International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 7, July 2014 **ISSN 2319 - 4847**

Analyzing the Sensitivity of Neighborhood size on Prediction Quality using Collaborative Filtering

Babu Reddy.M
Krishna University, Machilipatnam-521 001, India

ABSTRACT

Traditional statistical techniques are to be extended to handle the high dimensional data during machine learning and prediction tasks. In many of the applications, the size of data will be exorbitant: from retail marketing data to biomedical images and natural resource information. For such high-dimensional statistical explorations, Feature selection is essential. As the dimensionality of data gets increased day by day, there is an essential need to follow new mathematical approaches to deal with the high dimensional data. Unlike traditional statistical approaches, in machine learning and prediction scenarios, the risk minimization is considered as a key issue. In view of this, all the mathematical methods in connection with machine learning are moving around the understanding of the performance of learning theory. In these lines, this paper highlights the dynamics associated with the collaborative filters which are very much useful in prediction tasks. An effort has been made to analyze the relationship among the percentage of training samples used, learning rate applied and neighborhood size, especially in supervised learning environment. Bench mark micro array data set of English Alphabets has been used in the experimental setup.

1. INTRODUCTION

Feature Selection is a process of choosing a subset of original features so that the feature space is condensed according to a certain evaluation criterion. Feature selection is one of the prominent preprocessing steps to machine learning. Feature selection has been an important field of research and development since 1970's and proved very useful in removing irrelevant and redundant features, increasing the learning efficiency, improving learning performance like predictive accuracy, and enhancing comprehensibility of learned results [John & Kohavi, 1997; Liu & Dash,1997; Blum & Langley, 1997]. High dimensional data can contain high degree of redundant and irrelevant information which may greatly influence the performance of learning algorithms. Many researchers have come up with variety of Feature selection algorithms and these can be broadly divided into two categories, namely, the filter model and the wrapper model [Das, 2001; John & Kohavi, 1997]. The filter models will not depend on any specific learning algorithm and relies on general characteristics of the training data. Where as the wrapper model relies on a predetermined learning algorithm and the features will be evaluated through the performance of learning algorithm. For the high dimensional data, the filter model is usually a choice due to its computational efficiency. Feature selection and Feature Ranking algorithms to deal with high dimensional data are usually of Filter type. Feature filter is a function that returns a relevance index R(S/D ; C) which estimates, relevance between a given feature subset S of data D and the task C (usually classification of the data). The relevance index R(S/D ; C) is calculated directly from data, without depending on any induction algorithm. The relevance index may be estimated by an algorithmic procedure such as constructing a decision tree or finding nearest neighbors. Based on the relevance indices computed for individual features X_i; i = 1..N, ranking order $R(X_1)<R(X_2)<….<R(X_n)$ can be established. The features with lowest ranks are filtered out. This may be sufficient for independent features, but if features are correlated, many of them may be redundant. Sometimes, the best pair of features may not even include a single best feature [Langley P 1997, John GH, et al 1994]. And hence, ranking may not give guarantee that the largest subset of suitable features will be found. Filters can also be used to evaluate the usefulness of subsets of features. Conceptually, the filter methods have no direct dependence of the relevance index on the induction algorithms. Even then, by using relevance indices or by evaluation of the feature contributions by the final system, thresholds for feature rejection may be fixed. At first step, features are ranked by the filter approach, but by using adaptive system as a wrapper, the total number of features that are finally taken may be determined. Evaluation of the adaptive system performance (cross validation tests) are done only for limited selected feature sets, but still this approach may be rather costly. So, a simple filter method which can be applied on large dataset to assign weight/rank for all the features and removing redundant features, is to be parameterized in statistically well established way.

1.1 Collaborative Filter

Collaborative filtering is a neighborhood centered approach for prediction tasks. A target sample is matched against the feature set to discover the similar neighbors. Items that are recommended by neighboring samples will be supplied to the target user. Collaborative systems have been widely used in so many areas, such as Ringo system for music album selection[Upendar and Patti, 1995], MovieLens system for selection of movies, Jeter system for jokes filtering [Gupta et al., 1999] and an online radio system-Flycasting[Hauver,2001]. Collaborative filtering system overcomes some limitations of content-based filtering. Instead of contents of the items, the collaborative filters suggest to users based on the ratings of

items, and that can improve the quality of recommendations. In spite of the effectiveness of collaborative filtering, still needs to overcome some challenges to become an efficient information filtering technique.

i) **Cold start problem :**What about the recommendations for items that have not been rated yet ?
ii) Collaborative filtering works well with the recommendations of the neighborhood; it is against the information extraction from contents.

2. RISK MINIMIZATION AND PERSISTANCE

Especially while dealing with the machine learning applications in medical domain, misclassification and expected losses need more attention than the accuracy[Greenshtein. E 2006],[Greenshtein E. 2004 and et al]. This property is termed as Persistence**.** Consider a class model $m(x^T \beta)$ to predict the class model C and the loss function of the class model is defined as: $l(m(x^T \beta), C)$ Then the risk is: $L_n(\beta) = El(m(x^T \beta), C)$ where 'E' is evaluation function, and 'n' is used to stress the dependence of dimensionality. The minimum risk is obtained as $\beta_n^* = \min_\beta L_n(\beta)$. Let β_n' is an estimator, and the persistence requires $L_n(\beta_n^*) - L_n(\beta_n') = 0$, but there no restriction for the consistency between β_n^* and β_n'. Research results[Greenshtein E. 2004 and et al] shows that the LASSO (penalized L 1 least squares) is persistent under the quadratic loss. .It is shown there that for the quadratic loss, LASSO is persistent, but the rate of persistency is slower than a relaxed LASSO[Meinshausen. N 2005].This again shows the bias problems in LASSO.

2.1 Mean absolute error

The M**ean Absolute Error (MAE)** is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The Mean Absolute Error is given as[Hyndman, R. and Koehler A.,2005].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|\,.$$

As the name suggests, the mean absolute error is an average of the absolute errors $e_i = |f_i - y_i|$, where f_i is the prediction and y_i is the true value. Note that alternative formulations may include relative frequencies as weight factors.

3. EXPERIMENTAL SETUP AND RESULTS ANALYSIS

To determine the sensitivity of density of the data set an experiment was carried out with varying values of Learning rate(α) from 0.1 to 0.4 in an increment of 0.1 and percentage of training samples are varied from 10% to 90% in an increment of 10%. Three different sizes are considered as the size of neighborhood. For each of these train/learning ratio values, the experiment was run on bench mark micro array data set of English Alphabets using the Learning Vector Quantization technique. From the results, it has been observed that the quality of prediction increase as we increase percentage of training samples. The learning rate has the influence on classification performance time and the size of the neighborhood has its influence on the quality of classification. These relationships were studied by using the trend of MAE for varying values of chosen parameters like: learning rate, percentage of training samples and size of neighborhood. The data set contains 20,000 samples of alphabets with different fonts and sizes. Each character grid is converted into 16 primitive feature/attributes(edge counts and movements) and the feature values are scaled to a range of 0 to 15. Some portion of the given 20,000 samples can be used for training purpose and the rest for testing purpose.

Source of Information

- Creator: David J. Slate
- Odesta Corporation; 1890 Maple Ave; Suite 115; Evanston, IL 60201
- Donor: David J. Slate (dave@math.nwu.edu) (708) 491-3867
- Date: January, 1991 Number of Instances: 20000. Number of Attributes: 16 . Number of classes : 26 (from A to Z)

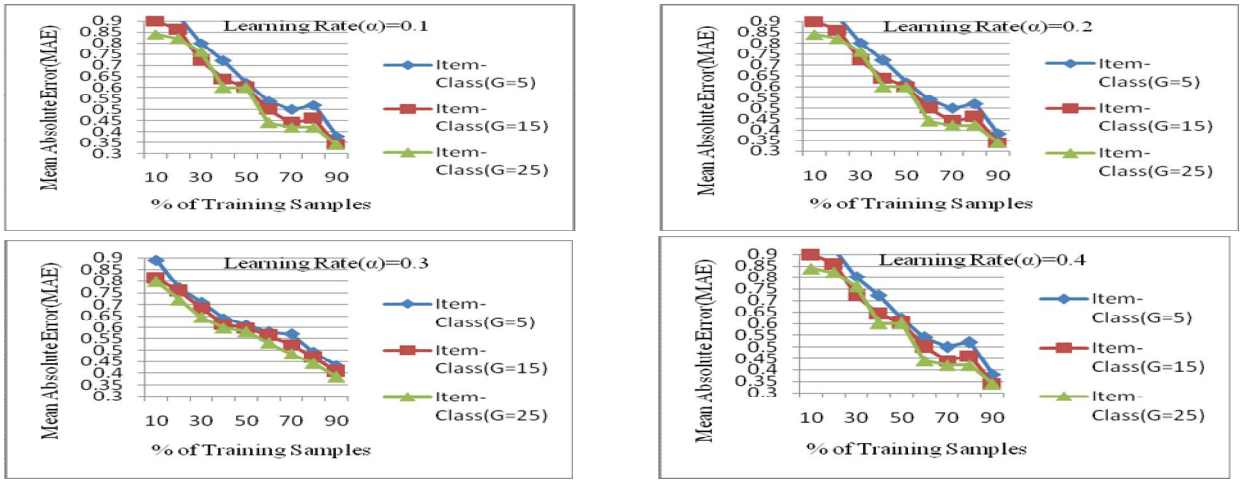
1.	x-box	horizontal position of box	(integer)
2.	y-box	vertical position of box	(integer)
3.	width	width of box	(integer)
4.	high	height of box	(integer)
5.	onpix	total # on pixels	(integer)
6.	x-bar	mean x of on pixels in box	(integer)
7.	y-bar	mean y of on pixels in box	(integer)
8.	x2bar	mean x variance	(integer)
9.	y2bar	mean y variance	(integer)
10.	xybar	mean x y correlation	(integer)
11.	x2ybr	mean of x * x * y	(integer)
12.	xy2br	mean of x * y * y	(integer)
13.	x-ege	mean edge count left to right	(integer)
14.	xegvy	correlation of x-ege with y	(integer)
15.	y-ege	mean edge count bottom to top	(integer)
16.	yegvx	correlation of y-ege with x	(integer)

Missing Attribute Values: None
Num Instances: 20000
Num Attributes: 17(16+1 class attribute)
The experimental results are tabulated in the following table and the trend has been analysed using the line graphs:

MEA –Mean Absolute Error (for varying Learning rate, percentage of training samples, size of neighborhood):

% of Training Samples	Learning Rate(α)=0.1 Neighborhood Size=G			Learning Rate(α)=0.2 Neighborhood Size=G			Learning Rate(α)=0.3 Neighborhood Size=G			Learning Rate(α)=0.4 Neighborhood Size=G		
	G=5	G=15	G=25	G=5	G=15	G=25	G=5	G=15	G=25	G=5	G=15	G=25
10 %	0.89	0.81	0.79	0.93	0.91	0.89	0.98	0.94	0.89	0.98	0.90	0.84
20 %	0.77	0.75	0.72	0.88	0.82	0.79	0.92	0.84	0.80	0.92	0.86	0.82
30 %	0.71	0.68	0.64	0.81	0.78	0.74	0.82	0.74	0.74	0.80	0.72	0.76
40 %	0.64	0.61	0.59	0.74	0.71	0.68	0.70	0.66	0.60	0.72	0.64	0.60
50 %	0.61	0.59	0.58	0.70	0.68	0.64	0.64	0.66	0.58	0.62	0.60	0.60
60 %	0.58	0.56	0.53	0.70	0.64	0.62	0.60	0.56	0.52	0.54	0.50	0.44
70 %	0.57	0.52	0.48	0.62	0.60	0.54	0.52	0.52	0.52	0.50	0.44	0.42
80 %	0.49	0.468	0.442	0.54	0.52	0.52	0.53	0.48	0.42	0.52	0.46	0.42
90 %	0.43	0.406	0.384	0.46	0.46	0.42	0.42	0.44	0.38	0.38	0.34	0.34

The following graphs analyze the experimental results of applying item-class based collaborative filtering technique for performing classification task:



In assessing the quality of classification, sensitivity of parameters like: size of neighborhood, learning rate and percentage of training samples was determined.

The size of the neighborhood has significant impact on the prediction quality. To determine the sensitivity of this parameter, an experiment was conducted with varying number of neighbors and computed MAE. Through the experimental results, it is evident that the size of neighborhood does affect the quality of classification tasks. It is observed that the classification performance is acceptable for moderate size of neighborhood.

4. CONCLUSIONS

In machine learning and prediction applications, the risk minimization is considered as a key concern. Considering this, all the mathematical methods in association with machine learning are moving around the understanding of the performance of learning theory. In this paper, experimental study has been conducted to analyze the dynamics associated with the collaborative filters which are very much useful in prediction tasks. An effort has been made to analyze the relationship among the percentage of training samples used, learning rate applied(α) and neighborhood size(G) especially in supervised learning environment. Bench mark micro array data set of English Alphabets has been used in the

experimental setup. More variations are found in the English Alphabet set because of varying font style and size of each alphabet. And hence, data set with large number of samples is preferable to ensure good classification or prediction accuracy. Possible heuristic measures which include "Rank Based", "Relevance based" and "Redundancy based" approaches to feature subset selection can also be used and the possible hybridization of two or more measures may also lead to better results.

ACKNOWLEDGEMENT

The author of this paper wish to acknowledge Creator and Donor : David J. Slate (dave @ math.nwu.edu), Odesta Corporation; 1890 Maple Ave; Suite 115; Evanston, IL 60201 (708) 491-3867, January, 1991, for his effort in building & maintaining and for his kind heart in donating the bench mark data set of English Alphabets which helped in conducting the experiments and validating the results.

REFERENCES

[1.] Aggarwal, C. C., Wolf, J. L., Wu K., and Yu, P. S. (1999). Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering. In *Proceedings of the ACM KDD'99 Conference*. San Diego, CA. pp. 201-212.

[2.] Langley P 1997 and Sage S, "Scaling to domains with many irrelevant features"; in R. Greiner(Ed), Computational Learning Theory and Natural Learning systems(Vol:4), Cambridge, MA:MIT Press.

[3.] Billsus, D., and Pazzani, M. J. (1998). Learning Collaborative Information Filters. In *Proceedings of ICML '98*. pp. 46-53.

[4.] Upendra, S. and Patti, M.. 1995, Social Information Filtering: Algorithms for Automating "Word of Mouth", In Proc. ACM CHI'95 Conf. on Human Factors in Computing Systems. pp.210-217.

[5.] Hauver, D. B. (2001). Flycasting: using collaborative filtering to generate a play list for online radio. In Proceedings of the international conference on web delivery of music

[6.] Breese, J. S., Heckerman, D., and Kadie, C. (1998), Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52.

[7.] Greenshtein E. 2004], and Ritov. Y, "Persistence in high dimensional predictor selection and the virtue of over parametrization"; Bernoulli 10 (2004), 971–988.

[8.] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D.(1992). Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM. December.

[9.] Good, N., Schafer, B., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., and Riedl, J. (1999). Combining Collaborative Filtering With Personal Agents for Better Recommendations. In Proceedings of the AAAI-'99 conference, pp 439-446.

[10.] Herlocker, J., Konstan, J., Borchers, A., and Riedl, J.(1999). An Algorithmic Framework for Performing Collaborative Filtering. In Proceedings of ACM SIGIR'99. ACM press.

[11.] Meinshausen. N 2005, "LASSO with relaxation" Manuscript, 2005

[12.] Ling, C. X., and Li C. (1998). Data Mining for Direct Marketing: Problems and Solutions. In Proc. of the 4th International Conference on Knowledge Discovery and Data Mining, pp. 73-79.

[13.] Reichheld, F. R., and Sasser Jr., W. (1990). Zero Defections: Quality Comes to Services. Harvard Business School Review, 1990(5): pp. 105-111.

[14.] Sarwar, B., M., Konstan, J. A., Borchers, A., Herlocker, J., Miller, B., and Riedl, J. (1998). Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System. *In* Proceedings of CSCW '98, Seattle, WA.

[15.] Ungar, L. H., and Foster, D. P. (1998) Clustering Methods for Collaborative Filtering. In Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence.

[16.] Babu Reddy.M, Thrimurthy. P and et al..Feature Selection for High Dimensional Data-Empirical Study on the usability of Correlation & Coefficient of Dispersion measures—ICWS 2009, 10-11 Jan,2009 organized by KL University Vijayawada, India.

[17.] Langley P & Sage S(1997): *Scaling to domains with many irrelevant features*- In R. Greiner(Ed), Computational Learning Theory and Natural Learning systems(Vol:4), Cambridge, MA:MIT Press.

[18.] Mark A. Hall – Correlation based feature selection for discrete & numeric class machine learning; pages: 359-366(2000); Publisher: Morgan Kaufmann

[19.] M. Dash and H Liu; Feature Selection for classification and Intelligent data analysis: An International Journal, 1(3),pages: 131-157, 1997.

[20.] SN Sivanandam, S. Sumathi and SN Deepa –Introduction to Neural Networks using Matlab-6.0; TMH-2006

[21.] Dr. M. Babu Reddy- "Independent Feature Elimination in High Dimensional Data: Empirical Study by applying Learning Vector Quantization Method" published in International Journal of Application or innovation in Engineering & Management(IJAIEM), Vol:2, Issue:10, October:2013.(Impact Factor: 2.379 as on 01-11-2013.

[22.] Blake C and Merz C(2006): UCI repository of Machine Learning Databases – Available at: http://ics.uci.edu/~mlearn/MLRepository.html

[23.] Hyndman, R. and Koehler A. (2005). "Another look at measures of forecast accuracy"

AUTHOR

Dr. M. Babu Reddy, has received his Master's Degree in the year 1999 and Doctor of Philosophy in the year 2010 from Acharya Nagarjuna University, AP, India. He has been actively involved in teaching and research for the past 15 years and now he is working as Asst. Professor of Computer Science, Krishna University, Machilipatnam, AP, India. His research interests include: Machine Learning, Software Engineering, Algorithm Complexity analysis and Data Mining.