Text Mining Assignment

Akash G CB.EN.U4CSE17105

Text Mining

Text mining, also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text." The definition strikes at the primary chord of text mining – to delve into unstructured data to extract meaningful patterns and insights required for exploring textual data sources. Text mining incorporates and integrates the tools of information retrieval, data mining, machine learning, statistics, and computational linguistics, and hence, it is nothing short of a multidisciplinary field. Text mining deals with natural language texts either stored in semi-structured or unstructured formats.

Steps for text mining

- Gathering unstructured data from multiple data sources like plain text, web pages, pdf files, emails, and blogs, to name a few.
- Detect and remove anomalies from data by conducting pre-processing and cleansing operations. Data cleansing allows you to extract and retain the valuable information hidden within the data and to help identify the roots of specific words.
- For this, you get a number of text mining tools and text mining applications.
- Convert all the relevant information extracted from unstructured data into structured formats.
- Analyze the patterns within the data via the Management Information System (MIS).
- Store all the valuable information into a secure database to drive trend analysis and enhance the decision-making process of the organization.

Applications of text mining

There are various applications of Text mining like automatic processing of messages and emails. For example, it is possible to "filter" out automatically "junk email" based on certain terms, such messages can automatically be discarded. Such automatic systems for classifying electronic messages can also be useful in applications where messages need to be routed automatically to the most appropriate department. Another application is Analyzing warranty or insurance claims, diagnostic interviews. In some business domains, the majority of information is collected in textual form.

Text mining applications

- 1 – Risk management
- 2 – **Knowledge management**
- 3 – Cybercrime prevention
- 4 – Customer care service
- 5 – Fraud detection through claims investigation
- 6 – Contextual Advertising
- 7 – Business intelligence
- 8 – Content enrichment.

Knowledge Management

- Knowledge management (KM) is an effort to increase useful knowledge within the organization. Ways to do this include encouraging communication, offering opportunities to learn, and promoting the sharing of appropriate knowledge artifacts
- Knowledge management process focuses on knowledge flows and the process of creation, sharing, and distributing knowledge. Each of the knowledge units of capture and creation, sharing and dissemination, and acquisition and application can be facilitated by information technology.
- Knowledge management in organizations is about ensuring that the right information is delivered to the right person at the right time. How can the right information be easily identified? This work demonstrates how text mining provides a tool for generating human understandable textual summaries that ease the task of finding the relevant information within organizational documents repositories.

Applications of Data Mining

- Health Care Organization
- Retailing
- Financial/Banking
- Small and Middle Businesses
- Entrepreneurial Science etc.

Data Mining Techniques used

- Within the context of articles reviewed, applications of data mining have been widely used in various enterprises ranging from public health-care, construction industry, food company, retailing to finance. Each field can be supported by different data mining techniques which generally include classification, clustering, and dependency modeling.
- Classification is one of the most common learning methods in data mining. This task aims at mapping a data item into one of several predefined classes.
- Examples of classification methods used as part of knowledge management include the classifying of the patients from primary health-care centers to specialists; the combination of the data mining and decision support approaches in planning of the regional health-care system; and the implementation of visualization methods to facilitate KM and decision-making processes.
- For small and middle businesses: food company domain, data mining can improve decision-making by knowledge cultivating method namely Extenics and Extension data mining (EDM).
- This method was the integration of data mining and knowledge management, to develop a decision support system platform for better decisions.
- In the business organizations with a large volume of works, such companies wanted to better understand what the hidden patterns between the KM and its performance, using the combination of data mining techniques: Bayesian Network (BN) classifier and Rough Set Theory (RST) in their business could help companies producing the KM to be performed effectively and achieve higher efficacy resulted.
- Common tools used for classification are decision trees, neural networks.
- Bayesian network and rough set theory.
- Clustering: This involved seeking to identify a finite set of categories and grouping together objects that are similar to each other and dissimilar to the objects belonging to other clusters.

Applications

- Healthcare: clustering categories and attributes used in analyzing the similarities between community health centers.
- Retailing: clustering the segmentation for possible product line and brand extension to identify market to customer clusters.
- Financial/Banking: identifying groups of corporate bond clusters according to the industry and a specific segment within an industry; then tuning cluster data for each industry as a template for predicting rating changes.
- Construction Industry: clustering textual data to discover groups of similar access patterns.
- Collaboration and Teamwork: identifying groups of workers with similar task-related information needs based on the similarities of workers' knowledge flow.

Conclusion

In organization, knowledge is an important resource. Management of knowledge resources has become a strong demand for development. Discovering useful knowledge has also a significant approach for management and decision making.