Data Analytics Success Stories : Data Mining

Hanshin University. Applied Statistics. Grade 4. 201452024 Park Sang Hee

Contents

1. What is data mining?

The dictionary meaning of data mining is “Finding statistical rules or patterns systematically and automatically in large amounts of stored data.” In other words, it is called "kDD(knowledge-discovery in databases)". Recently, as the amount of data we deal with has grown exponentially, we have begun to focus on using data efficiently. The data also became an important indicator in decision making. Data mining is not limited to one field, but is very useful in various fields. The most successful use of data mining is in the marketing sector. In addition, banking credit evaluation, telecommunications, and insurance sectors are covered. Application methods include classification, clustering, association analysis, Sequencing analysis, and prediction.

Classification infers classification and division through specific definitions of certain groups. An example of classification is the prediction of abandoners through user analysis. Clustering looks for clusters that share specific characteristics. Clustering differs from classification in that it does not have information about predefined characteristics. An example is a cluster model for identifying consumption patterns by customer type. Associations analysis define the relationships between concurrent events. For example, a customer who buys sweets at a grocery store knows what they usually buy together and adjusts the store's display position. Continuity identifies relationships that occur over a period of time. Similar to the correlation analysis except for the characteristics of the period. An example is an analysis that predicts recurring visits to supermarket and financial product use. Predictions predict the future based on patterns within large datasets. Prediction is the most widely used data mining method.

2. Success stories using data mining.

The automatic product category classification model of "Naver Shopping" is a representative success case using the classification technique. "Naver Shopping" is a shopping portal service provided by Naver. "Naver Shopping" organizes products and matches them to categories for easier searching by users. However, it is impossible for humans to match more than 20 million newly registered products to about 5,000 categories every day. When a user inputs a keyword to search for a product in "Naver Shopping", the category of the input keyword is first identified, and then the products matching the keyword category are listed according to the search sorting logic. In order for the user to quickly find the product they are looking for and get the search results they want, they must be matched to the correct category.

"NCSOFT", one of the famous game companies in Korea, uses clusters to manage customers. The

biggest feature of gamers is that heavy and medium users make up only 20% of the total, but they make up 70% of the total sales. The majority of light and other users are 80%, but only 30% of sales. Therefore, game companies need to screen heavy and medium users well. This is because it is necessary to cluster customers by using their behaviors and patterns, and to implement different user policies according to the clustering. In particular, the game called "Lineage" by "NCSOFT" is well known for its selection of very heavy users and their custom game policies. Other users are also known for their high level of satisfaction with the game, with events and policies tailored to their game patterns.

The most representative example of association analysis is shopping cart analysis. Wilmart carried out the famous "shopping cart analysis" that revealed the relationship between beer and diapers, an example that has always been a staple in data mining textbooks. WilMart analyzed vast amounts of customer purchase data and found that consumers had a big impact buying beer and diapers together. Based on this information that explains the customer's buying patterns, Wilmart puts beer side by side alongside the diaper. This shopping cart analysis is an example of establishing a sales strategy such as product display by providing a seller with information on the relationship between products in terms of a customer's purchasing behavior.

Compositions that used to be considered unique to humanity have now arrived. Recently, Sony unveiled "Flowmachines," an artificial intelligence composed by Sony. The AI analyzed more than 13,000 songs stored in the database and composed the song the way the user wanted. By analyzing the continuity of the songs in the database, we composed the most comfortable notes by humans, analyzed the patterns of those sounds, and predicted the next note when a specific code or note was given. Although it is still very difficult to hear songs written by AI, it has been greatly appreciated in terms of expanding the scope of AI. Sony plans to release an album composed of songs composed by AI in the future.

Google began to provide flu forecasting services by analyzing search terms. The flu forecast system was established by analyzing the search terms related to 'cold'. The frequency of the query queries related to the flu, such as the flu and influenza, was investigated on the Google homepage. Google was able to set up an early warning system for the spread of flu called Google Flu Trends. It also provided regional flu epidemics far ahead of US health authorities. Comparison with the US Centers for Disease Control and Prevention shows that there is a real correlation. Google has also begun researching the frequency of search queries left by users on search sites and providing information on the distribution and spread of flu patients. In addition, various search terms were analyzed to enable users to give meaningful feedback on culture, economy, and sports.

3. Conclusion and Suggestions.

However, data mining is not 100% correct. When classifying spam, important mail can be classified as spam, or the wrong users can be clustered as heavy users. Also, attempting to find something in the data without any specific goals and plans can be very reckless. If you just believe in the data and make the decision, it can hurt your company. However, it is clear that data mining is still a good weapon as it is being used and developed in various fields. Misuse of a weapon can be fatal. Therefore, to use the great weapon of data mining well, you need enough practice and effort.