Quantitative Measures of the Accuracy, Comprehensibility, and Completeness of a Fuzzy Expert System

Phayung Meesad         Gary G. Yen
meesad@okstate.edu     gyen@okstate.edu
Intelligent Systems and Control Laboratory
Electrical and Computer Engineering
Oklahoma State University

Abstract: Using optimization tools such as genetic algorithms to construct a fuzzy expert system (FES) focusing only on its accuracy without considering comprehensibility may result in a system that is not easy to understand or so called a black box model. To exploit the transparency features of FESs for explanation in higher-level knowledge representation, a FES should provide high comprehensibility while preserving its accuracy. The completeness of fuzzy sets and rule structures should also be considered to guarantee that every data point has a response output. This paper proposes some quantitative measures to determine the degree of the accuracy, the comprehensibility, and the completeness of FESs. These quantitative measures are then used as a fitness function for a genetic algorithm in optimally building a FES.

I. INTRODUCTION

Quantitative measures are essential and form the basis for making reliable decisions in information technology such as fuzzy expert systems (FESs). Quantitative assessment helps us to evaluate the quality of a FES that is not accessible to our intuitive ability. Generally, in constructing a FES, an accuracy measure is a goodness index that is usually a primary concern. The accuracy measure implies how well a FES can perform.

Comprehensible knowledge representation is a key advantage of FESs over black box schemes such as neural networks. However, the if-then rules of a FES may not be understandable without a careful design. Accuracy alone may not be sufficient to show the goodness of FESs [1]-[3]. A comprehensibility measure is an additional quantitative assessment that indicates whether a FES is understandable. In addition, a completeness measure is an indicator to evaluate a fuzzy system whether its linguistic variables and rule structure cover the entire possible data domain [4]-[6].

In literature, there have been some discussions regarding the accuracy and the comprehensibility [1]-[3], [9] as well as the completeness [4]-[6] of FESs; however, the quantitative measures of FESs have not been well considered yet. This paper proposes some quantitative measures as a guideline to design FESs. The quantitative measures developed are used as a fitness function to guide a GA optimization to construct a FES that preserves accuracy, comprehensibility, and completeness.

II. FUZZY EXPERT SYSTEMS

A fuzzy expert system (FES) of interest here is targeted for pattern classification applications. Its knowledge structure is based on a generalization of the Takagi-Sugeno (TS) fuzzy model [7]. The FES has linguistic representation in the antecedent part and constant numbers in the consequent part. The knowledge representation of the FES is as follows.

$$R_i: \text{If } x_1 = A_{i1} \text{ and } x_2 = A_{i2} \text{ and } \ldots \text{ and } x_M = A_{iM}, \text{ then class is } C_i \text{ with confidence } CF_i;$$

where $R_i$, $i = 1, \ldots, L$, is the label of the $i$th rule; $L$ is the number of rules; $x_j$, $j = 1, \ldots, M$, is the $j$th dimension of the input pattern; $M$ is the number of dimensions; $A_{ij}$ is the index to the linguistic terms of the $i$th rule in the $j$th dimension. $C_i$ is the class of the $i$th rule; and $CF_i$ is the confident factor for the $i$th rule.

III. QUANTITATIVE MEASURES

Three important quantitative measures of FESs include accuracy, comprehensibility, and completeness that are discussed in the following.

A. Accuracy Measure

The most common assessment of the performance of a classifier system is to examine its accuracy. Accuracy is a measure of a predictive model that reflects the number of times that the model is making the correct classification when it is applied to test data. It measures the probability that the system can correctly classify the data. The accuracy measure ($A_C$) can be determined from the following equation:

$$A_C = \frac{\text{correctly classified patterns}}{\text{total patterns}}$$  (1)
B. Comprehensibility Measures

Comprehensibility of fuzzy systems involves three important matters: the compactness of fuzzy systems, the similarity between linguistic terms, and the inconsistency of fuzzy rules.

**Compactness**: Compactness of fuzzy systems is associated with the comprehensibility of fuzzy systems. A compact fuzzy system implies that the fuzzy systems are easy to comprehend. Compactness of fuzzy systems relates to three aspects: a small number of linguistic terms in each dimension, a small number of fuzzy rules in the rule base, and a small number of conditions in the rule premise or antecedent part [3]-[4]. Figs. 1 and 2 illustrate the compactness concept of fuzzy systems concerning linguistic terms. Comparing Figs. 1 and 2, Fig. 2 has fewer linguistic labels. It is relatively easier for users to discern a fuzzy variable with three rather than seven linguistic labels.

A second aspect of compactness is the number of fuzzy rules. The number of fuzzy rules needed to represent a physical system depends on the structure of the fuzzy rules. In a standard structure of a fuzzy system with $M$ dimensions and each dimension partitioned into $N$ subspaces, there exist up to $N^M$ rules in the fuzzy system. For example, in Fig. 3, for a two-dimensional fuzzy system partitioned into 6 subspaces in each dimension, the number of fuzzy rules is 36. If all the possible rules are used then the system is considered not compact. For the same fuzzy system, a more compact fuzzy system is shown in Fig. 4. A compact rule set is easier to comprehend and recognize. Compactness of fuzzy rules becomes more important when the system involves a large number of dimensions [4], [8].

![Fig. 1: Comprehensible fuzzy variable](image1)

![Fig. 2: More comprehensible fuzzy variable](image2)

![Fig. 3: A structure of two-dimensional fuzzy system with too many fuzzy rules](image3)

![Fig. 4: A structure of two-dimensional fuzzy system with a more compact rule set](image4)

![Fig. 5: A two-dimensional fuzzy system with two conditions per rule](image5)

![Fig. 6: A more compact fuzzy system with 1.67 conditions per rule](image6)
A third aspect of compactness is the number of conditions in the antecedent part of fuzzy rules or the number of features used per rule. If some of the features are not used in rules then the system becomes more compact. The system structure can be easier to comprehend. Fig. 5 shows a two-dimensional fuzzy system with four rules. In Fig. 5, the number of conditions is two. Each rule uses both inputs as conditions in the antecedent part. Fig. 6 illustrates the same fuzzy system with three rules, but the number of conditions per rule is 1.67. Rules 1 and 3 use both inputs $x_1$ and $x_2$ in the antecedent part while Rule 2 uses only input $x_1$. Fig. 6 has a structure that is easier to comprehend and recognize. The compactness of a fuzzy system can be quantified into numerical values as follows.

\[
N_R = \text{counts of all the rules in the rule set} \quad (2)
\]

\[
N_a = \frac{\text{counts of all the antecedents in the rule set}}{N_R} \quad (3)
\]

\[
N_l = \frac{\text{counts of all the linguistic labels}}{M} \quad ; \quad (4)
\]

where $N_R$ is the number of rules; $N_a$ is the number of antecedents per rule; $N_l$ is the number of linguistic labels per dimension; and $M$ is the number of dimensions.

**Linguistic Similarity:** Similarity measure [2], [9] for fuzzy sets is used to quantify the comprehensibility of fuzzy knowledge base. The degree of linguistic similarity is considered the highest when two fuzzy sets are equal. When there are no overlapping fuzzy sets, the degree of linguistic similarity is zeros. The degree of linguistic similarity falls in [0, 1], if there are overlapping fuzzy sets. The degree of linguistic similarity ($LS$) of a fuzzy system can be determined by the following equations:

\[
LS_j(k_1, k_2) = \frac{\sum_{x} \min[m_{l_{j_{k_1}}}(x), m_{l_{j_{k_2}}}(x)]}{\sum_{x} \max[m_{l_{j_{k_1}}}(x), m_{l_{j_{k_2}}}(x)]} \quad (5)
\]

\[
LS_j = \frac{1}{(N_j - 1) + (N_j - 2) + \cdots + 1} \sum_{k_1 = 1}^{N_j - 1} \sum_{k_2 = k_1 + 1}^{N_j} LS_j(k_1, k_2) \quad ; \quad (6)
\]

\[
LS = \frac{1}{M} \sum_{j=1}^{M} LS_j \quad (7)
\]

where $LS_j(k_1, k_2) \in [0, 1]$ is the degree of linguistic similarity between linguistic labels $l_{j_{k_1}}$ and $l_{j_{k_2}}$; $N_j$ is the number of linguistic labels in the $j$th dimension; $k_1$ and $k_2$ are the indexes to linguistic labels; and $LS_j \in [0, 1]$ is the average of the degree of linguistic similarity in the $j$th dimension.

**Inconsistency of Fuzzy Rules:** Inconsistency of fuzzy rules can directly effect to the overall decision-making of the system. It can degrade the overall performance of the system.

Inconsistency of fuzzy rules should be avoided. Inconsistency of fuzzy rules occurs when there are two or more rules are conflicting. Fuzzy rules are conflicting if they have similar antecedents but different consequents. Measuring rule inconsistency is equivalent to measuring rule similarity. Degree of fuzzy rule similarity can be measured by using fuzzy similarity measure. Fuzzy rule similarity ($RS$) is divided into two parts: the similarity of the antecedents ($SA$) and the similarity of the consequents ($SC$). The similarity between the $j$th antecedents of the $i$th rule and the $k$th rule ($SA(R_i, R_k)$) can be determined from the following equation:

\[
SA(R_i, R_k) = \sum_{x} \max[m_{l_{i_{k_1}}}(x), m_{l_{i_{k_2}}}(x)] \quad (8)
\]

Using constant numbers as consequents, the similarity between the consequents of the $i$th rule and the $k$th rule ($SC(R_i, R_k)$) can be determined from the following equation:

\[
SC(R_i, R_k) = \begin{cases} 1 & \text{if the consequents are the same;} \\ 0 & \text{otherwise.} \end{cases} \quad (9)
\]

\[
RS(R_i, R_k) = \frac{1}{M + 1} \left( \sum_{j} SA(R_i, R_k) + SC(R_i, R_k) \right); \quad (10)
\]

for $i \neq k; \ i = 1, \ldots, L-1; \ k = 2, \ldots, L; \ j = 1, \ldots, M$. (11)

Where $RS(R_i, R_k) \in [0, 1]$ is the degree of rule similarity between rules $R_i$ and $R_k$; and $RS \in [0, 1]$ is the average of the degree of rule similarity.

**C. Completeness Measure**

Completeness is a property of deductive systems that has been used in the context of artificial intelligence to indicate that the knowledge representation scheme can represent every entity within the intended domain. In a fuzzy system, completeness is a fundamental issue since complete fuzzy systems can respond to any given input. A complete fuzzy system can achieve proper operation avoiding undesirable situations [5]-[6]. The completeness of fuzzy systems consists of two main factors: completeness of fuzzy partitions and completeness of fuzzy rule structure [4].

Suppose input variable $x$ in the universe of discourse $X$ is divided into $N$ fuzzy partitions represented by membership functions $m_i(x)$, for $i = 1, \ldots, N$. The completeness of the system is satisfied if

\[
\forall x \in X, \exists i : 1 \leq i \leq N \text{ such that } m_i(x) > 0. \quad (12)
\]

A certain level of completeness, $\delta$, give rises to the concept of strong completeness, as follows:
\( \forall x \in X, \exists i: 1 \leq i \leq N \text{ such that } m_i(x) > \delta. \)  \hspace{1cm} (13)

Fig. 7: a) a complete rule structure; b) an incomplete rule structure

Rule
\[
\begin{align*}
R_1 & : x_1 \quad \delta_1 \\
R_2 & : x_2 \quad \delta_2 \\
R_3 & : x_3 \quad \delta_3 \\
R_4 & : x_1 \vee x_2 \vee x_3 \\
\end{align*}
\]

Fig. 8: Antecedent structure of a complete fuzzy system

Fig. 9: Antecedent structure of an incomplete fuzzy system

Figs. 7a and 7b, respectively, illustrates complete and incomplete fuzzy rule structures. In Fig. 7a, the rule structure is complete because every partition in each dimension is used. Though the fuzzy partitions are complete, the rule structure in Fig. 7b is incomplete because some partitions are not used. The input in which its partitions are not used may cause a no-response or zero output.

To check whether or not the rule structure is complete, each dimension of fuzzy sets from all rules are mapped onto the same axis by OR operator \((\vee)\), as shown in Figs. 8 and 9. Fig. 8 shows the antecedent structure of a complete fuzzy system drawn from Fig. 7a, while Fig. 9 illustrates the antecedent structure of an incomplete fuzzy system derived from Fig. 7b.

A completeness measure of a fuzzy rule structure is defined as the proportion of the complete region and the region of interest. Similarly, an incompleteness measure is defined as the proportion of the incomplete region and the region of interest. Completeness degree in the \(j\)th dimension \((CD_j)\) and incompleteness degree in the \(j\)th dimension \((ID_j)\) are calculated from the following equations:

\[
CD_j = \frac{CR_j}{RI_j} = \frac{N_{x \in X | m(x) > \delta}}{N_{x \in X}} \hspace{1cm} (14)
\]

\[
ID_j = \frac{IR_j}{RI_j} = \frac{N_{x \in X | m(x) < \delta}}{N_{x \in X}} = 1 - CD_j \hspace{1cm} (15)
\]

\[
I_D = \frac{\sum_{j=1}^{M} ID_j}{M} \hspace{1cm} (16)
\]

where \(I_D\) is the overall incompleteness degree which is the average values of all the incompleteness degrees from each dimension. \(M\) is the number of the dimensions. \(CD_j\) and \(ID_j\) \(\in [0, 1]\) are completeness degree and incompleteness degree, respectively, in the \(j\)th dimension; \(CR_j\) is the length of the complete region in the \(j\)th dimension; \(IR_j\) is the length of the incomplete region in the \(j\)th dimension; and \(RI_j\) is length of the region of interest in the \(j\)th dimension or the universe of discourse \(X\). \(x \in X\) is the input elements. \(N_x\) is the number of element \(x\). \(m(x)\) the membership degrees of \(x\). \(\delta \in [0, 1]\) is the level of completeness.

IV. GENETIC OPTIMIZATION

There are many ways to optimize fuzzy rules to preserve the performance accuracy and the comprehensibility in terms of knowledge representation as well as the completeness of the system. One of the popular techniques is to use evolutionary computation such as genetic algorithms (GAs). In this study, we applied a GA to perform an optimization process of fuzzy rules by searching for good accuracy and comprehensibility as well as low incompleteness based on the quantity measures proposed above.

A. Genetic Algorithm (GA)

The genetic algorithm [10] is a stochastic search useful for optimization problems. It is motivated by the mechanisms of evolution in nature [11]. The genetic algorithm operates on populations of strings, with the string coded to represent some underlying parameter set. Reproduction, crossover, and mutation are applied to successive string populations to create
new string populations. These operators involve random number generation, string copying, and partial string exchange. This study employed the genetic rule encoding discussed in [12].

B. Fitness Functions

GAs have been widely used for helping in the generation of if-then rule bases of FESs. When a FES is constructed, accuracy, comprehensibility, and completeness should be a concern during the optimization process using the GAs. A fitness function is used to guide the evolutionary process to a satisfactory goal. The fitness function used is based on the accuracy performance, the comprehensibility, and the completeness of the rule set. The fitness function can be determined from the following equations:

\[ P_F = (N_R)(A_R) + (N_D)(L_D) + L_S + R_S + I_D \]  (17)

\[ F_N = (W_{AC})(A_C) - (W_{PF})(P_F) \]  (18)

where \( F_N \) represents fitness function; \( P_F \) represents a penalty function; \( A_C \) is the accuracy; \( N_R \) and \( A_R \) are the number of rules and antecedents per rule, respectively; \( N_D \) and \( L_D \) are the number of dimensions and linguistic terms per dimension; \( L_S \) is linguistic similarity; \( R_S \) is the rules similarity or the inconsistency; and \( I_D \) is the degree of incompleteness. \( W_{AC} \) is the weight for the accuracy and \( W_{PF} \) is the weight for the penalty term.

V. SIMULATION RESULTS

Three medical data domains which were provided by the Institute of Oncology, University Medical Center, Ljubljana, Yugoslavia were used to validate the quantitative measures proposed in the design of FESs in this study. The three data domains include Breast Cancer Data, Lymphography Domain, and Primary Tumor Domain [13]-[16].

Breast cancer data comprises of 9 attribute values. It is a two-class problem with 286 patterns: 201 patterns of non-recurrence and 85 patterns of recurrence. Lymphography data is a four-class problem and has 148 patterns with 18 attribute values. Primary tumor data is a 21-class problem and has 339 patterns with 17 attribute values.

In our experiment, the training patterns were randomly selected for 70% of all examples and the remaining 30% for testing.

We developed a graphical user interface (GUI) system, a hybrid intelligent system (HIS) [17] containing an incremental learning fuzzy neural network (ILFN) [18], a FES, and a GA. The GUI, as shown in Fig. 10, was developed to run under the Matlab environment. Fig. 11 illustrates an example of rule quality evaluation based on the accuracy measure and the comprehensibility measures as well as the incompleteness measure.

Fig. 11: Rule evaluation for Lymphography data

In this paper we focused on extraction of fuzzy knowledge bases from the three data domains. First, the ILFN was trained one pass, and then fuzzy rules were extracted from the trained ILFN [12]. Then, a GA was used to optimize the fuzzy rules to maintain accuracy, comprehensibility, and completeness, as discussed in this paper.

Using the three medical domains, FESs were constructed based on the trained ILFN networks and the GA [12]. The performance accuracy, the comprehensibility, and the incompleteness of the FESs are shown in Table I. Performance accuracy comparisons among other methods in literature are shown in Table II.

TABLE I

<table>
<thead>
<tr>
<th>Data</th>
<th>Accuracy (%)</th>
<th>Comprehensibility</th>
<th>( I_D )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
<td>Overall</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>75.5</td>
<td>73.25</td>
<td>74.83</td>
</tr>
<tr>
<td>Lymphography</td>
<td>94.17</td>
<td>82.22</td>
<td>90.54</td>
</tr>
<tr>
<td>Primary Tumor</td>
<td>52.02</td>
<td>26.51</td>
<td>39.53</td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (%)</th>
<th>Comprehensibility</th>
<th>( I_D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant86</td>
<td>72</td>
<td>77</td>
<td>46</td>
</tr>
<tr>
<td>Bayes</td>
<td>65</td>
<td>83</td>
<td>39</td>
</tr>
<tr>
<td>AQR</td>
<td>72</td>
<td>76</td>
<td>35</td>
</tr>
<tr>
<td>CN2</td>
<td>71</td>
<td>82</td>
<td>37</td>
</tr>
<tr>
<td>AQT-15</td>
<td>68</td>
<td>82</td>
<td>41</td>
</tr>
<tr>
<td>CLILP2</td>
<td>76</td>
<td>85</td>
<td>37</td>
</tr>
<tr>
<td>FES</td>
<td>74.83</td>
<td>90.54</td>
<td>39.53</td>
</tr>
</tbody>
</table>

Table I shows the simulation results based on performance accuracy, comprehensibility, and incompleteness. For the breast cancer data set, the extracted fuzzy rules had 75.5%, 73.26%, and 74.83% accuracy on the training set, testing set, and overall, respectively. The number of rules
(N_R) was 3. The number of antecedents per rule (A_R) was 5. The number of linguistic variables per dimension (L_D) was 2.4. The linguistic similarity (LS) or inconsistency degree was 0.08. The incompleteness degree (I_D) was 0.25. In this data set, using the proposed quantitative measures to guide the GA in searching for an optimal rule set, the resulting FES presented a high degree of comprehensibility. The incompleteness was also low. Using δ = 0.001, only 25% of the region of interest was incomplete.

For the lymphography data domain, the accuracies were 94.17%, 82.22%, and 90.54% on the training data, testing data, and overall, respectively. The comprehensibility measures were 8, 7.9, 2.5, 0.15, and 0.56 for N_R, A_R, L_D, LS, and RS, respectively. From the quantitative measures on the comprehensibility, it implied that the resulting fuzzy rules were comprehensible. The incompleteness degree (I_D) was low as 0.23.

For the primary tumor data set, the performance accuracies were 52.02%, 26.51%, and 39.53% on the training data, testing data, and overall available data, respectively. The comprehensibility measures based on N_R, A_R, L_D, LS, and RS were 40, 7.3, 2.4, 0.07, and 0.46, respectively. The incompleteness degree (I_D) was 0.22. The number of fuzzy rules for this data set was relatively high since this data set has 21 classes overall. The average number of rules generated was only two fuzzy rules per class.

The simulations have shown that the resulted fuzzy knowledge bases had the best accuracy on the lymphography data set. The fuzzy knowledge base extracted for the primary tumor data was the lowest accuracy. However, the fuzzy knowledge bases were easy to comprehend, since the number of rules, the number of antecedent per rule, the number of linguistic per dimension, the linguistic similarity degree, and the rule similarity degree were low.

Table II shows the comparisons among several methods based on the accuracy performance. Other methods such as Assistant-86, Bayes, AQR, CN2, AQ7-15, and CLILP2 were reported in [19], using the same data set. It is found that the fuzzy expert system achieved significant accuracy performance with other methods. However, in this study we were not focused mainly on the accuracy. Instead, to develop a method that can construct a fuzzy expert system, which preserves the accuracy performance of the system, enhances the comprehensibility of the fuzzy sets, and provides the completeness of the fuzzy rule structure is the top priority in the design process.

VI. CONCLUSIONS

In this paper, we proposed several quantitative measures pertaining to the accuracy performance of a fuzzy system, the comprehensibility of fuzzy sets, and the completeness of fuzzy rule structures. The quantitative measures were used as the fitness function to guide a genetic algorithm (GA) to search for an optimal fuzzy rule set.

In simulations on the three medical data sets, the resulting fuzzy rule based systems were able to perform a compatible accuracy to some state-of-the-art methods in literature. The resulting fuzzy knowledge bases show a high degree of comprehensibility of fuzzy sets and a low degree of incompleteness of the rule structure.

The quantitative measures on accuracy performance, comprehensibility, and completeness were developed for the quality evaluation of a fuzzy system for pattern classification applications. The fuzzy system has linguistic antecedents and constant numbers in consequent. However, the proposed quantitative measures can be easily extended to other fuzzy systems with a few modifications.

REFERENCES