ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 13, No. 1, 2018, *pp*.003-021

Efficient Approach for Land Record Classification and Information Retrieval in Data Warehouse

C.B. David Joel Kishore [1] and T. Bhaskara Reddy [2*]
[1] *Research Scholar, Department of Computer Science, Rayalaseema University, Kurnool, Andhra Pradesh, India.*
[2] *Professor, Department of Computer Science & Technology, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India.*
(Received October 21, 2017, accepted December 18, 2017)

Abstract. A data warehouse collects the recent and ancient data that are used for creating analytical reports and put together to produce useful information. Retrieve the information accurately from a large source of data is a challenging task. A novel ANN-FUZZY-CSO approach is proposed to predict and retrieve the information accurately. First, the artificial neural network (ANN) classifies the input data for ordering the information to construct a database as different classes. Then, the mongo database will store a large amount of data for facilitating easy maintenance, prompt updating of land records and security. After that, the optimized fuzzy ranking function is used to retrieve the information from the database based on the optimal fuzzy rules using cat swarm optimization algorithm. The fuzzy rules provide a ranking for an individual field in the database. The accurate results for the user query are retrieved using cat swarm optimization (CSO) algorithm. The optimized fuzzy rules allow the users for easy access to their records. Finally, the performance is evaluated for the retrieval results

Keywords: artificial neural network, cat swarm optimization, data warehouse, fuzzy, mongo database

1. Introduction

A data warehouse is a structure utilized for the data analysis and storing the amount of information. Also, the data warehouse is a central repository which can be placed all data relevant to the management of the organization and which provides the prominent information to effectively manage the organization [1]. The information stored in the warehouse is in the scheme of derived views of data from the sources [2]. The data warehouse defines how data's are composed, merged, elucidated, trained, and that should be training-based managerial information [3]. They can determine the facts that facilitate managers to analyze situations requiring action and to know the situation and its intention. They must be established that authorize a manager to place and utilize the relevant organizational knowledge which predicts and scope the impact of a declaration added. These are the significant challenges for the research community in data warehousing [4]. For finer declaration, a data warehouse is processed to keep the large database.

The big data technologies used for storing an infinite number of data and we make a good decision [5]. So, they proposed remarkable uses of the bid data technology in the data warehouse. Big data is a major technology to analyze the large volume of data and extract the necessary information or knowledge for future action [6]. As a rule, the knowledge extraction process must be extremely effective and near real-time on the grounds that putting away all watched information is about hopeless. Also, the big data application using the large data sets, has a complex structure which is trouble of storing, analyzing and visualizing the results are varied [7-10]. The action of research into big amounts of data to reveal invisible patterns and private correlations named as big data analytics. These data's are essential to improve achieving richer and deeper insights and getting an advantage over the competition in the companies or Institutions. For this reason, big data implementations are a demand to be analyzed and completed as exactly as possible.

In existing, the large-scale real-world taxi trace data of a big city with a population of 6 million that are utilized to store in big data technologies. There the big data improved clustering algorithm (iterative DBSCAN) is presented to extract regions, respectively to the quality of the data. Six kinds of features are designed, four classifiers are combined into the evaluation. The achievement of different features and classifiers is evaluated, and the best feature combination is also achieved. But the system considers only current addresses regions with pure land use that does not consider regions with multiple land-use classes. So the Information of the passenger cannot be accessed clearly [11]. Another one sales forecasting for utilizing fuzzy logic and Naïve Bayesian for large data. There, the sales forecasting for any product is

'Poor', then it gives hint that some preventive measures like discount or offer should be taken to enhance the sale. If sales forecasting for any product is 'Good' then it gives hint that enough amount of stock should be made available in the store. So the system having low preparing time to given poor accurate results [12]. A fast Messy Genetic Algorithm used in an Advanced Hybrid Genetic Algorithm with support vector machine (SVM) system was presented. The early prediction of conflict disposition in the initial stage of public-private partnership projects using fmGA. The system providing a term of accuracy, precision, sensitivity, specificity, and area under the curve. But the system has some classification problem so that doesn't give the efficient result [13].

Another current technique is the land cover classification with better resolution remote sensing information coordinating worldly elements extricated from time arrangement coarser resolution information. The coarser resolution vegetation index information is initially intertwined with better resolution information to acquire time arrangement better resolution information. The coarser resolution vegetation index data is first combined with finer declaration data to obtain the time series. The fleeting elements are removed from the intertwined information and added to enhance classification accuracy. But the classification accuracy is lower than the other classification techniques [13]. Another one the big data using service-oriented decision support systems. The system given better results such as reduction in unit service costs due to increase in operational size (scale), reduction in unit service costs due to increase in number of services being developed and provided (scope) and reduction in unit costs as a result of increase in number of benefits are put through supply/demand chain (speed). But the result is not accurate and does not efficient [14].

The major contribution of the proposed work:

1) A classification model is proposed for gathering the facts to arrange a database concerning land profit, grazing method, and land use.
2) A novel optimized fuzzy based ranking approach for facilitating easy maintenance and prompt updating of land records.
3) The Cat Swarm Optimization (CSO) algorithm will allow the users to easily access their records optimally.
4) The mongo database used in our work will make the land records secure.

The remaining paper is organized by the following sections. Related work is described in section 2. Section 3 defines the proposed methodology in detail. The Experimental result is discussed and comparison of various methods are explained in section 4. Finally, Section 5 provides the conclusion.

2. Related Work

Torsten Priebe *et al.* [15] have proposed a combined methodology to structure and describe business requirements in large data accelerated projects, e.g. data warehouse employment, that precise and explicit data definition suitable to further accession and assignment of data governance responsibilities. The information can be placed in the center of the business model. The Data Administrator gave the information with the help of end-to-end analysis, design, development, testing to data quality checks. In addition, they display that the method is proper beyond conventional data warehouse platform. Also, the big data landscapes and information science initiatives are applied, the wherever business necessities analysis is commonly ignored.

Vikas S Shah *et al.* [16] have proposed an approach in consideration of multi-faced compliance-aware data services (DS) that enables the degree of distinction in business rules. They presented an approach to continuously monitor regulatory updates and rationalization to translate them into CS. The research also presents runtime structure to evolve and govern compliance-aware DS. The categorization and corresponding implementation of CS into the DS are identified and implied in association with business rules. Formulae to evaluate the degree of distinction and assessment criteria to monitor governance of deployed compliance-aware Data Services are illustrated with an example implementation and validated in the number of actual deployment iterations.

Osden Jokonya *et al.* [17] developed and validated an IT adoption structure to comfort organizations with IT adoption administration. The framework useful for improving IT adoption administration in organizations as it addresses different concerns during IT adoption such as stakeholder buy-in, discrimination, oppression, and stakeholder participation. The structure promoter for the need to understand the problem context before selecting suitable approaches for intervention. On that note, the properties of a