CO-012

CASED-BASED REASONING FOR INTELLIGENT THEMATIC CARTOGRAPHY SYSTEM

LI H., HAN J., WANG Y., LUO B., YU Z.
Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science, BEIJING, CHINA

1. BACKGROUND AND OBJECTIVES

Several important concepts are presented one after another in the field of Cartography and GIS (geography information system) recently. The idea of “Citizens as Voluntary Sensors” describes the cooperative production of geo-spatial information and sharing of geo-spatial knowledge in a more persuasive way (Goodchild 2007). The trend towards the "Wikification of GIS" and geography without geographers (Daniel Z. Sui 2008) lets everybody participate in the developing of the four components of GIS including geo-spatial data, software, hardware and people. It also changes geo-spatial applications and even basic GIScience questions dramatically. With the arrival of "the New Era of Geography" (Li D R and Shao Z F 2009), serviced users are extended from professionals to all public users and geospatial data are measurement-on-demand through smart sensor web. Google Earth/Maps is changing the way we see the world (Ratliff E. 2007). A volunteer army of amateur cartographers is collaborated in a sprawling, networked geoweb atlas. Mapmaking power is within the grasp of a 12-year-old with support of tools from Google, Microsoft, and Yahoo nowadays. The map designers usually are the same person as the map readers. Wikification of map i.e. making and sharing maps collaboratively has been gradually become a trend in the field of cartography.

Thematic map design and compilation is a knowledge-based decision-making process which needs lots of map-making experience and creativity of cartographers. But it is difficult to make excellent or relative good thematic maps for users if the GIS or cartography software can not provide effective decision support tools for map design. Thus it is important and necessary to acquire the map-making knowledge and to guide user’s design or provide intelligent recommendations even solutions automatically during map design process.

Knowledge based thematic cartography system is defined as computer based information system that represents the knowledge of cartographer experts and manipulates the knowledge to solve complex map design problems at an expert’s level of performance. Rule-based reasoning (RBR) and case-based reasoning (CBR) are the two fundamental and complementary reasoning methods in the knowledge-based information systems. RBR depends heavily on the rule base consisting of a set of production rules of the form: IF A, THEN B. CBR is a problem solving paradigm that utilizes the specific knowledge of previously experienced situations, called cases.

Since 1990s RBR approaches are already used in computer-assisted design of map symbols (Müller, J.C. and Wang Zeshen 1990), automatic configuration of map labels (Phefferkorn, C. et. al. 1985; Steven Z. 1991), automatic selection of map projections (Nyerges Timothy L. and Jankowski P. 1989; Khalid A. Eldrandaly 2006) and selection of representation methods of thematic features (Guo Q.S. and Zhou J.S. 2004; Tian J., Huang R.T. and Guo Q.S. 2007) etc.. Because of the difficulties in knowledge acquisition and formulization, the prototype systems are still in the laboratory and not used widely. Meanwhile previous study on RBR approaches lacks an integrative solution throughout the thematic mapping process.

Cased-Based Reasoning (CBR) has become a successful technique for knowledge-based systems in many domains such as medical science (Rainer Schmidt etc. 2001), bioinformatics, law (Edwina L. Rissland, Kevin D. Ashley and L. Karl Branting 2005) and so on. There are already some mature CBR systems such as MaSTerClass for classification of biomedical terms (Irena S., Sophia A. and Junichi T. 2005) and CBR development tools such as KATE from French company AcknoSoft. There are also attempts to use case based reasoning in automatic cartography generalization (Keller S. 1994) and geospatial modeling (A. Holt and G. L. Benwell. 1999) in the field of cartography and GIS. Few researches have been conducted on applying the complete CBR cycle to thematic mapping.

After comparing the appropriateness of CBR and RBR, a new framework of intelligent thematic mapping using CBR approach are presented in this paper. The thematic map design process is divided into several sequential steps. Case bases and similarity measures are set up. The complete CBR cycle (case retrieval, reuse, revise, retain) is applied in each step for decision making. Application examples on intelligent selection of representation methods of statistical maps are emphasized and show the potential of CBR approach in thematic mapping.

The article is organized into four sections. After a brief introduction on background and objectives, section 2 presents the CBR conceptual framework in our study, CBR design in selection of representation methods of statistic map and design of CBR-based intelligent thematic cartography prototype system. The results are described in section 3; Section 4 presents conclusions and future research plan.

2. APPROACH & METHODS

2.1 CBR approach to intelligent thematic cartography system

2.1.1 Comparison of CBR and RBR approach

Either the CBR or RBR approach to intelligent thematic cartography system has its own strengths and weaknesses. Thematic map design is complex and creative and needs professional knowledge to achieve a certain degree of intelligence in each step. The knowledge structure in rule base is easily understood and handled. The main impasses of RBR approach for intelligent thematic mapping are the knowledge acquisition and formulization as well as inefficiency of reasoning. Case-Based Reasoning as a model of human memory and remembering can avoid many knowledge acquisition and reasoning problems. But accurately concept expressing, little tolerance of noises and deficiency of case revision mechanism are shortcomings of CBR approach.

2.1.2 Conceptual framework for CBR-based intelligent thematic cartography

First and foremost, this paper seeks to consolidate a conceptual framework for CBR-based intelligent thematic cartography, which directed the future research. Figure 1 shows the designed conceptual framework in the study. The framework is composed of thematic map case base, CBR life cycle of each case base, case retrieval engine, case adaptation module and case learning mechanism. Similarity measures are essential in case retrieval engine and should be well designed for every kind of case.

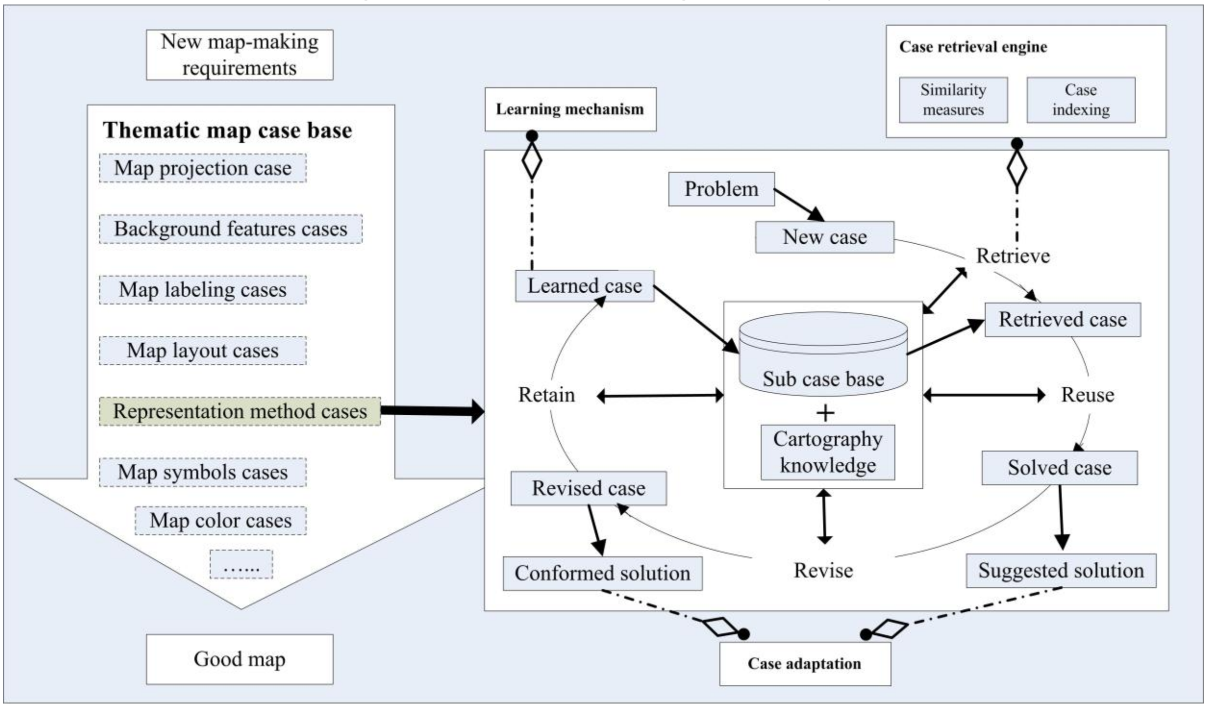


Fig. 1 framework design of CBR-based intelligent thematic cartography

Thematic map-making process consists of sequential map design tasks including selecting and styling background geographical features, determination of theme details, designing or selecting map projection, map labeling design, map layout design, selection of representation methods, map symbol design and color configuration. Design of map sheet system and similar other task is sometimes considered. It is difficult for us to construct one single case that support reasoning from one map. The divide-and-conquer strategy has been adopted. The thematic map design is divided into several sub-tasks and the case bases of each sub-task are constructed. The partition granularity depends on whether the sub-task is easily described and compared with designed similarity measures. The corresponding local similarity measures of respective attribute should be clearly defined at the same time. The formulas of global similarity measures are also provided for each sub design task. Thus, when there are new mapping requirements, the CBR-based intelligent thematic cartography system can provide integrative solution throughout the thematic mapping process based on reasoning capabilities of cases in each sub-task. Good maps will be generated under the automatic recommendations and interactive case adaptation.

Construction of case base is prerequisite for knowledge mining from existing atlas. It tries to cover different design tasks during the entire thematic map design process. Thus the thematic map case base is composed of map projection cases, background feature cases, map labeling cases, map layout cases, representation method cases, map symbol cases and map color cases etc.. For examples, a map from the Population Atlas of China titled as “the proportion of population whose educational level is above the primary school” can be divided into several design segments. There are big rivers, big lakes, national border lines and provincial border lines in background feature cases. The background feature cases should also include which feature in certain layers are selected and the geometry details according to the map scales. For instances, there are only three big rivers (the Yellow River, the Yangtze River and the Pearl River) in the river layer. The main subject of this map is education level of population in China. Map projection cases contain the projection name” Albers” and relative projection parameters. There is description about the inset map of South Sea Islands in map layout cases. The representation method of the map is choropleth map. Map color cases record the gradual change from orange to faint yellow on the map.

The problem-solving life cycle of a CBR system consists of essentially four prime components: retrieve, reuse, revise, and retain. The right part of Figure 1 is the CBR life cycle for cartography which is modified from (A. Aamodt and E. Plaza. 1994). The CBR-based thematic cartography system describes the old experiences, published maps and the design process as cases. Each case is composed of problem specification and corresponding solution. When the new map-making requirements arrives, the system retrieves past cases that are similar to the current one through the case retrieval engine. Then the system reuses the cases by copying or integrating the solutions from the cases retrieved or revising past successful solution(s) in an attempt to solve the new problem by case adaptation module. The current case can be retained as learned case and put into a case base. The case base enables one to store knowledge of the thematic map design, and can be continuously and easily upgraded through the addition of new cases and (possibly) the deletion of old ones that proved to be out of date according to the learning mechanism.

2.1.3 Construction of case base

Construction of initial case base is a prerequisite in case-based reasoning. Published paper atlas, electronic atlas as well as thematic maps on the web provides plenty of recognized thematic map cases. Case engineering (D. Aha and L. Breslow 1997) is a useful way to determine the essential information of a thematic map case, to define case representation method, and to extract cases from available data. Electronic atlases with database support can do great help feeding cases through case generating method while manual input work is needed for paper atlases. In addition, case learning mechanism can enrich the case base during the running periods of the CBR-based cartography system.

2.2 CBR design in Selection of Representation Methods of Statistic Map

Cases in a case base can represent many different types of knowledge that can be stored in many different representational formats. Generally speaking, cases are usually represented as a triple group with problem description, solution description, and effect description. Case base on representation methods of statistical maps is

expressed as $CB = \{C_1, C_2, ..., C_n\}$. $C_i = C(F_i, S_i, V)$ is the ith case in the case base. $F_i = \{f_{i1}, f_{i2}, ..., f_{im}\}$ is the feature set of case C_i with attribute–value pairs. f_{im} is the mth feature of case C_i. $S_i = \{s_{i1}, s_{i2}, ..., s_{ik}\}$ is the solution set of case C_i. s_{ik} is the kth solution. V is the comprehensive evaluation score of map quality by experts.

2.2.1 Database schema design of representation method case

The well designed database schema is very important for passing old cartographers’ knowledge on paper atlas to new users of intelligent thematic cartography system. The case base schema has been designed and listed in Tab 1.

Tab1. Schema design of case base of representation methods on statistics map

Table name	Table fields
Map	MapID， Name， RepID1， RepID2， RepID3， GraphID, LegendPath， MapPath
Represention	RepID， FeatureId, SID, Value
Feature	FeatureId, Dimension, MeasuringScale, Characteristics, NumberOfIndex, NumberOfGroup, NumberOfSerial, Backup
Solution	SID, MID, ParamID, LegendPath， MapPath
Method	MID， Name, Desc, Range, AttributeList， AttributeTypeList， Backup
parameter tables	Each representation method has its own table recording the parameters.

Table name and their fields are introduced as follows. Table named Map is a table of map description including the IDs of different representation method, map name and map/legend image path. In order to visualize multiple aspects of a subject, there are usually several representation methods in the same map. Matching principles of two or more representation methods are very important and should be integrated into thematic map case base. Three fields of representation method identification reflect the fact that there may be three representation methods on one map at the most. Each record in Representation table is a representation method case. Solution table stores the index of representation method, parameter and map/legend image path.

The problem specification consists of a set of attributes and values which should be sufficient to predict a solution for that case. Problem of selection of representation method of thematic maps are specified in Feature table. The attributes of problem space can be divided into two categories. One is integer description such as number of index, group and serial. The other is symbol description such as dimension, measure and characteristics of thematic data. Attributes in the former category is straightforward. Attributes and their values in the latter category are discussed here. The dimension of thematic data reflects the spatial characteristics, one important factor for method selection of thematic maps. The values of dimension attribute are “point”, “line”, “area” and “volume”, similar with traditional GIS layer. Representation method of statistical maps and characteristics of statistical data have close relationship. The main characteristics of statistical data are the scales of measurement: “nominal”, “ordinal”, “interval” and “ratio”. When visualizing more than one group of data, the overall characteristic, proportion of each groups, contrast and dynamics are taken into account. This can be represented by value “quantity”, “class”, “composition”, “sum”, “contrast”, “sequence”, “percent” and their combination (e.g. “percent & contrast”).

Method table and corresponding parameters tables are complementary. Method table stores the ID, name, description, applicability, attribute list, attribute type list and so on. The two fields AttributeList and AttributeTypeList are stored as string text and can be resolved to acquire the attribute information of the method. They define the parameters used in different representation methods of statistical maps including the subtype of certain method. Each representation method is designed as one or more tables. For examples, the cartodiagram method has been divided into nine main types with 36 hypo-types (Yu Zhuoyuan etc.. 2005) such as horizontal/vertical bar chart, pie chart, ring chart, fan chart, and pyramid map etc.. Each kind of chart has the stacked version, percent version, three-dimensional version and gradient color version. Parameters of each type of chart should be stored in independent table.

2.2.2 Case retrieval and similarity measures

Case retrieval is the process of finding, within the case base, those cases that are the closest to the current case with the help of indexing procedures. Retrieval methods developed by researchers and implementers are extremely diverse. The simple nearest-neighbor search has been employed for case retrieval. In nearest-neighbor retrieval, the case retrieved is chosen when the global similarity is greater than other cases in the case base.

A similarity metric that allows closeness among the representation method cases to be measured has been developed. Local similarity measures of attributes in representation method cases are based on the data type such as numerical (int, float, double), single value and multi-value symbolic (bool, string).

$$sim(a,b) = 1 - \frac{|a-b|}{range} \quad (1)$$

$$sim(a,b) = \begin{cases} 1, a = b; \\ 0, a \neq b; \end{cases} \quad (2)$$

$$sim(a,b) = \frac{card(a \cup b) - card(a \cap b)}{card(a \cup b)} \quad (3)$$

$$sim(A,B) = [\frac{1}{\sum_{i=1}^{p} w_i} \sum_{i=1}^{p} w_i [sim(ai,bi)]^2]^{1/2} \quad (4)$$

Where range is the difference between the maximum and the minimum; *card* is the cardinality of the value set; a and b is the attribute value of Case A and Case B. The number of attribute of each case is p. w_i is the weight of the ith attribute. $sim(a_i, b_i)$ is the local similarity measures of the ith attribute ranging from 0 to 1. $sim(A, B)$ is global similarity measure and is weighted sum of its local similarity measures.

Similarity measures of number of index, group and serial are calculated by formula 1 while symbol description such as dimension, measure and characteristics of thematic data by formula 2 and 3. The determination of weights of attributes is very important for global similarity measure. In this study, the weights of dimension and measure scale are 0.2 while weight of characteristics of thematic data is 0.1. The weights of number of index and group are 0.2 while the weight of number of serial is 0.1.

2.3 Design of CBR-based intelligent thematic cartography prototype system

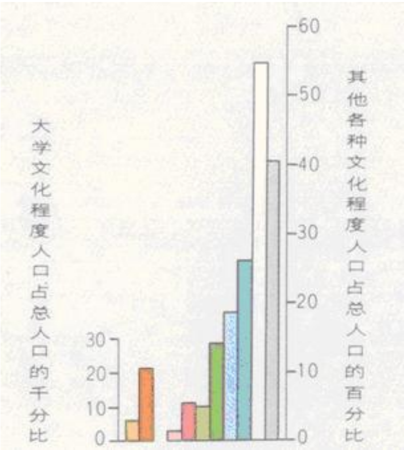
Design of CBR-based intelligent thematic cartography prototype system (CBR-ITCS) has been studied. The basic technical framework is to enrich the web-based thematic cartography system with case based reasoning capabilities. Several open source case-based reasoning tools including myCBR plug-in for Protégé (Thomas R. Roth-Berghofer and Daniel Bahls 2008) and jCOLIBRI (Juan Antonio Recio García 2008) are reviewed for use in CBR-ITCS. The myCBR tools features comfortable graphical user interfaces for modeling various kinds of attribute-specific similarity measures and for evaluating the resulting retrieval quality. It is used as similarity-based retrieval engine in the case design and reasoning phase.

The CBR-ITCS intends to provide recommendations and ranking of different solutions in different design tasks throughout the entire thematic map design process for novice map-making users. Preliminary implementation has integrated the selection of representation method of statistical maps into a CBR cycle.

3. RESULTS

Representation method case base of statistical maps is constructed from the Chinese version of the Population Atlas of China (1987). English version of the atlas was published by Oxford University Press in the same year. The atlas is a comprehensive, systematic portrayal of the third census in the July 1982. The Atlas's 137 full-color maps, clear explanatory text, and 32 pages of statistical tables summarize and make readily available for the first time a wealth of authoritative information of population distribution and change; nationalities; age and sex; marriage and fertility; education levels; occupations; and more. It received CAS Science & Technology Achievement Award (first class) and National Science & Technology Progress Award (second class) in China. There are one hundred and eight case bases stored in the case base. These cases come from the 6 groups of maps of the atlas with the help of myCBR software.

One example is presented as evidence for the success of case based reasoning approach. Selection of representation methods of population migration data in statistical chart map-making has been studied. When designing a map on population migration by province in China, there are 106 attribute data from 1954 to 2006 grouped by two indices (immigration rate and outmigration rate). Based on analysis of data characteristics and case retrieval from representation method case base, two representation methods with high similarity value are recommended. One of them is grouped horizontal bar method which has a similarity value 0.83 (see map legend in Fig.2. a). Another one is pyramids method which has a similarity value 0.94 (see map legend in Fig.2. b). Thus the pyramid method of population migration is used as solution of representation method of the new case (see Fig.3).



2.a) grouped vertical bar 2.b) pyramid

Figure 2. Legend images in case base

Figure 3. Draft pyramid map of population migration

4. CONCLUSION AND FUTURE PLANS

4.1 Conclusions

In this paper this paper proposes a new concept framework of knowledge-based intelligent thematic cartography system by introducing the CBR methodology. The framework divides the map design task into smaller, more manageable pieces, and then takes control of those pieces one by one using the divide-and-conquer strategy. Map design tasks throughout thematic mapping process including background features selection, map symbols design, representation methods design, and color design etc. will be easily completed with the help of cases from existing maps. An example presented here indicates that case-based reasoning approach has several advantages and is expected to do better than rule-based reasoning approach.

The idea of discovering cartography knowledge from existing atlases is the key to the success of case-based reasoning approach. The capabilities and qualities of problem solving in map design depend heavily on the completeness, volume of thematic map case base as well as the reasoning engines. Although the case retaining strategy of CBR cycle provides incremental mechanism, feeding more cases of different map design tasks into database is a long term process. Retrieving most similar cases is still a challenge because of the difficulties in quantifying the similarity measures in different map design tasks and capturing authentic user requirements.

The contributions of this study are two-folds. Firstly, the introduction of case-based reasoning approaches diversifies the methods and sheds new light on expert systems of intelligent thematic cartography. One example given in the paper provides new algorithm of comparing two thematic maps by similarity measure modeling. Secondly, the thematic map case base is firstly constructed from existing atlases. The cartography knowledge contained in excellent maps can be acquired, formulized and stored in the case base for reference and decision support for novice map-makers. Like the atlas itself, the thematic map case is of great scientific significance and important for practical applications in cartography.

4.2 Open problem and future plans

There are still many open problems on case based reasoning approach to intelligent thematic cartography system to be addressed. The following will briefly outline some of them and the corresponding future plans.

Firstly, the workload of case engineering is very heavy. The size of case database on selection of representation methods is still very small. Enriching the case database with more cases on selection of representation methods and more cases in other map design tasks is a priority. Several social-economical published atlases will be used as chief source to enrich the case base for knowledge-based thematic cartography. They are the Economic Atlas of China (1990), the Population, Environment and Sustainable Development Atlas of China (2005), and the Atlas on Changing of Population and Environment (2008). Automatic methods of case engineering based on electronic atlas and database technology need further research.

Secondly, similarity measures are essential for case retrieval. Design of database schema and similarity measures on cases of other map design tasks need in-depth analysis. Other approaches to similarity measures such as semantic based measures should be integrated.

Thirdly, RBR and CBR are proved to be complementary to a large degree. We plan to explore different approaches integrating RBR and CBR for better reasoning results to improve intelligence in thematic map-making.

Fourthly, augmenting classic RBR and CBR with domain ontology may resolve possible impasses in the reasoning cycle. In particular, ontology could be used for capturing the new thematic mapping requirements, similarity measures and decision support during map-making.

Lastly, the integration of case-based reasoning and thematic cartography system needs intensive development for web-based solution to intelligent thematic cartography. With the support of the proposed prototype system, experienced cartographer can make more excellent maps enriching the case database; novice map-makers can learn basic cartography knowledge during map design process. As the maturation of CBR-based intelligent thematic cartography prototype system it will play an important role to fulfill the great vision "to enable everybody to be a professional cartographer" and to motivate the advent of the "Wikification of Maps".

References

A. Aamodt and E. Plaza. 1994. "Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches." *Artificial Intelligence Communications. 7*: 39- 59.

A. Holt and G. L. Benwell. 1999. "Applying case-based reasoning techniques in GIS." *International Journal of Geographical Information Science*. 13(1): 9-25.

C. Pfefferkorn, D. Burr, D. Harrison, B. Heckman, C. Oresky, and J. Rothermel. 1985. "ACES: A Cartographic Expert System." *Proceedings of Auto-Carto 7*. Washington DC. 399-407.

D. Aha and L. Breslow. 1997. "Refining conversational case libraries." *Proceedings of the Second International Conference on Case-Based Reasoning*. Berlin: Springer Verlag. 267-278.

Edwina L. Rissland, Kevin D. Ashley and L. Karl Branting. 2005. “Case-based Reasoning and Law”. *The Knowledge Engineering Review*. 20: 293-298.

Juan Antonio Recio García. 2008. “jCOLIBRI: A Multi-level Platform for Building and Generating CBR Systems.” PhD thesis, Universidad Complutense de Madrid, Madrid.

Keller S. 1994. “On the Use of Case-based Reasoning in Generalization.” *Proc. of 6th International Symposium on Spatial Data Handling*. Edinburgh: Taylor and Francis, 2: 1118-1132.

Goodchild M F. 2007. “Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2. 0.” *International Journal of Spatial Data Infrastructures Research*. 2: 24–32.

Guo Q. S. and Zhou J. S. 2004. “Reasoning about Choosing Representations of Thematic Map.” *Journal of Geomatics* (in Chinese). 29 (2): 31-33.

Irena S. , Sophia A. and Junichi T. 2005. “MaSTerClass: a Case-based Reasoning System for the Classification of Biomedical Terms.” *Bioinformatics*. 21(11): 2748-2758.

Khalid A. Eldrandaly. 2006. “A COM-based Expert System for Selecting the Suitable Map Projection in ArcGIS.” *Expert Systems with Applications*. 31(1): 94-100.

Li D R, Shao Z F. 2009. “The New Era for Geo-information.” *Science China Serial F - Information Science*. 52(7): 1233–1242.

Müller, J. C. , Wang Zeshen. 1990. “A Knowledge Based System for Cartographic Symbol Design.” *The Cartographic Journal*. 27: 24-30.

Nyerges Timothy L. , Jankowski Piotr. 1989. “A Knowledge Base for Map Projection Selection.” *Cartography and Geographic Information Science*. 16(1): 29-38.

Population Census Office of the State Council of the People's Republic of China and the Institute of Geography of the Chinese Academy of Sciences. 1987. *The Population Atlas of China* (in Chinese). Beijing: China Statistics Press. (Note: English version is published by Oxford University Press.)

Rainer Schmidt, Stefania Montani, Riccardo Bellazzi, Luigi Portinale and Lothar Gierl. 2001. “Cased-Based Reasoning for Medical Knowledge-based Systems.” *International Journal of Medical Informatics*. 64: 355–367.

Ratcliff E. 2007. Google Maps is changing the way we see the world. *Wired Magazine*. 15. Available at www. wired. com/techbiz/it/magazine/15-07/ff_maps.

Steven Zoraster. 1991. “Expert Systems and the Map Label Placement Problem.” *Cartographica*. 28(1):1-9.

Thomas R. Roth-Berghofer and Daniel Bahls. 2008. “Explanation Capabilities of the Open Source Case-Based Reasoning Tool myCBR.” *Thirteenth UK Workshop on Case Based Reasoning*. Cambridge, UK, 9th December.

Tian J. , Huang R. T. and Guo Q. S. 2007. “Study on intelligent choice o f representation methods in thematic map.” *Science of Surveying and Mapping* (in Chinese). 32 (5): 170-172.

Yu Zhuoyuan, Wang Yingjie, Suying, Tan Yuqi and Hong Wei. 2005. “System Design and Construction of Digital Atlas of Population of China.” *Geo-information Science*(in Chinese). 7(2): 23-29.