Expert Systems with Applications 39 (2012) 10544–10551

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Hybrid genetic algorithm and association rules for mining workflow best practices

Amy H.L. Lim [a], Chien-Sing Lee [a,*], Murali Raman [b]

[a] Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia
[b] Graduate Institute of Management, Faculty of Management, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

ARTICLE INFO

Keywords:
DSS development-functionality
Development-methodology-business models
Business intelligence
Genetic algorithm
Performance measurement
E-commerce

ABSTRACT

Business workflow analysis has become crucial in strategizing how to create competitive edge. Consequently, deriving a series of positively correlated association rules from workflows is essential to identify strong relationships among key business activities. These rules can subsequently, serve as best practices. We have addressed this problem by hybridizing genetic algorithm with association rules. First, we used correlation to replace support-confidence in genetic algorithm to enable dynamic data-driven determination of support and confidence, i.e., use correlation to optimize the derivation of positively correlated association rules. Second, we used correlation as fitness function to support upward closure in association rules (hitherto, association rules support only downward closure). The ability to support upward closure allows derivation of the most specific association rules (business model) from less specific association rules (business meta-model) and generic association rules (reference meta-model). Downward closure allows the opposite. Upward-downward closures allow the manager to drill-down and analyze based on the degree of dependency among business activities. Subsequently, association rules can be used to describe best practices at the model, meta-model and reference meta-model levels with the most general positively dependent association rules as reference meta-model. Experiments are based on an online hotel reservation system.

1. Introduction

The process of finding and performing in-depth analysis on patterns/anomalies within the data is increasingly crucial as these patterns describe relationships among important decision variables which can increase the manager's understanding of the current state. The decision outcome subsequently informs the determination of strategies, which create competitive advantage. Five competitive forces that most organizations have to deal with are "*threat of substitute products, the threat of established rivals, and the threat of new entrants, the bargaining power of suppliers and the bargaining power of customers*" (Porter, 1979).

Assessment outcomes often result in adjustments to goals and tactics to outwit existing entrants in the market. Continuous assessment means that the above procedures will be carried out during specific time intervals to understand the current status of the organization and to evaluate the effectiveness of the implemented strategy and next possible actions to be performed.

This paper is an extension to our previous three works aimed at introducing approaches to assist continuous assessment and support an organization's management in decision making. *The first earlier work* highlights the need to increase the effectiveness of workflow management systems (WFMS) to allow reusability of effective and efficient workflow. This means that WFMS should be extended to provide analysis and create a pool of best practices. It can be achieved by creating a pool of best workflow practices within a repository. This is addressed with the introduction of the Weighted and Layered WF evaluation (WaLwFA) methodology (Lee & Lim, 2007). WaLwFA adopts the concept of Model-Driven Architecture (MDA) with aims to capture best practices, which can be adopted and instantiated to many domains or various information systems. In WaLwFA, the business process models are evaluated using a set of weighted criteria and sub-criteria, which are derived by averaging the assignment of weights by a group of experts. The business models with higher scores are kept within the repository to form business meta-models and reference model respectively. The repository will constantly be updated when there is a new "good" business model, thus ensuring that the repository remains updated with the latest best practice.

The *second earlier work* highlights the need to have an extensive reusable business performance measurement framework that can pinpoint causal relationships between the organization's current business performance and its future directions as well as measure the organization's workforces. To address the *second* problem, Integrated Model-Driven Business Evaluation (IMoBe) methodology is proposed (Lim & Lee, 2008). In this framework, a model-driven

* Corresponding author.
E-mail addresses: amy.lim@mmu.edu.my (A.H.L. Lim), cslee@cl.ncu.edu.tw (C.-S. Lee), murali.raman@mmu.edu.my (M. Raman).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.183

knowledge base serves as repository to store two or more commonly used business performance strategies or elements to evaluate any business organization such as Sun Tzu's 13 themes of business management strategies (Lee, Roberts, Lau, & Bhattacharya, 1998) reported in Ko and Lee (2000) as well as concepts of critical success factor (CSF) and critical barrier factor (CBF) (Niazi, Wilson, & Zowghi, 2005). IMoBe's methodology is an integration of several business performance approaches consisting of the Balanced Scorecard (BSC) (Kaplan & Norton, 1992) and the Quality Function Deployment (QFD) methodology. The selected business performance model will serve as a predictor measure and guide organizations on the next course of actions. The selected model contains criteria and they represent the rows in the House of Quality (HOQ) matrix while the columns are represented by strategic objectives identified for each perspective of the BSC. Customization of criteria is allowed where other criteria can be added to the HOQ's row in order to assist any organization in making efficient and effective actions.

The *third previous work* describes the need to improve existing evaluation methods in the first problem to extend from weighted evaluation to incorporating DSS with OLAP to improve the process of decision making where rules consisting of criteria as attributes can be hierarchically arranged and sorted according to its degree of complexity with each association rule having scores of support, confidence and correlation. For the *third* problem, a decision support system architecture consisting of our business performance methodology, namely WaLwFA, is extended to incorporate business intelligence capabilities to assist decision-making (Lim & Lee, 2010). The C4.5 decision tree algorithm (Quinlan, 1993) is used to discover significant attributes and association rule, namely the Apriori algorithm, (Agrawal, Imielinski, & Swami, 1993) is used to derive simple and complex association rules as well as to perform correlation analysis to calculate the dependency among attributes.

This paper extends from our DSS-OLAP study (Lim & Lee, 2010) to address the problems below.

2. Problem statements

Business workflow analysis has become crucial in strategizing how to create a competitive edge. Consequently, deriving a series of positively correlated association rules from workflows is essential to identify strong relationships among key business activities. These rules can subsequently, serve as best practices. In this paper, strong rules refer to rules that have correlation score of more than "1".

Second, typical association rule algorithm (AR) such as Apriori supports only downward closure. If a set of *n*-items has reached minimum support value, the subsets of items also meet the minimum support (Liu, Hsu, & Ma, 1999). Support and confidence must be pre-specified by the user prior to applying AR.

Upward closure as mentioned by Brin, Motwani, and Silverstein (1997) is "*constructive*". Suppose we have *n*-items that are correlated, then the superset associated with *n*-items is correlated as well. The correlation score of the superset should minimally equal to the correlation score of the *n*-items. Confidence measurement does not support upward closure. This can be seen in the example provided by Brin et al. (1997) in their paper.

3. Approach

Hybridization of genetic algorithm (GA) (Holland, 1975) with association rules is used to mine interesting association rules. The efficiency of the association rule algorithm (AR) in delivering quality association rules can be enhanced through GA and through correlation in GA's fitness function. GA is chosen because it mimics the human's ability to adapt to any form of environment. Furthermore, GA can perform global search as it employs greedy algorithm and thus can help to reduce search time in non-deterministic environments.

By optimizing fitness values, it can serve as threshold to prune the list of association rules within search space in order to obtain positive association rules only. By introducing correlation in GA's fitness function, it replaces the user's role in specifying support and confidence as pruning measurements in typical AR.

4. Significance of study

First, the result from our proposed hybrid application of genetic algorithms and association rules allows the managers to view a set of association rules sorted according to rule length, allowing drill-down into many levels-of-details. In other words, we can view a set of rules consisting of parent rule, which is the most general rule having two item sets, followed by sub-parent rule and so forth. Each rule describes the level of dependency among activities within a workflow. We can regard the most general rules as reference meta-model; sub-parent rules as meta-model and complex rules as a model of best practices.

Knowing many levels of details within the sets of association rules is an advantage to the managers during the design phase of an information system. Depending on the type of information systems to be built, managers can choose to adopt any of these types of rules/models. For example, if a manager is given a project to build a new, standalone information system, it would save the manager and his team's design time if they adopt the complex rule (model) and instantiate it to a platform-specific model. If the manager is given a task of building an information system, which is consolidated into an existing information system, then adopting the most general rule (reference meta model) is sufficient, considering that the embedded system serves as a sub-module within an existing system.

In keeping with current trends, the set of association rules which represent best practices are updated by re-applying the proposed hybrid algorithms. Managers are given access to monitor and be notified of every updated best practice, so that their team can rapidly enhance the existing information system or build new ones. The result is a constantly updated, convenient and easy-to-use information system.

Second, our hybrid GA with association rule (GA-AssocRule) supports both upward and downward closure. The following explains how our solution supports upward closure. Suppose that we can find a few association rules having positive correlation score such as $A \rightarrow B$ (length = 2), $A\&C \rightarrow B$ (length = 3) and $A\&C \rightarrow B\&D$ (length = 4) with the correlation scores of 1.05 respectively. We can identify that association rule length of 3 and 4 are related to association rule length of 2 because all association rules share common basic items, i.e., *A* and *B*, and can be found in the association rule of length 2.

The ability to support upward closure allows us to sort the positively-mined correlated association rules based on rule length and allows us to see the progression and transformation of association rules from basic rules to more complex association rules of increasing rule length. In other words, we are able to identify parent association rule, sub-parent association rules and other complex rules.

The sections below cover the following: Section 4 introduces existing works on process improvement approaches and advantages of using data mining to assist the organization in strategic planning. The subsections in Section 4 describe the association rule algorithm and GA as well as application of GA for rule learning. Section 5 presents the application of hybrid GA with association rules (GA-AssocRule) to an online hotel reservation system. Subse-