JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A FRAMEWORK FOR KNOWLEDGE MINING IN MULTI-DISCIPLINARY FIELDS

[1]Karnam Akhil, [2]Dr. A. Brahmananda Reddy, [3]Abhinav Borad

[1]Student, [2]Associate Professor, [3]Student
[1]Computer Science and Engineering Department
[1]VNR VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY, HYDERABAD, INDIA

Abstract : In artificial intelligence (AI), knowledge mining employs modular services to learn from vast amounts of data quickly. Information can be understood and explored easier, hidden insights uncovered, and relationships and patterns found on a large scale. Artificial Intelligence (AI) has made it much easier to organize unstructured data and discover dark data. Knowledge mining can be automated by AI bots using advanced Machine Learning (ML) techniques. All your documents are crawled and understood, metadata is generated, and they are automatically classified and tagged. In addition to organizing your documents accurately, AI bots are more efficient and less expensive to run. Speed. Thus, enterprise search makes these documents easy to find. Knowledge mining is widely used in Deep-Learning, Machine- Learning, Artificial Intelligence, Data-Mining, Natural Language Processing, speech recognition, and many others. In this paper, different functionalities of knowledge mining are illustrated, and a framework is designed for the workflow of knowledge discovery in multi-disciplinary fields.

IndexTerms - Knowledge Mining, Knowledge evaluation, Outlier analysis, Time-series analysis, Prediction

I. INTRODUCTION

An emerging discipline in artificial intelligence (AI) called knowledge mining involves learning from huge amounts of data by combining intelligent services. By discovering hidden insights, finding relationships and patterns at scale, and exploring information in depth, organizations are able to fully understand and easily explore information. Knowledge mining, the next wave of artificial intelligence, generates dynamic understandings of relationships and patterns in a corpus of data by utilizing the data sources of a certain type, whereas AI's first wave focused on training models on a single data source for a single application. Having access to real-time information has become a critical part of enterprise digital transformation initiatives. Extraction of knowledge is the process of creating knowledge from both structured (relational databases, XML) and unstructured (text, documents, images) sources.

A language for retrieving resource description frameworks (RDF) from relational databases is currently being developed by the RDB2RDF W3C group. Wikipedia can also be transformed into structured data and mapped to existing knowledge as an example of knowledge extraction. Making ontologies pervasive is undoubtedly a great challenge. The technique associated with the automated construction of ontologies is called ontology learning, which involves the use of automated methods in ontological engineering. Information retrieval (IR) is a subdomain of this field. The traditional approach to building ontologies can be divided into two categories. The ontology so constructed generally has high quality, but is restricted in size and scope due to the handcrafted human resources. Using automated methods, however, it is possible to produce extremely large and comprehensive ontologies, but at the expense of quality. Inconsistencies and ambiguities are inherent in human language. In specific contexts and for specific tasks, statistical approaches (data mining) may be capable of emulating human intelligence. Whenever there is a need to generalize or move into a new domain and task, it may not be applicable. Using text mining techniques on semi-structured knowledge sources appears to be another approach that has been quite successful in recent years. Wikipedia is a good candidate for this knowledge source.

In recent years, research has focused on mining Wikipedia articles to uncover world knowledge on a wide range of topics. Wikipedia pages refer to concepts, and concepts are semantically related to one another. Ontologies can be described as such loosely structured knowledge bases. The URLs of Wikipedia entries serve as identifiers for articles, which can be viewed as ontology elements. Links within articles, category links, redirections, category names, and infoboxes are some of the ways Wikipedia's semistructured knowledge base can be mined. This knowledge base is constantly updated and checked by a huge community of users. This attribute is extremely important if it is to serve as a knowledge source for the construction of an ontology. The Wiki technology has been used to develop collaborative ontologies (Bloom 1956). Structured data is generated by information extraction (IE) from unstructured or semistructured text. Since the 2000s, numerous methods have been proposed to extract web information. It identifies six types of relationships between concept pairs: hyperonymy, hyponymy, holonymy, meronymy, antonymy, and synonymy. The pattern extraction process requires extracting patterns between each pair of concepts. During pattern generalization, similarities are

discovered between patterns and dissimilarities are removed. By considering the holonymy and meronymy that exist between concepts, New relations are identified that are not present in the WordNet.
[1]

II RELATED WORKS

knowledge mining in multi- disciplinary field

The performance of students has been analyzed using several statistical tools over time. A data mining approach is presented in this paper to identify students' failure patterns in an educational environment. To enhance the decision-making process of academic planners in higher institutions of learning, the identified patterns are analyzed, and constructive recommendations are made. As a result, the curriculum structure will also be improved and modified to reduce failure rates and improve students' academic performance. An analytical process was described, and software was developed for mining failed courses. [1]

It is possible to extract knowledge from clinical datasets due to the availability of clinical datasets and knowledge mining methodologies. Mathematical models have been developed using different data mining techniques to assist clinicians in making decisions. Based on a minimal set of attributes extracted from the clinical dataset, this research aims to build a classifier that predicts the presence or absence of disease. This study uses a backpropagation neural network (RS-BPNN) to study rough set indiscernibility. Two stages are involved in this project. Using the indiscernibility relation method, the first stage is handling missing values to obtain a smooth data set. Second, the selected reductions of the dataset are classified using backpropagation neural networks. Using datasets from the University of California at Irvine (UCI) machine learning repository, the classifier has been tested on hepatitis, Wisconsin breast cancer, and Starlog heart disease datasets. As a result of the proposed method, 97.3% of hepatitis cases, 98.6% of breast cancer cases, and 90.4% of heart disease cases are correctly diagnosed. A clinical dataset classification model is proposed in the proposed system. [2]

Many medical diagnoses and decision support systems use rule-based classification as a data mining task. In rule bases, rules are stored that affect the efficiency of classification. A meta-heuristic approach called Wind-driven Swarm Optimization (WSO) is used in this work to optimize rule sets extracted with data mining tools. Heuristic or meta-heuristic approaches are used to improve the quality of rule sets. Biological inspiration underlies this algorithm, making it unique. [3]

Mining engineering knowledge from electronic documents is essential for new business opportunities and devising new design solutions for emerging applications. In light of constantly proposed engineering innovations, it cannot be easy to continually and comprehensively review and assess all technical literature to identify the most relevant ideas and solutions. The development of high-performance analog circuits has been at the forefront of technological innovations due to their importance in telecommunications, mobile, and medical applications. Analog circuit design metaknowledge can be represented using a model and mining techniques in this paper. There are three components to the metaknowledge representation: a conceptual representation of the circuit hierarchy, a performance capability representation (e.g., tradeoffs, bottlenecks), and a causal model of the starting ideas and design plans. New applications can be analyzed using knowledge structures, existing circuits can be improved, and design errors can be detected and corrected using knowledge structures. In this case study, a set of 30 high-frequency analog circuits is used to illustrate the proposed knowledge structure. [4]

Business data is being stored in databases or log files on web servers due to the rapid development of internet technology. Making better decisions is an urgent task facing vast amounts of data. To support scientific decision-making, data mining becomes one of the most essential directions; however, most information gathered only describes static knowledge, not how-to-do knowledge. The paper describes our method for addressing this problem, using MCLP-based classification, which classifies data sets on the web, then a decision tree is used to find the rules that explain the models. However, it is only static know-what information, and we still do not know how to transfer the data from one class to another. A new methodology based on decision tree rules and Extension set theory can improve such situations by revealing actionable knowledge. In our second-level mining method, we can detect hidden patterns related to changing customers' behavior, so they will not churn and provide "from can't to can, from bad to good" rules that can be used to change customers' behavior. To reduce customer churn, it provides decision-making support. Our method has been demonstrated to be feasible and effective in a case study of a web company. Data mining in the web business also has a reference value. [5]

III SYSTEM ARCHITECTURE

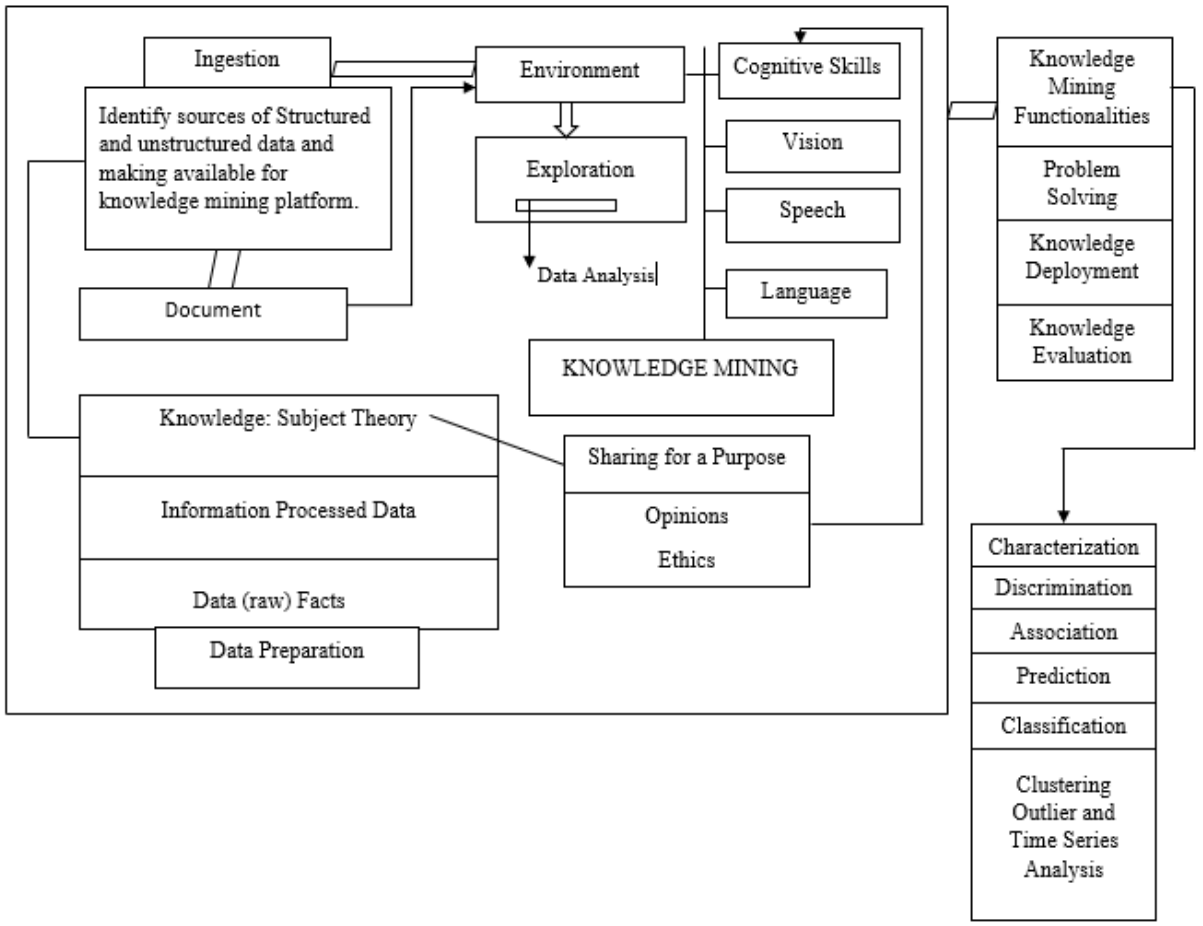

Fig 1: Architecture of Knowledge Mining

Patterns discovered during data mining tasks are represented by data mining functionalities. A data mining task can generally be classified into two types: descriptive and predictive. The descriptive mining tasks identify common characteristics of the database data, while the predictive mining tasks make predictions based on the current information. [9]

There are various data mining functionalities which are as follows −

Classification

Classification involves categorizing elements in a collection according to their predefined properties and functionalities. New instances can be classified using a classification model that is unknown to the user. A model is created using training data, which are specific instances. This type of mechanism uses methods such as if-then statements, decision trees, neural networks, or even rules. Identifying future data can be done by retrieving these methods.

Association Analysis

The term Market Basket Analysis can also be used to describe Association Analysis. Sales use this data mining methodology very often; analysis helps to find relations between elements frequently occurring together. There are a variety of components that make up the case model, as well as rules explaining how the elements are grouped within the cases. A database association rule is used to predict the presence of an element in the database and is based on the manifestation of a specific element that has been identified as being important. Association analysis is based on 2 parts rule –

antecedent (if)

consequent(then) –

An antecedent (if) points towards a degree of discovering a consequent (then) in the data set. It suggests that they are associated.

Cluster Analysis

Classification and cluster analysis are similar processes. The only difference between cluster analysis and classification analysis is that the class label is unknown. As a result of clustering algorithms, the data are divided into groups based on similarities, with each group of data being more similar to the others than the others. Among other things, cluster analysis is used in machine learning, deep learning, image processing, pattern recognition, and natural language processing.

Data Characterization

As part of the data characterization process, generic features of data are summarized, which can be used to define a target class according to specific rules. Graphs, charts, and tables can be used to visualize characterized data without much user interaction or intervention using an attribute-oriented induction technique.

Data Discrimination

Data discrimination occurs when a data set or source is treated differently than others, either intentionally or unintentionally. Based on ambiguity in attribute values, this data mining functionality helps separate peculiar data sets.

Prediction

Data mining is a popular method for determining any missing or unknown elements in a data set using prediction. Businesses use linear regression models to forecast the results of any given event using numeric predictions, based on previous data. There are two types of predictions –

Numeric Predictions – Predict any missing or unknown element in a data set

Class Predictions – Predict the class label using a previously built class model

Outlier Analysis

The outlier analysis technique is used if there is no way to group the data into any class. By analyzing outliers. In most cases, an outlier represents an abnormality in the data. Data sets with more outliers have a lower quality. A data set with a large number of outliers cannot be used to detect patterns or draw any conclusions. As a result of the outlier analysis process, check whether there is any data that can be used for analysis after some clean-up has been done. Nevertheless, keeping track of unusual data and activities is still important to detect anomalies and business impacts as soon as possible.

Evolution Analysis

Analyzing data sets that have undergone a transformation or change is known as evolution analysis. Multivariate time series data can be characterized, classified, or discriminated and clustered using evolution analysis models. [9]

DISCUSSIONS AND CONCLUSION

One of the most interesting things about data mining is that you can get information without asking specific questions. In addition to using statistics and algorithms, the entire process is largely predictive, predicting future trends or what will happen based on the stored data. In addition to predicting future events, data mining also identifies hidden information. Data mining functionalities contribute to finding trends in data mining, making it an important tool for data scientists. Cognitive skills are important to enhance the knowledge visualization and in extracting the features of the unstructured data. Framework of knowledge mining in multi-disciplinary field is presented after a deep study. In the future, Proposed framework can be extended to apply Web semantics and building Intelligent systems.

REFERENCES

[1] Chandra, E., and K. Nandhini. "Knowledge mining from student data." *European journal of scientific research* 47.1 (2010): 156-163.

[2] Nahato, Kindie Biredagn, Khanna Nehemiah Harichandran, and Kannan Arputharaj. "Knowledge mining from clinical datasets using rough sets and backpropagation neural network." *Computational and mathematical methods in medicine* 2015 (2015).

[3] Christopher, J. Jabez, H. Khanna Nehemiah, and Arputharaj Kannan. "A swarm optimization approach for clinical knowledge mining." *Computer methods and programs in biomedicine* 121.3 (2015): 137-148.

[4] Jiao, Fanshu, et al. "Analog circuit design knowledge mining: Discovering topological similarities and uncovering design reasoning strategies." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 34.7 (2015): 1045-1058.

[5] Li, Xingsen, et al. "An intelligent transformation knowledge mining method based on Extenics." *Journal of Internet Technology* 14.2 (2013): 315-325.

[6] https://www.tutorialspoint.com/what-are-the-functionalities-of-data-mining