

Review on Online Feature Selection

Miss. Ankita C. Deshmukh¹, Prof.Dr.S.S.Sane²

¹ K. K. Wagh Institute of Engineering Education & Research, Department of Computer Engineering,
Savitribai Phule Pune University
Nashik,India

² K. K. Wagh Institute of Engineering Education & Research, Department of Computer Engineering,
Savitribai Phule Pune University
Nashik,India

ABSTRACT

Feature selection is an important concept in data mining. Batch learning is the mostly used learning algorithm in feature selection. Unlike Batch learning, online learning proves to be the most promising, efficient and scalable machine learning algorithm. Most existing studies of online learning require accessing all the features of training data. But, accessing all attributes becomes a problem when we deal with high dimensional data. To avoid this limitation, we investigate an online learner which will maintain a classifier having small and fixed number of attributes. The key challenge of online feature selection is how to make accurate prediction for an instance using a small number of active features. This is in contrast to the classical setup of online learning where all the features can be used for prediction. We aim to develop novel OFS approaches which are compared with previous classification algorithms and to analyse its performance for real-world datasets with full and partial inputs.

Keywords: Classification, Feature Selection, High Dimensional Data, Online Learning

1. INTRODUCTION

The rapid advance of computer based high-throughput technique has provided unparalleled opportunities for humans to expand capabilities in production, services, communications, and research. Meanwhile, immense quantities of high-dimensional data are accumulated challenging state-of-the-art data mining techniques. Feature selection is an essential step in successful data mining applications, which can effectively reduce data dimensionality by removing the irrelevant (and the redundant) features. In the past few decades, researchers have developed large amount of feature selection algorithms. These algorithms are designed to serve different purposes, are of different models, and all have their own advantages and disadvantages. Feature selection, a process of selecting a subset of original features according to certain criteria, is an important and frequently used dimensionality reduction technique for data mining. It reduces the number of features, removes irrelevant, redundant, or noisy data, and brings the immediate effects for applications: speeding up a data mining algorithm, and improving mining performance such as predictive accuracy and result comprehensibility. For classification, the objective of feature selection is to select a subset of relevant features for building effective prediction models. By removing irrelevant and redundant features, feature selection can improve the performance of prediction models by alleviating the effect of the curse of dimensionality, enhancing the generalization performance, speeding up the learning process, and improving the model interpretability. Feature selection has found applications in many domains, especially for the problems involved high dimensional data. Despite being studied extensively, most existing studies of feature selection are restricted to batch learning, which assumes that the feature selection task is conducted in an offline/batch learning fashion and all the features of training instances are given a priori. Such assumptions may not always hold for real-world applications in which training examples arrive in a sequential manner or it is expensive to collect the full information of training data. For example, in an online spam email detection system, training data usually arrive sequentially, making it difficult to deploy a regular batch feature selection technique in a timely, efficient, and scalable manner. Another example is feature selection in bioinformatics, where acquiring the entire set of features/attributes for every training instance is expensive due to the high cost in conducting wet lab experiments.

2. FEATURE SELECTION

Feature selection is the technique of selecting subset of original features according to certain criteria. It is used for dimension reduction and hence, can be called as dimensionality reduction technique.

Fig.1. illustrates a unified view of feature selection process.

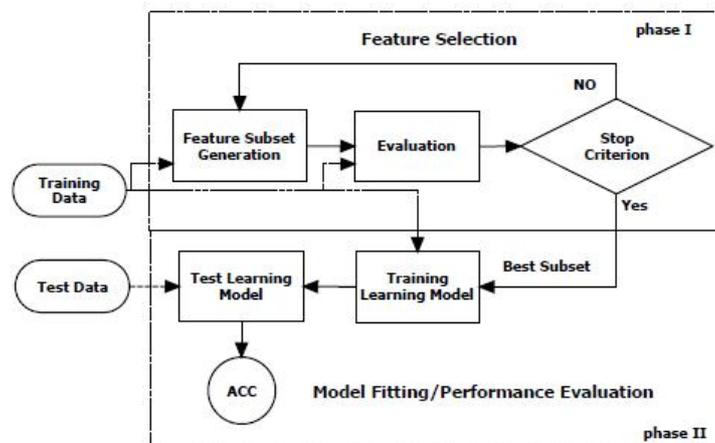


Fig. 1. A general view of Feature Selection Process

There four main components in a feature selection process: feature subset generation, subset evaluation, stop criterion, and results validation. These components work in 2 phases.

PHASE I:

Feature subset generation component will produce candidate feature subsets based on a certain search strategy. Then each candidate subset is further evaluated by a certain evaluation measure and it is compared with the previous best one with respect to this measure. If a new subset turns out to be better, it replaces the previous best subset. The process of subset generation and evaluation is repeated until a given stopping criterion is satisfied.

PHASE II: The finally selected subset is subject to result validation by some given learning algorithms.

2.1. Online Learning

Online learner is allowed to maintain a classifier by involving only a small fixed number of features. The challenge is how to make accurate prediction on an instance using a small number of active features. Online learning is preferred because of its following key features:

- Avoid re-training when adding new data
- High efficiency
- Excellent scalability
- Strong adaptability to changing environments
- Simple to understand
- Trivial to implement
- Easy to be parallelized
- Theoretical guarantee

2.2. Online Learning Methods

1. Online Feature Selection To select a subset of informative features in machine learning tasks for analyzing data with high dimensionality
- Online Collaborative Filtering To learn from a sequence of rating data (sparse rating matrix) for resolving recommender systems
2. Online Multiple Kernel Learning To fuse multiple types of diverse data sources by multiple kernels based machine learning where each kernel represents each type of data/representation

3.RELATED WORK

In this section, overviews of existing feature selection techniques are provided. The objective of this survey is to clearly understand the limitations of existing schemes.

3.1ONLINE PASSIVE-AGGRESSIVE ALGORITHMS [3]

Perceptron algorithm is one of the well known feature selection algorithm. Recently, a large number of online learning algorithms have been proposed in which many of them follow the criterion of maximum margin principle. For example, the Passive-Aggressive algorithm proposes to update a classifier when the incoming training example is either misclassified or fall into the range of classification margin. The PA algorithm is limited in that it only exploits the first order information during the updating. This limitation has been addressed by the recently proposed confidence weighted online learning algorithms that exploit the second order information. Despite the extensive investigation, most studies of online learning require the access to all the features of training instances. In contrast, we consider an online learning problem where the learner is only allowed to access a small and fixed number of features, a

significantly more challenging problem than the conventional setup of online learning. In this paper several online learning tasks are described and analyzed. Author has first introduced a simple online algorithm which we call Passive-Aggressive (PA) for online binary classification. Alternative modifications to the PA algorithm which improve the algorithm's ability to cope with noise are proposed. A unified analysis for the three variants is also proved. Building on this unified view, author show how to generalize the binary setting to various learning tasks, ranging from regression to sequence prediction.

3.2 DIMENSIONALITY REDUCTION via SPARSE SUPPORT VECTOR MACHINES [1]

This work is closely related to sparse online learning, whose goal is to learn a sparse linear classifier from a sequence of high-dimensional training examples. Our work however differs from these studies in that we are motivated to explicitly address the feature selection issue and thus impose a hard constraint on the number of nonzero elements in classifier w , while most of the previous studies of sparse online learning do not aim to explicitly address feature selection, and usually enforce only soft constraints on the sparsity of the classifier. Despite the difference between two kinds of problems and methodologies, we will show empirically in our experiments that our proposed online feature selection algorithm performs better than the cutting-edge sparse online learning algorithms for online classification tasks when the same sparsity level is enforced for the two algorithms. VS-SSVM proved to be very effective on problems in drug design. The number of variables was dramatically reduced while maintaining or even improving the generalization ability. This method outperforms SVMs trained using all the attributes and the attributes selected by correlation ranking. Chemists have found model visualization to be useful both for guiding the modeling process and for interpreting the effects of the descriptors used in the models (Song et al., 2002). Through model visualization, we discovered the simple rule of eliminating variables with weights that flip signs in distinct individual models. Automating this rule proved to be a valuable heuristic for further refining variable selection. VS-SSVM is not a general methodology suitable for all types of problems. This work has demonstrated its effectiveness on very high-dimensional problems with very little data. On problems where linear models cannot adequately capture relationships, the method would fail. Open research areas include a theoretical underpinning of the approach, characterization of the domains on which it is effective, and extension to nonlinear interactions.

3.3 ONLINE STREAMING FEATURE SELECTION [4]

We note that it is important to distinguish online feature selection addressed in this work from the previous studies of online streaming feature selection in. In those works, features are assumed to arrive one at a time while all the training instances are assumed to be available before the learning process starts, and their goal is to select a subset of features and train an appropriate model at each time step given the features observed so far. This differs significantly from our online learning setting where training instances arrive sequentially, a more natural scenario in real-world applications.

1.1 ONLINE FEATURE SELECTION AND ITS APPLICATIONS [5]

In this paper, online learning is introduced. Using it 2 methods are implemented viz., 1) Learning with full inputs 2) learning with partial inputs. Sparsity regularization and truncation techniques are used for developing algorithms. For the first method, it is assumed that the learner can access all the features of training instances, and goal is to efficiently identify a fixed number of relevant features for accurate prediction. In the second task, a more challenging scenario is considered where the learner is allowed to access a fixed small number of features for each training instance to identify the subset of relevant features. To make this problem tractable, the learner allowed to decide which subset of features to acquire for each training instance. OFS algorithm here uses binary classifier and this algorithm is compared with previous algorithms and it is proven that OFS algorithm is the promising family of efficient and scalable algorithms. Various experiments are performed using this algorithm. The proposed techniques are applied to solve two real-world applications: image classification in computer vision and microarray gene expression analysis in bioinformatics. The encouraging results show that the proposed algorithms are fairly effective for feature selection tasks of online applications, and significantly more efficient and scalable than batch feature selection technique.

4. CONCLUSION AND FUTURE WORK

In this paper, we reviewed different techniques of feature selection and efficiency of OFS algorithms against these techniques. OFS aims to select a small and fixed number of features for binary classification in an online learning fashion. OFS addresses two kinds of OFS tasks in two different settings:

1) OFS by learning with full inputs and 2) OFS by learning with partial input. When experimented the encouraging results show that the OFS algorithms [5] are fairly effective for feature selection tasks of online applications, and significantly more efficient and scalable than other batch feature selection techniques. In future, we can use Online Feature Selection novel approaches for online multiclass classification problems and compare the mistake rates thus generated with previous algorithms. Thus, we will come to know whether online multiclass classifier is as efficient as online binary classifier or not.

References

- [1] J. Bi, K.P. Bennett, M.J. Embrechts, C.M. Breneman, and M. Song, "Dimensionality Reduction via Sparse Support Vector Machines," J. Machine Learning Research, vol. 3, pp. 1229-1243, 2003.
- [2] S.C.H. Hoi, J. Wang, P. Zhao, and R. Jin, "Online Feature Selection for Mining Big Data," Proc. First Int'l Workshop Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine '12), pp. 93-100, 2012.
- [3] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, "Online Passive-Aggressive Algorithms," J. Machine Learning Research, vol. 7, pp. 551-585, 2006.
- [4] X. Wu, K. Yu, H. Wang, and W. Ding, "Online Streaming Feature Selection," Proc. Int'l Conf. Machine Learning (ICML '10), pp. 1159- 1166, 2010.
- [5] Jialei Wang, Peilin Zhao, Steven C.H. Hoi, Member, IEEE, and Rong Jin, "Online Feature Selection and its Applications," IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 3, March 2014

AUTHOR



Ankita C. Deshmukh received the B.E. degree in Information Technology from Sir Visvesvaraya Institute of Technology in 2013 and is pursuing M.E. in Computer engineering from K. K. Wagh Institute of Engineering Education and Research , Nasik, India.



Prof. Dr. S. S. Sane, M. Tech (CSE, IITB), Ph D (COEP, Univ. of Pune), Vice Principal, Professor & Head of Dept. of Computer Engineering, Prof. In-charge Central Library. K. K. Wagh Institute of Engineering Education & Research, Nasik, India.