Using Fuzzy Dependency-Guided Attribute Grouping in Feature Selection

Richard Jensen and **Qiang Shen**
{richjens,qiangs}@dai.ed.ac.uk

Centre for Intelligent Systems and their Applications
School of Informatics, The University of Edinburgh

Abstract. Feature selection has become a vital step in many machine learning techniques due to their inability to handle high dimensional descriptions of input features. This paper demonstrates the applicability of fuzzy-rough attribute reduction and fuzzy dependencies to the problem of learning classifiers, resulting in simpler rules with little loss in classification accuracy.

1 Introduction

The task of gathering information and extracting general knowledge from it is known to be the most difficult part of creating a knowledge-based system. The present work aims to induce low-dimensionality rulesets from historical descriptions of domain features which are often of high dimensionality. In particular, a recent fuzzy rule induction algorithm (RIA), as first reported in [2], is taken to act as the starting point for this. In order to speed up the RIA and reduce rule complexity, a preprocessing step is required. Fuzzy-rough attribute reduction (FRAR) with fuzzified dependency is introduced here to address this issue. This step reduces the dimensionality of potentially very large feature sets while minimising the loss of information needed for rule induction.

2 Fuzzy-Rough Attribute Reduction

Rough Set Attribute Reduction (RSAR) described in [6] can only operate effectively with datasets containing discrete values. As most datasets contain real-valued attributes, it is necessary to perform a discretization step beforehand. However, membership degrees of attribute values to fuzzy sets are not exploited in the process of RSAR. By using *fuzzy-rough* sets [3], it is possible to use this information to better guide attribute selection. Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, where $\mathbb{U}$ is a non-empty set of finite objects (the universe); $\mathbb{A}$ is a non-empty finite set of attributes such that $a : \mathbb{U} \to V_a$ for every $a \in \mathbb{A}$; V_a is the value set for attribute a. In a decision system, $\mathbb{A} = \{\mathbb{C} \cup \mathbb{D}\}$ where $\mathbb{C}$ is the set of conditional attributes and $\mathbb{D}$ is the set of decision attributes.

The fuzzy lower approximation is defined here as:

$$\mu_{\underline{P}X}(x) = \sup_{F \in \mathbb{U}/P} min(\mu_F(x), \inf_{y \in \mathbb{U}} max\{1 - \mu_F(y), \mu_X(y)\}) \qquad (1)$$

The tuple $< \underline{P}X, \overline{P}X >$ is called a fuzzy-rough set. FRAR builds on the notion of the fuzzy lower approximation to enable reduction of datasets containing real-valued attributes [4]. The membership of an object $x \in \mathbb{U}$, belonging to the fuzzy

positive region can be defined by

$$\mu_{POS_P(\mathbb{D})}(x) = \sup_{X \in \mathbb{U}/\mathbb{D}} \mu_{\underline{P}X}(x) \qquad (2)$$

Using the definition of the fuzzy positive region, the new dependency function can be defined as follows

$$\gamma'_P(\mathbb{D}) = \frac{|\mu_{POS_P(\mathbb{D})}(x)|}{|\mathbb{U}|} = \frac{\sum_{x \in \mathbb{U}} \mu_{POS_P(\mathbb{D})}(x)}{|\mathbb{U}|} \qquad (3)$$

More details of the fuzzy-rough attribute reduction and the fuzzy-rough QUICKREDUCT algorithm can be found in [4]. This employs the new dependency function γ' to choose which attribute to add to the current list of selected features (termed a *reduct candidate*) in the same way as the original QUICKREDUCT process [6]. The algorithm terminates when the addition of any remaining attribute does not increase the dependency (such a criterion could be used with the original QUICKREDUCT algorithm).

3 Fuzzy Dependency

By its definition, the degree of dependency measure (whether using crisp or fuzzy-rough sets) always lies in the range [0,1], with 0 indicating no dependency and 1 indicating total dependency. By fuzzifying the output values of the dependency function it is hoped that the problem of noise and non-generality can be successfully tackled. In addition to this, attributes may be *grouped* at stages in the selection process depending on their dependency label, speeding up the reduct search.

The goal of RSAR and FRAR is to find a (possibly minimal) subset of the conditional attributes for which the degree of dependency is at a maximum (typically the value 1). In the case of fuzzy equivalence classes, where an element of uncertainty is introduced, the maximum degree of dependency may be substantially less than this. In fact, the maximum dependency for different datasets may be quite different due to differing levels of uncertainty. The maximum for dataset A may be 0.9 whereas for dataset B the maximum may be only 0.2. Given a degree of dependency of 0.19, for dataset A this is quite a small value but for dataset B this is quite large, so a certain way of scaling the dependency value depending on the dataset is required. The following is one potential way of achieving this for a subset P of all conditional attributes $\mathbb{C}$:

$$\gamma''_P(\mathbb{D}) = \frac{\gamma'_P(\mathbb{D}) - \gamma'_{low}(\mathbb{D})}{\gamma'_{high}(\mathbb{D}) - \gamma'_{low}(\mathbb{D})}$$

The new fuzzy-rough QUICKREDUCT algorithm (FQR) which employs scaling and fuzzy dependencies can be seen in figure 1. In using fuzzy degrees of dependency, a variety of selection strategies may be used. Indeed, it is possible to change strategy at any stage of the attribute selection process. The main distinction to make in the set of possible strategies is whether features are chosen individually or in groups.

$$
\begin{aligned}
&(1) \quad R \leftarrow \{\},\ \gamma'_{best} \leftarrow 0,\ \gamma'_{prev} \leftarrow 0, \\
&(2) \quad \textbf{do} \\
&(3) \quad\quad Cands \leftarrow \{\},\ \gamma'_{prev} \leftarrow \gamma'_{best},\ \gamma'_{high} \leftarrow 0,\ \gamma'_{low} \leftarrow 1 \\
&(4) \quad\quad \forall x \in (\mathbb{C} - R) \\
&(5) \quad\quad\quad T \leftarrow R \cup \{x\} \\
&(6) \quad\quad\quad Cands \leftarrow Cands \cup (x, \gamma'_T(\mathbb{D})) \\
&(7) \quad\quad\quad \textbf{if } \gamma'_T(\mathbb{D}) > \gamma'_{high} \textbf{ then } \gamma'_{high} \leftarrow \gamma'_T(\mathbb{D}) \\
&(8) \quad\quad\quad \textbf{else if } \gamma'_T(\mathbb{D}) < \gamma'_{low} \textbf{ then } \gamma'_{low} \leftarrow \gamma'_T(\mathbb{D}) \\
&(9) \quad\quad Cands \leftarrow \mathrm{scale}(Cands, \gamma'_{high}, \gamma'_{low}) \\
&(10) \quad\quad R \leftarrow R \cup \mathrm{selectFeatures}(Cands) \\
&(11) \quad\quad \gamma'_{best} \leftarrow \gamma'_R(\mathbb{D}) \\
&(12) \quad \textbf{until } \gamma'_{best} = \gamma'_{prev}
\end{aligned}
$$

Fig. 1. The new fuzzy-rough SmallCaps{QuickReduct} algorithm with fuzzy dependencies

In order to evaluate the utility of the present work and to illustrate its domain-independence, a challenging test dataset was chosen, namely the Water Treatment Plant Database [1]. The dataset consists of historical data charted over 521 days, with 38 different input features measured daily. Each day is classified into one of thirteen categories depending on the operational status of the plant. However, these can be collapsed into just two or three categories (i.e. *Normal* and *Faulty*, or *OK*, *Good* and *Faulty*) for plant monitoring purposes as many classifications reflect similar performance. It is likely that not all of the 38 input features are required to determine the status of the plant, hence the dimensionality reduction step. However, choosing the most informative features is a difficult task as there will be many dependencies between subsets of features.

4 Results

In all experiments here, FQR employs the strategy where all attributes belonging to the highest dependency group are selected at each stage. For the 2-category problem, the FRAR feature selector returns 10 features out of the original 38, whereas FQR returns 12. Figure 2 compares the classification accuracies of the reduced and unreduced datasets on both the training and testing data. The best results for FQR were obtained in the range 0.85 to 0.88, producing a classification accuracy of 82.6% on the training set and 83.9% for the test data. For FRAR, the best accuracies were 83.3% (training) and 83.9% (testing). Compare this with the optimum for the unreduced approach, which gave an accuracy of 78.5% for the training data and 83.9% for the test data. As can be seen, both the FRAR and FQR results are almost always better than the unreduced accuracies over the tolerance range The 3-category dataset is a more challenging problem, reflected in the overall lower classification accuracies produced. Both fuzzy-rough methods, FQR and FRAR, choose 11 out of the original 38 features (but not the *same* features). The results of these approaches can be seen in figure 3. Again, it can be seen that both FQR and FRAR outperform the unreduced approach on the whole. The best classification accuracy obtained for FQR was 72.1% using the training data, 74.8% for the test data. For FRAR the best results were 70.0%
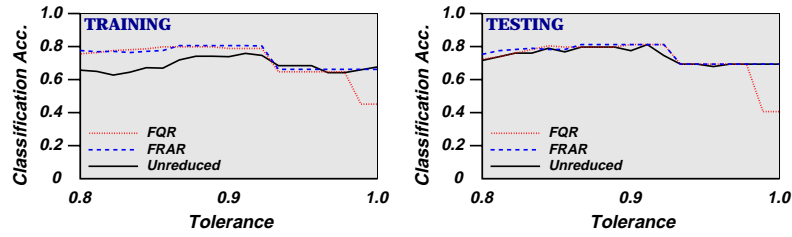
Fig. 2. Classification accuracies for the 2-category dataset.

(training) and 71.8% (testing). In this case, FQR has found a better reduction of the data. For the unreduced approach, the best accuracy obtained was 64.4% using the training data, 64.1% for the test data.
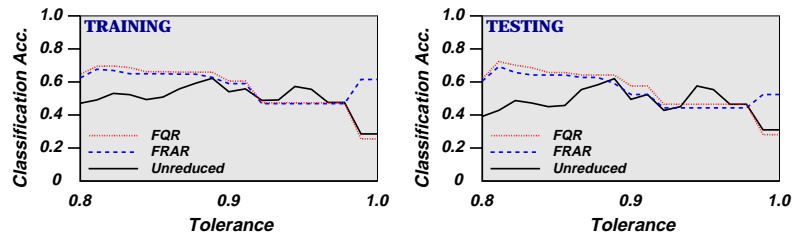

Fig. 3. Classification accuracies for the 3-category dataset.

5 Conclusion

This paper has shown the utility of employing FRAR and FQR as pre-processors to fuzzy rule induction. It has demonstrated the potential benefits of using fuzzy dependencies and attribute grouping in the search for reducts. Not only are the runtimes of the induction and classification processes improved by this step (which for some systems are important factors), but the resulting rules are less complex. Further investigations into FQR include optimization of the dependency fuzzification and selection strategy comparisons.

References

1. C.L. Blake and C.J. Merz. UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science. http://www.ics.uci.edu/~mlearn/
2. S. Chen, S.L. Lee and C. Lee. A new method for generating fuzzy rules from numerical data for handling classification problems. AAI, 15(7):645-664, 2001.
3. D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In Intelligent Decision Support, pp203-232. Kluwer Academic Publishers, Dordrecht. 1992.
4. R. Jensen and Q. Shen. Fuzzy-Rough Sets for Descriptive Dimensionality Reduction. Proceedings of the 11th International Conference on Fuzzy Systems, pp. 29-34, 2002.
5. Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht, 1991.
6. Q. Shen and A. Chouchoulas. A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. EAAI, 13(3):263-278, 2000.