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

\[ \text{decision IF } \langle \text{condition} \rangle \]

UNLESS \langle \text{censor} \rangle

GENERALITY \langle \text{general info} \rangle

SPECIFICITY \langle \text{specific info} \rangle \]

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 \( (P_{\text{pub}}) \) inherited by all the descendant classes, special properties \( (P_{\text{spl}}) \) inherited only by some designated descendant classes and private properties \( (P_{\text{pvt}}) \) 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:

\[ D_k \{ P_{\text{pub}} | P_{\text{spl}} \| P_{\text{pvt}} \} \]

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

- \( (P_{\text{pub}})_k = \{ P_{\text{pub}1}, P_{\text{pub}2}, \ldots \} \) is a non-empty set of public properties inherited by all the classes in the sub-trees of the root class \( D_k \).
- \( (P_{\text{spl}})_k = \{ S_1(C_1), S_2(C_2), \ldots \} \) 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 \( C_i \).
- \( (P_{\text{pvt}})_k = \{ P_{\text{pvt}1}, P_{\text{pvt}2}, \ldots \} \) is a set of private properties exclusively for a class \( D_k \) which will not be inherited.

Note that \( P_{\text{spl}} \) and /or \( P_{\text{pvt}} \) can be empty for any class at the branch nodes in the fuzzy hierarchy. However, the leaf
classes will have both the \(P_{spl}\) and \(P_{priv}\) empty.

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

\[
\begin{align*}
\text{High Level (k=1)} & : D_k \equiv (\{P_{pub,k}\} \cup \{P_{spl,k}\} \cup \{P_{priv,k}\}) \\
\text{Level (k)} & : D_k \equiv (\{P_{pub,k}\} \cup \{P_{spl,k}\} \cup \{P_{priv,k}\}) \\
\text{Low Level (k+1)} & : D_{k+1} \equiv (\{P_{pub,k+1}\} \cup \{P_{spl,k+1}\} \cup \{P_{priv,k+1}\})
\end{align*}
\]

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

Referring to Fig.1, the degree (\(d_i\), \(d_s\)) of inheritance (subsumption) between the general and specific classes gives the fuzzy hierarchical structure. For example, \(D_k \equiv D_{k+1} \subseteq D_{k+2}\). The contents of \((P_{pub,k}, \{P_{spl,k}\}, \{P_{priv,k}\})\) 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 \(D_{k+1}\). \(\alpha\) is a non-empty set of public properties belonging to the class \(D_k\) such that \((P_{pub,k}) \cap \alpha = \emptyset\). The set \((P_{pub,k})\) is disjoint from \(P_{spl}\) and \(P_{priv}\) for any class at any level.

(ii) \((P_{spl}) = \beta \cup \gamma\)

where, \(\beta \subseteq (P_{spl,k+1})\) i.e. \(\beta\) is a subset of special properties belonging to the class \(D_{k+1}\). \(\gamma\) is a set of special properties belonging to the class \(D_k\) and not to the class \(D_{k+1}\). The sets \(\beta\) and \(\gamma\) are disjoint and any of them can be empty. Also the set \((P_{priv,k})\) is disjoint from \((\beta \cup \gamma)\).

(iii) \((P_{priv,k}) = \beta \cup \delta\)

where, \(\beta\) (defined at (ii)) and \(\delta\) is a set of private properties belonging to the class \(D_k\) and not to the class \(D_{k+1}\). The sets \(\beta\) and \(\delta\) are disjoint and any of them can be empty. The set \((P_{priv,k})\) 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 \([D_{a1}(d_1), \ldots, D_{a_i}(d_i), \ldots, D_{a_j}(d_j)]\).

where, \(P = (P_{pub,k}) \cup (P_{spl,k}) \cup (P_{priv,k})\), and the specificity element \(D_{a_i}(d_i)\) means that \(D_{a_i}\) 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.

\[
\begin{array}{c}
\text{Living-Being [blood]} \\
\text{Mammals [milk]} & \text{Eggs, Animal [eggs | toothed (Amphibians)]} \\
\text{Bird [feathers]} & \text{Amphibians [aquatic | legs = 4]} \\
\text{Fish [fins]} & \text{0.66} & \text{1} & \text{0.8}
\end{array}
\]

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 \(\text{Freq}\) 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 \(\text{Freq}\) matrix will be denoted by property \(P_i\) (1 \(\leq i \leq m\)). The \(i\)-th tuple corresponding to the property \(P_i\) is \(t_i = (\text{Freq}[P_i, D_1], \text{Freq}[P_i, D_2], \ldots, \text{Freq}[P_i, D_n])\).

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 \(\epsilon [0...1]\) according to the following formula:

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

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 \(\text{Freq}\) matrix.

In the \(\text{Freq}\) matrix only the values greater than or equal to the user defined threshold value in the interval \([0.5...1]\) (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 \(P_1, P_2, P_3, \ldots\). 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 \(\text{Freq}\) matrix.

For a given set of properties \(R\) in \(\text{Freq}\) 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 \(t_{max} = (1, 1, \ldots, 1)\) (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 \(\text{dist}\) tuple (the best tuple) where,
Dist(t_max, t_i) = \sqrt{\frac{\sum_{k=1}^{n} (1 - \text{Freq}(t_i, D_k))^2}{n}} \tag{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_{pub}^k), are discovered by considering the properties P_i corresponding to the highlighted element(s) in the column D_k (Freq[P_i, D_k], 1 \leq i \leq m). With the help of formula 4, (P_{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_{pub}^k) of each class D_k we can extract (if any) the properties that can uniquely define class D_k such that (P_{unq}^k \subseteq (P_{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_{unq}^k) is based on the minimum probability values of P_i \in (P_{pub}^k). Dependent on the frequencies of the elements, Freq[P_i, D_j] \forall j, the probability 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_{unq}^k) and vice versa. If P_i a unique property that belongs only to class D_k i.e. Freq[P_i, D_k] \geq \text{threshold} and Freq[P_i, D_j] = 0 \forall j \neq k, then the probability 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_{unq}^k) is

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

The properties in (P_{pub}^k) are arranged in ascending order of probability. The first property P_1 \in (P_{pub}^k) is combined with the initialized (P_{unq}^k) = \emptyset. If this combination, with the help of formula 4, reduces the frequencies of all or some of the elements Freq[P_i, D_j] \forall j \neq k then (P_{unq}^k) = (P_{unq}^k) \cup \{P_i\} and (P_{pub}^k) = (P_{pub}^k) \setminus \{P_i\}. The process of increasing (P_{unq}^k) and decreasing (P_{pub}^k) is repeated until either the frequencies of all the elements Freq[P_i, D_j] \forall j \neq k are zero, or (P_{pub}^k) becomes empty. In case (P_{pub}^k) is empty and some elements Freq[P_i, D_j] \forall j \neq k are still non-zero then (P_{unq}^k) does not exist, so (P_{unq}^k) = \emptyset.

Any class D_k for which (P_{unq}^k) does not exist is considered as a weak class and will be removed during the construction of the fuzzy hierarchy.

VII. DISCOVERING THE ROOT CLASS

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 Freq[P_i, D_k] in the Freq matrix is defined as follows:

deg_element(Freq[P_i, D_k]) = 1 \text{ if the element is highlighted.} \\
deg_element(Freq[P_i, D_k]) = 0 \text{ 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}(Freq[P_i, D_k])}{n} \tag{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_root = \{P\}. Otherwise, finding the lowest_dist tuples (formula 5) helps in discovering P_root. That is, P_root = \{ all tuples having distance ds = y + x from t_max \} where y = Dist(t_{max}, 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:

root_deg (D_k) = \frac{|P_{root}|}{|P_{unq}|} \tag{8}

If there exist only single class D_k in the Freq matrix such that more than half of its public properties belong to P_root i.e. root_deg (D_k) \in (0.5...1], then class D_k 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 \text{ if } (x \not\subset y) \text{ and } (y \not\subset x) \text{ and } ((x = y) \text{ or } (x \subset y) \text{ and } (attribute of x = attribute of y).} \\
\text{subsume}(x, y) = 0 \text{ if } (x \subset y) \text{ or } (y \subset x) \text{ or } (x \not\subset y) \text{ or } (attribute of x \not= attribute of y).}

Given two classes D_a and D_b. Let \alpha_i is the i-th property in (P_{unq}^a), and \beta_j is the j-th property in (P_{unq}^b). The degree of the two classes is
\[ \text{deg}_\text{sub}(D_a, D_b) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{subsume}(\alpha_i, \beta_j) \]

\[ \text{deg}_\text{sub}(D_a, D_b) = \frac{|(P_{\text{unq}})_a|}{|(P_{\text{unq}})_b|} \]

\[ d_1 = \frac{\sum_{j=1}^{|\text{unq}_{\text{a}}|} \text{subsume}(\alpha_j, \beta_i)}{|(P_{\text{unq}})_a|} \]

\[ d_2 = \frac{\sum_{i=1}^{|\text{unq}_{\text{b}}|} \text{subsume}(\beta_i, \alpha_j)}{|(P_{\text{unq}})_b|} \]

Class \( D_b \) is specific class of \( D_a \) (\( a \neq b \)) if
- \( \text{deg}_\text{sub}(D_a, D_b) > \text{deg}_\text{sub}(D_a, D_b) \) if \( k \neq a \neq b \),
- \( \text{deg}_\text{sub}(D_a, D_b) = \text{deg}_\text{sub}(D_a, D_b) \) if \( k \neq a \neq b \), and
- \( D_b \) must inherit at least one property from its general classes at the higher levels i.e., classes on the path from root class \( D_a \) to \( D_b \).

A new class \( D_{\text{new}} \) is introduced as a direct ancestor class for both the classes \( D_a \) and \( D_b \) if
- \( \text{deg}_\text{sub}(D_a, D_{\text{new}}) > \text{deg}_\text{sub}(D_a, D_b) \) if \( k \neq a \neq b \), and
- \( \text{deg}_\text{sub}(D_a, D_{\text{new}}) = \text{deg}_\text{sub}(D_a, D_b) \) if \( k \neq a \neq b \).

Root class \( D_c \) is the direct ancestor (parent) of the class \( D_b \) if \( D_b \) cannot become specific class of \( D_a \) (\( a \neq b \)) 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 \( D_c \) along the path from root class \( D_i \) to \( D_b \) will be reorganized and the following two cases would be arise:

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

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

**IX. EXPERIMENTAL RESULTS**

The performance and behavior of the proposed DFPRH-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.

**X. CONCLUSION**

DFPRH-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 \( D_i \) in the fuzzy hierarchy with three categories of properties, \( D_i \rightarrow \{P_{\text{pub}}, P_{\text{spl}}, P_{\text{pvt}}\} \). One of the most important extension of the present work would be discovery of Hierarchical Production Rules with Exceptions and Fuzzy Hierarchy.

**REFERENCES**


