

Discovery of Production Rules with Fuzzy Hierarchy

Fadl M. Ba-Alwi and Kamal K. Bharadwaj

Abstract— In this paper a novel algorithm is proposed that integrates the process of fuzzy hierarchy generation and rule discovery for automated discovery of Production Rules with Fuzzy Hierarchy (PRFH) in large databases.

A concept of frequency matrix (Freq) introduced to summarize large database that helps in minimizing the number of database accesses, identification and removal of irrelevant attribute values and weak classes during the fuzzy hierarchy generation. Experimental results have established the effectiveness of the proposed algorithm.

Keywords— Data Mining, Degree of subsumption, Freq matrix, Fuzzy hierarchy.

I. INTRODUCTION

THE most predominant representation of the discovered knowledge is the standard production rules in the form *If P Then D*.

Much world knowledge is best expressed in the form of hierarchies because they give comprehensible knowledge structures that carry more information than simple collection of rules. Hierarchies are useful for incorporating knowledge about relationships between items, clear class boundaries [1]. Hierarchies find use in measuring linguistic relatedness or similarity [2]. Hierarchies allow the user to view the discovered rules at different levels of details, and to focus his/ her attention on those interesting aspects [3].

An important structure mechanism for knowledge bases is building an inheritance hierarchy of classes based on the context of their objects [4]-[6]. Fuzzy hierarchical structure is an object-oriented hierarchy representation which is getting more and more important for modern database applications. Automated discovery of fuzzy hierarchical structure from large datasets [7]-[9] plays fundamentally important role in data mining because it provides comprehensible results that captures real-life inheritance of objects.

Integrating hierarchy generation process into the data-mining algorithm can dramatically reduce the execution time [10]. Therefore, hierarchies are tremendously reduces the search time to find them as well [11].

As an extension of CPR, Bharadwaj and Jain [12] proposed a Hierarchical Censored Production Rule (HCPR), which is a CPR, augmented with a specificity and generality information. A HCPR is of the form

Fadl Muter Ba-Alwi is a Ph.D. scholar at School of Computer and Systems Sciences (SC&SS), Jawaharlal Nehru University (JNU), New Delhi, India, on leave from the Sana'a University, Republic of Yemen. Phone: +91-9871326272; (e-mail: fadl_ba_alwi@hotmail.com).

Kamal K. Bharadwaj, is a Professor at the SC&SS, JNU, New Delhi, India. (kamal_bharadwaj@hotmail.com).

Decision IF <condition>

UNLESS <cursor>

GENERALITY <general info>

SPECIFICITY <specific info> (1)

HCPR can be made to exhibit variable precision in the reasoning such that both the certainty of belief in a conclusion and its specificity may be controlled by the reasoning process.

In this paper we introduce the concept of frequency matrix (Freq) to summarize large dataset that helps in minimizing the number of dataset accesses, identification and removal of irrelevant attribute values and weak classes during the fuzzy hierarchy generation. Further, we proposed a Discovery of Production Rules with Fuzzy Hierarchy (DPRFH) algorithm that integrates the process of fuzzy hierarchy generation and rule discovery using a novel approach based on Freq matrix. The proposed fuzzy hierarchical structure takes into account a general view of the inheritance such that each class is associated with three categories of properties – public properties (P_{pub}) inherited by all the descendant classes, special properties (P_{spl}) inherited only by some designated descendant classes and private properties (P_{pvt}) exclusively associated with the class which will not be inherited by the descendant classes. Using the fuzzy hierarchical structure discovered, the DPRFH-algorithm finally produces Production Rules with Fuzzy Hierarchy.

II. PRODUCTION RULES WITH FUZZY HIERARCHY

Given dataset with n classes, each class D_k ($1 \leq k \leq n$) in the proposed fuzzy hierarchy has a general format with three categories of properties separated by single and double vertical lines:

$$D_k [P_{pub} | P_{spl} || P_{pvt}]$$

where P_{pub} , P_{spl} and P_{pvt} for a class D_k are defined as follows:

- $(P_{pub})_k = \{P_{pub1}, P_{pub2}, \dots\}$ is a non-empty set of public properties inherited by all the classes in the sub-trees of the root class D_k .
- $(P_{spl})_k = \{S_1(C_1), S_2(C_2), \dots\}$ is a set of special properties inherited by some / all direct descendants of the class D_k and may or may not be inherited by the classes at the lower levels depending on the property definition at the direct descendants, i.e. $S_i(C_i)$ means that set of properties S_i will be inherited by the set of classes C_i .
- $(P_{pvt})_k = \{P_{pvt1}, P_{pvt2}, \dots\}$ is a set of private properties exclusively for a class D_k which will not be inherited.

Note that P_{spl} and/or P_{pvt} can be empty for any class at the branch nodes in the fuzzy hierarchy. However, the leaf

classes will have both the P_{spi} and P_{pvt} empty.

The different levels of the proposed fuzzy hierarchy are presented in Fig.1.

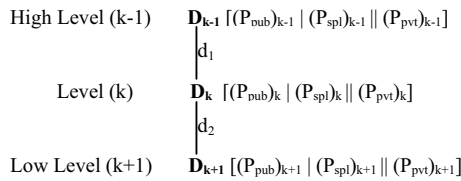


Fig.1: General structure of the proposed fuzzy hierarchy.

Referring to Fig.1, the degree (d_1 , d_2) of inheritance (subsumption) between the general and specific classes gives the fuzzy hierarchical structure. For example, $D_{k-1} \xrightarrow{d_1} D_k \xrightarrow{d_2} D_{k+1}$. The contents of $(P_{pub})_k$, $(P_{spi})_k$ and $(P_{pvt})_k$ with reference to class D_k at level k are defined as follows:

- (i) $(P_{pub})_k = (P_{pub})_{k-1} \cup \alpha$.
where, $(P_{pub})_{k-1}$ is a non-empty set of public properties belonging to the general class D_{k-1} . α is a non-empty set of public properties belonging to the class D_k such that $(P_{pub})_{k-1} \cap \alpha = \emptyset$. The set $(P_{pub})_k$ is disjoint from P_{spi} and P_{pvt} for any class at any level.
- (ii) $(P_{spi})_k = \beta \cup \gamma$.
where, $\beta \subseteq (P_{spi})_{k-1}$ i.e. β is a subset of special properties belonging to the class D_{k-1} . γ is a set of special properties belonging to the class D_k and not to the class D_{k-1} . The sets β and γ are disjoint and any of them can be empty. Also the set $(P_{pvt})_{k-1}$ is disjoint from $(\beta \cup \gamma)$.
- (iii) $(P_{pvt})_k = \beta \cup \delta$.
where, β (defined at (ii)) and δ is a set of private properties belonging to the class D_k and not to the class D_{k-1} . The sets β and δ are disjoint and any of them can be empty. The set $(P_{pvt})_k$ does not belong to the specific class D_{k+1} .

Based on the Hierarchical Censored Production Rules (HCPR) [5], [12], Production Rules with Fuzzy Hierarchy (PRFH) for the class D_k ($1 \leq k \leq n$) is defined as follows:

$$P \rightarrow D_k$$

Generality [general class]
Specificity [$D_{ki}(d_1), \dots, D_{ki}(d_i), \dots, D_{kj}(d_j)$].

(2)

where, $P = (P_{pub})_k \cup (P_{spi})_k \cup (P_{pvt})_k$, and the specificity element $D_{ki}(d_i)$ means that D_{ki} is a specific class of D_k with degree of subsumption d_i . Note that Generality would be empty for the root class and Specificity would be empty for leaf classes. The Generality-Specificity relation minimizes the redundancy of the properties (preconditions) of PRFH.

An example to the proposed fuzzy hierarchy is given in Fig.2.

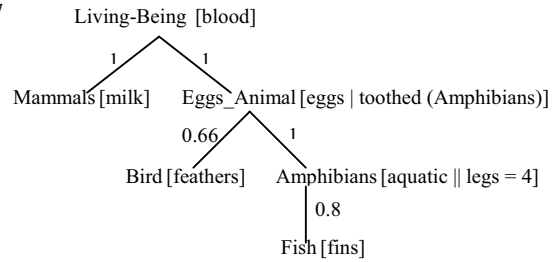


Fig.2: Fuzzy Hierarchical structure for the Living-Being.

III. FREQUENCY MATRIX

In order to summarize large datasets, we introduce a concept of frequency matrix (Freq) of size $(m \times n)$, where m is the number of distinct values of attributes and n is the number of distinct classes in the dataset. The m attribute values in Freq matrix will be denoted by property P_i ($1 \leq i \leq m$). The i -th tuple corresponding to the property P_i is $t_i = (\text{Freq}[P_i, D_1], \text{Freq}[P_i, D_2], \dots, \text{Freq}[P_i, D_n])$.

For each class D_k ($1 \leq k \leq n$) and property P_i ($1 \leq i \leq m$) the element $\text{Freq}[P_i, D_k]$ contains a frequency value $\in [0 \dots 1]$ according to the following formula:

$$\text{Freq}[P_i, D_k] = \frac{|P_i \wedge D_k|}{|D_k|} \quad (3)$$

The frequency value of the missing data in the dataset is zero. And for the attribute of Boolean type, only the true-value will be used in the Freq matrix.

In the Freq matrix only the values greater than or equal to the user defined *threshold* value in the interval $[0.5 \dots 1]$ (default is 0.5) will be considered and **highlighted**. All the tuples having no highlighted elements (irrelevant tuples) are removed and the properties corresponding to the remaining tuples are relabeled as P_1, P_2, P_3, \dots . The classes that are rare or sparse data i.e. classes which have no highlighted elements are irrelevant classes and hence, they will be removed from the Freq matrix.

For a given set of properties R in Freq matrix (arranged in decreasing order of their corresponding frequencies in column D_k), the frequency f of the combination of all the properties in R corresponding to the class D_k is

$$f = \left(\sum_{i=1}^{|R|} \text{Freq}[P_i, D_k] \right) - (|R|-1), \quad \forall P_i \in R \quad (4)$$

Using formula (4) will reduce the number of dataset accesses, such that if $f > 0$ then no need to verify this combination in the dataset. Otherwise, if $f \leq 0$ then only those properties which make $f \leq 0$, will be verified from the dataset.

IV. DISTANCE BETWEEN TUPLES IN Freq MATRIX

The Euclidean distance measure is used to determine the closest tuple to the $t_{\max} = (1, 1, \dots, 1)$ (fixed tuple with all the values equal to the maximum frequency 1). A tuple t_i of length n (where n is the number of the highlighted elements in t_i) having the minimum $\text{Dist}(t_{\max}, t_i)$ is called **lowest_dist** tuple (the best tuple) where,

$$\text{Dist}(t_{\max}, t_i) = \sqrt{\sum_{k=1}^n (1 - \text{Freq}[t_i, D_k])^2} \quad (5)$$

It is to be noted that only the highlighted elements in t_i (formula (5)) would be considered.

V. PUBLIC PROPERTIES

The public properties for each class D_k , $(P_{\text{pub}})_k$, are discovered by considering the properties P_i corresponding to the highlighted element(s) in the column D_k ($\text{Freq}[P_i, D_k]$, $1 \leq i \leq m$). With the help of formula 4, $(P_{\text{pub}})_k$ will contain only those properties whose combination belongs to the class D_k , and the highlighted element(s) in the column D_k corresponding to the properties which are not included in the combination are converted to unhighlighted element(s).

VI. UNIQUE CLASS DEFINING PROPERTIES

Amongst the public properties $(P_{\text{pub}})_k$ of each class D_k we can extract (if any) the properties that can uniquely define class D_k such that $(P_{\text{uniq}})_k \subseteq (P_{\text{pub}})_k$.

When no other information is given, an event with lower probability to occur gives more information, than an event with higher probability [13]. The discovery of $(P_{\text{uniq}})_k$ is based on the minimum probability values of $P_i \in (P_{\text{pub}})_k$. Depend on the frequencies of the elements, $\text{Freq}[P_i, D_j] \forall j$, the probability $\text{Prob}(P_i, D_k)$ is computed, such that if P_i has low frequencies in each class D_j then P_i in class D_k will have low probability which means high chance to be a member in $(P_{\text{uniq}})_k$ and vice versa. If P_i a unique property that belongs only to class D_k i.e. $\text{Freq}[P_i, D_k] \geq \text{threshold}$ and $\text{Freq}[P_i, D_j] = 0 \forall j \neq k$, then the probability $\text{Prob}(P_i, D_k) = 0$, and hence, class D_k cannot become part of the fuzzy hierarchy (irrelevant class) because class D_k cannot inherit any property from any other class and also no other class can inherit the unique property of D_k . The probability of occurrence of the property P_i in $(P_{\text{uniq}})_k$ is

$$\text{Prob}(P_i, D_k) = -0.5 * \log \left(\frac{\text{Freq}[P_i, D_k]}{\sum_{j=1}^n \text{Freq}[P_i, D_j]} \right) \quad (6)$$

The properties in $(P_{\text{pub}})_k$ are arranged in ascending order of probability. The first property $P_i \in (P_{\text{pub}})_k$ is combined with the initialized $(P_{\text{uniq}})_k = \emptyset$. If this combination, with the help of formula (4), reduces the frequencies of all or some of the elements $\text{Freq}[P_i, D_j] \forall j \neq k$ then $(P_{\text{uniq}})_k = (P_{\text{pub}})_k \cup \{P_i\}$ and $(P_{\text{pub}})_k = (P_{\text{pub}})_k - \{P_i\}$. The process of increasing $(P_{\text{uniq}})_k$ and decreasing $(P_{\text{pub}})_k$ is repeated until either the frequencies of all the elements $\text{Freq}[P_i, D_j] \forall j \neq k$ are zero, or $(P_{\text{pub}})_k$ becomes empty. In case $(P_{\text{pub}})_k$ is empty and some elements $\text{Freq}[P_i, D_j] \forall j \neq k$ are still non-zero then $(P_{\text{uniq}})_k$ does not exist, so $(P_{\text{uniq}})_k = \emptyset$.

Any class D_k for which $(P_{\text{uniq}})_k$ does not exist is considered as a weak class and will be removed during the construction of the fuzzy hierarchy.

If the root node of the hierarchy is not properly selected, the hierarchy can become lopsided and difficult to interpret [14]. First we will discover the set of root properties P_{root} for the root node such that P_{root} is included in all other nodes of the hierarchy (that is different from the root node in [15]). The degree of an element $\text{Freq}[P_i, D_k]$ in the Freq matrix is defined as follows:

$\text{deg_element}(\text{Freq}[P_i, D_k]) = 1$ if the element is highlighted.
 $\text{deg_element}(\text{Freq}[P_i, D_k]) = 0$ if the element is unhighlighted.

The degree of a tuple t_i corresponding to the property P_i is computed using the following formula:

$$\text{deg_tuple}(t_i) = \frac{\sum_{k=1}^n \text{deg_element}(\text{Freq}[P_i, D_k])}{n} \quad (7)$$

If there is only one tuple with the highest degree then the property P corresponding to that tuple is the root property i.e. $P_{\text{root}} = \{P\}$. Otherwise, finding the *lowest_dist* tuples (formula 5) helps in discovering P_{root} . That is, $P_{\text{root}} = \{\text{all tuples having distance } ds = y + x \text{ from } t_{\max}\}$ where $y = \text{Dist}(t_{\max}, \text{lowest_dist tuple})$, and $0 \leq x \leq 0.1$ (the value $x = 0.1$ gives the closest tuples to the *lowest_dist* tuple).

Removing the classes having no highlighted elements corresponding to the root properties will further reduce the size of Freq matrix. Amongst the remaining classes, the root class D_r (at level 0) can be discovered by computing the root degree:

$$\text{root_deg}(D_r) = \frac{|P_{\text{root}}|}{|(P_{\text{uniq}})_r|} \quad (8)$$

If there exist only single class D_r in the Freq matrix such that more than half of its public properties belong to P_{root} i.e. $\text{root_deg}(D_r) \in (0.5...1]$, then class D_r is the root class of the fuzzy hierarchy. Otherwise, a new class D_{new} will be introduced as root class with an appropriate label assigned in accordance with its properties.

VIII. CONSTRUCTING THE FUZZY HIERARCHY

Based on the subsumption relation we can discover the Generality-Specificity relation between the classes at the remaining levels of the hierarchy.

As a modification of the subsumption measure in [16]-[17], we define a subsumption measure between two values x and y of two attributes as follows:

$\text{subsume}(x, y) = 1$ if $(x \neq \emptyset)$ and $(y \neq \emptyset)$ and $((x = y)$ or $(x \subset y))$ and $(\text{attribute of } x = \text{attribute of } y)$.
 $\text{subsume}(x, y) = 0$ if $(x = \emptyset)$ or $(y = \emptyset)$ or $(x \neq y)$ or $(x \not\subset y)$ or $(\text{attribute of } x \neq \text{attribute of } y)$.

Given two classes D_a and D_b . Let α_i is the i -th property in $(P_{\text{uniq}})_a$ and β_j is the j -th property in $(P_{\text{uniq}})_b$. The degree of the two classes is

$$d_1 = \text{deg_sub}(D_a, D_b) = \frac{|\sum_{i=1}^{(P_{\text{unq}})_a}| \sum_{j=1}^{(P_{\text{unq}})_b} \text{subsume}(\alpha_i, \beta_j)|}{|(P_{\text{unq}})_a|} \quad (9)$$

$$d_2 = \text{deg_sub}(D_b, D_a) = \frac{|\sum_{j=1}^{(P_{\text{unq}})_b}| \sum_{i=1}^{(P_{\text{unq}})_a} \text{subsume}(\beta_j, \alpha_i)|}{|(P_{\text{unq}})_b|} \quad (10)$$

Class D_b is specific class of D_a ($a \neq b$) if

- $\text{deg_sub}(D_a, D_b) > \text{deg_sub}(D_k, D_b) \forall k \neq a \neq b$,
- $\text{deg_sub}(D_a, D_b) \neq \text{deg_sub}(D_b, D_a)$,
- $\text{deg_sub}(D_a, D_b) \geq \text{threshold}$, and
- Class D_b must inherit at least one property from its general classes at the higher levels i.e. classes on the path from root class D_r to D_a .

A new class D_{new} is introduced as a direct ancestor class for both the classes D_a and D_b if

- $\text{deg_sub}(D_a, D_b) = \text{deg_sub}(D_b, D_a) \geq \text{threshold}$,
- $\text{deg_sub}(D_a, D_b) > \text{deg_sub}(D_k, D_a) \forall k \neq a$, and
- $\text{deg_sub}(D_a, D_b) > \text{deg_sub}(D_k, D_b) \forall k \neq b$.

Root class D_r is the direct ancestor (parent) of the class D_b if D_b cannot become specific class of $D_a \forall a \neq b$, and a new class cannot be introduced.

During the insertion of class D_b into the fuzzy hierarchy as a specific class of D_a , the different categories of the properties, $(P_{\text{pub}}, P_{\text{spl}}, P_{\text{pvt}})$, of each class D_k along the path from root class D_r to D_a will be reorganized and the following two cases would be arise:

Case (a): D_a has no specific classes (i.e. leaf class) and some properties $R \subset (P_{\text{pub}})_a$ are not found in $(P_{\text{pub}})_b$ then R will be moved to $(P_{\text{pvt}})_a$.

Case (b): D_a has specific classes, so

- If some properties $R \subset (P_{\text{pub}})_a$ are not found in $(P_{\text{pub}})_b$ but R is found in other specific classes of D_a then R will move to $(P_{\text{spl}})_a$ and R is removed from $(P_{\text{pub}})_b$.
- As per the definition, $(P_{\text{spl}})_a = \{S_1(C_1), \dots, S_k(C_k), \dots\}$. If $S_k \cap (P_{\text{pub}})_b = R \neq \emptyset$ then $C_k = C_k \cup D_b$ and R is removed from $(P_{\text{pub}})_b$.
- If $(P_{\text{pvt}})_a \cap (P_{\text{pub}})_b = R \neq \emptyset$ then R from $(P_{\text{pvt}})_a$ is moved to $(P_{\text{spl}})_a$ as $R(D_b)$ and R is removed from $(P_{\text{pub}})_b$.
- All the properties R which the class D_a inherits and R is not found in $(P_{\text{pub}})_b$ will be moved to $(P_{\text{pvt}})_a$ and all the general classes of D_a up to the root class D_r will be reorganized.

IX. EXPERIMENTAL RESULTS

The performance and behavior of the proposed DFHPR-algorithm are tested on Zoo, Bridge, Lenses, Letter, Transportation-means datasets and large synthetic data. The experimental results are quite encouraging and have established the effectiveness of the algorithm.

DPRFH-algorithm is an attempt towards integrating process of fuzzy hierarchy generation and discovery of rules. A concept of frequency matrix is introduced to summarize large datasets that minimizes the number of dataset accesses and helps in eliminating irrelevant attribute values and weak classes. A generalized view of inheritance is incorporated into the proposed fuzzy hierarchy by associating each class D_k in the fuzzy hierarchy with three categories of properties, $D_k [P_{\text{pub}} | P_{\text{spl}} || P_{\text{pvt}}]$. One of the most important extension of the present work would be discovery of Hierarchical Production Rules with Exceptions and Fuzzy Hierarchy.

REFERENCES

- [1] F. M. Ba-Alwi and K. K. Bharadwaj, "Automated discovery of hierarchical ripple-down rules (HRDRs)", *Proc of In the Twenty-third IASTED International Conference on Artificial Intelligence and Applications (AIA 2005)*, Innsbruck, Austria, February 14-16, 2005.
- [2] S. Levachkine and A. Guzmán-Arenas, "Hierarchies measuring qualitative variables", Springer-Verlag Berlin Heidelberg 2004, A. Gelbukh (Ed.): CICLing 2004, 2004, pp. 262-274.
- [3] B. Liu, M. Hu and W. Hsu, "Multi-Level organization and summarization of the discovered rules", Boston, USA, SIGKDD-2000, Aug 20-23, 2000.
- [4] G. W. Mineau and R. Godin, "Automatic structuring of knowledge bases by conceptual clustering", *IEEE Transactions on Knowledge and Data Engineering*, vol. 7(5), 1995, pp. 824-828.
- [5] N. K. Jain, K. K. Bharadwaj and N. Marranghello, "Extended of hierarchical censored production rules (EHCPs) system: An approach toward generalized knowledge representation", *Journal of Intelligent Systems*, UK, vol. 9(3,4), 1999, pp. 259-295.
- [6] E. Tsiporkova and S. Garrett, "On different modes of inheritance in fuzzy class hierarchies", CIMCA'99, Vienna, 1999.
- [7] G. Chen, Q. Wei, E. Kerre and G. Wets, "Overview of fuzzy associations mining", in *Proc of ISIS2003*, Jeju, Korea, September 2003.
- [8] Q. Wei and G. Chen, "Mining generalized association rules with fuzzy taxonomic structures", in *Proc of the North America Fuzzy Information Processing Society (NAFIPS99)*, New York, 1999, pp. 477-481.
- [9] F. Boussemart, C. Lecoutre, S. Merchez and E. Grégoire, "Fuzzy hierarchies", CP99 Post-Conference Workshop on MODELLING AND SOLVING SOFT CONSTRAINTS, Alexandria, Virginia, USA, October 16, 1999.
- [10] R. Srikant, Q. Vu and R. Agrawal, "Mining association rules with item constraints", in *Proc of the Third International Conf on Knowledge Discovery and Data Mining (KDD97)*, 1997, pp. 67-73.
- [11] M. Suan, "Semi-Automatic taxonomy for efficient information searching", *Proceeding of the 2nd International Conference on Information Technology for Application (ICITA- 2004)*, 2004.
- [12] K. K. Bharadwaj and N. K. Jain, "Hierarchical censored production rules (HCPRs) Systems", *Data and Knowledge Engineering*, North Holland, vol. 8, 1992, pp. 19-34.
- [13] F. Hussain, H. Liu, E. Suzuki and H. Lu, "Exception rule mining with a relative interestingness measure", *Knowledge Discovery and Data Mining, Lecture Notes in Artificial Intelligence (PAKDD)*, Terano, T., Liu, H. & Chen, A.L.P. (Eds.), Kyoto, Japan. Springer-Verlag, vol. 1805, April 2000, pp. 86-97.
- [14] C. Westphal and T. Blaxton, "Data mining solution, methods and tools for solving real-world problems", Wiley Computer Publishing, 1998.
- [15] J. Domingo-Ferrer and V. Torra, "Approximating fuzzy measures by hierarchically decomposable ones", *ISIF 2002*, pp. 191-198.
- [16] H. Suryanto and P. Compton, "Learning classification taxonomies from a classification knowledge based system", In Staab, S., Maedche, A., Neldel, C. & Wiemer-Hastings, P. (Eds.), *Proceedings of the Workshop on Ontology Learning*, 14 Conference on Artificial Intelligence (ECAI'00), Berlin, August 20-25, 2000.
- [17] Z. Ding and Y. Peng, "A probabilistic extension to ontology language owl", in *Proc of the Hawai'i International Conf. on System Sciences*, 2004.