5. Data Mining and Knowledge Discovery

Shuliang Wang, Wenzhong Shi

In this chapter, data mining and knowledge discovery (DMKD) is presented with basic concepts, a brief history of its evolution, mathematical foundations, and usable techniques, along with the data warehouse and the decision support system (DSS). First, dataset and knowledge will be defined and elucidated as under DMKD. DMKD is a discovery process with different hierarchies, granularities, and/or scales. For a set of concepts that may be best understood if being viewed and explained from various perspectives, the chapter starts with a definition followed by a table explaining DMKD from different views (Sect. 5.1). The evolution of DMKD is then briefly tracked from the rapid advance in massive data to the birth of DMKD (Sect. 5.2). Some mathematical foundations are given in Sect. 5.3, i. e. probability theory, statistics, fuzzy set, rough set, data fields, and cloud models. Section 5.4 introduces some usable DMKD techniques. DMKD is used to discover a set of rules and exceptions with association, classification, clustering, prediction, discrimination, and exception detection. In Sects. 5.5 and 5.6, data warehouses and decision support systems are given. The first one mentioned is one of the data sources for DMKD, and DMKD is a new technique to assist the latter with a task. Finally, trends and perspectives are summarized and forecasted into two promising fields, web mining and spatial data mining (Sect. 5.7).

5.1 **Basic Concepts in Data Mining and Knowledge Discovery** 123
5.1.1 Dataset and Knowledge in DMKD 124
5.1.2 Hierarchy, Granularity, and Scale in DMKD 124
5.1.3 DMKD Understood from Different Views 125

5.2 **Evolution of Data Mining and Knowledge Discovery** 125
5.2.1 Rapid Growth of Data Volume 126
5.2.2 Data Usage 126
5.2.3 Birth of DMKD 127

5.3 **Mathematical Foundations of DMKD** 127
5.3.1 Probability and Statistics 127
5.3.2 Fuzzy Sets 128
5.3.3 Rough Sets 128
5.3.4 Data Fields 129
5.3.5 Cloud Models 130

5.4 **Techniques of DMKD** 130
5.4.1 Qualitative and Quantitative Transform 130
5.4.2 Association 131
5.4.3 Classification 132
5.4.4 Clustering 133
5.4.5 Prediction 135
5.4.6 Discrimination 135
5.4.7 Exception 136

5.5 **Data Warehouse for DMKD** 137
5.5.1 Data Warehouse Architecture 137
5.5.2 Data-Warehouse Design 138

5.6 **Decision Support System for DMKD** 138
5.6.1 DSS Architecture 139
5.6.2 DSS Models 139

5.7 **Trends and Perspectives** 140
5.7.1 Web Mining 140
5.7.2 Spatial DMKD 140

References 141

5.1 Basic Concepts in Data Mining and Knowledge Discovery

Data mining and knowledge discovery (DMKD) is the efficient extraction of interesting, previously unknown, potentially useful, and ultimately understandable knowledge from huge datasets under a given task with constraints [5.1–4]. The process of DMKD may be grouped into three significant steps [5.5, 6].

1. Data preparation (locating the mining target, collecting background information, cleaning data).
2. Data mining (reducing data dimensions, selecting mining techniques, discovering knowledge).
3. Knowledge application (interpretation, evaluation, and application of the discovered knowledge).

5.1.1 Dataset and Knowledge in DMKD

The dataset may be an untreated accumulation of noisy, fuzzy, random, disorderly and unsystematic data, consisting of, e.g., positions, attributes, texts, raster images, vector graphics, logs, voices, and multimedia in special circumstances. They are internal, external and in various formats. An integrated and shared collection of logically related data constitutes a database to meet the information needed by an organization. Because of the challenge of dealing with large amounts of data in the database, data warehousing that is subject-oriented, integrated, time-variant, and nonvolatile, are developed for advanced data analyses and further DMKD. So the datasets to be mined might be all of current information assets to access, e.g., databases, data cubes, data warehouses, data marts, and knowledge bases.

The knowledge to discover is a set of rules and exceptions, along with different hierarchies, granularities, and/or scales. With given data of sufficient size and quality, DMKD may be capable of predicting trends, describing patterns, or detecting exceptions [5.7]. The trends and patterns are the rules to show the intersection of two or more objects or attributes according to a particular set of procedures. The rules may be an association rule, a characteristics rule, a discrimination rule, a clustering rule, a classification rule, a serial rule, a predictive rule, and so on. The exception is the interesting outlier. The knowledge bases capture the experience derived from observations and interpretations of past events or phenomena, and the application of methods to past situations, in the form of rules, case studies, standard practices, and typical descriptions of objects and object systems that can serve as prototypes.

5.1.2 Hierarchy, Granularity, and Scale in DMKD

People may observe and analyze the same entity from very different cognitive levels, and actively jump between the different levels. As a computerized simulation of human cognition, DMKD may be implemented with different hierarchies, granularities, and/or scales for the same datasets for various needs. That is, it should discover the knowledge not only in worlds with the same hierarchy, granularity, and scale, but also in worlds with different hierarchy, granularity, and/or scale (Fig. 5.1).

1. The hierarchy reflects the level of cognitive discovery and describes the summarized transformation from the micro-view world to the macro-view world, e.g., knowledge with different demands.
2. The granularity reflects the precision of interior detail of DMKD and describes the combined transformation from the fine world to the coarse world, e.g., images with different pixel size.
3. The scale reflects the measurement of exterior geometry of DMKD, and describes the zoomed transformation from the large world to the small

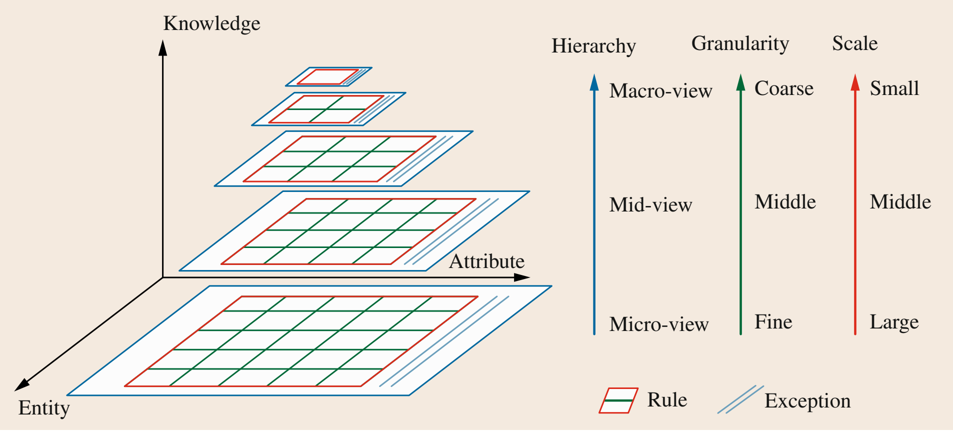

Fig. 5.1 DMKD with hierarchies, granularities, and/or scales

Table 5.1 DMKD from different views

View	Data mining and knowledge discovery (DMKD) is ...
Subject	An interdisciplinary confluence of database systems, statistics, artificial intelligence, machine learning, pattern recognition, network.
Cognition	An inductive process that leads from vast concrete datasets to a small set of knowledge, from special phenomena to generic rules.
Analysis	A process of uncovering knowledge from huge amounts of datasets via a sets of interactive, repetitive, associative, and data-oriented manipulations.
Logic	A data-mining process to discover knowledge but not to prove knowledge, and the knowledge is constrained by the data and tasks.
System	A computerized process from input to output of original data in a database, cleaned data in a data warehouse, background information.
Methodology	A method of matching the multidisciplinary philosophy of human thinking that suitably deals with the complexity, uncertainty, and variety when summarizing data and representing knowledge.

world, e.g., observing a map with the manipulation of zoom-in or zoom-out.

5.1.3 DMKD Understood from Different Views

As an interdisciplinary topic, DMKD may be understood from different views (Table 5.1) [5.3]. From the data to knowledge view, it was presented in a database system view by *Han* and *Kamber* [5.2], summarized in a geospatial table by *Thearling* [5.8], and integrated in a SDMKD (spatial DMKD) pyramid by *Wang* [5.9, 10].

In the context of hierarchy, granularity, and/or scale, the users at different levels want the knowledge from the different worlds, i. e. the most concrete knowledge with the smallest amount, a connecting knowledge between the preceding world and the following world, and the most abstract knowledge with the biggest amount. With micro-view hierarchy, fine granularity, or large scale, DMKD is to uncover the complicated external phenomenon in order to locally distinguish the detailed distinctions for gaining individual knowledge, the mining algorithms of which may be accurate. Contrarily, with macro-view hierarchy, coarse granularity, or small scale, DMKD is to ignore razor-thin distinctions in order to globally seek the generalized commonness for gaining public knowledge, the mining algorithms of which may be smooth. For instance, seen from the same monitoring databases, a landslide firstly shows DMKD continuously observed data instead of discrete symbolic parameters or qualitative concepts. The top decision-maker to guide the final direction may ask for the most generalized knowledge for the whole landslide. In order to bridge upper decision and lower decision the middle decision-maker may demand the common knowledge for each breaking-sector of the landslide. To monitor the exact deformation rules, the bottom decision-maker may want detailed knowledge of each monitoring point.

5.2 Evolution of Data Mining and Knowledge Discovery

The evolution of DMKD may be regarded as a part of the evolution of database system technology (Fig. 5.2). When the related data are collected, databases are created under database management systems (DBMS). DBMS is software to define and manipulate database. For a database system, it is necessary to provide specific database, DBMS, and related software. Under the umbrella of new science and technology, database systems have been developed in several directions [5.2]. One is advanced database systems to study new data models for new applications. Another is data analysis and cognition to interpret datasets for further applications, one representative of which is DMKD (Chap. 3).

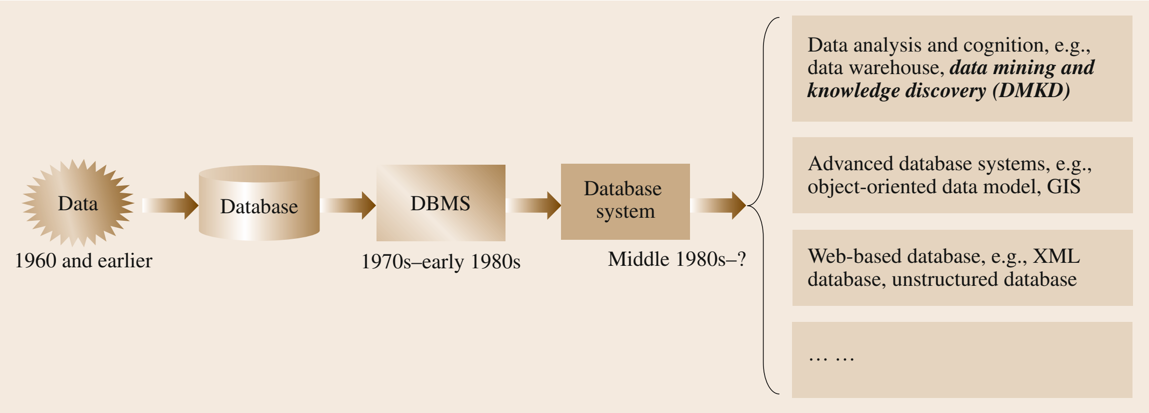
Data
1960 and earlier
Database
DBMS
1970s–early 1980s
Database system
Middle 1980s–?
Data analysis and cognition, e.g., data warehouse, *data mining and knowledge discovery (DMKD)*
Advanced database systems, e.g., object-oriented data model, GIS
Web-based database, e.g., XML database, unstructured database
… …

Fig. 5.2 DMKD evolution in the context of database system technology

5.2.1 Rapid Growth of Data Volume

The rapid increase in data acquisition, transmission, and storage has resulted in the growth of vast computerized datasets at unprecedented rates. These datasets come from various sectors, e.g., business, education, government, scientific community, Internet, or one of many readily available off-line and online data sources in the form of text, graphics, images, video, audio, animation, hyperlinks, markups, and so on. Moreover, they are continuously increasing and are amassed in both the attribute depth and the scope of instance objects every time. These phenomena may be much more serious in geospatial science.

Many data are georeferenced [5.11]. Spatial data may come from natural resource investigations, surveying and mapping, astronomical data, satellites, and spacecraft images. They include not only positional data and attribute data, but also spatial relationships among spatial entities. Moreover, the spatial data structure is more complex than the tables in an ordinary relational database. In addition to tabular data, there are raster images and vector graphics in a spatial database. Their attributes are not explicitly stored in the database. Contemporary Geographical Information Systems (GIS) analysis functionalities are not enough to make full use of spatial datasets.

Now, the huge amounts of computerized datasets have far exceeded human ability to completely interpret and use these datasets [5.3]. Many decisions are made on large spatial datasets, e.g., the National Spatial Data Infrastructure, Digital Earth, the National Aeronautics and Space Administration, the National Geospatial-Intelligence Agency, and the National Cancer Institute. These decisions are spread across many application domains including Earth science, ecology and environmental management, public safety, epidemiology, and climatology [5.4]. Therefore, it is possible and necessary to extract valuable and usable knowledge from the scalable repository of data and information.

5.2.2 Data Usage

In order to understand and make full use of these data repositories, some techniques have been investigated and tried, e.g., data analysis, machine learning, expert systems, artificial intelligence, and especially database system technology. In the database system, there are specific database, DBMS, and related software. Examples are data models (e.g., hierarchical, network, relational) along with tools (e.g. entity-relationship), indexing and accessing methods (e.g. B-trees, hashing), query languages (e.g. Structured Query Language), user interfaces, forms, reports, query processing and query optimization, transactions, online transactional processing (OLTP), and concurrency control and recovery. Advanced database systems are used to study new data models (e.g., extended relational, object-oriented) for new applications (e.g., spatial, temporal, multimedia, active, stream and sensor, scientific and engineering, knowledge-based).

Data analysis and cognition are used to identify and interpret datasets for further applications. When data were warehoused [5.12] OLTP was changed into online analysis and processing (OLAP). In 2008, *Kimball* [5.13] published his book on the data warehouse toolkit. With the development of data warehouses, some new terms were introduced, such as dimensions and facts, data marts in the 1960s, and business data warehouse in 1988. In 1991, Prism Solutions introduced Prism Warehouse Manager, software for developing a data warehouse [5.12–15].

In order to improve the decision-making process, a decision support system (DSS)was proposed in the late 1970s [5.16]. The first decision support tools included ad hoc query and reporting tools, optimization and simulation models, OLAP, DMKD and data visualization [5.17]. In parallel with DSS, spatial decision support system (SDSS) evolved, associated with the need to expand the GIS capabilities for tackling complex, ill-defined, spatial decision problems [5.18].

5.2.3 Birth of DMKD

Although the aforementioned techniques are used to manipulate data, it is ultimately up to humans to cast the data into a usable hypothesis. When these techniques are used for time-constrained analysis, some variables or instances that seem to be unimportant often have to be discarded. However, the discarded items may carry core information about unknown patterns. Second, the accompanying need for improved computational engines can be matched in a cost-effective manner with parallel multiprocessor computer technology. Third, web techniques develop very quickly, which broadens the scope of data of the same entity, i. e., from local data to global data. As the Internet provides dynamic sources of data and information, the web is becoming an important part of the computing community. Fourth, most techniques focus on a database or data warehouse, which is physically located in one place. However, many data may be distributed in heterogeneous sites. People directly integrate industry-standard data warehouses and OLAP platforms during the process of data utilization. Finally, the decision support system is an interactive software-based system intended to assist human decision-makers to compile useful information for identifying conflicts and making decisions, rather than replace them. It also provides the decision-making patterns instead of producing a solution to a problem [5.16].

Because of the challenge of dealing with large amounts of datasets during the process of decision-making, knowledge discovery in databases was proposed in 1989 to answer the questions that traditionally were too time-consuming to resolve. In 1995, the name data mining was introduced for the search for valuable knowledge based on similarities between sweeps through large databases. As both data mining and knowledge discovery in databases virtually point to the same techniques, people like to identify them together, i. e., data mining and knowledge discovery, which has also become the name of an international journal. Without pre-selecting a subset of variables or instances under DMKD, users can explore the full depth of a database and make inferences regarding small but important segments of a population because larger samples often generate lower estimation errors and variance.

Now, some techniques have been given for DMKD along with the applications, e.g., generalization, classification, association, clustering, outlier detection, and prediction. When DMKD is implemented on high performance parallel processing systems in the web, they can analyze massive databases in minutes. Faster processing allows users to experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield an improved precision of the predictions.

5.3 Mathematical Foundations of DMKD

The usable mathematical foundations in common for DMKD may be probability theory, statistics, fuzzy set, rough set. The theory of data fields and cloud models was developed primarily in China [5.9]. These are addressed as typical techniques in the following sections.

5.3.1 Probability and Statistics

Probability and statistics [5.19] are suitable for DMKD with randomness on the basis of stochastic probabilities in the contexts of adequate samples and background information.

The probability is extended to an interval instead of an exact value in evidence theory. Given the same information on an uncertain quantity, the plausibility function and the belief function can be regarded as upper probability and lower probability of possible values [5.20]. Spatial statistics (Chap. 2) primarily consists of geostatistics, spatial point patterns, and lattice data, with the algorithms of spatial autocorrelation, spatial interpolation, spatial regression, spatial interaction and simulation, and modeling [5.21]. Geostatistics is a spatial process indexed over a continuous space; spatial point patterns pertain to the location of *events* of interest, and lattice data are spatial data indexed over a lattice of points [5.22]. Based on probability and statistics, a set of physical or abstract objects may be grouped into classes of similar objects, the process of which is called clustering analysis [5.23].

5.3.2 Fuzzy Sets

A fuzzy set [5.24] is used in DMKD with fuzziness on the basis of a fuzzy membership function that depicts an uncertain probability [5.3].

As a consequence of the fuzzy characteristics of natural classes and concepts, a fuzzy set characterizes the fuzziness via a fuzzy membership function instead of the crisp characteristics function, and maps the uncertainty to a numerical value on the interval [0,1] instead to a set of the two values {0, 1}. The fuzzy membership function is the relationship between the values of an element and its membership belonging to a set [5.25]. Fuzzy sets allow an element to be partially in a set, and each element is given a fuzzy membership value ranging from 0 to 1, e.g., the membership of an element belonging to a concept. An element is assigned to a series of membership values in relation to the respective subsets of the universe of discourse since the concept of multiple and partial class membership is fundamental to fuzzy sets. If one only allowed the extreme membership values of 0 (not an element of the set) and 1 (an element of the set), this would actually be equivalent to a crisp set [5.26]. Simultaneously, fuzzy sets deal with the similarity of an element to a class. The accumulation of membership values for one element can exceed 1, but differs from the probability that always adds up to 1. The entity with mixed classification, indeterminate boundary, or gradual change may be described with more than one fuzzy membership value.

5.3.3 Rough Sets

A rough set [5.27] is used in DMKD with incompleteness via lower and upper approximation [5.27–29].

Given the universe of discourse U that is a finite and nonempty set. Suppose an arbitrary set $X \subseteq U$. X^{c} is the complement set of X, and $X \cup X^{\mathrm{c}} = U$; U/R is the equivalence class set composed of disjoint subsets of U partitioned by $R \subseteq U \times U$ on U, $[x]_R$ is the equivalence class of R including element x, and (U, R) formalizes an approximate space. The definitions for lower and upper approximation are

lower approximation (interior set) of X on U :

$$\mathrm{Lr}(X) = \{x \in U | [x]_R \subseteq X\} ,$$

upper approximation (closure set) of X on U :

$$\mathrm{Ur}(X) = \{x \in U | [x]_R \cap X \neq \Phi\} .$$

If the approximate space is defined in the context of region, then

positive region of X on U:

$$\mathrm{pos}(X) = \mathrm{Lr}(X) ,$$

negative region of X on U:

$$\mathrm{neg}(X) = U - \mathrm{Ur}(X) ,$$

boundary region of X on U:

$$\mathrm{bnd}(X) = \mathrm{Ur}(X) - \mathrm{Lr}(X) ,$$

where X is defined if $\mathrm{Lr}(X) = \mathrm{Ur}(X)$, while X is rough with respect to $\mathrm{bnd}(X)$ if $\mathrm{Lr}(X) \neq \mathrm{Ur}(X)$. A subset $X \in U$ defined with the lower approximation $\mathrm{Lr}(X)$ and upper approximation $\mathrm{Ur}(X)$ is called rough set. The lower approximation $\mathrm{Lr}(X)$ is the set of elements that surely belong to X, while the upper approximation $\mathrm{Ur}(X)$ is the set of elements that possibly belong to X. The difference of the upper approximation and the lower approximation is the uncertain boundary $\mathrm{bnd}(X)$. It is impossible to decide whether or not an element in $\mathrm{bnd}(X)$ belongs to the spatial entity because of the incompleteness of the set. The rough degree is

$$\begin{aligned} R_{\mathrm{degree}}(X) &= \frac{R_{\mathrm{card}}(\mathrm{Ur}(X) - \mathrm{Lr}(X))}{R_{\mathrm{card}}(X)} \times 100\% \\ &= \frac{R_{\mathrm{card}}(\mathrm{bnd}(X))}{R_{\mathrm{card}}(X)} \times 100\% , \end{aligned}$$

where $R_{\mathrm{card}}(X)$ denotes the cardinality of set X; X is crisp when $R_{\mathrm{degree}}(X) = 0$. Based on whether or not the statistical information is used, the existing rough set models may be grouped into two major classes such as algebraic and probabilistic models [5.30].

Spatial entities may be depicted with rough sets (Fig. 5.3). Suppose that there is an entity X and let the amount of element units be

$$\begin{aligned} U &= 36 , \\ \mathrm{Lr}(X) &= 4 , \\ \mathrm{Ur}(X) &= 20 , \\ \mathrm{pos}(X) &= \mathrm{Lr}(X) = 4 , \\ \mathrm{neg}(X) &= U - \mathrm{Ur}(X) = 36 - 20 = 16 , \\ \mathrm{bnd}(X) &= \mathrm{Ur}(X) - \mathrm{Lr}(X) = 20 - 4 = 16 , \\ R_{\mathrm{degree}}(X) &= \tfrac{16}{20} \times 100\% = 80\% . \end{aligned}$$

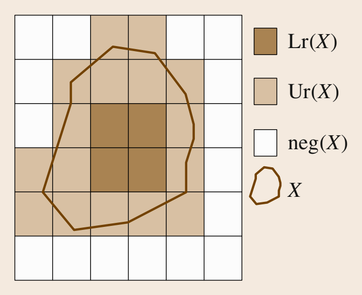

Fig. 5.3 Illustration of a rough set

5.3.4 Data Fields

A data field [5.9] is applied in DMKD with inadequately sampled data via diffusing the effect of observed data to the mining task from the sample to the universe of discourse. That is, understand the population of datasets with the sampled data.

In physics, an entity, e.g., the electric charge, radiates the effect from its location to the universe of discourse, which produces a physical field. Similarly, observed data may be regarded as diffusing the effect to DMKD from its sampled location to the universe of discourse in which the data field lives. In order to depict the power of the observed data in the universe of discourse, the potential function of the data field may be derived from physical fields with the assumption that all observed data diffuse the effects and are influenced by the others at the same time.

The data field is often classified into two categories according to its behavior under the symmetry transformations of space and time, named scalar field and vector field. In this chapter, a scalar field with its potential function is chosen as an example to express the properties. For a single data field created by sample A, the potential ϕ of a point x_1 in the number universe can be computed by

$$\varphi(x) = m \times \mathrm{e}^{-\left(\frac{\|x-x_1\|}{\sigma}\right)^2} ,$$

where $\|x - x_i\|$ is the distance between A and x_i, m ($m \geq 0$) denotes the power of A, and σ indicates influential factors. Usually, $\|x - x_1\|$ has a Euclidean norm. Many types of influential factors σ, e.g., radiation brightness, radiation gene, data amount, space between the neighbor isopotentials, grid density of Cartesian coordinates and so on, contribute to the data field.

In most cases, there is more than one sampling point in the universe of discourse. In order to obtain the power of any point under these circumstances, the effects of all samples should be considered. Because of overlap, the potential of each sample point in the number universe is the sum of all data potentials. Referring to the potential function, the potential can be calculated using

$$\varphi(x) = \sum_{i=1}^{n} \left(m_i \times \mathrm{e}^{-\left(\frac{\|x-x_i\|}{\sigma}\right)^2} \right) ,$$

where the sum is defined over all the sample points.

In a similar way to the distribution of a scalar physical field, the equipotential line or surface can be utilized to describe the spatial distribution of the potential function in the low-dimensional potential field. More specifically, given a potential value Ψ, the corresponding equipotential line or surface can be obtained, that is, according to the set of potential values $\{\Psi_1, \Psi_2, \ldots\}$ that satisfies $\phi(x) = \Psi$, a series of equipotential lines or surfaces can be the spatial distribution of potential function. The sketch map of a two-dimensional data field is shown in Fig. 5.4.

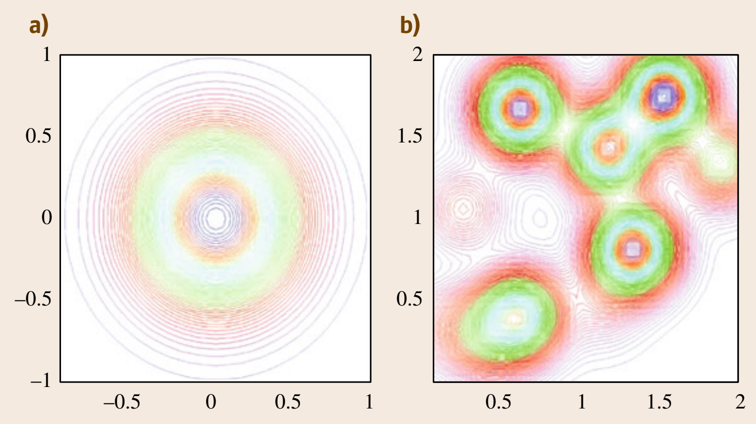

Fig. 5.4a,b Map of a two-dimensional data field. **(a)** Equipotential line of a one-point data field; **(b)** equipotential line of multipoint data field

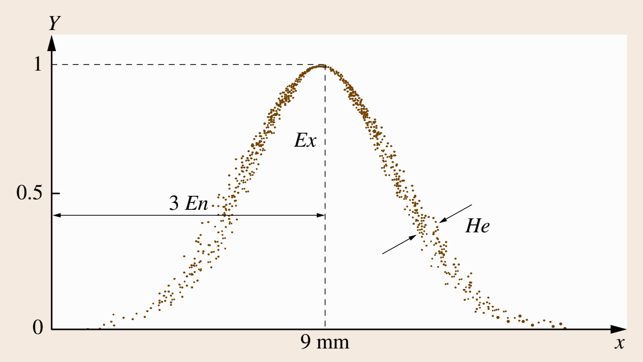

Fig. 5.5 The cloud model of *displacement is* 9 mm *around*; x is the displacement, y is the membership to the concept

5.3.5 Cloud Models

A cloud model [5.31] is a mathematical model to implement an uncertainty transition between a qualitative concept and its quantitative data to represent the concept in DMKD via integrating randomness and fuzziness.

The cloud model describes the fuzziness and randomness by means of a set of three characteristics {*Ex*, *En*, *He*} (Fig. 5.5). In the universe of discourse, *Ex* (expected value) is the position corresponding to the center of the cloud gravity, the elements of which are fully compatible with the linguistic concept; *En* (entropy) is to depict the concept coverage, i. e. a measure of the fuzziness, which indicates how many elements could be accepted to the linguistic concept; and *He* (hyperentropy) is to measure the dispersion on the cloud drops, which can also be considered as the entropy of *En*. Figure 5.5 shows that the cloud model as a whole is observable as a shape but fuzzy in detail, which is similar to the natural cloud in the sky. A piece of cloud is composed of many droplets represented by data, any one of which is a stochastic mapping in the universe of discourse from a qualitative fuzzy concept, along with the membership of the data belonging to the concept. Given {*Ex*, *En*, *He*} on a concept, the data may be generated to depict the cloud droplets of a piece of cloud, which is called the forward cloud generator. Given datasets, {*Ex*, *En*, *He*} may also be generated to represent a concept that is called the backward cloud generator.

The cloud model integrates randomness and fuzziness in a mathematical mapping with membership. When the cloud model is generated, the mapping of each cloud droplet is random, and its membership is fuzzy, the process of which integrates randomness and fuzziness. Take the displacement of landslide as an example, the cloud model of *displacement is around* 9 mm in Fig. 5.5 is composed of many cloud droplets, and each cloud droplet is a stochastic mapping with its membership, i. e., stochastic data {. . . , 8 mm, 9 mm, 10 mm, . . . } with the membership {. . . , 0.9, 1, 0.9, . . . }. The concept of *displacement is around* 9 mm is qualitative, while the data {. . . , 8 mm, 9 mm, 10 mm, . . . } and the membership {. . . , 0.9, 1, 0.9, . . . } are quantitative. The data {. . . , 8 mm, 9 mm, 10 mm, . . . } are the displacements from actual landslide monitoring, and the membership {. . . , 0.9, 1, 0.9, . . . } describes to what degree the data belong to the concept of *displacement is around* 9 mm under the given membership function.

5.4 Techniques of DMKD

The core components of data mining technique have been in use for decades in specialized analysis tools that work with relatively small volumes of data [5.32]. Today, the maturity of these techniques, coupled with high-performance relational database engines and broad data integration efforts, make these techniques practical in the environments of data warehouse and network [5.4, 33]. Because DMKD is an interdisciplinary subject, there are various techniques associated with different types of knowledge [5.34]. In order to discover useful knowledge, it is suitable to select more than one technique to mine the same datasets simultaneously under the given mining task. When they are applied, some techniques are developed further [5.35].

5.4.1 Qualitative and Quantitative Transform

Qualitative and quantitative transform is to link the numerical data and the conceptual knowledge when necessary during the process of DMKD. While the amount of data is huge, the volume of extracted knowledge is often very small. The more condensed the knowledge, the bigger the disparity.

The backward cloud generator is to extract the conceptual knowledge from numerical datasets, and the forward cloud generator is to represent the qualitative concept with quantitative data. Both of the issues are elementary in DMKD. The type of cloud model has to be chosen according to the type of data distribution. Since a normal distribution is ubiquitous, in the following the normal cloud model is taken as an example to present the algorithms of the forward cloud generator and backward cloud generator. In detail the algorithms read as follows.

Algorithm of the Forward Normal Cloud Generator

Input.

Ex, *En*, *He* – three characteristics of a concept;
N – the number of cloud droplets to be generated.

Output.

Cloud droplet (x_i, y_i) $(i = 1, 2, \dots, N)$ – the quantitative positions of N cloud droplets in data space and the membership of each cloud droplet to the concept.

Steps.

1. Produce a normally distributed random number En' with mean En and standard deviation He.
2. Produce a normally distributed random number x with mean Ex and standard deviation En';
3. Calculate
$$y = \mathrm{e}^{-\frac{(x-Ex)^2}{2(En')^2}} .$$
4. Drop (x_i, y_i) is a cloud drop in the universe of discourse.
5. Repeat steps 1–4 until N cloud-drops are generated.

Algorithm of the Backward Normal Cloud Generator

Input.

Cloud droplet (x_i, y_i) $(i = 1, 2, \dots, N)$ – the quantitative positions of N cloud droplets in data space and the membership of each cloud drop to the concept.

Output.

Ex, *En*, *He* – three characteristics of the concept.

Steps.

1. Calculate the mean value of x_i $(i = 1, \dots, N)$,
$$Ex = \frac{1}{N}\sum_{i=1}^{N} x_i .$$
2. For each pair of (x_i, y_i), calculate
$$En_i = \sqrt{-\frac{x_i - Ex}{2 \cdot \ln y_i}} .$$
3. Calculate the mean value of En_i $(i = 1, \dots, N)$,
$$En = \frac{1}{N}\sum_{i=1}^{N} En_i .$$
4. Calculate the standard deviation of En_i,
$$He = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(En_i - En)^2} .$$

Because it builds the formal relationship between the concept and its data independent of the application field, the cloud model overcomes the weakness of rigid and precisely defined specifications, which often conflict with the fuzzier human recognition process. Moreover, it performs the interdynamic transition between the qualitative concepts and its quantitative data through the use of mathematical functions under the data distribution. During the process of DMKD with the cloud model, the quantitative data first generate several essential cloud models. The more condensed cloud model may be generated from the essential cloud models by hierarchy. The top hierarchy of DMKD is the most generalized knowledge, while the bottom hierarchy of DMKD is the objective data in databases.

5.4.2 Association

Association is to discover a logic rule among different sets of entities that associate one or more objects with other objects, to study the frequency of items occurring together in datasets. It identifies the collections of data attributes that are statistically related in the underlying data, e.g., the identification of seemingly unrelated products that are often purchased together. Association rule mining is that, given a database of transactions, a minimal confidence threshold and a minimal support threshold, find all association rules whose confidence and support are above the corresponding thresholds. So along with an associated rule, it is necessary to give some parameters, e.g., support and confidence.

The association rule is of the form

$$X \Rightarrow Y \quad \text{[confidence, support]} ,$$

where X and Y are disjoint conjunctions of attribute–value pairs. The confidence of the rule is the conditional probability of Y given X, $P(Y|X)$, and the support of the rule is the prior probability of X and Y, $P(X \cap Y)$. Here, probability is taken to be the observed frequency in the dataset. For example,

$$\begin{aligned}&\text{Rain (location, amount of rain)} \Rightarrow\\ &\quad\text{Landslide (location, occurrence)}\\ &\quad[\text{confidence} = 98\%, \text{support} = 76\%]\,.\end{aligned}$$

5.4.3 Classification

Classification is to discover a rule that defines whether an entity belongs to a particular class with a defined kind and quantity or a set of classes. For example, classifying remotely sensed images based on spectrum and GIS data.

In DMKD, there are countless real examples where the probability of one event is conditional on the probability of a previous one. Bayes' theorem shows the relation between one conditional probability and its inverse [5.36–38]. When applied to large databases under Bayes' theorem, the Bayesian classifier can statistically predict a class membership probability that a given object belongs to a particular class, with high accuracy and speed [5.39]. A Bayesian network represents a joint generalized probability density function over the universe of discourse [5.40–42].

Bayes' Theorem

Suppose that X is an object with a tuple of attributes $(a_1, a_2, \ldots, a_m)$, i.e. $X = (x_1, x_2, \ldots, x_m)$ in which $x_1, x_2, \ldots, x_m$ are the values of $a_1, a_2, \ldots, a_m$. The discourse of class C is partitioned into $C = (C_1, C_2, \ldots, C_n)$ and their prior probabilities $P(X) > 0$, $P(C_i) > 0$, $(i = 1, 2, \ldots, n)$. Then, the probability $P(C_i|X)$ that X belongs to C_i is

$$\begin{aligned}P(C_i|X) &= \frac{P(X|C_i)P(C_i)}{\sum_{j=1}^{n} P(X|C_j)P(C_j)}\\ &= \frac{P(X|C_i)P(C_i)}{P(X)}\,,\end{aligned}$$

where $P(C_i|X)$ is the posterior probability of C_i conditioned on X. $P(X|C_i)$ is the posterior probability of X conditioned on C_i, and $P(X)$ is constant for all classes. If $P(C_j)$ are not known, it is commonly assumed that $P(C_1) = P(C_2) = \ldots = P(C_n)$, or $P(C_j)$ are estimated by the ratio of the number of training tuples of class C_j in D to the number of training tuples in D. Given data sets with many attributes it would be extremely computationally expensive to compute $P(X|C_j)P(C_j)$.

Bayesian Classifiers

In order to simplify the computations of Bayes' theorem, a naïve Bayesian classifier assumes that the effect of an attribute value on a given class is independent of the values of the other attributes, which is called class conditional independence

$$\begin{aligned}P(X|C_j) &= \prod_{k=1}^{m} P(x_k|C_j)\\ &= P(x_1|C_j) \times P(x_2|C_j) \times \cdots \times P(x_m|C_j)\,.\end{aligned}$$

From the training tuples, the probabilities $P(x_k|C_j)$ can be estimated along with attribute a_k. If a_k is categorical, then $P(x_k|C_j)$ is the number of tuples of class C_j in D having the value x_k for a_k, divided by the number of tuples of class C_j in D. If a_k is continuous-valued, then it is typically assumed to have a Gaussian distribution defined by

$$P(x_k|C_j) = \frac{1}{\sqrt{2\pi}\sigma_{C_j}} \mathrm{e}^{-\frac{(x_k-\mu_{C_j})^2}{2\sigma_{C_j}^2}}\,,$$

where μ_{C_j}, σ_{C_j} are, respectively, the mean and the standard deviation of the values of attributes a_k for training tuples of class C_j. Therefore, the classifier is

$$\begin{aligned}&\text{if } P(C_i|X) = \max_{j=1}^{n}\left[P(C_j|X)\right] \quad \text{then } X \in C_i\,,\ \text{or}\\ &\text{if } P(X|C_i)P(C_i) = \max_{j=1}^{n}\left[P(X|C_i)P(C_i)\right]\\ &\quad\text{then } X \in C_i\,.\end{aligned}$$

When the assumption of class conditional independence holds true, the naïve Bayesian classifier is the most accurate in comparison with all other classifiers [5.2].

If the assumption proves false in practice, Bayesian belief networks may be used [5.43].

Bayesian Belief Networks

The Bayesian network representation consists of a set of local conditional generalized probability density functions combined with a set of conditional independence assertions that allow the construction of a global generalized probability density function from the local generalized probability density function. A belief network is defined by a directed acyclic graph and a set of conditional probability tables. The conditional probability table for an attribute variable $a_i (i = 1, 2, \ldots, m)$ specifies the conditional distribution $P(a_i|\text{Parents}(a_i))$, where $\text{Parents}(a_i)$ are the parents of a_i. Given its parents, each variable is conditionally independent of its nondescendants in the network graph. This allows the network to provide a complete representation of the existing joint probability distribution with

$$P(x_1, x_2, \dots, x_m) = \prod_{i=1}^{m} P(x_i \,|\text{Parents}(a_i)) = \prod_{i=1}^{m} P(x_i \,|x_{i+1}, x_{i+2}\ , \dots, x_m)\ ,$$

where $P(x_1, x_2, \dots, x_m)$ is the probability of a particular combination of the values of X, and the values for $P(x_i|\text{Parents}(a_i))$ correspond to the entries in the conditional probability table for a_i. Rather than returning a single class label, the classification of Bayesian belief networks process can return a probability distribution that gives the probability of each class.

5.4.4 Clustering

Clustering is to discover a segmentation rule that groups a set of objects by virtue of their similarity without the knowledge of what causes the grouping and how many groups exist. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. For example, grouping crime locations to find distribution patterns, and partitioning facial expressions to identify various persons.

With different clustering distance, some objects may be close to one another according to one distance but farther away according to the other distance [5.7]. Such distances are, eg, Minkowski distance, Manhattan distance, Euclidean distance (Manhattan and Euclidian distance are covered in Chap. 6). Because they are derived from the matching matrix, some measures are given to compare various clustering results when different clustering algorithms perform on a set of data [5.23].

Minkowski Distance

Minkowski distance is a metric defined on Euclidean space. Between two points $P = (x_1, x_2, \dots, x_n)$ and $Q = (y_1, y_2, \dots, y_n) \in \mathbb{R}^3$, the Minkowski distance of order p is defined as

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} .$$

When $p = 1$, Minkowski distance is Manhattan distance. When $p = 2$, it becomes the Euclidean distance. In the limiting case of p reaching infinity, Chebyshev distance (or maximum value distance) is obtained

$$\lim_{p\to\infty} \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} = \max_{i=1}^{n} |x_i - y_i|\ .$$

So it can be considered as a generalization of some distances, e.g., Euclidean distance or Manhattan distance.

Chebyshev Distance

Chebyshev distance is a metric defined on a vector space where the distance between two vectors is the largest among all their differences along any standard coordinate. In DMKD, it examines the absolute magnitude of the differences between coordinates of a pair of objects. Since the minimum number of moves needed by a king to move from one square on a chessboard to another, equals the Chebyshev distance between the centers of the squares in the game of chess, it is also known as chessboard distance.

Mahalanobis Distance

Formally, the Mahalanobis distance (or *generalized squared interpoint distance*) of a multivariate vector $\boldsymbol{x} = (x_1, x_2, x_3, \dots, x_N)^{\mathrm{T}}$ from a group of values with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^{\mathrm{T}}$ and covariance matrix $\mathbf{S}$ is defined as

$$D_M(x)\sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{S}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})}\ .$$

Mahalanobis distance can also be defined as a dissimilarity measure between two random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of the same distribution with the covariance matrix $\mathbf{S}$.

Fuzzy Comprehensive Clustering

In fuzzy sets, fuzzy comprehensive evaluation and fuzzy clustering analysis are two essential techniques. Integrating them into DMKD, knowledge may be discovered reasonably well [5.44].

- Firstly it acquires the fuzzy evaluation matrix on each influential factor.
- Then all fuzzy evaluation matrices multiply the corresponding weight matrices, the product matrix of which is the comprehensive matrix on all factors.
- Third, the comprehensive matrix is further used to create a fuzzy similar matrix, on the basis of which a fuzzy equivalent matrix is obtained.
- Fourth, fuzzy clustering is implemented via the proposed maximum remainder algorithms.

Table 5.2 Land value is influenced by the factors and their detailed factors, along with weights

Factors	$u_1 =$ point factors			$u_2 =$ linear factors		$u_3 =$ polygon factors			
Factor weights	$a_1 = 0.5$			$a_2 = 0.2$		$a_3 = 0.3$			
Detailed factors	$u_{11} =$ hospital	$u_{12} =$ station	$u_{13} =$ park	$u_{21} =$ road	$u_{22} =$ river	$u_{31} =$ power	$U_{32} =$ disaster	$U_{33} =$ pollution	$U_{34} =$ geology
Detailed factor weights	$a_{11} =$ 0.15	$a_{12} =$ 0.05	$a_{13} =$ 0.3	$a_{21} =$ 0.15	$a_{22} =$ 0.05	$a_{31} =$ 0.1	$a_{32} =$ 0.05	$a_{33} =$ 0.1	$a_{34} =$ 0.05

In the next paragraph we use these sets and matrices:

U	set of influential factors
$\mathbf{A}$	matrix of weights
u_i	set of subfactors
$\mathbf{A}_i$	matrix of weights
V	set of grades
$\underset{\sim i}{\mathbf{X}}$	fuzzy membership matrix
$\underset{\sim i}{\mathbf{Y}}$	fuzzy evaluation matrix of u_i
$\underset{\sim}{\mathbf{Y}}^{(p)}$	the fuzzy evaluation matrix of U, i. e., all factors, on the spatial entity p
$\mathbf{Y}_{l\times n}$	a total matrix of fuzzy comprehensive evaluation on all entities
$\underset{\sim}{\mathbf{R}}$	fuzzy similar matrix
$t(\underset{\sim}{\mathbf{R}})_{l\times l}$	fuzzy equivalent matrix.

Suppose that the set of influential factors is $U = \{u_1, u_2, \dots, u_m\}$ with the matrix of weights $\mathbf{A} = (a_1, a_2, \dots, a_m)^{\mathrm{T}}$, and the set of detailed influential factors $u_i = \{u_{i1}, u_{i2}, \dots, u_{iki}\}$ $(i = 1, 2, \dots, m)$ with the matrix of weights $\mathbf{A}_i = (a_{i1}, a_{i2}, \dots, a_{iki})^{\mathrm{T}}$ $(i = 1, 2, \dots, m)$ simultaneously. For example, land value is influenced by the factors and their detailed factors, along with their weights (Table 5.2).

The set of evaluating grades is $V = \{v_1, v_2, v_3, \dots, v_n\}$, including n grades. In the context of the given grades, the fuzzy membership matrix $\underset{\sim i}{\mathbf{X}}$ on the detailed influential factors of factor u_i may be described as

$$\underset{\sim i}{\mathbf{X}} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{k_i1} & x_{k_i2} & \dots & x_{k_in} \end{pmatrix} .$$

When multiplying the fuzzy evaluation matrix and its weight matrix, the fuzzy evaluation matrix of u_i becomes $\underset{\sim i}{\mathbf{Y}} = \mathbf{A}_i \cdot \underset{\sim i}{\mathbf{X}}$. The fuzzy evaluation matrix of U, i. e., all factors, on the spatial entity p is computed via

$$\underset{\sim}{\mathbf{Y}}^{(p)} = \mathbf{A} \cdot \underset{\sim}{\mathbf{Y}} = \mathbf{A} \cdot (\underset{\sim 1}{\mathbf{Y}}, \underset{\sim 2}{\mathbf{Y}}, \dots, \underset{\sim i}{\mathbf{Y}}, \dots, \underset{\sim m}{\mathbf{Y}})^{\mathrm{T}} .$$

When p $(p \geq 1)$ spatial entities are evaluated at the same time, a total matrix of fuzzy comprehensive evaluation can be obtained. Let the number of spatial entities be l, then

$$\underset{\sim l\times n}{\mathbf{Y}} = \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \dots & \dots & \dots & \dots \\ y_{l1} & y_{l2} & \dots & y_{l_n} \end{pmatrix} ,$$

where $p = 1, 2, \dots, l$. Taking the element $y_{ij}(i = 1, 2, \dots, l;\ j = 1, 2, \dots, n)$ of the matrix $\underset{\sim l\times n}{\mathbf{Y}}$ as original data, the fuzzy similarity matrix can be created via

$$r_{ij} = \frac{\sum_{k=1}^{n} (y_{ik} \times y_{jk})}{\left(\sum_{k=1}^{n} y_{ik}^2\right)^{1/2} \times \left(\sum_{k=1}^{n} y_{jk}^2\right)^{1/2}} , \quad \text{i.e.,}$$

$$\underset{\sim}{\mathbf{R}} = (r_{ij})_{l\times l} ,$$

which indicates the fuzzy similarity relationships among the entities.

The fuzzy similarity matrix $\underset{\sim}{\mathbf{R}} = (r_{ij})_{l\times l}$ has to be changed into the fuzzy equivalent matrix $t(\underset{\sim}{\mathbf{R}})$ by the self-squared method [5.44] when clustering. There is k $(k = 1, 2, \dots, l$, and $k \leq l)$ that makes the equivalent matrix $t(\underset{\sim}{\mathbf{R}}) = \underset{\sim}{\mathbf{R}}^{2^k}$ if and only if $\underset{\sim}{\mathbf{R}}^{2^k} = \underset{\sim}{\mathbf{R}}^{2^{k+1}} = \underset{\sim}{\mathbf{R}}^{2^{k+2}} = \dots = \underset{\sim}{\mathbf{R}}^{2^l}$. Then $t(\underset{\sim}{\mathbf{R}})$ can be calculated by

$$\begin{aligned} \underset{\sim}{\mathbf{R}} &\to \underset{\sim}{\mathbf{R}} \cdot \underset{\sim}{\mathbf{R}} = \underset{\sim}{\mathbf{R}}^{2^1} \to \underset{\sim}{\mathbf{R}}^{2^1} \cdot \underset{\sim}{\mathbf{R}}^{2^1} = \underset{\sim}{\mathbf{R}}^{2^2} \to \underset{\sim}{\mathbf{R}}^{2^2} \cdot \underset{\sim}{\mathbf{R}}^{2^2} \\ &= \underset{\sim}{\mathbf{R}}^{2^3} \to \dots \to \underset{\sim}{\mathbf{R}}^{2^k} , \end{aligned}$$

where $i = 1, 2, \dots, l;\ j = 1, 2, \dots, l;\ i \neq j;\ \underset{\sim}{\mathbf{R}} \cdot \underset{\sim}{\mathbf{R}} = \max(\min(r_{i1}, r_{j1}), \min(r_{i2}, r_{j2}), \dots, \min(r_{ik}, r_{jk}), \dots, \min(r_{il}, r_{jl}))$; $\max(.)$, $\min(.)$ are to choose the maximum or minimum of two elements, respectively. The complexity of the calculus is $2^{k-1} < l \leq 2^k$, that is,

$k < \log_2 n + 1$. The fuzzy equivalent matrix $t(\underset{\sim}{\mathbf{R}})$ is

$$t(\underset{\sim}{\mathbf{R}})_{l\times l} = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1l} \\ t_{21} & t_{22} & \dots & t_{2l} \\ \dots & \dots & \dots & \dots \\ t_{l1} & t_{l2} & \dots & t_{l_l} \end{pmatrix} ,$$

with $t(\underset{\sim}{\mathbf{R}})$ and α. The entities can be clustered by using maximum remainder algorithms; α is a fuzzy confidential level that is the fuzzy probability that two or more than two entities belong to the same cluster.

Maximum Remainder Algorithms

Input. $t(\underset{\sim}{\mathbf{R}})$, α.

Output. A set of clusters.

Steps.

1. Select fuzzy confidential level α.
2. Summarize the elements column by column in the fuzzy equivalent matrix $t(\underset{\sim}{\mathbf{R}})$, excluding the diagonal elements

$$T_j = \sum_{i=1}^{l} t_{ij} , \quad (i \neq j, i, j = 1, 2, \dots, l) .$$

3. Compute the maximum and the ratio

$$T_{\max}^{(1)} = \max(T_1, T_2, \dots, T_l) , \quad K_j^{(1)} = \frac{T_j}{T_{\max}^{(1)}} .$$

4. Put the $K_j^{(1)} \geq \alpha$ of entity j into the first cluster.
5. Repeat the steps 3 and 4 in the remaining T_j until the end.

5.4.5 Prediction

Prediction is to discover an inner trend that forecasts future values of some variables when the temporal or spatial center is moved to another one, or predicts some unknown or missing attribute values based on other seasonal or periodical information. Examples are a forecast bankruptcy and other forms of default, a forecast movement trend of landslide based on available monitoring data, and an identification of the segments of a population likely to respond similarly to given events.

There are times when the use of prior knowledge would be a useful contribution. For example, in gathering data from deep-space observatories and planetary probes, people do not always know what to expect or even have hypotheses for what to test when gathering such data. Classical inferential models do not permit the introduction of prior knowledge into the calculations, which is an appropriate response to prevent the introduction of extraneous data that might skew the experimental results. Bayesian inference [5.45] is useful because it allows the inference system to construct its own potential systems of meaning upon the data. Once any implicit network is discovered within the data, the juxtaposition of this network against other datasets allows for quick and efficient testing of new theories and hypotheses.

5.4.6 Discrimination

Discrimination is to discover a different feature that distinguishes one entity from another, to compare the core features of objects between a target class and a contrasting class. For example, comparing land prices in a suburban area with land prices in an urban center.

Characterizing both certainties and uncertainties, rough sets are incompleteness-based reasoning in the form of a decision-making table. Rough sets-based DMKD is also a process of intelligent decision-making under the umbrella of given datasets (Fig. 5.3). $\mathrm{Lr}(X)$ is certainly *Yes*, $\mathrm{neg}(X)$ is surely *No*, while both $\mathrm{Ur}(X)$ and $\mathrm{bnd}(X)$ are uncertainly *Yes or No*. With respect to an element $x \in U$, it is sure that $x \in \mathrm{pos}(X)$ belongs to X in terms of its features, but $x \in \mathrm{neg}(X)$ does not belong to X; while it cannot be ensured by means of available information whether or not $x \in \mathrm{bnd}(X)$ belongs to X. So it can be seen that $\mathrm{Lr}(X) \subseteq X \subseteq \mathrm{Ur}(X) \subseteq U$, and $\mathrm{Ur}(X) = \mathrm{pos}(X) \cup \mathrm{bnd}(X)$.

Furthermore, spatial entities may be depicted with rough sets under different granularities or different dimensional spaces. In Fig. 5.6a, Fig. 5.3 is used to study granularity. When the granularity becomes double fine it looks like in Fig. 5.6b,c and shows Fig. 5.3 in three dimensions.

Probabilistic rough sets can be formulated on the basis of the notions of rough membership functions as

$$\begin{aligned} &\mu_X(x), \mu_X(x) \in [0, 1] , \\ &\mu_X(x) = \frac{R_{\text{card}}\,(X \cap [x]_R)}{R_{\text{card}}\,([x]_R)} \\ &\quad = \begin{cases} 1 & x \in \mathrm{pos}(x) \\ (0, 1) & x \in \mathrm{bnd}(x) \\ 0 & x \in \mathrm{neg}(x) \\ 1 - \mu_{X^c}(x) & x \in X^c \end{cases} \end{aligned}$$

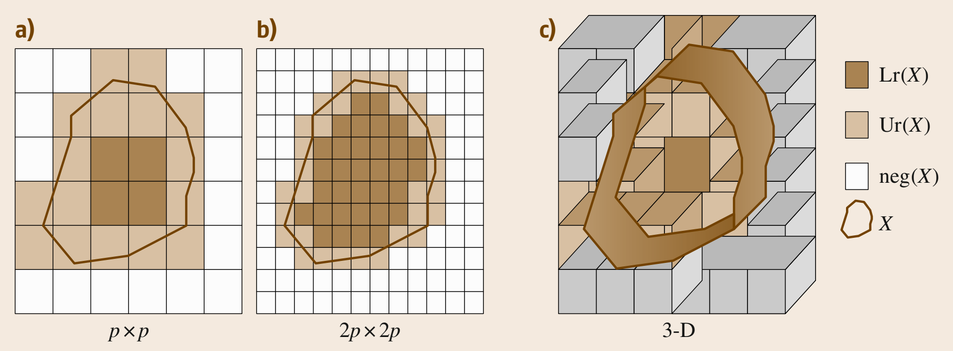

Fig. 5.6a–c Illustrations of rough sets (after [5.46])

The rough membership value may be regarded as the probability of $x \in X$ given that x belongs to an equivalence class. That is, it is taken for a conditional probability to illustrate a certain degree of x belonging to X, $\mu_X(x) + \mu_X^c(x) = 1$. Let $P(X|[x]_R) = \mu_X(x)$ and $\alpha \in [0, 1]$, a probabilistic rough set in α context is defined

$$\begin{aligned}
&\mathrm{Lr}_\alpha(X) = \{x | P(X|[x]_R) \geq 1 - \alpha\} ,\\
&\mathrm{Ur}_\alpha(X) = \{x | P(X|[x]_R) > \alpha\} .
\end{aligned}$$

In this sense, $\mu_X(x)$ gives a probabilistic rough space of X via a pair of upper approximation and lower approximation.

An entity is described by a set of attributes $A = (a_1, a_2, \ldots, a_m)$. In the context of DMKD, the attributes may be divided into conditional attributes $C = (a_1, a_2, \ldots, a_c)$ and decision attributes $D = (a_{c+1}, a_{c+2}, \ldots, a_m)$, $A = C \cup D$, $C \Rightarrow D$. There is a subset S of C, $S = (a_p)(1 \leq p \leq c)$, $S \subseteq C$. If $C - S \Rightarrow D$, then S is superfluous; $a_p(1 \leq p \leq c)$ in S is called the superfluous attribute on the decision attributes D, under which a_p can be reduced. The attribute reduction may decrease the number of dimensions in DMKD.

5.4.7 Exception

Exception is to detect an outlier that is isolated from common rules or derivates from other data observations substantially, used for identifying anomalous attributes and entities. For example, a monitoring point detecting exceptional movement predicts landslide and detects fraudulent credit card transactions.

During the discovery process, some observations may deviate so much from other data observations [5.47] that they identify and explain exceptions to the rules. For example, spatial trend predictive modeling first discovers the centers that are a local maximum of some nonspatial attribute, and then determines the theoretical trend of some nonspatial attribute when moving away from the centers. Finally a small number of deviations are found that do not follow the predicted trend. These deviations may be noise or may be generated by a different mechanism. So besides the commonly used rules, outlier detection is used to extract the interesting exceptions from datasets in DMKD via statistics, clustering, classification, and regression. Outlier detection can also identify system faults and fraud before they escalate with potentially catastrophic consequences [5.9, 47].

Traditionally, outlier detection has been studied using statistics, and a number of discordance tests have been developed. Most of these treat the outliers as *noise* and try to eliminate their effects by removing them or by developing some outlier-resistant methods [5.48]. In fact, these outliers may remedy the rules. In the context of DMKD, they are treated as meaningful input signals rather than noise. In some cases, outliers represent unique characteristics of the objects that are important to an organization. Therefore, a piece of generic knowledge is the form of a rule plus an exception.

Although outlier detection has been used for centuries to detect and remove anomalous observations from data, there is no rigid mathematical definition of what constitutes an outlier. Ultimately, it is a subjective exercise to determine whether or not an observation is an outlier [5.49, 50]. There are three fundamental approaches to outlier detection [5.50].

1. The unsupervised approach to ascertaining the outliers without prior knowledge of the data, which processes the data as a static distribution, pinpoints the most remote points, and flags them as potential outliers.

2. The semi-supervised approach to modeling only normality, in which the normal class is taught but the algorithm learns to recognize abnormality.
3. The supervised approach to modeling both normality and abnormality, which requires pre-labeled data, tagged as normal or abnormal.

A normal distribution of the data is assumed to identify observations that are deemed unlikely on the basis of mean and standard deviations. Distance-based methods frequently use the distance to the k nearest neighbors to label observations as outliers or nonoutliers [5.51].

5.5 Data Warehouse for DMKD

A data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of the management's decision-making process [5.12, 14, 15].

Different from an organization's existing operational database, the data warehouse is a centralized repository that integrates data from several heterogeneous data sources. Its integration is used for reorganization by forming new combinations of data, while historical and stored data in the warehouse are generally not revised. The new requirements of the data warehouse often demand gathering, cleaning and integrating new data from data marts that are tailored for ready access by users. A data mart is a repository of data gathered from operational data and other sources and is designed to serve a particular community of knowledge workers, while the data warehouse is a repository of an organization's electronically stored data [5.13, 14].

5.5.1 Data Warehouse Architecture

In data-warehouse architecture, there may be four interconnected layers that are data sources, data access, warehouse repository, and informational access (Fig. 5.7).

The data sources layer contains the internal and external data of the data warehouse. Large data sources may include scanned graphs, text documents, hypertext documents, spreadsheets, images, sounds and video. The example documents are policies and procedures, product specifications, and catalogs, and corporate historical documents, including minutes of meetings and correspondence. Flat file data are extracted to a temporary intermediate layer for cleaning, transformation and integration, and are finally loaded into the data warehouse or data mart, e.g., operational databases, and organization's enterprise resource planning (ERP) system.

The data access layer is the interface between the data sources layer and the informational access layer. The data warehouse serves as a foundation for improved DMKD and decision making. It should adequately guarantee accurate and error-free data, a technically easy access, incremental change, protection against misuse and loss of data, and a well-defined model. In Fig. 5.7, after data cleaning, users extract the required data from

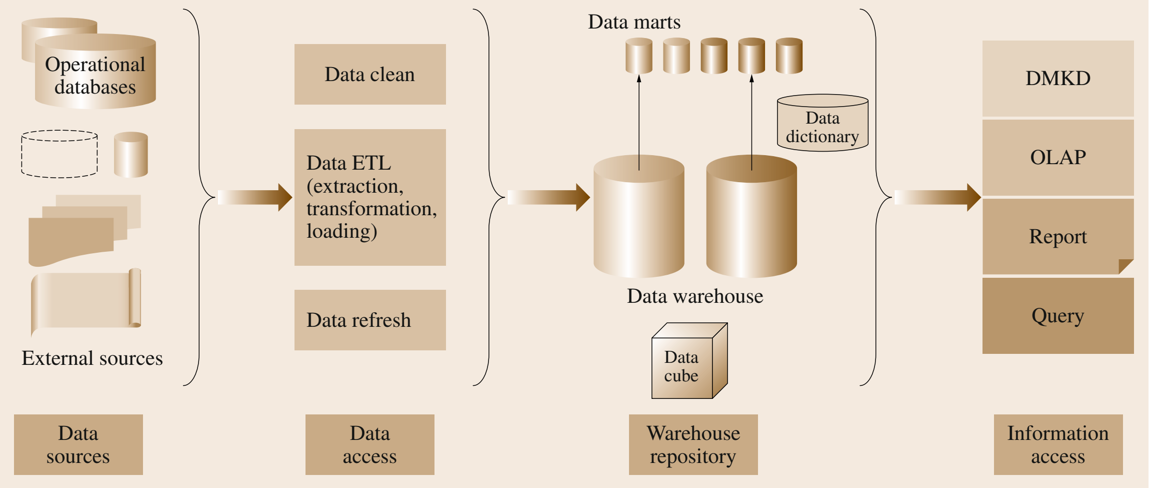

Fig. 5.7 Data warehouse architecture

the data source and load the data into the data warehouse in accordance with a pre-defined data-warehouse model. ETL (extraction, transformation, and loading) is responsible for the data in distributed heterogeneous data sources such as relational databases. Because the stored data are filtered and transformed, it avoids using the wrong data analysis tools and incorrect analysis results. Furthermore, data warehouses are now accessed on the web.

The warehouse repository layer is the core of data warehouse, including the data warehouse for an organization and the data dictionary for the metadata of the data directory. Data marts are further necessary when the data warehouse is considered in a department of the organization. Data cubes may be derived from the data warehouse under multidimensional constraints. The data warehouse transforms the integrated data into a multidimensional data model for efficient querying and analysis so that all data elements relating to the same real-world event or object are linked together. There are two leading approaches to storing data in a data warehouse: the dimensional approach and the normalized approach. A data warehouse uses the database structure for modeling, where each dimension corresponds to one or a set of properties. Each unit is stored as a value that represents the kind of aggregation. The actual physical structure of the data warehouse can be a relational data store or a data cube, which provides a multidimensional view of data and allows for pre-computational and fast access to aggregated data.

The informational access layer contains the data accessed and the tools for reporting and analyzing them. The data warehouse contains a standardized, consistent, clean and integrated form of data coming from various operational systems in use in the organization, structured in a way that specifically addresses the reporting and analytical requirements. When decision-making, the data warehouse manipulated by computerized tools and operators provide additional functionality, OLAP provide the higher level of functionality and decision support that is linked to analysis of large collections of historical data, and DMKD provide the highest implicit patterns.

5.5.2 Data-Warehouse Design

There are two main methods for designing a data warehouse, namely bottom-up design [5.13] and top-down design [5.14]. They are interrelated in the informational access layer of the data-warehouse architecture (Fig. 5.7). Additionally, a hybrid design has evolved to take advantage of the fast turn-around time of the bottom-up design and the organization-wide data consistency of top-down design.

In the bottom-up design, firstly data marts are created to provide reporting and analytical capabilities for specific business processes. These data marts, which contain atomic data and, if necessary, summarized data, can eventually be merged to create a comprehensive data warehouse. The combination of data marts is managed through the implementation of a data-warehouse bus architecture. Business values can be returned as soon as the first data marts have been created. Maintaining tight management over the data warehouse bus architecture is fundamental to maintaining the integrity of the data warehouse. It is important to make sure that dimensions among data marts are consistent for management.

In top-down design, a data warehouse is a centralized repository for the entire organization, in which the data warehouse is designed to use a normalized organization data model to provide a logical framework for delivering business intelligence (BI) and business management capabilities. Atomic data at the lowest level of detail are stored in the data warehouse. Dimensional data marts containing data needed for specific business processes or specific departments are created from the data warehouse as it enables a relational database to emulate the analytical functionality of a multidimensional database. Since all data marts are loaded from the centralized repository, the top-down design methodology generates highly consistent dimensional views of data across data marts. It is a relatively simple task to generate new dimensional data marts against the data stored in the data warehouse. The main disadvantage of the top-down approach is that it represents a very large project with a very broad scope.

5.6 Decision Support System for DMKD

The decision support system (DSS) is an information system that supports business and organizational decision-making activities.

DSS is a set of tools rather than solutions to a pre-determined set of problems. It is an information application along with a human component that can sift through large amounts of data picked from many choices and presents it suitably so that users make decisions more easily, while an operational application collects the data in the course of normal business operation [5.52]. Furthermore, decision-making is a process from theory to practice, while DMKD is a process from practice to the theory. DMKD is thus one of the technologies to aid DSS, as well as the database and the data warehouse.

The decision-making process may be regarded as an outcome of cognitive process to select a course of action among several alternatives. Both human and computer-based resources and capabilities working interactively may come up with the best solution. The computerized results of DSS are only referenced for decision-making in the real world. The more important the decision, the more carefully the DSS results are referenced. For example, the financial crisis of 2008/2009 has demonstrated that a final decision always requires the weighting of interest done by a responsible human and that the blind belief in proposals generated by a computer DSS can become extremely dangerous.

5.6.1 DSS Architecture

In the DSS architecture, there are four elements [5.17], input, user knowledge and expertise, output, and decisions (Fig. 5.8). The DSS taxonomy uses the mode of assistance as the criterion [5.17] and user interface should be graphical in nature together with online help.

The input components are the factors, numbers, and characteristics that shall be analyzed. These components are combined with the user's knowledge and expertise, which are specialized problem-solving skills consisting of knowledge about a particular domain, understanding problems within that domain, and skill at solving some of these problems. The output components are the agglomerated data from which the decisions are derived. The decisions are the results generated by the DSS based on the user's criteria and the decision context, e.g., environments, documents, business models, comparative sales figures, projected revenue figures, personal knowledge, and past experience.

In a DSS, the actual application for users allows decision-making in a particular problem area and the hardware/software environment allows for developing specific applications with case tools or systems. The tools include lower-level hardware/software, e.g., special languages, function libraries, and linking modules. An iterative developmental approach allows the DSS to be changed and redesigned at various intervals. Once the system is designed, it will need to be tested and revised for the desired outcome.

When applied, DMKD is a technical supportable tool for all data-referenced decision-making processes, e.g., business, retails, police, medicine, transportation, robot, navigation, GIS, remote sensing, and GNSS. It is known that the data stored in the data warehouse are in support of management's decision-making processes. Making a right decision is usually based on the quality of data and the ability to sift through and analyze the data to find trends from which solutions and strategies can be created.

5.6.2 DSS Models

There are several types of DSS models. The models may be passive, active and cooperative under the relationship of the user in mind. The models may also be organizational, departmental, and single user in the context of the decision-making scope. In general, DSS models are data-driven, model-driven, communications-driven, document-driven, or knowledge-driven with the mode of assistance as the underlying basis [5.17, 53].

A data-driven model focuses on faster, real-time access to and manipulation of a time-series of internal and sometimes external real-time data in larger, better integrated databases. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality. Data warehouse, OLAP, and DMKD may provide much higher functionality. A model-driven model emphasizes the access to and the manipulation of mathematical models or business

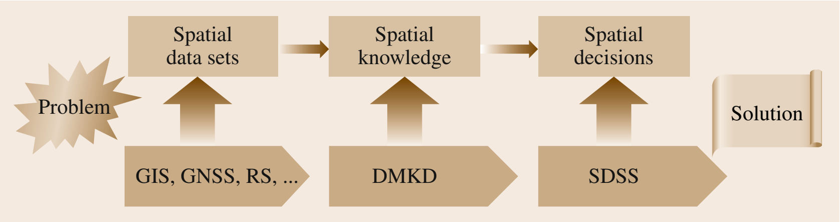

Fig. 5.8 Architecture of a spatial decision support system RS: remote sensing; DMKD: data mining and knowledge discovery; SDSS: spatial decision support system

models to produce the patterns for supporting a solution with limited data and parameters. Simple quantitative models provide the most elementary level of functionality. A model-driven DSS will be more complex, yet understandable, and systems built using simulations and their accompanying visual displays will be increasingly realistic. A communications-driven model focuses on facilitating decision-relevant collaboration and communication when many collaborators work together to come up with a series of decisions in an office or on the web, by using network and communication technologies. The communication technologies are the dominant architectural component whose tools include groupware, real-time video conferencing and computer-based bulletin boards. A document-driven (text-oriented) model focuses on providing document retrieval and analysis for making decisions as well as further manipulating the information to refine strategies, by using computer storage and processing technologies. A search engine is a primary decision-aiding tool associated with a document-driven DSS. A document-driven DSS will access larger repositories of unstructured data, and the systems will present appropriate documents in more useable formats. A knowledge-driven (suggested or knowledge-based) model recommends actions to decision-makers by using more sophisticated and more comprehensive rules or facts. The advice from knowledge-driven DSS will be better and the applications will cover broader domains. A web-based DSS is a computerized system that delivers decision support information or tools to a user with a web browser, i. e. a web-based collection of bookmarks.

5.7 Trends and Perspectives

Currently, new techniques and equipment are creating new challenges for DMKD. For example, the Internet of Things [5.54] enables the Internet to reach out into the real world of physical objects. Ubiquitous devices are mutually connected and extensively used, e.g., laptops, palmtops, cell phones, and wearable computers, in which the enormous quantity of data is being created, transmitted, maintained, and stored. Digital Earth globally provides spatial data for DMKD, along with further applications. If the Internet of things is spatially localized, a smart planet with instruments, interconnection and intelligence may come into being, in which the amount of data resources will be much larger. Under the Internet environment, more generic and useful knowledge can be discovered from the localized data together with global data in the context of a global model. Now, more and more new DMKD applications are happening, i. e. networked topological mining, community discovery, stream data mining, bio-data mining, time serial mining, text mining, intrusion detection, and privacy-preservation with DMKD. Sections 5.7.1 and 5.7.2 present two promising trends in DMKD.

5.7.1 Web Mining

Web mining is to extract useful patterns and implicit information from artifacts or activities related to the World Wide Web under the Internet [5.55, 56]. The Internet provides a technology platform for further extending the capabilities and deployment of computerized decision support. And the World Wide Web broadens data resources. The release of the HTML 2.0 specifications was a turning point in the development of DMKD. Many DBMS vendors shifted their focus to web-based analytical applications and business intelligence solutions. The enterprise knowledge portals combined information portals, knowledge management, business intelligence, and communications-driven DSS in an integrated web environment.

Web mining may include web-content mining, web-structure mining, and web-usage mining. Web-content mining is to discover the knowledge from the content of web-based data, documents, and pages or their descriptions. Web-structure mining is to uncover the knowledge from the structure of websites and the topological relationship among different websites [5.57]. Web-usage mining is to extract web-user behavior or modeling and predicting how a user will use and interact with the web [5.58]. Because people are almost being buried by reports, correspondence, memos, and other paperwork, it is greatly promising to extract new, never-before encountered knowledge from a body of textual sources in the web. In DMKD, web mining is one of the most promising areas. *Srivastava* et al. [5.59] have given a taxonomy of different web-mining applications, for example, personalization, system improvement, site modification, business intelligence, and usage characterization.

5.7.2 Spatial DMKD

Spatial data are more complex, more dynamic and bigger than common affair datasets. The explosive growth of spatial data and widespread use [5.60] of spatial databases emphasizes the need for considering spatial dimensions in DMKD. Spatial dimension means each item of data has a spatial reference [5.61], where each entity occurs on the continuous surface, or where the spatial-referenced relationship exists between two neighbor entities. In order to discover spatial knowledge, a DMKD branch in geospatial science has been developed further, i. e., spatial data mining and knowledge discovery (SDMKD) [5.7, 11]. SDMKD was also regarded as an interdisciplinary area at the intersection of computer science and GIS [5.4].

The complexity of spatial data and intrinsic spatial relationships limits the usefulness of conventional data mining techniques for extracting spatial patterns. Given proper analysis discovery, spatial data can represent a great deal of spatial information as many of them are image-oriented [5.62]. With spatial databases, spatial data warehouses, data cubes, spatial OLAP, spatial knowledge bases, and related types of data, SDMKD may help people to understand spatial data, find out the relationships between two spatial data items or between nonspatial items and spatial items, and characterize spatial entities. Examples could be the association of a road and a bridge, the clustering of similar grassland, the analysis of the spatial distribution of bank locations, the predication of future land use, the creation of buffers related to service regions of government agencies, and the overlapping regions of several neighboring retail shops. The knowledge discovered may further enrich the knowledge bases [5.1].

Because the spatial knowledge discovered can support and improve spatial data-referenced decision-making, spatial data mining has attracted significant attention from computer scientists as well as spatial statisticians. Moreover, it is clear that the acquisition, storage, manipulation, and visualization of geospatial data are special and require substantially different approaches and assumptions to those in other fields [5.2, 34, 62–64].

References

5.1 J. Wang: *Encyclopedia of Data Warehousing and Mining* (Idea Group Reference, Hershey 2006)

5.2 J. Han, M. Kamber: *Data Mining: Concepts and Techniques*, 2nd edn. (Academic, San Francisco 2001)

5.3 D.R. Li, S.L. Wang, D.Y. Li: *Theories and Applications of Spatial Data Mining* (Science Press, Beijing 2006)

5.4 S. Shekar, H. Xiong (Eds.): *Encyclopedia of GIS* (Springer, New York 2007)

5.5 M.W. Berry: *Survey of Text Mining: Clustering, Classification, and Retrieval Scanned by Velocity* (Springer, Berlin Heidelberg 2004)

5.6 T. Dasu: *Exploratory Data Mining and Data Cleaning* (Wiley, New York 2003)

5.7 M. Ester, A. Frommelt, H.-P. Kriegel, J. Sander: Spatial data mining: databases primitives, algorithms and efficient DBMS support, Data Min. Knowl. Discov. **4**, 193–216 (2000)

5.8 K. Thearling: *An Introduction to Data Mining* (Vertex Business Services, Richardson 2001)

5.9 S.L. Wang: Data field and cloud model-based spatial data mining and knowledge discovery. Dissertation, Wuhan University, Wuhan (2002)

5.10 J. Wang: *Data Mining: Opportunities and Challenges* (Idea Group Reference, Hershey 2002)

5.11 D.R. Li, T. Cheng: KDG: Knowledge discovery from GIS – Propositions on the use of KDD in an intelligent GIS, Proc. ACTES, Can. Conf. GIS (1994)

5.12 W.H. Inmon: *Building the Data Warehouse* (QED, London 1992)

5.13 R. Kimball: *The Data Warehouse Lifecycle Toolkit* (Wiley, New York 2008)

5.14 W.H. Inmon: *Tech Topic: What is a Data Warehouse?*, Vol. 1 (Prism Solutions, Brighton 1995)

5.15 W.H. Inmon: *Building the Data Warehouse*, 4th edn. (Wiley, New York 2005)

5.16 F. Burstein, C.W. Holsapple: *Handbook of Decision Support System* (Springer, Berlin Heidelberg 2008)

5.17 D.J. Power: *A Brief History of Decision Support Systems*, (DSSResources.COM, Cedar Falls 2007) available at http://DSSResources.COM/history/dsshistory.html, version 4.0 (March 10, 2007)

5.18 P.J. Densham, M.F. Goodchild: Spatial decision support systems: A research agenda, Proc. GIS/LIS'89, Orlando (1989) pp. 707–716

5.19 A.M. Arthurs: *Probability Theory* (Dover, London 1965)

5.20 G. Shafer: *A Mathematical Theory of Evidence* (Princeton Univ. Press, Princeton 1976)

5.21 S.K. Thompson: *Sampling* (Wiley, New York 1992)

5.22 N. Cressie: *Statistics for Spatial Data* (Wiley, New York 1993)

5.23 J. Grabmeier, A. Rudolph: Techniques of cluster algorithms in data mining, Data Min. Knowl. Discov. **6**, 303–360 (2002)

5.24 L.A. Zadeh: The concept of linguistic variable ant its application to approximate reasoning, Inform. Sci. **8**, 199–249 (1975)

5.25 Z.Y. Wang, G.J. Klir: *Fuzzy Measure Theory* (Plenum, New York 1992)

5.26 L. Polkowski, S. Tsumoto, T.Y. Lin: *Rough Set Methods and Applications* (Physica, Heidelberg 2000)

5.27 Z. Pawlak: *Rough Sets: Theoretical Aspects of Reasoning About Data* (Kluwer, Dordrecht 1991)

5.28 L. Polkowski, A. Skowron: *Rough Sets in Knowledge Discovery 1* (Physica, Heidelberg 1998)

5.29 L. Polkowski, A. Skowron: *Rough Sets in Knowledge Discovery 2* (Physica, Heidelberg 1998)

5.30 Y.Y. Yao, S.K.M. Wong, T.Y. Lin: A review of rough set models. In: *Rough Sets and Data Mining Analysis for Imprecise Data*, ed. by Y. Lin, N. Cercone (Kluwer, London 1997) pp. 47–75

5.31 D.Y. Li, Y. Du: *Artificial Intelligence with Uncertainty* (National Defense Industry Press, Beijing 2005)

5.32 D.L. Olson, D. Dursun: *Advanced Data Mining Techniques* (Springer, Berlin Heidelberg 2008)

5.33 K.C. Di: *Spatial Data Mining and Knowledge Discovery* (Wuhan Univ. Press, Wuhan 2001)

5.34 D.R. Li, S.L. Wang, D.Y. Li, X.Z. Wang: Theories and techniques of spatial data mining and knowledge discovery, Geomat. Inf. Sci. Wuhan Univ. **27**(3), 221–233 (2002)

5.35 D.T. Larose: *Data Mining Methods and Models* (Wiley, New York 2006)

5.36 T. Bayes: An essay toward solving a problem in the doctrine of chances, Philos. Trans. R. Soc. Lond. **53**, 370–418 (1764)

5.37 J. Stutz, P. Cheeseman: *A Short Exposition on Bayesian Inference and Probability* (NASA Ames Research Centre, Data Learning Group, Moffett Field 1994)

5.38 J. James: *Bayes' Theorem, Stanford Encyclopedia of Philosophy* (Metaphysics Res. Lab, Stanford 2003)

5.39 N. Friedman, D. Geiger, M. Goldszmidt: Bayesian network classifiers, Mach. Learn. **29**, 131–163 (1997)

5.40 Daryle Niedermayer I.S.P.: An introduction to Bayesian networks and their contemporary applications, *Innovations in Bayesian Networks* (Springer, Berlin Heidelberg 2008) pp. 117–130

5.41 N. Friedman, M. Goldszmidt: *Learning Bayesian Network from Data* (SRI International, Menlo Park 1998)

5.42 D. Heckerman, D. Geiger: *Learning with Bayesian Networks*, Tech. Rep. MSR-TR-95-06 (Microsoft Research, Redmond 1995) available at http://research.microsoft.com/apps/pubs/default.aspx?id=69588

5.43 D. Heckerman: Bayesian networks for data mining, Data Min. Knowl. Discov. **1**, 79–119 (1997)

5.44 S.L. Wang, X.Z. Wang: *A Fuzzy Comprehensive Clustering Method ADMA 2007*, Lecture Notes in Artifical Intelligence, Vol. 4632 (Springer, Berlin Heidelberg 2007) pp. 488–499

5.45 R.L. Winkler: *An Introduction to Bayesian Inference and Decision* (Holt Rinehart Winston, Toronto 1972)

5.46 S.L. Wang, H.N. Yuan, G. Chen, D.R. Li, W.Z. Shi: Rough spatial interpretation, Lect. Notes Artif. Int. **3066**, 435–444 (2004)

5.47 S. Shekar, C.T. Lu, P. Zhang: A unified approach to detecting spatial outliers, GeoInformatica **7**(2), 139–166 (2003)

5.48 D. Hawkins: *Identifications of Outliers* (Chapman Hall, London 1980)

5.49 P. Rousseeuw, A. Leroy: *Robust Regression and Outlier Detection*, 3rd edn. (Wiley, New York 1996)

5.50 V.J. Hodge, J. Austin: A survey of outlier detection methodologies, Artif. Int. Rev. **22**, 85–126 (2004)

5.51 S. Ramaswamy, R. Rastogi, K. Shim: Efficient algorithms for mining outliers from large datasets, Proc. ACM SIGMOD Conf. Manag. Data, Dallas (2000) pp. 427–438

5.52 A. Jøsang, R. Ismail, C. Boyd: A survey of trust and reputation systems for online service provision, Decis. Support Syst. **43**, 618–644 (2007)

5.53 D.J. Power: *Decision Support Systems: Concepts and Resources for Managers* (Quorum, Westport 2002)

5.54 The Internet of Things Council: http://www.theinternetofthings.eu/ (last accessed July 1, 2010)

5.55 B. Liu: *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data* (Springer, Berlin Heidelberg 2007)

5.56 Z. Markov, D.T. Larose: *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage* (Wiley, New York 2007)

5.57 A. Barabási, E. Bonabeau: Scale-free networks, Sci. Am. **288**, 60–69 (2003)

5.58 D.J. Watts, S.H. Strogatz: Collective dynamics of small world networks, Nature **393**, 440–442 (1998)

5.59 J. Srivastava, R. Cooleyz, M. Deshpande, P.-N. Tan: Web usage mining, ACM SIGKDD Explor. **1**(2), 12–23 (2000)

5.60 D.R. Li, Z.Q. Guan: *Integration and Realization of Spatial Information System* (Wuhan Univ. Press, Wuhan 2002)

5.61 R. Haining: *Spatial Data Analysis: Theory and Practice* (Cambridge Univ. Press, Cambridge 2003)

5.62 H.J. Miller, J. Han: *Geographic Data Mining and Knowledge Discovery*, 2nd edn. (CRC, Boca Raton 2009)

5.63 D.R. Li, S.L. Wang, W.Z. Shi, X.Z. Wang: On spatial data mining and knowledge discovery (SDMKD), Geomat. Infor. Sci. Wuhan Univ. **26**(6), 491–499 (2001)

5.64 F. Giannotti, D. Pedreschi: *Mobility, Data Mining: Geographic Knowledge Discovery* (Springer, Berlin Heidelberg 2008)