

USING DATA MINING APPROACHES TO IDENTIFY HIDDEN PATTERNS IN THE DATA SET

Turetskyi A. O., Zubenko A.O, Nazirova T.O.

***National Technical University "Kharkiv Polytechnic Institute",
Kharkiv***

Data mining is also known under many other names, including knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing. Data possibly be in different or extraordinary formats as it comes from different sources, irrelevant attributes and missing data. Consequently, data needs to be prepared before applying any kind of data mining.

Data size are generally growing from day to day and across all the fields, data are being collected and accumulated at a vivid pace. Analysis and evaluation of stored data may lead to the discovery of patterns hidden and trends within the data that could significantly strengthen understanding. The exemplary data mining process includes transmitting data originally accumulated in production systems into a data warehouse, cleaning or scrubbing the data to remove errors and check for succession of formats, and for further searching the data using neural networks or statistical queries or other methods machine learning.

A spatial database is a database that is enhanced to store and access spatial data or data that defines a geometric space. These data are often associated with geographic locations and features, or constructed features like cities. Data on spatial databases are stored as coordinates, points, lines, polygons and topology. Some spatial databases handle more complex data like three-dimensional objects, topological coverage and linear networks. Such spatial databases present new invocation to data mining algorithms.

Within data mining, there is a group of tools that have been developed by a research community and data analysis enthusiasts; they are offered free of charge using one of the existing open-source licenses. Open-source software development is the process by which open-source software, or similar software whose source code is publicly available, is developed by an open-source software project. These are software products available with its source code under an open-source license to study, change, and improve its design. The development and application of data mining principle requires use of very powerful software tools. As the number of available tools continues to grow, the choice of most suitable tool becomes increasingly difficult.

This survey paper summarizes various approaches and tools for medical data for E-health. Algorithms applied in this area would be valuable for researchers in medical assessment and medical Information Technology specialists to develop a decision support system aggregate classification and clustering techniques in information systems. It also highlighted the importance of locality techniques used for the processing medical information as specific data.

The studies presented are intended to highlight the motive force perceived as contributing to the success or otherwise of the data mining approach and as the same way an approach that can be used to expedite the initial stages of information analysis in order that the results obtained may be more thoroughly investigated.