

A Survey on Online Feature Selection and Different Methods

Catherine Mathew¹, Varghese chooralil²

Abstract— This survey paper talks about the online feature selection. The paper talks on online learning method working on partial and full inputs. The paper discuss on various sparsity regularization methods and how they prevent from over fitting of datasets. Truncation techniques are used to reduce the computation time of processing and discovering relevant datasets. The major advantages is that data used are data points rather than whole datasets. The data sets taken are from UCI machine learning repository. The results of various experiment depicts that the proposed algorithms are fairly effective for feature selection tasks of online applications and more efficient and scalable than some state of the art batch selection techniques

Index Terms— Batch Learning, Datamining , Online feature selection.

1 INTRODUCTION

DATAMINING is field in computer science used for analysing patterns in large datasets. Feature selection is method to select a subset of relevant features[1]. The features so selected are used to build prediction models. The main characteristics of datamining is latent information delivery. The technologies supporting datamining are advanced algorithms multi processor computers and massive databases. Online learning requires access to all features of training instances.

This setup is not always available in real-world applications. When data instances are of high dimensionality or the access to such data are valuable [1]. To overcome this limitation, the concept of online feature selection was introduced. Using this concept the online system is permitted to maintain a classifier consist of features. The disadvantage of this method is that all features are active and can be used for prediction [1]. We overcome this disadvantage by introducing the concept of sparsity regularization and truncation techniques.

The different approaches in various selection are defined using relevant variables. Different approaches are based on different way to analyse which information is relevant. The basic methods are filter methods, wrapper and embedded methods.

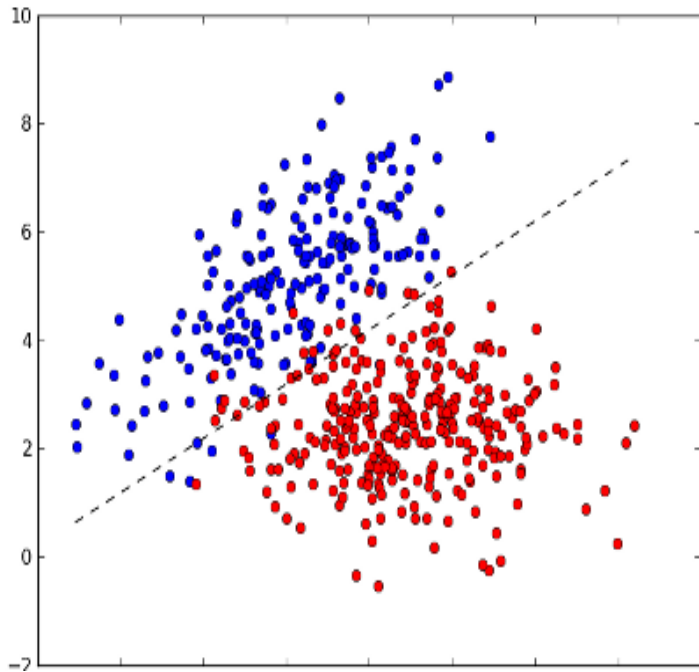
The basic type of sparsity regularization are L1 norm and L2 norm uses the basic idea by minimizing the sum of absolute and between target and between target value and estimated values. The basic characteristics of these regularization techniques are :

1. **Robustness:** L2 norm squares the error when compared to L1 norm. The model is highly sensitive which means disparaging the error when considering the outlier[7].
2. **Stability :** regularization is very important technique to prevent overfitting in this method. It adds a regularization in order to prevent the coefficient to fit so perfectly to overfit. The solution uniqueness is an important property[7]. As in L1 norm is considered L1 norm is considered L1 has unique solution. The L2 norm don't has built in feature selection and L2 norm.

has non-sparse coefficient. The basic intention of sparsity is that a few entries in a matrix or vector form is non-zero. L1 norm has the property of producing many coefficients with zero values or small values with few large coefficients[8]. The computational efficiency of L1 and L2 when studied shows that L1

1. Catherine Mathew is currently pursuing masters degree program in Computer Science and Information System, Rajagiri School of Engineering and Technology, Kochi, Kerala, India. E-mail: catherinemathew167@gmail.com

2. Varghese chooralil is currently working as Asst. Professor in the Dept. of Computer Science and Engineering, Rajagiri School of Engineering and Technology, Kochi, Kerala, India.



has no analytical solution. And L2 has a analytical solution.

FIGURE 2:LARGE LINEAR CLASSIFICATION

2 RELATED WORK

Huan liu and Lei in there paper discuss feature se-

| NO: | PAPER | AUTHOR | APPROACH |
|-----|---|---|--|
| 1 | Toward Integrating Feature Selection Algorithms for Classification and Clustering | Huan Liu, Senior Member, IEEE | Filter Algorithm and Wrapper Algorithm |
| 2 | LIBOL: A Library for Online Learning Algorithms | Steven C.H. Hoi | Journal of Mach First order learning |
| 3 | Large Margin Classification Using the Perceptron Algorithm | yoav@research.att.com AT&T Labs, Shannon | voted-perceptron algorithm. |

lection as NP hard problem [NP hard problem is hardest problem in NP]. The feature are selected by

following step

1. Based on search strategy a subset is generated via search procedure.

2. Once the candidate subset is selected it is evaluated and compared with previous set of best one according to certain evaluation criteria

3 if a new subset is found out to be better then it replace the older.

4. A stopping criteria is to satisfied

A short description of related work is given in tabular form in figure 1

The paper by yoav and Robert discuss on how kernel function uses high dimension spaces .The concept of vavnik on support vector machines[2] .We have you use map the instances into a very high dimensionality space in which the two classes are linearly seperable[8].The use of such functions along with perception algorithm yields a dramatic improvement in performance.The information we obtain during training is the set or list of all predictions [2] .

The paper by steven ch hoi and jailei wang discuss on open source library for large scale online learning.The algorithm

used in LIBOL is based on linear classification.Figure 3 gives a bird view of LIBOL .

The linear classification is method of obtaining decision based on the value of a linear combination is method of obtains decision based on the value of a linear combination of the characteristics[9]. The primary setup for implement linear classifier to take x_+ and x_- separate. each point is described by a set of numbers[3]. The point at index S in the datasets as x_i . In building mathematical model for classify, We are going to focus on dividing these points with straight line. the figure 2 depicts the large linear classification[10]

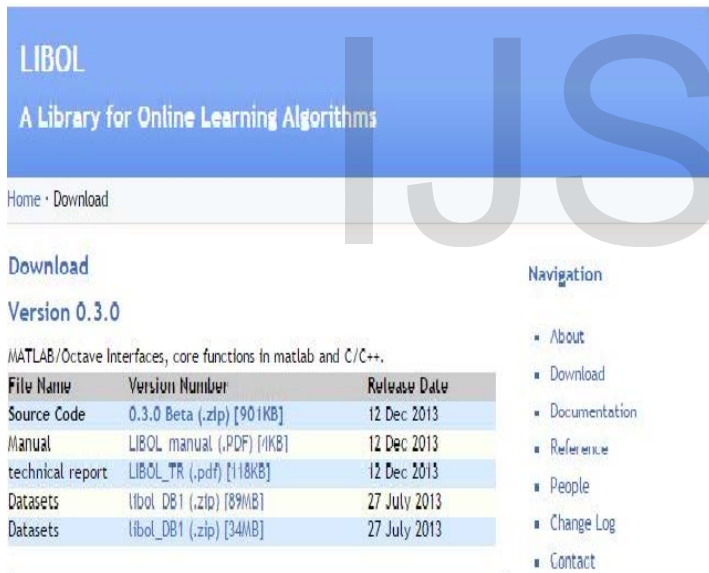


FIGURE 3:LIBOL

3 DIFFERENT METHODS IN DATMINING FOR FEATURE SELECTION

A) Classification with labeled data:

Data are generated in mixture model. we have to classify the data roughly called as documents such as large enough for training data set we can efficiently classify data if large

chunks of examples are provided

B) Perceptron Learning Algorithm:

The perceptron learning rule is Training patterns are presented to the network's inputs; the output is calculated.

The algorithm[5] is as follows:

1. Initialize the weights and threshold to small random numbers.
2. Present a vector x to the neuron inputs and calculate the output.
3. Update the weights
4. Repeat steps 2 and 3 until:

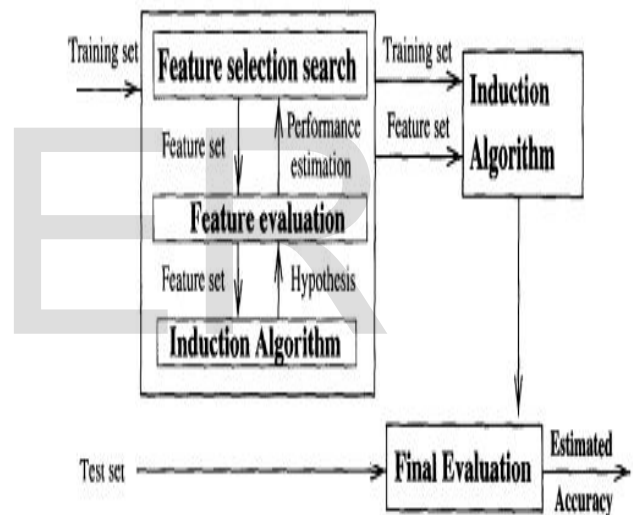


Figure 4:Filter Method

- the iteration error is less than a user-specified error threshold or

- a predetermined number of iterations have been completed

C] Wrapper Model:

Using feature subset selection, a learning algorithm is de-

veloped with the demerit of selecting a relevant subset of features. the wrapper approach discuss on the feature subset

D) Filter Model:

Filter Model are applied over artificial data sets with different number of relevant features, interaction between features and increasing number of samples are taken to evaluate the output [11]. Filter methods, in which the selections of features are independent of the classifier used. The main disadvantage of the filter approach is that it totally rectify the effects of the selected feature subset on the performance of the induction algorithm [3].Figure 4 shows the filter method.

Filter methods that attempt to define relevance autonomously of the learning algorithm. These problems [3]include

1. Inability to remove a relevant feature in symmetric targets concepts
2. Inability to include irrelevant features that helps in performance

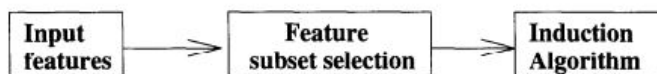


Figure 5:filter method

Examples are methods that select variables either by ranking them through compression techniques (like PCA or clustering) else by computing correlation with the output

E) Hybrid Model:

They takes variable selection as part of the learning procedure and are usually unambiguous to given learning machines Instances are classification trees, random forests, and methods based on regularization techniques (e.g. lasso)

F) Voted perceptron algorithm:

The voted perceptron method is based on the perceptron algorithm of Rosenblatt and Frank. The algorithm takes merit of data that are linearly separable with large margins. This method is simple to implement and much more efficient in terms of computation time as compared to Vapnik's SVM. The algorithm can also be used in very high dimensional spaces using kernel functions.

The experimental comparison of voted perceptron with SVM on classification tasks in NLP and observed that the voted perceptron is comparable to SVM in terms of accuracy and, in addition, as to learning time and prognosis speed of voted perceptron is considerable better than SVM.perceptron is more accurate and voted perceptron is more superior to SVM when taking time into consideration ,prediction and memory footprint

4 CONCLUSION

In this paper we prepensed some well known algorithms concerned with data mining [6]. Feature selection is used basically for online learning. The paper presents an idea on how to improve the performance of searching was improved by using truncation techniques .The paper discuss on how L1 and L2 norms are applied to various known algorithms to improve the performance. The paper depicts how a new research problem was dealt ,OFS which aims to select a small and known number of features for binary classification in an online learning fashion[5].

REFERENCES

- 1] Online Feature Selection for Mining Big Data Steven C.H. Hoi, Jialei Wang, Peilin Zhao, Rong JinSchool of Computer Engineering, Nanyang Technological University, Singapore
- 2] Toward Integrating Feature Selection Algorithms for Classification and Clustering Huan Liu, Senior Member, IEEE, and Lei Yu, Student Member, IEEE
- 3] Wrappers for feature subset selection Ron Kohavi a,*,

George H. John b,l a Data Mining and Visualization, Silicon Graphics, Inc., 2011 N. Shoreline Boulevard, Mountain view, CA 94043, USA b Epiphany Marketing Sofhyare, 2141 Landings Drive, Mountain View, CA 94043, USA

4] Large Margin Classification Using the Perceptron Algorithm yoav@research.att.com AT&T Labs, Shannon Laboratory, 180 Park Avenue, Room A205, Florham Park, NJ 07932-0971

5] Online Feature Selection and Its Applications Jialei Wang, Peilin Zhao, Steven C.H. Hoi, and Rong Jin

6] A Survey Of Data Mining Techniques)1st International Conference on Research in Science, Engineering & Management (IOCRSEM 2014) <http://goo.gl/bwzny5>

7][https://en.wikipedia.org/wiki/Regularization_\(mathematics\)](https://en.wikipedia.org/wiki/Regularization_(mathematics))

8] <http://www.chioka.in/differences-between-the-l1-norm-and-the-l2-norm-least-absolute-deviations-and-least-squares/>

9] <http://www.cs.ucf.edu/~mtappen/cap5415/lects/lec6.pdf>

10] <http://mlpy.sourceforge.net/docs/3.5/liblinear.html>

11] <https://goo.gl/dvzWiS>

IJSER