Knowledge Extraction and Sharing in External Communities of Practice

Ajumobi Udechukwu
University of Calgary, Canada

Ken Barker
University of Calgary, Canada

Reda Alhajj
University of Calgary, Canada

INTRODUCTION AND BACKGROUND

Communities of practice (CoPs) may be described as groups whose members regularly engage in sharing and learning, based on common interests (Lesser & Storck, 2001). Traditional communities of practice exist within organizations and are centered on work functions. These CoPs may be self-organizing or corporately sponsored. They exist to encourage learning and interaction, create new knowledge, and identify and share best practices for the organization's processes (Wenger, 1998). The members of a community of practice may be collocated (within an office) or spatially dispersed (e.g., a group may interact via electronic chat). There may also be communities of practice that are not centered on work functions. For example, several online groups exist for enthusiasts of new technology, politics, environment, and so forth. These groups qualify as bona fide CoPs. We classify the CoPs discussed so far as *active* communities of practice because the members actively seek to learn and share from each other. In this work, however, we examine *passive* communities of practice in which the members do not actively interact with each other. This class of CoPs shares the core characteristics of traditional communities of practice—the members can learn from each other, and the organization can gain useful knowledge capital and best practices. Our discussions will be based on user communities using cable-TV viewers as a case in point. In contrast to work-centered CoPs whose members share knowledge and learn how to perform their work tasks better, members of user-centered CoPs learn how to maximize the utility from the product/service of interest. In both cases, a learning organization can extract useful knowledge capital and best practices to improve its processes and products/services. In this work, we use the case of cable-TV viewers to show how useful knowledge can be learned and shared in passive user-centered communities of practice. Our techniques will be based on data mining and knowledge discovery, which are introduced in the subsection that follows.

DATA MINING AND NAVIGATIONAL PATTERN DISCOVERY

The widespread use of computers and the increased abilities to collect and store massive amounts of data have led to phenomenal growth in the popularity and use of data mining techniques. Data mining is the analysis of data with the goal of uncovering hidden patterns. Historically, technological advances that improve the collection of data have led to new domains for data mining. For example, advances in bar code technology and the ability to collect and store transaction data logs led to the growth of association rule mining (Agrawal, Inielinski & Swami, 1993) and its many variants (Fayyad, 1998). More recently, the widespread use of the World Wide Web and the ability to collect and store Web logs of user sessions have driven research interest in Web usage mining (Cooley, Mobasher & Srivatava, 1997; Srivastava et al., 2000). An interesting problem in Web usage mining that has attracted the attention of several researchers is the discovery of traversal patterns of Web users (Chen, Park & Yu, 1998; Nanopoulos & Manolopoulos, 2000). Mining path traversal patterns involves identifying how users access information of interest to them and travel from one object to another using the navigational facilities provided. Tracking user-browsing habits provides useful information for service providers and businesses, and ultimately should help to improve the effectiveness of the service provided. Previous works on identifying path-traversal patterns have been directed at traversals between relatively static objects (e.g., Web pages). By static, we mean information that can be regenerated by the user as required. Thus, dynamic Web pages fall under our definition of static objects because the user may regenerate the dynamic Web pages on each visit.

In this research, our focus is on navigational patterns in environments where the objects are continuously changing in time (i.e., streaming content). An example of such a system is cable-TV where the program sequence is continuously changing. The viewers of cable-TV are able to navigate from one object (channel/station) to another. However, if viewers navigate away from a station/channel and later return to that channel, the content displayed may have changed. Thus, there is a strong temporal component in the systems studied in this research. The temporal component in our framework motivates new techniques to capture navigational patterns, as existing techniques in the literature do not take temporal semantic information into consideration. Our framework can be applied to identifying navigational patterns in any environment with streaming content. However, the discussions in this article will be motivated by cable-TV viewing patterns. The choice of cable-TV viewing patterns is due to recent technological innovations that enable the collection of anonymous logs of viewing data through digital video recorders attached to cable-TV receivers. The logs are kept anonymous to protect the privacy of the viewers. This is similar to the ethical standards that have long been adopted in analyzing Web and transaction logs. In the past, there has been very limited ways to collect data on the viewing patterns of cable subscribers. The advent of digital video recorders and the ability to track and report logs on the channels viewed by subscribers (on a second-by-second basis) opens up several interesting areas for data mining. Digital video recorders are growing in popularity (with Tivo Inc. reporting over 700,000 subscribers in the US in 2003 and a projection of over a million subscribers by the year-end), and a massive deployment is expected in the near future (Tivo Inc., 2003). Digital video recorders keep track of the channels viewed through the cable receiver. The view logs are uploaded to the service provider daily. The challenge is to extract interesting patterns from all the view logs submitted to the service provider.

There are several interesting questions that can be addressed by analyzing the view logs. For example, an advertiser may be interested in knowing if more viewers stay tuned during the commercial breaks of prime-time programs than for regular programming. It may also be of interest to know the advertising slot that is most effective; that is, is it more likely for an advert to be viewed if it is the first ad during the commercial break or if it has the last slot, middle slot, and so on. It may also be interesting to discover the percentage of viewers that return to a program once they tune off during a commercial break. Several other interesting patterns may also be discovered. In our framework, we propose a novel technique that categorizes the dynamic content of sites into distinct event sequences and then explores the navigational patterns of users relative to the distinct event sequences. The behavioral/navigational patterns discovered may be used to improve the program sequencing for future broadcasts. The analysis may also be given a spatiotemporal dimension so that appropriate programming is directed at users based on their locations and times of broadcast. Viewers may be grouped or profiled based on common navigational behavior. In interactive TV environments, this would enable personalized programming and program recommendations tailored to particular viewer groups or individual viewers.

RELATED WORK

Several authors have studied communities of practice (CoPs) in organizations (Brown & Duguid, 1991; Hildreth, Kimble & Wright, 2000; Lesser & Prusak, 1999; Lesser & Storck, 2001; Wenger, 1998; Wenger & Snyder, 2000). These works are centered on work-related CoPs and differ from user-centered CoPs discussed in this article.