Linguistic Rulesets Extracted from a Quantifier-based Fuzzy Classification System

Rasmani, Khairul; Garibaldi, Jonathan M.; Shen, Qiang; Ellis, Ian O.

Published in:
Proceedings of the 18th International Conference on Fuzzy Systems (FUZZ-IEEE'09)

DOI:
10.1109/FUZZY.2009.5277081

Publication date:
2009

Citation for published version (APA):
Rasmani, K., Garibaldi, J. M., Shen, Q., & Ellis, I. O. (2009). Linguistic Rulesets Extracted from a Quantifier-based Fuzzy Classification System. In *Proceedings of the 18th International Conference on Fuzzy Systems (FUZZ-IEEE'09)* (pp. 1204-1209). IEEE Press. https://doi.org/10.1109/FUZZY.2009.5277081

Linguistic Rulesets Extracted from a Quantifier-based Fuzzy Classification System

Khairul A. Rasmani, Jonathan M. Garibaldi, Qiang Shen and Ian O. Ellis

Abstract—**The use of linguistic rulesets is considered one of the greatest advantages that fuzzy classification systems can offer compared to non-fuzzy classification systems. This paper proposes the use of fuzzy thresholds and fuzzy quantifiers for generating linguistic rulesets from a data-driven fuzzy subsethood-based classification system. The proposed technique offers not only simplicity in the design and comprehensibility of the generated rulesets but also practicality in the implementation. Additionally, the use of fuzzy quantifiers makes it easier for the user to understand the classification process and how such classifications were reached. The effectiveness of the proposed method is demonstrated using a medical dataset which provides evidence that rules generated by the proposed system are consistent with the expert-rules created by clinicians.**

I. Introduction

The application of fuzzy rule induction algorithms (FRIAs) to solve various real world problems has been widely reported in the literature [1]. Of particular interest to this paper are data-driven FRIAs for handling classification tasks, referred to here as data-driven fuzzy classification systems (FCSs). The main feature of such a system is the capability to learn from data, with the use of fuzzy sets to provide structural knowledge that can be used to classify new instances [2].

There are many non-fuzzy classification algorithms currently available (see for example [3]). However, many of these classification algorithms may be very good in generalisation ability and so be very useful for classifying new instances, but lack of comprehensibility of the generated models. In fact, most of the models generated by non-fuzzy classification algorithms contain numerical values and may not be linguistically interpretable. This makes it harder for the user to utilise the models for decision making purposes. Note that an automated-system, also known as a computer assisted system, is normally considered as a tool to assist experts or non-experts in decision making. Hence, interpretability of such a system should be regarded as highly important [4].

This paper proposes the use of a rule simplification technique to extract linguistic rulesets from a data-driven subsethood-based fuzzy rule induction algorithm, namely FuzzyQSBA [5]. The main intention of the proposed technique is to build a model that can be easily interpreted by a non-expert in classification systems in general or FCSs in particular. To demonstrate the generalisation ability and comprehensibility of the generated rulesets, these rules are compared with rules generated by non-fuzzy rule-based classification algorithms. Additionally, comprehensibility of some selected models is compared to the rules created by clinicians who are considered experts in the field of study.

The rest of this paper is organized as follows. Section II describes the background theory and the systems overview of the proposed algorithm which includes the rule induction and simplification techniques. This is followed by Section III which presents the comparative study, experimental results and discussions of the findings. Finally, conclusions and suggestions for further research are outlined in Section IV.

II. Background Theory and System Overview

A. Fuzzy Subsethood Measures

A fuzzy subsethood measure was originally defined as the degree to which a fuzzy set is a subset of another. However, the definition of fuzzy subsethood value can be extended to calculate the degree of subsethood for linguistic terms in an attribute variable V to a decision class D [6]. For linguistic terms $\{A_1, A_2, \ldots, A_n\} \in V$ and $(V, D) \subseteq U$:

$$S(D, A_i) = \frac{\sum_{x \in U} \nabla(\mu_D(x), \mu_{A_i}(x))}{\sum_{x \in U} \mu_D(x)} \qquad (1)$$

where ∇ can be any *t-norm* operator. It should be noted that, to be used for classification problems, both V and D must be defined under the same universe of discourse U [6]. Although the decision class is represented by fuzzy sets, this definition allows the decision class with zero fuzziness where the membership value is either one or zero.

B. Rule Induction

FuzzyQSBA is a rule induction algorithm that was developed by extending the Weighted Subsethood-based

Manuscript received February 5, 2009.

Khairul A. Rasmani is with the School of Computer Science, University of Nottingham, UK. He is currently on sabbatical leave from Faculty of Information Technology and Quantitative Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia (phone: 00603-55435492 fax: 00603-55435501; e-mail: khairulanwar@ tmsk.uitm.edu.my).

Jonathan M. Garibaldi is with the Intelligent Modelling and Analysis Research Group, School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham NG8 1BB UK (e-mail: jmg@cs.nott.ac.uk).

Qiang Shen is with the Advanced Reasoning Research Group, School of Computer Science, Aberystwyth University , Ceredigion SY23 3DB, UK (email: qqs@aber.ac.uk).

Ian O. Ellis is with the School of Molecular Medical Sciences, Nottingham University Hospital and University of Nottingham, Queens Medical Centre, Derby Road, Nottingham, NG7 2UH, UK. (email: Ian.Ellis@nottingham.ac.uk)

978-1-4244-3597-5/09/$25.00 ©2009 IEEE

Algorithm (WSBA) [7]. WSBA has the significant advantage, as compared to previous subsethood-based methods, of not relying on the use of predefined threshold values in generating fuzzy rulesets. The development of WSBA was based on fuzzy subsethood values as defined in Equation (1). Given a training dataset, WSBA induces a fixed number of rules according to the number of possible classification outcomes. To avoid the use of any threshold values in the rule generation process, crisp weights generated using fuzzy subsethood values are created for each of the linguistic terms appearing in the resulting fuzzy rule antecedents. In FuzzyQSBA, fuzzy quantifiers are applied to replace the crisp weights within the rules learned by WSBA. As small changes in the training dataset might cause a change to the entire ruleset, developing a fuzzy model that employs continuous fuzzy quantifiers may be more appropriate compared to two-valued or multi-valued crisp quantifiers [5]. Vila et al. [8] proposed a continuous fuzzy quantifier which applies linear interpolation between the two classical, extreme cases of the existential quantifier $\exists$ and the universal quantifier $\forall$. In particular, the quantifier was defined such that:

$$Q(A_{ij}, D_k) = (1-\lambda_Q).\mathrm{T}_{\forall,A/D} + \lambda_Q.\mathrm{T}_{\exists,A/D} \qquad (2)$$

where Q is the quantifier for fuzzy set A relative to fuzzy set D and λ_Q is the *degree of neighborhood* of the two extreme quantifiers. The truth values of the existential quantifier $\mathrm{T}_{\exists,\mathrm{A}/\mathrm{D}}$ and the universal quantifier $\mathrm{T}_{\forall,\mathrm{A}/\mathrm{D}}$ were defined as:

$$\mathrm{T}_{\exists,A/D} = \underset{k=1}{\overset{N}{\Delta}} \mu(a_k) \nabla \mu(d_k) \qquad (3)$$

$$\mathrm{T}_{\forall,A/D} = \underset{k=1}{\overset{N}{\nabla}} (1-\mu(d_k)) \Delta \mu(a_k) \qquad (4)$$

where a_k and d_k are the membership functions of fuzzy sets A and D respectively, ∇ represents a *t-norm* and Δ represents a corresponding *t-conorm*. By using fuzzy subsethood values as the *degree of neighborhood* (λ_Q) of the quantifiers, any possible quantifiers that exist between the existential and universal quantifiers can be created in principle. Initially, all linguistic terms of each attribute are used to describe the antecedent of each rule. This may look tedious, but the reason for keeping this complete form is that every linguistic term may contain important information that should be taken into account. The continuous fuzzy quantifiers are created using information extracted from data and behave as a modifier for each of the fuzzy terms. The resulting FuzzyQSBA ruleset can be simply represented by

$$R_k = \underset{i=1..m}{\nabla} \left(\underset{j=1..n}{\Delta} (Q(A_{ij}, D_k) \nabla \mu_{A_{ij}}(x)) \right),\ k = 1, 2, ..., n \qquad (5)$$

where $Q(A_{ij}, D_k)$ are fuzzy quantifiers and $\mu_{A_{ij}}(x)$ are fuzzy linguistic terms.

As both the quantifiers and the linguistic terms are fuzzy sets, choices of *t-norm* operators can be used to interpret $\nabla(Q(A_{ij}, D_k), \mu_{A_{ij}}(x))$ whilst guaranteeing that the inference results are fuzzy sets. Based on the definitions of the fuzzy subsethood value (1), fuzzy existential quantifier (3), and fuzzy universal quantifier (4), it can be proved that if λ_Q is equal to zero then the truth-value of quantifier Q will also equal zero. Thus, during the rule generation process, the emerging ruleset is simplified as any linguistic terms whose quantifier has the truth-value of zero will be removed automatically from the fuzzy rule antecedents, reducing considerably the seeming complexity of the learned ruleset. As commonly used in rule-based systems for classification tasks, the concluding classification will be that of the rule whose overall weight is the highest amongst all.

C. Rule Extraction

Fuzzy quantifiers have been employed in FuzzyQSBA with the intention to increase the readability of the resulting fuzzy rules and to improve the transparency of the rule inference process. However, the structure of the rules is still very complex. Thus, although the use of quantifiers will make the rules more readable, it seems that it does not increase the comprehensibility of the fuzzy rules. As an alternative, a rule simplification process that is based on fuzzy quantifiers is proposed below. In [9], fuzzy quantifiers are suggested to be used as a *fuzzy threshold*. The basic idea of a *fuzzy threshold* is extended here to conduct the rule simplification process for FuzzyQSBA. This is to offer flexibility in accepting or rejecting any particular linguistic term to represent a particular linguistic variable in a fuzzy rule.

To employ the rule simplification, the following *fuzzy quantifiers* and *fuzzy antonym quantifiers* are proposed:

$$\mathrm{T}_{\mathrm{Q}}(\eta) = \begin{cases} 1 & \text{if } \mathrm{T}_Q(\lambda) \geq \eta \\ \dfrac{\mathrm{T}_Q(\lambda)}{\eta} & \text{if } \mathrm{T}_Q(\lambda) < \eta \end{cases}, \qquad (6)$$

$$\mathrm{T}_{\mathrm{antQ}}(\eta) = \begin{cases} 1 & \text{if } \mathrm{T}_Q(\lambda) \leq 1-\eta \\ \dfrac{1-\mathrm{T}_Q(\lambda)}{\eta} & \text{if } \mathrm{T}_Q(\lambda) > 1-\eta \end{cases} \qquad (7)$$

where $\mathrm{T}_Q(\lambda)$ is the truth value of quantifier (TVQ) associated with each linguistic term (Equation 5) and η is a threshold value that can be defined as:

$$\eta = p \times \omega \qquad (8)$$

where p is a multiplication factor for the maximum TVQ, ω. In this technique, the decision to accept a particular linguistic term is made locally without affecting other variables. The aim of using a fuzzy threshold is to soften the decision boundary in the process of accepting or rejecting any terms to be promoted as antecedents of a fuzzy rule, whilst at the same time significantly reducing the number of terms in the induced fuzzy rules. The fuzzy quantifiers mentioned above can be interpreted as 'at least η' and its antonym 'at most 1-η'.

The following technique is proposed to perform the rule simplification:

(1) For each variable, select the maximum TVQ and calculate $T_Q(\eta)$ and $T_{antQ}(\eta)$ for each linguistic term.
(2) For i = 1, 2, …, l where l is the number of linguistic terms for a variable, and for $m \neq n$, calculate:
   - $\delta(T_{Q_i}(\eta)) = |T_{Q_m}(\eta) - T_{Q_n}(\eta)|$ (9)
   - $\delta(T_{antQ_i}(\eta)) = |T_{antQ_m}(\eta) - T_{antQ_n}(\eta)|$ (10)
(3) Conduct the following test:
   If $\min_i \{\delta(T_{Q_i}(\eta))\} \geq \{\delta(T_{antQ_i}(\eta))\}$:
   - choose the negation of terms with the lowest TVQ to represent the conditional attribute,
   - otherwise, choose the term with the highest TVQ.
(4) Create a simplified rule using the accepted linguistic terms (or negation of the terms).

Note that when η = 1, the fuzzy quantifier and its antonym will become 'most' and 'least', and when $\eta = 0$ the quantifier and its antonym will become 'there exists at least one' and 'for all'. By using the technique proposed above, the primary terms with higher TVQs are accepted to represent the antecedents of the fuzzy rules. By lowering the value of η, the primary terms with a lower TVQ will gradually be accepted. The idea behind this technique is that only one of the linguistic term or negation of the linguistic term that is dominant will be chosen to represent a particular linguistic variable.

III. Comparative Study

The data used for this comparative study is an extract from the Nottingham Tenovus Breast Cancer Dataset [10]. The dataset which contains 663 instances is currently classified into six classes. For the purpose of this research, the ten protein expression variables were labelled as B1, B2, B3, B4, B5, B6, B7, B8, B9 and B10. Besides the comparison of classification accuracy and the number of rules generated from algorithms involved in this study, the comprehensibility of some selected models was also compared with a ruleset created by clinicians, presented in the form of the decision tree shown in Fig. 1. In the clinician rulesets, the plus and minus sign indicates a 'high' or 'low' value respectively for the corresponding variable. Note that the rules created by the clinician are considered as the benchmark for comparative purposes.

A. Selection of Non-fuzzy Rule-based Classification Algorithms

Seven non-fuzzy classification algorithms that generate models in the form of a ruleset were chosen to perform the classification tasks. These algorithms were chosen from the algorithms categorised as rule-based classifiers in the WEKA machine learning toolbox [3]. The selected algorithms are Conjunctive Rule, Decision Table, JRip, NNge, OneR, PART and Ridor. Note that these algorithms can be used for classification of datasets with nominal class labels. The DTNB algorithm was excluded because the classification accuracy is calculated based on a combination of a Decision Table with a Naïve Bayes classifier, and hence is not a purely rule-based algorithm. The ZeroR algorithm was also excluded because it is simply the default classifier (predicting the majority class). Hence, the selection of algorithms was made based on the capability to make a fair comparison, especially on the comprehensibility of the generated rulesets. Brief details of the selected algorithms are given in the WEKA software. In conducting the experiments using WEKA, default parameter settings were chosen for all of the algorithms. However, for the purposes of comparing the classification accuracy obtained with the same number of rules, some parameters were adjusted in order to obtain the desired number of rules.

B. Generation of FuzzQSBA Simplified Rulesets

One of the most important tasks in the fuzzy rule generation process is the creation of a partition for each of the variables that will be used to generate the model and for the testing purposes. Hence, a simple method to create such a fuzzy partition was developed based on the median and the percentile values of each variable. These values were used to create triangular or trapezoidal membership functions which represent fuzzy terms 'low', 'medium' and 'high' for each of the variables. These partitions are then used for fuzzification of the dataset.

From the initial analysis of the rulesets generated by the FuzzyQSBA rule simplification method, it has been observed that most antecedents (93.33%) of the rulesets are either the linguistics terms 'low' or 'high'. These initial partitions were then simplified further. For eight of the ten variables, for which the median value is non-zero, the term 'medium' was combined with 'high' whereas for the other two variables for which the median value is zero, the term for 'medium' was combined with 'low'. These new partitions were used to create the final simplified rulesets. Sigmoid functions were used rather than the piece-wise linear functions for the implementation of the final model in order to avoid non-classified instance due to the use of Mamdani-type inference. The first simplified ruleset, denoted as Simplified FuzzyQSBA(I) consisted of all variables in the dataset, whereas the second simplified ruleset, Simplified FuzzyQSBA(II), was further simplified. This final ruleset was created by removing any variable that featured the same fuzzy linguistic term for all rules.

C. Comparison of the Generalisation Ability and the Generated Rulesets

Ten non-fuzzy rule-based models generated from the seven non-fuzzy classification algorithms were compared with three fuzzy rule-based models obtained using the method proposed in this study. The average classification accuracy were calculated based on ten-fold cross-validation and presented in Table I. It can be observed that although JRip(I), NNge, PART(I) and Ridor(I) produced higher classification accuracy compared to FuzzyQSBA, Simplified FuzzyQSBA(I) and Simplified FuzzyQSBBA(II) models,

TABLE I
COMPARISON OF NUMBER OF RULES AND CLASSIFICATION ACCURACY

Classification Model	Number of Rules	Classification Accuracy
Conjunctive Rule	1	41.48
Decision Table	79	69.08
JRip(I)	14	87.78
JRip(II)	6	79.03
NNge	78	89.59
OneR	1	48.72
PART(I)	19	89.14
PART(II)	6	80.69
Ridor(I)	127	86.88
Ridor(1I)	6	65.61
FuzzyQSBA	6	79.63
Simplified FuzzyQSBA(I)	6	79.03
Simplified FuzzyQSBA(II)	6	82.50

the non-fuzzy models contain a higher number of rules. It also has been discovered that the reduction on the number of rules through parameter setting for some of these non-fuzzy algorithms has resulting in significant reduction on the classification accuracy (see Table I for JRip(II), PART(II) and Ridor(II) classification accuracies). The results also show that models generated based on FuzzyQSBA algorithm in general produced comparable classification accuracy compared with the other non-fuzzy classification algorithms that utilise six classification rules, and show that the Simplified FuzzyQSBA(II) model produced the highest classification accuracy (for six rules).

The ability of a classification algorithm in producing rulesets that can be explicitly expressed in the form of linguistic rules is an important aspect to assist decision-making process. Analysis conducted from models generated by the Conjuctive Rule and OneR algorithms shows that these algorithms generate only one default rule which obviously cannot be utilised to explain the characteristic of each possible class. For the Decision Table and NNge algorithms, the main setback is that these algorithms generate a large number of rules which cannot be reduced further through parameter setting.

Apart from the models generated based on the FuzzyQSBA algorithm, once again JRip, PART and Ridor were the only non-fuzzy algorithms that are capable of producing rulesets with a small number of rules which can be translated further into the form of linguistic rules (Fig. 2 – Fig. 4). However, in general, these models contain a default rule that is not an explicitly explained characteristic of the ruleset. For example, in the model generated by PART (Fig. 3), the last rule was created to represent any cases that are not covered by the rules for Classes 2 to 6. On the other hand, the simplified FuzzyQSBA model (Table III) contains a fixed number of linguistic rules created to represent each of the possible classification outcomes. It can be observed that these rules are easier to interpret compared to the other non-fuzzy models because these rules were explicitly expressed in the form of natural language and explicitly explain the characteristic of each generated rule.

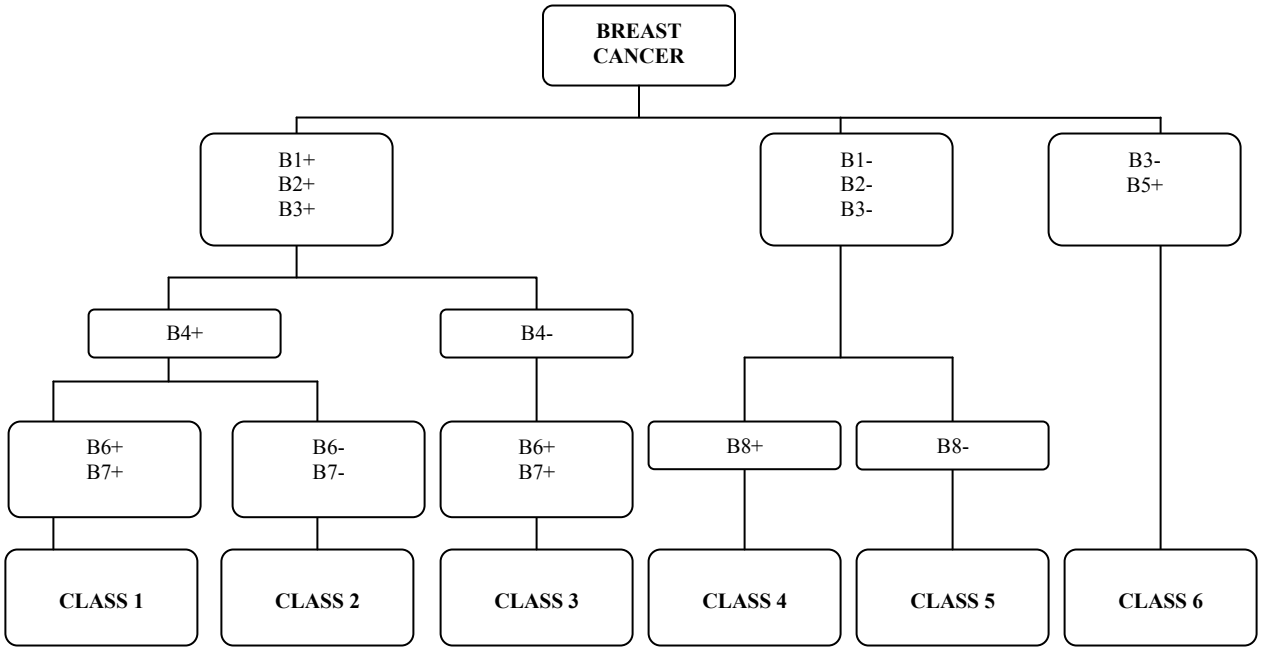

Fig. 1. Decision tree created by (Expert) Clinician.

TABLE II
TRUTH VALUE OF FUZZY QUANTIFIERS FOR EACH LINGUISTIC TERM

Classification	Terms	Variables									
		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
Class 1	**Low**	0.0617	0.0315	0.0650	0.1260	0.6339	0.0634	0.0648	0.9066	0.1656	0.0626
	High	0.9383	0.9685	0.9350	0.8740	0.3661	0.9366	0.9352	0.0934	0.8344	0.9374
Class 2	**Low**	0.1611	0.2246	0.1129	0.0712	0.6364	0.7874	0.8206	0.9464	0.4151	0.1045
	High	0.8389	0.7754	0.8871	0.9288	0.3636	0.2126	0.1794	0.0536	0.5849	0.8955
Class 3	**Low**	0.2003	0.2311	0.1158	0.5604	0.5431	0.3487	0.2081	0.8382	0.2454	0.9329
	High	0.7997	0.7689	0.8842	0.4396	0.4569	0.6513	0.7919	0.1618	0.7546	0.0671
Class 4	**Low**	0.9299	0.9608	0.9626	0.9834	0.7377	0.3092	0.2580	0.0000	0.4444	0.7446
	High	0.0701	0.0392	0.0374	0.0166	0.2623	0.6908	0.7420	1.0000	0.5556	0.2554
Class 5	**Low**	0.9787	0.9874	0.8455	0.8972	0.7154	0.5529	0.4356	0.7351	0.4438	0.7245
	High	0.0213	0.0126	0.1545	0.1028	0.2846	0.4471	0.5644	0.2649	0.5562	0.2755
Class 6	**Low**	0.3122	0.2512	0.8460	0.8602	0.0698	0.2695	0.1274	0.4592	0.1584	0.1057
	High	0.6878	0.7488	0.1540	0.1398	0.9302	0.7305	0.8726	0.5408	0.8416	0.8943

TABLE III
LINGUISTIC RULESETS EXTRACTED FROM FUZZYQSBA

Classification	Variables									
	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
Class 1	High	High	High	High	Low	High	High	Low	High	High
Class 2	High	High	High	High	Low	Low	Low	Low	High	High
Class 3	High	High	High	Low	Low	High	High	Low	High	Low
Class 4	Low	Low	Low	Low	Low	High	High	High	High	Low
Class 5	Low	Low	Low	Low	Low	Low	High	Low	High	Low
Class 6	High	High	Low	Low	High	High	High	High	High	High

```
B2 <= 92 AND B8 > 125: Class4 (52.0/5.0)
B5 <= 95 AND B2 > 92 AND B10 > 135 AND B7 > 85: Class1
(128.0/15.0)
B5 <= 95 AND B1 > 190 AND B10 > 125: Class2 (104.0/19.0)
B5 <= 95 AND B1 > 190: Class3 (64.0/16.0)
B5 > 90: Class6 (52.0/6.0)

Else : Class5 (42.0/3.0)
```

Fig. 2. JRip classification rules.

```
(B1 <= 150) and (B8 <= 125) and (B5 <= 60) => class=C5 (56.0/2.0)
(B5 >= 100) and (B4 <= 165) => class=C6 (66.0/3.0)
(B10 <= 120) and (B8 <= 130) => class=C3 (89.0/20.0)
(B8 >= 150) and (B2 <= 110) and (B6 >= 20) => class=C4 (74.0/0.0)
(B7 <= 80) and (B6 <= 180) and (B4 >= 5) and (B5 <= 90) =>
class=C2 (120.0/4.0)

Else => class=C1 (258.0/67.0)
```

Fig. 3. PART classification rules.

```
Class = C6  (663.0/586.0)

Except (B5 <= 72.5) => class = C4  (372.0/4.0) [189.0/4.0]
Except (B8 <= 112.5) => class = C5  (299.0/0.0) [150.0/0.0]
Except (B1 > 247.5) => class = C3  (227.0/0.0) [103.0/0.0]
Except (B10 > 177.5) => class = C2  (149.0/0.0) [83.0/0.0]
Except (B7 > 112.5) => class = C1  (74.0/0.0) [31.0/0.0]
```

Fig. 4. Ridor classification rules.

Additionally, in order to classify new instances, the firing strength for each rule in the form of membership value degree is used to determine the final classification outcome. This characteristic is one of the significant advantages of the fuzzy rules that the non-fuzzy classification algorithms are unable to offer. Furthermore, the original FuzzyQSBA model induces fuzzy quantifiers where the truth value of the quantifiers (Table II) can be interpreted directly from the generated values. This is very useful to assist the user in order to choose the appropriate linguistic terms 'high' or 'low' to represent each of the variables. This is particularly useful for those non-expert in fuzzy systems to understand

how the rules are created. These values are also easy to interpret as they represent the characteristics of the quantifier. Furthermore, these linguistic rules which are in the form of *complete rules* that can be further simplified by removing any of the variables. This can be done manually based on the truth values of the quantifiers, expert knowledge or using any appropriate automated technique. For example, in Simplifed FuzzyQSBA(II), the variable 'B9' which features the linguistic term 'high' in all the rules was manually removed (Table III). This further simplification process which may be termed *rule-pruning* can make the rulesets simpler while also creating better classification results.

Another important finding in this study is that all the linguistic rulesets extracted from the FuzzyQSBA models through the 10-fold cross-validation are consistent across the ten training datasets. Analysis conducted for the models generated by non-fuzzy algorithms shows that different sets of rules were created for each training datasets. Hence, it may be concluded that the FuzzyQSBA algorithm generates more consistent models compared to the non-fuzzy classification algorithms. Although the rulesets are different compared to the clinician's ruleset in terms of the variables chosen as the antecedents of each rule, it is consistent from the point of view of using the linguistic terms 'high' or 'low' to represent each of the variables. Obviously, these linguistic rulesets can be very useful and helpful for any clinician in creating 'expert rulesets' that can be used to assist decision making.

IV. Conclusion

This paper has presented a novel rule simplification technique to extract linguistic rulesets from a data-driven quantifier-based fuzzy classification system. The comparative study shows that the models generated from the proposed fuzzy technique are capable of producing comparable classification accuracy compared to models generated by non-fuzzy algorithms, and even higher classification accuracy in some cases, for rulesets with similar numbers of rules. Furthermore, although some of the non-fuzzy models can be considered as very comprehensive, it is very obvious that the linguistic ruleset generated from the fuzzy technique have far better advantages compared to their non-fuzzy counterpart in terms of 1) readability of the rulesets, 2) explicitly expressed characteristics of each rule and, 3) providing information on the firing strength of each rule. Additionally, the linguistic rulesets created using the proposed technique were found to be not only consistent throughout the 10-fold cross validation training datasets but also consistent with the expert rulesets in terms of the linguistic terms used to represent each of the variables.

Note that there are many other fuzzy approaches that generate understandable fuzzy linguistic rulesets in addition to the method presented in this paper (see for example [4, 11]). However, the main difference of the proposed system in this research with the other systems is the use of fuzzy quantifiers that make it easier for non-experts in fuzzy systems to understand how the linguistic terms are chosen as the antecedents of the rulesets.

The findings of this research also show that the rulesets generated from an automated classification system can be utilised to an assist an expert in creating their own 'expert-rules'. However, it should be noted that such a task can be considered as a rule-pruning task with the involvement of expert knowledge or additional empirical knowledge. Future research may find that an automated system will perform better when information given by experts is employed in the pruning process rather than in the knowledge discovery process. Hence, further research on this key issue is a worthy topic to be explored. Additionally, it would be very interesting to know the consistency of the final pruned rulesets when the simplified FuzzyQSBA rulesets are pruned manually by different experts.

Acknowledgment

The authors are grateful to all members of the Nottingham Breast Cancer Research Team [10] for the contribution of the dataset used in this research.

References

[1] H. Roubos, and M. Setnes: 'Compact and transparent fuzzy models and classifiers through iterative complexity reduction', *IEEE Transactions on Fuzzy Systems*, 2001, vol. 9, (4), pp. 516-524.

[2] L.I. Kuncheva: '*Fuzzy Classifier Design*' Physica-Verlag, 2000.

[3] I.H. Witten, and E. Frank: '*Data Mining: Practical Machine Learning Tools and Techniques*' Morgan Kaufman, 2005, 2nd edn.

[4] G. Castellano, A.M. Fanelli, and C. Mencar: 'Classifying data with interpretable fuzzy granulation'. *Proceedings of the 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on Advanced Intelligent Systems 2006*, Tokyo, 2006, pp. 872-877.

[5] K.A. Rasmani, and Q. Shen: 'Modifying fuzzy subsethood-based rule models with fuzzy quantifiers'. *Proceedings of the 13th IEEE International Conference on Fuzzy Systems*, Budapest, Hungary, 2004, pp. 1679-1684.

[6] Y. Yuan, and M.J. Shaw: 'Induction of fuzzy decision trees', *Fuzzy Sets and Systems*, 1995, vol. 69, (2), pp. 125-139.

[7] K.A. Rasmani, and Q. Shen: 'Data-driven fuzzy rule generation and its application for student performance evaluation', *Applied Intelligence*, 2006, vol. 24, pp.305-309.

[8] M.A. Vila, J.C. Cubero, J.M. Medina, and O. Pons: 'Using OWA operators in flexible querry processing', in '*The Ordered Weighted Averaging Operators: Theory, Methodology and Applications*', 1997, pp. 258-274.

[9] G. Bordogna, and G. Pasi: 'Application of OWA operators to soften information retrieval systems', in '*The Ordered Weighted Averaging Operator: Theory, Methodology and Applications*', 1997, pp. 275-292.

[10] D. Abd El-Rehim, G. Ball, S. Pinder, E. Rakha, C. Paish, J. Robertson, D. Macmillan, R. Blamey, and I. Ellis: 'High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses.' *International Journal of Cancer*, 2005, vol. 116, pp. 340-350.

[11] H. Ishibuchi, T. Nakashima, and T. Murata: 'Three-objective genetics-based machine learning for linguistic rule extraction', *Information Sciences*, 2001, vol.136, pp. 109-133.