An Improved Ad Hoc Approach based on Active Help Method to Detect Data Flow Anomalies in a Loop of Business Process Modeling

N. Chadli, M. I. Kabbaj and Z. Bakkoury
Mohammed V University, Avenue des Nations Unies, Rabat, Morocco
Mohammadia Engineering School, Avenue Ibn Sina, Rabat, Morocco

Keywords: Data Flow modeling, Active Help, Loop Modeling, Verification, Validation.

Abstract: The data flow in business process modeling is created and distributed by the exchange of data moving from a task to another in information systems. Among open issues in workflow modeling is the detection of errors for data flow and control flow correctness. Recently researchers have focused on detecting errors by applying an active help with a concept of Data-Record. However, this method does not support a loop modeling yet. The goal of this paper is to apply an active help with a Data-Record concept in order to detect data flow anomalies in loop modeling. We propose to improve the active help approach by some suitable Rules for loop modeling. In this context, a decision node, using a data connection as an input data, replaces the connector Xor-split. The input data of the decision node is returned to the last activity by a feedback when the error message is found. The proposed approach is validated using a deterministic finite state process model which uses a logic Boolean predicate (Yes or No) to specify the routing of an input data. Moreover, anomalies such as Missing Data, Conflicting Data and redundant Data are used. The verification is triggered when an anomaly is detected, where the system is locked until a correction is performed. The results show that Missing Data anomalies are efficiently handled by the proposed approach.

1 INTRODUCTION

Currently, many business functions such as purchasing, manufacturing, marketing, engineering, and accounting have been automated by most organizations, (A Basu, RW Blanning, 2000). To use these functions, the data exchange in an information system is necessary for each task to the next in the business process management. In this sense, each task requires input and output data. Specific collection of tasks, resources and information elements make up a workflow system as in (A Basu, RW Blanning, 2000). Indeed, the business process activities are realized throughout tasks in the information systems. Also, business process activities can be achieved by information systems without any human involvement (TH Davenport - 1993 – books). In fact, in the business process management, it is necessary to use a workflow system in order to interplay between data flow created by a data exchange of information systems and the control flow in a workflow. Consequently, data flow is important for business process integration because data is always classified when conducting inter-organizational business and data errors could still happen even given syntactically correct activity dependence. However, the focuses of control perspective and data perspective to describing the logical order of tasks and the information exchange between tasks on verification is for most of the techniques, i.e., on the discovery of design errors. Certainly, the flow-oriented nature of workflow processes styles the Petri net formalism is a natural contender for the modeling and analysis of workflows (LIU Cong, Q ZENG, D Hua ,2014). In the time of processing, there are many issues found when a continuous passage the data flow of a process models from an activity to the next. Many solutions have been proposed by researchers to resolve the data flow anomalies, as each activity needs operational information to define the state of data, that is Read, Write or Destroy. Therefore, this operation can specify the state of data in the activity to another that can cause missing data, conflicting data and redundant data (SX Sun, JL Zhao, JF Nunamaker,2006). An approach ad hoc uses active help (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015) is proposed in a linear model and Xor-split with two branches. This approach helps as to verify in each

Chadli, N., Kabbaj, M. and Bakkoury, Z.
An Improved Ad Hoc Approach based on Active Help Method to Detect Data Flow Anomalies in a Loop of Business Process Modeling.
DOI: 10.5220/0009775402840290
In *Proceedings of the 1st International Conference of Computer Science and Renewable Energies (ICCSRE 2018)*, pages 284-290
ISBN: 978-989-758-431-2
Copyright © 2020 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

time the task and corrected the errors in the same time used a locked system and a concept DataRecord (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015). Moreover, they didn't use the loop when the system ad hoc had a problem in a send message in a linear model or model with Xor branches. In this way, the same approach is used when adding a loop modeling in the linear model with an Xor split in order to detect the data flow modeling anomalies. Indeed, the Xor - split is used to feedback an existing message errors at a proceeding of modeling, this message errors returned to the source activity where is created it up to proceed of correction. Therefore, employ active help (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015), and the rules for verification in the model, that is triggered when some issue in the time of modeling has occurred. However, the loop couldn't assimilate this approach to detect the anomalies not because the active help is insufficient but because the rules of this approach could only create and update. Subsequently, it's proposed to enrich this approach with some enhancements in rules and model in order for the approach to be adapted by the loop. A decision node is proposed like a connector that has a data connection at the input data. In this case, it requires a Boolean predicate (Yes=true, No=false) in a finite-state automaton determinist, so we used the guarding (i.e. blocking) tasks solely on the DataState (N Trcka, W van der Aalst, N Sidorova,2008). In this context, we implemented DataState to verify the last record state of the dataset for each input and output in the activity. In this manner, this data connection is a decision variable that is a routing decision can be made based on a set of data items inputted to the decision node. Each of such data items involved in a routing decision is called a decision variable (SX Sun, JL Zhao, JF Nunamaker,2006). Also, this decision variable is allowed to change the state of DataState that can be initialized in each iteration of connection. Moreover, there isn't the problem in the first iteration however when the iteration is high requires an initialization of the DataState.

The remaining of the paper is organized as follows. Section 2 presents some approach and concepts used in this paper. Section 3 shows that the loop modeling cannot integrate assimilate the approach with active help. In Section 4 presents the new visualization of the approach. Section 5 concludes the paper and discusses the perspective.

2 RELATED WORK

Modeling in the business process has become very important in recent years, with data-flow modelling and verification being the two important challenges in workflow system management. It had many works stakeholders in this problem of anomalies of data-flow and control flow in the workflow. Recently, data flow formalization in process modeling has been investigated by many researchers. In most organisation, it is particularly important that the responsible of key processes feel their interests are represented during the latter phase. To achieve this, the main stakeholders such as the heads of key functions intersected by the process, the managers with operational responsibility for the process, suppliers of important change resources (e.g., the IT, human resource, and financial functions), and process customers and suppliers, both internal and external should participate in the team during the design phase. (TH Davenport - 1993 – books). In graph-based approaches to business process modelling, data dependencies are represented by data flow between activities. Each process activity is given a set of input and a set of output parameters. Upon its start, an activity reads its input parameters, and upon its termination, it writes data it generated to its output parameters. These parameters can be used by follow up activities in the business process (M Weske p.100 ,2012). The importance of data-flow verification in workflow processes was first mentioned in (S Sadiq, M Orlo, W Sadiq, C Foulger, 2004). The information perspective defines what data are expended and produced with reverence to each activity in a business process. Thus, the operational perspective requires what tools and applications are used to execute a particular task (SX S, JL Z, JF N, 2006). Many approaches have been proposed for for data-flow verification, these approaches enable systematic and automatic elimination of data-flow errors as in (SX S, JL Z, JF N, 2006). An approach of ad hoc that treated the anomalies of data-flow for each activity by an active help using a conception of dataRecord which stored data with their state read, write and destroy presented in (MI K, A B, Z Ba, A R ,2015). Indeed, data flow perspective approach formally discovers the correctness criteria for data-flow modeling. Petri Net based approach is proposed for modeling the control flow of workflow. We extended this model by including the input and output of data flow and added a complexity of algorithm for detecting the anomalies of data flow as in (LIU C, Q Z, D H ,2014). Our approach extends and generalizes data flow verification methods that have been recently proposed. It also makes use of the concept of corresponding pairs lately introduced in control flow verification. It has, therefore, the potential to be developed into a unified algorithmic procedure for the concurrent detection of control flow and data flow errors. The aim of this paper is to present an algorithm called GTforDF, for data flow verification through the detection of lost data. The paper also will explain through practical examples how GTforDF detects data flow errors in workflows and define an important new error category called redundant data in loops that can lead to data loss in some situations (HS Meda, AK Sen, A Bagchi, 2007). The approach focuses on the discovery of data flow errors in workflows such as Redundant Data, Lost Data, Missing Data, Mismatched Data, Inconsistent Data, and Misdirected Data. To achieve this, we an analysis which uses "The RWD Boolean Table Technique" that is expressed in steps, to split data-flow from control flow and to create Boolean table for each data elements, and also to compare RWD Boolean table for current task and next task until it gets to the end of workflow as in (AE Rgibi, SZ Yao, JJ Xu,2012) . A three-layer workflow model for designing a workflow was proposed in (FJ Wang, CL Hsu, HJ Hsu ,2006). They characterized the behavior of an artifact by its state transition diagram and identified six inaccurate usages affecting workflow execution and a set of algorithms a set of algorithms to detect these inaccurate usages in workflow specification is presented (FJ Wang, CL Hsu, HJ Hsu ,2006). An approach in data flow issues proceedings for mapping BPMN to Petri-Net to provide a systematic technique of possible flows related to the data flow of business process Data flow issues and BPMN mapping to Petri Net: Road map as (AES Rgibi,2015). In a nutshell, the objective of this paper is merging the Decision Node with an input data and a logic Boolean by the guard to find a new solution to solve the problems of data flow anomalies in the business process with a loop modeling in a linear model and Xor split.

3 VERIFICATION APPROACH WITH A LOOP MODELING

In the case where an ad hoc method is applied in a simple linear model with a loop, the system is triggered when an error message of transmission is produced. So, a feedback is structured to return error to correct it. Otherwise, the modeler continues to execute the next task as in figure 1. So, the feedback requires the verification to detect data flow anomalies in each system workflow instance. Consequently, the output data became an input data in an information exchange system of data flow. Indeed, at the moment of the error message is returned, the approach is triggered, and the verification is required to detect the anomalies in each system workflow instance. As a result, in the feedback and in each activity the output data becomes an input data in an information exchange system. Additionally, in a workflow, each activity performs a comparison operation on a data element. Thus, data operations are spontaneous, when an activity A is reading data, the item d is an input data. The same, when an activity A carry out a writing operation on d, d is the output data from A as in (SX Sun, JL Zhao, JF Nunamaker,2006). Furthermore, it is proposed to verify some anomalies such as Missing data, conflicting data, redundant data for loop modeling applying the approach rules (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015).

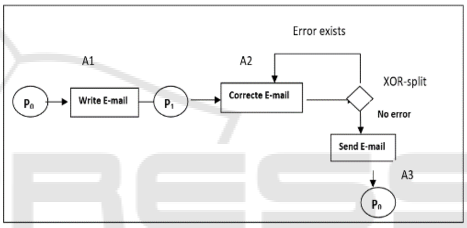

Figure 1: Description of linear model and loop modeling.

3.1 Description of Symbols Used in the Model Looping

The tables 1, 2 and 3 define and describe the symbols and operations used in the model in figure 1.

Table 1: Description of Data item.

Data item	Description
m	Message1(email)
m’	Message2(email)
d	destination
e	error
a	Accuse

Table 2: Description of activities.

Activities	Description
A1	Write email
A2	Corrected email
A3	Send email
N1	Decision Node

Table 3: Description of Operation.

Operation	Description
R	Read
W	Write
D	Destroy
I	Input
o	Output

3.2 Verification the Model with a Loop

We apply the Rule 1 of approach (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015) in the example below in figure 2. Rule 1:

"For an activity, a given data d with the state (x,y,z), if d is inserted for the first time in the DataRecord and x ≠ 0 we have an error ==> uninitialized data (missing data)."

The conclusion is that data item d is detected as missing data in activity A2 in figure 2 & table 4 below.

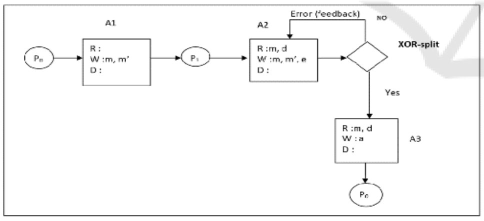

Figure 2: Data Flow modeling in linear model.

Table 4: Simple state without iteration.

Data	State
m	(0, W_{A1},0)
m'	(0, W_{A1},0)

The modeler chooses two solutions to correct the anomalies of missing data, either not to read the data item d in activity A2 (to destroy) or write data d in the activity A2. Then, data d is not read in activity A2. In this instance, before the first iteration, an error e has occurred at processing in the activity A2. Consequently, after drawing the model and using a connector Xor-split to have the conditions for the error e to occur. This model is considered as a finite state determinist model with Xor-split as a node. Indeed, if the error e is written in activity A2, the feedback loop returns the message for correction, otherwise, the process continues modeling. In fact, it is required to verify the anomalies at modeling to detect missing data in the two cases applying this approach (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015). Applying the Rule 1:

if e≠ ∅ it is noteworthy that the DataState (DataRecord) contains the latest data set, and activities as in table 5. At the verification, the loop cannot analyse the anomalies of data flow existed in processing, so no missing data is detected.

Table 5: The first iteration.

Data	State
m	(R_{A2}, W_{A1},0)
m'	(0, W_{A2},0)
e	(0, W_{A2},0)

Otherwise, if e=∅, the system continues with the modeling, and applies the Rules 4 and the process verification continues as in table 6, and the redundant data a is detected.

Table 6: The new dictionary with iteration.

Data	State
m	(R_{A3}, W_{A1},0)
m'	(R_{A3}, W_{A2},0)
d	(0, W_{A3}, 0)
a	(0, W_{A3},0)

We conclude, that for loop modeling with a connector Xor-split, even if they are based on data this will not change anything in this approach (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015), as the Rules are not practicable. Therefore, it's proposed to enrich this approach (MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi, 2015) to take into consideration the loop when the error in the message has occurred.

4 DESCRIPTION OF THE NEW APPROACH

4.1 Description and Definition

Dataflow is often defined using a model or diagram in which the entire process of data movement is

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

mapped as it passes from one component to the next within a program or a system, taking into consideration how it changes form during the process. Thus, data item of data flow has two roles; one is data link which is connects an activity to another by an input data and an output data. The two roles are to transmit the information from one task to another. In our situation data has the two roles and this data is extended by read and write/update and destroy. Consequently, when data transmit information there are many errors to be tackled such asMissing data, conflicting data and redundant data.

4.1.2 Decision Nodes

The decisions node is a conditional construct which can also be modeled with a conditional node using a logic Boolean predicates with a function guard that can allow us to model decision points in which the choice is made based on some data elements. When the model uses decision nodes, usually their edges have guards that are Boolean logic value specification evaluated at runtime to discover if control and data can be evaluated along the edge. Additionally, for each individual control and data token evaluated by the guards at the decision node to get precisely the edge that the token will be extended across. we can say that a decision nodes are Task nodes that represent atomic manual automated activities or subprocess that must be completed to fulfil the below business process objectives (SX S, JL Zhao, JF N,2006).

4.2 The New Approach

The approach based on method ad hoc applying an active help attended with the concept DataRecord for verification would be improved in order to be compatible with the loop. Consequently, it is suggested to enrich this method in order to be able to apply Rules on a loop. Additionally, an Xor-split is used by the model as a connector to feedback the exchange data flow when the error message has occurred. Indeed, it is proposed to use the Xor-split as a decision node with a data connector at an input data I(d). Moreover, this data connection is a routing decision which can be made based on a set of data items inputted to the decision node for a function to return the data when the feedback with a read operation in the activity A2 in which the error message occurred. But this proposition needs the concept DataRecord that it changed by a DataState which is to be initialized for each iteration and carry the latest activity and state of a dataset. The verification of detection data flow anomaly is the same, we keep an active help in an ad hoc approach. Indeed, each activity has an input data I(d) formalized as (RA,0,0), d is read in the activity A, and an output. O(d') in activity A' as (0,wa',0) d' is writen in activity A', e.g. if I(d) is accessed by activity A that produces an output data O(d') by activity A'. Otherwise, an input I(d) has to be processed for the next task and data will not updated. Thus, an input data I(d) is read in the decision node and data d is chosen to be read in the activity A2 in figure 3. In order to solve the above problem in the precedent section, the Decision Node N1 that is considered as an activity required by adding an input data I(d) as a decision variable. Additionally, a Boolean logic predicate (Yes, No) with a guard to specify the routing of a Decision Node N1 that (Yes =true & No=false) is added.

4.2.1 The Solution with a Decision Node

We began drawing the model for each task at the moment of modeling applying the proposed improvement of the approach. Therefore, if the error e exists ($e \neq \emptyset$) the guard is "No", in this instance, the system must be on feedback loop and read data d (destination) in the activity A2 which has been I(d) in the activity N1 that it couldn't be writen or updated in A2. Otherwise, ($e = \emptyset$) the guard is Yes, and no change in the next activity A3 the model continues processing. Consequently, the loops and Xor can only read the data; it can't delete nor create nor update data. In this case, the given data flow anomaly is only missing data. In this case, the missing data rule would be ameliorated, and the concept DataState manages the state of each data in the activities which have been reinitialised in each iteration, as in figure 3.

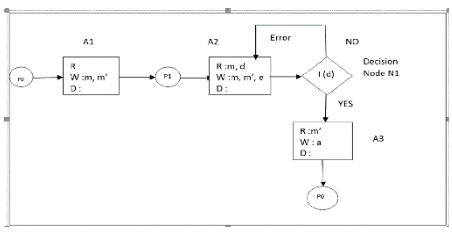

Figure 3: The conditional Node decision.

4.2.2 Rule Missing Data

For each iteration, DataState is initialized to record the new stored data set and activity and the decision node N1 has a guard No, the data d is in read in activity A2, involving the missing data.

if $d \rightarrow I(d)$ & $Pred\ (No)$, then d is read in the next activity; therefore, involving missing data.

4.2.3 Rule of Message Error

When an error e is created in a task, an output O(e) occurs, then the process cannot continue their task until the next activity due to the error e. Consequently, the loop starts, and if the guard is false (No), the system is loaded to correct the error at the next activity A2 (corrected email) in time of looping by destroying it.

4.2.4 Interpretation

We suppose that the error e is corrected when the system returns to activity A2 by applying Rule 2, and the error e is destroyed. At the same time when the preceding task is connected to node N1 by an input data d, the system has a choice to return the message by No to correct the error, that involve data d in reading in activity A2. Applying Rule 1, the system has detected an anomaly d in A2 missing data, after verification and locking the system, the modeler chooses to correct the anomaly with write data d in activity A2 as in DataState in table 7.

Table 7: The new DataState with loop and Decision Node.

Data	State	Iteration
m	(R_{A2}, W_{A2},0)	1
m’	(0, W_{A2},0)	1
e	(0, W_{A2}, D_{A2})	1
d	(0, W_{A2} ,0)	1

The decision Node has a connector input data I(d) which is not depended on the number of iterations. Consequently, our approach with a (rules 1 & rules 2) are valid for n iteration:
when $n > 1, n \epsilon N - \{0\}$ n iteration as in table 8.

Table 8: The DataState for n iteration.

Data	State	Iteration
m	(R_{A2}, W_{A2},0)	n
m’	(0, W_{A2},0)	n
e	(0, W_{A2}, D_{A2})	n
d	(0, W_{A2} ,0)	n

4.3 Future Work

In the future, we will try to find a solution for the other anomalies i.e conflicting data, redundant data and to verify this method with a model checking to verify of each instant the model by an active help at the moment the anomalies are detected, a computation tree logic CTL* in a temporal logic used subset LTL is to be investigated, and to find an example on which we can validate the model . (N Tr, W van der Aal, N Si, 2008; EM C, O Gr, D P - 1999; EA E JY Halper,1986).

5 CONCLUSIONS

The main focus of this paper is to use ad hoc approach and active help and a concept of DataRecord to verify data flow anomaly issues in loop modeling. In this case, we ameliorate existing rules by adding a decision node to make the grade by Yes or No. Also, we reinitialize DataState in each iteration. As such we have detected some anomalies of missing data, but we could not detect the others anomalies i.e. conflicting data and redundant data, because in the loop we can only read but not create, or modify. Consequently, we must ameliorate this approach in future work to detect more anomalies of data flow modeling in the business process.

REFERENCES

A Basu, RW Blanning - Information Systems Research, 2000 - pubsonline.informs.org A Formal Approach to Workflow Analysis.

TH Davenport - 1993 - books.google.com Process innovation: reengineering work through information technology.

M Weske - Business Process Management, 2012 – Springer Business process management architectures.

S Sadiq, M Orlowska, W Sadiq, C Foulger - Proceedings of the 15th …, 2004 - dl.acm.org, Data flow and validation in workflow modelling.

SX Sun, JL Zhao, JF Nunamaker… - Information Systems …, 2006 – pubso, nline.informs.org, Formulating the data-flow perspective for business process management.

LIU Cong, Q ZENG, D Hua - International Journal of Science and …, 2014 - ijsea.com, Formulating the data-flow modeling and verification for workflow: A petri-net based approach.

MI Kabbaj, A Bétari, Z Bakkoury, A Rharbi - IJCSA, 2015 - researchgate.net, Towards an active help on detecting data flow errors in business process models.

HS Meda, AK Sen, A Bagchi - 2007 - papers.ssrn.com, Detecting data flow errors in workflows: A systematic graph traversal approach.

AE Rgibi, SZ Yao, JJ Xu - Applied Mechanics and Materials, 2012 - Trans Tech Publ Dataflow Errors Detection in Business Process Model.

FJ Wang, CL Hsu, HJ Hsu - Computer Software and …, 2006 - ieeexplore.ieee.org, Analyzing inaccurate artifact usages in a workflow schema.

AES Rgibi - … and Industrial Engineering: Proceedings of the …, 2015 - content.taylorfrancis.com, Data flow issues and BPMN mapping to Petri Net: Road map.

N Trcka, W van der Aalst, N Sidorova - Computer science report, 2008 - research.tue.nl, Analyzing control-flow and data-flow in workflow processes in a unified way.

EM Clarke, O Grumberg, D Peled - 1999 - books.google.com, Model checking.

EA Emerson, JY Halpern - Journal of the ACM (JACM), 1986 - dl.acm.org, on branching versus linear time temporal logic.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS