Research on Spatial Conceptual Model Based on Natural Language Processing

Liu Xiaobo

School of Information Science and Technology
Nanjing Xiaozhuang University
Nanjing, China
e-mail: 4487758@qq.com

***Abstract* — The article studied spatial ontology library construction technology based on ontology and theory of machine learning. Specific content contained an instance method of merging ontology of space-based semi-automatic spatial relation extraction method, spatial ontology library based on linear classifier, method of object space based on genetic algorithm. On the basis of these studies, the article constructed a small space-based ontology modeling of natural language spatial concept prototype and gave visual evaluation system based on t test. The article also verified the validity of conceptual modeling based on natural language.**

Keywords - Ontology; Spatial Relationships; Information Extraction; Conceptual Model; Natural Language

I. INTRODUCTION

It is well known that natural language is a lingua franca, such as Chinese, Japanese, English, and so on. They are tools for people learning and communication (Satheesh S K. et al, 2013). In the course of human development, most of knowledge is documented and circulated in the form of preserved rely on the language which is accounted for more than 80% of current human knowledge (Gunn A. et al, 2012). For the application of computer, its function for mathematics calculation only accounted for 10% and accounted not to 5% for process control. The remaining of 85% is for currently language information processing, mainly including natural language processing of sound, shaped and meaning. That is the text entry generally (M. Ma. et al, 2004).

Natural language processing is actually a kind of natural language understanding. Natural language processing is to allow the computer to really understand our language (H. Tanaka, T. et al, 2002). For example, if we tell the computer that we want to find information which requires the computer to really understand what we need. Understanding of computers is to make sure that the machine is how to understand our language. After so years of development course, most scientists use AI technology and then express out natural language with computer program to constitute a natural language of system (A. Mukerjee, K. et al, 2000). It is to put on computer of output and entered compared as a reaction computer understanding of degree of a standard in th function description of the system. If the output and entered is almost same, this system is high in efficiency, otherwise this system still needs to improve and be more perfect (Mott D. et al, 2012).

Mylopoulos defined the conceptual modeling as the formal description of activities for the objective and certain aspects of the social world around us (Parnas D L, 1972). The purpose of which is to provide understanding and exchanges in the field. Conceptual modeling plays an important role in various fields of computer science including information systems design, artificial intelligence, knowledge representation, organization and environment modeling, business processing, software development, software requirements and son on (Moody D L. et al, 2003).

Conceptual modeling originated early in the 1970 of the 20th century (Ross D, 1977). Database technical development has laid a solid foundation for concept modeling. The study of parnas provide accurate and complete instructions of data abstraction. He made the software fragments to interact and proposed the concept of data abstraction, then hidden the implementation details from the user's perspective (S. Yusuke. et al, 2000). Most of the database design modeling relies on data structure, and the model is stored in the actual file system. Hierarchical and Network models which are typical focus on the physical dimensions of the model (Preece A. et al, 2012). Codd separated the logical data and physical data and captured more demand for semantics which laid the groundwork for conceptual modeling.

The rest of the paper is organized as follows. In Section 2, spatial conceptual model is summarized briefly. In Section 3, k-nearest neighbor algorithm is described. In Section 4, numerical experiments are presented and the results are discussed. Finally, a conclusion is provided in Section 5.

II. OVERVIEW OF SPATIAL CONCEPTUAL MODEL

Current conceptual modeling systems can be divided into two broad categories: prototype system based on script and prototype system based on natural language. Fixed format phrase and high-level language modeling system is a prototype system based on script. Model based on dialogue or discourse is prototype system based on natural language. Figure 1 shows a combined libraries and machine-learning algorithms based on spatial ontology model of conceptual modeling framework. The left part is spatial ontology acquisition process including ontology concepts, relationships, and other system definitions. The

right part is the spatial ontology containing ontology including hownet semantic knowledge, such as semantic support library, location body, solid body and graphical ontology.

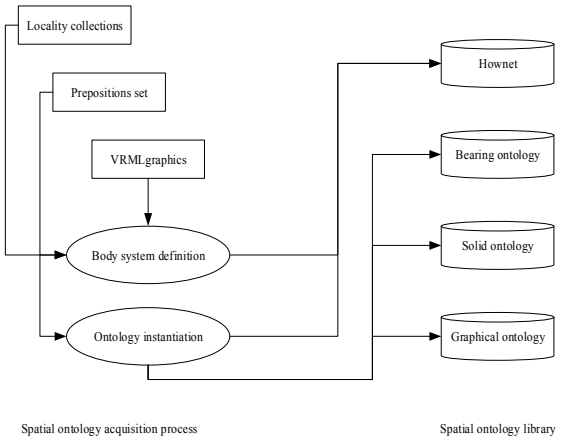

Fig. 1: Framework of Conceptral Model System Based on Spatial Ontology.

Conceptual modeling process is a process of extraction and quantification. Figure 2 shows a piece of text creation corresponds to display scene modeling process, as well as the ontology used in the modeling process. For a given chapter, we truncated to get substring the substring set first, then identified spatial expression border for each substring circular to get the spatial expression borders collection. We recognized landmarks, prepositions and nouns of extraction in each space within the borders to complete extraction of spatial expression and get the spatial expression borders collection. For each space corresponds to the expression, we first recognized its entire trajectory to get the spatial relationships and make the formation of spatial relationship set. For each spatial relationships, we conducted semantic disambiguation of nouns in the first, and then placed shooting range, finally get the location of each trajectory based on genetic algorithm and progress towards it. By looping through each spatial relationship, we can get each points for all entities.

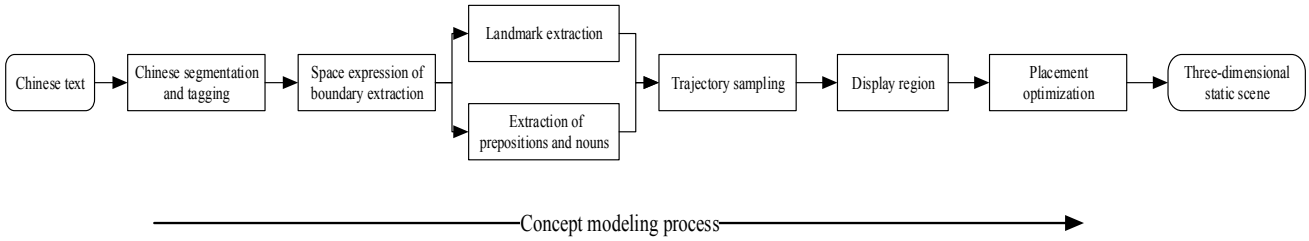

Fig. 2: Concept Modeling Process.

III. NEW K-NEAREST NEIGHBOR ALGORITHM

A. *Definition of algorithm*

K-nearest neighbor (KNN) is a machine learning method based on case study, and it can use for approximate functions of real-valued and discrete objects. It simply stores the initial learning process of training data. If there have new classification examples, the algorithms will take the original training data which is similar to the new instance and classify this as a reference to the new instance. When the objective function is complex and it is not so complex in partial approaches, KNN methods have a significant advantage for machine learning.

KNN algorithm assumes that all instance corresponds to a point in n-dimensional space and the nearest neighbor of an instance is defined according to the standard Euclidean distance. In other words, an instance space eigenvector x is represented as the following:

$$\langle a_1(\mathrm{x}), a_2(\mathrm{x}) \ldots a_n(\mathrm{x}) \rangle \tag{1}$$

$a_r(\mathrm{x})$ represents the r-th attribute values of an instance. The distance between two instances can be defined as follows:

$$d(\mathrm{x}_i, \mathrm{x}_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \tag{2}$$

B. *Algorithm implementation*

K-nearest neighbor algorithms have three steps which are as follows: (1) training algorithms. We should put the training examples $\langle x, f(\mathrm{x}) \rangle$ into the list of training. (2) classification algorithms. For a given instance for unspecifying, we must chose the first k instances in the training list which are the most near to x_q and indicate with $x_1 \ldots x_k$. (3) return the follow value:

$$f(x) \leftarrow \arg \max \sum_{\substack{i=1 \\ v \in V}}^{k} \delta(v, f(x)) \tag{3}$$

If $a=b$, we can get $\delta(a,b)=1$. Otherwise $\delta(a,b)=0$.

C. *Evaluation indicator*

Evaluation indicator contains averages, statistics, two-tailed probability of the statistic and the one-tailed probability evaluation. Evaluation rating refers to the arithmetic mean of rated scenes. The formula is as follows:

$$\overline{x_k}=\frac{\sum_{i=1}^{m}\left(\left(\sum_{j=1}^{n_i}x_{ij}\right)/\mathrm{n}_i\right)}{m} \tag{4}$$

The formula of statistics is as follows:

$$t=\frac{(\overline{x_1}-\overline{x_2})}{S_{\overline{x_1}-\overline{x_2}}} \tag{5}$$

$$S_{\overline{x_1}-\overline{x_2}}=\sqrt{\frac{\sum(x_1-\overline{x_1})+\sum(x_2-\overline{x_2})}{(n_1-1)+(\mathrm{n}_2-1)}\times\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \tag{6}$$

x_{ij} indicates the j-th reviewer scores for the I-th scene. n_i indicates the number of the i-th scenarios. x_k indicates the sample mean of the k-th scene.

IV. COMPUTATIONAL RESULTS AND NUMERICAL EXPERIMENTS

A. *Experimental sample*

We select one material with 3,000 Chinese text to test classification algorithms for the information system and conduct comparative analysis of efficiency and results through experiment.

Most of the texts in the corpus collected from Government newspapers and related media which are the relevant press releases. We classify the releases by experts in various fields to divided into 35 categories according to their experience about social, economic, political, and so on. We need to choose a training set and a testing set firstly. And then data was divided into 20 equal parts which experts have classified before, then select one of them as the opening test set where the rest 19 serve as training sets and closed sets. Each one can become a testing set to run sorting algorithm. And the average for 20 classification operation is calculated. The results as shown in table 1.

From test results in the form above, KNN algorithm in classification effect is more closed to real value and costs the least time in training process. But its cost in classification process is the highest, and this make total time increased which is not conducive to timely text processing.

TABLE. 1: THREE ALGORITHMS COMPARISON

Algorithm	Closed test			Open test		
Indicator	Full rate	Exact value	F1 value	Full rate	Exact value	F1 value
Bayesian	83%	84%	83%	77%	78%	77%
General distances	88%	88%	88%	81%	81%	81%
KNN	90%	92%	91%	84%	86%	86%

B. *Experimental Results*

Under the above definition of feature space, performance indicators in the training set is shown in Figure 3.

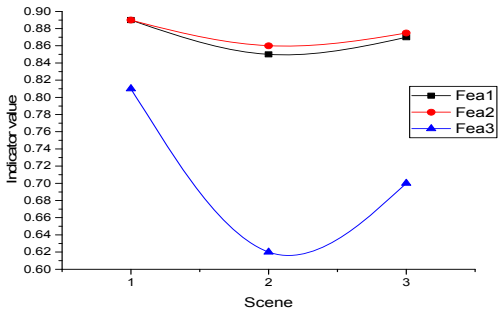

Fig. 3: The Open Results on Three Feature Spaces.

Under a non-speech feature space, overall performance is lowest, particularly recall rates falls fastest which indicates that the speech feature space borders are very important in the expression. However performance between Fea1 and Fea2 has not so much difference which indicates that the semantic contribution is not very big. So we also conduct opening test in the both feature space, the results are shown in Figure 4.

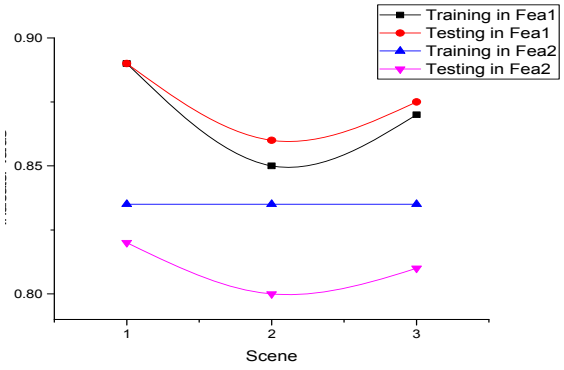

Fig. 4: The Closed and Open Results on the First Two Feature Spaces.

To study the effectiveness of feature, we respectively conduct closing and opening test for landmark extraction, the results are shown in Figure 5. Fea1 and Fea2 have the

better performance than Fea3 which indicates importance of context features.

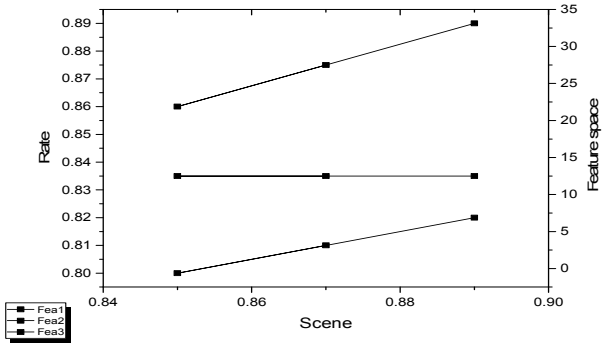

Fig. 5: The Close and Open Tests Results on Winnow or Perceptron.

V. CONCLUSION

In this paper, automatic and effective conceptual modeling methods of natural language description of space is studied. Based on the natural language spatial relations as the research object, a set of space concept modeling based on natural language is conducted. As an important part of the conceptual modeling, the article studied spatial ontology library building and space relation extraction technology. Also the article researched and developed two-step object space. Visual evaluation method based on t tests on prototype system of space conceptual modeling scenarios were evaluated and verified the effectiveness of small-scale spatial ontology library. The proposed space conceptual modeling based on natural language descriptions was feasible and of great value.

It is desirable to further apply KNN algorithm to solving those more complex real-world optimization problems and it will be our further work.

REFERENCES

[1] A. Mukerjee, K. Grupta and S. Nautiyal, 2000. Conceptual description of visual scenes from linguistic model. Image and Vision Computing. 18(2), pp: 173-187.

[2] Gunn A, Poole K G, Nishi J S, 2012. A conceptual model for migratory tundra caribou to explain and predict why shifts in spatial fidelity of breeding cows to their calving grounds are infrequent. Rangifer, 32(2), pp: 259-267.

[3] H. Tanaka, T. Tokunaga and Y. Shinyama, 2002. Animated Agents that Understand Natural Language and Perform Actions. Proceedings of PRICAI-02 Workshop on Liflike Animated Agents. pp: 89-94.

[4] M. Ma and P. Kevitt, 2004. Visual Semantics and Ontology of Eventive Verbs. The First International Joint Conference on Natural Language Processing (IJCNLP-2004). Hainan, China. pp: 187-196.

[5] Mott D, Braines D, Poteet S, et al, 2012. Controlled Natural Language to facilitate information extraction[C]//Proceedings of the Sixth Annual Conference of the International Technology Alliance, London, UK.

[6] Moody D L, Shanks G, 2003. Improving the quality of data models: Empirical validation of a quality management framework. Information Systems, 28(6), pp: 619-650.

[7] Parnas D L, 1972. A technique for software module specification with examples. Communications of the ACM, 15(5), pp: 330-336.

[8] Preece A, Pizzocaro D, Braines D, et al, 2012. Integrating hard and soft information sources for D2D using controlled natural language[C]//Information Fusion (FUSION), 2012 15th International Conference on. IEEE, pp: 1330-1337.

[9] Ross D, 1977. Structured analysis (SA): A language for communicating idea. IEEE Transactions on Software Engineering. SE-3(1), pp: 16-34.

[10] S. Yusuke, T. Takenobu and T. Hozumi. Kairai, 2000. Software Robots Understanding Natural Language. Third International Workshop on Human-Computer Conversation. Bellagio, Italy, pp: 158-163.

[11] Qingyuan Zhou, Jianjian Luo. The Risk Management Using Limit Theory of Statistics on Extremes on the Big Data Era [J]. J. Comput. Theor. Nanosci. 12, 6237-6243 (2015).

[12] Satheesh S K, Moorthy K, 2013. AEROSAT-a space-borne sensor for continental aerosols: evaluation of the conceptual model. Current Science, 104(8), pp: 1008-1011.

[13] Z. Haiming, S. Xiaoxiao, Study on prediction of atmospheric pm2. 5 based on rbf neural network, in: Digital Manufacturing and Automation (ICDMA), 2013 Fourth International Conference on, IEEE, 2013, pp. 1287‑1289.

[14] X. Luo, Z. Xu, J. Yu, and X. Chen. Building Association Link Network for Semantic Link on Web Resources. IEEE transactions on automation science and engineering, 8(3):482-494, 2011.

[15] Qingyuan Zhou, Jianjian Luo. Artificial neural network based grid computing of E-government scheduling for emergency management. Comput Syst Sci & Eng. Vol 30, No 5, p 327-335, September 2015.

[16] Y. Zheng, F. Liu, and H. Hsieh. U-Air: When Urban Air Quality Inference Meets Big Data. In Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1436-1444, 2013.

[17] L. Linderoth, S.J. Wright, Decomposition algorithms for stochastic programming on a computational grid, Computational Optimization and Applications 24 (2003) 207-250.

[18] H. Casanova, J. Dongarra, Netsolve: Network enabled solvers, IEEE Computational Science and Engineering 5 (3) (1998) 57-67.

[19] J. Yuan, Y. Zheng, X. Xie, and G. Sun. 2013. T-Drive: Enhancing driving directions with taxi drivers' intelligence. Transactions on Knowledge and Data Engineering 25, 1, 220–232.

[20] W. C. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. 2010. Collaborative filtering meets mobile recommendation: A user-centered approach. In Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 236–241.