Active Mode Incremental Nonparametric Discriminant Analysis Learning

Shaoning Pang[1], Kshitij Dhoble[1], Ye Chen[1], Nik Kasabov[1], Tao Ban[2] and Youki Kadobayashi[2]

[1] Knowledge Engineering & Discovery Research Institute, Auckland University of Technology, Private Bag 92006, Auckland 1020, New Zealand

[2] Information Security Research Center, National Institute of Information and Communications Technology, Tokyo, 184-8795, Japan

{spang,hhw9255,zhf0202,nkasabov}@aut.ac.nz, bantao@nict.go.jp, youki-k@is.aist-nara.ac.jp

Abstract: This paper presents a novel active mode incremental nonparametric discriminant analysis (aIncNDA) learning method, in which traditional passive incremental NDA is extended with a one-pass online data selective sampling, then enabled for active online discrimination analysis over data streams. Given an incoming instance y, the proposed aIncNDA computes a simple discrimination residue ratio ν between within-class and between-class, which imitates the fundamental NDA $tr(S_b^{-1}S_w)$ criterion for a maximum separation between classes and minimum separation within classes. In the experiment, we described how the discriminative instances can be significantly selected in line with the discrimination residue with, at most, minor sacrifices in learning rate and classification accuracy. The experimental results show that the proposed aIncNDA is capable of estimating the discriminant contribution for every newly presented instance under different condition of discriminant redundancy, and performing incremental learning gracefully with smaller number of instances learned, but often an improved discrimination performance than the passive incremental NDA.

Keywords: Nonparametric discriminant analysis, incremental NDA, active learning, active mode incremental NDA learning

1 Introduction

Active learning technique is crucial for classification as it iteratively selects distinctive information for training the classifier. Active rather than passive learning is preferred as it performs selective sampling, which enables the learning, immune to noise and data scarcity problems. Owing to its adaptive, evolving and dynamic characteristics it is potentially useful for targeted learning tasks and works well particularly for nonlinear dataset/data stream. By now, active learning has been successfully used in the field of internet security, bioinformatics [9] and text classification [14].

Active learning fundamentally consists of two main components namely the selective sampling engine and the base classifier. Selective sampling is carried out based on a certain criterion, which selects informative instances from the given chunk of data to better the learning function. Thus active learning technique is principally more accurate and computationally efficient than passive learning.

In supervised machine learning for class discrimination, the nonparametric discriminant analysis (NDA) is similar to Linear Discriminant Analysis (LDA) [21], which seeks a transformation towards a maximum separation between classes and minimum separation within classes. Classic NDA is a passive batch learning approach, assumes the entire dataset for training is truly informative and is presented in advance. However in real world applications, data is often being presented at different times in a stream of random chunks, and the quality of data is often not guaranteed due to noise affection. Incremental NDA (IncNDA) [19] somehow has solved the difficulty of NDA and empowered the NDA with an flexibility of incremental learning that accommodate a data stream sequentially. But in spite of that, IncNDA still conducts a rigid learning because IncNDA does not make any instance choices before actual learning, just passively learns whatever instances that are confronted/provided.

In order to overcome NDAs passive learning limitation, we have proposed an active mode incremental NDA learning approach, which incorporates incremental NDA (IncNDA) and selective sampling technique together to form an online active learning. The proposed aIncNDA allows constant informative update of NDA eigenspace obtained from the incoming data.

2 Related Researches and Motivations

The concept of Active learning has only been explored recently. The key to active learning lies in its adaptive selective sampling technique, which selects the most informative instances or data, and eventually boosts the performance of the classifier. The selected data will be assimilated into the training set to retrain the classifier in order to achieve improved level of performance. This procedure can be iterative, since the objective is to achieve a targeted level of performance with least amount of data and high number of informative instances. In our method, incremental NDA is addressed for

Proceedings of the Seventh International Conference on Information and Management Sciences, Kunming & Banna, China, July 20-28, 2009, pp. 216-221.

active learning implementation.

2.1 Approaches of Active Learning

There are varieties of selective sampling approaches used in active learning models. Amongst them, one of the most commonly used technique is Pool-based active learning. However it suffers from multiple drawbacks. Most of the pool-based active learning iteratively selects random samples from the pool which may be informative or irrelevant [2]. Moreover, selecting the samples to be included in the pool itself is a time consuming process. Another selective sampling approach is membership query which selects samples directly from the dataset. Membership query scheme does not have the drawbacks posed by the pool-based scheme. It also reduces the predictive error rapidly and is less computationally intensive.

Clustering [6] and Batch mode active learning [7] are some of the other common flavors of active learning which aims at decreasing the redundancy amongst the selected instances, consequently providing more unique instances for the refinement of classifiers. Lastly, Query by Committee technique [8] is an effective approach, where selective sampling is based on the disagreement amongst ensemble of hypotheses. Some of the frequently used ensemble in this type of active learning includes techniques such as Bagging and Boosting.

For application, incorporation of active learning with support vector machine has been commonly used especially in the field of bioinformatics and text categorization [14]. However majority of them have made use of pool-based technique, which suffers from multiple drawbacks stated above, therefore it is recommended that though incorporation of active learning with SVM is good, other approaches such as membership querying or batch mode active learning should be used as they negate the drawbacks introduced by pool based learning.

2.2 Incremental Discriminant Analysis Approaches

It is well known that Linear Discriminant Analysis (LDA) [21] seeks a transformation towards a global maximum separation between classes and minimum separation within classes. In contrast, another known discriminant analysis approach, Nonparametric Discriminant Analysis (NDA) relies on local eigenvectors for obtaining discriminant knowledge from the entire dataset. The advantage of NDA over LDA is that, NDA does not rely on assumptions that instances are drawn from a given probability distribution, therefore are more robust than parametric methods such as LDA, and suits particularly on those nonlinear datasets. Similar to LDA, NDA requires the entire dataset for training presented in advance, thus is often called batch NDA in the literature. For incremental learning of NDA, Raducanu et. al [19] proposed an incremental version of NDA, which allows us to maintain a constantly updated NDA eigenspace. However, both batch NDA and incremental NDA are merely a passive learning approach, learning passively whatever data is being given/confronted.

2.3 Motivation of Active Mode Incremental NDA Learning

To enable active learning of NDA, we incorporated incremental NDA and selective sampling technique together to form a new active learning technique, which delivers constant informative updating of NDA eigenspace, therefore minimizing concept drift and computational cost.

3 Passive NDA Learning Approaches

Classic NDA [1] assumes that the entire training dataset is provided in advance, the learning is passively done in one batch. Incremental NDA (IncNDA) is capable of learning incoming instance continuously, but IncNDA also learns inactively whatever instances are confronted. The computation of Batch NDA and IncNDA are briefed as follows.

3.1 Nonparametric Discriminant analysis (Batch NDA)

Assuming that the data samples we have belong to N classes. Let C_i represents samples belonging to one of the class i, $i = 1, 2, 3, ..., N$. Then, a NDA discrimination eigenspace according to [19] can be computed to express the class separability of data,

$$\Omega = tr(S_w^{-1}.S_b) \tag{1}$$

In above Ω, S_w is the within class covariance matrix defined as:

$$S_w = \sum_{i=1}^{C_N} \sum_{j \in C_i} (x_j - \mu C_i)(x_j - \mu C_i)^T; \tag{2}$$

S_b is the between class covariance matrix defined as,

$$S_b = \sum_{i=1}^{C_N} \sum_{j=1, j \neq i}^{C_N} \sum_{q=1}^{n_{C_i}} W_{ijq}(x_q^i - \mu NN(x_q^i, C_j)) \tag{3}$$
$$(x_q^i - \mu NN(x_q^i, C_j))^T,$$

where μC_i is the mean vector of class C_i, and w_{C_i} is the number of samples in class C_i.

In S_b, $\mu NN(x_q^i, C_j)$ is defined as a local K-NN mean,

$$\mu NN(x_q^i, C_j) = \frac{1}{k} \sum_{t=1}^{k} NN_t(x_q^i, C_j)` \tag{4}$$

where $NN_t(m_q^i, C_j)$ represents the tth nearest neighbor from vector m_{q^i} to class C_j. $W_{ijq}$is defined as a weighting function,

$$w_{ijq} = \frac{d^\alpha(x_q^i, NN_t(x_q^i, C_i))(x_q^i, NN_t(x_q^i, C_j))}{d^\alpha(x_q^i, NN_t(x_q^i, C_i)) + (x_q^i, NN_t(x_q^i, C_j)))}. \tag{5}$$

where α denotes control parameter for sample weights which can be selected between zero and infinity.

3.2 Incremental Nonparametric Discriminant Analysis (IncNDA)

Consider new instances are presented in the future. Incremental NDA [19] incorporates the discriminant knowledge presented in the new coming sample as: given new instance y is coming in, then the current NDA model Ω is required to be updated as,

$$\Omega' = f(\Omega, y) = tr(S_w'^{-1} \cdot S_b') \tag{6}$$

This means that S_w and S_b are required to be updated respectively.

According to , the updated between class $S_b^{'}$ and within class $S_w^{'}$ covariance matrix can be calculated as follows:

$$S_b^{'} = S_b - S_b^{in}(C_L) + S_b^{in}(C_{L'}) + S_b^{out}(y^{C_L}) \tag{7}$$

$$S_w^{'} = \sum_{j=1, j\neq L}^{C_N} S_w(C_j) + S_w(C_L^{'}) \tag{8}$$

where $S_b^{in}(C_L)$ represents the covariance matrix between the existing class and the class newly presented, $S_b^{out}(y^{C_L})$ gives the covariance matrix between the existing class and the updated class $C_{L'}$, and $S_w(C_L^{'})$ signifies the updated within class covariance matrix. For further computation approaches on $S_b^{in}(C_L)$, $S_b^{out}(y^{C_L})$, and $S_w(C_L^{'})$, please refer to [19].

The above IncNDA can be used to construct an agent capable of updating the current discriminant knowledge $\Omega(t)$ by $\Omega(t+1) = \mathcal{F}(\Omega(t), \boldsymbol{y})$ whenever a new instance $\boldsymbol{y}$ is confronted by the agent in the future. However, the IncNLDA is counted as a passive learning approach, because the IncNDA learns passively every instance confronted, even if the instance is confirmed redundant or noise data.

4 The proposed Active IncNDA (aIncNDA)

For active learning, we consider here an active learning way (aIncNDA) to empower the IncNDA with the ability of detecting the discriminative interestingness of data before it is delivered for IncNDA learning. That is, the above IncNDA can be renovated to conduct incremental learning in an active learning way,

$$\Omega(t+1) = \begin{cases} \mathcal{F}_c(\Omega(t), \boldsymbol{y}) & \text{if} L(t) > \xi \\ \Omega(t) & otherwise. \end{cases} \tag{9}$$

where only discriminative instances are delivered for IncNDA learning. ξ is the threshold identifying discriminative criterion of NDA. The smaller ξ leads to the bigger number of instances learned by IncNDA.

Recall that the nature of NDA learning lies at the discriminability difference between the NDA transformed space and the original space. Straightforwardly, $L(t)$ can be represented as a type of mathematical residue that reflects the discriminability difference between the NDA transformed space and the original space.

Given one new instance presented at one time, similar to [20], the discriminability difference between the NDA transformed space and the original space of the IncNDA at time t by a classification performance evaluation as,

$$L(t) = Ad(t) - Ao(t), \tag{10}$$

where $Ad(\cdot)$ is the classification accuracy on discriminant eigenspace, and $Ao(\cdot)$ is the accuracy on original space. It could be any type of classification performance evaluation by any classifier.

However, such performance-based residue calculation involves a serious problem. That is, the $L(t)$ is highly classifier dependent. For example, suppose a K-NN method is used for performance evaluation $Ad(\cdot)$ and $Ao(\cdot)$, then the selected instances for incremental learning is meaningful only for K-NN classification and the category of prototype-based methods, but may not for the classification using any other methods such as hyperplane-based support vector machines (SVM) and decision-tree based C4.5.

4.1 Discrimination Residue Ratio

The idea of discrimination residue ratio is adapted from the weighting function (i.e. Eq. (5))used in NDA, where $NN_k(x^i, C_i)$ and $NN_k(x^i, C_j)$ emphasize local within class distances and local between class distances. As we know, the principle of NDA, similar to LDA, seeks simultaneously minimizing within class distances and maximizing between class distances. The difference between NDA and LDA is, LDA is global model, whereas NDA focus on local instances distribution.

Given M new instances $Y = \{y_1, y_2, ...y_M\}$ presented as one chunk at time t, for each instance $y_i \in Y$, we can quickly estimate the within-class residue to the class mean vector μC_i:

$$\|NN_k(y^i, C_i) - \mu C_i\|, \tag{11}$$

also the between-class residue to any other the class mean vector $\mu C_j, j = 1, ..., C_N, j \neq i$:

$$\|NN_k(y^i, C_j) - \mu C_j\|. \tag{12}$$

Thus, the contribution of incoming instance y_i to the NDA fundamental maximum $tr(S_w^{-1}.S_b)$ criterion can be estimated as the following discrimination residue ratio of with-class to between-class scatter estimates

$$\nu(y_i) = \frac{\|NN_k(y^i, C_i) - \mu C_i\|}{\|\frac{1}{C_N - 1}\sum_{j=1, j\neq i}^{C_N} NN_k(y^i, C_j) - \mu C_j\|} \tag{13}$$

if $\nu(y_i) > 1$, then the contribution of y_i to NDA discrimination is positive, otherwise is negative.

Table 1: Comparison of aIncNDA versus IncNDA on incremental learning of instances over 8 UCI datasets.

	aIncNDA			IncNDA		
Datasets	ξ	No. Instances(rate[%])	Acc.[%]	No. Instances	Acc.[%]	Diff.[%]
Iris	0.75	56 (37.3)	94.5	150	92.0	+2.5
Liver-disorder	0.8	51 (22.2)	63.3	345	62.4	+0.9
Vehicle	3.0e-3	251 (29.7)	77.6	846	75.4	+2.2
Glass	0.98	50 (23.4)	60.1	214	52.5	+7.6
Wine	0.95	162 (92.7)	83.7	178	78.5	+5.2
Wisconsin	0.95	443 (95.7)	89.7	463	84.3	+5.4
Ionosphere	0.7	291 (83.1)	76.2	350	76.1	+0.1
Heart	0.65	33 (11.1)	53.2	297	52.3	+0.9

However, it is noticeable that the above discrimination residue ratio varies in practice largly depending on individual dataset. Thus, it is hard for us to determine a suitable threshold value for a given dataset. To overcome this difficulty, we compute the discrimination residue ratio for every instance of the Y, then the above $\nu(y_i)$ can be normalized as,

$$\nu_{y_i} = \frac{\nu - \bar{\nu}}{\sqrt{\frac{1}{M}\sum_{m=1}^{M}(\nu_m - \bar{\nu})^2}} \tag{14}$$

where $\bar{\nu} = \frac{1}{M}\sum_{m=1}^{M}\nu_m$ is the chunk mean discrimination residue ratio. Thus, $L(t)$ in Eq. (9)can be implemented by ν_{y_i} as a chunk data filter.

$$\Omega' = \begin{cases} \mathcal{F}_c(\Omega, \boldsymbol{y}) & \text{if}\nu(y) > \xi \\ \Omega & otherwise. \end{cases} \tag{15}$$

5 Experiments and Discussions

In this section, we have examined the efficiency and accuracy of the proposed aIncNDA method, and compared to IncNDA. Particularly, we investigate the relationship between 1) the discriminability and number of instances, 2) the redundancy and number of instances. To experiment on data with different discriminative characterization, we used datasets from two database resources. One resource is from UCI Machine Learning Repository [23], where we selected 8 datasets that have different application backgrounds and the features 100% of continuous/integer values and no missing value. The other resource is the MPEG-7 face database [24], which consists of *pose* and *light* two subsets, total 1355 face images of 271 persons, 5 different face images per person and each face image has the size of 56 $\times$ 46.

5.1 Experimental Setup

To implement the proposed aIncNDA for incremental learning, we select randomly, for each dataset, 10% for initial batch NDA training, and divide the remaining data into 10 random chunks for incremental learning test. We collect every instance learned by aIncNDA, and evaluate the performance of aIncNDA and IncNDA on discrimination contribution at every learning stage. For performance evaluation, we compared the eigenspace from the proposed aIncNDA with the eigenspace from IncNDA by a leave-one-out kNN (k=1) classification over all data presented by current learning stage. Note that we use the term *learning stage* instead of the usual time scale since the events of data arriving in the above incremental learning may not happen in a regular time interval. Here, the number of learning stages is equivalent to the number of instances that have been learned by incremental models.

In the experiment, parameter ξ is relevant to the number of curiosity instances and the discriminability of the resulting NDA. For each experiments, we fixed ξ by the rule that the instances are significantly selected with, at most, minor sacrifices in discriminability.

5.2 UCI Datasets

Table 1 gives an comparison of aIncNDA versus IncNDA on the incremental learning of 8 UCI datasets. In the table, ξ is fixed for each dataset by the rule described above, the number of instances and the percentage to the number of all instances is denoted as ‘No. Instances(rate)’, and the classification accuracies is denoted as ‘Acc.’. The discriminability difference (denoted as ‘Diff.’) is calculated as the proposed aIncNDA minus IncLDA in terms of the K-NN LOO classification performance at the final learning stage.

As seen in the table, the proposed aIncLDA method, ignores 4.7%-88.9% instances of the whole dataset, constructs discriminant eigenspaces on the remaining 11.1%-95.3% selected instances. However, the discriminability of the obtained eigenspace from composed instance subset, compared to the eigenspace from all instances (using IncLDA), has no decrease, reversely, most of case has a slight increase. This suggests that the proposed active IncNDA learning is valid, and the selected instances by aIncNDA have the expected discriminative representativeness.

5.3 Performance under different discriminative redundancy

To test the performance of the proposed method under different level of discriminative redundancy, we carried out face recognition (FR) and face membership authentication (FMA) experiments [25–27] using the same face database described above. FMA is to distinguish the membership class (cls. 1) from the non-membership class (cls. 2) in a total group through a binary class classification. FMA involves more discriminative redundancy than face recognition problem, because the size of membership in FMA is often smaller than that of nonmembership, which indicates that not every instance are discriminatively important for FMA.

Over the 271 persons 1355 faces data, we conducted FR and FMA, respectively. For the FMA experiment, we set the membership size as 71 (cls. 1/cls. 2 is 71/200) without loss of generality. Thus, we compared the proposed aIncNDA with the IncNDA on incremental learning of 271 classes (i.e. FR) and 2 classes (i.e. FMA) data, respectively.

Fig. 1(a) shows the comparison of NDA discriminability between the proposed aIncNDA and the IncLDA for both FR and FMA experiments, and Fig. 1(b) reports corresponding the number of instances learned by aIncNDA.

As seen in Fig. 1(a), the proposed aIncNDA learns NDA for FR on 1331 of total 1355 instances, only 24 instances are found redundant. Whereas for FMA, aIncNDA learns 1093 of 1355 which is only about 20.0% of total 1355 instances are reduced. However, the performance of the proposed aIncNDA for both FR and FMA as given in Fig. 1(a) outperforms in most cases, the performance of the IncNDA on all 1355 instances. This indicates that the proposed aIncNDA is able to suit itself automatically to data with discriminative redundancy, and select a suitable number of instance to build an correct NDA model. This also can be reflect from Fig. 1(b), where aIncNDA is shown actively selecting different number of instance for incremental learning.

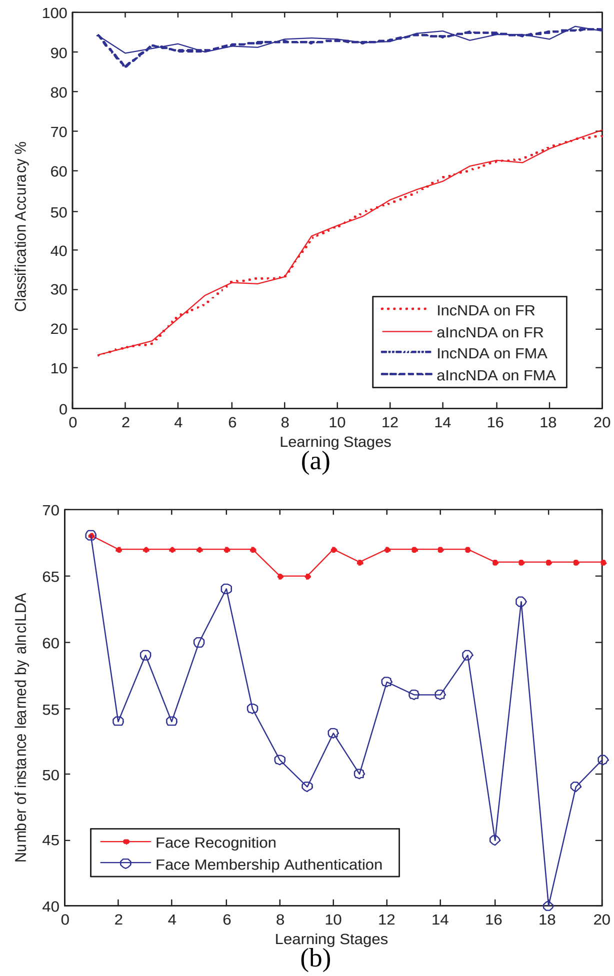

Figure 1: Comparison of aIncNDA and IncNDA on FR and FMA, (a) the performance of aIncLDA versus IncNDA on incremental learning; (b) the number of learned instances by aIncNDA at every learning stage.

6 Conclusion and Future Works

Method based on passive learning prove to be inadequate in real world application. To overcome this limitation, we have developed active mode incremental NDA which performs adaptive discriminant selection of instances for incremental NDA learning. Performance evaluation carried out on benchmark UCI datasets show that Active Mode Incremental NDA performs on par and in many cases better then incremental NDA with less number of instances. Given the nature of network data which is large, streaming, and constantly changing, we believe that our method can find practical application in the field of internet security.

Over the datasets from different resources, the proposed aIncNDA learning method is evaluated on: (1) aIncNDA versus IncNDA, and (2) performance under different level redundancy, where face recognition and face membership authentication are studied, respectively. The experimental results demonstrate that the proposed aIncNDA learning helps more efficient NDA learning with fewer instances, but with no performance deduction. One limitation of the proposed method concerns, as the original IncNDA retains raw data at every step of incremental learning, the data processing in aIncNDA is not one-pass.

As future work, the presented methods application in intrusion detection system will be exploited along with added enhancements to the selective sampling criterion. Also, the use of incremental classifier will be researched to serve as an extension to our present model which will eliminate the need for retraining further enhancing the processing speed while been computationally efficient.

Acknowledgments

This work was supported in part by KEDRI project titled ”High Speed Algorithms for Outlier Detection and Classification over Huge-size Network Data Streams (PI: S. Pang), through the FY2009 funding program from National Institute of Communication and Technology (NICT), Japan.

References

[1] Fukunaga K, Introduction to statistical pattern recognition. *2nd edn. Academic Press,* Boston, 1990.

[2] Charles, X. L. and D. Jun, Active learning with direct query construction. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining,* Las Vegas, Nevada, USA, ACM, 2008.

[3] Kai, Y., Z. Shenghuo, et al. Non-greedy active learning for text categorization using convex transductive experimental design. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval,* Singapore, Singapore, ACM, 2008

[4] Ming-yu, C., C. Michael, et al., Putting active learning into multimedia applications: dynamic definition and refinement of concept classifiers. *Proceedings of the 13th annual ACM international conference on Multimedia,* Hilton, Singapore, ACM, 2005.

[5] Sanjoy, D. and H. Daniel, Hierarchical sampling for active learning. *Proceedings of the 25th international conference on Machine learning,* Helsinki, Finland, ACM, 2008.

[6] Hieu, T. N. and S. Arnold, Active learning using pre-clustering. *Proceedings of the twenty-first international conference on Machine learning,* Banff, Alberta, Canada, ACM.2004

[7] Steven, C. H. H., J. Rong, et al., Large-scale text categorization by batch mode active learning. *Proceedings of the 15th international conference on World Wide Web,* Edinburgh, Scotland, ACM, 2006.

[8] Prem, M. and J. M. Raymond, Diverse ensembles for active learning. *Proceedings of the twenty-first international conference on Machine learning,* Banff, Alberta, Canada, ACM, 2004

[9] Danziger, S.A., Zeng, J., et al., Choosing where to look next in a mutation sequence space: Active Learning of informative p53 cancer rescue mutants. *Bioinformatics,* 23(13), 104-114, 2007.

[10] Liu, Y. , Active Learning with Support Vector Machine Applied to Gene Expression Data for Cancer Classification. *J. Chem. Inf. Comput. Sci.,* 44(6): 1936-1941, 2004.

[11] Warmuth, M. K., J. Liao, et al., Active Learning with Support Vector Machines in the Drug Discovery Process. *J. Chem. Inf. Comput. Sci.,* 43(2): 667 -673, 2003.

[12] Greiner, R., Grove, and D. Roth, Learning cost-sensitive active classifiers. *Artificial Intelligence,* 139(2): 137-174, 2002.

[13] Kapoor, A. and R. Greiner, Learning and classifying under hard budgets, *Lecture Notes in Computer Science,* 3720, 170-181, 2005

[14] Tong, S. and D. Koller, Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research,* 2, 2001.

[15] Yu, K. Bi, J. and V. Tresp, Active learning via transductive experimental design. *Proc. of International Conference on Machine Learning (ICML),* 2006.

[16] Zhang, W. V., X. He, B. Rey, and R. Jones, Query rewritting using active learning for sponsored search. *ACM SIGIR Conference,* 2007.

[17] Park, J. Convergence and application of online active sampling using orthogonal pillar vectors. *IEEE T. Pattern Anal. Mach. Learn.* 28, 1197-1207, 2004.

[18] Bor-Chen, K. and D. A. Landgrebe, Nonparametric weighted feature extraction for classification. *IEEE Trans. on Geoscience and Remote Sensing,* 42(5): 1096-1105, 2004

[19] Raducanu, B. and J. Vitria, Online nonparametric discriminant analysis for incremental subspace learning and recognition. *Pattern Analysis Application,* 11: 259 - 268, 2008

[20] Naoki Shimo, Shaoning Pang, Nikola Kasabov and Takeshi Yamakawa, Curiosity-driven Multi-agent Competitive and Cooperative LDA Learning. *International Journal of Innovative Computing, Information and Control,* 4(7): 1537-1552, 2008

[21] P. Belhumeur, et al. Eigenfaces vs. Fisher-faces: recognition using class specific linear projection. *IEEE Trans. on PAMI,* 19(7): 711-720, 1997.

[22] Shaoning Pang, Seiichi Ozawa and Nik Kasabov, Incremental Linear Discriminant Analysis for Classification of Data Streams. *IEEE Trans. on System, Man, and Cybernetics-Part B,* 35(5), pp.905-914, 2005.

[23] http://www.ics.uci.edu/mlearn/MLRepository.html

[24] M. Kim, D. Kim, S. Bang, and S. Lee, Face Recognition Descriptor using the Embedded HMM with the 2nd-Order Block-Specific Eigenvectors. Jeju, Korea, ISO/IEC/ JTC1/SC21/WG11/M7997, 2002.

[25] Hyun-Chul Kim, Shaoning Pang, Hong-Mo Je, Daijin Kim,and Sung Yang Bang, Constructing support vector machine ensemble. *Pattern Recognition*, 36(12): 2757-2767, 2003.

[26] Shoning Pang, D. Kim and S. Y. Bang, Membership authentication in the dynamic group by face classification using SVM ensemble. *Pattern Recognition Letters*, Vol. 24, pp. 215-225, 2003.

[27] Shaoning Pang, D. Kim, and S. Y. Bang, Face Membership Authentication Using SVM Classification Tree Generated by Membership-based LLE Data Partition. *IEEE Trans. on Neural Network,*16(2): 436-446, 2005.