A MODEL FOR REPRESENTING VAGUE LINGUISTIC TERMS AND FUZZY RULES FOR CLASSIFICATION IN ONTOLOGIES

Cristiane A. Yaguinuma, Vinícius R. T. Ferraz, Marilde T. P. Santos, Heloisa A. Camargo
Department of Computer Science, Federal University of São Carlos, Rod. Washington Luis Km 235, São Carlos, SP, Brazil

Tatiane M. Nogueira
Institute of Mathematics and Computer Sciences, University of São Paulo
Av. Trabalhador São-carlense, 400, São Carlos, SP, Brazil

Keywords: Knowledge representation, Fuzzy set theory, Fuzzy ontology, Fuzzy reasoning, Classification.

Abstract: Ontologies have been successfully employed in applications that require semantic information processing. However, traditional ontologies are not able to express fuzzy or vague information, which often occurs in human vocabulary as well as in several application domains. In order to deal with such restriction, concepts of fuzzy set theory should be incorporated into ontologies so that it is possible to represent and reason over fuzzy or vague knowledge. In this context, this paper proposes a model for representing fuzzy ontologies covering fuzzy properties and fuzzy rules, and we also implement fuzzy reasoning methods such as classical and general fuzzy reasoning, aiming to support classification of new instances based on fuzzy rules.

1 INTRODUCTION

Ontologies have been widely used in applications regarding knowledge representation, such as the Semantic Web, which investigates the semantics of content and services over the Web. In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse (Gruber, 2009). As ontologies model semantic elements and support reasoning, they have been applied to improve semantic information processing among humans and computational systems.

Although expressive, the background formalism of traditional ontologies is not able to represent imprecise or vague information, which often occurs in human language and in several application domains. For example, it is difficult to model concepts like *young*, *dark*, *hot* and *large*, for which a clear and precise definition is not possible (Straccia, 2006). Fuzzy sets (Zadeh, 1965) provide a meaningful and powerful representation of vague concepts expressed in natural language. Several fuzzy sets representing linguistic concepts, such as *low*, *medium* and *high*, are often employed to define states of a variable (*fuzzy variable*) (Klir and Yuan, 1995). Such variables are often used in production rules, which support knowledge inference based on fuzzy reasoning methods, e.g. classical and general methods (Cordón et al., 1999). Therefore, such rules can be used to infer the classification of elements into specific categories according to the values of fuzzy variables.

In this sense, it is important that ontologies represent vague or fuzzy information and support approximate reasoning mechanisms. Hence, fuzzy reasoning methods should be associated to fuzzy ontologies, improving the expressiveness of concepts and relationships modelled. Specifically, classification methods based on fuzzy rules are useful to several applications, including Semantic Web and Text Mining, by classifying contents (Web pages and documents) into concepts of a domain depending on fuzzy properties associated to linguistic terms.

In order to face these issues, we propose a model for representing fuzzy ontologies that considers the classic and general fuzzy reasoning methods for classification based on rules containing fuzzy properties. This paper is organized as follows. Section 2 discusses related work on fuzzy ontology representation. Next, Section 3 describes the proposed model for fuzzy ontology representation, followed by a case study presented on Section 4. Finally, Section 5 concludes this paper and points future works.

A. Yaguinuma C., R. T. Ferraz V., T. P. Santos M., A. Camargo H. and M. Nogueira T. (2010).
A MODEL FOR REPRESENTING VAGUE LINGUISTIC TERMS AND FUZZY RULES FOR CLASSIFICATION IN ONTOLOGIES.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Artificial Intelligence and Decision Support Systems*, pages 438-442
DOI: 10.5220/0002976204380442

2 RELATED WORK ON FUZZY ONTOLOGY

Among the approaches for fuzzy ontology representation, some researches consider ontologies composed of fuzzy classes and fuzzy relationships to represent the semantics of domains. In this case, fuzzy class and fuzzy relationship respectively correspond to fuzzy set and fuzzy relation of the fuzzy set theory (Zadeh, 1965). *Fuzzy OWL* (Stoilos et al., 2006) and *f-OWL DL* (Pan et al., 2008) extend OWL language elements in order to represent such features in ontologies.

Recently, fuzzy ontology approaches are moving forward to express fuzzy properties, membership functions and linguistic hedges (Straccia, 2006; Calegari and Ciucci, 2007). Fuzzy properties or fuzzy datatypes extend attributes of traditional ontologies, corresponding to fuzzy variables whose values can be vague linguistic terms defined as fuzzy sets. For example, the *age* attribute can assume fuzzy linguistic values such as *young*, *adult* or *old* that correspond to fuzzy sets. Linguistic hedges are special linguistic terms by which other terms are modified (Klir and Yuan, 1995). Terms like *very*, *more or less*, *fairly* are classic examples of hedges. Any linguistic hedge may be interpreted as a unary operation on the unit interval $[0,1]$. By representing these fuzzy semantic elements, ontologies can model a semantics closely related to the vagueness of human vocabulary.

Some researches have also pointed to fuzzy rule representation. For example, Damásio *et al.* (Damásio et al., 2008) extend the *RuleML* language to express degrees of truth assigned to propositons and rules, however their approach do not support fuzzy properties and membership functions. *Fuzzy DL system* (Bobillo and Straccia, 2008) has achieved better expressiveness, as it represents fuzzy rules containing linguistic terms and implements reasoning mechanisms based on Mamdani method.

As the review of the literature shows, there is a growing trend to increase expressiveness in fuzzy ontologies. Following this direction, it is interesting that fuzzy ontologies support reasoning mechanisms towards classification based on rules containing fuzzy properties. Aiming to support these features, it is fundamental to represent fuzzy properties, membership functions and rules in a suitable way. In this sense, Section 3 describes the proposed model for representing fuzzy properties and classification rules for further fuzzy reasoning in ontologies.

3 MODEL FOR FUZZY PROPERTIES AND RULES IN ONTOLOGIES

In order to represent fuzzy class, fuzzy property and fuzzy rule, we present some definitions to comprehend how we have incorporated them in ontologies:

Definition 1. *Fuzzy class*, which corresponds to a fuzzy set, is a class whose individuals can belong to it with a membership degree between the interval $[0,1]$. When a fuzzy class corresponds to a fuzzy linguistic term associated to a continuous attribute, it can be defined by a parameterized membership function, such as triangular and trapezoidal functions (Klir and Yuan, 1995).

Definition 2. *Fuzzy property* is defined as a property or attribute that can assume vague linguistic values represented by fuzzy classes. This definition is based on the fuzzy variable definition, which corresponds to a variable whose values are fuzzy linguistic terms.

Definition 3. *Fuzzy rule* is defined as an if-then rule whose antecedent contains fuzzy properties and the consequent has a class defined in the ontology (classification rule). Fuzzy reasoning methods can be applied over these rules to derive the class of an individual based on the values of its fuzzy properties.

From these definitions, we propose the model illustrated as a graph in Figure 1. In this model, the *Fuzzy Variable* element is responsible for connecting a *Fuzzy Property* to its linguistic terms represented as *Membership Function Fuzzy Class*. Such connection is respectively made by *hasFuzzyProperty* and *hasFuzzyValue* relationships. The *Membership Function Fuzzy Class* is a fuzzy class that is defined by a parameterized membership function. This class can represent a linguistic term, whose label is described by *linguisticLabel* property of type *String*. The *Trapezoidal* and *Triangular* subclasses represent two possible parameterized membership functions that can be used to model a fuzzy class. To define more parameterized functions, subclasses can be added as well. Finally, the parameters of membership functions are represented by *leftZeroParameter, leftOneParameter, rightZeroParameter, rightOneParameter* properties of type *float*. These parameters determine which values of the fuzzy property that the membership function assumes minumum (zero) and maximum (one) values, both considering left and right sides of the trapezium. In case of triangular functions, only three parameters are required, with *leftOneParameter* correspon-

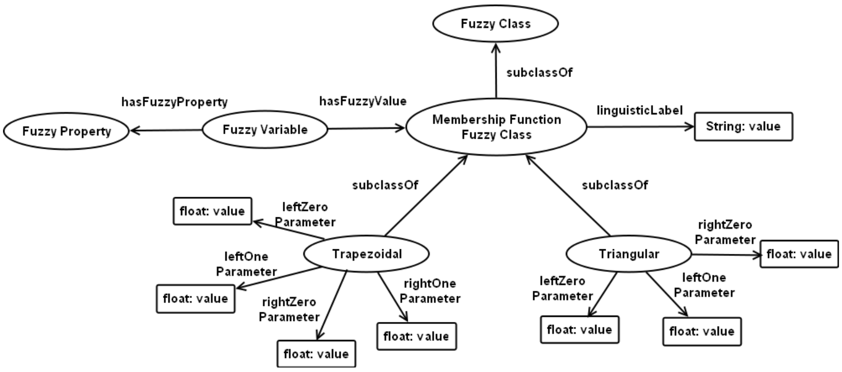

Figure 1: Model for representing fuzzy properties and their linguistic terms as parameterized fuzzy classes.

ding to the center of the triangle.

An important feature of the proposed model is that it is an abstract representation, thus independent of ontology language syntax. So, it can be instantiated using any traditional ontology language that models representational primitives such as classes, instances, attributes and relationships. This is a relevant contribution in comparison to related work, since existing approaches usually extend ontology language syntax to incorporate fuzzy concepts, consequently losing compatibility with reasoners and applications.

Regarding fuzzy rules, we have used Jena Framework (Carroll et al., 2004) to implement classical and general fuzzy reasoning methods, considering fuzzy properties modelled in OWL and rules defined according to Jena rule syntax. The choice for OWL and Jena Framework was just a matter of implementation, other languages and technologies could also be used. Next section depicts a case study involving text mining, illustrating a real application of the proposed model and implemented fuzzy reasoning methods.

4 CASE STUDY: CLASSIFICATION OF TEXT DOCUMENTS

The conducted tests involved the classification of scientific documents considering Artificial Intelligence research area and its subareas. Fifty documents were used, whose attributes firstly corresponded to the frequency of keywords that appear on them, obtained by applying Text Mining techniques (study reported in (Nogueira et al., 2009)). However, as the generated document-attribute matrix was too sparse and high dimensional, we have clustered the documents with fuzzy c-means algorithm (Bezdeck, 1981), thus reducing the dimension of the matrix. In this case, attributes represent the relationship between documents and identified clusters, with a degree corresponding to the strength of such relation. At the end of this pre-processing step, we generated a document-cluster matrix (snippet in Table 1) that represents the fuzzy properties of documents. Last column contains the document class assigned by domain experts, where FL stands for Fuzzy Logics, M for Mining and ML for Machine Learning. All these categories correspond to classes defined in the ontology.

Table 1: Document-cluster matrix generated by fuzzy c-means.

ID	C0	C1	C2	C3	C4	C5	Class
5	0,173	0,0378	0,143	0,271	0,179	0,193	FL
22	0,059	0,222	0,262	0,144	0,163	0,145	M
35	0,335	0,485	0,037	0,036	0,022	0,082	ML

Next step is modelling fuzzy properties in the ontology. In this experiment, we consider that fuzzy properties correspond to relationships between documents and identified clusters, named from *C0* to *C5*, which can assume *low*, *medium* and *high* linguistic values, defined by uniformly distributed triangular membership functions. Listing 1 illustrates how *C0* property and its *low* linguistic value are modelled according to the proposed model, instantiated in OWL.

```
<fuz:FuzzyVariable rdf:ID="c0_fuzzy_variable">
    <fuz:hasFuzzyProperty rdf:resource="#c0"/>
    <fuz:hasFuzzyValue>
        <fuz:TriangularFuzzyClass rdf:ID="c0_low">
            <fuz:linguisticLabel rdf:datatype="&xsd;string">low
            </fuz:linguisticLabel>
            <fuz:leftZeroParameter rdf:datatype="&xsd;float">3.13E−8
            </fuz:leftZeroParameter>
            <fuz:leftOneParameter rdf:datatype="&xsd;float">3.13E−8
            </fuz:leftOneParameter>
            <fuz:rightZeroParameter rdf:datatype="&xsd;float">0.489177
            </fuz:rightZeroParameter>
        </fuz:TriangularFuzzyClass>
    </fuz:hasFuzzyValue>
    . . .
</fuz:FuzzyVariable>
```

Listing 1: *C0* property and its *low* linguistic value.

Once we have modelled all fuzzy properties, they can be used in fuzzy rules for classification of scientific documents. In this experiment, fuzzy rules should contain *C0* to *C5* properties in the antecedent part, and a class in the consequent that corresponds to the inferred class of the ontology. In order to automatically learn these rules, we have applied the Wang and Mendel method (Wang and Mendel, 1992) over the document-cluster matrix. As a result, we have obtained nineteen rules in total, which were modelled using Jena rule syntax. Due to space limitations, Listing 2 shows only two of them, just to illustrate how they are specified according to the proposed model.

```
[rule1: (?x c0 'low'), (?x c1 'low'), (?x c2 'high'), (?x c3 'low'),
        (?x c4 'low'), (?x c5 'low') -> (?x rdf:type Machine_Learning)]

[rule2: (?x c0 'low'), (?x c1 'medium'), (?x c2 'low'), (?x c3 'low'),
        (?x c4 'medium'), (?x c5 'low') -> (?x rdf:type Fuzzy_Logics)]
```

Listing 2: Fuzzy rules generated by Wang and Mendel.

At this stage, fuzzy reasoning methods can be applied to classify scientific documents regarding Artificial Intelligence subareas. When a new document needs to be classified, it passes through the preprocessing step in order to obtain its fuzzy property values. After that, the user chooses a fuzzy reasoning method (classical or general) to obtain the inferred class. In our tests, general method performed better than classical one, as the former considers all fired rules by combining their association degree. However, this research does not intend to evaluate the best method, our goal is making them available for ontology-based applications that require the representation of fuzzy properties and classification rules.

Concluding this case study, we have observed that the proposed model contributed to represent the vagueness present in textual information content. Moreover, fuzzy reasoning methods can automatically infer the classes of new documents, an information that can be analyzed by document retrieval systems for improving query results.

5 CONCLUSIONS AND FUTURE WORK

We have proposed a model for representing fuzzy properties and fuzzy rules in ontologies, which can be instantiated using any traditional ontology language that models basic representational primitives. Furthermore, fuzzy reasoning methods (classical and general) were implemented in order to support classification of new instances according to their fuzzy property values. The results obtained from the study case demonstrate that the proposed model contributed to manage vagueness on text documents, representing a good approach not only for classification but also for organization of the text documents.

Finally, we plan to incorporate more fuzzy set concepts, such as linguistic hedges, fuzzy relations etc. We intend to support other types of rules and their fuzzy reasoning methods (e.g. Mamdani and Larsen methods), as well as defuzzification methods.

ACKNOWLEDGEMENTS

We thank CAPES and INEP agencies for supporting this research, inside the scope of WebPIDE project.

REFERENCES

Bezdeck, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.

Bobillo, F. and Straccia, U. (2008). fuzzydl: An expressive fuzzy description logic reasoner. In *FUZZ-IEEE*, pages 923–930.

Calegari, S. and Ciucci, D. (2007). Fuzzy ontology, fuzzy description logics and fuzzy-owl. In *International Workshop on Fuzzy Logic and Applications*, pages 118–126.

Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., and Wilkinson, K. (2004). Jena: implementing the semantic web recommendations. In *WWW*, pages 74–83.

Cordón, O., Jesus, M. J. D., and Herrera, F. (1999). A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning*, 20(1):21 – 45.

Damásio, C. V., Pan, J. Z., Stoilos, G., and Straccia, U. (2008). Representing uncertainty in ruleml. *Fundamenta Informaticae*, 82(3):265–288.

Gruber, T. R. (2009). Ontology. In *Encyclopedia of Database Systems*, pages 1963–1965. Springer.

Klir, G. J. and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic - Theory and Applications*. Prentice Hall PTR, Upper Saddle River, USA.

Nogueira, T. M., Camargo, H. D. A., and Rezende, S. O. (2009). Management of imprecision and uncertainty in the identification of similar textual documents. *Congress of the Tri-national Academy of Sciences*, pages 1–10. C3N Annals (in portuguese).

Pan, J. Z., Stamou, G., Stoilos, G., Taylor, S., and Thomas, E. (2008). Scalable querying services over fuzzy ontologies. In *WWW*, pages 575–584.

Stoilos, G., Simou, N., Stamou, G., and Kollias, S. (2006). Uncertainty and the semantic web. *IEEE Intelligent Systems*, 21(5):84–87.

Straccia, U. (2006). A fuzzy description logic for the semantic web. In *Fuzzy Logic and the Semantic Web*, pages 73–90. Elsevier.

Wang, L. and Mendel, J. (1992). Generating fuzzy rules by learning from examples. *IEEE Transaction on Fuzzy Systems, Man and Cybernetics*, 22:414–427.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.