



ELSEVIER

Available online at www.sciencedirect.com



Decision Support Systems 42 (2006) 745–758

Decision Support
Systems

www.elsevier.com/locate/dsw

Integrating information retrieval and data mining to discover project team coordination patterns

Fu-ren Lin ^{a,*}, Kuen-jin Huang ^b, Nian-shing Chen ^b

^a*Institute of Technology Management, National Tsing Hua University, 101 Sec. 2 Kuang-fu Rd., Hsinchu 300, Taiwan*

^b*Department of Information Management, National Sun Yat-sen University, 70 Lien-hai Rd., Kaohsiung, 804, Taiwan*

Received 21 June 2004; received in revised form 17 April 2005; accepted 28 April 2005

Available online 6 September 2005

Abstract

This study integrates information retrieval and data mining techniques to discover project team coordination patterns from project documents written in Chinese. The coordination pattern of a project team describes the project execution process, including task category, execution sequence and duration, as well as the team member cooperation. The integration comprises two phases. The first phase extracts the most relevant keywords describing tasks executed by projects from unstructured or semi-structured documents using the mutual information estimate and the term weighting system. A concept hierarchy tree generated using the hierarchical clustering technique represents multiple levels of task categories. The second phase discovers project team coordination patterns through sequential pattern analysis. The proposed approach obtains encouraging results by mining coordination patterns from information system development projects. In the present era of the knowledge economy, the application of groupware to facilitate team coordination and collaboration streamlines the collection and analysis of project documents throughout the project life cycle. A project manager can visualize the project execution process of a team, and can anticipate the project outcomes based on discovered team coordination patterns. Accordingly, the proposed approach can be adapted to team projects that share certain characteristics with information system development projects.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Information retrieval; Data mining; Sequential pattern analysis; Coordination pattern

1. Introduction

A teamwork approach for executing projects involves various roles and tasks conducted within a period of time. Numerous business practices, for example new product development, information system

development, and so on, adopt the teamwork approach. Applying groupware to facilitate team coordination and collaboration streamlines the collection and analysis of project documents throughout the project life cycle. Team coordination patterns can be retrieved from these electronic documents, which may be either semi-structured or unstructured. The coordination pattern of a project team describes the project execution process, including the task category, se-

* Corresponding author.

E-mail address: frlin@mx.nthu.edu.tw (F.-r. Lin).

quence and duration, and the cooperation degree among team members. For example, consider an information system (IS) development project, which generally includes such tasks as requirement determination, modeling data and processes, database and interface design, implementation, testing, installation, and documentation [11], that has been completed by seven team members (e.g., A_1, A_2, \dots, A_7) in six weeks. During the first two weeks, two system analysts (e.g., A_1, A_2) worked together to identify customer requirements, and design data flow diagrams (DFD) and entity relation diagrams (ERD). The next two weeks were used for prototyping the system after designing the database and completing the structure charts. The other four members (e.g., A_2, A_3, A_4, A_5) took charge of these tasks. During the final two weeks, the finished system was tested, installed, and cut over to users. The main tasks, namely system testing, document editing, and user training, were performed by team members A_1, A_6 and A_7 . The discovery of coordination patterns may help the project management to assign tasks, allocate resources, and evaluate performance.

Techniques adopted for identifying team coordination patterns require joint efforts from researchers on information retrieval and data mining. Efforts to design methods for retrieving information from unstructured documents were presented in the Text Retrieval Conference (TREC, <http://trec.nist.gov/>). Data mining involves the exploration and analysis of large quantities of data to discover meaningful patterns and rules using automatic or semiautomatic methods [2]. Knowledge discovery in databases (KDD) is a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [7]. The KDD process can be divided into five phases: *selection, preprocessing, transformation, data mining, and interpretation and evaluation*. Common model functions in the current data mining practice include classification, regression, clustering, summarization, dependency modeling, link analysis, and sequential pattern analysis. However, KDD research has focused on structured data; thus existing KDD tools are of limited usefulness in mining knowledge from unstructured or semi-structured documents [8,9].

This work attempts to integrate information retrieval and data mining techniques for generalizing, visualizing, and forecasting project team coordination

patterns. The integration of information retrieval and data mining techniques is conducted in two phases. During the first phase, approaches developed in research on information retrieval are employed to label each document with a set of keywords that represent the main concepts extracted from the document. These keywords are converted to a structured format, and represent the tasks, participants, resources, and time involved in the project execution. After obtaining the structured project description, the second phase is conducted, which applies sequential pattern analysis to generalize the coordination patterns, and then displays them in graphic user interfaces. Sequential pattern analysis is a data mining technique that generalizes sequential patterns from transactions occurring in different time periods. A project manager can compare the progress of on-going projects with the generalized coordination patterns to predict the project outcomes. The integration of information retrieval and data mining techniques contributes to the discovery of project team coordination patterns by using sequential pattern analysis. The integration of the aforementioned two techniques for mining knowledge from unstructured or semi-structured documents sheds light on project management.

Research on the PLANMINE system relates closely to the present study [23]. The PLANMINE sequence mining algorithm predicts plan failures based on extracted event patterns. Each plan is labeled as good or bad depending on whether it succeeds or fails to achieve its goals. PLANMINE attempts to find event sequences that can be used to confidently predict plan failure. First, the planner uses simulation to generate a database of good or bad plans. These plans are fed into the mining engine to identify event patterns of bad plans, which are used to fix the plan and prevent failures. The pattern generation and plan modification loop is executed repeatedly until no further improvement is obtained. Second, the high confidence patterns are used to generate a plan monitor that notifies a plan manager before the failure of a new plan. PLANMINE was demonstrated using two planning applications: TRIPS and IMPROVE. TRIPS is a collaborative planning system for designing an evacuation plan by using simulation and data mining for plan analysis. IMPROVE first simulates a plan repeatedly and calls PLANMINE to extract high confidence rules for predicting plan failure. IMPROVE then

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات