ABSTRACT

Mobile user data mining is about extracting knowledge from raw data collected from mobile users. There have been a few approaches developed, such as frequency pattern (Goh & Taniar, 2004), group pattern (Lim, Wang, Ong, et al., 2003; Wang, Lim, & Hwang, 2003), parallel pattern (Goh & Taniar, 2005) and location dependent mobile user data mining (Goh & Taniar, 2004). Previously proposed methods share the common drawbacks of costly resources that have to be spent in identifying the location of the mobile node and constant updating of the location information. The proposed method aims to address this issue by using the location dependent approach for mobile user data mining. Matrix pattern looks at the mobile nodes from the point of view of a particular fixed location rather than constantly following the mobile node itself. This can be done by using sparse matrix to map the physical location and use the matrix itself for the rest of mining process, rather than identifying the real coordinates of the mobile users. This allows performance efficiency with slight sacrifice in accuracy. As the mobile nodes visit along the mapped physical area, the matrix will be marked and used to perform mobile user data mining. The proposed method further extends itself from a single layer matrix to a multi-layer matrix in order to accommodate mining in different contexts, such as mining the relationship between the theme of food and fashion within a geographical area, thus making it more robust and flexible. The performance and evaluation shows that the proposed method can be used for mobile user data mining.

Keywords: mobile knowledge extraction; mobile user data mining

INTRODUCTION

Data mining (Agrawal & Srikant, 1994, 1995; Chen & Liu, 2005; Xiao, Yao, & Yang, 2005) is the field of research which aims to extract useful and interesting patterns out from source datasets supplied to the algorithm. Data mining is an emerging field which allows organisations such as business and government who have a huge amount of datasets stored in
very large database to be able to benefit from the algorithms by converting datasets into patterns and eventually studied and becomes useful knowledge. Data mining is still an ongoing research, and previously available outcomes from data mining include association rules, sequential patterns which derives useful patterns by analysing market basket (Agrawal & Srikant, 1994, 1995), which is the list of items customers buy in a supermarket. Other previously proposed methods in data mining includes time series analysis (Barbar’a, Chen, & Nazeri, 2004; Han, Dong, & Yin, 1999; Han, Gong, & Yin 1998), brain analysis (Claude, Daire, & Sebag, 2004), Web log pattern analysis (Christophides, Karvounarakis, & Plexousakis, 2003; Eirinaki & Vazirgiannis, 2003; Wilson & Matthews, 2004), increasing overall efficiency of data mining in very large databases (Han, Pei, & Yin, 2000; Li, Tang, & Cercone, 2004; Thiruvady & Webb, 2004), data mining on data warehouses (Tjioe & Taniar, 2005), security of private data in data mining (Oliveira, Zaiane, & Saygin, 2004) and spatial, location dependent data mining (Hakkila & Mantyjarvi, 2005; Koperski & Han, 1995; Lee, Xu, Zheng, & Lee, 2002; Tse, Lam, Ng, & Chan, 2005).

Mobile user data mining (Goh & Taniar, 2004a, 2004b, 2005; Lee, Xu, Zheng, & Lee, 2002; Lim, Wang, Ong, et al., 2003) is an extension of data mining which specializes in looking at how useful patterns can be derived from the raw datasets collected from mobile users. In a mobile environment, two types of entities can usually be found: static nodes, which are fixed entities such as the wireless access points, and mobile nodes, which are the mobile entities which have the flexibility to move along in the environment, such as the personal digital assistant, mobile phones, and laptop computers. The raw datasets from mobile users comes from the physical movement logs of mobile users, the items that mobile users purchased over time, the location of static nodes and their properties and the context in which the mobile users went into over a timeframe.

This article aims to propose a new method for finding relationships among two locations in a mobile environment in order to determine the nature of how mobile users visit them. This mobile user data mining method could reduce the consumption of resources for mobile user data mining by using the strategy of gathering data only to the extent that is relevant for the desired accuracy. In this proposed method, the covered area for mobile user data mining is first surveyed and mapped into a matrix, which could be a dense or sparse matrix depending on the amount of items marked into the matrix. Once this mapping is done, the matrix is used when the mobile user starts visiting the physical locations and the visiting behaviours are recorded based on the position in the matrix which the mobile users have contacted. This data is then used for data mining purposes, thus reducing the requirement for constantly identifying and gathering of the latest position information of the mobile users. The elimination of the