International Journal of Approximate Reasoning 69 (2016) 35–57

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Online streaming feature selection using rough sets

S. Eskandari *, M.M. Javidi

Shahid Bahonar University of Kerman, Kerman, Iran

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 9 October 2015
Accepted 10 November 2015
Available online 14 November 2015

Keywords:
Feature selection
Online streaming feature selection
Rough sets theory
Significance

ABSTRACT

Feature Selection (FS) is an important pre-processing step in data mining and classification tasks. The aim of FS is to select a small subset of most important and discriminative features. All the traditional feature selection methods assume that the entire input feature set is available from the beginning. However, online streaming features (OSF) are an integral part of many real-world applications. In OSF, the number of training examples is fixed while the number of features grows with time as new features stream in. A critical challenge for online streaming feature selection (OSFS) is the unavailability of the entire feature set before learning starts. Several efforts have been made to address the OSFS problem, however they all need some prior knowledge about the entire feature space to select informative features. In this paper, the OSFS problem is considered from the rough sets (RS) perspective and a new OSFS algorithm, called OS-NRRSAR-SA, is proposed. The main motivation for this consideration is that RS-based data mining does not require any domain knowledge other than the given dataset. The proposed algorithm uses the classical significance analysis concepts in RS theory to control the unknown feature space in OSFS problems. This algorithm is evaluated extensively on several high-dimensional datasets in terms of compactness, classification accuracy, run-time, and robustness against noises. Experimental results demonstrate that the algorithm achieves better results than existing OSFS algorithms, in every way.

1. Introduction

Although a larger feature set gives more information about a concept, in the supervised classification and learning tasks we need to keep the feature set as small as possible to reduce computational complexity, gain good generalization performance and increase accuracy [16,47]. In the classical feature selection problem, a dataset of m feature vectors consisting of n input features and one or more output feature is given. The task of feature selection is to select the smallest subset of the most important and discriminative input features. A variety of feature selection algorithms have been developed during the last three decades [27,34,1,38,11,6,39,55,3,33,46,54,5,21].

All the traditional feature selection methods assume that the entire input feature set is available from the beginning. However, online streaming features (OSF) is an integral part of many real-world applications. In OSF, the number of feature vectors is fixed while feature set grows with time. There are two major reasons for OSF:

1. The feature space is unknown or even infinite. For example, in bioinformatic and clinical machine learning problems, acquiring the entire set of features for every training instance is expensive due to the high cost of lab experiments

* Corresponding author.
E-mail addresses: eskandari@math.uk.ac.ir (S. Eskandari), javidi@uk.ac.ir (M.M. Javidi).

http://dx.doi.org/10.1016/j.ijar.2015.11.006
0888-613X/

[51]. As another example, in texture-based image segmentation problems, the number of different texture filters can be infinite and therefore acquiring the entire feature set is infeasible [40,53,15]. In all these scenarios, we need to incrementally update the feature set as new features are available over time.

2. The feature space is known but feature streaming offers many advantages. This scenario is common when the feature space is very large, which makes exhaustive search over the entire feature space expensive or even infeasible. For example, a dataset in document classification problems may contain a set of hundreds of thousands of potential features [48] and therefore, considering a batch feature selection algorithm over the entire dataset needs considerable storage and computational time capabilities.

Online streaming feature selection (OSFS) is the task of selecting a best feature subset from so-far-seen features in OSF. Any OSFS method must satisfy three critical conditions; First, it should not require any domain knowledge about feature space, because the full feature space is unknown or inaccessible. Second, it should allow efficient incremental updates in selected features. We usually have a fixed, limited amount of computational time available between each feature arrival, and so we want to use a method whose update time does not increase without limit as more features are seen [40]. Third, it should be as accurate as possible at each time instance to allow having reliable classification and learning tasks at that time instance.

We would like to distinguish OSFS in this work from the previous studies of incremental feature selection (IFS) in [51,19,32,20,31,28]. In IFS, which is also known as standard online feature selection (SOFS) [51,19] and dynamic feature selection (DFS) [31], the number of training instances (feature vectors) grows with time while the number of attributes (features) is assumed to be fixed. IFS is common in monitoring environments using sensor networks such as network traffic monitoring for net controlling applications, streams of stock data reported from various stock exchanges monitoring for financial analysis applications, and query monitoring in an environment like the Internet. This is in some sense dual to our OSFS scenario in this work where features arrive sequentially and training instances are all available from the beginning.

Although several research efforts have been made to address OSFS [40,48,57,53,50], none of these algorithms satisfy all the critical conditions for an OSFS algorithm. In the paper, the OSFS problem is considered from the rough sets (RS) perspective. The main motivation for this consideration is that RS-based data mining does not require any domain knowledge other than the given dataset. This property seems to be an ideal tool in OSF scenarios. Several successful RS-based feature selection algorithms are proposed in the literatures [18,17,29,30,35,45,56,22,36,26]. However, all these algorithms consider the batch feature selection problem and are not applicable to OSF scenarios. In this paper, a new OSFS algorithm, called OS-NRRSAR-SA, is proposed. This algorithm adopts the classical RS-based feature significance concept to eliminate irrelevant features. The efficiency and accuracy of the proposed algorithm is demonstrated using several experimental results.

The remainder of the paper is organized as follows: Section 2 discusses related works. Section 3 summarizes the theoretical background of RSFS along with a look at the rough set extensions and modifications. Section 4 discusses the new OSFS algorithm. Section 5 reports experimental results and Section 6 concludes the paper.

2. Related work

Traditional feature selection methods can be classified into wrapper, filter and embedded methods. In the wrapper approaches [27,34], the classifier or induction algorithm is a part of feature subset evaluation process. For each subset of input features, a classifier is trained and the subset with minimum classification error is selected. Although this method has high accuracy, the exponential number of subsets makes the method computationally expensive. Moreover, using a classifier to evaluate feature subsets, biases the selected subset to that classifier. In the filter approaches, the feature subset evaluation is independent of the classifier or induction algorithm. For each candidate subset of features, evaluation measures such as information [1,3,38], consistency [11] and relevance [38,53] are applied and the best feature subset is selected. These methods are less accurate than wrapper methods. However, they are simple, fast and unbiased to any special classifier [53,16]. In the embedded approaches [39,55], the feature selection is considered as an integral part of a model training process. In such approaches the trainer tries to make a trade off between model complexity and model accuracy by selecting or removing features. Embedded methods are typically more efficient than the wrappers [53]. However the results are biased to the classifier.

The traditional methods discussed above, examine a batch of all candidate features at each iteration to select the best feature subset. Therefore they consider that all the candidate features are available before starting the feature selection process. Adopting the traditional feature selection methods in OSF scenarios poses critical challenges, because the feature space is not available from the beginning in this scenarios.

Motivated by these challenges, several research efforts have been made to address OSFS. Perkins and Theiler proposed a grafting based algorithm for this problem [40]. Grafting is an iterative gradient descent algorithm, which treats the feature selection task as part of a regularized risk minimization problem [39]. In the online version of this algorithm, a newly seen feature is added to the selected features if the improvement in the model accuracy is greater than a predefined threshold λ. While this algorithm is able to handle streaming features, it is ineffective in dealing with true OSF scenarios for three reasons; 1) choosing a suitable λ requires information about the global feature space. 2) This algorithm suffers from the so-called nesting effect [41]. If a previously chosen feature is later found to be redundant, there is no way for it to be