Pattern Recognition 46 (2013) 2770–2782

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

FRPS: A Fuzzy Rough Prototype Selection method

Nele Verbiest [a,*], Chris Cornelis [a,b], Francisco Herrera [b]

[a] Ghent University, Department of Applied Mathematics, Computer Science and Statistics, Krijgslaan 281 S9, 9000 Gent, Belgium
[b] University of Granada, Department of Computer Science and Artificial Intelligence, E-18071 Granada, Spain

ARTICLE INFO

Article history:
Received 29 February 2012
Received in revised form
26 February 2013
Accepted 5 March 2013
Available online 15 March 2013

Keywords:
Classification
Fuzzy rough sets
Instance selection
k NN
Prototype Selection

ABSTRACT

The k Nearest Neighbour (k NN) method is a widely used classification method that has proven to be very effective. The accuracy of k NN can be improved by means of Prototype Selection (PS), that is, we provide k NN with a reduced but reinforced dataset to pick its neighbours from. We use fuzzy rough set theory to express the quality of the instances, and use a wrapper approach to determine which instances to prune. We call this method Fuzzy Rough Prototype Selection (FRPS) and evaluate its effectiveness on a variety of datasets. A comparison of FRPS with state-of-the-art PS methods confirms that our method performs very well with respect to accuracy.

1. Introduction

Classification methods aim to predict the class $d(t)$ of a new target instance t, based on the knowledge in the given decision system (the training data X). That is, the attribute values $a_1(t),\ldots,a_m(t)$ of t are given and $d(t)$ needs to be determined, making use of the instances X in the decision system and their attribute and class values.

Many classification methods are available. In this work, we focus on the use of k-Nearest Neighbours (k NN, [1]). It determines the k instances in X closest to t and then assigns t to the class that is best represented among these k neighbours. In case of ties, a class is assigned at random from the candidate classes.

k NN is a simple classification method that does not impose assumptions on the data. Due to its local nature it has low bias; more specifically, the error rate of 1NN asymptotically never exceeds twice the optimal Bayes error rate [2]. On the other hand, the local nature also results in a high variance, that is, k NN is highly susceptible to noisy data [3]. Furthermore, k NN needs high storage requirements and has low efficiency caused by multiple computations of similarities between the test and training samples.

A technique that deals with these weaknesses of k NN is Prototype Selection (PS, [4]). It first selects a subset of instances $S \subseteq X$ and then classifies a new instance t using the k NN rule acting over S instead of over X. PS should not be confused with instance selection [5]. Instance selection methods are designed to serve as a general data reduction technique for all kinds of machine learning methods, whereas PS methods are instance selection methods specifically designed to improve k NN classification.

Rough set theory [6], initiated by Pawlak in the early 80s, is a mathematical approach that deals with imperfect knowledge. It has been used widely for feature selection [7–16]. Extending rough sets to fuzzy rough sets [17] and using them for feature selection has been explored extensively [18–28], but using fuzzy rough sets for instance selection is still in its infancy.

A preliminary attempt to use fuzzy rough sets for instance selection can be found in [29], presenting the Fuzzy Rough Instance Selection (FRIS) technique. It uses fuzzy rough set theory to express for each instance its membership to the fuzzy positive region, that is, the extent to which instances indiscernible from it belong to the same class. Only instances belonging to the positive region more than a certain threshold are retained. As we will discuss in Section 2.2, FRIS has some shortcomings.

The aim of this paper is to present a new PS method based on fuzzy rough set theory that we call Fuzzy Rough Prototype Selection (FRPS). First, the instances are ordered according to a measure based on fuzzy rough set theory that evaluates the lack of predictive ability of the instances, and the instances for which the value exceeds a certain threshold are removed from the training set. To determine this threshold, we consider the values of all instances and use each of them as threshold. The final threshold is the threshold for which applying 1NN to the corresponding reduced training set results in the highest training accuracy.

In order to make our method more robust, we replace the strict max operator in the fuzzy rough measure by the Ordered

* Corresponding author. Tel.: +32 9 264 47 72; fax: +32 9 264 49 95.
E-mail addresses: Nele.Verbiest@Ugent.be (N. Verbiest), chriscornelis@ugr.es (C. Cornelis), herrera@decsai.ugr.es, herrera@ugr.es (F. Herrera).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.03.004

Weighted Average (OWA) operator. These aggregation operators, introduced by Yager in [30], associate weights to the ordered positions of the values and can hence be used to generalize the max operator to a more robust operator.

The remainder of this work is structured as follows: in Section 2, we summarize the related work on PS methods and fuzzy rough approaches to data reduction. In Section 3, we introduce four versions of our new algorithm, FRPS. In Section 4 we first select the best performing method among these four versions and then demonstrate its good performance by applying it on 58 real datasets from the KEEL dataset repository, as in [31], and compare it to 21 state-of-the-art PS methods and FRIS. Finally, we conclude in Section 5.

2. Related work

In this section we briefly present research related to the FRPS method. In Section 2.1, we review the literature on Prototype Selection methods, while in Section 2.2, we discuss data reduction techniques based on (fuzzy) rough set theory.

2.1. Prototype Selection

In [31], an extensive taxonomy on PS methods can be found. In this section we summarize the conclusions of that paper, recall the FRIS algorithm and position our new approach FRPS in the taxonomy.

2.1.1. Type of selection

Below, we list three types of PS methods that can be distinguished based on the sort of instances they select, together with some important representatives.

1. A first class of techniques are *editing methods*. The main goal of these techniques is not to reduce the size of the decision system, but to improve the classification quality of the k NN rule by removing noisy instances. A simple example of such a technique is Edited Nearest Neighbours (ENN, [32]). It considers every instance in the training set and removes it whenever the class predicted by using the k NN rule over the other instances in the training set is different from its true class. Methods derived from ENN include the Modified Edited Nearest Neighbour (MENN, [33]) method and the All k Nearest Neighbour (AllKNN, [34]) method. One of the most effective editing techniques is the Relative Neighbourhood Graph (RNG, [35]) method. The general idea is that after construction of a proximity graph, instances misclassified by their neighbours in this graph are removed. Another editing technique is the Model Class Selection (MoCS, [36]) method that uses a feedback system to incorporate knowledge about the dataset in a tree-based classifier.
2. *Condensation techniques* try to remove superfluous instances. In general, these methods are good at reducing the dimensionality of the decision system. A well-known condensation technique designed specifically for 1NN is Condensed Nearest Neighbours (CNN, [37]). This technique starts off with an empty set $S=\phi$. Then it runs through all instances in the training set and adds an instance to S if it is wrongly classified when applying the 1NN rule over the current set of instances S. As a result, all instances in the decision system will be classified correctly when applying 1NN over S. A more advanced technique is the Reduced Nearest Neighbour (RNN, [38]) technique. This technique first applies CNN to the entire training set X, resulting in a subset $S \subseteq X$. Next, all instances $x \in S$ are considered iteratively. The instance x is temporarily removed from S and it is verified whether all instances in X are classified correctly when applying the 1NN rule over the subset S. If at least one instance is classified incorrectly, x is re-added to S, otherwise, x is removed from S. This is repeated until all instances $x \in S$ have been considered. Other methods derived from CNN are the Fast Condensed Nearest Neighbour (FCNN, [39]) and Modified Condensed Nearest Neighbour (MCNN, [40]) method. Patterns by Ordered Projections (POP, [41]) finds patterns in the training dataset without calculating distances and eliminates instances not satisfying these patterns. Modified Selective Subset (MSS, [42]) retains a consistent subset of instances such that for each instance in the original training set, there is an instance in this subset closer than any other instance. Reconsistent [43] aims to replace neighbouring instances by a single instance.
3. Finally, *hybrid techniques* aim to simultaneously remove noisy and superfluous instances. They are designed to reduce the dimensionality of the decision system and meanwhile improve the classification using the k NN rule. Many of these techniques are based on evolutionary algorithms. For instance, the Generational Genetic Algorithm (GGA, [44,45]), Random Mutation Hill Climbing (RMHC, [46]), Steady-State Memetic Algorithm (SSMA, [47]) and CHC Evolutionary Algorithm (CHC, [48]) are genetic algorithms where the chromosomes correspond to the instances currently selected, and the fitness function depends both on the current reduction rate and the accuracy of the k NN rule over the current chromosome. The Hit Miss Network Edition Iterative (HMNEI, [49]) is a non-evolutionary hybrid PS algorithm. It represents the decision system as a hit and miss network, for which the structural properties correspond to properties of the instances related to the decision of the k NN rule, such as being a noisy or superfluous instance. The Decremental Reduction Optimization Procedure (DROP, [50]) removes instances if this does not cause a decrease of the training accuracy of the current (reduced) training set. The Class Conditional Instance Selection (CCIS, [51]) method introduces the class conditional nearest neighbour to remove instances. C-Pruner [52] computes the order in which instances should be removed and then removes them if this does not result in a drop of training accuracy. The Instance Based 3 (IB3, [53]) method uses a wait and see evidence gathering method to determine which of the saved instances are expected to perform well during classification. Iterative Case Filtering (ICF, [54]), starts off with the ENN algorithm and then employs neighbours and associates to smooth the decision boundaries.

2.1.2. Evaluation of search

Besides labelling PS methods based on the kind of instances they remove, one can also distinguish between filter and wrapper methods.

In the context of PS methods, *filter techniques* use the k NN rule to decide for partial data if they should be removed or added to the selected instances. CNN is such a filter method: an instance is selected if the 1NN rule over the current subset of instances classifies it wrong. ENN is also a filter method: an instance is removed when k NN applied over the universe of instances classifies it incorrectly.

Wrapper methods on the other hand use the k NN rule for the complete training set: many subsets of instances are generated, and each subset is evaluated using a leave-one-out validation scheme. That is, given a subset of instances S, each instance x in the training set X is classified as follows: In case the instance x is in S, the k NN rule is applied over S without the instance x, that is, the neighbours of x are looked up in S but have to be different from x itself. In case x is not in S, the k NN rule is applied over the