Detection of Dataflow Anomalies in Business Process An Overview of Modeling Approaches

Najat Chadli, Mohammed Issam Kabbaj and Zohra Bakkouri

June 2, 2018

Detection of Dataflow Anomalies in Business Process An Overview of Modeling Approaches *

N. CHADLI
Mohammadia Engineering Schoo
Mohammed V University, Rabat
Morocco

najatchadli@research.emi.ac.ma

M. I. KABBAJ
Mohammadia Engineering School
Mohammed V University, Rabat,
Morocco

kabbaj@emi.ac.ma

Z. BAKKOURI
Mohammadia Engineering School P.O.
Mohammed V University, Rabat
Morocco

bakkoury@emi.ac.ma

ABSTRACT

Most research focus on control flow modeling when modeling and analyzing business process models, but less attention is paid to data flow. However, both of data flow and control flow are essential in process modeling. Thus, the verification and data flow modeling have an important in detecting anomalies. However, when they began to effect data flow modeling in process model there are several errors discovered. In this study, some recent approaches for anomalies detection are reviewed. Eventually, the approaches are: first an analytical approach for detecting and eliminating the three types of data-flow errors, that formally establishes the correctness criteria for data-flow modeling. Second, formulate the data-flow modeling and verification using a Petri Net based approach. Third, an ad hoc approach detecting data modelling errors in business process models apply for an active help. indeed, we explain for each approach its proper method and tools. We then compare and analyze them in order to discover the added-value of each approach.

KEYWORDS

data-flow modeling; verification; data-validation; data anomalies.

1 INTRODUCTION

A business process consists of a set of activities structured to collaboratively achieve common business goals using a clearly identified input, output, a beginning and an end of the process. Therefore, the business process is considered as the specific ordering of work activities across time and place [1-2]. So, in Business Process Management, the standard paradigm for process modeling is the workflow concept. However, the workflow specification requires to characterize the other perspectives, including time, data, etc... [3]. As consequently, for workflow modeling, analyzing and verification the time management is an important aspect [3-5]. Similarly, data is also important for data-flow modeling and routing choices in a process that is typically determined by certain data items [2]. While the business process becomes increasingly complex, the workflow technology constitutes a standard solution to manage it. In workflow design, most efforts dedicate to control-flow to identify errors. As the workflow system gets more complex, data flow is becoming more and more prominent in modeling, as for verification. Indeed, the activity sequencing is controlled by some of the operational constraint that determined by the relation among data elements. Consequently, data flow perspective had an important in workflow management [4]. The consumed and produced data, with respect to each activity in a business process, is defined by the "information perspective", i.e. How technological changes interact with individual behaviors, organizations and society to affect the availability and uses of information in governance processes [4]. Therefore, to detect the errors of data, it is being necessary to include data flow analysis with control flow into workflow structured. In this concept, several approaches have been proposed for detecting the errors of data flow modeling in a workflow. To model and analyze the data requirements in workflow systems, some informal and formal modeling tools have been developed [6]. In business process modeling literature, the data flow modeling anomalies are tackled by different approaches. Although, their methods to analyze the issues is different from one of the other, these methods are verified by rules and lemmas and theorem of the correctness criteria and an algorithmic verification. Indeed, there are many errors types in data flow modeling. The three basic types are missing data, conflicting data and redundant data. Consequently, the choice of these three types definitions of data flow anomalies seems to be sufficient in analyzing the data flow requirements at the conceptual level [2-3-4]. The main motivation of this work is to compare several approaches using the aforementioned errors types in order to extract advantages of each approach. This will allow to build an effective new approach based on the hybridization of the compared methods. The first method [4] used in this study, formally establishes the correctness criteria besides data-flow specification by proposing an additional component for data flow analysis. The second one is a Petri Net based approach using the Petri-net incidence matrix with a polynomial complexity algorithm to formulate the Data-Flow Modeling and Verification [3]. The last one [2] is an ad-hoc approach using an active help method for detecting data modelling errors in workflow and a DataRecord concept to manage datasets. The purpose of this paper is to provide symptomatic descriptions of modeling methods and tools using the above error types. Afterwards, we aim to compare the different strong points and weaknesses between the three-approach proposed. The interest in this work is to provide new insights and study the added value of each approach.

This paper is organized and structured as follows. Section 2 presents a data flow modeling and detection of anomalies in business process and make a critical analysis of each approach of an article selected. Section 3 we give a statement that provides something clear of comparative approaches. Section 4 Conclusion and Perspectives.

2 DETECTION OF DATAFLOW ANOMALIES IN BUSINESS PROCESS

2.1 Introduction.

2.1.1 Dataflow modeling:

Data modeling involves a progression from conceptual model to logical model to physical schema. So, Data flow modeling is one of the foundations of the Structured Systems Analysis and Design Methods. Data-flow modeling verification concentrate on identifying data-flow errors by means of a set of well-defined correctness criteria. Consequently, verification aims to define the problems caused by incorrect data flow modeling.

*2.1.*2 The anomalies of data flow modeling:

There are several anomalies of data flow modeling that is: Redundant Data, conflicting data, missing data, Mismatched data, Inconsistent data, Misdirected data, Insufficient data [4]. The basic data flow anomalies are missing data, conflicting data and redundant data [7].

- Missing Data.

when data has never been created before or accessed without being initialized during modeling process, in this case a missing data error occurs.

- Conflicting Data.

Conflicting data occurs if some data elements are written by an activity, however, activities cannot confirm an update due to the existence of several versions of the same data elements, which causes a conflict of version should be updated.

- Redundant Data.

If a data element is written by an activity, but has never been read in all possible continuations, then, this data element is a redundant data. Maintaining the Integrity of the Specifications.

2.2 Anomalies detection approaches.

In the workflow, detecting data flow modeling anomalies has many approaches. However, these three approaches cited below are selected in this study because they are recent and have applied different methods and concepts.

2.2.1 formulating the data flow perspective for business process modeling:

This approach provides a data-flow framework for detecting data-flow anomalies. This framework includes two basic components: data-flow specification and data-flow analysis. With these components, the business process management has a more analytical stiffness, and have an interesting aspect towards a formal methodology for data flow modeling. A data-flow verification, which is a theoretical foundation criterion, is also proposed to eliminate systematic and automatic data-flow anomalies [4].

- Explication of approach:

This approach can be tested in the real-world application; it needs a prototype data Flow modeling to develop the formal manner of correcting data flow anomalies. To reform these errors, this approach requires to modify not only the data flow but also some cases of the control flow. Consequently, a new workflow conception based on the data flow analysis is elaborated containing a new operational perspective that specifies methods used in the workflow system such as data flow operations and data flow matrix. So, they had illustrated the concepts that have introduced a property loan approval process shown as a UML activity diagram. the correctness criteria for data flow modeling is formally established as a theoretical foundation for the data flow verification. These criteria enable systematic and automatic elimination of data flow errors as in the Figure 1 [4].

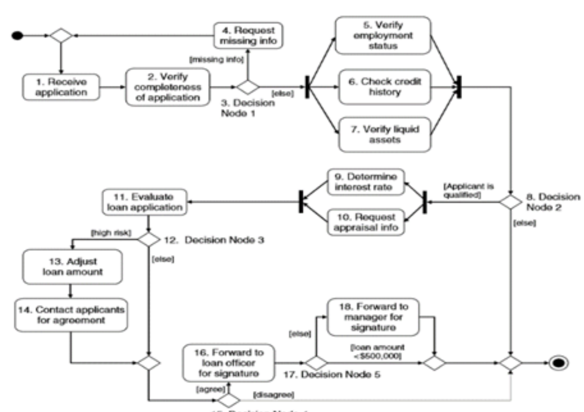

Figure 1: Property Loan Approval Process.

- Data-flow operations.

Operations called data flow operations performed in the activity need a data item. Additionally, data items are produced, accessed and modified by means of these operations. Data-flow operations are classified into six operations notably initializing, approving, updating, referring and verifying according to the semantic meanings of data flow operations in the business process [4].

- Data-flow matrices.

The concept of the data-flow matrix is a two-dimensional table specifying data flow in workflow applications. Indeed, the matrix records the data-flow operations each activity performs on data items in a workflow. Moreover, the decision nodes must be included in the data-flow matrix as activities since they require input data. Finally, with such dataflow matrices, it is easier to find out how each data item is processed in a workflow [4].

- Integration of dataflow in the workflow model.

Input and output for an activity can be described by using the object flows in a UML activity diagram. Indeed, to indicate the relationship action-object, each object is connected to one or more activities. However, as no details are provided on how different data items connected with one object are processed differently in a workflow, the object flow for modeling data flow in workflow management is insufficient. consequently, it is possible to integrate Data-flow information the control flow model [4].

2.2.2 Formulating the Data-Flow Modeling and Verification for Workflow:

In this study, a Petri-net [8-9] based approach is proposed. As mentioned by Aalst [12]" there are at least three reasons for using Petri nets to model and analyze workflows: (1) Graphical nature and formal semantics have been defined for Petri nets; (2) Petri nets are state-based instead of event-based, so the state of the case can be modeled explicitly in Petri nets; and (3) Petri nets are characterized by the availability of many analysis techniques". Indeed, to model the control flow and the data flow information, each activity has been extended with its read and write data-sets. At the moment of modeling, this approach applies a verification using a polynomial complexity algorithm exploiting a data-activity incidence matrix to detect the anomalies of data flow [3].

- Explication of approach.

This approach is mainly based on firing rules in Petri-Net. It formulates the data flow modeling and verification in a workflow structured with data elements. Therefore, to formulate data flow anomalies, an input for reading and an output for writing the data-sets is required. Furthermore, a polynomial complexity algorithm and the activity-data incidence-matrix of the WFIO-net are used for detecting the above-mentioned three kinds of basic data-flow anomalies [3].

- Firing rules:

The firing rules for classic Petri nets, have some basic static notions that need to be explained. Let N be an arbitrary Petri net with a set of places P, a set of transitions T and a flow relation F. All places which are connected to a transition by an arc form the set of pre-places and post-places of a specific transition. Accordingly, the rules describing the possible firing rules, the occurring change itself is called a firing [10].

- From WF-Net to WFIO-Net.

"Modeling a workflow process definition in terms of a Petri net is rather straightforward, i.e. tasks are modeled by transitions, conditions are modeled by places, and cases are modeled by tokens". A Petri net which models a workflow process definition is called a workflow net (WF-net) and defined as follows [11-12]:
A Petri net $PN, PN = (P, T;\ F)$ is a WF-net if and only if.
1) PN has two special places: $i\ and\ o$, place i is a source place and o is a sink place: $\cdot i = 0\ place\ o\ is\ a\ sink\ place{:}\ o \cdot = 0$.
2) A transition $t*$ is added to PN which connects place o with place, then the resulting Petri net is strongly connected. The workflow modeled by WF-net is a sub-set of a Petri Net which is used to characterize formal languages [9-12]. Nevertheless, the WF-net can exclusively signify the logic relation of the workflow, i.e. the control-flow aspect. The workflow input/output net "WFIO-net" is prospective for modeling the data flow elements of a workflow that is a WF-net while extending each activity with its input and output data sets.
As in [3], the figure 2 and figure 3 represent a model by WF-net and WFIO-net.

TA is an activity transition set.
PL is logic place set.
TL is a logic transition set.
Source place ps.

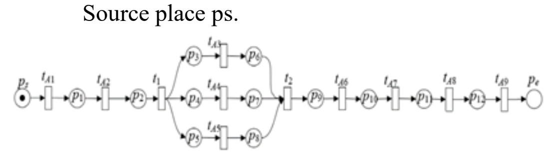

Figure 2: WF-net of the Property Loan Approval BP.

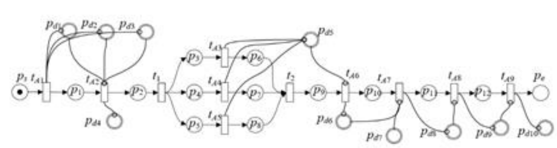

Figure 3: WFIO-net of the Property Loan Approval BP.

2.2.3 Towards an active help on detecting data-flow errors in business process models:
In this study, an ad-hoc approach intended to discover the data flow modeling anomalies is proposed. Indeed, to solve the issues of data flow analysis using an active help, a verification process is applied permanently for each fragment of the model. However, to store the last set of each data and the last activities that have read, update or destroy this data, a DataRecord concept is used [2].

- Explication of approach:

The business process modeler first designs model and then check it to discover the data flow errors. However, fixing these errors does not signify that the model is correct. In this case, a re-validation of the entire model is essential to ensure that there is no error which causes a loss of time and cost. Thus, an active help tool for real-time analysis to anticipate the error is proposed. Actually, the process of verification is triggered whenever a fragment is added to the model. This has given a free error fragment during modeling. Additionally, to store the last set of data in the model and the last activities that have read, update or destroy this data, the concept of "DataRecord" is introduced. DataRecord is an $(n \times p)$ matrix where n is the number of data and p is the number of branches XOR he model. DataRecord is initially empty, as and when the business process model is drawn, data is inserted according to the rules of data flow anomalies notably missing data, conflicting data and redundant data errors. The various of data-items in the workflow are incrementally recorded in the matrix by passing from an activity to the other sequentially [2].

- Example of Model with an XOR split.

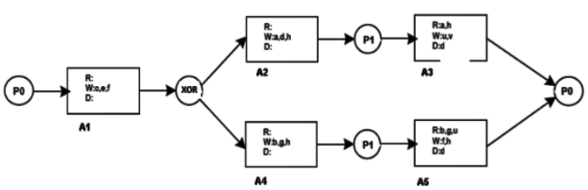

Figure 4: Model with an XOR split.

In this above example of figure 4, the approach results in an analyze like development in the table. As consequently, the column "state" of DataRecord reflects the latest state of each data item and the next column is the last state of each xor branch's. indeed, applying the rule 1, data item u, detected like missing data, since this data created only in the first xor branch's of activity $A3$, but didn't create in the second xor branch's. However, the modeler blocks the model to correct the error, that can choose the state to destroy u. Thus, applied the rule 2 in the same example data item f and h detected like conflicting data and in the rule 4 data item v, c and e detected as redundant data.

Table 1: The Last State of DataRecord of the model with an XOR split.

Data	State	State branch 1	State branch 2
c	$(0,W_{A1},0)$		
e	$(0,W_{A1},0)$		
f	$(0,W_{A1},0)$		$XOR_2\ (0,W_{A5},0)$
a	$(0,W_{A2},0)$	$XOR_1\ (R_{A3},W_{A2},0)$	
d	$(0,W_{A2},0)$	$XOR_1\ (0,W_{A2}, D_{A4})$	$XOR_2\ (0,W_{A2}, D_{A5})$
h		$XOR_1\ (R_{A3},W_{A2},0)$	$XOR_2\ (0,W_{A5},0)$
u		$XOR_1\ (0,W_{A3},0)$	
v		$XOR_1\ (0,W_{A3},0)$	
b			$XOR_2\ (R_{A5},W_{A4},0)$
g			$XOR_2\ (R_{A5},W_{A4},0)$

3 GENERAL DISCUSSION OF THE THREE APPROACHES

In this section, we aim to explain the comparative study of these approaches by analyzing the similar and the opposed ideas in each approach. This analysis has gone through three approaches that verify and model the flow of data in a workflow.

3.1 Similarities of the three studies

The authors have pursued the same goals to detect data flow anomalies in business process modeling. They used Keywords e.g; data-flow modeling, workflow, data-flow verification, data-flow anomalies, data-flow analysis; they made data-flow analysis more manageable and covers the key issues of ensuring data-flow integrity in the workflow at the conceptual level beside the control flow to address the data flow anomalies. Moreover, workflow systems have become a standard solution for managing complex processes in business domains such as supply chain management. The three approaches used the same types of data errors to analyze their work, which are: Missing data, Conflicting data and Redundant data.

3.2 The differences

Each approach uses a matrix that contains data and activities, but there are major differences between their purpose. In the first approach, data flow matrix is used to record the data-flow operations that each activity performs on various data items. The second approach used Activity-Data Incidence Matrix of WFIO-net because this type of matrices shows the control-flow relation but no data-flow information is reflected whereas the third approach uses DataRecord matrix to records the various data items and last activities in the workflow.

In more details, first approach aims to present a data perspective including two basic components: data-flow specification and data-flow analysis. A simple predicate logic format is used instead of following the event-role-object-condition-action format. The presented lemmas and theorems gave rise to data-flow verification rules. In this study also, algorithms have been developed to be used as a roadmap for the implementation of the data-flow perspective. A data- flow matrix and an extension of the unified modeling language (UML) activity diagram is introduced. However, the authors have limited their attention to five patterns of workflow: sequences, AND-Split -AND-Join -XOR-Split - XOR-Join.

Second approach is based on the Petri-net system. A WFIO-net extending the WF-net as its basic concepts. The authors used control flow and data flow to formulate the data-flow modeling verification in workflow, they also proposed for interpreting, to introduce the activity task incidence matrix of a WFIO-net enriched by the rules of data flow anomalies. However, the author didn't use a time factor in WFIO-net to give a more accurate verification. There is also a lack of detailed the taxonomy for each kind of data flow errors.

While the abovementioned approaches provide a passive help to the designer as they need to check the correctness of the model at the end of the modeling phase, repairing detected errors doesn’t ensure that the result is a correct model, it is obligatory to revalidate the model. The third study, an ad hoc approach uses an active help method and a “DataRecord” concept to store the last state of each data item in the model and the last activity that has performed read, update and destroy operations. This approach applies verification at the modeling time. The authors tested this approach on a simple linear model and an XOR-Split model of two branches. However, they didn't apply a loop modeling. In this case, for using the loop, it is necessary to enrich the approach by special looping rules.

4 CONCLUSIONS

In this paper, an overview of detecting anomalies approaches in a workflow is provided. The first approach’s goals are to formulate the data flow perspective by means of dependency analysis. The data-flow matrix and an extension of the UML activity diagram are proposed to specify the data flow in the business process. The second approach had a goal to formulate the data-flow modeling and verification, a Petri Net based approach, applied a polynomial complexity algorithm and the activity-data incidence matrix of the WFIO-net. The third approach is to introduce an ad hoc approach for an active help on detecting data modeling errors. They used a “DataRecord” concept to store the last state of each data item in the model and the last activity that has read, update or destroy this data.

REFERENCES

[1] Davenport, T.H. (1993). Process Innovation: reengineering work through information technology, Harvard Business School Press, Boston, MA, USA.

[2] M. I, Kabbaj; A. R; B, A; Z., Bakkoury. Towards an active help on detecting data flow errors in business process models.

[3] Cong LIU . Qingtian ZENG Formulating the Data-Flow Modeling and Verification for Workflow: A Petri Net based Approach.

[4] Formulating the Data-Flow Perspective for Business Process Management.

[5] Jianqiang Li, Yushun Fan, Mengchu Zhou, “Performance modeling and analysis of workflow”, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol.34, no.2, pp. 229- 242, March 2004.

[6] P.Lang G.Kappel W. Retshitzegger S.Ransch ,Workflow management based on objects rules and role January 1995.

[7] Sherry Sun Leon Zhao Olivia Sheng “Data Flow Modeling and Verification in Business Process Management”.

[8] W. M. P. van der Aalst, K. M. van Hee, A. H. M. ter Hofstede et al., “Soundness of workflow nets: classification, decidability, and analysis,” Formal Aspects of Computing, vol. 23, no. 3, pp. 333–363, 2011.

[9] T. Murata, “Petri Nets: Properties, Analysis and Applications,” Proceedings of the IEEE, vol. 77, no. 4, pp. 541–580, April, 1989.

[10] Time and Petri Nets, Louchka Popova-Zeugmann ; Chapter 2 The Classic Petri Net.

[11] W. Reisig, Understanding Petri Nets: Modeling Techniques, Analysis Methods, Case Studies, Springer, 2013. 23.1 A Brief History of Petri Nets p .207.

[12] van der Aalst, W.M.P.: The Application of Petri Nets to Workflow Management. The Journal of Circuits, Systems and Computers 8(1), 21–66 (1998).