between good and bad projects, as shown in Figure 3. Feature Selection (FS) technique is “the process of reducing dimensionality by removing irrelevant and redundant features” (Guyon & Elisseeff, 2003 apud Refaeilzadeh et al., 2007)(Blum & Langley, 1997 apud Refaeilzadeh et al., 2007) reducing “the complexity of the problem, transforming the data set into a data set of lower dimensions” (Nisbet et al., 2009). Figure 3 shows that among the 12 variables of the dataset created in the last step, there are 5 that have a p-value of less than 0.01, i.e., that stand out as the most important predictors variables to determine whether a project would be finalized in the production/mature or in the inactive status.

Starting from the most relevant to the less relevant, these 5 variables are: 1-Project duration; 2-Specify requirements duration; 3-Inflation rate; 4-Unemployment rate; 5-Perform Beta Test Duration. Note that 2 of these relevant variables are process activities outcomes; 1 is a process outcome; and the third and the fourth most relevants variables are from the external context.

	Best predictors for categorical dependent var: status_final	
	Chi-square	p-value
Project duration	83,34835	0,000000
Specify requirements duration	75,96964	0,000000
Inflation rate	49,45075	0,000000
Unemployment rate	33,22903	0,000009
Perform beta test duration	18,63327	0,002249

Figure 3: Best predictors variables for categorical dependent status_final ordered top to bottom on basis of lowest p-value to highest (Stratified Random Sampling).

Step 8.4 (Data Mining) - Decision trees are powerful tools for classification and prediction. The decision tree C&RT (Standard Classification Trees with Deployment) of Figure 4 was run using STATISTICA Data Miner (StatSoft, 2010) considering the relevant variables found in the previous step. We used the V-fold cross validation and a 30% sample of dataset for testing to assess the accuracy of the model. Based on the 1087 projects of the full dataset, initially we used a training data sample to build the decision tree (training phase), then, a testing data sample to refine and evaluate the decision tree (testing phase), and finally, we used another dataset with different projects to re-evaluate the accuracy of the decision tree (re-evaluation phase).

In the training phase the decision tree had an error rate of 19.12%; in the testing phase, 21.53%; and in the re-evaluation phase, 20%. The error rate of 19.12% (training phase) means that the decision tree C&RT can predict correctly with an accuracy of 80.80% whether a project will be finalized in the production/mature or in the inactive status. The percent of correct predictions for the bad projects (final status = inactive) is 77.44%; and for the good projects (final status = production/mature) is 82.10%.

Step 8.5 (Interpretation/Evaluation) - The decision tree C&RT (Standard Classification Trees with Deployment) of Figure 4 show the relation between the relevant external contexts, the relevants process activities outcomes and the relevants process outcomes. This decision tree shows that the process outcome “Project duration” is related to the Perform beta test activity by its outcome “Perform beta test duration” and that these outcomes are related to the external context “Inflation rate”, as we can see in nodes 1, 2, 4, 6, 8 and 11. Node 11 clearly shows the relevance of the external variable to the Perform beta test activity. It evidences that, when the inflation rate raises below or equal 2.705 and greater than 1.67, then there is a higher probability of the projects, that have Project duration <=4.5 and Perform beta test duration <=0.5, to be deactivated, i.e., to be concluded as inactive.

The project manager or the decision maker can use the decision tree when, for example, he will decide to develop a new software project that will last less than 0.5 month in the “Perform beta test” activity, so he can see the estimate for the USA inflation rate when this project is supposed to be concluded. If this rate raises below or equal 2.705 and greater than 1.67, so there is a higher probability of this project to be concluded as inactive, i.e., the decision maker can decide not to start this project or he can make actions to maximize the chances of this project to be deployed and minimize the chances of it be inactive. This same scenario can happen with an on going project, that is why the relevant external contexts must be monitored because it may fire a change during the process execution or before the project start.

5 CONCLUSIONS

5.1 Analysis and Discussion

It is important to note that external variable relevance is discovered based on the process log. As

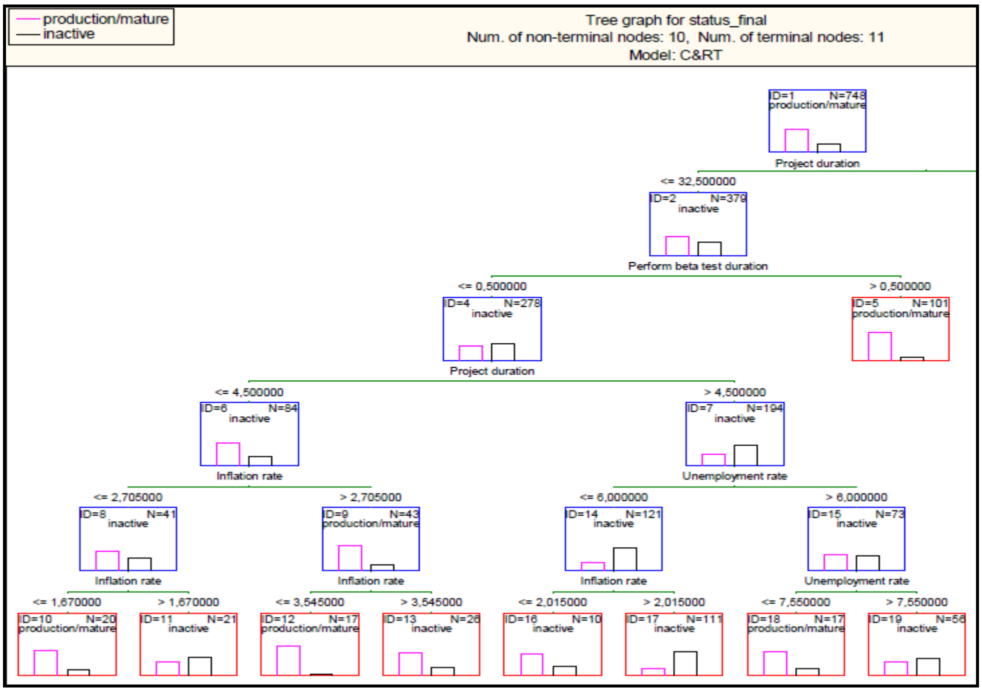

Figure 4: Part of the decision tree C&RT (training phase) for the SF projects dataset considering the best predictor variables to the dependent variable “final status”.

with any data mining approach, the discovered knowledge depends on the amount of detailed information available in the log. This is a limitation of the proposed method. Therefore, when our approach discovers that a specific external variable is not-so-relevant, it does not mean that it is not relevant at all; instead, it means that the process log did not include enough evidences pointing to the relevance of this external variable to the historical process log, when compared to other variables.

Therefore, it is important to take into account a process log with enough information to run our method and to consider other methods and the experience and feelings of the specialists and of the decision makers when deciding which external variables are relevant to be scanned. At least, it must contemplate the relevant variables found in our proposed method. Another limitation in the method is that transforming some KIQs into external variables may be very difficult, as well as collecting these variables.

In this explanatory case study, our goal was not to get the most relevant external variables that exist, but our goal was to confirm the relevance of the defined variables identified applying our proposed method. It explains why we could do some limitations in this case study, such as, interviewing people that were not involved in none of the 1087 projects of this dataset neither had experience in OS.

Our method differs from existing approaches in the literature (Rosemann *et al*., 2008); (Soffer *et al.*, 2010) since it suggests new external context variables that may not be part of the organizational memory and that can be very relevant to the organization achieve the process goals; and shows which specific process activities are impacted by the external context variables to the organization achieve the process goal.

5.2 Conclusion and Future Work

Successful organizations are those able to identify and answer appropriately to changes in their internal and external environments. The organizations´ decision makers need to make important decisions in order to carry this out.

In this paper we described a method for supporting the identification and prioritization of variables to be considered in the context of the external environment that impacts process execution. This method also shows which specific process activities are impacted by these variables to the organization achieve its process goals. An explanatory case study illustrated the application of our method in a software development process using real data from projects of SourceForge.net. This method is based on CI and data mining techniques and provides the process manager with a fact-based understanding on which are the most relevant external variables that really influenced previous process executions, among the several variables that could be taken into consideration unnecessarily. This case study showed that changes in relevant variables of the external context may fire a decision of the decision maker to quickly responding to these changes, by adapting the process specification, or creating other business rules to be followed by the business process.

As future work we suggest applying our proposed method: in others different scenarios, such as oil&gas and risk management; applying to larger samples of process log and with more variables; interviewing decision makers of the same process log organization. We also suggest refining the model evaluation of our method.

REFERENCES

Azevedo, A., Santos, M. F., 2008. KDD, Semma and CRISP-DM: A Parallel Overview, *European Conference Data Mining-IADIS*.

Comino, S., Manenti, F., & Parisi, M., 2007. From planning to mature: On the success of open source projects. *Research Policy*, 36(10), 1575-1586. Retrieved from http://www.scopus.com.

Crerie, R., 2009. A method for discovering of business rules by using mining. *UNIRIO*. Master degree.

BizAgi Process Modeler., Version 1.6.1.0, 2011. BPMN Software. http://www.bizagi.com. May/2011.

Cook, M., Cook, C. 2000. Competitive Intelligence. London: Kogan Page Limited.

Dey, A. K. 2001. Understanding and using context', Personal and Ubiquitous Computing, 5(1), pp 4–7.

Fayyad, U. M., Piatetsky-Shapiro, G., Smith, P. e Uthurusamy, R. 1996. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press.

Herring, J. P. 1999. Key Intelligence Topics: A Process to Identify and Define Intelligence Needs. *Competitive Intelligence Review*, Vol. 10, No. 2.

Herring, J. P., Francis, D. B. 1999. "Key Intelligence Topics: A Window on the Corporate Competitive Psyche", *Competitive Intelligence Review* 10(4).

IndexMundi. USA Unemployment and Inflation rate. Available at http://www.indexmundi.com. April/2011.

Jackson, J., 2002. Data Mining: a Conceptual Overview, *Comm. Association for Information Systems* 8(19). Available at: http://aisel.aisnet.org/cais/vol8/iss1/19.

Jung, J., Choi, I., Song, M. 2006. An integrated architecture for knowledge management systems and business process management systems. *Computers in Industry* 58, pp 21–34.

Kimball, R., Ross, M. 2002. The Data Warehouse Toolkit. New York, *Wiley Computer Publishing*.

Liebowitz, J., 2003. A set of frameworks to AID the Project manager in conceptualizing and implementing knowledge management initiatives. *Sciencedirect*.

Mack, D., Chawla, N. V., Madey, G., 2005. Activity Mining in Open Source Software, *In NAACSOS 2005*.

Nisbet, R., Elder, J., Miner, G. 2009. Handbook of statistical analysis And Data Mining Applications. California, *Elsevier Inc*.

Nunes, V. T., Santoro F.M., Borges R. B. 2009. A Context-based Model for Knowledge Management embodied in Work Processes, *Information Sciences* 179, pp 2538-2554.

OMG-Object Management Group/Business Process Management Initiative. *BPMN Specification Releases: BPMN 1.2*. http://www.bpmn.org. October/2010.

PMI. 2008. A guide to the project management body of knowledge (PMBOK® Guide) – (Fourth ed.). Newtown Square, PA: Project Management Institute.

Porter, Michael E., 1979. How competitive forces shape strategy, *Harvard business Review*, March/April 1979.

Refaeilzadeh, P., Tang, L., Liu, H., 2007. On Comparison of Feature Selection Algorithms, *Association for the Advancement of Artificial Intelligence* (www.aaai.org).

Ramos, E. C., Santoro, F. M.. 2010. A Model to Support Knowledge Management based on External Context. *In: Workshop of Theses and Dissertations– Brazilian Symposium on Information Systems*, Marabá, Brazil.

Recker, J. C., Rosemann, M. 2006. Context-aware Process Design: Exploring the Extrinsic Drivers for Process Flexibility. *In: The 18th International Conference on Advanced Information Systems Engineering. Proceedings of Workshops and Doctoral Consortium*.

Rosemann, M., Recker, J., Flender, C. 2008. "Contextualization of Business Processes," *International Journal of Business Process Integration and Management*, vol. 3, pp. 47-60.

Saidani.O, S. Nurcan. 2007. Towards Context Aware Business Process Modelling, *Workshop on Business Process Modelling, Development, and Support* (BP MDS), Trondheim, Norway.

Soffer P., Ghattas J., Peleg M. 2010. A Goal-Based Approach for Learning in Business Processes, In: Nurcan et al (eds), *Intentional Perspectives on Information Systems Engineering*, Springer.

StatSoft Inc. 2010. STATISTICA Data Miner. http://www.StatSoft.com. May/2011.

Greg Madey, ed., The SourceForge Research Data Archive (SRDA). University of Notre Dame. Available at http://srda.cse.nd.edu. July/2011.

The Times, SLEPT analysis. 100 Edition. 2010. www.thetimes100.co.uk. Last accessed Apr/2010.

Vuori, V., Pirttimäki, V. 2005. Identifying of Information Needs in Seasonal Management, Frontiers of E-business Research, pp. 588-602.

Yin, R. K., 2009. Case Study Research: Design and Methods. *Fourth Ed. SAGE Publications*. California.