

Fuzzy Set Theory is also called **Parailly** Theory. This theory was proposed by Lotfi Zadebin 1965 as an alternative the two-value logic and probability theory. The theory allows us to work at a high level of abstraction. It also provides us the means for dealing with imprecise measurement of data.

The fuzzy set theory also allows us to deal with vague or inexact facts. For example, being a member of a set of high incomes is in exact (e.g. if \$50,000 is high then what about \$49,000 and \$48,000). Unlike the traditional CRISP set where the element either belong to S or its complement but in fuzzy set theory the element can belong to more than one fuzzy set.

For example, the income value \$49,000 belongs to both the medium and high fuzzy sets but to differing degrees. Fuzzy set notation for this income value is as follows –

 $m_{medium_{income}}$ (\$49k)=0.15 and $m_{high_{income}}$ (\$49k)=0.96

• The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

Applications of Cluster Analysis

- Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.
- Clustering can also help marketers discover distinct groups in their customer base. And they can characterize their customer groups based on the purchasing patterns.
- In the field of biology, it can be used to derive brant and animal taxonomies, categorize series with similar functionalities and gain oright into structures inherent to population

• Clustering also helps in identification of areas of similar

- and use in around th observation database. It also helps in the identification of groups of houses in a city according to house type, value, and geographic location.
 - Clustering also helps in classifying documents on the web for information discovery.
 - Clustering is also used in outlier detection applications such as detection of credit card fraud.
 - As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

Mining Web page layout structure

The basic structure of the web page is based on the Document Object Model (DOM). The DOM structure refers to a tree like structure where the HTML tag in the page corresponds to a node in the DOM tree. We can segment the web page by using predefined tags in HTML. The HTML syntax is flexible therefore, the web pages does not follow the W3C specifications. Not following the specifications of W3C may cause error in DOM tree structure.

The DOM structure was initially introduced for presentation in the browser and not for description of semantic structure of the web page. The DOM structure cannot correct Didentify the semantic relationship between the different parts of a web page.

Vision-based page segmentation (VIPS) 28

- The phylose of VIPS is the extract the semantic structure of a web pagebased on its visual presentation.
 - Such a semantic structure corresponds to a tree structure. In this tree each node corresponds to a block.
 - A value is assigned to each node. This value is called the Degree of Coherence. This value is assigned to indicate the coherent content in the block based on visual perception.
 - The VIPS algorithm first extracts all the suitable blocks from the HTML DOM tree. After that it finds the separators between these blocks.
 - The separators refer to the horizontal or vertical lines in a web page that visually cross with no blocks.

mining systems do not share underlying data mining query language.

Trends in Data Mining

Data mining concepts are still evolving and here are the latest trends that we get to see in this field –

- Application Exploration.
- Scalable and interactive data mining methods.
- Integration of data mining with database systems, data warehouse systems and web database systems.
- SStandardization of data mining query language.
- Visual data mining.
- New methods for mining complex types of data.
- Biological data mining ______
- Data mining and coftware enginesing.
- Web mining.
- DreNstribute dato Rining.
 - Real time data mining.
 - Multi database data mining.
 - Privacy protection and information security in data mining.