ABSTRACT

The pattern is special kinds of data which are created through various data mining techniques and stored in the pattern warehouse through a specialized pattern management system (PMS). Pattern warehouse makes the pattern non-volatile or persists. Now a day’s persistent pattern retrieval is a very new and important issue. This paper focuses on problems and challenges with pattern retrieval. One can see the applicability of classification in pattern retrieval as an opportunity and trying to bring attention to probable issues and challenges behind the physical implementation of this concept. This paper concluded that the applicability of classification in pattern retrieval is well feasible. It has also discussed that how of pattern classification is different with data’s classification. Classification method should be defined in such a way that it can handle pattern efficiently. So far, little emphasis has been posed on developing an overall classification system for pattern retrieval. This paper concerns only association kinds of patterns. It has presented some issues regarding (i) Decision boundary of pattern classes. (ii) Problem of calculating a reliable estimate of pattern classes. (iii) How to define class boundary (iv) How to handle overlapping of pattern classes (v) Parameter selection for pattern classes estimation (vi) Preprocessing of patterns (vii) How to handle classification on demand. (vii) Updating of pattern classes (vii) Finding optimal test conditions.

Keywords: Classification, Data Models, Knowledge Warehousing, Pattern-Base Management Systems, Pattern Bases, Pattern Retrieval, Pattern Warehouse

1. INTRODUCTION

Now a day’s many applications generating huge volumes of data (Vazirgiannis et al., 2003). Many techniques have been developed to extract knowledge, especially in the context of data mining. The results of such operations are abstract and compact representations of the original data, which called patterns (Catania, 2004). The pattern gives the semantic representation of raw data. There is required raw data for pattern and vice-versa mapping (Zdenka, 2012).

DOI: 10.4018/ijbir.2014070101
The volume of extracting patterns from various knowledge discovery applications is increasing rapidly, so there is need for effective and efficient pattern management system (Fernández et al., 2011; Mohammad et al., 2009; Vaziriannis et al., 2003). Patterns can come from clustering, association rules, classification, probabilistic rules and many other ways. Moreover, patterns can be found in images, signals, text and music and of course in the World Wide Web (Barbara and Anna, 2005). Theses all kinds of patterns are stored in pattern warehouse through PBMS (Pattern Base Management system) (Catania, 2004; Jaesoon et al., 2002; Manolis et al., 2007). There are no tools or systems that are designed to deal exclusively with pattern retrieval. All current technologies just try to extend the existing query system to get patterns, but this is not effective because of a special characteristic of patterns and its diversity (Evangelos et al., 2005; Rizzi, et al., 2003; Rizzi, 2004). Pattern retrieval can be improved through applying data mining techniques itself. Problem of diversity or complex structure can be solved by grouping relevant patterns. It is called classification of patterns. This task is not easy because patterns are semantically rich and have different and specialized structure. The main issue of this paper is revealing the difficulties with classification of association type patterns in the context of pattern retrieval. Association rule mining is a popular method for discovering interesting relations between variables in large databases (Thakur et al., 2006; Tiwari and vipin, 2010). Patterns can be a set of association rules discovered from databases using different measures of interestingness (Riccardo et al., 2011).

The objectives of this paper are to discuss the need of pattern management, benefit of pattern classification in context of retrieval and discuss the problems and challenges with pattern classification. The aim only point out the problems which may occur in applying data mining techniques for pattern retrieval on pattern warehouse rather than generating patterns itself. Classification is one of the powerful data mining techniques. For simplicity, research work revolves around the challenges with the classification of patterns.

2. PROBLEM FORMULATION

Pattern warehouse is very new concepts and little emphasis has been given till date. A pattern warehouse is as attractive as data warehouse as the main repository of an organization’s historical pattern and is optimized for reporting and analysis (Manolis and Vassiliadis, 2003; Mohammad et al., 2011). By nature, patterns are not persistent. Means each time when you need patterns need to execute pattern generating method again and again (Tiwari and Thakur, 2014). Pattern warehouse is a way to make the pattern persistent by storing them permanently. In this work, we try to bring the attention on pattern retrieval by incorporating existing data mining technique for pattern retrieval. Classification is one of the famous and tested data mining techniques. In this section, we are discussing various issues and challenges behind the applicability of classification in context of pattern retrieval. Pattern discovery system must be reliable and efficient for end-user point of view (Catania, 2004). Pattern classification is not a single or independent process. It consists of mainly four parts:

1. Data to pattern conversion;
2. Make patterns, persistent through pattern warehouse;
3. Classify pattern from pattern warehouse;
4. Query processing on classified pattern to result/ knowledge retrieval.

This work is carried out for discussing challenges with pattern classification only. For simplicity association type patterns are taken into consideration. Table 1(a) shows the transaction database. This is stored in the data warehouse. Patterns are generated through applying data mining techniques (association mining) on this transaction database. Patterns
Production Planning Based on DEA Profit Efficiency Models
[www.igi-global.com/article/production-planning-based-on-dea-profit-efficiency-models/181780?camid=4v1a](www.igi-global.com/article/production-planning-based-on-dea-profit-efficiency-models/181780?camid=4v1a)

The Performance Mining Method: Extracting Performance Knowledge from Software Operation Data
[www.igi-global.com/article/the-performance-mining-method/132821?camid=4v1a](www.igi-global.com/article/the-performance-mining-method/132821?camid=4v1a)