

# Workflow Mining of More Perspectives of Workflow

Peng Liu<sup>1</sup>, Bosheng Zhou<sup>2</sup>

<sup>1</sup>School of Computer Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100191, China, <sup>2</sup>School of Computer Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100191, China  
Email: childbiggo@hotmail.com

Received November 27<sup>th</sup>, 2008; revised November 30<sup>th</sup>, 2008; accepted December 1<sup>st</sup>, 2008.

## ABSTRACT

*The goal of workflow mining is to obtain objective and valuable information from event logs. The research of workflow mining is of great significance for deploying new business process as well as analyzing and improving the already deployed ones. Many information systems log event data about executed tasks. Workflow mining is concerned with the derivation of a graphical process model out of this data. Currently, workflow mining research is narrowly focused on the rediscovery of control flow models. In this paper, we present workflow mining of more perspectives of workflow to broaden the scope of workflow mining. The mining model is described with GBMS's VPML and we present the entire model's workflow mining with the GBMS's VPML.*

**Keywords:** Workflow Mining, GBMs, VPML

## 1. Introduction

Workflow technology continues to be subjected to on-going development in its traditional application areas of business process modeling and business process coordination, and now in emergent areas of component frameworks and inter-workflow, business-to-business interaction. Addressing this broad and rather ambitious reach, a large number of workflow products are commercially available, which see a large variety of languages and concepts based on different paradigms.

In 1993, the standardization organization of workflow technology—Workflow Management Coalition (WFMC) was built. The definition of workflow in WFMC is as follows: workflow is concerned with the automation of procedures where documents, information or tasks are passed between participants according to a defined set of rules to achieve, or contribute to an overall business goal [1]. So, all kinds of activities of enterprises are organized and corresponded by business processes modeling and defined business logic relation. Workflow model is the process model that can be executed in Computer and its performance can be analyzed with the executing result. In selection of workflow engine, it must review the analysis ability of the process model.

Workflow mining means the knowledge discovery of workflow system, which can be induced by the definition of data mining. The ultimate target of workflow mining is to mine the transactional logs of workflow system, and to discover the knowledge of workflow including workflow models mining, workflow performance mining and workflow models improving.

Most workflow mining research is aimed at the rediscovery of explicit control flow models, which are used to specify the behavior of a process. We believe that

this approach limits the scope and utility of workflow mining by neglecting the important notion that workflow is much more than control flow. In fact, the behavior of a process is but one of its perspectives. In [2], Wang Lei and Zhou Bosheng present four major perspectives (Behavior Model, Information Model, Resource Model, and Coordination Model) of a process model that have emerged from the disciplines that helped shape the workflow area.

## 2. The Research Basis

### 2.1 GBMS and VPML

VPML [3] (Visual Process Modeling Language) is a graphic language that supports a special process definition. Process is a set of transformations of input elements into products with specific properties, characterized by transformation parameters; VPML describes the process with visual process diagram and relevant text specification. The diagram shows the structure of the process and the specification shows the attribute of all items in the diagram. The process has a higher degree in visualization and formalization. The process model built in VPML can simulate. It was proved that VPML not only describes a whole complete process model, but also is a process modeling language with rich functions, visual diagram, easy learning, flexible application.

GBMS [4] (Government Business Modeling System) is a modeling system oriented E-government. GBMS not only supports the Government business's process handling, modeling, simulating and optimal restructure

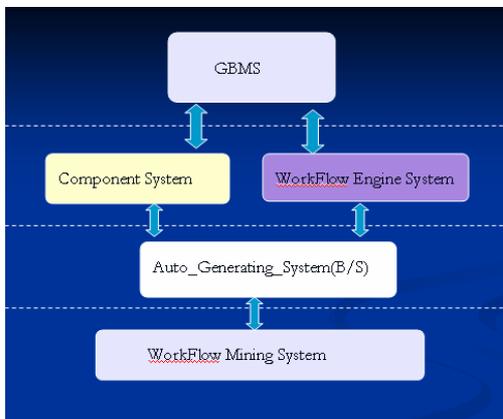
that implements interdepartmental business integration and the sharing of information resources, but also organic integrates the government business modeling with the requirement extraction and analysis of the E-government Business Application System. GBMS completely describes the business model in the view aspect of process, organization, resource, information and collaboration and is process-centered, organic integrates five views and keeps each view's consistency and integrity. GBMS lays the foundation for design and implementation of the Business Application System in E-government area.

**2.2 The Background and the System Architecture**

The research background of this paper is that The Construction of Oriented E-government Software Component Library. First, using GBMS, we build process model, organization model, information model, collaboration model, resource model and behavior model. The process model is the core of these models, which clearly describes the government business and finds out the shared resources. Meanwhile by accumulating plenty of the government business models, it not only gradually establishes E-government business pattern base, designs and builds component library oriented E-government and sharing in departments, but also unifies the standard of the government business process and avoids the duplicate investment of E-government. On the base of the component library, it executes the models on jBPM's workflow engine and takes the rapid response to the new requirement in the E-government area, assembles the components of the E-government component library, and rapidly builds relevant the E-government business application system. It can build the system on demand and reduce production costs.

The system's architecture is given in Figure 1. The whole project has 5 parts:

- (1) GBMS; (2) Component System; (3) jBPM Workflow Engine System; (4) Automatically Generating System; (5) Workflow mining System.



**Figure 1. The system architecture**

This paper extends the paper “The Research on the Modeling Transformation From GBMS to jBPM” [5], and presents how to mine the entire model from the E-government business system in order to validate the business model built in GBMS.

**2.3 Semantics of GBMS and VPML**

**DEFINITION 1.** GBMS as a 6-tuple

$PM = \{P, A, R, Control, Support, Input, Output\}$  where  $P = \{Product_1, Product_2, \dots, Product_j\}$  is a set of products;  $A = \{a_1, a_2, \dots, a_m\}$  is a set of Activities;  $R = \{r_1, r_2, \dots, r_k\}$  is a set of resources; *Control* is a set of relations of controlling; *Support* is the relations of resources supporting activities,  $Support \subseteq AxR$ ; *Input* is the relations of activities and input products,  $Input \subseteq AxP$ ; *Output* the relations of activities and output products,  $Output \subseteq AxP$ .

We need to define some assistant objects based on the above definition.

*State (i):*  $i \in P \cup R$ , represents the states of element i. It is enumeration type:  $enum = \{able, disabled\}$  means the states can be either able or disabled.

*Status (a):* where  $a \in A$ , represents the status of the activity a. In GBMS Model, activity has 2 statuses, *START* and *END*.

*Source(c):* where  $c \in Input \cup Output$ , means the set of source objects related by c.

*Target(c):* where  $c \in Input \cup Output$ , is a set of target object related by c.

*Prod\_Source(a):* where  $a \in A$ , represents the set of input products related with activity a.

*Prod\_Target(a):* where  $a \in A$ , means the set of all output products related with activity a

*Resource(a):* where  $a \in A$ , the set of all resources related with activity a.

*Role(a):* where  $a \in A$ , represents the set of all roles related with activity a.

**DEFINITION 2.** The GBMS Data Model is used to describe the information perspective of workflow. GBMS Data Model is the 4-tuple(P, A, Input, Output).

*Input(p, a):* is that the Activity a need Product p to run.

*Output(p, a)* is that Activity a is running to produce the Product p.

**DEFINITION 3.** The GBMS Organization Model is used to describe the organization perspective of workflow. GBMS Organization Model is the triple (A, R, Support): A is the set of GBMS's Activities; R is the set of resources in GBMS; Support is the relations of resources supporting activities;

*Support (a, r)* is that the Resource r supports the Activity a to run.

**DEFINITION 4.** S is a subset of P, which represents the set of all source products.

$$S = \{S \subset P \mid \forall c \in \text{Input} \cup \text{Output}, \\ \text{Source}(c) \cap S = \Phi\}$$

**DEFINITION 5.** E is a subset of P, which represents the set of all the final products.

$$E = \{E \subset P \mid \forall c \in \text{Input} \cup \text{Output}, \\ \text{Target}(c) \cap E = \Phi\}$$

**DEFINITION 6.** O(a) represents the constraints.

$$O(a) = \{a \in A, p \in \text{Prod} \_ \text{Source}(a), \\ r \in \text{Resource}(a), u \in \text{Role}(a) \mid \\ \text{Or} \quad \text{State}(p) = \text{able} \ \&\& \ \text{State}(r) = \text{able} \\ \&\& \ \text{State}(u) = \text{able}\}$$

### 3. The Workflow Mining Algorithm

It is theoretically proved that VPML is equivalent Petri-Net [6]. So this paper uses the mining algorithm based on the Aalst's  $\alpha$ -algorithm [7,8,9].

(1) Constructing dependence frequency table (D/F-table);

By given activities a and b, I) activity a and activity b's appearance frequency #a and #b; II) b is directly ahead of a : #b<#a; III) b directly succeeds a : #a>#b; IV) b is ahead of a: #b<<<#a; V) a is ahead of b: #a>>>#b; VI) the degree of dependence between a and b: #a→#b;

(2) Mining activities relation table (R-table) by D/F-table;

Based on the D/F-table, mining the basic activities relations ( $a>_wb, a\rightarrow_wb, a\#_wb, a//_wb$ ) [8];

(3) Reconstructing the workflow net by R-table and  $\alpha$ -algorithm.

### 4. More Perspectives of Workflow Mining

In this section, we present a sample of more perspectives of Workflow Mining. The workflow event log is shown in Table 1. First, we do some reasonable assumptions because of the workflow mining algorithm. We assume that events are logged in temporal order, the event logs do not contain noise, and the event logs are theoretically complete.

#### 4.1 Workflow Mining in the Behavior Model

The behavior model shows the behavior perspective of the workflow. This perspective is the basic and important of workflow. It is the performance of workflow system.

In the workflow event log table, it contains 3 cases. We use the algorithm in the 3<sup>rd</sup> section to mine the behavior model of GBMS.

(1) Definitude the basic relations between the activities using the D/F-table.

In the log , there are 3 cases,  $\sigma_1=\{A,B,C,D,E\}, \sigma_2=\{A,B,D,C,E\}, \sigma_3=\{A,F,G\}$ .

Then we get the relations between activities:

$$a>_wb: A>_wB, A>_wF, B>_wC, B>_wD, C>_wD, C>_wE, D>_wE, \\ F>_wG;$$

$$a\rightarrow_wb: A\rightarrow_wB, B\rightarrow_wC, B\rightarrow_wD, A\rightarrow_wF, F\rightarrow_wG, C\rightarrow_wE, \\ D\rightarrow_wE;$$

$$a\#_wb: B\#_wF$$

$$a//_wb: C//_wD, D//_wC;$$

(2) Using the  $\alpha$ -algorithm to mine the behavior model.

$$1) T_w = \{A, B, C, D, E, F, G\};$$

$$2) T_i = \{A\};$$

$$3) T_o = \{E, G\};$$

$$4) X_w = \{(A,B), (B,C), (B,D), (C,E), (D,E), (A,F), (F, G)\};$$

$$5) Y_w = X_w;$$

$$6) P_w = \{p(A,B), p(B,C), p(B,D), p(B,D), p(C,E), p(D,E), p(A,F), p(F,G)\} \cup \{i_w, o_w\};$$

$$7) F_w = \{(A, p(A,B)), (p(A,B), B), (B, p(B,C)), (p(B,C), C), (B, p(B,D)), (p(B,D), D), (C, p(C,E)), (p(C,E), E), (D, p(D,E)), (p(D,E), E), (A, p(A,F)), (p(A,F), F), (F, p(F,G)), (p(F,G), G)\} \cup \{i_w, o_w\};$$

$$8) \alpha(W) = \{T_w, P_w, F_w\}.$$

(3) In GBMS, the behavior model is described in the Figure 2.

**Table 1. Workflow event log**

Case	Activity	Status	Need	Produced	Resource
1	A	START	DOC1		ROLE1
1	A	END		DOC2	
2	A	START	DOC1		ROLE1
2	A	END		DOC3	
3	A	START	DOC1		ROLE1
3	A	END		DOC2	
1	B	START	DOC2		ROLE2
1	B	END		DOC4, DOC5	
3	B	START	DOC2		ROLE2
3	B	END		DOC4, DOC5	
1	C	START	DOC4		ROLE4
3	D	START	DOC5		ROLE4
1	D	START	DOC5		ROLE5
3	C	START	DOC4		ROLE5
1	C	END		DOC6	
3	D	END		DOC7	
1	D	END		DOC7	
3	C	END		DOC6	
1	E	START	DOC6, DOC7		ROLE6
1	E	END		DOC9	
3	E	START	DOC6, DOC7		ROLE6
3	E	END		DOC9	
2	F	START	DOC3		ROLE3
2	F	END		DOC8	
2	G	START	DOC8		MACHINE1
2	G	END		DOC9	

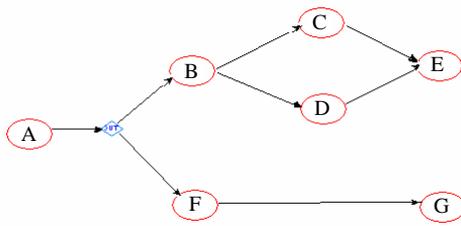


Figure 2. The GBMS behavior model

4.2 Workflow Mining in the Information Model

In workflow mining area, there have been no more researches in informational perspective. In this section, we introduce the workflow mining in the information model. The information model is used to describe the products consumed and produced in the business process and the relations between the products. In GBMS, the activity can be executed when it has products and roles that the states of those must be able to support. So the information model’s mining is very important and can show the data flow of the process.

In the workflow event log Table 1, it contains 7 activities and 9 kinds of products. In activity’s START status, it needs product to run; In END status, it can produce the new products. So in Table 1 of event log, we can get the formulas:

- 9) Prod\_Source(A)={DOC1};Prod\_Target(A)={ DOC 2, DOC3}
- 10) Prod\_Source(B)={DOC2};Prod\_Target(B)={ DOC4, DOC5}
- 11) Prod\_Source(C)={DOC4};Prod\_Target(C)={ DOC6}
- 12) Prod\_Source(D)={DOC5};Prod\_Target(D)={ DOC7}
- 13) Prod\_Source(E)={DOC6,DOC7};Prod\_Target (E)={ DOC9}
- 14) Prod\_Source(F)={DOC3};Prod\_Target(F)={ DOC8}
- 15) Prod\_Source(G)={DOC8};Prod\_Target(G)={ DOC9}

Input={Input(A,DOC1),Input(B,DOC2),Input(C,DOC 4), Input(D,DOC5), Input(E,DOC6), Input(E,DOC7), Input(F,DOC3), Input(G,DOC8)}

Output={Output (A,DOC3), Output (A,DOC2), Output (B,DOC4), Output (B,DOC5), Output (C,DOC6), Output (D,DOC7), Output (E,DOC9), Output (F,DOC8), Output (G,DOC9)}

Figure 3 illustrates the integration of the GBMS behavior model and data model.

4.3 Workflow Mining in the Organization Model

In GBMS, the organization model is used to define the government organization; it concludes all kinds of resources: person, machine, place, etc. In mining this aspect, a workflow analyst can discover if the participants of a business process are being used efficiently and effectively.

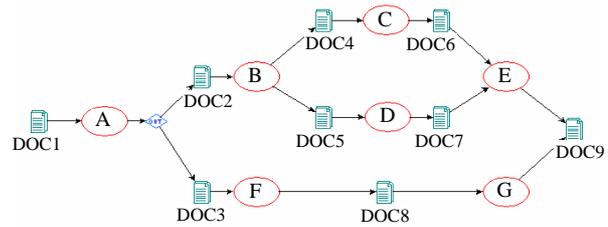


Figure 3. The rediscovered of GBMS behavior model and data model

In the workflow event log Table 1, it contains 7 activities, 6 roles and 1 machine. Activity needs the resource to support to run in the workflow system. From Table 1, we can get the formal representation of the rediscovered organization model:

- 1) Role (A)={ROLE1}
- 2) Role (B)={ROLE2}
- 3) Role (C)={ROLE4}
- 4) Role (D)={ROLE5}
- 5) Role (E)={ROLE6}
- 6) Role (F)={ROLE3}
- 7) Resource (G)={MACHINE1}

Support = {Support(A,ROLE1), Support(B,ROLE2), Support (C,ROLE4), Support(D,ROLE5), Support (E,ROLE6), Support (F,ROLE3), Support (G,MACHINE1)}

The Figure 4 shows the integration of the GBMS behavior model, data model and organization model.

This sample is a business process of Beijing Xuanwu government shown in Figure 5. It is a process of business application. First the applicant fills in the table of application, and sends the table for the approval. If the application is the type of common service, then it is assigned to Windows’ transaction, and distribute the app to different leaders to deal with, last it arrives the director to disposal. If the application is the type of administration permission, it is directly sent to the administrator to deal with and recorded into the database of Government. It was built by GBMS.

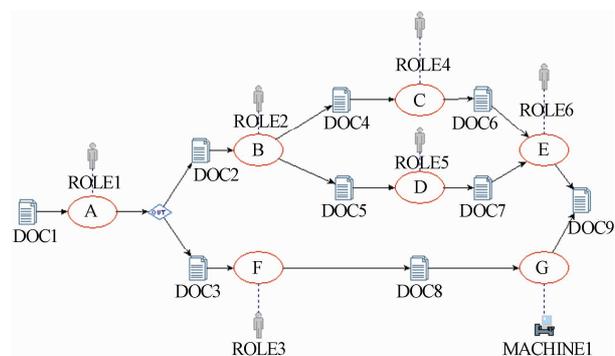


Figure 4. The rediscovered of GBMS behavior model, data model and organization model

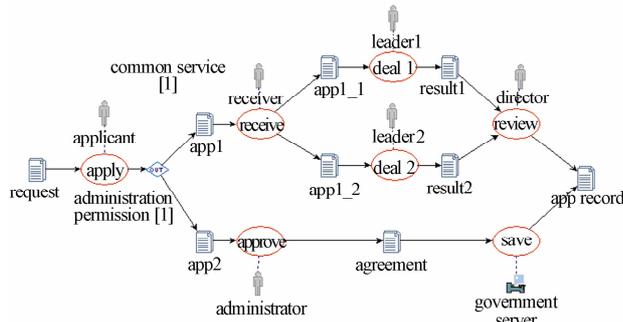


Figure 5. the application business process of GBMS

## 5. Conclusions and Future Work

Currently, most workflow mining algorithms have the goal of rediscovering a control flow model. We feel that this approach limits the scope and utility of workflow mining; it neglects the important point that workflow is much more than control flow. So this paper has presented the workflow mining in more perspective of workflow and given a full sample to show how to mine the entire GBMS model from the event log.

In future, we will study the incomplete log with noise and improve the mining algorithm. This is one research of the whole project. We also need to validate the mining model with the model design by the GBMS.

## 6. Acknowledgement

This project is funded by the Ministry of Science and Technology, P.R.C.

We should like to thank Beijing Cyber Technology for supporting the project.

## REFERENCES

- [1] Workflow Management Coalition, "Workflow management coalition terminology and glossary," Technical Report, WfMCTC-1011, Brussels: Workflow Management Coalition, 1996.
- [2] L. Wang and B. S. Zhou, "The study of enterprise model," *Computer Engineering and Application*, 1002-8331-(2001)12-0005-05.
- [3] B. S. Zhou, H. X. and L. Zhang, "The Principle of Process Engineering and an Introduction to Process Engineering Environments," *Journal of Software (supplement)*, pp. 519-534, August 1997.
- [4] B. S. Zhou and S. Y. Zhang, "Visual Process Modeling Language VPML," *Journal of Software (supplement)*, pp. 535-545, August 1997.
- [5] P. Liu and B. S. Zhou, "The research on the modeling transformation from GBMS to jBPM," 2008 International Conference on Computer Science and Software Engineering.
- [6] A. H. Ren, "Research on the Concurrent Software Developing Method Based on Object Oriented Petri Nets," BHU, The school of computer science, pp. 116-128, 2001.
- [7] W. M. P. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering*, 16(9), pp. 1128-1142, 2004.
- [8] A. J. M. M. Weijters and W. M. P. van der Aalst, "Process mining: Discovering workflow models from event-based data," in: B. Kroose, M. de Rijke, G. Schreiber, M. van Someren (Eds.), *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001)*, pp. 283-290, 2001.
- [9] A. J. M. M. Weijters and W. M. P. van der Aalst, "Workflow mining: Discovering workflow models from event-based data," in: C. Dousson, F. Hooppner, R. Quiniou (Eds.), *Proceedings of the ECAI Workshop on Knowledge Discovery and Spatial Data*, pp. 78-84, 2002.