Scientific Ontology Construction Based on Interval Valued Fuzzy Theory under Web 2.0

Na Xue, Suling Jia, Jinxing Hao and Qiang Wang
School of Economics and Management, Beihang University, Beijing, China
Email: xuena12530@163.com, jiasuling@126.com, gordonhao@gmail.com, wang6965@sina.com

Abstract **—Construction of the unified and shared domain ontology is significant for effective knowledge management. For the acquisition and sharing of scientific research knowledge under Web2.0, a novel approach of building Interval Valued Fuzzy Ontology (IVFO) in scientific research domain is presented. Through interval valued fuzzy theory, the definition and constructing framework of IVFO is proposed. Then IVFO is applied to semi-automatic extraction of information retrieval research domain. The preliminary constructing of research domain ontology is an essential base for the knowledge management system of scientific research. It can be effective methods for enhancing the efficiency and productivity of researching.**

Index Terms **—Scientific Research Ontology, Interval Valued Fuzzy Ontology, Semi-automatic Construction, Interval Valued Fuzzy Sets**

I. Introduction

Scientific research is the born and communication base of new science and new technology. It requires effective knowledge management to improve the efficiency of scientific research. The advent of Web 2.0 provides the technical platform for wide communication and cooperation. Coupling with the scientific community, Web 2.0 brings about the open science reformation in the Internet and information age. Open Science (or Open Research, Open Source Science, Science 2.0) becomes research hot spot problems.

Communication and collaboration by network for knowledge sharing and innovation would promote the development of scientific research significantly. As formal specification of a shared conceptualization, ontology is important for knowledge representation and sharing. Therefore, the construction of unified and shared scientific domain ontology is particularly urgent.

Concepts in knowledge resources have various kinds of relations. The boundary of concepts and their relations is difficult to be specified. To deal with the uncertainty, fuzzy set theory was introduced into ontology by many researches. Construction of fuzzy ontology makes it more flexible to adapt to the actual requirement. Tomohiro Takagi et al.[1] fused conceptual fuzzy sets with ontology for representing common concepts, and proposed a conceptual matching method for information retrieval. Valerie V. Cross and Clare Voss [2] performed fuzzy ontology query for multilingual document exploitation. The fuzzy sets theory was introduced into the concept and relationship in the construction of ontology since then. And researches on the fuzzy ontology are emerging gradually [3-13].

However, existing fuzzy ontology models were not fully capable of modeling concepts and objects in a way compatible with the human perception. Traditional type-1 fuzzy sets employed in fuzzy ontology could resolve some uncertainties. But concepts could not be divided by a single criterion for different people have different measures. Besides, views on a certain object from different researchers are difficult to be unified. Especially in the circumstance of open science, the unified cognition needs to be reached. Therefore, interval valued fuzzy concepts and relations that can embrace the different viewpoints would fulfill domain knowledge sharing. Based on this, we proposed a novel method of building scientific domain ontology based on interval valued fuzzy sets, which is a kind of type-2 fuzzy sets. It can well handle the notion of people and a high degree of uncertainties.

A framework of building interval valued fuzzy ontology is proposed in this paper based on the interval valued fuzzy sets theory. The approach is used in scientific domain ontology for knowledge representation. The remainder of this study is organized as follows. Section 2 reviews the related work of scientific domain ontology. An approach of building interval valued fuzzy ontology is proposed in Section 3. Section 4 describes the scientific domain ontology extraction method. Section 6 concludes the paper.

II. Related Work

A. Ontology and Fuzzy Ontology

Ontology is a conceptualization of a domain into a human understandable, machine-readable format consisting of entities, attributes, relationships, and axioms [14]. It is engineered by - but often for - members of a domain by explicating a reality as a set of agreed upon terms and logically-founded constraints on their use [15]. Ontology is widely used for knowledge representation in artificial intelligence, information retrieval and semantic web.

doi:10.4304/jsw.8.8.1835-1842

However, typical ontology with crisp description could not handle the uncertainties in real applications for the lack of clear-cut distinctions of domain concepts. For example, a manuscript can be quite innovative, general innovative or less innovative. Therefore, expert opinions on the manuscript can be different or even opposite. To handle with this problem, several researchers employed fuzzy set theory to construct fuzzy ontology.

Lee C. S. et al. [16] proposed a fuzzy ontology and applied it to news summarization. Quan Thanh Tho et al. [17] incorporated fuzzy logic into ontology to represent uncertainty information, and proposed Fuzzy Ontology Generation frAmework for automatic generation of fuzzy ontology on uncertainty information. LAU R. et al. [18] illustrated a novel concept map generation mechanism which is underpinned by a fuzzy domain ontology automatic extraction algorithm. C. A. Yaguinuma et al. [19] tried to use fuzzy logic concepts in crisp ontologies for a more expressive representation of vague information relevant to some domains, and presented DISFOQuE system for data integration based on fuzzy ontology.

B. Scientific Ontology

Scientific knowledge management has attracted attention from many researchers to join the study. Here we reviewed the works by means of ontology technique and method.

M. Ettorre et al. [20] developed an experimental knowledge management tool serving the ICAR-CNR, a research institute of the Italian National Council of Research. It applied in the formalized groups (such as project groups) where collaborative work takes place and informal groups (communities of practice) that may arise around common problems, interests and objectives. The tool provides a web-based environment for knowledge creation, sharing and access. It helps researchers quickly set up their own portals, create and organize knowledge items (such as news, papers, links, announcements, projects, etc.), share them within workgroups, search for a specific document or browse through a set of related documents (ontology-driven browsing). This should be done by basing on the built of the Research Ontology and the Knowledge Item Ontology, which contain the formal specification of application domain concepts, relations and constraints.

Q.T. Tho et al. [21] employed context-based cluster analysis (CCA) and context-based ontology generation framework (COGA) to develop a significantly improved citation-based document retrieval system. COGA aims to generate ontology from clusters relationship built by CCA. The improved retrieval system is applied to find research domain experts.

J.C. de Almeida Biolchini et al. [22] developed the ontologies to describe knowledge regarding systematic reviews of experimental studies. As an explicit specification of conceptualization, the scientific research ontology can be useful to guarantee the terminological homogeneity of the concepts that are to be used by different researchers, contributing to a higher consistency between the retrieved information and the consequent results.

Sheng-Yuan Yang [23] developed an ontology-supported information integration and recommendation system for scholars. It can extract important information from domain documents by information integration and recommendation ranking. Ontology database is built to serve the webpage crawler for querying the webpage of related scholars, and to assist the webpage classifier in processing webpage classification, etc.

Bian Wen-yu et al. [24] discussed the problems in present knowledge management for scientific research in China, such as complicated knowledge sources, scattered storage and innovated knowledge's delayed storage, etc. Ontology is adopted to build knowledge model of scientific research. The study constructed an ontology system of scientific research knowledge and showed the formal expression of scientific research knowledge ontology.

Hong Na and Zhang Zhi-xiong [25] chosen science individuals as analysis objects, using ontology construction tool (Protégé) to create Science Ontology and reason on a simple relation. And the study also summed up some problems still existing in large scale ontology storage and management.

Zhang Qiang [26] proposed the literature ontology structure and semantic dictionary structure through combining ontology construction tools (Protégé) and ontology inference machine (Racer). OWL ontology description language is employed to express the semantic information of literature domain ontology. The research realized the semantic and intelligent retrieval of literatures.

The existing researches are mainly concerned the scientific domain objects like literature retrieval and finding experts. But concepts of scientific research theme and their uncertain boundary are ignored. Besides, there also has the problem of unifying different personnel cognitive under the Web2.0 environment. Even many researches on fuzzy ontology only consider a real number to describe the concept hierarchy. But assigning an exact number to an expert's opinion is too restrictive, and that the assignment of an interval of values is more realistic. On the other hand, interval fuzzy sets could be used to handle group decision process as they can model different between expert preferences. This requires the construction of scientific research domain ontology with type-2 fuzzy theory, as it deals with human knowledge representation.

Chang-Shing Lee et al. [27] combining type-2 fuzzy sets with ontology for the first time to develop the personal diet recommend system. The type-2 fuzzy ontology improved satisfaction of experts and users at the classification of personal diet. Interval fuzzy sets are used as a special situation of type-2 fuzzy sets. Based on this, we promote interval valued fuzzy sets in scientific research domain. The uncertainties mainly concerned include the follows:

(1) uncertainties associated with sensed measurements and the level of scholars;

(2) uncertainties associated with changing relations and application context;

(3) uncertainties of linguistic meaning to different people;

(4) uncertainties associated with the experts' opinions and criterions.

Type-2 fuzzy sets could be used to handle the above uncertainties in group knowledge sharing as they can model the uncertainties between expert preferences. Interval valued fuzzy sets are most widely used type-2 fuzzy sets. Therefore we employ interval valued fuzzy sets to model the scientific concepts for knowledge representation.

III. Construction of Interval Valued Fuzzy Ontology

A. Interval Valued Fuzzy Theory

Current fuzzy ontology employs the type-1 fuzzy sets theory, in which the membership of each element is a crisp number in [0, 1].

Definition 1. A (type-1) fuzzy set A in X is a set of ordered pairs

$$A = \{(x, \mu_A(x))\}, x \in X$$

where $\mu_A(x)$ is the grade of membership of x in A and $\mu_A : X \to [0,1]$ is called the membership function [28].

Using type-1 fuzzy sets to deal with vague knowledge, the correlation value of concepts in fuzzy ontology is defined as a number in [0, 1]. It breaks the limitations of classic ontology that can only determine whether the concepts are related or unrelated. Fuzzy sets are required to define the descriptive concepts in scientific domain ontology.

Although type-1 fuzzy sets can alleviate the impact of uncertainty to a certain extent, the information from the membership degree is not complete, especially for modeling expert group decision process. Nevertheless, it is still a great challenge to specify a crisp membership degree to describe domain concepts and their boundary in scientific ontology.

For example, consider a type-1 fuzzy set representing the linguistic evaluation of a certain research article from three experts in Fig. 1. As shown, the experts give different scores on the same aspect. However, if we have to assign a number to evaluate the practicality, like defining the mean of all scores as the final score, we will miss significant information about the reviews.

Actually, type-1 fuzzy sets are not capable to deal with high uncertainties in several situations [29]:

(1) Linguistic expression whose meaning is different from different reviewers;

(2) Group opinions which are not unified on the same object.

These uncertainties are difficult to be described by s crisp number with type-1 fuzzy sets. To deal with these uncertainties, we employed type-2 fuzzy sets whose membership functions are fuzzy set instead of a crisp number.

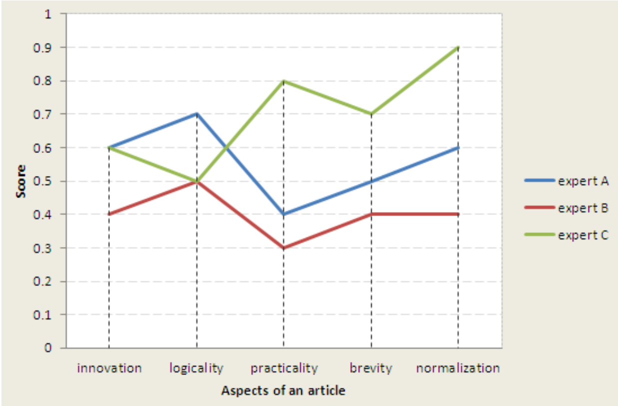


Fig. 1 Type-1 fuzzy sets for linguistic evaluation on an article from three experts

Different from the type-1 fuzzy sets, type-2 fuzzy sets have the membership functions mapping from U to [0, 1]. That is to say, the membership function of type-2 fuzzy sets ranges over type-1 fuzzy sets.

Definition 2. A type-2 fuzzy set $\tilde{A}$ is characterized by a type-2 membership function $\mu_{\tilde{A}}(x, u)$, where $x \in X$ and $u \in J_x \subseteq [0,1]$, i.e.,

$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) \mid \forall x \in X, \forall u \in J_x \subseteq [0,1]\}$$

in which $0 \leqslant \mu_{\tilde{A}}(x, u) \leqslant 1$ [29].

When the secondary grades all equal one, the resulting type-2 fuzzy sets are called interval valued fuzzy sets. I.e., for $\forall u \in J_x \subseteq [0,1], f_x(u) \equiv 1$. Such sets are the most widely used type-2 fuzzy sets to date. Interval valued fuzzy sets are the most widely used type-2 fuzzy sets because they are simple to use.

Let I be a closed unit interval, i.e., $I=[0,1]$, and $[I] = \{\bar{a} = [a^-, a^+] : a^- < a^+, a^-, a^+ \in I\}$.

Definition 3. Let X be an ordinary set, mapping $A : X \to [I]$ is called an **interval valued fuzzy set (IVFS)** on X.

B. Interval Valued Fuzzy Ontology (IVFO)

A fuzzy ontology is extended domain ontology with fuzzy concepts and fuzzy relationships [5]. Type-2 fuzzy ontology is a knowledge representation model to describe the domain knowledge with uncertainty. It is an extension of the domain ontology [27]. We will give the formal definition of type-2 fuzzy ontology.

Definition 4. A **Type-2 Fuzzy Ontology (T2FO)** is a 6-tuple $O_{2F} = \{C, A^C, R, A^R, H, X\}$, where C is a set of concepts; A^C is the set of attributes describing each concept; $R = (R_T, R_N)$ represents the type-2 fuzzy relation, including the taxonomy relation R_T and non-taxonomy relation R_N; A^R defines the attributes of type-2 fuzzy relation R; H describes the concept hierarchy; X is a set of axioms.

When $R = (R_T, R_N)$ are the interval fuzzy relations, the type-2 fuzzy ontology is the Interval Valued Fuzzy Ontology (IVFO).

Definition 5. An **Interval Valued Fuzzy Ontology (IVFO)** is a 6-tuple $O_{[I]}=\{C, A^C, R, A^R, H, X\}$, where C is a set of concepts; A^C is the set of attributes describing each concept; $R=(R_T, R_N): C\rightarrow[I]$ represents the interval valued fuzzy relation, including the taxonomy relation $R_T: C\rightarrow[I]$ and non-taxonomy relation $R_N: C\rightarrow[I]$; A^R defines the attributes of type-2 fuzzy relation R; H describes the concept hierarchy; X is a set of axioms.

When it comes specifically to scientific ontology, C can refer to a certain form of scientific outcomes (such as journal article); then A^C is the attributes describing the outcomes (for example, the title, keywords, abstract, author, reference of the article and so on). If C is the research topic concept, then A^C contains topic name, research object, method, related publication and so on.

The general scientific domain ontology covers research topic, researcher, activity, outcome, organization, and facility. Each category can be further classified. For example, the scientific outcome includes patent, book, report, article and so on. This study mainly aims at discussing how to construct knowledge ontology system of research domain topic. Part of data mining research concept hierarchy is shown in Fig. 2.

C. Construction Framework of Interval Valued Fuzzy Ontology

Generally, there are two ways to build ontology: manually deriving ontology from knowledge and data; semi-automatic extracting ontology [8,10] from data. Manually deriving ontology requires expert group to predefine concepts and give out the annotation. Automatically extracting ontology from corpus applies information retrieval approach. Manual generation of ontology is difficult and tedious. Therefore, automatic generation of concept is highly desirable.

Ontology generation data sources include non-structured, semi-structured and structured data. Between these, text data are the most used. In this study, we present the generation of interval valued fuzzy ontology of scientific domain from the text data. The framework is shown in Fig. 3.

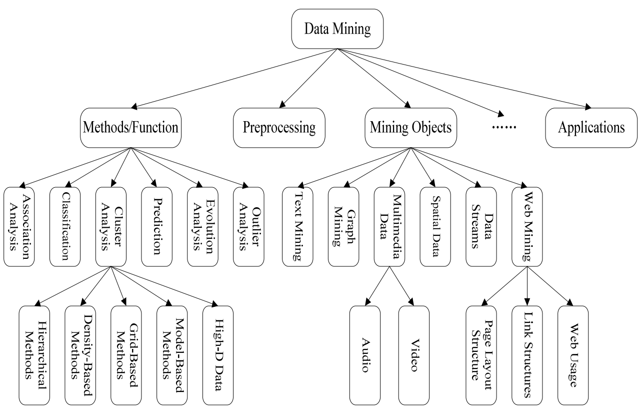


Fig. 2 Data mining research concept hierarchy

IV. Semi-Automatic Generation of Scientific Ontology

A. Concept Extraction

The scale of the core concept in scientific domain ontology is small than upper ontology. For example, the number of concepts in different ontology is distributed as: General Field > Scientific Domain > Business Intelligence Discipline > Data Mining Topic.

In the scientific domain, the concepts can be divided into three categories: basic concept, evaluative concept and exploratory concept. Basic concept is the normal domain specification, such as the journal, author, and algorithm and so on. Evaluative concept comprises a significant degree of descriptive content and is evaluatively loaded. For example, expert, top journal, hot subject etc. Exploratory concept is innovated research which still being argued. This kind of classification represents the evolution of researching.

According to the fuzzy degree of concept, there are crisp concept, (type-1) fuzzy concept and interval fuzzy concept in interval valued fuzzy ontology. Crisp concept has clear definition of intension and extension. The membership degree of a fuzzy concept is a number in [0, 1]. Semantic annotation of fuzzy concept membership can be the significance of term [2], and etc. The element of interval valued fuzzy concept is fuzzy concept. Fuzzy concept and interval valued fuzzy concept are suitable for overlapping concepts.

The domain knowledge item with specific attribute is always what scholar considers. For example, the concept of "book" of research outcome could have attributes as reference value, advantage. While in teaching field, it can have learnability, universality and applicability. Domain knowledge is the intelligent reasoning part combines with relevant concept data, which can most reflect the value of knowledge ontology. Therefore, it is fuzzier and more complex, which is applicable using interval valued fuzzy theory.

The three main methods of concept extraction are rule-based approach, machine learning based on statistics and method based on structured data. The measures that are used to estimate the membership degree include Jaccard (JA), conditional probability (CP), Kullback-Leibler divergence (KL), Expected Cross Entropy (ECH), Nomalized Google Distance (NGD) and Mutual Information (MI). We employ the balanced Mutual Information (BMI) [8] to obtain the membership degree of the term to concept (see Eq. (1)).

$$\begin{aligned}\mu_{c_i}(t_j) &\approx BMI(t_i,t_j)\\ &=\beta\times[P(t_i,t_j)\log_2(\frac{P(t_i,t_j)+1}{P(t_i)P(t_j)})\\ &\quad+P(\neg t_i,\neg t_j)\log_2(\frac{P(\neg t_i,\neg t_j)+1}{P(\neg t_i)P(\neg t_j)})]\\ &\quad-(1-\beta)\times[P(t_i,\neg t_j)\log_2(\frac{P(t_i,\neg t_j)+1}{P(t_i)P(\neg t_j)})\\ &\quad+P(\neg t_i,t_j)\log_2(\frac{P(\neg t_i,t_j)+1}{P(\neg t_i)P(t_j)}]\end{aligned} \tag{1}$$

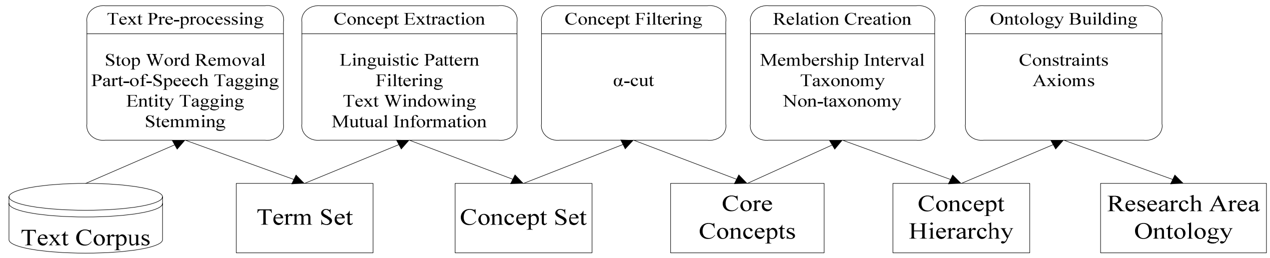


Fig. 3 The generation framework of interval valued fuzzy ontology

Where $\mu_{c_i}(t_j)$ is the membership function to calculate the degree of term $t_j \in A^C$ belonging to concept $c_i \in C$. $P(t_i, t_j)$ is the joint probability that both terms appear in a text window. $P(t_i) = |w_{t_i}|/|w|$ is the probability of term t_i appears in a text window, in which $|w_{t_i}|$ is the number of windows containing the term t_i and $|w|$ is the total number of windows constructed from a corpus. $\neg t_i$ represents contain no t_i. $\beta > 0.5$ is used to control the relative importance of two kinds of evidence.

In order to get more information, we tune the window length to be 5 and 10. After the text data been processed, the intersection of two term sets is the final result of the concept set. The two membership values are end point of the interval. For example, for $\delta = 5$, the degree of the term "recommendation" belonging to concept "intelligence" is 0.62; for $\delta = 10$, the degree of the term "recommendation" belonging to concept "intelligence" is 0.81. Therefore, the interval of "recommendation" belonging to "intelligence" is [0.62, 0.81].

B. Interval Valued Fuzzy Relation Extraction

The hierarchical relation of domain ontology includes hypernymy relation and hyponymy relation. Non-hierarchical relation includes synonymy relation and other operational relations. As the concept hierarchy is a partially ordered set [10], which can be generated by the probable order of terms appear in text windows.

The similarity of terms is effectively measured by semantic dictionary, for it cannot be reflected by the co-occurrence. In our study, we turn to WordNet - a lexical network of English words. WordNet has the semantic senses of the words and has become one of the most widely adopted resources for semantic analysis. In WordNet, nouns, verbs, adjectives, and adverbs are each organized into networks of synonym sets (synsets) that each represents one underlying lexical concept and are interlinked with a variety of relations. A polysemous word will appear in one synset for each of its senses. With WordNet, the calculation of semantic similarity of two different terms is shown in Eq. (2).

$$S(t_i, t_j) = \alpha S_s + (1-\alpha) S_r = \alpha \frac{s_i \cdot s_j}{\|s_i\| \cdot \|s_j\|} + (1-\alpha) \frac{\|r_i - r_j\|}{\|r_i + r_j\|} \qquad (2)$$

Where $\alpha \le 1$ decides the relative importance and contributions of semantic information S_s and word order information S_r; α is suggested to 0.85. s_i and r_i are semantic vector entropy and word order vector of t_i respectively [30].

C. Interval Valued Fuzzy Ontology Axiom

Ontology axiom is the constraint on the concept's and relation's attribute value, or the relations between concepts. Axioms cover concept axiom, attribute axiom and instance axiom. The format describing axiom is SWRL [31], as following:

$$antecedent \Rightarrow consequent$$

Expanded to fuzzy format for fuzzy ontology, denoted by FSWRL, as shown:

$$antecedent_{[a]} \Rightarrow consequent_{[b]}$$

Further expanded to interval fuzzy format for interval valued fuzzy ontology, denoted by IFSWRL, as following:

$$antecedent_{[a^-, a^+]} \Rightarrow consequent_{[b^-, b^+]}$$

where $a \in [0, 1], b \in [0, 1]$, and $[a^-, a^+] \subseteq [0, 1]$, $[b^-, b^+] \subseteq [0, 1]$.

In scientific area knowledge ontology, the main rules and axioms are endowed as follows.

Implies (Antecedent (consist-of (C_1, C_2) [a,b])
Consequent (belong-to (C_2, C_1) [a,b]))
Implies (Antecedent (belong-to (C1, C2) [a,b])
Consequent (consist-of (C_2, C_1) [a,b]))
// the relation of "consist-of" and "belong-to" are reverse
Implies (Antecedent (super-area (C1, C2) [a,b])
Consequent (sub-area (C_2, C_1) [a,b]))
Implies (Antecedent (sub-area (C1, C2) [a,b])
Consequent (super-area (C_2, C_1) [a,b]))
// the relation of "super-area" and "sub-area" are reverse
Implies (Antecedent (relevant (C1, C2) [a,b])
Consequent (relevant (C_2, C_1) [a,b]))
Implies (Antecedent (relevant (C1, C2) [a,b] relevant (C2, C3) [c,d])
Consequent(relevant(C_1, C_3)[a*c, b*d]))
// the relation of "relevant" are transitive

V. Performance Evaluation

Experiments on datasets are conducted to show the effectiveness of the proposed method. We try to evaluate the generated ontology with the ontology developed by different experts. The experiments are carried out on a PC running Windows 7 with Intel Pentium Dual-Core CPU E5200 (2.50GHz), 3.0G RAM.

The data sets are 100 papers of 10 topics about information management and knowledge engineering from online Web of Knowledge digital database (http://apps.webofknowledge.com/). Each topic has 10 papers. The concept extraction algorithm is EntropyConceptExtraction, relation extraction algorithm is SubcatRelationExtraction. Our evaluation metrics are precision, recall and F-measure of the ontology construction method.

$$Precision_C = \frac{|C_E \cap C_S|}{|C_S|}$$

$$Recall_C = \frac{|C_E \cap C_S|}{|C_E|}$$

$$Precision_R = \frac{|R_E \cap R_S|}{|R_S|}$$

$$Recall_R = \frac{|R_E \cap R_S|}{|R_E|}$$

$$Precision_O = \omega_P \times Precision_C + (1\text{-}\omega_P) \times Precision_R$$

$$Recall_O = \omega_R \times Recall_C + (1\text{-}\omega_R) \times Recall_R$$

where C_E and C_N represent the set of concept found from the ontology created by human experts and that generated by our method respectively. Similarly, R_E and R_S are the set of relations of the ontology given by human experts and that generated by our method respectively. The parameter ω_P is used to tune the ontology precision measure based on a weighted sum of the concept precision and relation precision respectively. Similarly, the parameter ω_R is used to tune the ontology recall measure. For the experiments conducted in this study, we adopt $\omega_P = \omega_R = 0.5$. The standard F-measure is shown in the following equation.

$$F_\eta = \frac{(1+\eta^2) Precision \times Recall}{\eta^2 Precision + Recall}$$

If we assume that precision is as important as recall (i.e. $\eta = 1$), then the ontology F-measure is calculated by:

$$F_O = \frac{2 \times Precision_O \times Recall_O}{Precision_O + Recall_O}$$

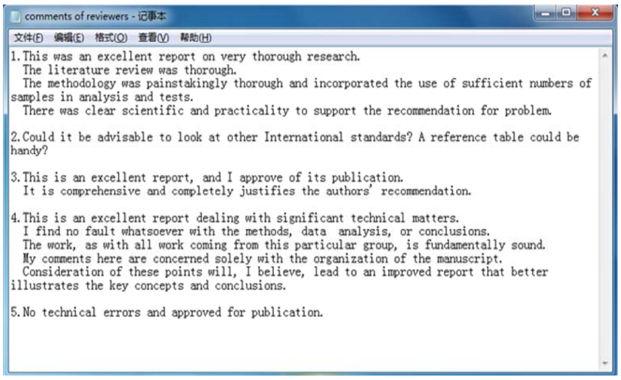
The precision, recall and F-measure of the ontology achieved over the 10 topics are shown in Table 1. The divided topic results are shown in terms of ontology precision, ontology recall and ontology F-measure. The second and third columns refer to the number of concepts and relations generated by the method.

As can be seen from the average results, the ontology F-measure means the ontology generated by our method can mainly represent the domain knowledge as perceived by the domain expert.

The second experiment we conducted is to model the comments from different experts. We selected comments of 5 reviewers on a certain topic research on the web. Fig. 4 shows a sample of an article reviewed comments from different reviewers.

TABLE 1.
THE RESULTS OF THE GENERATED ONTOLOGY OF DIFFERENT TOPICS

Topic	Concept	Relation	Ontology Precision	Ontology Recall	Ontology F-measure
1. Web mining	12	82	0.410	0.696	0.516
2. Multimedia mining	29	91	0.293	0.614	0.397
3. Metadata and ontologies	16	96	0.317	0.728	0.442
4. Time series forecasting	23	90	0.450	0.783	0.572
5. Link and sequence analysis	19	85	0.385	0.789	0.517
6. Text mining	22	98	0.415	0.773	0.540
7. Knowledge representation	27	103	0.444	0.741	0.555
8. Multidimensional query optimization	18	89	0.368	0.778	0.499
9. Clustering methods	31	109	0.488	0.677	0.567
10. Information extraction methodologies	26	95	0.514	0.731	0.604
Avg.	22.3	93.8	0.408	0.731	0.521


comments of reviewers - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

1.This was an excellent report on very thorough research.
The literature review was thorough.
The methodology was painstakingly thorough and incorporated the use of sufficient numbers of samples in analysis and tests.
There was clear scientific and practicality to support the recommendation for problem.

2.Could it be advisable to look at other International standards? A reference table could be handy?

3.This is an excellent report, and I approve of its publication.
It is comprehensive and completely justifies the authors' recommendation.

4.This is an excellent report dealing with significant technical matters.
I find no fault whatsoever with the methods, data analysis, or conclusions.
The work, as with all work coming from this particular group, is fundamentally sound.
My comments here are concerned solely with the organization of the manuscript.
Consideration of these points will, I believe, lead to an improved report that better illustrates the key concepts and conclusions.

5.No technical errors and approved for publication.

Fig. 4 The sample of comments on a certain topic from 5 reviewers.

The descriptions on the study from the expert are assigned as numbers in [0, 1]. For example, descriptive comments can be an element of {poor, not sufficient, general, well organized, and excellent} on a specific aspect, which is translated to be {0.1, 0.3, 0.5, 0.7, and 0.9} respectively. Each element represents the set of synonymy term group. That is to say, "outstanding" and "excellent" both represent the score of 0.9 as they are in the same synonymy term group. According to the description category, scientific objects evaluation ontology can be obtained. With the score generated in detail, the aggregate assessment from one expert could be generated. That is expressed with the type-1 fuzzy ontology. In other words, type-1 fuzzy ontology can be used to represent the review result from one expert. It supports both the overall evaluation and drilling down details. The type-1 fuzzy ontology generated from the fourth reviewer describing the report is shown in Fig. 5. As can be seen from the figure, the illustration of the key concepts and conclusions are considered to be improved according to the reviewer. And the entire assessment can be inferred from the bottom results.

After the opinion of one expert is delivered by type-1 fuzzy ontology, we can generate the annotations of a certain object from different experts. And Fig. 6 represents the aggregate result of the comments with interval valued fuzzy ontology. The comments are illustrated with the interval value, in which the lower bound and upper bound are minimized and maximized of the comments on a certain aspect of the object. The super interval is the weighted average of its sub value intervals from the bottom to the top.

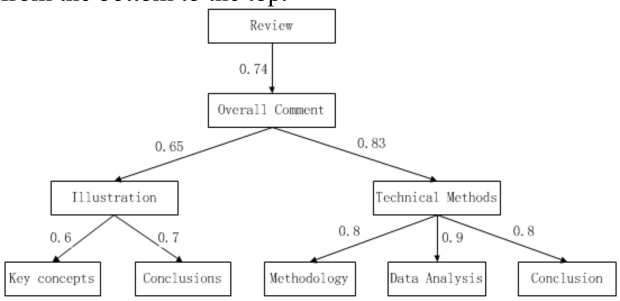


Fig. 5 The type-1 fuzzy ontology obtained from the fourth reviewer.

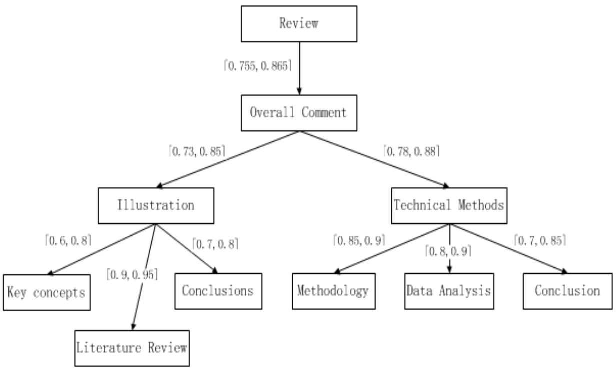


Fig. 6 The interval valued fuzzy ontology obtained from all reviewers.

As is shown in Fig. 5 and Fig. 6, the comment of a reviewer on a certain research outcome can be well described in detail by the fuzzy ontology. Extended to more general situations, online scholars can make annotations on a certain document and other scientific ontology objects. For example, researchers can evaluate from as tiny as an article theorem, to as large as a subject's top journals or leading scientists. With the aggregate online comments, then valuable domain knowledge is available to scholars.

VI. CONCLUSIONS

With the diversity of scientific research subject and the surge of the research achievements, effectively arrangement of research knowledge is urgently needed to make full use of scientific intellectual assets. Web 2.0 technologies are promoting the scientific research through more online communication and collaboration to share knowledge. The construction of shared and reused scientific ontology could promote research efficiency.

This paper proposed a novel approach to generate the scientific ontology through interval valued fuzzy sets. The interval valued fuzzy theory is introduced to deal with the uncertainty of concepts and different comprehension under the context of Web 2.0. To achieve the united domain knowledge, we employed the ontology construction method to generate domain concept and relations. Represented by the research topic, the semi-automatic generation framework is illustrated, which includes concept extraction, relation extraction and axioms. The methodology proposed in this study provides a valuable reference to ontology building in other domains in Web 2.0 ages.

Scientific knowledge ontology building is a complicated project. In addition, ontology of the researcher, activity, outcome, organization and facility sub area still need to be studied. The work of standardization of the process of formal ontology and its inference should be conducted in the future.

ACKNOWLEDGMENT

The work reported in this paper has been funded in part by the China National natural science foundation youth science fund project (71101005), doctoral program of higher education by the specialized research fund project (20111102120022) and the special fund of central university basic scientific research business expenses.

REFERENCES

[1] Tomohiro Takagi, Shinichi Kasuya, Masao Mukaidono, Toru Yamaguchi, Takashi Kokubo, “Realization of Sound-scape Agent by the Fusion of Conceptual Fuzzy Sets and Ontology,” *Proceedings of IEEE International Fuzzy Systems Conference*, 1999.

[2] Valerie V. Cross, Clare Voss, “Fuzzy ontologies for multilingual document exploitation,” *18th International Conference of the North American Fuzzy Information Processing Society- Nafips*, 1999: 392-397.

[3] Chang-Shing Lee, Mei-Hui Wang, Huan-Chung Li, and Wen-Hui Chen, “Intelligent Ontological Agent for Diabetic Food Recommendation,” *IEEE International Conference on Fuzzy Systems*, 2008: 1805-1812.

[4] Chang-Shing Lee, Yea-Juan Chen, Zhi-Wei Jian, “Ontology-based fuzzy event extraction agent for Chinese e-news summarization,” *Expert Systems with Applications*, 2003, 25(3): 431-447.

[5] Chang-Shing Lee, Zhi-Wei Jian, Lin-Kai Huang, “A fuzzy ontology and its application to news summarization,” *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 2005, 35(5): 859-880.

[6] Umberto Straccia, “Towards a Fuzzy Description Logic for the Semantic Web,” *The Semantic Web: Research and Applications*, 2005: 73-123.

[7] Wang, Mei-Hui, Chang-Shing Lee, Zhi-Wei Chen, Hani Hagras, Su- E. Kuo, Hui-Ching Kuo, and Hui-Hua Cheng. "A Type-2 Fml-Based Fuzzy Ontology for Dietary Assessment." Chap. 9 In *On the Power of Fuzzy Markup Language*, edited by Giovanni Acampora, Vincenzo Loia, Chang-Shing Lee and Mei-Hui Wang. Studies in Fuzziness and Soft Computing, 149-68: Springer Berlin Heidelberg, 2013.

[8] Raymond Lau, Dawei Song, Yuefeng Li, Terence Cheung, Jin-Xing Hao, “Towards a fuzzy domain ontology extraction method for adaptive e-learning,” *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(6): 800-813.

[9] JIANG, B., ZHU, M., WANG, J., “Ontology-Based Information Extraction of Crop Diseases on Chinese Web Pages,” *Journal of Computers*, 2013, 8(1): 85-90.

[10] Tho Q.T., Hui S.C., Fong ACM, Cao T.H., “Automatic Fuzzy Ontology Generation for Semantic Web,” *IEEE Transactions on Knowledge and Data Engineering*, 2006, 18(6): 842-856.

[11] ZHAI, J., LI, J., LIN, Y., “Semantic Retrieval Based on SPARQL and Fuzzy Ontology for Electronic Commerce,” *Journal of Computers*, 2011, 6(10): 2127-2134.
[12] Zhao Jian, Liu Lei, “Construction of concept granule based on rough set and representation of knowledge-based complex system,” *Knowledge-based Systems*, 2011, 24(6):809-815.
[13] GONçALVES, E., “Specifying Knowledge in Cognitive Multiagent Systems Using a Class of Hierarchical Petri Nets,” *Journal of Software*, 2012, 7(11): 2405-2414.
[14] Guarino, N., Carrara, M., Giaretta, P., “Ontologies and knowledge bases: towards a terminological clarification,” *Amsterdam IOS Press*, 1995.
[15] Mika, P, “Ontologies are us: A unified model of social networks and semantics,” *Journal of Web Semantics*, 2007, 5(1): 5-15.
[16] Lee, C.S., Z.W. Jian, and L.K. Huang, “A fuzzy ontology and its application to news summarization,” *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 2005, 35(5): 859-880.
[17] Tho, Q.T., Hui, S.C., Fong, A.C.M., Cao, T.H., “Automatic Fuzzy Ontology Generation for Semantic Web,” *IEEE TTransactions on Knowledge and Data Engineering*, 2006, 18(6): 842-856.
[18] LAU, R., SONG, D., LI, Y., CHEUNG, T., HAO, J., “Towards a fuzzy domain ontology extraction method for adaptive e-learning,” *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(6): 800-813.
[19] Yaguinuma, C.A., et al., “A Fuzzy Ontology-Based Semantic Data Integration System,” *Journal of Information & Knowledge Management*. 2011, 10(3): 285-299.
[20] M. Ettorre, L.P., M. Ruffolo, P. Rullo, D. Saccà., “A Prototypal Environment for Collaborative Work within a Research Organization,” *14th IEEE International Workshop on Database and Expert Systems Applications (DEXA’03)*. 2003.
[21] Tho Q.T., Hui S.C., Fong ACM., “A citation-based document retrieval system for finding research expertise,” *Information Processing and Management*, 2007, 43(1): 248-264.
[22] J.C. de Almeida Biolchini, P.G. Mian, A.C.C. Natali, T.U. Conte, G.H. Travassos., “Scientific research ontology to support systematic review in software engineering,” *Advanced Engineering Informatics*, 2007, 21(2): 133-151.
[23] Yang S.-Y., “Developing an ontology-supported information integration and recommendation system for scholars,” *Expert Systems with Applications*, 2010, 37(10): 7065-7079.
[24] BIAN Wen-yu, BAO Zhen-qiang, ZHOU Lei, WANG Chao, “Ontology-based knowledge management modeling of scientific research,” *Application Research of Computers*, 2007, 24(8): 93-96.
[25] HongNa, Zhang Zhixiong, “Practice of Creating and Reasoning Science Ontology by Protégé,” *Modern library and information technology*, 2009, 181/182(7/8): 1-5.
[26] Zhang Qiang, “Research on literature domain ontology construction and reasoning based on the Semantic Web,” *Software Guide*, 2010, 9(11): 139-141.
[27] Chang-Shing Lee, Mei-Hui Wang, Hani Hagrasm, “A Type-2 Fuzzy Ontology and Its Application to Personal Diabetic-Diet Recommendation,” *IEEE Transactions on Fuzzy Systems*, 2010, 18(2): 374-395.
[28] GLEB BELIAKOV, “Fuzzy Sets and Membership Functions Based on Probabilities,” *Information Sciences*, 1996, 91(1-2):95-111.
[29] Mendel, J. M., John, R. I. B., “Type-2 fuzzy sets made simple,” *IEEE Transactions on Fuzzy Systems*, 2002, 10(2): 117-127.
[30] Li Y. H., McLean D., Bandar Z. A., O'Shea J. D., Crockett K, “Sentence similarity based on semantic nets and corpus statistics,” *IEEE Transactions on Knowledge and Data Engineering*, 2006. 18(8): 1138-1150.
[31] SWRL: A Semantic Web Rule Language Combining OWL and RuleML, *http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/*. 2004.

Na Xue, born in 1985, is currently a Ph. D candidate in management science and engineering at School of Economics and Management from Beihang University, Beijing, China. She received her master degree of Science from Qufu Normal University in 2010. Her main research interests include semantic web, knowledge management and ontology learning.

Suling Jia, born in 1954, is a professor of School of Economics and Management, Beihang University, Beijing, China. Her current research interests include information management and information system, system dynamics.

Jinxing Hao, born in 1978, is an Assistant Professor at the School of Economics and Management, Beihang University. He received his Ph. D in Business Information Systems from City University of Hong Kong and Master in Informatics from Wuhan University. His research interests include knowledge management, business intelligence and mobile commerce. He has published articles in journals such as Journal of the American Society for Information Science and Technology, IEEE Transactions on Knowledge and Data Engineering, Decision Support Systems and others.

Qiang Wang, born in 1966, is an associate professor of School of Economics and Management, Beihang University, Beijing, China. His current research interests include information management and information system, system dynamics.