Available online at www.sciencedirect.com

ScienceDirect

IERI Procedia 6 (2014) 45 – 51

Procedia IERI

www.elsevier.com/locate/procedia

2013 International Conference on Future Software Engineering and Multimedia Engineering

A Simulation Approach in Process Mining Conformance Analysis. The Introduction of a Brand New BPMN Element.

Converso Giuseppe, Menicocci Valerio, Murino Teresa, Santillo Liberatina Carmela

**DICMAPI, Scuola politecnica e delle scienze di base, Piazzale Tecchio80,80125, Napoli-Italy*

Abstract

The computerization of organizational processes provides several tools for evaluating the quality of mapping. An executed process produces, in each instance, log files that can be used to reconstruct the actual procedure carried out by the system as well as to highlight deviations or "move" from the mapped path. These "moves" represent a loss that we consider unacceptable for organizations, which manifests its effects in various modes and weights. From this arises, in our view, the need for tools and logic approach to manage and limit the "conformance risk", including as well as a proper consideration and methodological evaluation of the same risk, but also practical solutions and mapping tools that might influence the occurrence. In this paper we propose a methodology and a simulation model of "Conformance Risk Aware Desing" in order to support the modeler in moving from a diagnostic to a preventive and design view of the conformance's matter.

Keywords: BPMN; Conformance Risk; Process Mining;Business Process Modelling

1. Business Process Modelling: state of art

The modeling phase is the starting point for any implementation of techniques and theories of business process management. Looking at the process' life cycle is easy to recognize how the quality of and the choices made in the modeling phase influence the system's performance and the enforceability of tests carried out on them. Starting from the Petri nets, first standard for the description of complex relational systems, the

doi:10.1016/j.ieri.2014.03.008

computerization of organizational reality and the ever increasing need for integration between proprietary systems has led to the development and dissemination of graphical modeling standards of success and to the emergence of the modeling as an independent use case in the research field. Despite looking at the “complex object” process with different logics, standards such as EPC and IDEF, close to a declarative and object-oriented vision of the process and clearly influenced by Petri nets and BPMN and UML AD, instead mostly oriented by procedural logics, have achieved high acclaim and market shares. The statistical data, in particular, confirm the huge success of BPMN (77% of the user’s choices),on the way to become the de-facto standard for the business modeling. Recent BPM institutes and research community efforts are focusing on the issue of compliance of executed processes compared to those mapped. The support of business information systems has allowed the development of algorithms and process mining techniques that allow the identification of any deviations from the prescribed procedure enabling interesting diagnostic analysis.

2. Process Mining and Conformance

The Process Mining is a relatively young area of research that arises from the confluence of data mining, from one side, and computational intelligence and, secondly, the modeling and analysis of processes, on the other. The assumption of process mining is to deduce, monitor and improve business processes by extracting real knowledge from log files, now widely available in the enterprise information systems. It is useful, however, to group the purposes of operating a system of process mining in three fundamental problems: Discovery; Conformance checking; Enhancement.

The discovery techniques take as input an event log and produce a model. The extracted model is typically a process model (for example, a Petri net, a BPMN model, a model EPC or a UML diagram of activities). However, the model may also describe other perspectives (such as a social network). The conformance checking techniques take as input an event log and a model. The output consists of a series of diagnostic information that show the differences between the model and the log. The techniques of enhancement, finally, (revision or extension) require an event log and a model input. The output is the same model improved or extended. For the purpose of the study here, is functional focus on the second category of use of process mining.

3. Conformance Risk Aware Design

3.1 Conformance Risk Assessment

Any policy of risk management related to the deviation by the mapping procedural nodes should start from a proper evaluation phase of the "conformance risk" refers to the specific mapped element. Each node represents a breakdown of the entire work, of the choices and the global rules included into the process, the non-execution or delayed execution of which has repercussions on the entire organization. It is proposed, therefore, to assign to each node of the map a risk value R(i) equal to the product of the deviation’s probability Pr(i), calculated on the basis of time series obtained from log files of similar processes and similar nodes, and an Imp(i) measurement of the effects on the organization of a deviation from the specific node, and related to potential fines, recovery costs, affects on the whole process. The output of this stage is a differentiation between the nodes of the map according to their criticality respect to possible deviations or executions not in line with the mapped procedure. The identification of critical nodes is made easier by the use of a Pr(i)-Imp(i) graph with a curve of the maximum acceptable level of risk, chosen by the modeler.

3.2 Conformance Controller

It is proposed in this paper a method of prevention of "conformance risk" that is firmly based on the process mapping and design phase of the process itself. The methodological part of the proposal consists in managing the risk of diversion through the systematic inclusion, in the design phase of the process, of BPMN pattern consisting of a sequence of elements "critical node" - "conformance controller". Acquired as input the assessment results and ,then, the list of "critical" nodes, the methodology routinely inputs testing conformance patterns that identify the critical upstream node, communicating the system about the deviation's danger, and constrain the node enforcement by creating specific control loop. To reach this aim we propose to enlarge the catalog of controls available in BPMN with a new implicit element, specifically designed for conformance risk and that we call "conformance controller". A first key characterization given to the proposed new item is graphic and symbolic one. The main objective of the creation of that element and the inclusion of the same in a mapping of business process is to communicate to the recipient about the severity of the mapping and the importance of upstream node and in addition to indicate the obligation for the one who has performed the activities to communicate the result to a monitoring system. Great emphasis will be placed, therefore, on the immediacy and the comprehensibility of the new element, launched mainly with communicative purposes and to lighten and synthesize any representations of the same extended monitoring mechanisms. To achieve these purposes it is planned to include the "controller" in the category of "Flow object" BPMN elements and in particular, among the controls, identified graphically by lozenges. This choice was made because the "controller" is definitely an active element and an element of the flow of activities.

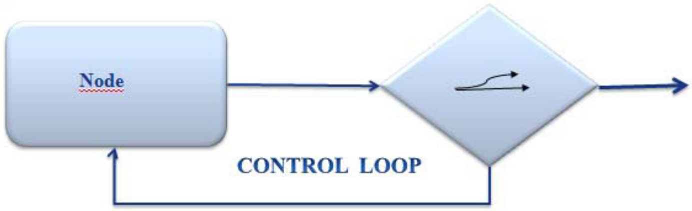

Fig. 1. Conformance Controller loop

From a semantic point of view, this type of diamond would be an unique within its category, going to assume an implicit meaning and unambiguous, away from the operating conditions and data unlike the other controls available in BPMN, whose semantics depend on the process type and the conditions imposed by itself. Overcoming any kind of misunderstanding or interpretation of operation, the "controller" looks at the mapping itself, and is looming as a guarantee for non-material aspects of the process but for the respect of what was decided at the design stage, well marrying with a modern view of the analysis of variances and the challenge that we set ourselves personally here, which is to treat to conformance checking not as a diagnostic situation but as something that can be treated at the design stage of the case, or at least the what we might think, already in the design phase of the process, to mitigate using consolidated and spread tool of the business process mapping. As already pointed out the inclusion of the new element BPMN would benefit to the mapping of the process also in terms of quality of the mapping itself, going to indicate with a single implicit and unequivocal graphic symbol a organizational situation for which would, however, be used several syntactic elements, which weigh down modeling itself, raising the likelihood of errors (statistically correlated to the number of nodes of a mapping). Recent research on the matter, in fact, show how factors related to the quality of the graphic design of the mapping can influence in a decisive way the behavior of the operators. Among these factors are, in particular, the number of nodes, the ambiguity inherent in the symbology graphic notations, the number of paths and the "diameter" of the mapping.

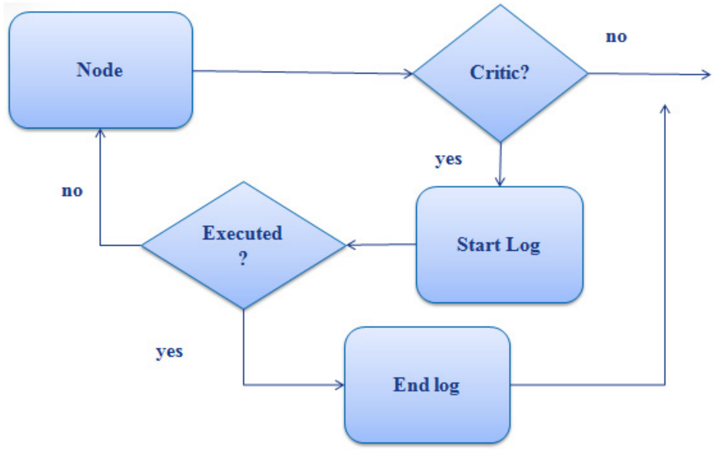

Fig. 2. Explicit Conformance pattern

3.3 Process Model Conformance Index (PMCI)

The quality of conformance control pattern mapping is measurable through an indicator that we define Process Model Conformance Index (PMCI) that intends to measure the quality of the choices made in the design stage of the process in reference to critical tasks bound to the control. The indicator is the ratio between the value of the "conformance risk" associated to the entire process map, and specifically equal to the sum of the results of conformance risk assessment of individual nodes, and a value equal to the actual losses incurred in the specific instance of the process, given by the sum of the costs of implementation of the control system on the folded single instance, expressed as a percentage of the R(i) value of the node to which the control refers, and the value of deviations incurred in the specific instance, equal to the sum of R (i) of the diverted nodes. A value greater than 1 indicates a positive situation of PMCI and therefore, an advantage for the organization relates to the specific mapping. Values less than unity testify differently, bankruptcy instances from the point of view of compliance. In the following we will produce a simulation model that will make possible the prevision of this indicator for the purpose, not only witness and document-level operational benefits of the introduction of "conformance controller" in a mapping process, but also to provide additional support to the modeler about the design choices in the field of conformity control.
PMCI = (MAXIMUM RISK OF MAP) / (RISK SINGLE INSTANCE)

3.4 Conformance Tester

We, finally, propose the interaction between the fuzzy logic inference and the dynamic simulation in order to produce a simulation model as support to the modeler's decision. With this aim we propose the use of Matlab Fuzzy Inference System and Powersim Studio to replicate the relationship and comparison between the behavioral benefits imposed to the system by the conformance controller's introduction and the increasing mapping complexity, linked by literature with the error's and, so deviation's probability. Accepting as design-input three values:
1. Maximum level of acceptable risk conformance;
2. Type of graphic control configuration: Conformance Pattern vs Conformance Controller;
3. Cost Index: average ratio between the control's cost of the single node and its deviation's cost;
the model returns as output the value PMCI of the single simulated instance : $\Sigma R(i)/\Sigma c{*}R(i)c+\Sigma D(i)$ where:

ΣR(i) is equal to the sum of the results of the assessment phase.

C is the cost index above-introduces.

R(i)c is the conformance risk assessment of the controlled nodes.

D(i) is equal to the R(i) of the deviated nodes in the simulated instance of the process.

Entering in the dynamic logic of the model it's to underline that in each Time-step the system replicate the occurrence of a mapped node and assigns to the same an R(i) through a scrolling-vector containing all the assessment value and identify it as a "critic node" on the basis of the comparison with the Maximum level of acceptable risk conformance, imposed by the modeller. The system includes conformance controls, in the binary graphic choice, downstream the critical nodes. Controls, in every configuration, increase the total number of nodes to be processed and so the error probability represented by the value of a specific variable called "error". Comparing this value with the one assumed by a variable called "skill", representing the behavioral benefits linked with the control's inclusion and modelled through the fuzzy inference of two verbal evaluations, the system, finally, determinates in-line execution of the node or deviation of the same, updating the PMCI value in both the cases.

3.5 The "Error" variable

We have modeled this variable by referring to recent research by Jan Mendeling, Gustaf Neumann and Will Van der Aaalst on quality of business process modeling in the design phase, which provide an interesting model for the evaluation of the evolution of the statistical probability of error in terms of model's quality, in dependence on the number of nodes included in the model. Looking at this information in a of "conformance checking" perspective is possible to consider these probabilities usefull to be used as an average probability (with a d.v. introduced by the system) of having a "move" in the model and therefore a deviation from the mapping.

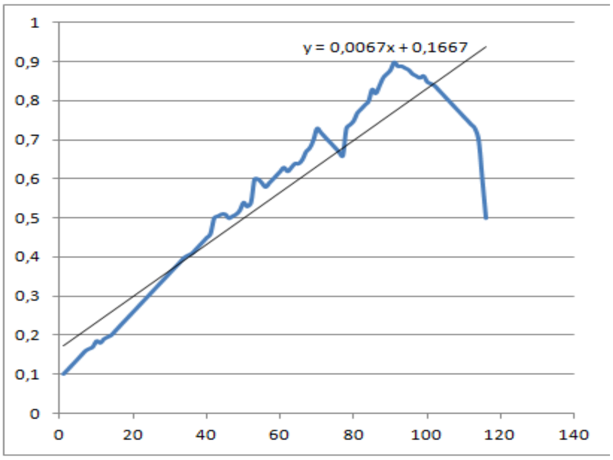

Fig. 3. Error Probability

The correlation coefficient calculated on the dispersion of the data allows to find an obvious linearity and to be able to draw a characteristic equation that well approximates the trend: y = 0.0067 +0.1667 * x.

3.6 The "Skill" variable

The inclusion of control points in the mapping has the aim to prevent deviation from the requirements and to make more robust the mapping process itself. The existence of control points downstream "critical" nodes should push the system to not deviate from what mapped and, therefore, prevent deviations that may have an impact on business performance. We propose the use of Fuzzy Logic to simulate the average numeric value (with a standard deviation assigned by the system) of the mapping robustness and pass from verbal feedback regarding the mapping to a randomly extracted value to compare, at each step, with the probabilistic variable "error". We identify as influencing elements of strength of the single mapped node: the level of perceived criticality and the distance from the nearest control point. Assuming, so, that the compliance of the operator in respect of the mapping is as greater as the level of risk of the node and as the distance from a control point mapped with the control element decreases. The evaluation of the numerical value is performed via the fuzzy logic tool of MATLAB and the results are exported to the simulation software. The methodology involves two key moments: Fuzzyfication and Defuzzyfication. The first consists in building in the Matlab command window a three-column matrix containing the historical values refer to the probability of successful execution of the single node at the risk values and distances from controls. Working, then, on the graphic fuzzy editor fuzzy, with the command "Anfisedit", and loading on it the array of data from the workspace. Then setting the shape and number of Membership function for each input variable (being no particular systematic in the data, we usually opt is typically 3 MF with bell shape) and the shape (linear or constant) output. Start, therefore, the numerical estimate of the data from the "train FIS", selecting: the estimation method (hybrid to more accurate estimate), fault tolerance and the number of epochs of adapting the system to the data provided (in general > 10). In this phase, the system creates functions and fuzzy rules using a mixed technique of "least squares" and "backpropagation". It's possible, finally, check the data quality with the fixed test section, checking the data generated by the model with those imported from the workspace. The second phase, namely the defuzzyfication consists in providing the software with a simple line of code: skill = evalfis ([lr, d], "name rescue") where the values in brackets are those of the node.

4. Conclusions and Future works

The Business Process Modelling has become an established science and finds application in all organizations, in a continuous and cross way. Nevertheless, we have identified the need for modelers to be able to raise the quality of the mapping in order to manage the risk of deviations from the procedure laid down, which may prove, in reference to critical tasks, certain deleterious. Our proposal methodological and ontological application has confirmed a remarkable success in the simulation scenarios, particularly in the more prescriptive ones (which represents the current trend of the business reality!), and opens a new theme and a new challenge in the context of process mapping. In the future we expect to test the system, the methodology, the application and the "conformance controller" in the design of brand new real process, as well as to propose the whole logic and tools to institutions and organizations that bring forward BPM research and standardization within OMG, OASIS, EIT. We plan, then, to be able to improve the simulation model, including the inside of the variable dependencies "Error" from further characteristics of the process mapped, on which the research is slowly pushing, with the main objective to provide more support to the modeler work.

References

[1] Paul Harmon P, Wolf C. Business Process Modelling Survay, 2011. http://www.bpminstitute.org,
[2] Will van der Aalst et al. Manifesto del Process Mining, Springer-Verlag; 2011.

[3] Mendeling J, Neumann G, Van der Aalst W. Understanding the Occurrence of Errors in Process Models based on Metrics; 2007.
[4] White SA, Miers D. BPMN – Modeling and references guide. Future Strategies Inc., Lighthouse Pt; 2008.
[5] Mendeling J, Reijers HA, Van der Aalst W.Seven Process Modeling Guidelines (7PMG); 2009.
[6] OMG, “Business Process Model and Notation (BPMN),” 2011, version 2.0. [Online]. Available: http://www.omg.org/spec/BPMN/2.0 , March 2013
[7] De Leoni M, Maggi FM, Van der Aalst W. Aligning Event Logs and Declarative Process Models for Conformance Checking; 2013.
[8] Ramezani E, Fahland D, Van der Aalst W. Where Did I Misbehave? Diagnostic Information in Compliance Checking; 2013.