Chapter 9. Artificial intelligence and other applications

[presentation] (./pdf/ppt9.pdf) **[book] (./pdf/book9.pdf)**

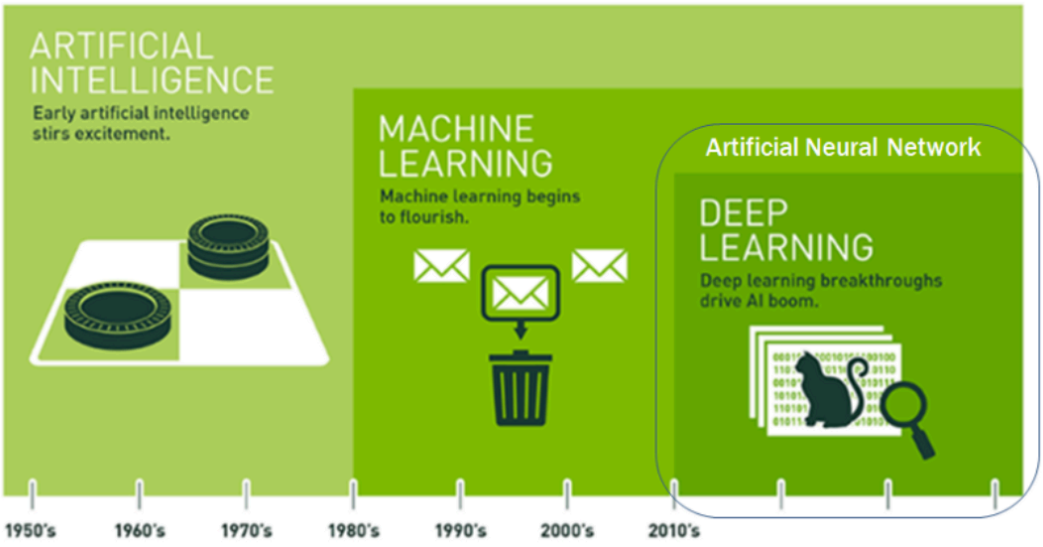
CHAPTER OBJECTIVES

We introduce the following in this chapter.

- Applications of data science into artificial intelligence in section 9.1.
- Text mining in section 9.2.
- Web data mining in section 9.3.
- Multimedia mining in section 9.4.
- Spatial data analysis in section 9.5.

9.1 Artificial intelligence, machine learning, and deep learning

Artificial intelligence (AI) refers to machines that have the intelligence to imitate human intelligence and perform complex tasks like humans. An important method for implementing the artificial intelligence is the **machine learning** models. In this book, we have discussed several models of the machine learning such as the supervised learning in Chapters 6 and 7, the unsupervised learning in Chapter 8, but the **artificial neural network** model is the most important model for the artificial intelligence. One method for finding the solution to the neural network model is a **deep learning**. Therefore, the artificial intelligence, machine learning, artificial neural network, and deep learning can be expressed as a set relationship as in <Figure 9.1.1>.



<Figure 9.1.1> Relationship of the artificial intelligence, machine learning, artificial neural network, and deep learning

A. Artificial Intelligence

In 1943, Warren McCullough and Walker Pitts, who were neuroscientists in the United States, first proposed the idea of an artificial intelligence model by proposing the operating principles of human neurons based on binary code. Nowadays, the definition of artificial intelligence varies from person to person. According to the book widely used as a textbook of the artificial intelligence, 'Artificial Intelligence – A Modern Approach' by Stuart Russel and Peter Norvig, it is defined as a machine that "**thinks humanly, acts humanly, thinks rationally, and acts rationally**."

In the early days of artificial intelligence, efforts were made to create machines similar to humans by imitating the human brain and way of thinking. In 1955, Marvin Minsky and others at Dartmouth College in the United States built the first neural network, the SNARC system. Around the same time, computer scientist Viktor Glushkov in the Soviet Union created the All-Union Automatic Information Processing System (OGAS). With the development of computer science, rather than implementing human intelligence itself, AI machines were made to efficiently solve real-world problems that can be broadly divided into four types of stages.

Level 1: Simple control machine – wash machine, cleaning machine

Level 2: Classical artificial intelligence with diverse patterns – numerous and sophisticated ways to establish input-output relationships, utilizing a knowledge base

Level 3: Artificial intelligence that accepts the machine learning models – uses machine learning algorithms that learn based on data

Level 4: When performing machine learning, the machine directly learns the features of the input values without a person inputting them

At the early stage of AI, a single layer neural network, perceptron, was used. However, due to the limitations of information processing capabilities, the applications of the perceptron were limited, and its popularity waned for a while. In 1974, Paul Warboss proposed a backpropagation algorithm that could solve multilayer neural networks, and research on AI using multilayer neural network models was actively conducted, resulting in visible results such as character recognition and speech recognition. However, until then, it was closer to an answering machine than a conversation with a human. In 2006, Geoffrey Hinton announced the deep confidence neural network that is capable of unsupervised learning methods which were considered impossible, and as a result, the methodology called **deep learning** replaced the higher-level concept of the artificial neural network and became the only methodology. In particular, in 2012, Hinton's disciples Alex Krizhevsky and Ilya Sutskever built AlexNet, a convolutional neural network architecture, and won the computer vision competition called "ILSVRC" with overwhelming performance, and deep learning became an overwhelming trend, surpassing the existing methodology. In 2016, Google DeepMind's AlphaGo popularized deep learning methods and showed results that surpassed human levels in several fields. AI research has continued to conduct innovative research to solve problems that only humans could do, such as natural language processing and complex mathematical problems, using computers. Awareness spread that AI could quickly surpass human capabilities in the field of weak artificial intelligence.

Artificial intelligence began to change significantly in 2022 with the emergence of **generative AI**. OpenAI's ChatGPT and Drawing AI, which are representative of the generative AI, have finally begun to be applied to actual personal hobbies and work applications, and the practical application of AI,

which had seemed like a dream, has finally begun. The background of this generative AI was the transformer structure. However, generative AI has led to active discussions about AI, and among them, theories of caution and threats to AI have also begun to emerge.

Technologies of AI

Most of the current artificial intelligence is made up of artificial neural network models where numerous nodes have their intelligence which performs their own calculations. The artificial intelligence performs a large number of simulations on a neural network with many hidden layers with various weights and learning methods until it produces a successful result. When a question (signal) is given to the artificial intelligence, each node responds to the question and transmits a signal to the next node. So each node that receives the signal filters the signal according to its own given criteria, the bias, and re-produces it. The sum of the filtered signals becomes the 'answer' that artificial intelligence delivers to us.

For example, if we give a picture of a partially obscured puppy and ask the artificial intelligence whether the object in the image is a puppy, the artificial intelligence filters the signal for the picture according to its own criteria, which is the characteristics of each puppy remembered by each node. If it is a picture of a puppy with its eyes obscured, the node that learned the puppy's eyes determines that the picture is an incorrect image (signal) because there are no puppy eyes, and outputs 0, meaning false or incorrect. However, nodes that have learned other parts of the dog, such as the nose, mouth, ears, and legs, judge that the photo is the correct image and output 1, meaning true or correct. At this time, since the values of 1 are produced more than 0 by parts other than the eyes, the AI ultimately calculates the answer for the photo as 'puppy'. This characteristic of AI contributes to automatically ignoring errors and producing the correct answer, just like a human, even if there are typos or incorrect words in the question. That is why some experts call AI's answers probabilistic answers.

In modern times, **research on probability and random algorithms** is the most popular. In general, problems that can be determined as "if A, then B." can be approached relatively easily by computer programs. However, if there are cases where multiple answers can exist, such as 'art' can be 'art' or 'technology', the surrounding circumstances such as 'context' must be considered. But it is difficult to give a clear answer such as "if these words appear before and after, then it is 'art', otherwise it is 'technology'." This type of problem is solved using complex mathematics that deals with statistics and probability. In fact, modern artificial intelligence research is said to be proceeding by assigning categories to each word and interpreting the meaning of the sentence as a whole with more categories. As an extremely simple example, when 'Music is an art' is said, it guesses the category 'art' that includes the two meaningful words in the sentence, music and art, and interprets it according to the context. AlphaGo also belongs to this method.

In artificial intelligence, if a given problem can be solved, all techniques and technologies are used without distinction. If the quality of the results is excellent, technologies with no theoretical background are applied and recognized. Below is a list of only some of the famous technologies and techniques.

Maze exploration algorithm: This is the most basic artificial intelligence algorithm that allows robots (micro mice) or self-driving cars to recognize adjacent terrain features and find their way to a specific destination. It is an algorithm that can operate without machine learning. Naturally, more complex machine learning algorithms are used in commercial products.

Fuzzy theory: A concept that expresses ambiguous states in nature into quantitative values, such as ambiguity in natural language, or, conversely, to change quantitative values into ambiguous values in nature. For example, when a human feels "cool," it determines the temperature and uses it.

Pattern recognition: This refers to finding specific patterns in various linear and nonlinear data such as images, audio sources, and text. In other words, to put it simply, it is a technology that allows computers to judge data similarly to humans and distinguish what kind of data it is.

Machine learning: As the name suggests, it is a field that studies giving computers artificial intelligence that can learn.

Artificial neural network: One of the learning algorithms being studied in the field of machine learning. It is a technology mainly used for pattern recognition, and it reproduces the connections between neurons and synapses of the human brain as a program. To put it simply, it can be seen as 'simulating' 'virtual neurons' (it is not exactly the same as the operating structure of actual neurons). Generally, appropriate functions are given by creating a neural network structure and then 'learning' it. Since the human brain has the best performance among the intelligent systems discovered so far, artificial neural networks that imitate the brain can be seen as a discipline that has developed with a fairly ultimate goal. For more information, refer to the machine learning document. In the 2020s, the computing power of computers has developed at a frightening rate, and the amount and type of data being poured in has also increased accordingly, so artificial neural network technology, which has an excellent ability to process unstructured data [13], is receiving the most attention among all artificial intelligence technologies and is being treated as a technology that will receive even more attention in the future. By now, those who are interested in artificial intelligence will have noticed that a deep neural network is an artificial neural network that is made by connecting numerous artificial neural networks and stacking them in layers, and this is what we commonly know as deep learning.

Genetic algorithm: Finding the appropriate answer to a specific problem by repeating generations through the process of natural evolution, that is, the process of crossover and mutation of populations that make up a certain generation. While most algorithms express the problem as a formula and find the maximum/minimum through differentiation, the purpose of a genetic algorithm is to find the most appropriate answer, not the exact answer, to a problem that is difficult to differentiate.

Energy and computer architecture issue

A program called the deep learning of an artificial neural network is a type of simulation. Running a computer simulation of a large number of cases for a long time consumes a lot of energy for computer operation. In fact, the cost of

electricity was not a big issue until just before 2020, when artificial neural networks began to develop in earnest. However, as various weights and learning methods were introduced to artificial neural networks and parameters were increased, the demand for parallel processing devices, which are much more expensive than the central processing units (CPU), and the corresponding electricity consumption increased vertically to improve performance, which became an issue.

The fundamental problem is that the current von Neumann structure computer components were not initially designed for artificial neural networks. Current computers are separated from the CPU, which is a serial processing device, to the data storage and processing lines and auxiliary data processing devices. In particular, this auxiliary data processing device refers to the graphic processing units (GPU), which perform parallel processing. Since artificial neural networks are designed based on matrix mathematics, simultaneous calculations must be performed in each node (computational unit/coded neuron) that makes up the artificial neural network, and to do so, the parallel processing device must be the main device. This problem of inefficiency is directly linked to the problem of heat generation, so there is a limit to simply increasing the number of transistors through process miniaturization without changing the architecture to improve performance.

For this reason, many semiconductor companies are currently jumping into the development of AI chips to reduce power consumption and increase computational efficiency. The industry predicts that the landscape of AI development will change again when AI chips completely replace GPUs in the future. However, even if an AI chip is developed, it will take a considerable amount of time to establish a software ecosystem that can utilize the chip.

B. Machine Learning

Teaching artificial intelligence to have good intelligence is called **machine learning**. Machine learning is a type of program that gives artificial intelligence the ability to learn without being explicitly programmed. A simple definition of a computer program is "If input data is A and condition B is satisfied, the answer is C." However, machine learning is "If input data is A, train a machine to learn condition B given correct answer C, then make it

conclude C". The comparison between a computer program and machine learning is as follows.

Table 9.1.1 Comparison between a computer program and machine learning

	Computer program	Machine learning
Input	A = 3, B = 2	coefficients 3, 2, 1, 8, 3, 5
Program	C = A * B	3 * X + 2 * Y = 1 8 * X + 3 * Y = 5
Output	C = 3 * 2 = 6	X = 1, Y = -1

There are various machine learning methods for artificial intelligence, and models such as supervised learning (decision tree, Bayes classification, kNN, support vector machine, etc.) and unsupervised learning (k-means clustering, etc.) studied in chapters 6 to 8 of this book are widely used. And recently, the **reinforcement learning** method, which gives rewards instead of outcome values and makes action selection, is also introduced using the **Markovian Decision Process** (MDP) theory. In reinforcement learning, a reward is given every time an action is taken in the current state, and learning proceeds in the direction of maximizing this reward. The video below shows how to teach a robotic arm to play table tennis using reinforcement learning.

https://www.youtube.com/watch?v=SH3bADiB7uQ
(https://www.youtube.com/watch?v=SH3bADiB7uQ)

<Figure 9.1.2> Robotic arm to play table tennis using reinforcement learning

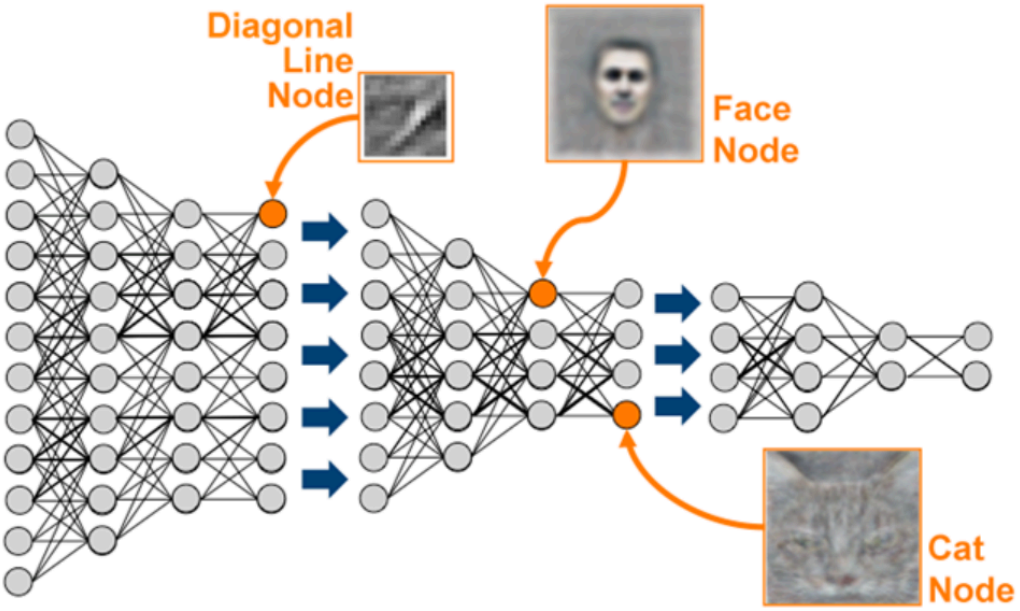
C. Deep learning

The multilayer neural network model learned in Chapter 7 is a model that can handle various data among various machine learning models, but it has limitations in application because it cannot find an accurate solution when solving a complex nonlinear optimization problem using the back-propagation algorithm. In cases where there are multiple hidden layers, the back-propagation algorithm has problems such as vanishing gradients where data disappears and learning does not proceed well, and it also has limitations in processing new data, although it processes learned content well.

However, in 2006, Professor Geoffrey Hinton of the University of Toronto, Canada, solved the vanishing gradient problem through pretraining of neural networks and dropout data, and began calling the neural network that applied this algorithm as **deep learning**. The recent popularity of deep learning is due to the development of an algorithm that overcomes the limitations of the existing back-propagation, the increase in learning materials due to the accumulation of a large amount of data, and the development of graphic processing unit (GPU) hardware suitable for learning and calculation using neural networks. Deep learning has been popular since 2010, and there are several important events that led to this popularity.

1) The error rate of Microsoft's voice recognition program was over 10% until 2010, but after the introduction of the deep learning algorithm, the error rate was drastically reduced to 4%.
2) IMAGENET, a competition in the field of image recognition, is a competition to recognize objects in photos by looking at them, and the error rate was close to 26% until 2011. However, in 2012, the SuperVision team of the University of Toronto in Canada recorded an incredible error rate of 16.4% with AlexNet, which applied a deep convolutional neural network. After that, IMAGENET participants reduced the error rate to 3.6% using the deep learning method. This is a figure that surpasses the error rate of humans.

3) In 2012, Google successfully recognized cats and humans by learning 10 million screens (200×200) of captured YouTube videos. This was an important event in that it trained machines using an unsupervised learning method, not the existing supervised learning method. This project was led by Professor Andrew Ng and was the result of training a neural network with 16,000 CPU cores and 9 layers and 1 billion parameters for 3 days.



<Figure 9.1.3> Pattern recognition of humans and cats using unsupervised learning by Google

These three major events confirmed the performance of deep learning and became an opportunity to be actively used in various fields of society.

9.2 Text mining

With the development of computer technology, a large amount of useful information is being accumulated, and a considerable amount of this is text data. For example, useful information collected from various sources such as books or research papers in digital libraries, news from media outlets, email messages, and web pages is being rapidly stored in text databases. A method of searching for useful information in such text databases in a more efficient way is called **text mining**. Unlike existing databases, text databases have characteristics in that the form and size of documents are unstructured, and therefore require analysis methods different from existing methods.

A traditional method of searching for useful information in text databases is the **keyword information retrieval** method used in libraries. Libraries classify books or reference materials by subject or author and create a database system. When users query using the keywords of the information they are looking for, the computer system searches for documents or books containing these words. Recently, keyword search methods are also used when searching web documents on portal sites such as Google. However, as the amount of digital documents increases, keyword search methods can cause users to be confused about which documents are useful, as the number of web documents searched for a given query reaches thousands. Therefore, it became necessary to create measures of the importance or relevance of the documents being searched and to prioritize the information being searched. This type of research is an example of text mining. Let's discuss various issues for text data mining.

A. Preprocessing of data for text mining

In order to perform text mining, **preprocessing** of data is required. First, **stop words** that are not directly needed for text mining are removed from each document's data. For example, articles, prepositions, and verb endings changed by tense are removed, and prefixes and suffixes are removed from nouns and only the stem is extracted. A document is expressed as a set of terms consisting of such stems and nouns.

B. Measure of the appropriateness of information retrieval

Let R be the set of relevant documents in a text database for a given query, and $n(R)$ be the number of relevant documents. Let T be the set of documents retrieved by a system, $n(T)$ be the number of retrieved documents, and $n(R \cap T)$ be the number of relevant documents in the retrieved documents. To evaluate the documents retrieved for a given query by an information retrieval system, there are precision and recall as measures as follows.

$$\text{precision} = \frac{n(R \cap T)}{n(T)}$$

$$\text{recall} = \frac{n(R \cap T)}{n(R)}$$

Precision is the proportion of relevant documents in the retrieved document set, and recall is the proportion of documents retrieved among the relevant documents. The higher both measures are, the better the retrieval system is.

C. Information retrieval method

Information retrieval methods can be broadly divided into **keyword-based** and **similarity-based** search. Keyword-based search displays each document as a set of keywords, and when a user queries with multiple keywords, the system finds appropriate documents using the queried keywords. The difficult problem with this method is how to find documents that are synonyms or have multiple meanings of the keywords.

Similarity-based search uses the frequency and proximity of keywords to find similar documents. The following cosine distance measure is often used to measure the similarity of documents. Let's display a document as a set of terms excluding stop words. If the terms appearing in the entire document set are $x_1, x_2, \dots x_m$, then each document can be represented as a binary data vector indicating whether the terms appear ('1') or not ('0'), for example, $\boldsymbol{x}_1 = (x_1 = 1, x_2 = 0, \dots x_m = 1)$ by the occurrence of each term. The cosine distance between two document vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is as follows.

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) \ = \ \frac{\boldsymbol{x}_1 \cdot \boldsymbol{x}_2}{||\boldsymbol{x}_1||||\boldsymbol{x}_2||}$$

Here, $\boldsymbol{x}_1 \cdot \boldsymbol{x}_2$ is the inner product of two vectors, which means the number of terms that both documents contain, and $||\boldsymbol{x}_1||$ which means the number of terms in document $\boldsymbol{x}_1$. In other words, the cosine distance means the ratio of both terms included to the average number of terms in each document.

A document can be represented as a binary vector indicating the occurrence of a term, or as a term frequency vector indicating how many times a word appears. Or, it can be represented as a relative frequency of the occurrence frequency of all terms to analyze similarity. For more information, please refer to the references.

Association analysis and classification analysis

Using the binomial vector data of documents for keyword-based retrieval, the relevance of terms can be studied through association analysis. Terms that frequently occur in each document are likely to form phrases that are complexly related to each other, and association analysis can be used to find association rules of interesting terms.

Recently, as the amount of newly created documents increases, the problem of classifying these documents into appropriate document sets has been gaining attention. To this end, a set of pre-classified documents is used as training data, and new documents are classified using the classification analysis.

9.3 Web data mining

The development of network technology has led to the construction of knowledge and information held by individuals, companies, and organizations in the form of web pages, allowing them to share such information in real time. The web, which provides useful information and convenience in daily life, also presents many tasks that must be solved.

- The number of individuals and organizations providing information through the web is constantly increasing, and the information they provide is frequently updated, so the problem of efficiently managing this.
- The web can be thought of as a huge digital library, but each web has a non-standardized structure and various creative forms. In this huge and diverse digital library, the information that users want is a very small portion of less than 0.1%, so the problem of efficiently searching for the information they want.
- Web users are diverse people with different backgrounds, interests, and purposes. However, most users do not have detailed network or search skills, so they can get lost on the web, so the problem of easily guiding them is the problem.

All methods studied to solve such problems are called **web data mining**. Recently, portal sites such as Google have appeared that organize web pages well on behalf of users and search for necessary web pages, competing with users with their own search engines. The search method used is mainly a keyword-based search method used in text mining. As the web becomes larger, there is a problem that when a common keyword is given, thousands of web contents with low relevance are searched, or web contents with synonyms for the keyword are not searched. Web content mining, which searches for websites with good information that users want in web databases that have a more complex structure than text databases, is an important research field. In addition, there are various research fields such as web document classification and web log mining.

A. Web content mining

Studying whether web pages retrieved from a web database are useful can be viewed as a similar concept to text mining. However, the web contains not only pages but also additional information, including hyperlinks that connect other pages within a page. Although text documents may have references, web links contain more reliable information. In particular, web pages called **hubs**, which collect links to reliable web pages on a single topic, do not stand out on their own but provide useful information on common topics.

Many studies have been developed using hyperlinks or hub pages to search for reliable web content, and a useful search method called HITS (Hyperlink Induced Topic Search) is introduced.

1) Search web pages for a given query to form a root set, and form a base set including all pages linked to the web pages in the root set. The pages in the base set are denoted as $\{1,2, ...,n\}$, and the pages linked to each page are denoted as a $n \times n$ matrix $A = \{a_{ij}\}$. The element a_{ij} of the matrix is 1 if page i and page j are linked, and 0 if not.
2) Each page in the base set is given an initial weight $\boldsymbol{w} = (w_1, w_2, \dots w_n = 1)$ and a hub weight $\boldsymbol{h} = (w_1, w_2, \dots w_n = 1)$. At this time, the sum of the squares of the weights is set to 1, and the relationship between the two weights is defined as follows.

$$\boldsymbol{h} = A \cdot \boldsymbol{w}$$
$$\boldsymbol{w} = A'\boldsymbol{h}$$

The first equation means that if a page points to multiple trusted pages, its hub weight should increase. The next equation means that if a page is recommended by multiple hub pages, its weight should increase.

3) The above two equations are applied repeatedly l times, which can be mathematically expressed as follows.

$$\boldsymbol{h} = A \cdot \boldsymbol{w} = (AA') \cdot \boldsymbol{h} = (AA')^2 \cdot \boldsymbol{h} = \cdots = (AA')^l \cdot \boldsymbol{h}$$
$$\boldsymbol{w} = A' \cdot \boldsymbol{h} = (A'A) \cdot \boldsymbol{w} = (A'A)^2 \cdot \boldsymbol{w} = \cdots = (A'A)^l \cdot \boldsymbol{w}$$

That is, each weight is the eigenvector of $(AA')^l$ and $(A'A)^l$.

4) It searches pages with large weights and hub weights of each page first.

The HITS search method has the disadvantage of relying only on links, but it is useful for web page search. It is used in Google's search engine and is reported to obtain better search results than other search engines.

B. Web page classification

We have seen that classification analysis of documents is a major research topic in text mining. There are many studies on the problem of classifying new pages using training pages consisting of pages already classified into specific topics on web pages. For example, portal sites such as Google and Yahoo classify web pages into frequently used topics such as ‘movies’, ‘stocks’, ‘books’, and ‘cooking’ for the convenience of users, and when a new web document appears, it is classified into one of the existing classification topics. The term-based classification method used in text mining also shows good results in the classification of web documents. In addition, the concept of hyperlinks included in web documents can be usefully used in the classification.

C. Web log mining

When a web page server for e-commerce accesses a web page and searches for information, it stores basic information such as each user's IP address, requested URL, and requested time. This is called web log data. The amount of web log data recorded can easily reach hundreds of terabytes (10^{12} bytes), and

analyzing this data is called web log mining. By studying user access patterns for web pages, it is possible to design web pages efficiently, identify potential customers for e-commerce, develop marketing strategies that consider customers' purchasing tendencies, and deliver customized information to users.

In order to analyze web log data, preprocessing processes such as refining, simplifying, and converting the data studied in Chapter 3 are necessary. Then, a multidimensional analysis of IP addresses, URLs, time, and web page content information is performed to find potential customers. By studying association rules, sequential patterns, and web access trends, we can understand users' reactions and motivations, and use the results of these analyses to improve the system, such as page design. We can help build personalized web services for users and attempt to rank web pages.

9.4 Multimedia data mining

The popularization of audio and video equipment, CD-ROMs, and the Internet has led to the emergence of multimedia databases that contain voice, images, and video in addition to text. Examples include NASA's Earth Observation System, the Human Chromosome Database, and the audio-video databases of broadcasting stations. Multimedia data mining is the exploration of interesting patterns within these multimedia databases. It has some similarities to text data mining and web data mining, but it can be seen as involving a unique analysis of image data. Here, we will learn about similarity exploration, classification, and association mining of image data.

A. Similarity Search for Image Data

Similarity search for image data can be divided into description-based search and content-based search. **Description-based** searches for similar images based on image descriptions such as keywords, size, and creation time, and is similar to text mining. However, search based only on image descriptions is generally vague and arbitrary, and the results are not good.

Content-based searches for images based on visual characteristics such as color, texture, and shape of the image, and is widely applied in medical diagnosis, weather forecasting, and e-commerce. Content-based search can be

divided into a method of finding similar image data in a database using image samples and a method of finding data with similar characteristics in a database by describing the characteristics of the image in detail, and the latter is more commonly used. The following methods are available for representing the characteristics of images:

1) Color histogram-based features - This is a method of drawing a histogram for all the colors in an image, which allows you to understand the overall flow, but it may be a meaningless feature because it does not include information about the shape, location, or texture of the image.
2) Multifeature-composed features - These include a combination of the color histogram, shape, location, and texture of the image. A distance function is also introduced for each feature.
3) Wavelet-based features - It expresses shape, texture, and location information using the wavelet transform of the image. However, this method may fail to search for local images because it uses a single wavelet transform for the entire image.

So, it also expresses one image using multiple local wavelet transforms.

B. Classification of image data

Classification of image data is widely used in scientific research such as astronomy and geology. For example, in astronomy, models are built to recognize planets using properties such as size, density, and direction of movement of stars. Since the size of image data is usually large, data transformation, decision tree models, or complex statistical classification models are introduced for classification.

C. Association analysis of image data

Since image data has many features such as color, shape, texture, text, and spatial location, various association analyses are possible, which can be broadly divided into the following three types.

1) Association analysis of words that can characterize the content of an image

For example, an association rule such as ‘If more than 30% of an image is blue → sky’ can be given.

2) Association analysis of other images related to the content of a space in an image
   For example, an association rule such as ‘If there is green grain in a field → there is likely to be a forest nearby’ can be given.
3) Association analysis of other images with the content of an image that is not related to spatial relationships
   For example, an association rule such as ‘If there are two circles → there is likely to be a square’ can be given.

In image association analysis, resolution can also be important. That is, there are cases where features appear to be the same up to a certain resolution but differ at a finer resolution. In this case, association analysis is first performed at a low resolution, and then association analysis is performed while gradually increasing the resolution for patterns that satisfy the minimum support criterion.

9.5 Spatial data analysis

Spatial data analysis is the process of finding interesting patterns in a database containing spatial information and can be used to understand spatial data, discover spatial relationships, and discover relationships between spatial or non-spatial data. This type of spatial data analysis is widely used in medicine, transportation, the environment, management of multinational companies, and geographic information systems.

Spatial data often exists in various data formats and in different regions as databases. In such cases, it is recommended to create a spatial data cube after organizing a data warehouse by subject, time, or other classifications for analysis. Multidimensional analysis can be used to facilitate spatial data analysis.

Spatial data analysis can be used to describe, associate, classify, cluster, analyze spatial trends, and analyze spatial data. For example, spatial association analysis is to find an association rule that says, ‘If the average

monthly income is 3 million won or more and there are many sports centers in an area with a high concentration of apartments.' Spatial cluster analysis generalizes geographical points that are detailed into cluster areas such as commercial, residential, industrial, and agricultural areas according to land use. Spatial classification analysis can be used as an example to classify regions into high-income, low-income, etc. according to average income per household. This type of spatial classification analysis is usually classified in relation to spatial objects such as administrative districts, rivers, and highways. Spatial trend analysis studies changes in spatial data that change over time when there is a variable called time in spatial data and can examine population density according to economic development, etc.