Journal of Computational Information Systems 7: 14 (2011) 5050-5057
Available at http://www.Jofcis.com

Multimedia High-level Semantic Structure and Application

Yuan SUN, Wentao SUN, Gang HOU †, Hexin CHEN

Center for Computer Fundamental Education, Jilin University, Changchun, China,

Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun , China

College of Communication Engineering, Jilin University ,Changchun, China,

Abstract

The paper presents a new structure which organizes high-level semantic data for retrieving multimedia. For solving the problem of semantic gap in retrieving multimedia data, the paper gives the definition and operation laws of matrix division method, which is a mathematical model and describes the high-level semantic structure for many kinds of multimedia data. Based on the structure, the paper gives one specific application which is semantic branch structure to organize the high-level semantic data for retrieving key frames of video. The semantic branch structure is a triple level structure and every level structure has different model configuration. The experimental results show that the effectiveness of the proposed semantic branch structure is good.

Keywords: High-level Semantic; Semantic Branch Structure; Matrix Division Method

1. Introduction

Along with a great deal of digital multimedia data appears on the international network, a lot of researchers devote to retrieving all kinds of multimedia information. However, many retrieval methods are based on low-level features of multimedia data (audio, image, video etc), which is far from the way people well know—based on high-level semantic. It is called "semantic gap". So the region for searching multimedia data based on high-level semantic is research hotspot.

Sang Keun Rhee et al. [1] shows that if there are sufficiently many concepts, even low detection accuracy, the retrieval results is good. Gustavo Carneiro et al. [2] proposed a probabilistic formulation for semantic image annotation and retrieval. Annotation and retrieval are posed as classification problems where they define each class as the group of database images labeled with a common semantic label. Pradhan a.S. et al. [3] gives one relational semantic model, but the relational model is not exactly fit for non-structural data, especially not fit for multimedia data. N. Ruan et al. [4] proposed a framework based on domain-dependent ontology to perform semantic retrieval in image archives. In their framework, ontology is used to provide a sharable and reusable concept set as infrastructure for high level extension. Michael G. Christel et al. [5] shows high-level semantic instance for retrieval video. Chun-Yi Lin et al. [6] proposed a multi-level semantic modeling method, which integrates Support Vector Machines (SVM) into

† Corresponding author: Gang Hou
Email addresses: sy@jlu.edu.cn (Gang Hou).

December, 2011

hybrid Bayesian networks (HBN). Bing Wang et al. [7] put forward a self-organizing image description model for describing the image high-level semantic content. This model is a hierarchical architecture, which includes primitive image layer, image feature layer, image semantic layer, multi-level semantic pattern layer and semantic labeling layer. Lei WANG al. [8] presents a methodology to use object-oriented concept database to offer users' semantic search. A demo system named Intelligent Content Search Engine (ICSE) is developed to validate the proposed recommendation method

This paper is organized as follows. In section 2 we give the definition and operation laws of the matrix division method for high-level semantic structure. Semantic branch structure based on the matrix division method is given in section 3. In section 4, we validate the effectiveness of the semantic branch structure by experiments. The conclusion for the paper is in section 5.

2. Mathematical Model--Matrix Division Method for High-level Semantic Structure

The paper gives the definition and operation laws to describe the high-level semantic structure of multimedia data by using matrix below.

2.1. The Definition of Matrix Division Method

We $C_{M\times N} = A_{J\times K} +\propto B_{P\times Q}$ where $C_{M\times N} = [c_{mn}]_{M\times M}, A_{J\times K} = [a_{jk}]_{J\times K}, B_{P\times Q} = [b_{pq}]_{P\times Q}$,

$$\left[c_{mn}\right]_{M\times N} = \left[a_{jk}\right]_{J\times K} +\propto \left[b_{pq}\right]_{P\times Q} \quad c_{mn} = \begin{cases} a_{in}, (1 \le i \le j, 1 \le n \le K) \\ b_{jn}, (1 \le j \le P, K+1 \le n \le K+Q) \end{cases}$$

Where $+\propto$ is the operator of matrix division method.

$$i = [(m-1)\div P]+1, \; j = (m-1)\%P+1, \; M = J\times P, N = K+Q, \; 1 \le m \le M, 1 \le n \le N \; .$$

The operation laws of matrix division method

1) matrix division method addition associative law.

$$\mathrm{C}_{\mathrm{M}\times\mathrm{N}} = \mathrm{A}_{J\times K} +\propto B_{P\times Q} +\propto X_{RS}$$

Where $C_{M\times N} = [c_{mn}]_{M\times M}, A_{J\times K} = [a_{jk}]_{J\times K}, B_{P\times Q} = [b_{pq}]_{P\times Q}, X_{RS} = [x_{rs}]_{R\times S}$

$$\left[c_{mn}\right]_{M\times N} = (\left[a_{jk}\right]_{J\times K} +\propto \left[b_{pq}\right]_{P\times Q}) +\propto \left[x_{rs}\right]_{R\times S}$$

$$c_{mn} = \begin{cases} a_{in}, i = [(m-1)\div P]+1, M = J\times P, N = K+Q \; (1 \le i \le J, 1 \le m \le J\times P, 1 \le n \le K) \\ b_{jn}, j = (m-1)\%P+1, M = J\times P, N = K+Q \; (1 \le j \le P, 1 \le m \le J\times P, K+1 \le n \le K+Q) \end{cases} +\propto x_{rs} \quad (1)$$

$$= \begin{cases} a_{in}, \; i = (m\text{-}1)\div(P\times R)+1 \; (1 \le i \le J, 1 \le m \le J\times P\times R, 1 \le n \le K) \\ b_{jn}, j = (m\text{-}1)\div R\%P+1 (1 \le j \le P, \; 1 \le m \le J\times P\times R, K+1 \le n \le Q+K) \\ x_{kn}, k = (m\text{-}1)\%R+1 (1 \le k \le R, \; 1 \le m \le J\times P\times R, Q+K+1 \le n \le K+Q+S) \end{cases}$$

where final value of M,N is $M = J\times P\times R, N = K+Q+S$. $C_{M\times N} = A_{J\times K} +\propto (B_{P\times Q} +\propto X_{RS})$.

$$C_{MN} = [c_{mn}]_{M\times N} = A_{JK} + \propto (B_{PQ} + \propto X_{RS})$$
$$= A_{JK} + \propto ([b_{pq}]_{P\times Q} + \propto [x_{rs}]_{R\times S})$$
$$c_{mn} = a_{jk} + \propto \begin{cases} \mathrm{b}_{jn}, j = [(m-1)\div R]+1, M = P\times R, N = Q+S(1\le j\le P, 1\le n\le Q) \\ x_{kn}, k = (m-1)\%R+1, M = P\times R, N = Q+S(1\le k\le R, Q+1\le n\le Q+S) \end{cases} \quad (2)$$
$$= \begin{cases} a_{in},\ i = (m\text{-}1)\div(P\times R)+1\ (1\le i\le J, 1\le n\le K) \\ b_{jn}, j = (m\text{-}1)\div R\%P+1(1\le j\le P,\ \mathrm{K}+1\le n\le K+Q) \\ x_{kn}, k = (m\text{-}1)\%R+1(1\le k\le R,\ Q+K+1\le n\le Q+K+S) \end{cases}$$

Where final value of M,N is $M = J\times P\times R, N = K+Q+S$.

Due to (1)=(2), we can prove matrix division method addition associative law.

2) The definition of matrix division method unit matrix.

$$\mathrm{C}_{\mathrm{M\times N}} + \propto \Phi_{1\times 0} = \mathrm{C}_{\mathrm{M\times N}} \text{,where } \mathrm{C}_{\mathrm{M\times N}} = [c_{mn}]_{M\times N}, \Phi_{1\times 0} = [\ ]_{1\times 0} \text{,namely } [c_{mn}]_{M\times N} + \propto [\ ]_{1\times 0} = [c_{mn}]_{M\times N}$$

$$\Phi_{1\times 0} + \propto \mathrm{C}_{\mathrm{M\times N}} = \mathrm{C}_{\mathrm{M\times N}} \text{,where } \mathrm{C}_{\mathrm{M\times N}} = [c_{mn}]_{M\times N}, \Phi_{1\times 0} = [\ ]_{1\times 0} \text{,namely } [\ ]_{1\times 0} + \propto [c_{mn}]_{M\times N} = [c_{mn}]_{M\times N}$$

Where $\Phi_{1\times 0}, [\,]_{1\times 0}$ is unit matrix.

3) Matrix division method multiplication distributive law.

$$k(A_{J\times K} + \propto B_{P\times Q})$$

Where

$$A_{J\times K} = [a_{jk}]_{J\times K}, B_{P\times Q} = [b_{pq}]_{P\times Q}$$

$$k(a_{jk} + \propto\!\downarrow b_{pq}) = \begin{cases} ka_{in}, i = [(m-1)\div P]+1(1\le i\le J, 1\le n\le K) \\ kb_{jn}, j = (m-1)\%P+1(1\le j\le P, K+1\le n\le K+Q) \end{cases} \quad (3)$$

Where $M = J\times P, N = K+Q, 1\le m\le M, + \propto\!\downarrow$ is element plus.

$kA_{J\times K} + \propto kB_{P\times Q}$ namely $k[a_{jk}] + \propto k[b_{pq}]$,

$$ka_{jk} + \propto\!\downarrow kb_{pq} = \begin{cases} ka_{in}, i = [(m-1)\div P]+1, (1\le i\le J, 1\le n\le K) \\ kb_{jn}, j = (m-1)\%P+1, (1\le j\le P, K+1\le n\le K+Q) \end{cases} \quad (4)$$

Where $M = J\times P, n = K+Q$, $1\le m\le M$. Because (3)=(4), namely $k(A_{J\times K} + \propto B_{P\times Q}) = kA_{J\times K} + \propto kB_{P\times Q}$.

We can prove matrix division method distributive law and describe the structure as Figure 1 based on the matrix division method.

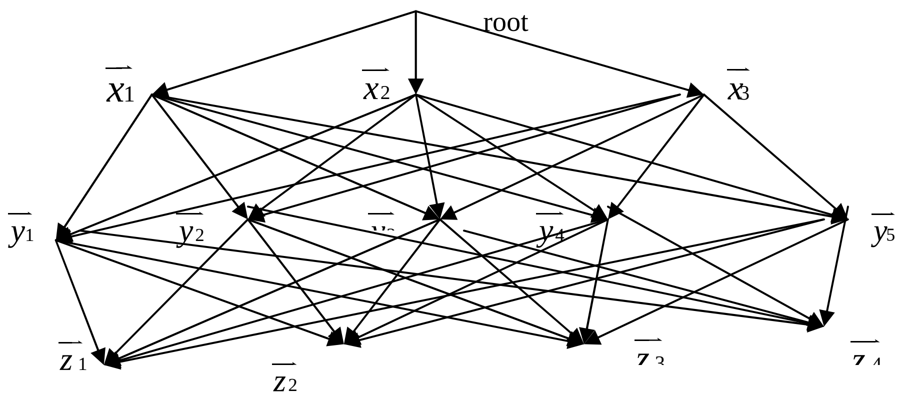

Fig.1 The Structure Described by Matrix Division Method

2.2. Semantic Branch Structure

The semantic vector is to describe the concepts contained in key frames. It can not contain the relationship among those semantic concepts, nor does it derive the high level semantics implied in the key frame. Based on the matrix division method, we give the semantic branch structure to describe the semantic model for the related key frames on the high-level semantic data; Figure 2 shows the semantic branch structure.

In semantic branch structure, the top level is the semantic class level, which describes the content of some semantic class by particular model parameters, it is the abstract level.

The middle level is the semantic module level, whose contents are more concrete semantics contained in a key frame than the top level. When we analyze a key frame by high-level semantics, we often divide the contents of a key frame into two parts: background and foreground. The background of a key frame is taken as environment, which contains a lot of basic semantic concepts; the foreground of a key frame contains objects. Maybe there are many objects in the foreground. However, it is not sufficient to represent key frame semantics by dividing a key frame into objects and environment simply. Semantics of key frames also include the relationships among objects, environment and as well as the relationship among objects and environment. Therefore, in semantic branch structure, we define the behavior semantic module to describe the relationships above.

The bottom level is the semantic concept level, which contains the concepts appeared in one key frame. It is an entity in semantic vector space. In this semantic branch structure, the higher semantic content can be derived from the lower semantic content by some operational rules and the lower semantic level supports the higher semantic level.

2.3. A Semantic Branch Structure for Conversation Category

Based on the semantic branch structure introduced in section above, we give an instance based on figure 2 as shown in figure 3.

According to semantic branch structure, the contents of interview category for key frames in the news video can be divided into two parts: environment and characters. Environment include outdoor or indoor, but mostly are indoor. Characters usually include several national leaders, officials of some organizations or government officials. However, it is not sufficient to represent key frames only by environment and characters. It should contain the interaction among characters. In interview category of key frames in video, the interaction of characters should contain behavior, such as handshaking, eye contact and so on.

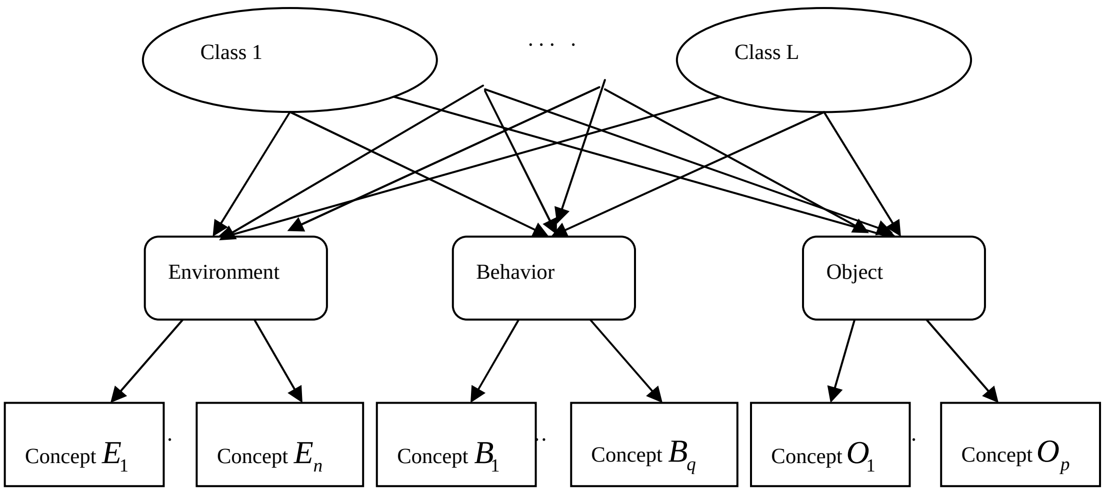

Fig.2 Semantic Branch Structure

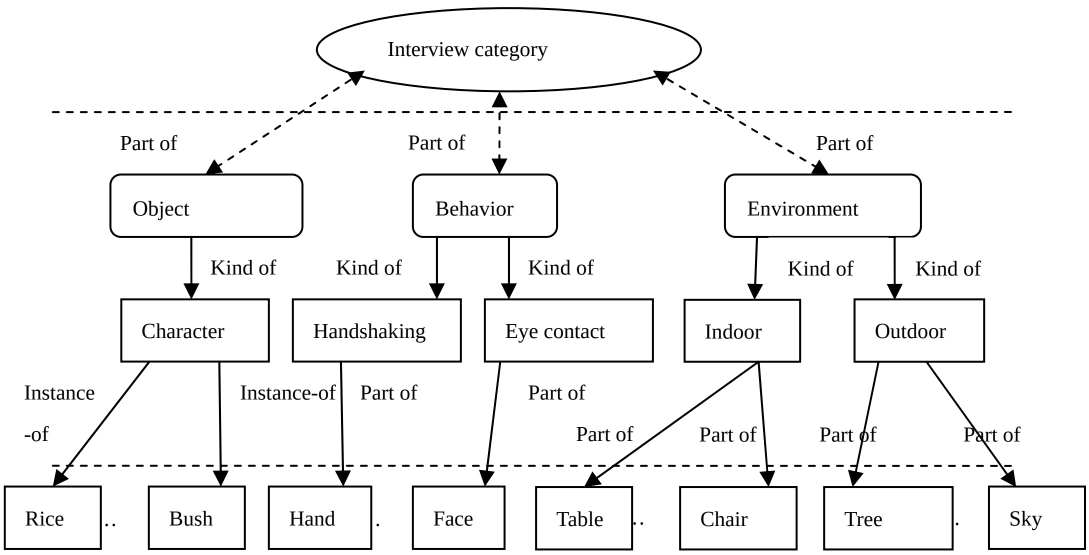

Fig.3 Semantic Branch Structure for Interview Category

3. Operational Rules for Semantic Branch Structure

Describing semantic category by semantic branch structure will generate different semantic branch structures according to different requirements. The difference is reflected on the semantic branch structure parameters.

3.1. Semantic Branch Structure Concept Mask

According to different domain knowledge, we can give semantic concept space. The annotation for key frames in video is expressed by a semantic vector. In this paper, for a specific semantic category, the corresponding semantic concept mask is determined. During the course of derivation, only the semantic concepts relevant to that specific semantic category are considered, and other semantic concepts are ignored. Thus, it will enhance the semantic vector representation accuracy for the semantic categories.

3.2. Operational Rules for Semantic Branch Structure

The probability of semantic branch structure can be derived from its corresponding semantic concepts by certain operations. The rules include the following calculations.

1) Sum operation. We can obtain the probability of a semantic branch structure by summing the probability of all the relevant semantic concepts. This kind of operation can be used in environment semantic module. For environment semantics, the more environment semantic concepts appear, the higher the probability of this environment semantics is.

2) Maximum operation. A lot of semantic concepts are included in one semantic module. We use the maximum of probability of all those semantic concepts as the probability of the semantic branch structure.

3) Value operation. The value of a semantic module is binary number. If the relevant concept disappears, the value is 0, and vice versa. This kind of calculation method is used in object module. For a specific semantic object, we consider only some specific semantic concepts.

The above operational rules can also be adopted in other semantic modules according to other requirements.

3.3. Derivation Rules for Semantic Category

The semantic branch structure includes object module, behavior module and environment module, from which the probability of a key frame in video belongs to a specific semantic category. For a specific semantic category, we give the probability of every semantic module with a value, and derive the probability that a key frame belongs to a specific semantic category.

In semantic branch structure for semantic category L_i, if the probability of object semantic is P_m, the probability of behavior semantic is P_n and the probability of environment semantic is P_q, then the probability that the key frame in video belongs to semantic category C_i can be calculated as $P_{L_i} = aP_m + bP_n + cP_q$, where $a + b + c = 1$. The semantic category of this key frame can be taken as L^θ if $L^\theta = \arg\max(P_{L_i})$, where $1,\cdots,I$, and I is the number of semantic categories. For every specific semantic category, the values of a, b and c are different. They can be set empirically.

4. Experiments

There are three compared values in the experiment. One is based on semantic branch structure, another is based on real-value semantic vector representation method, and the last one is based on binary-value semantic vector representation method. The experimental videos are news videos. The length of the video is almost 4 hours. There are 1065 key frames extracted from the video. The main semantic categories are meeting, interview, basketball match, and weather.

For news video, we give a semantic concept space. The key frame is annotated by the experimental staff. In the news videos, there are four semantic categories in total. For every category, we determine its semantic branch structure concept mask, define its operations and set empirically weights. The content of one key frame is represented by one semantic vector. One key frame is represented by a binary-value semantic vector, if a concept disappears; the corresponding semantic component value is set to 0, otherwise 1. One key frame is represented by a real-value semantic vector, if a concept appears; the relevant semantic component value is added by 1.

In the paper, we use recall rate and precision rate to evaluate the effectiveness of the three representation methods. During the course of experiments, we calculate separately the probabilities of a key frame in four semantic branch structures, and take the semantic category with the maximum probability as the semantic category of the key frame. We calculate separately the centers of four semantic categories by the binary-value semantic vector representation method and the real-value semantic vector representation method. Then we calculate the distances between one key frame and four centers, and take the semantic category with the minimum distance as the semantic category of that key frame. The results are shown as figure 4 and figure 5.

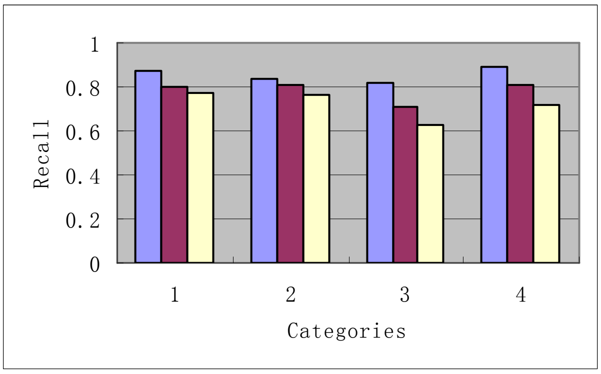

Fig.4 Recall Rates of the Classification Results

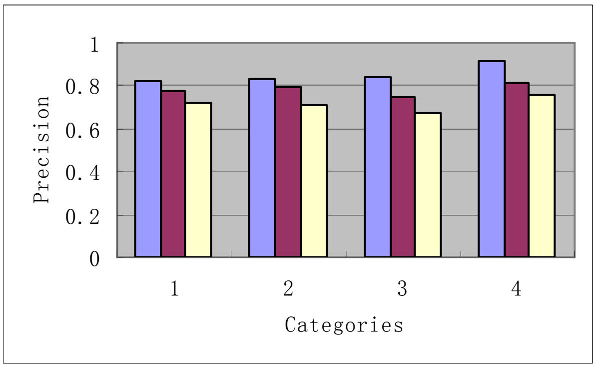

Fig.5 Precision Rates of the Classification Results

In figure 4 and figure 5, the first column is the result of the semantic branch structure, the second column is the result based on real-value semantic vector, and the third column is the result based on binary-value semantic vector. Because the binary-value semantic vector representation method only considers whether a semantic concept appears or not, its effectiveness is the worst. The real-value semantic vector representation method takes into account the contribution of every concept to semantic category for a key frame, so its effectiveness is better. Among all three methods, the effectiveness of our proposed semantic branch structure is best. This structure concerns not only the concepts relevant to specific semantic category, but also it considers the relationship between concepts as well as the relationship between semantic modules and concepts to explore semantics. So its performance is best.

5. Conclusion

For solving the problem of "semantic gap", the paper gives one mathematical model--the matrix division method, which constructs the structure for high-level semantic. According to the matrix division method, we give a semantic branch structure to organize the high-level semantic data for key frames in video. Meanwhile, according to the semantic branch structure we give an experiment contrasted by the real-value semantic vector and the binary-value semantic vector. The results show that the proposed structure is the best one.

Acknowledgement

This work is supported by Scientific and Technological Development Plan Project in JiLin Province (Grant NO. 20090113), Projects of International Cooperation, Exchanges NSFC (Grant No. 609111301281).

References

[1] Sang Keun Rhee, Jihye Lee and Myon-Woong. Park "Semantic Relevance Measure between Resources Based on a Graph Structure". 2008 Proceedings of the International Multiconference on Computer Science and Information Technology, 2008 pp. 229-236.

[2] Gustavo Carneiro, Antoni B. Chan, Pedro J. Moreno and Nuno Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 3, March 2007, pp.394-410.

[3] Pradhan a.S., Hovy E., Marcus M.S., Palmer M., Ramshaw L. and Weischedel R.. "OntoNotes: A Unified Relational Semantic Representation", International Conference on Semantic Computing (ICSC-07), 2008 pp. 517-526.

[4] N. Ruan, N. Huang and W. Hong, "Semantic-Based Image Retrieval in Remote Sensing Archive: An Ontology Approach," Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on, pp.2903-2906.

[5] Michael G. Christel and Alexander G. Hauptmann. "The Use and Utility of High-Level Semantic Features in Video Retrieval", International Conference on Image and Video Retrieval (CIVR'05), 2005 pp. 134-144.

[6] Chun-Yi Lin, Jun-Xun Yin, Xue Gao, Jian-Yu Chen and Pei Qin, "A Semantic Modeling Approach for Medical Image Semantic Retrieval Using Hybrid Bayesian Networks," Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06), vol. 2, 2006, pp.482-487.

[7] Bing Wang, Xin Zhang, Xiao-Yan Zhao, Zhi-De Zhang and Hong-Xia Zhang, "A semantic description for content-based image retrieval," Machine Learning and Cybernetics, 2008 International Conference on, pp.2466-2469.

[8] Lei WANG, Wanjing CAI, Yongli LI, Hongbing WANG, "Understanding Users' Searching Purpose by the Object-oriented Semantic Concept Knowledge Base," Journal of Computational Information Systems, vol. 6(1) 2010, pp. 221-227.