A Study on Techniques for Online Feature Choice Method

B.Ajay Babu, Nimala.K *(Asst. Professor)*
Dept. of Information Technology, SRM University Chennai, India

Abstract—*Feature choice is a vital technique for data processing. Despite its importance, most studies of feature choice are restricted to batch learning. in contrast to ancient batch learning ways, on-line learning represents a promising family of economical and ascendable machine learning algorithms for large-scale applications. Most existing studies of on-line learning need accessing all the attributes/features of coaching instances. Such a classical setting isn't invariably acceptable for real-world applications once data instances square measure of high spatial property or it's overpriced to accumulate the total set of attributes/features. to deal with this limitation, we investigate the matter of on-line Feature choice (OFC) during which an internet learner is simply allowed to keep up a classifier concerned only atiny low and glued range of options. The key challenge of on-line Feature choice is a way to create correct prediction for associate degree instance employing a tiny range of active options. will be in distinction to the classical setup of on-line learning wherever all the options can be used for prediction. we tend to decide to tackle this challenge by finding out meagerness regularization and truncation techniques. Specifically, this article addresses 2 totally different tasks of on-line feature selection: (1) learning with full input wherever associate degree learner is allowed to access all the options to make a decision the set of active options, and (2) learning with partial input wherever solely a restricted range of options is allowed to be accessed for every instance by the learner. we tend to gift novel algorithms to unravel every of the 2 issues and provides their performance analysis. we tend to value the performance of the projected algorithms for on-line feature choice on many public data base, and demonstrate their applications to real-world issues as well as image classification in laptop vision and microarray gene expression analysis in bioinformatics. The encouraging results of our experiments validate the effectualness and potency of the proposed techniques.*

Keywords— Feature Choice; on-line Learning; Large-scale knowledge Mining; Classification.

1. INTRODUCTION

Feature choice is a vital topic in data processing and machine learning, and has been highly studied for many years in literature. For classification, the target of feature choice is to pick out a set of relevant options for building effective prediction models. By removing extraneous and redundant features, feature choice will improve the performance of prediction models by assuaging the impact of the curse of spatial property, enhancing the generalization performance, dashing up the educational method, and improving the model interpretability. Feature choice has found applications in several domains, particularly for the problems concerned high dimensional information. Despite being studied highly, most existing studies of feature choice area unit restricted to batch learning, which assumes the feature choice task is conducted in an off-line/batch learning fashion and every one the options of training instances area unit given a priori. Such assumptions may not forever hold for real-world applications in which coaching examples arrive in a very ordered manner or it is expensive to gather the total info of coaching data. for instance, in a web spam email detection system, coaching information typically arrive consecutive, making it troublesome to deploy a daily batch feature choice technique in a very timely, efficient, and ascendable manner. Another example is feature choice in bioinformatics, where effort the complete set of features/attributes for every coaching instance is pricey attributable to the high value in conducting wet science lab experiments. Unlike the present feature choice studies, we study the problem of on-line Feature choice (OFC), aiming to resolve the feature choice drawback in a web fashion by effectively exploring on-line learning techniques. Specifically, the goal of on-line feature choice is to develop on-line classifiers that involve solely atiny low and fixed variety of options for classification. on-line feature selection is especially necessary and necessary once a real-world application must subsume ordered training information of high spatial property, like on-line spam classification tasks, wherever ancient batch feature selection approaches cannot be applied directly. In this paper, we tend to address 2 differing types of on-line feature choice tasks: (i) OFC by learning with full inputs, and (ii) OFC by learning with partial inputs. For the first task, we tend to assume that the learner will access all the options of coaching instances, and our goal is to efficiently establish a set variety of relevant options for correct prediction. within the second task, we tend to take into account a tougher situation wherever the learner is allowed to access a set little variety of options for every coaching instance to spot the set of relevant options. To make this drawback magnetized, we tend to enable the learner to decide that set of options to amass for every training instance. The major contributions of this paper include: (i) we tend to 2 propose novel algorithms to unravel each of the on top of OFC tasks; (ii) we tend to analyze their theoretical properties of the projected algorithms; (iii) we tend to validate their empirical performance by conducting an intensive set of experiments; (iv) finally, we tend to apply our technique to unravel real world problems in text classification, laptop vision and bioinformatics. we tend to note that a brief version of this work had been appeared within the rest of this paper is organized as follows. Section two reviews related work. Section three presents the matter and therefore the proposed algorithms further as their theoretical analysis. Section four discusses our empirical studies and Section five concludes this work.

2. CONCEPT EXTRACTION

Our work is closely associated with the studies of on-line learning and have choice in literature. Below we review vital connected works in each areas. One classical on-line learning technique is that the well known Perceptron formula. Recently, a large number of on-line learning algorithms are projected during which several of them follow the criterion of most margin principle for instance, the Passive-Aggressive formula proposes to update a classifier once the incoming coaching example is either misclassified or fall into the vary of classification margin. The PA formula is limited there in it solely exploits the primary order info during the change. This limitation has been addressed by the recently projected confidence weighted online learning algorithms that exploit the second order information. Despite the intensive investigation, most studies of on-line learning needs the access to all the options of coaching instances. In distinction, we consider an internet learning downside wherever the learner is only allowed to access alittle and stuck variety of features, a considerably more difficult downside than the conventional setup of on-line learning. Feature choice (FS) has been studied extensively in the literatures of knowledge mining and machine learning. the prevailing FS algorithms usually are often grouped into 3 categories: supervised, unattended, and semi-supervised FS. supervised FS selects options according to tagged coaching information. supported totally different selection criterions and methodologies, the prevailing supervised FS ways are often more divided into 3 groups: Filter ways, Wrapper ways, and Embedded methods approaches. Filter ways choose important options by measure the correlation between individual options and output category labels, while not involving any learning algorithm; wrapper ways rely on a planned learning formula to come to a decision a subset of vital options. though wrapper ways generally tend to outdo filter ways, they are sometimes additional computationally high-ticket than the filter ways. Embedded ways aim to integrate the feature choice method into the model coaching method. they're sometimes quicker than the wrapper ways and able to offer appropriate feature subset for the training formula. once there's no label info accessible, unattended feature choice attempts to pick out the vital options that preserve the original information similarity or manifold structures. Some representative works embrace Laplacian Score Spectral Feature choice, and also the recently projected ℓ2,1-Norm regular Discriminative Feature choice. Feature choice has found several applications including bioinformatics, text analysis and image annotation. Finally, recent years conjointly witness some semisupervised feature choice ways that exploit each labeled and untagged information info . Our OFC technique usually belongs to supervised FS. We note that it's vital to tell apart on-line Feature Selection addressed during this work from the previous studies of on-line streaming feature choice in . In those works, options are assumed to arrive one at a time whereas all the coaching instances are assumed to be accessible before the training method starts, and their goal is to pick out a set of options and train an appropriate model at on every occasion step given the options observed to this point. This differs considerably from our on-line learning setting wherever coaching instances arrive consecutive, a additional natural situation in real-world applications. Our work is closely associated with distributed on-line learning, whose goal is to be told a distributed linear classifier from a sequence of high-dimensional coaching examples. Our work but differs from these studies in that we tend to ar actuated to expressly address the feature selection issue and therefore impose a tough constraint on the number of non-zero parts in classifier w, while most of the previous studies of distributed on-line learning do not aim to expressly address feature choice, and usually enforce solely soft constraints on the spareness of the classifier. Despite the distinction between 2 forms of problems and methodologies, we'll show by trial and error in our experiments that our projected on-line Feature choice algorithm performs higher than the fashionable sparse on-line learning algorithms for on-line classification tasks once a similar spareness level is enforced for the two algorithms. Finally, we'd wish to distinguish our work from budget on-line learning which aims to be told a kernel-based classifier with a delimited number of support vectors. a typical strategy behind many budget on-line learning algorithms is to get rid of the “oldest” support vector once the most variety of support vectors is reached, that but is not applicable to on-line feature choice. Our work is different from some existing on-line learning work for online dimension reduction, like the net PCA algorithm. not like on-line feature choice that's a supervised learning, on-line spatiality reduction is completely unattended and needs the access to the full options.

3. ONLINE FEATURE CHOICE

3.1 Drawback Setting

In this paper, we have a tendency to think about the matter of on-line feature selection for binary classification. Let t =1,2, . . . , T be a sequence of input patterns received over the trials, wherever every disturbance ∈ Rd could be a vector of d dimension and yt ∈ . In our study, we have a tendency to assume that d could be a large number and for procedure potency we'd like to select a comparatively little range of options for linear classification. additional specifically, in every trial t, the learner presents a classifier wt ∈ Rd which will be wont to classify instance disturbance by a linear operate sgn(w⊤t xt). Instead of using all the options for classification, we have a tendency to need the classifier wt to own at the most B non-zero components, i.e., kwtk0 ≤ B where B &gt; zero could be a predefined constant, and consequently at most B options of disturbance are going to be used for classification. We refer to this drawback as on-line Feature choice (OFC). Our goal is to style a good strategy for OFC that can create alittle range of mistakes. Throughout the paper, we have a tendency to assume kxtk2 ≤ one, t = 1, . . . , T .

3.2 OFC: Learning with Full Input

In this task, we have a tendency to assume the learner is given full inputs of each coaching instance (i.e. x1, . . . , xT). To motivate our formula, we have a tendency to initial gift an easy however non effective formula that merely truncates the options with little weights. The failure of this easy formula motivates United States of America to develop effective algorithms for OFC.

3.2.1 An Easy Truncation Approach

A straightforward approach to on-line feature choice is to change the Perception formula by applying truncation. Specifically, within the t-th trial, once being asked to make prediction, we'll truncate the classifier wt by setting everything however the B largest (absolute value) components in wt to be zero. This truncated classifier, denoted by wB t, is then wont to classify the received instance disturbance. Similar to the Perception formula, once the instance is misclassified, we'll update the classifier by adding the vector ytxt wherever (xt, yt) is that the misclassified coaching example. Formula one shows the steps of this approach. Unfortunately, this easy approach doesn't work: it cannot guarantee alittle range of mistakes. to check this, consider the case wherever the input pattern x will solely take two potential patterns, either xa or xb. For xa, we set its first B components to be one and also the remaining components to be 0.

3.2.2 A distributed Projection Approach

One reason for the failure of formula one is that though it selects the B largest components for prediction, it doesn't guarantee that the numerical values for the unselected attributes square measure sufficiently little, that might doubtless lead to several classification mistakes. we will avoid this problem by exploring the sparseness property of L1 norm, given within the following proposition from. Proposition one: For alphabetic character &gt; 1 and x ∈ Rd, we have kx – x mkq ≤ ξqkxk1(m + 1)1/q−1,m = 1, . . . , d where ξq could be a constant relying solely on alphabetic character and xm stands for the vector x with everything however the largest elements set to zero. Proposition one indicates that once a vector x lives in a very L1 ball, most of its numerical values square measure focused in its largest components, and thus removing the littlest elements can end in alittle modification to the initial vector measured by the Lq norm. Thus, we'll prohibit the classifier to be restricted to a L1 ball, i.e., R = (1) Based on this concept, we have a tendency to gift a replacement approach for Online Feature choice (OFC), as shown in formula three. The online learner maintains a linear classifier wt that has at the most B non-zero components. Once a coaching instance (xt, yt) is misclassified, the classifier is initial updated by on-line gradient descent so projected to a L2 ball to make sure that the norm of the classifier is bounded. If the ensuing classifier bwt+1 has over 4 B non-zero components, we'll merely keep the B components in bwt+1 with the most important solute weights. Finally, Theorem one provides the error certain of formula three.

3.3 OFC: Learning with Partial Inputs

In the higher than discussion, though the classifier w solely consists of B non-zero components, it needs the total knowledge of the instances, namely, each attribute in xt has got to be measured and computed. we will more constrain the matter of on-line feature choice by requiring no over B attributes of disturbance once soliciting input patterns. we have a tendency to note that this might be vital for a number of applications once the attributes of objects are pricey to amass [36], [4]. Evidently, we can not just acquire the B attributes that have non-zero values in the classifier wt. this can be as a result of during this method, the classifier will ne'er be ready to modification the set of attributes with non-zero components, and it's straightforward to get a sequence of training examples that cause a poor classification performance for this approach. To address this challenge, we have a tendency to propose associate ε-greedy online feature choice approach with partial input data by using a classical technique for creating tradeoff between exploration and exploitation. In this approach, we'll pay ε of trials for exploration by randomly selecting B attributes from all d attributes, and the remaining 1−ε trials on exploitation by selecting the B attributes that classifier wt has non-zero values.

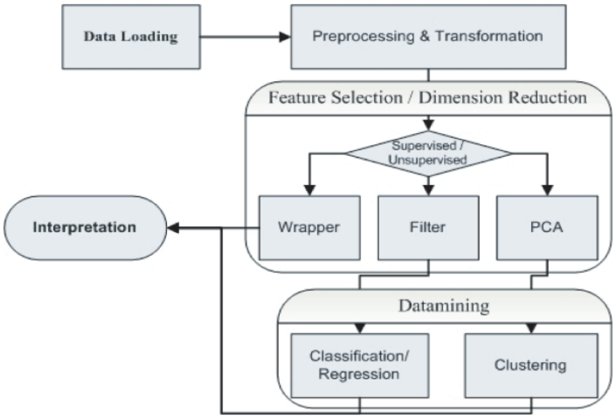
Data Loading
Preprocessing & Transformation
Feature Selection / Dimension Reduction
Supervised / Unsupervised
Wrapper
Filter
PCA
Interpretation
Datamining
Classification/ Regression
Clustering

Fig.1: System Architecture of Proposed System

4. EXPERIMENTAL RESULTS

In this section, we tend to conduct an intensive set of experiments to evaluate the performance of the projected online feature choice algorithms. we'll 1st assess the online prognosticative performance of the 2 OFC tasks on many benchmark datasets from UCI machine learning repository. We will then demonstrate the applications of the projected on-line feature choice technique for two real-world applications by comparison the projected OFC techniques with progressive batch feature choice techniques in literature. we'll conjointly compare the projected technique with regular the present on-line learning technique .

4.1 Experiment I: OFC with Full Input

In this section, we'll introduce the empirical results of the projected on-line Feature choice algorithms in full info setting.

4.1.1 Experimental Tested On UCI And Text Classification Datasets

We check the projected algorithms on variety of in public available benchmarking datasets. All of the datasets will be downloaded either from LIBSVM web site one or UCI machine learning repository a pair of. Besides the UCI information sets, we conjointly adopt 2 high-dimensional real text classification datasets supported the bag-of-words representation: (i) the Reuters Corpus Volume one (RCV1) 3; (ii) twenty Newsgroups datasets 4, we tend to extract the “comp” versus “sci” and “rec” versus “sci” to create 2 binary classification tasks.

4.1.2 Experimental Setup and Baseline Algorithms

We compare the planned OFC algorithmic rule against the following 2 baselines:

• the changed perceptron by the straightforward truncation step shown in algorithmic rule one, denoted as “PEtrun” for short;

• a irregular feature choice algorithmic rule, which randomly selects a set range of active options in a web learning task, denoted as “RAND” for short. To make a good comparison, all algorithms adopt the same experimental settings. we have a tendency to set the quantity of hand-picked features as round(0.1 ∗ dimensionality) for each dataset, the regularization parameter λ to zero.01, and the learning rate η to zero.2. a similar parameters square measure employed by all the baseline algorithms. After that, all the experiments were conducted over twenty times, every with a random permutation of a dataset. All the experimental results were according by averaging over these twenty runs.

4.1.3 Analysis Of On-Line Prognosticative Performance

Table two summarizes the web prognosticative performance of the compared algorithms with a set fraction of hand-picked features (10% of all dimensions) on the datasets. Several observations may be drawn from the results. First of all, we have a tendency to found that among all the compared algorithms, the RAND formula has the best mistake rate for all the cases. This shows that it's necessary to learn the active options in AN OFC task. Second, we found that the easy “PEtrun” formula will outdo the RAND formula significantly, that additional indicates the importance of choosing informative options for on-line learning tasks. Finally, among the 3, we have a tendency to found that the OFC formula achieved the smallest mistake rate, that is considerably smaller than that. This shows that the planned is in a position to significantly boost the performance of the easy “PEtrun” approach.

5. INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing requirement and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input targets on controlling the amount of input required, controlling the errors, avoiding interruption, avoiding extra steps and keeping the process simple. The input is planed in such a way so that it provides security and ease of use with retaining the privacy.

6. OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of dealing are communicated to the users and to other system through outputs. In output design it is divined how the information is to be displaced for immediate need and also the hard copy output. It is the most important source information to the user. Efficient and intelligent output plan improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the correct output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis of planning computer output, they should identify the correct output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system. The output form of an facts of system should accomplish one or more of the objectives.

7. CONCLUSION

This paper investigated a new research problem, Online Feature Selection (OFC), which aims to select a small and fixed number of features for binary classification in an online learning fashion. In particular, we addressed two kinds of OFC tasks in two different settings: (i) OFC by learning with full inputs of all the dimensions/ attributes, and (ii) OFC by learning with partial inputs of the attributes. We presented a family of novel OFC algorithms to solve each of the OFC tasks, and offered theoretical analysis on the mistake bounds of the proposed OFC algorithms. We extensively examined their empirical performance and applied the proposed techniques to solve two real-world applications: image classification in computer vision and microarray gene expression analysis in bioinformatics. The incresing 13 results show the proposed algorithms are fairly effective for feature selection tasks of online applications, and considerably more efficient and scalable than some state-of-the-art batch feature selection technique. Future work could extend our framework to other settings, e.g., online multi-class classification and regression problems, or to help tackle other emerging online learning tasks, such as online transfer learning or online AUC maximization.

References

[1] R. Bekkerman, R. El-Yaniv, N. Tishby, and Y.Winter. Distributional word clusters vs. words for text categorization. Journal of Machine Learning 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 Machine Learning Research, pages 2857–2878, 2011.

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

[6] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. 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] M. Dash and H. Liu. Feature selection for classification. Intell. Data Anal., 1(1-4):131–156, 1997.

[9] 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.

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

AUTHORS PROFILE

B. AJAY BABU Studying Masters of Technology in stream of Information Technology in SRM University, Chennai,India.

Ms.K.NIMALA Assistant Professor (Sr.G), Department of Information Technology, SRM University, Best Teacher Award in 2003.