ISSN:0975-9646

Hybrid Classification Models Integration of Data for Business Integration

Pornima Rathi

M.-Tech Researcher, Department of Computer Science,
Astral Institute Of Technology And Research, Indore
Rajiv Gandhi Praudyogiki Vishwavidyalaya, M.P. India

***Abstract*— In today's era, we observe major changes in how managers use computerized support in creation decisions. As further quantity of decision-makers become computer literate, decision support systems (DSS) is evolving from its beginning as a personal support tool and is becoming the shared resource in an association. Data mining has been an vigorous area of study in most recent two decades. Integration of data mining and decision support systems (DSS) can lead to the improved performance and can enable the tackling of new type of problems. In the current past, there has been an growing curiosity in applying evolutionary methods to Knowledge Discovery in Databases (KDD) and a number of successful applications of Genetic Algorithms (GA) and Genetic Programming to KDD have been established. No single algorithm has been found to be superior over all others for all data sets. This research we propose on the selection of evolutionary approach base hybrid classification models in diversity of datasets from different domains. To use grouping of genetic algorithm and other technique as neural network, identity developing GA and fuzzy learning have been test on datasets base on selected quality measures like predictive accuracy and training time. To provide insights into business applications of social network analysis and mining methods. Our propose research directions in social network analysis and mining from the perspective of business applications.**

***Keywords*— Decision Support Systems, Genetic Algorithms, Fuzzy Learning, Data Integration.**

I. INTRODUCTION

Business Intelligence (BI) includes a wide range of applications, practices, and technologies for the extraction, translation, integration, analysis and presentation of data to support improved decision making.

It is important to note that BI is not currently technology. It is people, practice and technologies approaching together so that organizations can make better decisions on a day-to-day basis. The Value of Business Intelligence Data integration and integrity - assure that data is reliable, precise and timely crosswise the endeavour simple access to decision making Quicker recognition of critical information More informed, fact-based decisions The ability to predict events in the future based on historical analysis Information plays a vital role in business organizations. Today's business is information ravenous. Information can be used by the top level management for decision making to make future policies. Due to increasing size of organizations data rapidly, manual interpretation of data for information discovery is not feasible. The impact of computer technology on organizations and society is increasing as new technologies evolve and current technologies expand. More and more aspects of organizational activities are characterized by interaction and cooperation between people and machines. All executives know the fact that information technology is vital to their business. Computerized systems are now penetrating complex managerial areas ranging from the design and management of automated factories to the application of artificial intelligence methods to the evaluation of proposed mergers and acquisitions. Such computerized systems have their roots in various disciplines such as statistics, economics, and operations research. Various methods have been developed for making rational choice. extra lately, these technique, often enhanced by a variety of techniques originating from, cognitive psychology, information science and artificial intelligence, have been apply in the form of computer program, also as separate tools or as integrated computing environments for complex decision creation. Such settings are frequently given the common name of decision support scheme. According to Gorry and Morton (1971), "A DSS is an interactive computer based system that helps decision makers utilizes data and models to solve unstructured problems [1]. Data mining is stare as the key constituent of a a lot extra complex process called Knowledge Discovery in Databases (KDD) which is defined as the non – trivial process of identifying valid, novel, and ultimately understandable patterns in large databases [2]. Over the last three decades, Data mining has been providing several exciting technologies based on supervised and unsupervised machine learning techniques that assist decision makers. These include genetic algorithms and their hybrids with fuzzy logic and artificial neural networks.

The bridging of data mining and decision support has a significant impact on the developments of together field, mostly by improving approach for problem solving in real settings, enabling the fusion of knowledge from experts and knowledge extracted from data, and consequently enabling the successful solution of new types of problems [5] have done excellent work for the integration of these two research areas. A lot of data mining algorithms have been proposed by the machine learning and statistics communities. No single algorithm has been found to be superior over all others for all data sets. A study, called the compares the predictive accuracy of several decision tree algorithms against some non decision tree algorithms on a large number of datasets . To propose hybrid classification models integration of data for business intelligence We apply self evolve GA assist top level management in decision making in light of predictive accuracy and training

time as selected evaluation parameter. To propose on the selection of evolutionary approach base hybrid classification models in diversity of datasets from different domains. To use grouping of genetic algorithm and other technique as neural network, nature developing GA and fuzzy learning have been test on datasets base on selected quality measures like predictive accuracy and training time.

II. Related Work

Ping-Tsai Chung in at al[1] two case studies on data integration and data mining were obtainable. The first case is for the traditional data analytics using relational database techniques such as Oracle database and Cognos BI tool for integrating and mining a company's web site. The subsequent case is for multimedia data analytics by Monago database and Pentaho BI tool for integrating and mining multimedia data presented for the travel related analytics of Food & Wine web site. They have compared both cases in aspects of Data incorporation, Metadata, Query concert and Data Analytics. In these studies reveal that NoSQL database management systems are very useful when working with a huge quantity of data when the data's nature does not require a relational model. The data can be prepared, but NoSQL is use while what really matters is the ability to store and retrieve great quantity of data, not the relations among the elements.

Alfredo Cuzzocrea in at al[2]they have significantly extended contribution , where the ClustCube framework has been introduced, by means of a novel distance function over complex database objects extracted from distributed settings that takes into account the typical tree-like nature of such objects. A comprehensive experimental campaign of ClustCube algorithms for computing ClustCube cubes has represented another relevant contribution of our research.

Ladislav Hluchy in at al[3] In this paper they have presented an overview of one of a set of use cases for environmental management, which form a pilot application of the data integration and data mining-targeting project ADMIRE. This use case is the most advanced of the set in terms of its completion – its deployment is occurring right now .The other use cases target different sub-domains of environmental management, such as short-term rainfall prediction by analyzing radar imagery, the analysis of long-term effects of deforestation along a river basin on the thread of floods on the river, or prediction of O3 concentrations (which is actually a more traditional data-mining problem.

Lars Linsen in at al[4] they have presented an approach to visualize multi-field smoothed particle hydrodynamics data. It consists of feature space and corresponding object space operations. In feature space, they have compute a density function, which they use for automatic detection of hierarchies of high compactness clusters. These clusters are predictable into a 3D star coordinate space. The projection is defined such that it optimizes cluster distribution in stipulations of extend beyond, compactness, and shape. A nested level set visualization for the high density area with respect to different density levels allows for an interactive exploration of the hierarchical clusters and to correlate the clusters to the original dimensions.

Stefan Anderlik and Reinhard Stumptner in at al [5] proposed describes an approach for integration of Decision Support Systems in the area of Structural Health Monitoring. The used and discussed integration ontology consists of descriptions from process steps and DSSs, which are classified as candidates for certain tasks. The mediator component, which implements all communication and interaction processes that are needed for interacting with the systems and with the end-user, searches the inferred ontology for appropriate systems for certain input parameters. These systems are suggested to the end-user who decides which system will be executed next. After computation, the results are evaluated by the end-user and concerning this decision the mediator either increases the acceptance count for the system within the actual process step, takes the output and searches for systems which work with these parameters or the previous process step will be repeated and the system's refuse count will be increased.

Kalinka Mihaylova Kaloyanova[6] The paper presents the idea of using data mining methods for supporting the most important and costly tasks of data warehousing – data integration. Tools that use data mining techniques for this process still are rare. Building of such kind of tools is an important direction for more efficient implementation of data warehousing.

III. Proposed Methodology

Data integration and mining is a complex iterative process. The process includes gaining permission to admission data, take out chosen subsets of data, cleaning and transforming data, performing analyses and delivering results to destination in the form required by users. However, as the data available for an application are continuously growing in volume as well as numbers of sources and formats, the data analysis process become more and more complicated. It is very important to have a uniform access to all data in a consistent way, integrate them together and create new knowledge from the data. Goal understanding: understanding scenarios and transforming them into the definition of a data mining task (data identification, time series, and mining methods).

Data understanding: examining the data and their key attributes (quality, frequency, and statistics).

Data preparation: this is the most complex step including data extraction (e.g. reading interesting items from), cleaning (measured data usually have errors), transformation, interpolation (time and space synchronization), integration.

Modelling : the core mining process, including training and validation Our proposed system is a classifier scheme in which every classifier preserve a prediction of predictable induce, except the classifier's fitness is recognized by

A calculate of the calculation Accuracy. The scheme executes the genetic algorithm in position distinct by the competition sets, as an alternate. Accuracy based Fitness, in grouping with a niche GA, consequences in population treatment to form a whole and precise map. From inputs and events to bribe predictions.

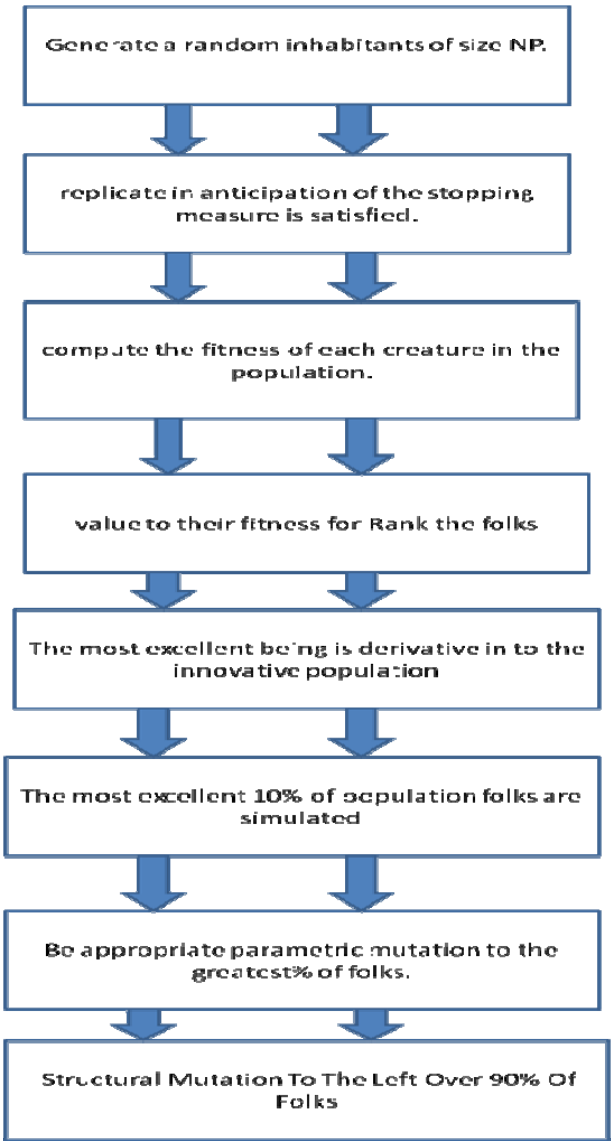

Fig 1: Proposed algorithm flow chart

Traditional Classifier system has not supposedly emphasizes or in information created such mappings, which can create payoff-maximizing Action-selection simple. Additional, tend to develop Classifiers that are maximally universal subject to a correctness measure, so that the mapping gains representative efficiency. Additional feature of the algorithm is obtainable. Our proposed is base on the genotype-phenotype preparation of Fuzzy rule in genetic algorithm have realize a replicated annealing based technique for induct together parameters and structure of a fuzzy Classifiers. The adjacency performance in simulated annealing has been alternate with a comprehensive change taken from tree shaped genotype additional feature of the algorithm can be available. Interested reader may also access the entire KEEL inbuilt algorithmsat [12].

Evaluation: the results from modelling process are evaluated if they meet the criteria of scenarios.

Processing elements are the basic execution units. Each processing element will perform an elementary operation over data streams. Our propose platform include basic operations as follows: Manipulation with data sources: performing SQL queries, reading data from files (from local disks, HTTP, FTP servers, cloud storages) Manipulation with data streams: splitting, merging, filtering Data transformation: conversion, transformations data pre-processing, association rules classification, clustering, regression, Data delivery: to clients, repository, FTP servers or cloud storages .Developers can create new processing elements and deploy them to execution servers. Each processing elements can have several inputs of three main types: input streams, literal parameters (e.g. strings, numbers), and data sources (e.g. databases, file systems (local or remote), cloud storages). Data are processed in the streaming manner; the processing elements read a portion of data from input streams, process them and generate data to output streams before reading next portion of input.

The presentation layer Web services platform on data mining, users can search the Web services as their needs. Identify the data we wish to use to build our model with a URL that points to that data. Specify the type of model we want to build, and parameters to the construct process. Such parameters are term construct settings. The most important build setting is the definition of the data-mining job, such as clustering, categorization, or correlation rules. Select certain attributes of the physical data and then map those attributes to logical values. We can specify such mappings in our build settings. Specify the parameters to the data-mining algorithms. Make a construct task and be relevant to that task the physical data references and the build settings. Finally execute the task. The outcome of that execution is the data model. That model describes the possible input attributes for later applying the model to additional data. Formulation of a formal framework for multi-agent systems that allows extensibility, reusability, integrity of system components varied upon particular task. Providing a platform for data mining researches which enhances research processes and providing a platform for agent researches to extend functionalities of the system incorporated with data mining capability. Our present preliminary results of our ongoing work on the data integration engine for environmental data that is being developed in the scope. We first describe scenarios dealing with the integration and mining of environmental data. The main challenge that the environmental data required by scenarios are maintained and provided by different organizations and are often in different formats. Our work concentrate on providing a platform that would allow integration of data from heterogeneous resources.

- Improve accuracy for Identifying customer buying behaviors.
- Identifying effective customer purchasing patterns and trends.
- Improving the Quality of Service (QoS)
- Attain improved consumer preservation and satisfaction.
- humanizing possessions expenditure relation.
- Design and developing effective goods transportation and distribution policies. For the Telecommunication and many other industries: we could share many similar goals and expectations of retail data mining.

CONCLUSION

We have present on the selection of evolutionary approach base hybrid classification models in diversity of datasets from different domains and data integration and mining. The platform has fully distributed data-streaming architecture with high level abstraction of data sources and processing elements. The abstraction will separate the presentation layer of data analysis process from enactment layer, allow experts to focus on their interests make the process of data integration and mining easier.

REFERENCES

[1] Ping-Tsai Chung, Sarah H. Chung," On Data Integration and Data Mining for Developing Business Intelligence" Systems, Applications and Technology Conference (LISAT), IEEE -2013.

[2] Alfredo Cuzzocrea, Paolo Serafino," Enhanced Clustering of Complex Database Objects in the ClustCube Framework" DOLAP'12, November 2, 2012, Maui, Hawaii, USA.

[3] Ladislav Hluchy, Ondrej Habala, Viet Tran, Marek Ciglan," Hydro-meteorological Scenarios Using Advanced Data Mining and Integration" Sixth International Conference on Fuzzy Systems and Knowledge Discovery- 2009.

[4] Lars Linsen, Tran Van Long, Paul Rosenthal, and Stephan Rosswog," Surface Extraction from Multi-field Particle Volume Data Using Multi-dimensional Cluster Visualization" IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 14, NO. 6, NOVEMBER/DECEMBER 2008.

[5] Stefan Anderlik and Reinhard Stumptner, Bernhard Freudenthaler , Martin Fritz," A Proposal for Ontology-based Integration of Heterogeneous Decision Support Systems for Structural Health Monitoring" iiWAS2010, 8-10 November, 2010, Paris, France.

[6] Kalinka Mihaylova Kaloyanova," Improving Data Integration For Data Warehouse: A Data Mining Approach"- 2005.

[7] Daniel Gruhl · Meena Nagarajan · Jan Pieper ," Multimodal social intelligence in a real-time dashboard system" The VLDB Journal (2010) 19:825–848 DOI 10.1007/s00778-010-0207-5.

[8] Henrike Berthold , Philipp Rösch, Stefan Zöller , Felix Wortmann, Alessio Carenini, Stuart Campbell, Pascal Bisson, Frank Strohmaier," An Architecture for Ad-hoc and Collaborative Business Intelligence" EDBT 2010, March 22–26, 2010, Lausanne, Switzerland.

[9] Kaashief Hartley, Lisa F Seymour," Towards a framework for the adoption of Business Intelligence in public sector organisations: the case of South Africa" SAICSIT '11, October 3–5, 2010, Cape Town, South Africa.

[10] Leslie Dolman, Waterloo , John Mylopoulos," Next Generation Business Intelligence (BI) Tools"

[11] Malu Castellanos, Song Wang, Umeshwar Dayal, Chetan Gupta," SIE-OBI: A Streaming Information Extraction Platform for Operational Business Intelligence" SIGMOD'10, June 6–10, 2010, Indianapolis, Indiana, USA.

[12] http://www.keel.es/.