Improved Latent Semantic Indexing-Based Big Data Mining Methods for Customer Relationship Management

Junzo Watada, Jianxiong Yang, Mohd Hilmi Hasan, Yukio Kodono

INTRODUCTION

Knowledge representation and manipulation play a pivotal role in many Artificial Intelligence (AI) methodologies. The knowledge representations have been the subject of various studies in the AI community. The purpose of knowledge representation is to make AI programs behave as if humans do in problem-solving. Knowledge representation schemes include logic programming, semantic networks, procedural interpretation, production systems, and frames. Knowledge representations play a pivotal role in automated optimization systems, expert systems, knowledge bases, and inference machines [1].

As AI applications, sometimes, data mining and text data mining refer to the process of deriving useful information from the data set. High-quality useful information can be derived by statistical data analysis and through data classification. Text mining usually includes the process of structuring an input text (parse the addition of linguistic attributes and others), deriving the patterns with structured data, interpreting and evaluating the output of patterns. In data mining, “high quality” used to refer to some completed relevance and novelty. Characteristic data mining tasks include data categorization, data extraction, data clustering, purpose analysis, document summarization, and create an entity of relation models [2]-[5].

The relational equations of fuzzy clustering play important roles in many applications, such as intelligence technology [6]-[7]. So, how should we compute the fuzzy relational equation solution is a fundamental problem? Numerous research works are found to investigate the solvability of the fuzzy relation equation by extending and generalizing the original results of various directions [6],[8]. Optimization problems have been widely studied with the constraints of the fuzzy relation equation [5], [9]-[11].

The method is comprised of rough set models [12] and fuzzy clustering which together deals with the uncertainty included in data. In the proposed method, we employed the rough set model to optimize the data mining knowledge, and then sample data are applied for data extracting, Finally, fuzzy clustering is applied to cluster the target customers by data as samples to extract some useful customers. The process of provided algorithm and its applicability are illustrated by using an application of the proposed method to the knowledge mining of Customer Relationship Management (CRM), and this shows that the improved fuzzy clustering approach should have wide-ranging applications in industries [13].

Latent semantic indexing (LSI) or Probabilistic latent semantic indexing (PLSI) can realize a higher accuracy rate for multi-attributes data analysis. So, this improved LSI method based on a rough set enables us to analyse customers' preferences and to give them appropriate services based on their preference for firmly catching the heart of faithful customers. If we can catch the heart of faithful customers successfully, we can remain invincible in the business [14]-[16].

LATENT SEMANTIC INDEXING (LSI) BASED FUZZY CLUSTERING ANALYSIS

Latent semantic indexing (LSI) is the extracting and indexing methodology that uses the mathematic method of Singular Value Decomposition (SVD) to distinguish models between the documents and terms in the relationships. Based on the principle, the words used in similar documents tend to have some similarities. By creating the associations between those terms that appear in similar documents, the pivotal attribute of LSI is an attribute that can extract conceptive contents in the principal part of documents [8].