International Journal of Engineering Trends and Technology- Volume4Issue2- 2013

Ontology Learning Process Using Fuzzy Formal Concept Analysis

[1]J.JELSTEEN, [2] D.EVANGELIN, [3] J.ALICE PUSHPARANI, [4]J.NELSON SAMUEL JEBASTIN

[1]*Assistant Professor, Nehru Group Of Institutions, Coimbatore. Tamil Nadu , India*

[2]*Assistant Professor, Sri Vidhya College Of Eng. & Tech. Virudhunagar, Tamil Nadu , India*

[3]*Assistant Professor, Christ The King Engineering College, Coimbatore. Tamil Nadu , India*

[4]*Assistant Professor, Annamalai University, Chidambaram. Tamil Nadu , India*

ABSTRACT

Currently reliable and appropriate information is difficult to find on the Internet. Bayesian networks were used earlier for probabilistic reasoning of unknown values and for determining knowledge representation. Various probabilistic approaches were used to represent uncertainty information. Typically, fuzzy ontology is generated from a predefined concept hierarchy. However, to construct a concept hierarchy for a certain domain manually can be a difficult and tedious task. To tackle this problem, this paper proposes the FOGA (Fuzzy Ontology Generation Framework) for automatic generation of fuzzy ontology on uncertainty information. The FOGA framework comprises the following components: Fuzzy Formal Concept Analysis, Concept Hierarchy Generation, and Fuzzy Ontology Generation. We also discuss approximating reasoning for incremental enrichment of the ontology with new upcoming data. This project describes some evaluation of information retrieval system designed to support fuzzy ontology based search refinement. The objective is to implement generation and learning of knowledge representation using fuzzy logic and ontology for reasoning. Fuzzy logic can be incorporated to ontology to represent uncertainty information. Finally automatic fuzzy ontology generation is proposed for knowledge domains like semantic web.

KEY TERMS—Intelligent Web services and semantic Web, ontology design, uncertainty, “fuzzy,” knowledge representation formalisms and methods, concept learning.

INTRODUCTION

ONTOLOGY is a conceptualization of a domain into a human understandable, machine-readable format consisting of entities, attributes relationships, and axioms [1]. It is used as a standard knowledge representation for the Semantic Web [2]. However, the conceptual formalism supported by typical ontology may not be sufficient to represent uncertainty information commonly found in many application domains due to the lack of clear-cut boundaries between concepts of the domains. For example, a document can be very relevant, relevant, or irrelevant to a research area. In addition, keywords extracted from scientific publications can be used to infer the corresponding research areas. However, it is inappropriate to treat all keywords equally as some keywords may be more significant than others.

To tackle this type of problems, one possible solution is to incorporate fuzzy logic [3] into ontology to handle uncertainty data. Traditionally, fuzzy ontology is generated and used in text retrieval [4] and search engines , in which membership values are used to evaluate the similarities between the concepts in a concept hierarchy. However, manual generation of fuzzy ontology from a predefined concept hierarchy is a difficult and tedious task that often requires expert interpretation. So, automatic generation of concept hierarchy and fuzzy ontology from uncertainty data of a domain is highly desirable.

In this paper, we propose a framework known as FOGA (Fuzzy Ontology Generation frAmework) that can automatically generate a fuzzy ontology from uncertainty data based on Formal Concept Analysis (FCA) theory. The generated fuzzy ontology is mapped to a semantic representation in OWL (Web Ontology Language) . The rest of this paper is organized as follows: Section 2 discusses related work on ontology generation and FCA. Section 3 gives some basic definitions and operators of the fuzzy theory. The FOGA framework is presented in Section 4.Section 5 discusses the approximating reasoning technique to incrementally furnish the generated ontology with new instance. The problem of integrating extra attributes in database to the ontology is given in Section 6. Performance evaluation of the proposed FOGA framework is given in Section 7. Finally, Section 8 concludes the paper.

2 RELATED WORK

2.1 ONTOLOGY GENERATION

Although editing tools have been developed to help users to create and edit ontology, it is a troublesome task to manually derive ontology from data. Typically, ontology can be generated from various data types such as textualdata [5], dictionary, knowledge-based, semi structured schemata, and relational schemata. Compared to other types of data, ontology generation from textual data has attracted the most attention. Among techniques used for processing textual data, clustering is one of the most effective techniques for ontology learning. Conceptual clustering techniques such as COBWEB [6] and CLASSIT are powerful clustering techniques that can conceptualize clusters for ontology generation.

2.2 FORMAL CONCEPT ANALYSIS

FCA is a formal technique for data analysis and knowledge representation. It defines formal contexts to represent relationships between objects and attributes in a domain. From the formal contexts, FCA can then generate formal concepts and interpret the corresponding concept lattice, so that information can be browsed or retrieved effectively.FCA is widely used for various applications, such as text processing, ontology merging , e-mail manager ,e-learning, Web navigation, and expert system. However, as most concept lattices are quite complicated in terms of the number of concepts generated, it is necessary to simplify the lattice generated.

Traditional FCA-based conceptual clustering approaches are hardly able to represent such vague information. To tackle this problem, fuzzy logic can be incorporated into FCA to handle uncertainty information for conceptual clustering and concept hierarchy generation. Many have proposed the L-Fuzzy context as an attempt to combine fuzzy logic with FCA. The L-Fuzzy context uses linguistic variables, which are linguistic terms associated with fuzzy sets, to represent uncertainty in the context. However, human interpretation is required to define the linguistic variables. Moreover, the fuzzy concept lattice generated from the L-fuzzy context usually causes a combinatorial explosion of concepts as compared to the traditional concept lattice.

We propose a new technique that combines fuzzy logic and FCA as Fuzzy Formal Concept Analysis (FFCA), in which the uncertainty information is directly represented by a real number of membership value in the range of [0,1]. As such, linguistic variables are no longer needed. Compared to the fuzzy concept lattice generated from the L-fuzzy context, the fuzzy concept lattice generated using FFCA will be simpler in terms of the number of formal concepts. It also supports a formal mechanism for calculating concept similarities.

3. FUZZY THEORY

In this section, we review some fundamental l knowledge of fuzzy theory.

DEFINITION 1 (FUZZY SET). A fuzzy set A on a domain U is defined by a membership function μ from U to [0.1], i.e., each item in A has a membership value given by μ. We denote Φ(S) as a fuzzy set generated from a traditional set of items S. Each item in S has a membership value in [0, 1]. S can also be called as a crisp set.

DEFINITION 2 (FUZZY RELATION). A fuzzy set A on a domain G×M, where G and M are two crisp sets is a fuzzy relation on G;M.

DEFINITION 3 (FUZZY SETS INTERSECTION). The intersection of fuzzy sets A and B, denoted as A∩B, is defined by $\mu_{A\cap B}(x)=\min(\mu_A(x), \mu_B(x))$.

DEFINITION 4(FUZZY SETS UNION). The intersection of fuzzy sets A and B, denoted as AUB, is defined by $\mu_{AUB}(x)=\max(\mu_A(x), \mu_B(x))$.

DEFINITION 5 (FUZZY SET CARDINALITY). Let S be a fuzzy set

on the domain U. The cardinality of S is defined as S , $|S|=\sum \mu(x)$,where μ(x) is the membership of x in S.

DEFINITION 6 (FUZZY SETS SIMILARITY). The similarity between two fuzzy sets A and B is defined as $E(A, B) =|A\cap B|/|AUB|$

4 THE FOGA FRAMEWORK

Fig. 1 shows the proposed FOGA (Fuzzy Ontology Generation frAmework), which consists of the following components.

4.1 FUZZY FORMAL CONCEPT ANALYSIS

The Fuzzy Formal Concept Analysis incorporates fuzzy logic into Formal Concept Analysis to represent vague information. Fuzzy formal context can also be represented as a cross table as shown in Table 1. The context has three objects representing three documents, D1, D2, and D3. It also has three attributes, “Data Mining,” “Clustering,” and “Fuzzy Logic” representing three research topics. The relationship between an object and an attribute is represented by membership value in [0, 1].An ∞-cut can be set to eliminate relations that have low membership values. Table 2 shows the cross-table of the fuzzy formal context given in Table 1 with ∞- ¼ 0:5.Generally, we can consider the attributes of a formal concept as the description of the concept. Thus, the relationships between the object and the concept should be the intersection of the relationships between the objects and the attributes of the concept. Since each relationship between the object and an attribute is represented as a membership value in fuzzy formal context, the intersection of these membership values should be the minimum of these membership values.

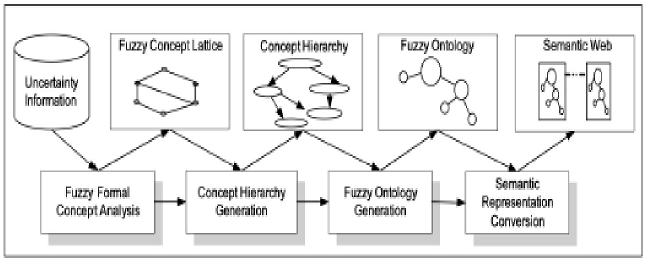

Figure:1 The FOGA framework

TABLE 1

A CROSS – TABLE of a FUZZY Formal Context			
D1	0.8	0.12	0.61
D2	0.9	0.85	0.13
D3	0.1	0.14	0.87

TABLE 2

FUZZY Formal Context in Table 1 With an a-cut a=0.5			
	Data Mining	Clustering	Fuzzy Logic
D1	0.8	-	0.61
D2	0.9	0.85	-
D3	-	-	0.87

In a formal context, a concept can have many superconcepts and subconcepts. However, the similarities of a concept to its superconcepts and subconcepts are different. Such information cannot be shown in a traditional concept lattice. With fuzzy concept lattice, we can make use of the fuzzy set theory to calculate the similarities between a concept and its subconcepts.Fig. 2 shows the traditional concept lattice generated from Table 1 without membership values. Fig. 3 shows the fuzzy concept lattice generated from the fuzzy formal context given in Table 2, in which the similarities between the concepts are given. Fuzzy formal concept lattice can provide additional information, such as membership values of objects in each fuzzy formal concept and similarities of fuzzy formal concepts, which are important for the construction of concept hierarchy.

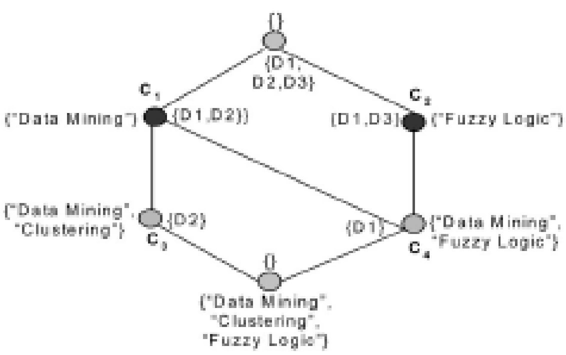

Figure 2: A concept lattice generated from traditional FCA

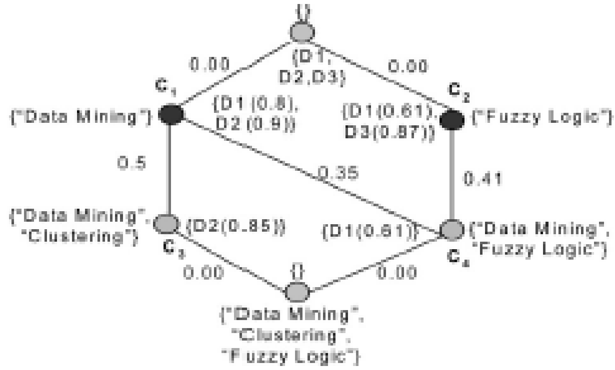

Figure 3: A fuzzy concept lattice generated by FFCA

4.2 CONCEPT HIERARCHY GENERATION

Concept Hierarchy Generation clusters the fuzzy concept lattice generated by FFCA to construct a concept hierarchy in the two following steps.

4.2.1 FUZZY CONCEPTUAL CLUSTERING

As in traditional concept lattice, the fuzzy concept lattice generated using FFCA is sometimes quite complicated due to the large number of fuzzy formal concepts generated. Since the formal concepts are generated mathematically, objects that have small differences in terms of attribute values are classified into distinct formal concepts. Such objects should belong to the same concept when they are interpreted by human. Thus, we cluster formal concepts into conceptual clusters using fuzzy conceptual clustering. Compared to traditional clusters, the conceptual clusters generated have the following properties:

1. Each conceptual cluster is considered as a human interpretable concept in the domain of the fuzzy concept lattice.
2. Each conceptual cluster is a sublattice extracted from the fuzzy concept lattice.
3. A formal concept must belong to at least one conceptual cluster. For example, a scientific document can belong to more than one research area.

Conceptual clusters are generated based on the premise that if a formal concept A belongs to a conceptual cluster R, then its subconcept B also belongs to R if B is similar to A. We can use a similarity confidence threshold Ts to determine whether two concepts are similar or not.Fig. 5 show the conceptual cluster generated from the fuzzy concept lattice given in Fig. 3 with similarity confidence thresholds Ts ¼ 0:4.

A conceptual cluster can be considered as a set of fuzzy formal concept. Each concept is associated with a set of objects and attributes. As such, each conceptual cluster can also be represented as sets of objects and attributes. Moreover, each object in each conceptual cluster should have a membership value implying the uncertainty degree of the fact "the object belongs to the conceptual cluster."

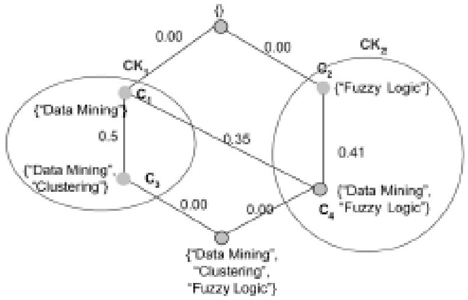

Figure 5: Conceptual clusters generated from Fig. 3 with confidence threshold Ts ¼ 0:4

4.2.2 HIERARCHICAL RELATION GENERATION

As discussed earlier, fuzzy conceptual clustering generates set of conceptual clusters SC. To construct a concept hierarchy from the conceptual clusters, we need to find the hierarchy relations from the clusters. Fig. 8b illustrates the hierarchical relations constructed from the conceptual clusters. Each concept in the concept hierarchy is represented by a set of its attributes. The supremum and infimum of the lattice are considered as Thing" and "Nothing" concepts, respectively.

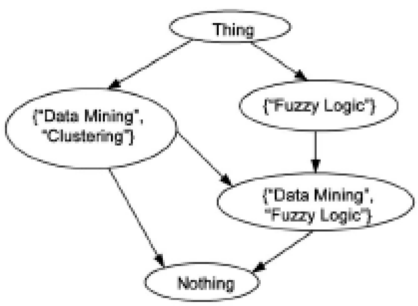

Figure 6: Concept Hierarchy

4.3 FUZZY ONTOLOGY GENERATION

Here, we construct fuzzy ontology from a fuzzy context using concept hierarchy created by clustering. This is done based on the characteristic that both FCA and ontology support formal definitions of concepts. However, a concept defined in FCA has both extensional and intensional information in a balanced manner, whereas a concept in ontology emphasizes on its intensional aspect.

To construct the fuzzy ontology, we need to convert both intensional and extensional information of FCA concepts into the corresponding classes and relations of the ontology. Thus, we define the fuzzy ontology as follows:

A fuzzy ontology FO consists of four elements ($C;A^C;R;X$), where C represents a set of concepts, A^C represents a collection of attributes sets, one for each concept, and $R=(R_T;R_N)$ represents a set of relationships, which consists of two elements: R_N is a set of nontaxonomy relationships and RT is a set of taxonomy relationships. Each concept ci in C represents a set of objects, or instances, of the same kind. Each object oij of a concept ci can be described by a set of attributes values denoted by $A^C(ci)$. Each relationship ri(cp;cq) in R represents a binary association between concepts cp and cq, and the instances of such a relationship are pairs of (cp; cq) concept objects. Each attribute value of an object or relationship instance is associated with a fuzzy membership value between [0,1] implying the uncertainty degree of this attribute value or relationship. X is a set of axioms. Each axiom in X is a constraint on the concept's and relationship's attribute values or a constraint on the relationships between concept objects. The constraints can be described using the SWRL format.

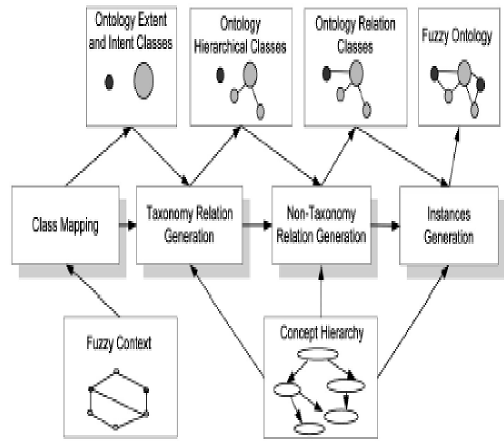

Figure 7: Fuzzy Ontology Generation

5. APPROXIMATING REASONING FOR ONTOLOGY ENRICHMENT

A common problem with ontology generation is how to incrementally deal with new data. As discussed earlier, apart from the fuzzy formal context, the fuzzy ontology is also generated from the concept hierarchy acquired using conceptual clustering. To perform conceptual clustering again to incorporate the new data into the ontology would be time-consuming. To avoid this, we propose to use the fuzzy-based approximating reasoning technique to assign new data into appropriate conceptual clusters, which consists of the two following steps.

5.1 PROPOSITION EXTRACTION

A proposition can be represented as a statement "x is A," where x is a variable and A is a value. In fuzzy logic, a proposition can be represented as a fuzzy set U, which implies "x is U." For reasoning, the proposition that has widely been used in fuzzy logic is the "IF-THEN" proposition, which can be represented as follows: IF <proposition> THEN <proposition>.

The aim of the IF-THEN proposition is to conclude if an object belongs to conceptual clusters, the THEN proposition should be a fuzzy set FS on the domain DC, where DC is the set of conceptual clusters. The problem is how to calculate the membership for each conceptual cluster Cj in DC when we construct the IF-THEN proposition from a certain conceptual cluster Ci. Since the membership value implies that "If an object belongs to Ci, then how much does that object belong to Cj," the membership value of Cj in the THEN part of the proposition

5.2 APPROXIMATING REASONING

After the proposition extraction step, we have a set of propositions as fuzzy rules. The next step is to use the generated rules for reasoning new data. For example, assume that we have a fuzzy rule "IF x is A THEN y is B," where A and B are fuzzy sets. Then, if we have a new proposition "x is A," we need to find what conclusion we can get about y. Theoretically, a proposition IF <FP1> THEN <FP2>, where FP1 and FP2 are two fuzzy propositions that can be interpreted as a relation connecting FP1 and FP2. In classical propositional logic, the rule "IF x THEN y" means "x implies y" .

6. INTEGRATING OF EXTRA ATTRIBUTES FROM DATABASE TO ONTOLOGY

In the previous section, we presented a technique for constructing fuzzy ontology from a fuzzy formal context. Such fuzzy formal context can be generated automatically from database schemata. However, apart from the attributes that are used in the fuzzy formal context, there are probably some other significant attributes available in the database. For example, besides the keywords, a document may have some other important attributes, or extra attributes, such as its authors, publisher, publication dates, etc. To make the generated fuzzy ontology more effective, it is necessary to integrate these extra attributes to the ontology. Thus, we propose a mathematical model to incorporate extra attributes into the fuzzy ontology generated using FOGA.

7 PERFORMANCE EVALUATION

7.1 GENERATING ONTOLOGY FROM CITATION DATABASE

To evaluate the proposed FOGA framework for ontology generation, we have collected a set of 1,400 scientific documents on the research area "Information Retrieval". The downloaded documents are preprocessed to extract related information such as the title, authors, citation keywords, and other citation information. The extracted information is then stored as a citation database.

7.2 PERFORMANCE EVALUATION OF ONTOLOGY GENERATION

Performance of the ontology generation is evaluated based on the generated Research Area Hierarchy. First, we measure the typical recall, precision, and F-measure to evaluate the clustering results. Second, we use the relaxation error and the corresponding cluster goodness measure to evaluate the goodness of the conceptual clusters generated. We also show whether the use of fuzzy membership instead of crisp value can help improve cluster goodness. Finally, we use the Average Uninterpolated Precision (AUP), which is a typical measure for evaluating a hierarchical construct, to evaluate the goodness of the generated concept hierarchy.

8 CONCLUSION

In this paper, we have proposed the FOGA framework for fuzzy ontology generation on uncertainty information. FOGA consists of the following steps: Fuzzy Formal Concept Analysis, Fuzzy Conceptual Clustering, Fuzzy Ontology Generation, and Semantic Representation Conversion. In addition, we have also proposed an approximating reasoning technique that allows the generated fuzzy ontology to be incrementally furnished with new instances. Finally, we have also proposed a technique to integrate extra attributes in a database to the ontology. The proposed FOGA framework would be useful to construct ontology from uncertainty data as it can represent

Uncertainty information and construct a concept hierarchy from the uncertainty information in automatically. Apart from constructing scholarly ontology from citation database as previously stated, FOGA has also been used to generate Machine Service Ontology for Semantic Help-desk

In addition, the scholarly ontology generated in Section 7 has been partially used to construct the Scholarly Semantic Web, a Semantic Web-based information retrieval system to support scholarly activities in the Semantic Web environment.

REFERENCES

1. N. Guarino and P. Giaretta, Ontologies and Knowledge Bases: Towards a Terminological Clarification. Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing. Amsterdam: IOS Press, 1995.
2. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific Am.,htttp://www.sciam.com/2001/0501issue/0501berners-lee.html, 2001.
3. L.A. Zadeh, "Fuzzy Logic and Approximate Reasoning," Synthese, vol. 30, pp. 407-428, 1975.
4. M.Z. Islam and L. Brankovic, "A Framework for Privacy Preserving Data Mining," Proc. Australasian Workshop Data Mining and Web Intelligence (DMWI '04), pp. 163-168, 2004.
5. B. Bachimont, A. Isaac, and R. Troncy, "Semantic Commitment for Designing Ontologies: A Proposal," Proc. Int'l Conf. Knowledge Eng. and Knowledge Management, pp. 114-121, 2002.Fuzzy Systems, pp. 1291-1294, 2001.
6. D. Fisher, "Knowledge Acquisition via Incremental Conceptual Clustering," Machine Learning, vol. 2, pp. 139-172, 1987.
7. W3C, "Web Ontology Language Overview, "http://www.w3.org/ TR/owl-features/, 2006.