International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.05,Issue.51
December-2016,
Pages:10528-10530

www.ijsetr.com

Online Feature Assortment for Model-Based Strengthening Learning

S. LAKSHMI RAGA SANDHYA[1], M. V. R. NARASIMHA RAO[2], S. SATISH KUMAR[3]

[1]PG Scholar, Dept of CSE, BVC College of Engineering, JNTUK, AP, India, E-mail: sanjusunkara@gmail.com.

[2]Associate Professor, Dept of CSE, BVC College of Engineering, JNTUK, AP, India, E-mail: bvcmnr@gmail.com.

[3]Assistant Professor, Dept of CSE, BVC College of Engineering, JNTUK, AP, India, E-mail: sskumar789@gmail.com.

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 dealing with high dimensional data. To avoid this limitation, and investigating 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. Proposed System aim to develop novel OFS approaches which are compared with previous classification algorithms and to analyze its performance for real-world datasets with full and partial inputs.

Keywords: Feature Selection, Online Learning, Large-Scale Data Mining, Classification, Big Data Analytics.

I. INTRODUCTION

The rapid advance of computer based high-throughput technique has provided unparalleled opportunities for humans to expand capabilities in production, services, communi-cations, 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.

II. 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. 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. Than 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.

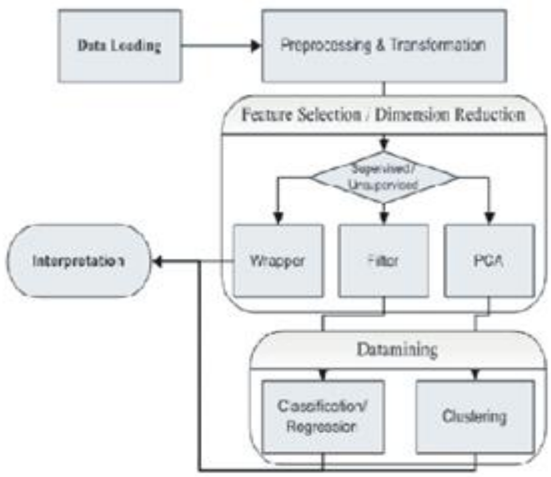

Fig.1. A general view of Feature Selection Process.

Algorithm 1 Modified Perceptron by Truncation for OFS

Input
- B: the number of selected features

Initialization
- $\mathbf{w}_1 = 0$

for $t = 1, 2, \ldots, T$ **do**
  Receive $\mathbf{x}_t$
  Make prediction $\mathrm{sgn}(\mathbf{x}_t^\top \mathbf{w}_t)$
  Receive y_t
  if $y_t \mathbf{x}_t^\top \mathbf{w}_t \leq 0$ **then**
    $\widehat{\mathbf{w}}_{t+1} = \mathbf{w}_t + y_t \mathbf{x}_t$
    $\mathbf{w}_{t+1} = Truncate(\widehat{\mathbf{w}}_{t+1}, B)$
  else
    $\mathbf{w}_{t+1} = \mathbf{w}_t$
  end if
end for

Algorithm 2 $\mathbf{w} = Truncate(\widehat{\mathbf{w}}, B)$

if $\|\widehat{\mathbf{w}}\|_0 > B$ **then**
  $\mathbf{w} = \widehat{\mathbf{w}}^B$ where $\widehat{\mathbf{w}}^B$ is $\widehat{\mathbf{w}}$ with everything but the B largest elements set to zero.
else
  $\mathbf{w} = \widehat{\mathbf{w}}$
end if

A. 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

III. 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.

A. Online Passive-Aggressive Algorithms

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 m is classified 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 project 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.

VI. CONCLUSION AND FUTURE SCOPE

In this project, by reviewing 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 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.

VII. REFERENCES

[1] R. Bekkerman, R. El-Yaniv, N. Tishby, and Y. Winter. Distributional word clusters vs. words for text categorization. Journal of MachineLearning Research, 3:1183–1208, 2003.
[2] J. Bi, K. P. Bennett, M. J. Embrechts, C. M. Breneman, and M. Song. Dimensionality reduction via sparse support

vector machines. Journal of Machine Learning Research, 3:1229–1243, 2003.

[3] G. Cavallanti, N. Cesa-Bianchi, and C. Gentile. Tracking the best hyperplane with a simple budget perceptron.Machine Learning 69(2-3):143–167, 2007.

[4] N. Cesa-Bianchi, S. Shalev-Shwartz, and O. Shamir. Efficient learning with partially observed attributes. Journal of MachineLearning Research, pages 2857–2878, 2011.

[5] A. B. Chan, N. Vasconcelos, and G. R. G. Lanckriet. Direct convexrelaxations of sparse svm. In ICML, pages 145–153, 2007.

[6] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y.Singer.Online passive-aggressivealgorithms.J.Mach. Learn. Res. (JMLR),
7:551–585, 2006.

[7] K. Crammer, M. Dredze, and F. Pereira.Exact convex confidenceweighted learning. In NIPS, pages 345–352, 2008.

[8] K. Crammer, A. Kulesza, and M. Dredze.Adaptive regularization of weight vectors. In NIPS, pages 414–422, 2009.

[9] M. Dash and V. Gopalkrishnan. Distance based feature selection for clustering microarray data. In DASFAA, pages 512–519, 2008.

[10] M. Dash and H. Liu. Feature selection for classification. Intell.Data Anal., 1(1-4):131–156, 1997.

[11] O. Dekel, S. Shalev-Shwartz, and Y. Singer. The forgetron:A kernel-based perceptron on a budget. SIAM J. Comput.,37(5):1342–1372, 2008.

[12] C. H. Q. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. J. Bioinformaticsand Computational Biology, 3(2):185–206, 2005.

[13]D.Donoho.Compressedsensing.IEEETrans.onInformation Theory, 52(4):1289 – 1306, 2006.

[14] M. Dredze, K. Crammer, and F. Pereira.Confidence-weighted linear classification. In ICML, pages 264–271, 2008.

[15] J. Duchi and Y. Singer. Efficient online and batch learning using forward backward splitting. J. Mach. Learn. Res., 10:2899–2934,2009.

[16] Y. Freund and R. E. Schapire.Large margin classification using the perceptron algorithm.Machine Learning, 37(3):277–296, 1999.

[17] C. Gentile. A new approximate maximal margin classificationalgorithm.J. Mach. Learn. Res. (JMLR), 2:213–242, 2001.

[18] K. A. Glocer, D. Eads, and J. Theiler. Online feature selection forpixel classification. In ICML, pages 249–256, 2005.

[19] I. Guyon and A. Elisseeff. An introduction to variable and featureselection.Journal of Machine Learning Research, 3:1157–1182, 2003.

[20] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. In NIPS, 2005.

Author's Profile:

Ms.S.Lakshmi Raga Sandhya, M.Tech student (C.S.E.) Department Computer Science and Engineering, BVC College of Engineering, Rajahmundry, JNTUK, A.P, India.

Mr.M V R Narasimha Rao Associate Professor, M.Tech(Ph.D) Department: Computer Science and Engineering, BVC College of Engineering, Rajahmundry, JNTUK, A.P, India

Mr. S Satish Kumar, Assistant Professor, M.Tech Department Computer Science and Engineering, ,BVC College of Engineering, Rajahmundry, JNTUK, A.P, India.