Fuzzy-Rough Set based Semi-Supervised Learning

MacParthalain, Neil; Jensen, Richard

Published in:
2011 IEEE International Conference on Fuzzy Systems (FUZZ)

DOI:
10.1109/FUZZY.2011.6007483

Publication date:
2011

Citation for published version (APA):
MacParthalain, N., & Jensen, R. (2011). Fuzzy-Rough Set based Semi-Supervised Learning. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ)* (pp. 2465-2472) https://doi.org/10.1109/FUZZY.2011.6007483

Fuzzy-Rough Set based Semi-Supervised Learning

Neil Mac Parthaláin
Dept. of Computer Science
Aberystwyth University
Aberystwyth, Ceredigion, Wales. UK
Email: ncm@aber.ac.uk

Richard Jensen
Dept. of Computer Science
Aberystwyth University
Aberystyth, Ceredigion, Wales. UK
Email: rkj@aber.ac.uk

***Abstract*—Much work has been carried out in the area of fuzzy-rough sets for supervised learning. However, very little has been accomplished for the unsupervised or semi-supervised tasks. For many real-word applications, it is often expensive, time-consuming and difficult to obtain labels for all data objects. This often results in large quantities of data which may only have very few labelled data objects. This paper proposes a novel fuzzy-rough based semi-supervised self-learning or self-training approach for the assignment of labels to unlabelled data. Unlike other semi-supervised approaches, the proposed technique requires no subjective thresholding or domain information. An experimental evaluation is performed on artificial data and also applied to a real-world mammographic risk assessment problem with encouraging results.**

***Index Terms*—Rough sets, fuzzy sets, mammographic analysis, semi-supervised learning**

I. Introduction

Each year worldwide, more and more data is collected, and it is estimated that the amount of data collected and stored at least doubles every two years. This collection and storage is facilitated by the fact that the actual process of doing so has become easier in recent times. The result is enormous collections of data. Of this data, a large percentage is unlabelled or has labels which are incomplete or missing. It is because this data is so large that it becomes very difficult, time-consuming and expensive for humans to manually assign labels to data objects. Additionally, many real-world application datasets (such as gene expression data, text classification, etc.) are also of large dimensionality. This further frustrates the process of label assignment for domain experts as large dimensionality can make the task of label assignment intractable.

Supervised learning operates on labelled data and attempts to learn its underlying functional relationships. It is the most common paradigm in machine learning and is typically concerned with the learning of classifiers which can accurately reflect the predictive regularities of the underlying model from the feature values and decision class labels. In unsupervised learning, decision class labels are unavailable and the task is to construct/reconstruct the class information from some inherent structure in the data (also known as cluster analysis). These techniques attempt to find groups in the data such that objects in the same group are similar to each other in some way. The notion of similarity is however subjective and as such unsupervised learning approaches are forced to make subjective assumptions about groupings as well as the number of groups into which objects belong. Semi-supervised learning (SSL) [1] lies somewhere between supervised learning and unsupervised learning. SSL is typically employed when some (but not all) of the data is labelled. The primary aim of semi-supervised learning (SSL) is to try to combine both the labelled and the unlabelled data. SSL has attracted much interest in the field of machine learning due (as mentioned previously) to the abundance of unlabelled data which is available for many real-world problems. SSL can also play an important part as a quantitative method in trying to reason about human category-learning, where most of the data is obviously unlabelled.

There have been many attempts to adapt existing supervised and unsupervised methods for use in the semi-supervised paradigm. These include typical approaches such as neural networks [2], EM algorithm [3], FCM clustering [4], Support Vector Machines [5], Fuzzy Systems [6], and others [7]. One of the main drawbacks of many of the existing semi-supervised approaches however is that they require the specification of at least one additional subjective tunable parameter.

Rough sets [8] and fuzzy-rough sets [9] have enjoyed much attention for the task of supervised learning [10], due to their domain independence and in the case of fuzzy-rough sets the additional ability to handle real-valued data. The vast majority of the work that has been carried out in the areas of rough sets and fuzzy-rough sets has been focused on supervised learning approaches, i.e. where the class labels are known. Very little work has been carried out which employs fuzzy-rough sets for the task of unsupervised learning, and even less still for that of semi-supervised learning. Although there has been some use of rough sets for the task of semi-supervised learning [11] as well as the use of fuzzy sets and rough sets in isolation of each other [12], no use has been made of *fuzzy-rough* sets. The motivation for a fuzzy-rough based semi-supervised approach is based on the success of the supervised approaches [10], [13], and the fact that no additional user-supplied thresholding is required.

This paper presents a novel fuzzy-rough based semi-supervised algorithm for the task of predicting labels for unlabelled data. The algorithm is based on the well-known self-training approach [14], [15], and uses the upper and lower approximation membership values of fuzzy-rough sets for each of the concepts of the labelled objects to iteratively label unlabelled objects.

The remainder of this paper is structured as follows. Section

2 summarises the theoretical basis and concepts of rough sets and fuzzy rough sets. Section 3 describes the fuzzy-rough semi-supervised self learning approach and corresponding algorithm. Section 4 shows the results of applying the fuzzy-rough semi-supervised self-learning approach to a number of artificial datasets, as well as a real-world mammographic risk assessment task. Section 5 concludes the paper with some suggestions, as well as a discussion of ideas for future work.

II. Rough Set and Fuzzy Set Hybridisation

Rough set theory (RST) [8] provides a means by which knowledge can be extracted from a domain in a concise manner. RST can retain information content whilst reducing the amount of information involved. At the core of RST is the concept of indiscernibility. Let $I = (\mathbb{U}, \mathbb{S})$ be an information system, where $\mathbb{U}$ is a non-empty set of finite objects (the universe of discourse) and $\mathbb{S}$ is a non-empty finite set of attributes so that $a : \mathbb{U} \rightarrow V_a$ for every $a \in \mathbb{S}$. V_a is the set of values that a can take. For any $P \subseteq \mathbb{S}$, there exists an associated equivalence relation $IND(P)$:

$$IND(P) = \{(x, y) \in \mathbb{U}^2 | \forall a \in P, a(x) = a(y)\} \tag{1}$$

The partition generated by $IND(P)$ is denoted $\mathbb{U}/IND(P)$ and is calculated as follows:

$$\mathbb{U}/IND(P) = \otimes\{\mathbb{U}/IND(\{a\}) : a \in P\} \tag{2}$$

where,

$$P \otimes Q = \{X \cap Y : \forall X \in P, \forall Y \in Q, X \cap Y \neq \emptyset\} \tag{3}$$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]_p$. Let $X \subseteq \mathbb{U}$. X can be approximated using only the information contained in P by constructing the P-*lower* and P-*upper* approximations of X:

$$\underline{P}X = \{x | [x]_p \subseteq X\} \tag{4}$$

$$\overline{P}X = \{x | [x]_p \cap X \neq \emptyset\} \tag{5}$$

The tuple $\langle \underline{P}X, \overline{P}X \rangle$ is known as a rough set. The problem with RST however, is that because of the definite equivalence imposed by the equivalence relation, it cannot be used effectively on real-valued data. As most data is real-valued, this usually involves a discretisation step prior to the application of RST. Discretisation usually results in a loss of information however and this has prompted the development of a more intuitive and flexible approach by hybridising fuzzy sets and rough sets to give fuzzy-rough sets [9].

A fuzzy-rough set [9] is defined by two fuzzy sets, fuzzy lower and upper approximations, obtained by extending the corresponding crisp rough set notions. In the crisp case, elements that belong to the lower approximation (i.e. have a membership of 1) are said to belong to the approximated set with absolute certainty. In the fuzzy-rough case, elements may have a membership in the range [0,1], allowing greater flexibility in handling uncertainty.

Definitions for the fuzzy lower and upper approximations can be found in [16], where a T-transitive fuzzy similarity relation is used to approximate a fuzzy concept X:

$$\mu_{\underline{R_P}X}(x) = \inf_{y \in \mathbb{U}} \mathcal{I}(\mu_{R_P}(x, y), \mu_X(y)) \tag{6}$$

$$\mu_{\overline{R_P}X}(x) = \sup_{y \in \mathbb{U}} \mathcal{T}(\mu_{R_P}(x, y), \mu_X(y)) \tag{7}$$

Here, $\mathcal{I}$ is a fuzzy implicator and $\mathcal{T}$ a t-norm. A fuzzy implicator is any $[0, 1]^2 \rightarrow [0, 1]$-mapping $\mathcal{I}$ satisfying $\mathcal{I}(0, 0) = 1, \mathcal{I}(1, x) = x$ for all x in $[0, 1]$. R_P is the fuzzy similarity relation induced by the subset of features P:

$$\mu_{R_P}(x, y) = \mathcal{T}_{a \in P}\{\mu_{R_a}(x, y)\} \tag{8}$$

$\mu_{R_a}(x, y)$ is the degree to which objects x and y are similar for feature a, and may be defined in many ways, for example:

$$\mu_{R_a}(x, y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|} \tag{9}$$

$$\mu_{R_a}(x, y) = \max(\min(\frac{(a(y) - (a(x) - \sigma_a))}{\sigma_a}, \frac{((a(x) + \sigma_a) - a(y))}{\sigma_a}, 0) \tag{10}$$

where ${\sigma_a}^2$ is the variance of feature a. As these relations do not necessarily display $\mathcal{T}$-transitivity, the fuzzy transitive closure can be computed for each attribute. The choice of relation is largely determined by the intended application. For example, for the task of feature selection, a relation such as (10) may be best suited, as this allows only small differences between feature values of differing objects. For the task of classification however, a more gradual and inclusive relation such as (9) can be employed. Indeed, this is the relation which is used in the semi-supervised approach described in this paper.

III. Fuzzy-Rough Semi-Supervised Learning

The algorithm employed in this paper is a variation of the popular semi-supervised self-learning or self-training approach [14], [15]. Self learning has been widely employed for the task of semi-supervised learning in text mining applications [7], [11], [14], [15]. It is fast and converges sooner than most other semi-supervised approaches. The basis for the algorithm is that given a set of labelled objects and a set of unlabelled objects, the labelled objects are used to make predictions about the class membership of the unlabelled objects. In self learning/training a classifier (or induction method) is initially trained with a (usually small) amount of labelled data. The classifier is then used to predict labels for the unlabelled data. Typically the most 'confident' unlabelled points, along with their corresponding predicted labels, are then added to the training set. The classifier is then re-trained and the procedure repeated. Notice that the classifier uses its own predictions in order to 'train' itself. One of the criticisms that is often levelled at self-learning approaches is that they can reinforce classification errors that are made early-on in the label assignment. However, it is shown in [14] that this is not strictly the case and the self-learning algorithm can perform equally as well as co-training [17] - another well-known SSL approach. In this paper, a variation of self-learning is employed in that only objects which belong fully to the fuzzy-rough lower approximation of any of the decision classes are added to the labelled data at each iteration, and therefore no subjective confidence level threshold is employed.

A. Fuzzy-Rough Semi-supervised Self Learning Algorithm

The rationale behind the fuzzy-rough self learning (FRSL) algorithm is that the fuzzy-rough lower and the upper approximation memberships of each of the (labelled) decision classes, (calculated by means of all of the labelled objects) to the unlabelled test object y, provide good clues in order to predict class membership.

In practice, the algorithm starts out by examining each object in the unlabelled data. If the unlabelled object belongs to the lower approximation of a given class with fuzzy-rough dependency equal to 1.0, then the label of that class is added to that object. The object is then added to the training data and will be included when generating the fuzzy-rough partitions in the following iterations. The algorithm continues until either: 1) all unlabelled objects have been labelled, or 2) until any of the remaining unlabelled objects do not belong to the lower approximation of any of the labelled decision classes with certainty (fuzzy-rough dependency < 1.0). If there are remaining unlabelled objects, a naive approach is adopted in order to assign labels by calculating the combination of lower and upper approximation memberships and using this value to assign class labels. Adopting such a strategy means that labels are (as far as possible) propagated via only those objects which are 'certain' thus minimising the risk of reinforcement of classification errors at an early stage in the algorithm.

It is important to note that the propagation of labels happens through the addition of previously unlabelled objects to the labelled training data. This has the effect of altering the fuzzy-rough lower and upper approximation memberships of the remaining unlabelled objects.

B. Worked Example

In order to demonstrate the fuzzy-rough concepts employed in the previous section, a small worked example is presented. Note that only a small dataset is included in this example, and as such the formal concepts are only demonstrated rather than the execution of the algorithm. The example dataset has 3 real-valued conditional attributes (a,b, and c) and a single crisp discrete-valued decision attribute (q) as the *labelled training data*, shown in Table I. A further dataset shown in Table II containing 2 objects is used as the *unlabelled* data, again with the same number of conditional attributes but with unknown decision attribute values.

FRSL(L, UL, C, y)
L, the set of all labelled data objects; UL, the set of all unlabelled data objects; C, the set of decision classes in L; y, the unlabelled data object to be labelled

while $(\exists y \in UL : \mu_{\underline{R}C}(y) == 1.0)$
  $buildFuzzyRoughPartition(L)$
    $\forall C \in L$
      if $(\mu_{\underline{R}C}(y) == 1.0)$
        $y \leftarrow y \cup \{C\}$
        $L \leftarrow L \cup \{y\}$
while $UL \neq \emptyset$
  $buildFuzzyRoughPartition(L)$
  $\gamma_{best} \leftarrow 0$, $\gamma_{prev} \leftarrow 0$, $bestClass \leftarrow \emptyset$
    $\forall C \in L$
      if $(\mu_{\underline{R}C}(y) + \mu_{\overline{R}C}(y))/2 > \gamma_{best})$
      $\gamma_{best} \leftarrow \gamma_{prev}, C \leftarrow bestClass$
      $y \leftarrow y \cup \{bestClass\}$
      $L \leftarrow L \cup \{y\}$
end

Fig. 1. The FRSL Algorithm

Object	a	b	c	q
1	-0.4	-0.2	-0.5	yes
2	-0.4	0.1	-0.1	no
3	0.2	-0.3	0	no
4	0.2	0	0	yes

TABLE I
EXAMPLE LABELLED TRAINING DATA

Referring to the FRSL algorithm described in the previous section, the first step is to calculate the fuzzy upper and lower approximations for all decision classes. In Table I there are 4 objects and as noted previously a decision attribute which has 2 classes ($\{yes\}$, and $\{no\}$).

Using the fuzzy similarity measure as defined in (17) the similarity of each unlabelled object is compared to all of the objects in the training data. For instance, consider the unlabelled object $ul1$:

$$\mu_{R_P}(ul1, 1) = T(\mu_{R_{\{a\}}}(ul1, 1), \mu_{R_{\{b\}}}(ul1, 1), \mu_{R_{\{c\}}}(ul1, 1)) = 0$$

$$\mu_{R_P}(ul1, 2) = T(\mu_{R_{\{a\}}}(ul1, 2), \mu_{R_{\{b\}}}(ul1, 2), \mu_{R_{\{c\}}}(ul1, 2)) = 0$$

Object	a	b	c	q
$ul1$	0.3	-0.3	0	-
$ul2$	-0.4	0	-0.3	-

TABLE II
UNLABELLED DATA

$$\mu_{R_P}(ul1,3) = T(\mu_{R_{\{a\}}}(ul1,3), \mu_{R_{\{b\}}}(ul1,3), \mu_{R_{\{c\}}}(ul1,3)) = 0.83$$

$$\mu_{R_P}(ul1,4) = T(\mu_{R_{\{a\}}}(ul1,4), \mu_{R_{\{b\}}}(ul1,4), \mu_{R_{\{c\}}}(ul1,4)) = 0.23$$

These similarity values can then be used to generate the lower and upper approximations. Note that the fuzzy connectives chosen for this example are the Lukasiewicz t-norm ($\max(x+y-1,0)$), and Lukasiewicz fuzzy implicator ($\min(1-x+y,1)$). For the labelled data decision concept $X = yes$ these are:

$$\mu_{\underline{R_P}}X(ul1) = \inf_{y\in\mathbb{U}} \{I(\mu_{R_P}(ul1,y), \mu_X(y)\}$$
$$= \inf\{I(0,1), I(0.0,0), I(0.83,0), I(0.23,1)\} = 0.17$$

and,

$$\mu_{\overline{R_P}}X(ul1) = \sup_{y\in\mathbb{U}} \{I(\mu_{R_P}(ul1,y), \mu_X(y)\}$$
$$= \sup\{T(0,1), T(0,0), T(0.83,0), T(0.23,1)\} = 0.23$$

Similarly for the decision concept $X = no$:

$$\mu_{\underline{R_P}}X(ul1) = \inf\{I(0,0), I(0,1), I(0.83,1), I(0.23,0)\} = 0.77$$

$$\mu_{\overline{R_P}}X(ul1) = \sup\{T(0,0), T(0,1), T(0.83,1), T(0.23,0)\} = 0.83$$

Note that for each of the unlabelled test objects, the object is allowed to belong to the class under consideration to degree 1.0 as the 'true' class is unknown. It can be seen that the upper and lower approximation membership values for test object $ul1$ for the class label $X = yes$ are equal to 0.17, and 0.23 respectively. When $X = yes$ the corresponding values are 0.77 and 0.83. Note that the FRSL algorithm in the first instance will only label objects with membership $\mu_{\underline{R_P}}\ X$ = 1.0 and that the example employed here is to show how these membership values are obtained. If we are to assume that ul1 and ul2 are remaining objects after all others have been labelled then, the algorithm will therefore label $ul1$ naïvely. This is done using the upper and lower approximation memberships and given that: (0.77 + 0.83)/2 = 0.80 for $X = no$ and (0.17 + 0.23)/2 = 0.20 for $X = yes$. Data object $ul1$ will be therefore labelled as belonging to the class $X = no$.

Similarly, the procedure is then repeated for object $ul2$ which results in the following upper and lower approximation values for $X = no$:

$$\mu_{\underline{R_P}}X(ul2) = \inf\{I(0.2,1), I(0.4,0), I(0,0), I(0,1)\} = 0.6$$

$$\mu_{\overline{R_P}}X(ul2) = \sup\{T(0.2,1), T(0.4,0), T(0,0), T(0,1)\} = 0.2$$

And, $X = yes$:

$$\mu_{\underline{R_P}}X(ul2) = \inf\{I(0.2,0), I(0.4,1), I(0,1), I(0,0)\} = 0.8$$

$$\mu_{\overline{R_P}}X(ul2) = \sup\{T(0.2,0), T(0.4,1), T(0,1), T(0,0)\} = 0.4$$

IV. Experimental Evaluation

There is no clear consensus or commonly agreed benchmark datasets for the comparison of semi-supervised learning techniques. The most common way of testing new techniques involves taking the labelled datasets available in [18] and removing labels from a percentage of the data objects. The problem with such an approach is that well-distributed absent labels are rarely the case in the real-world. Indeed, the removal of labels in such a manner may not demonstrate sufficiently the true utility of most semi-supervised approaches.

In light of the above statement, two different approaches have been adopted for the experimental evaluation: (i) Two artificial datasets are examined and compared with some existing methods (ii) a real-world labelling problem for mammographic risk assessment is presented along with some prototype labelled examples.

1) Experimental Setup: For the artificial data, two datasets have been generated which reflect some common realistic problems, and demonstrate how the FRSL method copes with such problems. The data objects contained in the datasets have 2 conditional attributes and belong to two separate classes. The labels of only 5% of the objects are known for the problem shown in Fig. 2 (and represented by ■ for class $+$ and ● for class ○) - thus 2.5% for one class and 2.5% for the other, giving a total of 410 labelled and unlabelled objects. For problem 2 in Fig.6, 10% of the labels are known (again represented by ■ and ●), thus giving a total of 420 labelled and unlabelled objects for this dataset. In the case of the mammographic data there are 322 objects in total, and four decision classes based on the BIRADS [19] labelling scheme to reflect risk of development of breast cancer. 10% of the data (32 data objects) has been labelled using the consensus of three expert radiologists - i.e. where all three agree completely on the classification for a given mammogram.

The FRSL method is compared with some common SSL techniques - SSFCM [4] and another self-learning technique based on FNN [20]. SSFCM was proposed as an extension to the popular FCM clustering algorithm. It uses the labelled data as seed points for the cluster prototypes which are essentially ignored when the clusters are updated. The rationale behind the approach is that all labelled examples should be trusted, which is a reasonable assumption but this can affect the objective function of FCM.

The FNN-based self-learning approach uses a fuzzy classifier to estimate class membership for unlabelled examples and learns in an iterative fashion by adding learned examples to the classifier to the data at the end of each iteration and then retraining. This approach has a number of parameters which must be tuned however: number of nearest neighbours, fuzzifier value, and threshold for class membership. For the experimental evaluation carried out in this paper, the fuzzifier value (m) was set to 2.0, the labelling threshold was varied between 0.8-0.95, and the number of nearest neighbours was varied between 1 and 20. The parameter settings which resulted in the best performance are presented here. No restrictions were imposed algorithm itself in terms of number of iterations, number of examples selected from the unlabelled data for labelling (per iteration), or number of most confidently labelled examples that are added at each iteration to the labelled data.

Although the FRSL approach does not have any subjective tunable parameters, a similarity relation must be selected. The fuzzy similarity relation defined in (9) is employed here as this has been shown to produce good results for the task of supervised learning.

A. Artificial Data

The first artificial dataset as shown in Fig. 2, represents two clusters of data, and has only two conditional features. The problem is that for the given class labels, the clusters do not characterise the classes very well and a supervised classifier will be easily misled. The labelled data objects have been chosen at random from the distribution.

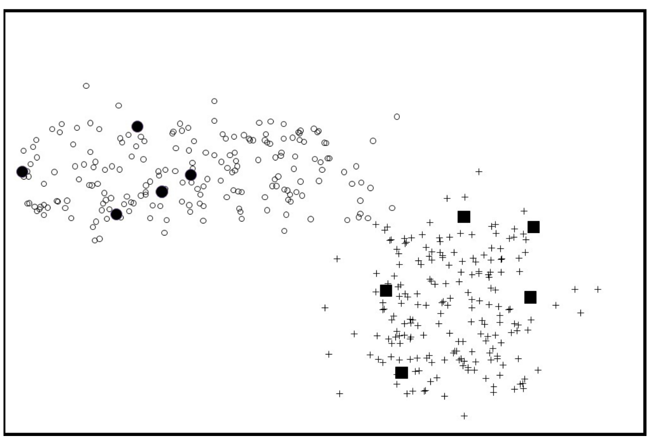

Fig. 2. Artificial Data - Problem 1

SS-FCM was applied to this problem and the results are shown in 3. It can be seen that SS-FCM manages to classify 76% of the objects correctly and fails mostly around the decision boundary between the two classes. Given that the two concepts are very close at this point, and there is some overlap between classes, this performance is quite good.

The FNN wrapper method seems to produce better results that that of SS-FCM but still suffers from poor performance around the decision boundary. It manages to increase classification accuracy to 88% correct classification however.

FRSL manages to label 92% of all objects correctly. Once again the objects are at the boundary between two classes which demonstrates that the overlapping boundary presents a challenge for all methods.

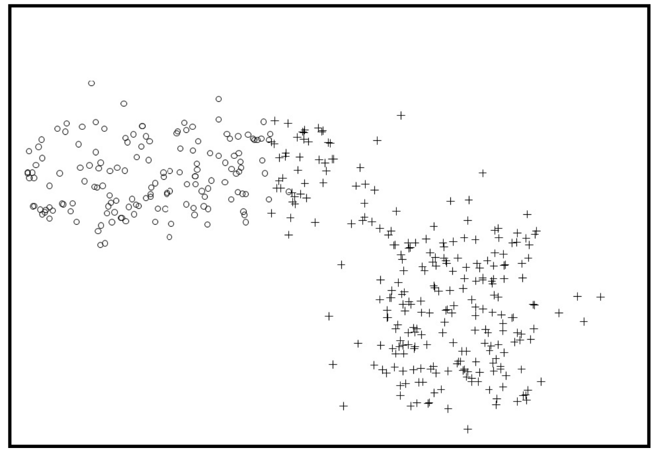

Fig. 3. Artificial Data Problem 1 - SS-FCM

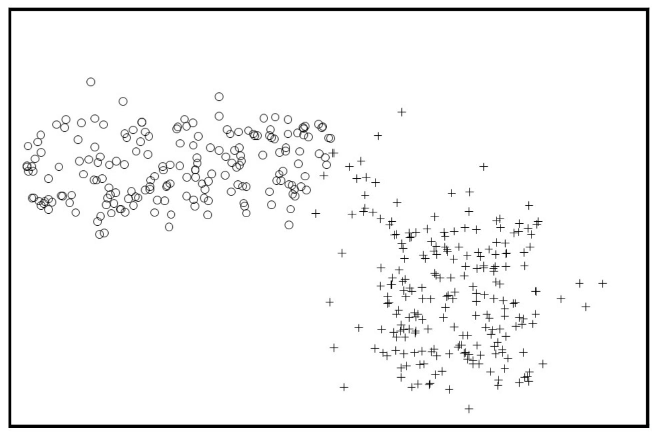

Fig. 4. Artificial Data Problem 1 - FNN Wrapper

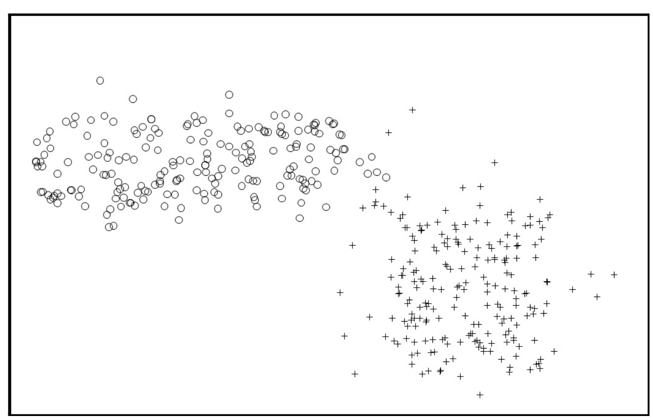

Fig. 5. Artificial Data Problem 1 - FRSL

The second artificial dataset shown in Fig.6 is more complex than that of the previous example. In this case one of the classes has been split into two regions. Once again the labelled objects have been chosen at random, although they now represent 10% of the overall data (20 objects) rather then 5 used in the previous example. This dataset has particularly badly defined decision boundaries and it is certain that a completely unsupervised method would fail in this case.

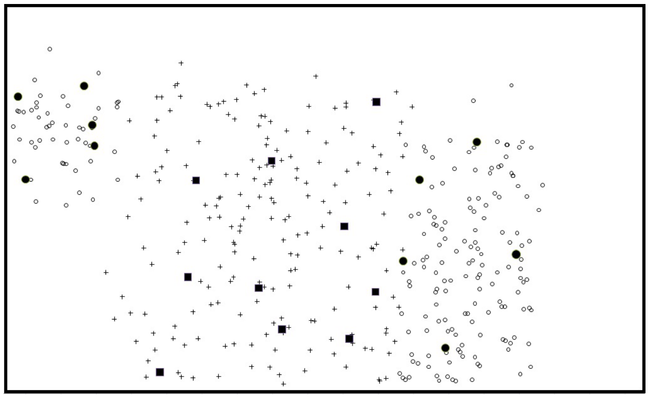

Fig. 6. Artificial Data Problem 2

The SS-FCM method failed completely to correctly label any of the unlabelled data objects that lie in the second cluster which is part of class 1 represented by "O" in Fig.7. The reason for this failure is that SS-FCM does not allow for the induction of more than one cluster for each concept. It also manages to mis-label a number of objects on the decision boundary which is to be expected to a certain degree as the boundary is extremely cluttered. The overall accuracy of the method for correctly labelled instances is 66%.

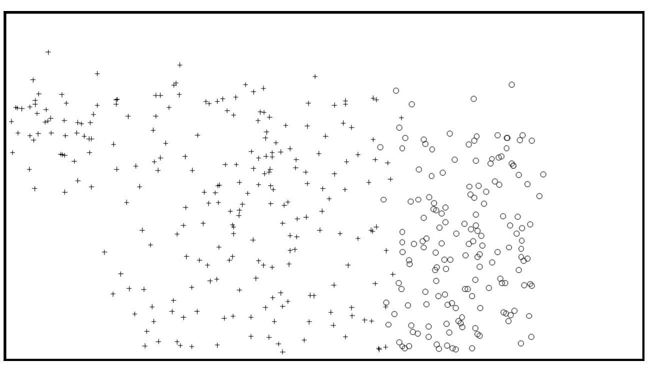

Fig. 7. Artificial Data Problem 2 - SS-FCM

In the case of the FNN wrapper method, performance is quite good and both clusters are identified. However there is a significant amount of mis-labelling of objects. Despite this, the method correctly labels 78% of all objects. Interestingly, some of the assignments appear to be quite strange around the boundaries, with some objects that are distant from the relevant labelled examples. This may be more to do with the sequence of self-learning and value for the number of neighbours than any inherent properties of FNN however.

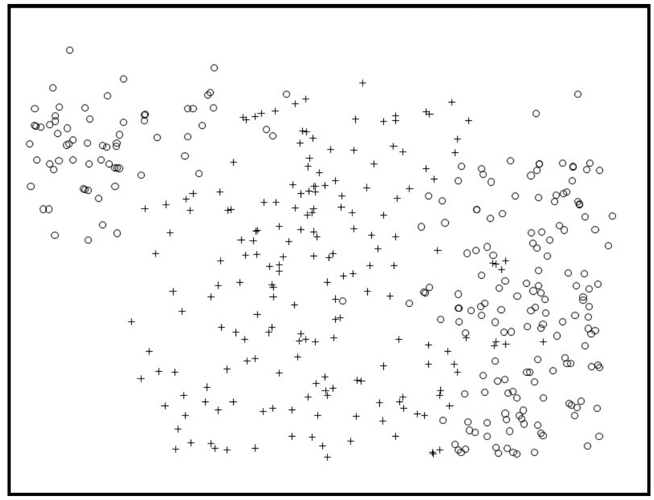

Fig. 8. Artificial Data Problem 2 - FNN

FRSL correctly labels 83% of all of the unlabelled objects for problem 2. Perhaps more importantly however, it manages to label all of those objects represented by the class "O" correctly, the only confusion that arises is with regard to incorrectly labelling objects of the other class.

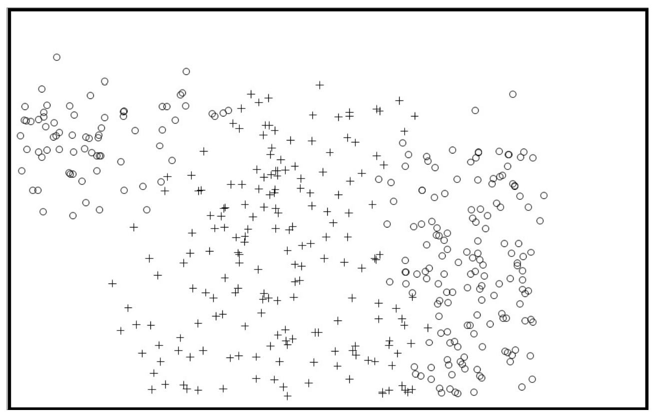

Fig. 9. Artificial Data Problem 2 - FRSL

B. *Mammographic Risk Assessment Data*

Although artificial data often provide useful performance indicators, a mammographic risk assessment problem has been included here in order to demonstrate that FRSL is also useful for such real-world tasks.

Breast cancer is a major health issue, and perhaps the most common amongst women in the EU. It is estimated that between 8% and 13% of all women will develop breast cancer at some point during their lives. Breast tissue characteristics are widely accepted as important indicators of the likelihood of the developing breast cancer [21]. The BIRADS [19] classification scheme allows the division of breast tissue density into 4 distinct classes where 1 represents tissue that is fatty and 4 represents tissue that is dense. The data that is used here has been obtained from [22] and the features have been extracted

TABLE III

CONFUSION MATRICES AND LABEL CLASSIFICATION ACCURACIES FOR THE MIAS DATASET.

FNN (*Classification accy* = 51.50%)

Consensus opinion

	1	*2*	*3*	*4*
1	62	15	8	2
2	40	48	15	0
3	32	22	35	6
4	6	4	6	21

FRSL (*Classification accy* = 65.22%)

	1	*2*	*3*	*4*
1	74	13	0	0
2	15	68	17	3
3	0	36	44	15
4	0	7	6	24

using the method described in [21]. Each of the labelled objects has been labelled using the consensus of three expert radiologists (i.e. where they *all* agree on the classification).

The confusion matrices in Fig. III show the results obtained by both the FNN Wrapper method and the FRSL method. The results for SS-FCM method was not presented for this problem as it fails almost completely to identify any distinct classes. This probably due in the most part to the fact that this is not a clustering task.

The true accuracy for this dataset is 68% when all of the objects have been annotated by experts, FRSL manages 65.21% correct label assignment which is close to this figure. FNN seems to fail to label classes 1 and 2 in particular, whereas one would expect the confusion to lie between classes 2 & 3 as these are conceptually close and are the classes upon which experts disagree most. FNN however shows a large amount of confusion between classes 1 and 3, and also between 1 and 4 which of course are not neighbouring classes. FRSL demonstrates results which do seem to follow a diagonal trend with low non-neighbouring class confusions and a better correct number of label assignments.

V. CONCLUSION

This paper has presented a novel algorithm for the semi-supervised labelling of unlabelled data based on the fuzzy-rough upper and lower approximation membership. Note that no user-defined thresholds are required for the FRSL method, although a choice must be made regarding fuzzy similarity relations and connectives.

Further work in this area will include a more in-depth experimental investigation of the proposed method and the impact of the choice of fuzzy relations, connectives as well as the learning speed of FRSL and number of instances labelled at each iteration etc. Another area worthy of investigation is the combination of the use fuzzy discernibility matrices on the unlabelled data and the use of the fuzzy-rough method on the labelled data, such that the information in both could be aggregated to build classifiers. This may help to address the current assumption for the FRSL method that the labelled objects are fully representative of the class distribution which of course may not be the case. One area is particular which may yield improvements in performance is that of multi-view learning or ensemble semi-supervised classification, where a number of weakly supervised classifiers are aggregated in order to label unlabelled data.

ACKNOWLEDGMENT

Neil Mac Parthaláin would like to acknowledge the financial support for this research through *The Research Institute of Visual Computing - Wales (RIVIC)*.

REFERENCES

[1] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006.

[2] A. Verikas, A. Gelzinis, and K. Malmqvist, Using unlabeled data for learning classification problems. In New learning paradigms in soft computing, Lakhmi C. Jain and Janusz Kacprzyk (Eds.). *Physica-Verlag Studies In Fuzziness And Soft Computing Series*, Vol. 84, Physica-Verlag GmbH, Heidelberg, Germany, pp. 367–402, 2002.

[3] K. Nigam, and R.Ghani, Analyzing the effectiveness and applicability of co-training. *Ninth International Conference on Information and Knowledge*, Management, pp. 86-93, 2000.

[4] A.M. Bensaid, L.O. Hall, J.C. Bezdek, and L.P. Clarke, Partially supervised clustering for image segmentation, *Pattern Recognition*, vol. 29 no.2, pp. 859-871. 1996.

[5] K.P. Bennett, A. Demiriz, Semi-supervised support vector machines, *Proceedings of Neural Information and Processing System*, pp. 368-374. 1998.

[6] B. Gabrys, A. Bargiela, *General fuzzy minmax neural network for clustering and classification*, IEEE Transactions on Neural Networks, vol.11, no. 3, pp.769-783, 2000.

[7] X.J. Zhu, and A.B. Goldberg, Introduction to semi-supervised learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 3, no.1, pp. 1-130, 2009.

[8] Z. Pawlak, Rough sets. *Int. J. Comput. Inf. Sci.* 11, pp. 341–356. 1982.

[9] D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In R. Slowinski, editor. *Intelligent Decision Support*. Kluwer Academic Publishers, Dordrecht. 1992., pp. 203–232. 1992.

[10] R. Jensen and Q. Shen. *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, IEEE Press and Wiley & Sons, 2008.

[11] L. Shi, X. Ma, L. Xi, Q. Duan, and J. Zhao, Rough set and ensemble learning based semi-supervised algorithm for text classification, *Expert Systems with Applications*, In Press, Corrected Proof, Available online 19 November 2010, ISSN 0957-4174, DOI: 10.1016/j.eswa.2010.11.069.

[12] Z. Xue, Y. Shang and A. Feng, Semi-supervised outlier detection based on fuzzy rough C-means clustering, *Mathematics and Computers in Simulation*, vol. 80, no. 9, pp. 1911–1921, 2010.

[13] R. Jensen and Q. Shen. New Approaches to Fuzzy-Rough Feature Selection. *IEEE Transactions on Fuzzy Systems*, vol.17, no.4, pp. 824–838, 2009.

[14] R. Mihalcea. Co-training and Self-training for Word Sense Disambiguation, *Proceedings of the Conference on Natural Language Learning* (CoNLL 2004), Boston, 2004.

[15] D. Yarowsky, Unsupervised word sense disambiguation rivaling supervised methods. *Proc. of the 33rd Annual Meeting of the Association for Computational Linguistics*, pp. 189-196, 1995.

[16] A.M. Radzikowska and E.E. Kerre, A comparative study of fuzzy rough sets, *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.

[17] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training. *COLT: Proceedings of the Workshop on Computational Learning Theory*, Morgan Kaufmann, pp. 92–100, 1998.

[18] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, *UCI Repository of machine learning databases* [http://www.ics.uci.edu/ mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science, 1998.

[19] American College of Radiology. *Illustrated Breast Imaging Reporting and Data System BIRADS*. 3rd Edition, American College of Radiology, 1998.

[20] J.M. Keller, M.R. Gray and J.A. Givens. A fuzzy K-nearest neighbor algorithm. *IEEE Trans. Systems Man Cybernet.*, vol. 15, no. 4, pp. 580-585, 1985.

[21] A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E.R.E. Denton, R. Zwiggelaar. A Novel Breast Tissue Density Classification Methodology. *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 55–65, 2008.

[22] J Suckling, J. Partner, D.R. Dance, S.M. Astley, I. Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.L. Kok, P. Taylor, D. Betal, and J. Savage. The Mammographic Image Analysis Society digital mammogram database. *International Workshop on Digital Mammography*, pp 211–221, 1994.