

Semantic Annotation of Ontology by Using Rough Concept Lattice Isomorphic Model

Hongsheng Xu, Ruiling Zhang*, Chunjie Lin and Wenli Gan

*College of Information Technology, Luoyang Normal University
Henan LuoYang, 471022, China*

**corresponding author, zhangruilingls@163.com*

Abstract

Semantic annotation is the process based on ontology annotation concept class, attribute and other metadata for cyber source and its various parts. Ontology mapping is to calculate the similarity between two ontology elements. Ontology merging is two or more source ontology merging into a goal Ontology. The basic principle of the concept lattice isomorphic generating is isomorphic to the background of the isomorphic concept lattice, and as concept lattice isomorphic background can generate the concept lattice. This paper analyzes the methods of ontology mapping and merging based on rough concept lattice isomorphic model and presents semantic annotation of ontology by using rough concept lattice isomorphic model. Experiments show that this method is better than the traditional method in semantic annotation accuracy and breadth.

Keywords: *Rough concept lattice; Ontology mapping; Ontology merging; Semantic annotation; Isomorphic model*

1. Introduction

Ontology mapping is the key technology to solve the information exchange between heterogeneous and distributed ontologies, ontology mapping is to calculate the similarity between ontology elements on the two elements, the similarity between ontology instances, and the definition of ontology and ontology rules or constraints, so the mapping method is also divided into methods, with the ontology structure and examples correlation and related methods and fields.

Ontology mapping language is used to represent the ontology mapping in M language. Similar in different substantive norms, the mapping language is important [1]. Here said the entity (entities) refers to the ontology itself and a real ontology mapping. Although an ontology mapping truly illustrates the relationship between ontology entities, ontology similarity of common standards for substantial similarity credibility to characterize (usually in 0 to 1 between the values). This is a typical precise specification; simple measurement of similarity is more powerful than it.

Ontology merging is two or more source ontology merging into a goal. For ontology merging, the use of traditional editing tools for manual and time-consuming and laborious, but also prone to error. According to the traditional theory, the usual method ontology integration can be divided into manual, semi-automatic and automatic method. Although artificial ontology fusion accuracy is higher, but time-consuming, so the industry is looking for semi automatic and automatic solutions. The working principle of automatic method basically is with semantic knowledge and text similarity calculation as the foundation, by calculating the concept similarity and similarity to achieve integration of original instance.

Semantic annotation is the process based on ontology annotation concept class, attribute and other metadata for cyber source and each part, is the foundation of semantic inference. Semantic annotation of the input is to be annotated document and Ontology (ontology is domain ontology user defined and can also be an upper ontology general), the output is the result of semantic description of the document, namely semantic content of Semantic Content. Add semantic information in the traditional Web, the Web state from machine readable to machine understandable, is the basis for the realization of the semantic Web. The development of semantic annotation technology will promote the practical application of semantic Web. This paper presents research on semantic annotation based on ontology mapping and merge.

2. Ontology Mapping and Merging based on Rough Concept Lattice Isomorphic Model

In the study of domain ontology construction, ontology based on UNSPSC core, on the one hand into the field of expert knowledge, on the other hand, considering the existing of structured and unstructured data (relational database, business forms, access log etc.), the model building method semi automatic extension of the core ontology. In the research of ontology mapping, the rough concept lattice model research, introducing the trapezoid fuzzy number similarity calculation of Complex Multi Strategy between ontology concepts, early focus on previous achievements on similarity in fuzzy decision problems, and the purpose is to improve the matching effect.

Ontology merging method based on rough concept lattice isomorphic generating: its core technology is to be merged ontology into attribute order independent semantic ontology background block number less, according to the similarity between the positioning module to module, according to the word UNSPSC in the hierarchy to ontology merging module.

2.1. Research on Rough Concept Lattice Isomorphic Model

The basic principle of the concept lattice isomorphic generating basis is isomorphic to the background of the concept lattice is isomorphic, *i.e.* based on concept lattice isomorphic background can generate new concept lattice. The method requires the presence of homogeneous background, must the system as background, the isomorphism identification become the core problem of this method is to consider, therefore, the variable precision rough reduction of formal context, analysis the isomorphic model based on rough formal context.

Concept lattice and rough set is analyzed and processed two powerful tools for data mining, this project will be the reduction in variable precision rough set of β , distribution reduction method is applied in the form of background, proposes to construct the concept lattice model based on variable precision rough set [2]. Its core idea is the variable precision rough set strong ability on attribute reduction for the reduction of concept lattice, which enhances the number of nodes in concept lattice generated is greatly reduced, the robustness of the system and the ability of resisting noise.

Approximate reduction is to consider the classification accuracy exist β ($0.5 < \beta \leq 1$), based on the quality of approximation classification standard of attribute reduction. For a decision table $S = \langle U, C \cup D, V, f \rangle$, which U is domain, C is a set of conditional attribute, D is the decision attribute set, is defined as follows:

Definition 1: $RED^\beta(C, D)$ approximate about called, its definition is: for a given value of $\beta \in (0.5, 1]$, assume: $Z \in E(D)$, $P \subseteq C$, satisfy the following two conditions.

$$\gamma^\beta(C, D) = \gamma^\beta(RED^\beta(C, D), D) \quad (1)$$

For the definition of the decision.

$$\beta = \{ \beta(X_i, Y_j) \mid 0 < i \leq U/C, 0 < j \leq U/D, \beta(X_i, Y_j) > 0.5 \}, \text{ for } \forall \beta \in (\beta_i, \beta_{i+1}],$$

$$\gamma(C, D, \beta) = \gamma(C, D, \beta_{i+1}), \beta^p : (\beta_{1,1}^p, \beta_{1,2}^p], (\beta_{2,1}^p, \beta_{2,2}^p], \dots, (\beta_{|\beta^p|,1}^p, \beta_{|\beta^p|,2}^p] \quad (2)$$

Divide the process of an arbitrary background includes two parts, object classification and attribute partition. Object division in the calculation of each object (for) the weight basis, according to the weight relationship will object set into equivalence classes, and makes the object equivalence classes according to their weight in descending order. In order to determine the uniqueness of the background form, in the transformation process, always want to maintain an orderly equivalence class. Similar attribute partition. So, the heavy division background object equivalence class and attribute equivalence class is divided into multiple sub background, only the K1 and K2 sub background intuitive algorithm for decision. Because of the heavy division, the complexity of the algorithm will be greatly reduced.

$$R_2 = R_1 \cup \{c_j\} \rightarrow \min(\Delta \gamma_{P \cup C}^\beta(R_1 \cup \{c_j\}, D), c_j \in C - R_1) \quad (3)$$

First calculate the semantic similarity, with the help of WordNet, by using the calculation method of concept similarity is proposed in this project, the similarity calculation between ontology concepts are property attributes, mapping unit set, then according to the areas with similar element relation of ontology, ontology merging method based on rough concept lattice isomorphism generated by this project is put forward, the unit of ontology merging, and finally to UNSPSC ontology element model as the core ontology, combined with expert knowledge, through the method of core ontology semi automatically extended extraction and the establishment of production ontology full Bureau.

In order to verify the proposed algorithm, developed the concept of variable precision rough set based on the generation system, visual interface, each step of constructing concept lattice directly displayed to the user [3]. In order to show more image hierarchy of concept lattice, this system can realize the layout strategy based on three dimensional spaces to show the relationship between concepts in concept lattice.

2.2. Analysis of Ontology Mapping based on Concept Similarity

Ontology mapping (ontology mapping): an ontology mapping M is a set of two OS Ontology (ontology) and OT (target ontology) (released) semantic overlap between specifications. This mapping can be unidirectional or bidirectional. In an injective mapping rules and it is how to use OS in terms of the expression of OT in terms. In a bijective mapping, the expression of OT in terms of the OS in the use of terminology, and can act in a diametrically opposite way.

This paper introduced the trapezoidal fuzzy numbers fuzzy similarity, put forward the distribution mapping method, combined with attribute and Multi Strategy at the same time, in order to improve the effect of sample space mapping by using variable precision rough set model fitting attribute deletion.

Ontology mapping / fusion (Specifying Mapping/Merging) is found between ontology (potential) similarities, mappings between ontology need to be specified. The formulation of the standard is usually rely on manually, but can be by PROMPT tools. For example, come up with specific suggestions for fusion operation, so that in many operations, the user only need to say "executive" can, instead of having to explain the complete operation.

Ontology mapping is to establish the mapping relationship between the two ontology, so in the map before, we must check the correlation between the two ontology, if not two ontology,

do not have to spend time mapping between them in the calculation, if the two ontology is only a part of the described in the field, it only needs to establish mapping relations between relevant parts.

The similarity between ontology concepts, needs many ontology model in the complex feature information into consideration, and the method of fuzzy similarity to the said, according to the characteristics of different information will be a series of expressed by different types of fuzzy similarity calculation between results obtained, these intermediate results need to be unified fuzzy similarity definition and description method, the fuzzy similarity in different types of calculation indicates that the intermediate results are transformed into fuzzy similarity unity, then the comprehensive weighted, non fuzzy operation, finally get the composite similarity between concepts of heterogeneous wooden ontology [4].

Based on the concept of E1 and E2 all attributes (I1, I2) for Cartesian product, namely P (I1, I2) = {<a1, b2>..... Something an, bn>}, and two two one groups of similarity calculation, and each will be similar objects and attributes than by weight (W), calculate concept similarity.

$$x_1^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t) * x_i^1(t) - s_0 \alpha N^1(t) \quad (4)$$

Firstly, it is the input Ontology (Import of ontologies). Ontology can be defined by a variety of language. This means that, in order to illustrate the mapping, the need to convert them to a common format. In addition, the ontology needs to be implanted tools; these tools are used to specify the mapping.

Secondly, it is looking for similarities (Finding Similarities). Some systems use the matching operator (Match operator), similarity in schema or ontology (semi) automatic search.

$$Sim((R_1, O_1), (R_2, O_2)) = \frac{|R_1 \cup R_2|}{o} \times (\beta + \lambda) + \left[\frac{\beta - 1}{R} \min \left(\sum_{R \in p(O_1, O_2)} ms(R, O) \right) \right] \times \beta \quad (5)$$

Name based mapping strategy is mainly based on the semantic similarity. Similarity between ontology and the element name (using WordNet as the vocabulary, the edit distance calculation of the two words in the dictionary to calculate semantic similarity); structure based mapping strategy, the main idea is that adjacent nodes two concept node is similar, they also tend to be similar, the similarity flooding algorithm to calculate similarity of ontology elements in the structure; example based mapping strategy is to calculate the similarity between ontology elements examples from the perspective of text types, properties of the naive Bias classifier to calculate the similarity, the attribute value types using KL distance the similarity calculation, categorical attributes (properties with two or more than two attribute values) using mutual information to compute similarity.

In this paper, using the characteristics of rough set and concept similarity algorithm was improved on the existing ontology mapping approaches combined with semantic attributes and objects, the concept of attribute similarity in semantic, analyzed the properties of ontology representation and similarity method; then do further processing of the form background with the improved algorithm, the to retain the information more complete; the corresponding extension of the connotation of the concept of similarity to measure the similarity between ontology concepts.

2.3. Ontology Merging Strategy based on Concept Lattice Isomorphism Theory

A Method for ontology merging is obtained after mapping, according to the application needs to perform the mapping ontology merging, and post treatment after execution. This paper considers two kinds of ontology merging method is promising is based on category theory (Category) and the method based on formal concept analysis (Formal Concept Analysis FCA) method. Category theory and FCA are abstract processing structure and structure relationship theory. Among them, FCA is only suitable to deal with lightweight ontology.

Definition 1: Two the concept of ontology of is-a may exist in two forms in the background, such as the two concepts of M, N, m is-a n, in the form of juxtaposed background appears only n. If $(g, m) \in I$ and M is-a n means: $(g; n) \in I$.

Definition 2: If the two formal context has two different names but the meaning of ontology concepts such as: a, b the same, this is because the domain experts from different concepts of the different naming methods cause. You can express the concept with "a\b" in the juxtaposition in the formal context, which means that a or b.

Definition 3: If there are two names, two forms the background meanings are different, but their distribution from the field in the document. Such as: P, Q, the document does not provide an adequate description of the ability to distinguish these two concepts, then in the juxtaposition in the formal context to P (q) to indicate, meaning P and Q.

Definition 4: Let $(A1, B1)$ and $(A2, B2)$ are two concepts a background, if $A1 \supseteq A2$ (equivalent to $B2 \supseteq B1$), then $(A1, B1)$ is called the $(A2, B2)$ sub concept, $(A2, B2)$ is called the $(A1, B1)$ - the concept, and denoted by $(A1, B1) \supseteq (A2, B2)$. Relationship is called the concept of partial order, and (G, M, R) the set of all concepts in this manner orderly representation for $B(G, M, R)$, Also known as the background (G, M, R) concept lattice.

Here the different magnitude of ontology, take a different approach. The lightweight ontology then combined with FCA in combination with Rough Set combined technology of ontology; if it is heavyweight ontology is the ontology data volume is huge, the project proposed ontology merging method based on rough concept lattice isomorphic generating.

First introduced the ontology merging based on rough concept lattice isomorphic generating [5]. The process is:

Step1: the background is decomposed into several appropriate scale sub background;

Step2: for an arbitrary background $K1$, system in the background database to check for the presence of homogeneous background, if it exists, is set to $K2$, the mapping between the two in the mapping database, otherwise it will be $K1$ into the background library, constructing the concept lattice $B(K1)$ and stored in the concept base;

Step3: according to user needs to form the ontology, this step is mainly the use of mapping and isomorphic lattices library $B(K2)$, $B(K1)$ isomorphic generating, then after the necessary reconstruction or combined to generate a complete ontology;

Step4: the user cans the ontology view for browsing or subsequent processing in the interactive interface (such as rule extraction). Distribution mapping method combined with attribute and multi strategy is proposed, with sub ontology B all (O_i) , finally, according to the ontology merging method based on FCA to get ontology B end (O) .

Definition 5: Background (G, M, R) a formal concept is an ordered pair (A, B) , where $A \subseteq G$, $B \subseteq M$, and meet: $A' \text{ and } B' = A = B$. Where A, B are called formal concept (A, B) epitaxial (extent) and content (intent). (G, M, R) (G, M) indicating the background, R the set of all formal concept. Without confusing situation, the following sections of this article will be referred to as the concept of formal concept.

The core idea is: for any ontology background O , can be decomposed into attribute order less ontology background O . Variable precision rough set attribute reduction, distribution of each sub ontology O_i , in the same order ontology context kernel exists in the isomorphic element H_j , and then based on the ontology structure of H_j concept lattice of the derived O_i , O_i ontology B thus obtained (O_i). Finally, it is combined with all of the B (O_i) B (O).

Theory basis: ontology merging can be applied to the ontology construction process, a plurality of small ontology into one ontology, also can be applied to ontology maintenance phase, small ontology integration to the original ontology of new knowledge is produced in the field. In addition, ontology merging is seldom used.

(1) Small ontology background ontology background larger into attribute order less;

When the background is large in scale, constructing the concept lattice will spend a lot of time [6]. Therefore, in addition to the reduction of the background, also need the background into smaller parts, respectively to construct the concept lattice to form a small ontology, finally merge into big ontology final. Background is the decomposition of the object set into subsets or attribute set into subsets, and the corresponding relationship between decomposition to complete. More generally, to investigating the relationship between the background and any of its sub background, as well as the relationship between the concepts lattice.

$$D_2^{*\beta} = \{([x]_C, [y]_C) : G_A^\beta(x) \neq G_C^\beta(y)\} \quad (6)$$

Generally speaking, can also decompose the object set and the attribute set to complete decomposition of the background, but for the convenience of study, we only consider the attribute set decomposition. There are four reasons: first, the practical application of users for data attribute often expressed concern degree, good stability, does not need the reduction in general; secondly, the lower order context kernel size is smaller, the background to quickly locate in the same order kernel isomorphic background; third, attribute set decomposition, the "a simplified" background will get good results; fourth, background has duality (Duality), attribute set decomposition is also applicable to the object set decomposition.

Definition 6: For the two ontology information (G_1, M_1, R_1) and (G_2, M_2, R_2) , if there exists $\alpha : G_1 \rightarrow G_2$, $\beta : M_1 \rightarrow M_2$ allows for any $g \in G_1$, $m \in M_1$, $gR_1m \Leftrightarrow \alpha(g)R_2\beta(m)$ was established, then we call these two ontology isomorphism.

(2) Variable precision rough set attribute reduction, distribution of each sub ontology background O_i .

Every finite background can be in the precondition of not changing the structure of the corresponding concept lattice under the transformation to a reduction, the reduction is the only (in the isomorphic premise), namely the standard background. Based on the forming principle concept, method to obtain the standard background is based on the method of variable precision rough set attribute reduction.

For background nuclear base in domain ontology, and find the ontology element and the isomorphic H_j , in the same order ontology background nuclear therefore, need to be further simplified n order context base [7]. Enough for (n) of a subset, can express use it (n) , which is (n) in each background concept lattice are concentrated and the son of a background of the concept lattice isomorphism.

Merging method using mapping method of fuzzy similarity and RFCA for ontology merging, according to ontology structure of H_j concept lattice of the derived O_i , O_i ontology B thus obtained (O_i). For the two to be merged ontology in the system of sub ontology mapping method, joint distribution according to the property of multi strategy

presented in this project, combining the RFCA-Merge method with the entire sub B (O_i), get the ontology the final B (O).

It is for background core library, and finds the ontology element and the isomorphic H_j in the same order context kernel in Ontology.

Step1: N attribute called n order background. Obviously, for a given n, all the n order background constitutes a is an infinite set, denoted as (n) . In order to facilitate the study of the nature and relationship between the background, from the (n) selected Co. a "representative" background. (n) for the n order background base ($n \geq 1$), there $\| (n) \| = 2^{2^n}$.

Step2: When the n increases, 2^{2^n} will rapidly become astronomical, this difficulty will make the relevant algorithms and techniques have little practical value, obtained by the finite background benefits will only stay in the theoretical sense. Therefore, it is to simplify the n order context base. We hope to be able to find out (n) of a subset, can express use it (n) , which is (n) in each background concept lattice are concentrated and the son of a background of the concept lattice isomorphism.

$$\phi(x) - \phi(a) = \phi'(\eta)(x - a) \tag{7}$$

Step3: The (n) is introduced, and then the reduction after the background according to the isomorphic relation between the set of equivalence class partition, take a background from each equivalence class. We can get the background of nuclear (n) , here is the definition and background of nuclear. This definition is not only ensure the nuclear background without redundancy, and ensure the completeness of the nucleus. N order context kernel is composed of n order background based simplification, elements of which two are not isomorphic, certain elements and arbitrary elements of order n background medium will be isomorphic to n order context kernel. In other words, given an arbitrary n order background K_i , will find $K_j \in (n)$, which $\underline{B}(K_i) \cong \underline{B}(K_j)$. Then, you can use a smaller (n) represents the infinite order n background.

Step4: According to the ontology structure of H_j concept lattice of the derived O_i, O_i ontology B thus obtained $\underline{B}(O_i)$. Finally, according to the concept hierarchy of vocabulary thesaurus is to merge all of the $\underline{B}(O_i)$ B (O).

[Is $X_i \subseteq R_p^\beta(Y_j)$?] If Yes then $X_i \cup M_1[i] \rightarrow M_1[i]$, go to Step3 to check the next Y_j .

Concrete is an element, either the n order background will be isomorphic to n order context kernel in it; either the n order background can be an element in the equivalent n order context kernel representation [8]. Because the concept lattice is isomorphic context isomorphism, according to the isomorphic mapping, concept lattice can generate the n order background with the concept lattice nuclei in the background, and then the concept lattice into a corresponding ontology, finally use lexical relationships in the thesaurus to ontology concept, from the ontology of concepts and their relations to obtain after the merger, finally get the merged Ontology.

The combined method of lightweight ontology, this project proposed merger based on rough concept lattice and ontology technology to. Methods: (1), first identified to be combined or cross domain ontology is the same field, for the two no crossover between fields, with little significance; (2), analysis of the corresponding two ontologies in the form of background, ontology merging key for the same field is the corresponding form background apposition, and Ontology merging for cross domain is the formal context apposition and overlay processing; (3), the combined formal context, the context of attribute variable

precision rough set, beta distribution and the corresponding identification matrix reduction; (4), aiming at the reduction of formal context, making use of the lattice construction algorithm of concept lattice, then in accordance with the combination of ontology and concept lattice, the concept lattice into ontology, namely the merged ontology. The whole processes of using FCA technology combined with the ontology such as a whole are described.

3. Semantic Annotation based on Rough Concept Lattice Model

Semantic Web, semantic Web content generation machine understandable semantic annotation and become a hot research field in. In the context of the semantic Web, many scholars are defined from different angles on the semantic annotation. From the perspective of ontology, semantic annotation is the ontology based metadata; semantic annotation based on Web content of the semantic category and semantic category relationship ontology based on semantic tagging is described; the Web content annotation ontology information.

Manual annotation as the name suggests is a specialized personnel as the core of tagging, tagging is determined using what kind of ontology personnel needs, how to parse Webpage content structure, choice of metadata elements, a semantic data RDF or HTML language tag. Manual annotation is intelligence intensive and labor-intensive process, only suitable for small range of Webpage annotation, facing the vast sea, changing cyber source, manual annotation is difficult to it. The manual annotation accuracy is generally higher than the automatic annotation, because, manual annotation is completed by the familiar ontology and professional knowledge.

From a labeled object is currently research focus in image semantic annotation method, annotation, semantic information of multimedia Web data annotation method; the first design of an ontology based knowledge management system, it can be on the cognitive map labeling, and put forward an ontology knowledge map framework, shared cognitive map collaborative tagging. By using ontology cognitive map warehouse; then using extended ontology, especially the OWL-S service ontology, semantic annotation to solve problems; finally, the automatic induction of semantic role labeling methods projection to overall framework Tagging Based on this framework, using syntactic and semantic information model, some contents which is semantic annotation it.

The field of document type definition (DTD) and document model of concept mapping and tagging. Because the DTD or Schema defines the document content structure and the content of element SGML and XML documents, we can establish mapping between certain concepts and specific DTD / Schema, which will automatically SGML / XML documentation for the DTD / schemR content element tags into the corresponding concept of metadata annotation. But due to the DTD / Schema application field and its architecture are often not coordinated is not compatible, so it is difficult to accurately map, also need to manually review and revise. The Semantic annotation based on ontology mapping and merging Structure diagram is shown by Figure 1 [9].

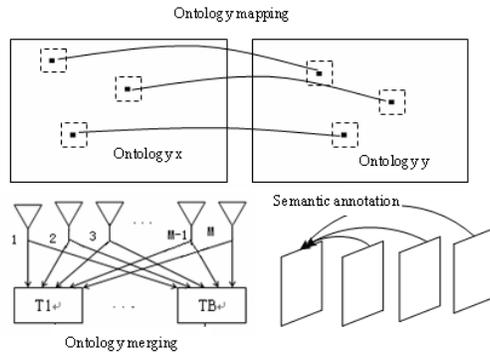


Figure 1. Semantic Annotation based on Ontology Mapping and Merging Structure Diagram

Existing problems: in the aspect of semantic annotation, supervision system need to learn from user labeled examples, but collect enough training examples is a very difficult task, unsupervised system uses a multi strategy learning how to mark, but its accuracy is limited.

In the semantic Web, Webpage and annotation information should be stored separately, because Webpage and annotation information belonging to different individuals or organizations, are put in different places. So the system uses storage method of philosophy knowledge engineering marking information, the tagging information and the original separation, which can not only share the annotation information, convenient document and annotation information updates, and provide the foundation for semantic retrieval. If the annotation information, this paper uses three tuple: $(R) = (R, D, O)$, B represents Webpage resources, D represents the ontology, E represents the concept ontology.

Semantic concept lattice model annotation will mark problems as classification problems based on the use of statistical learning methods, (such as: support vector machine Support Vector Machines (SVMs), concept lattice, etc.) to learn annotation model. The rough concept lattice is as an example to introduce the classification model team.

Consider two class classification problems in the first place. To make $\{(XL, YL), \dots, (xN, yN)\}$ represent the training data set, wherein X_i represents a training sample (usually represented as a feature vector); $Y_i \in \{-1, +1\}$ categorization is two class problem. Classification of formal description is as follows:

Classification problem: given a training set of $T = \{(XL, YL), \dots, (xN, yN)\}$ $(X \times Y) \in N$, $X_i \in X = R^n$, $Y_i \in Y = \{-1, +1\}$, $i = 1, \dots, N$, classification learning goal is to find the $X = R^n$ of a real valued function $g(x)$ in order to use the decision function of $F(x)$ can be inferred from any sample x corresponding y value. The decision function $f(x)$ is defined as follows.

$$F(t) = p(T \leq t) = \int_0^t u e^{-ut} dt = 1 - e^{-ut} \quad (8)$$

Based on rough concept lattice model is to construct ontology. Variable precision rough set attribute reduction of β selection algorithm and the identification matrix algorithm is improved, the beta reduction, distribution reduction method is suitable for formal context, semantic similarity calculation of the concept, classification and ontology information to rough concept lattice model, provides the classification basis for semantic annotation.

The measure to be labeled a three in a set to affect the overall annotation can be considered from two aspects: 1, three tuples themselves difficulty, represented by the symbol. This is mainly from the three itself into consideration, reflect the individual characteristics of each tuple in the collection; the impact on the set of three tuples, other 2 mark this three, represented by the symbol trC. This can be calculated through the three in the center to be annotated in the collection to reflect the degree of. For the sake of convenience, we first give some formal definitions, and then given it.

This paper puts forward a method of learning semantic annotation ontology based on semantic information to mining in social tagging implication, put forward a kind of implicit subsumption hierarchy to describe the structure underlying the label space, and this model is derived from the ontology based learning algorithm.

The concept of contact is a classification to semantic concept, because ontology describes the concepts in a field, the relationship between the concepts, characteristics and attributes of concepts, find out the concept connection to use ontology, as is shown by equation (9)

$$\mu(x, y) = e^{\int \phi(y) dy} = e^{\int -\frac{2}{y} dy} = y^{-2} \quad (9)$$

The core idea of this mark is: use classification model of rough concept lattice can learn the characteristics of distribution of label information itself, can also learn the distribution of label information and context boundary (usually including: start and end boundary), so as to realize the automatic identification label information. Mainly for the candidate labels for feature extraction, and then use the model training discrimination each candidate whether the object is labeled examples [10]. Finally, using two characteristic models distinguish each candidate is started to label the instance boundaries and end boundary; the best will begin to identify the boundary and the boundary between the end of the text as the annotation results.

Automatic semantic annotation main thought origin of the information extraction technology, realize the automation mode can be divided into four categories. The most basic way is to use the hand crafted rules or wrapper to capture and annotation information corresponding to the model. To realize the tagging research includes two kinds of annotation system through learning: supervised and unsupervised system. A supervised system marking information learning from user labeled examples. But for the supervised method, collect enough training examples, construct the set of training examples is a difficult task. Semantic annotation based on rough concept lattice model is as be shown by Figure 2.

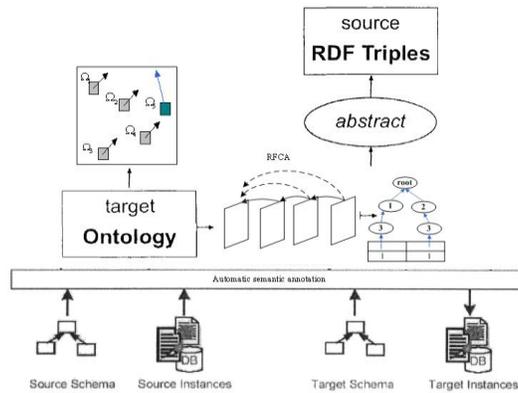


Figure 2. Semantic Annotation based on Rough Concept Lattice Model

The core idea: first acquiring knowledge of this field, in the cooperative expert and founder of the definition of ontology, domain ontology and create characteristic words from the document; then selected to represent the domain content, form the characteristic set of words; the containing these features vocabulary document and corresponding characteristic vocabulary related vocabulary, features associated with the document, combined with domain ontology, according to rough concept lattice based on domain ontology model method (RFCA isomorphic) storage, form the characteristic vocabulary set and its weight; thus, established the concept mapping between the ontology between the document and field.

With the traditional classifier is different, in the training process, SVM is trying to find an optimal separating hyperplane, which can separate the maximization of these two types of training samples (more precisely, should be between a maximum two sample boundary), this hyperplane corresponds to a linear SVM classifier (i.e.). Another focus on automatic annotation is tool case. The system of RFCA based text annotation method, using the similarity function to identify the sample context similar entity and annotation. For large-scale automatic annotation, this method is very suitable, as follows.

1) The word out markup document corresponds to the concept of ontology concept instance, as the corresponding, usually with RDF (RDF resource) form.

2) Find these examples exist and the corresponding attribute relationships in ontologies, usually between two examples and a case of association between expressed as (R1, P, R2), a RoF statement (RnF statement). P 2 instances of R1, relationship among R2, corresponding to the ontology of an attribute for any ontology V an element of V, $v_* := \bigvee \{x \in V | x < v\}$ definition and $v^* := \bigwedge \{x \in V | v < x\}$. If $V \neq v_*$, then V is called - not reduction (pronounced "Supremum can not reduction"), if $V \neq v^*$, then V is called - not reduction (pronounced "infimum can not reduction"). Using J (V) denote the set of all - irreducible elements, using M (V) denote the set of all - irreducible elements.

The semantic annotation idea is divided into two main steps: first, the characteristics of pattern extraction: second, according to the extraction rule generation model and mark.

Semantic annotation method based on rough concept lattice, this method can effectively annotation data between the level of dependency, the rough concept lattice based on support vector machine can effectively improve the accuracy of tagging, but it cannot be the brotherhood annotation. The syntax in the concept lattice isomorphic dependency increases to the adjacent edge support tree can be achieved on the annotation of long distance dependencies, can improve the accuracy of annotation.

4. Using Rough Concept Lattice Isomorphic Model to Semantic Annotation of Ontology

In the semantic annotation study, the focus is the field Webpage pretreatment, selection of Webpage classifier, Webpage extraction duplicate removal, such as denoising to research, using edit distance to conduct research in the rule mapping entropy.

Rule learning method for labeling the training data given automatic tagging rules based on learning, and then learning the rules for the new document annotation. Rule learning method reduces the manual definition rules difficulty, can to some extent to meet the needs of large-scale semantic annotation.

User driven automatic semantic annotation tool, it provides two main strategies for the user annotation [11]. On the one hand, it provides an adaptive information extraction system, through the generalization of user annotation information to learn how to label file. Under this strategy, tagging becomes such a process: in the beginning of all labeling need user involvement in tagging, tagging and in later, the user's work has been simplified to simply label recommendations on the verification system. On the other hand, it also provides the ability to write a regular expression based on rules, this allows advanced users can self define the extraction rules.

Current information extraction using the classification model of a lot of research work, but the classification model of semantic automatic tagging research is not many. Here the design of a rough concept lattice model automatic marking system of document types, semantic category which use classification learning model identification document (similar to the text automatic classification), and then using statistical learning method for identifying named entities in the Web pages, by the end of the annotation knowledge analysis don't results and the realization of the Web page entity relationship, as is shown by equation (10).

$$x_i^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t) * x_i^1(t) - s_0 \alpha N^1(t) \quad (10)$$

The rough concept lattice model is extended here unsupervised RDF knowledge base. Their method is through the use of name classes and sub classes in ontology to construct the sample, then to learn, in order to solve the lack of collection of tagged training examples of problems. Constructed sample, sample the context will also be learning. By similarity rough concept lattice isomorphism concept, some examples and the ontology inconsistent terminology can be identified.

Ontology background (G, M, R) is known to be clear (Clarified), if $\forall g, h \in G, g' = h'$ total $g=h$ was established; the same for $m, n \in M, m' = n'$ total $m=n$ was established.

The RDF framework is open, it uses Client/Serve: model will be shared Web page tagging information stored in the annotation server, which makes application of tagging content will be more extensive and flexible. System structure based on component, it by adding AJEP (Java Annotation Patterns Engine, for tagging provides regular expression patterns / behavioral rules based on Annotation and Diff tools).

Step1: First select the V glossary frequency highest w as keyword, generation mode is added to the set P.

Step2: According to the definition of 3 from V W and remove the words associated with W. Related variables defined in the following algorithm:

V={ (W, f) | w for the sample set of vocabulary, f to w frequency }

P={ (p, s) | P three tuple (prefix, keyword, suffix), s support P (support) }

Step3: Let $K := (G, M, R) \in (n)$, wherein the attribute set $M = \{M_1, m_2, \dots, M_n\}$. For any object $G \in G, G \in P (M')$, $P (M)$ that the power set of M

Step4: Take the same number of all content in L lattice node C into the same group, denoted as $B[i] := \{C: ||Intent(C)||=i\}; size := \max(i)$

Step5: Deep Web labeling by tagging the mapping between database schemas and ontologies to achieve Deep Web. So the user can establish a query conditions directly query the database data based on ontology.

Step6: Set up a mapping between heterogeneous ontologies $\{I_1, I_2, \dots, I_n\}$, feature involved the collection. Ladder which are corresponding to the mapping characteristics

of fuzzy weight $\{\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_n\}$. After the non fuzzy, determine the weight set $\{F_3(I_1), F_3(I_2), \dots, F_3(I_n)\}$. The absolute weight of any non fuzzy value $F_3(I_i)$ $1 \leq i \leq n$.

Step7: According to the ontology structure of Hj concept lattice of the derived Oi, Oi ontology B thus obtained (Oi). Finally, according to the concept hierarchy of vocabulary thesaurus to merge all of the B (Oi) B (O).

Step8: Semantic annotation, two nested rough concept lattice isomorphism elements Cl and CZ at the same time take verb role. We use the method of dynamic programming, from the bottom to the syntax tree leaf node to the root node end; calculation of the current syntactic component is not block root formation. Tree and current syntactic constituents were empty and non empty role of probability and, so as to achieve semantic annotation of Chinese vocabulary.

5. Experiments and Analysis

This paper chose to test under closed conditions, data we used CoNNL 2012 as training, provide development and test set, the acquisition of electronic commerce a portal through the web crawler (website includes concept and various electronic commerce example) containing detailed product information and the corresponding web portal web site. In order to reflect the layout structure of an HTML document complex and rich in content characteristics, this paper extracted from sites with data. In the sample data set, the corresponding Ontology Conceptual HTML files, calculate 9 indices of each group Webpage, ingredients to portal Webpage with products and the corresponding Webpage.

The experimental hardware environment: Pentium (R) 4 CPU2.40GHz, memory 2GB; The experimental software environment: operating system using Windows 7, the algorithm is realized by Java language, running in the JDK1.6 environment, development platform using IBM's Eclipse open-source software, server using Bea, WebLogic10, and database system for SQL Server2008.

In order to eliminate the syntactic analysis and accuracy of the identification center word tagging, the effects on the semantic, syntactic provide rough concept lattice isomorphic model dependencies and the center for the word semantic annotation based on the experimental results, as shown below.

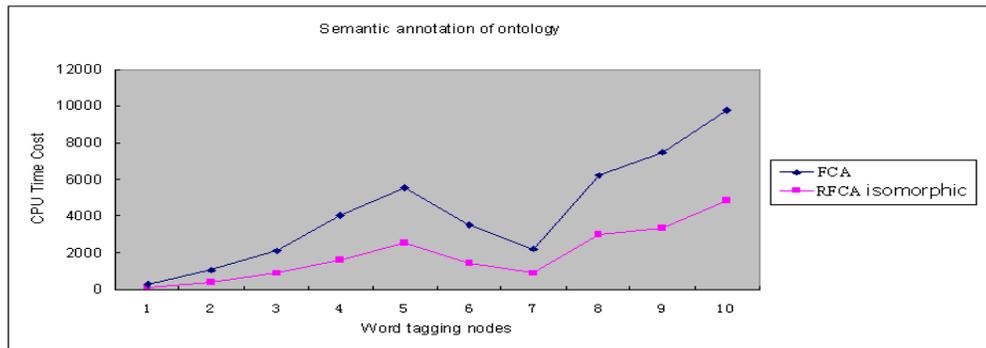


Figure 3. Experimental Results

We can see from the experimental results, the rough concept lattice isomorphic model semantic annotation relative to the concept lattice model is improved based on, mainly

reflected in the annotation on the relationship between semantic roles of internal structure; to avoid the formal concept analysis model will cause the syntactic structure tree splitting process of semantic role deletion. Semantic role labeling coarse lattice isomorphic model with respect to the accuracy of the concept lattice model based on semantic annotation is improved, by introducing the support vector machine (T), according to the F function importance evaluation time adjustment support, then based on the SVM function $f(T)$ score improved proposed similarity between users, from the experiment can be proved, linear classifier is obtained by maximizing the concept lattice is isomorphic with minimal generalization error. Linear SVM through the kernel function (such as: Gauss or polynomial kernel function) is extended to obtain nonlinear SVM considering influence of grammar and syntax of the user similarity, semantic is superior to the traditional in terms of accuracy of the labeling algorithm.

6. Summary

Automatic semantic annotation relative to the manual annotation can expand tagging; tagging can achieve a larger range of. However, semantic annotation is eventually completed in the ontology under the guidance of the general ontology, does not exist, the automatic semantic annotation can only be based on domain ontology, cannot be used for tagging applications. Automatic annotation to achieve a greater range of, first of all need to construct general ontology mass, but also to allow machines to establish the correct link automatic text in the instance of information and ontology, to the machine understand natural language level, to achieve this goal also requires researchers to continuous efforts, the proposed semantic rough concept lattice the isomorphic model based annotation is used for automatic semantic annotation of thought.

After semantic annotation corpus collection is mainly used for training in the various machine learning software, in order to complete the various research tasks, such as syntactic parsing, word sense disambiguation, co reference resolution, text summarization and information extraction. In addition, the semantic annotation of intelligent content creation include electronic dictionary, cross linguistic data, data knowledge expression etc..

Semantic annotation will be built in numerous small ontology, but also very obvious shortcomings of this view, a large number of small ontology is heterogeneous, which will cause the frequent treatment of heterogeneous ontology using semantic information in the process of operation, integration and expansion and ontology is a non process often difficult. Another view is that should be by a few large institutions to provide a comprehensive and universal large-scale ontology to guide people to mark, but this will cause the ontology structure of the monopoly, limiting the user freedom of expression, because we propose rough concept lattice isomorphic model based on mapping and ontology merging, so as to achieve semantic label effect, experiments show that, this method is better than the traditional methods in semantic annotation accuracy and breadth are superior.

Acknowledgements

This paper is supported by the National Natural Science Funds of China (61272015), and also is supported by the science and technology research major project of Henan province Education Department (13B520155) and Henan Province basic and frontier technology research project (142300410303).

References

- [1] C.-C. Yeh, D.-J. Chi and M.-F. Hsu, "A hybrid approach of DEA, rough set and support vector machines for business failure prediction", *Expert Systems with Applications*, vol. 37, no. 2, (2010), pp. 1535-1541.
- [2] C. Calero, F. Ruiz and A. Baroni, "An ontological approach to describe the SQL: object-re-lational features", *Computer Standards & Interfaces*, vol. 28, (2006), pp. 695-713.
- [3] R. Navigli, Velardip and A. Gangemi, "Ontology learning and its application to automated terminology translation", *IEEE Intelligent Systems*, vol. 18, no. 1, (2003), pp. 22-31.
- [4] H.-s. Xu, X.-j. Shen and Z.-t. Liu, "Construction and Presentation of Ontology on Semantic Web Based on Formal Concept", *Journal of Computer science*, vol. 34, no. 2, (2007), pp. 171-174.
- [5] J. J. Jung, "Reusing ontology mappings for query routing in semantic peer-to-peer environment", *Information Sciences*, vol. 180, (2010), pp. 3248-3257
- [6] L-D. Wang and X-D Liu, "Concept analysis via rough set and AFS algebra", *Information Sciences*, vol. 178, no. 21, (2008), pp. 4125-4137.
- [7] A. Maedche and S. Staab, "Ontology Learning for the Semantic Web Special Issue on the Semantic Web", *IEEE Intelligent System*, vol. 16, no. 2, (2001), pp. 72-79.
- [8] H.-s. Xu and R.-l. Zhang, "Application of ART and concept similarity in E-commerce recommendation systems", *Journal of Computer engineering and Applications*, vol. 46, no. 7, (2010), pp. 213-217.
- [9] Y. Y. Yao, "Concept lattices in rough set theory", In: *Proceedings of 23rd International Meeting of the North American Fuzzy Information Processing Society*, (2004).
- [10] R.-l. Zhang and H.-s. Xu, "Building and mapping ontology of e-business based on fuzzy rough concept lattices", *Journal of Convergence Information Technology*, vol. 6, no. 9, (2011), pp. 81-88.
- [11] J. J. Jason, "Ontology mapping composition for query transformation on distributed environments", *Expert Systems with Applications*, vol. 37, no. 12, (2010), pp. 8401-8405.

Authors



Hongsheng Xu, He was born on December 28, 1979.
Educational background: master, Henan University, Kaifeng, China, 2007;
Major field of study: data mining, Knowledge discovery, artificial intelligence, the Semantic Web..



Ruiling Zhang, She was born on December 23, 1964.
Educational background: master, Northwestern Polytechnical University, Xian, China, 2007;
Major field of study: data mining, Knowledge discovery, artificial intelligence, the Semantic Web.

