Decision Support Systems xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Enabling effective workflow model reuse: A data-centric approach

Zhiyong Liu [a], Shaokun Fan [b,*], Harry Jiannan Wang [c], J. Leon Zhao [d]

[a] Faculty of Management and Economics, Dalian University of Technology, China
[b] College of Business, Oregon State University, United States
[c] Lerner College of Business and Economics, University of Delaware, United States
[d] Department of Information Systems, College of Business, City University of Hong Kong, China

ARTICLE INFO

Article history:
Received 4 July 2014
Received in revised form 5 September 2016
Accepted 6 September 2016
Available online xxxx

Keywords:
Workflow model reuse
Workflow model management
Data flow perspective
Data dependency

ABSTRACT

With increasingly widespread adoption of workflow technology as a standard solution to business process management, a large number of workflow models have been put in use in companies in the era of electronic commerce. These workflow models form a valuable resource for workflow domain knowledge, which should be reused to support workflow model design. However, current workflow modeling approaches do not facilitate workflow model reuse as a fundamental requirement, leading to a research gap in effective workflow model reuse. In this paper, we propose a novel approach called Data-centric Workflow Model Reuse framework (DWMR) to provide a solution to workflow model reuse. DWMR compliments existing control-flow-focused workflow modeling approaches by explicitly storing workflow data information, such as data dependency, data task relationships, and data similarity scores. DWMR also provides data-driven workflow model search and composition algorithms to satisfy user query requirements by automatically combining multiple workflow models. We demonstrate the feasibility of the DWMR approach by applying it to data from a well-known industry workflow model repository.

1. Introduction

With increasingly widespread adoption of workflow technology as a standard solution to business process management (BPM), much effort is focused on designing appropriate workflow models that satisfy the business requirements for a given company. Workflow models are designed to support complex process management in many business domains such as supply chain management, knowledge management and e-commerce [1,2]. How to efficiently design workflow models to satisfy particular business requirements is one of the key factors of BPM success [3].

Workflow models are used to capture domain knowledge about business processes in organizations. Many modeling methods have been proposed and applied by both academic researchers and business practitioners, such as Petri nets [4], UML activity diagrams [5], metagraphs [6] and the dataflow-based approach [7]. The dataflow-based approach is a relatively new method, in which designers develop workflow models by analyzing the input and output data of tasks and their dependency relationships [7].Compared with other workflow modeling methods that require users to provide plenty of information about specific processes, such as description of business function, sequence of tasks and structured representation of business rules, dataflow-based modeling methods focus on the input and output data in the business processes. Thus, it is easier to gather information required for the dataflow-based modeling method, because input and output data items are usually contained in the documents that business users deal with on a daily basis.

When designing workflow models for a specific business context, workflow designers always confront with the difficulty of collecting domain knowledge [8]. A typical business process involves many functional departments of an organization, such as purchasing, production, and sales, and requires knowledge from these varied business domains. However, process designers usually possess expertise in modeling, system design or software engineering, but little business domain knowledge like sales or production. Process designers always devote considerable effort to collecting information from workers at the front line. Even though domain knowledge can be acquired from interviews, surveys and field studies, there is always a significant gap between designers' and users' understanding, which is caused by their different knowledge backgrounds.

Rather than starting from scratch, designers can gain valuable insights from the existing process records [9], which can serve as a valuable source of domain knowledge. To a certain degree, the existing models reflect the operational mechanisms of an organization. A formal approach is needed to utilize the knowledge embedded in existing workflow models in a systematic manner. Model reuse will certainly

* Corresponding author at: 302 Austin Hall, Oregon State University, Corvallis, OR 97331, United States.
E-mail address: Shaokun.fan@OregonState.edu (S. Fan).

http://dx.doi.org/10.1016/j.dss.2016.09.002
0167-9236/

improve the quality and efficiency of business process design. Business practitioners have documented many workflow models in various business domains based on their knowledge and experiences, called Process Reference Models (PRMs), such as SAP Process Reference Model and Oracle Best Practice Processes [10]. So far, workflow model reuse mainly relies on function or control structure. However, little research has been done on workflow model reuse based on the data dependency, which takes the data flow of workflow models as the focus of model management.

We adopt a data-centric perspective for workflow model reuse for three reasons. First, data flow information can be directly acquired from users. Data collection is the starting point of business requirement collection when building workflow models. The users are usually familiar with the data that they are using and producing on a daily basis. By contrast, the control flow information that reflects complex logic is not explicitly documented in working environments; it is summarized by users based on their domain knowledge. Second, dataflow of workflow models is more intuitive and less error-prone than control flow [11]. The control flow-based models require integrating designers' in-depth understanding of model semantics and logic. Even the error rate of control flow models designed by professional companies is reported to be more than 10% [11]. Third, the data flow information is independent from workflow modeling paradigms and supports cross-paradigm model reuse. The solutions for workflow model reuse based on the control flow perspective are vitally influenced by the model specification paradigms [12,13]. The control flow-centric models from different sources can be expressed following different paradigms such as Petri nets [14], UML activity diagram [5] and BPMN. Considerable effort is spent on transforming control flow models and adjusting model management algorithms. Workflow models with different paradigms can be managed in a unified way following a data-centric perspective.

In this paper, we fill the research gap by proposing a framework for data-centric workflow model reuse (DWMR), which consists of model storage, model search and model composition via a data-centric approach. Our contributions are as follows. First, we define the formal data structure of a workflow model repository and propose a data-centric indexing method for workflow models based on data dependency relationships. Second, we propose a method for matching user requirements with candidate workflow models based on data similarity and develop a flooding algorithm to search for model groups that satisfy a particular user query. Third, we develop a formal method to compose candidate groups of models when needed to meet a specific user requirement. We evaluate the approach by applying it to a business case and comparing the DWMR approach with existing approaches and human experts.

2. Literature review

2.1. Workflow modeling

The control flow perspective of workflow modeling focuses on the sequence and coordination of tasks. The objectives of control flow modeling are controlling, monitoring, optimizing and supporting business processes [14]. Petri nets have become one of the most popular control flow-based modeling paradigms since van der Aalst [14] introduced it into the workflow modeling field. In addition to providing an appropriate language for workflow specification, Petri nets also serve as a powerful analytical tool for verification of workflow models. Several important concepts have been introduced to support the analysis of the correctness of control flow models, such as structuredness [15,16] and soundness [17]. Vanderfeesten et al. [3] introduced cohesion and coupling metrics to evaluate workflow model design. Van der Aalst has extended Petri nets to propose a workflow net concept and the YAWL language [18]. Some extended forms of Petri nets are also proposed to add more data information to workflow models, including colored Petri nets [19] and timed Petri nets [20].

Most of the modeling methods from the control flow perspective or graph-based perspective ignore the importance of data flow. They just focus on the coordination of tasks or rules, while the data requirements of models, data exchange between tasks and data dependency relationships are often overlooked or simplified.

2.2. Data flow perspective

The data flow perspective was first proposed in the software engineering field [21,22] and more recently introduced into the business process modeling field to detect data errors and correct workflow models [23,24].

Compared with traditional modeling methods that focus on the structural issues [17], data flow is considered as an important complement to the workflow specification. Sadiq et al. [24] identify the importance of data flow issues to workflow research and introduce modeling [25], specification and validation of data flow. They also illustrate the essential requirements for data flow modeling and seven types of data anomalies. Russell et al. [26] describe a series of workflow data patterns to show the different ways to represent and use data flow in workflow systems. Data flow analysis is a useful tool for workflow verification, which has been used to validate the correctness of control flows. Sun. et al. [7] propose a formal data flow specification and develop algorithms to detect data anomalies based on data dependency. The data flow modeling method is further extended to support process integration [27]. Vanderfeesten et al. [2] introduced a product-based workflow design approach, which adopts a *Product Data Model* to capture the data dependency relationships. A formal workflow language based on Petri nets and nested relational calculus has been proposed to support data flow modeling [28].

2.3. Workflow model reuse

Research on management and reuse of models has a long history in fields such as software engineering [29,30] and operation research [31]. Existing models form an abundant source of domain knowledge and best practices. In the BPM field, they are referred to as Process Reference Models (PRMs) [32]. Many commercial companies have provided packages of their process designs and best practice for different industries, including SAP Process Reference Models and Oracle Best Practice Processes [10]. Reference Models help model designers avoid starting from scratch. These packages cover many business functions that can satisfy most of a company's basic requirements. Because of the unique characteristics of different companies, Process Reference Models must be adapted to fit a particular context before being adopted [33]. Models with a higher abstract level are more generalizable, and more effort is needed to adapt them to meet the specific requirements of the end user.

Researchers have studied workflow model reuse from different angles. Wang and Wu [10] propose a collaborative approach to integrate and manage process reference models (PRMs). Their approach is based on a "spiral" model for organizational knowledge creation and leverages Web 2.0 technologies such as social tagging and classification to maintain PRMs. The model patterns (or model components) are also reusable. Zhuge [34] defines workflow components with four characteristics: independency, encapsulation, completeness and consistency. Thom et al. [35] classify workflow patterns into nine categories, and Altintas et al. [36] abstract workflows into several components according to their service functions. Cao et al. [37] define the workflow components, which consist of function, quality of service, control flow model and business domain.

The model design cases are also a reusable resource. Madhusudan et al. [38] propose a framework to reuse two categories of workflow cases, i.e., prototypical cases and instance cases. They also introduce a similarity-based flooding algorithm to support case retrieval. Similarity-based model matching is another important aspect of model reuse. Madhusudan et al. [38] support case retrieval by matching the NAME properties of nodes in models through language processing and string