Correlation-based Feature Ordering for Classification based on Neural Incremental Attribute Learning

Ting Wang, Sheng-Uei Guan, and Fei Liu

***Abstract*—Incremental Attribute Learning (IAL) is a novel supervised machine learning approach, which sequentially trains features one by one. Thus feature ordering is very important to IAL. Previous studies on feature ordering only concentrated on the contribution of each feature to different outputs. However, besides contribution, correlations among input features and output categories are also very important to the final classification result, which has not yet been researched in feature ordering but has confirmed in multivariate statistics. This study aims to find out the relations between feature ordering and feature correlations. This paper presents a new method for feature ordering calculation which is based on correlations between input features and outputs. Experimental results confirm that correlation-based feature ordering can produce better classification results than contribution-based approaches, feature orderings with theoriginal sequence sorted in the database, and conventional methods where all features are trained in one batch.**

***Index Terms*—Machine learning, incremental attribute learning, pattern classification, feature ordering, correlation.**

I. Introduction

In pattern classification, the number of features (attributes) indicates the complexity of a problem. The more features in a problem, the more complex it is. To solve complex classification problems, some dimensional reduction strategies like feature selection have been employed [1, 2]. However, these methods are invalid when the feature number is huge and almost all features are crucial simultaneously. Thus feature reduction is not the ultimate technique to cope with high dimensional problems.

A strategy for solving high-dimensional problems is “divide-and-conquer”, where a complex problem is firstly separated into smaller modules by features and integrated after each module is tackled independently. Incremental Attribute Learning is an example of that. It is applicable for solving classification problems in machine learning [3-6]. Previous studies show that IAL based on neural networks obtains better results than conventional methods [3], [7]. For example, in Guan’s studies, compared with traditional methods [5],[6], classification error rates of Diabetes, Thyroid and Glass, three machine learning datasets from University of California, Irvine (UCI), derived by neural IAL were reduced by 8.2%, 14.6% and 12.6%, respectively [8].

However, because IAL incrementally imports features into systems, it is necessary to know which feature should be introduced in an earlier step. Thus feature ordering becomes a new preprocess apart from conventional preprocess like feature reduction. Previous studies of neural IAL presented contribution-based feature ordering method, where feature ordering was derived after each feature is solely employed to classify all outputs by neural networks. The result of each denotes every feature’s ability for discrimination. However, such a wrapper is more time-consuming than filter [9]. Thus it is necessary to study on feature ordering based on filter methods.

In this paper, a new contribution-based feature ordering metric is presented. It is derived by correlations between input and output. Such a metric will be checked for applicability and accuracy by a neural IAL algorithm calledIncremental neural network Training with an Increasing input Dimension (ITID). In Section 2, ITID will be reviewed and the contribution-based feature ordering method will be presented in Section 3; three benchmarks will be validated by neural IAL and analyzed in Section 4; conclusions will be drawn in section 5 with outlines of future works.

II. IAL and Feature Ordering

A. IAL

Based on some predictive methods like neural networks, IAL has exhibited its feasibility in solving multi-dimensional classification problems in a number of previous studies. ITID [12], a representative of neural IAL based on Incremental Learning in terms of Input Attributes (ILIA) [7], is shown applicable for classification. It is different from conventional approaches which train all features in one batch. It divides all input dimensions into several sub-dimensions, each of which corresponds to an input feature. After this step, instead of learning input features altogether as an input vector in training, ITID learns inputs through their corresponding sub-networks one after another and the structure of neural networks gradually grows with an increasing input dimension as shown in Figure1. During training, information obtained by a new sub-network is merged together with the information obtained by the old network. Moreover, based on ILIA, ITID has a pruning technique which is adopted to find the appropriate network architecture. With less internal interference among input features, ITID achieves higher generalization accuracy than conventional methods [12].

Manuscript received September 25, 2012; revised November 11, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 61070085.

Ting Wang is with the Department of Computer Science, University of Liverpool, Liverpool,L69 3BX, UK, and the Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China (e-mail: ting.wang@ liverpool.ac.uk).

Sheng-Uei Guanis with the Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China. (e-mail: steven.guan@xjtlu.edu.cn).

Fei Liu is with the Department of Computer Science and Computer Engineering, La Trobe University, Victoria, 3086, Australia (e-mail: f.liu@latrobe.edu.au).

10.7763/IJMLC.2012.V2.242

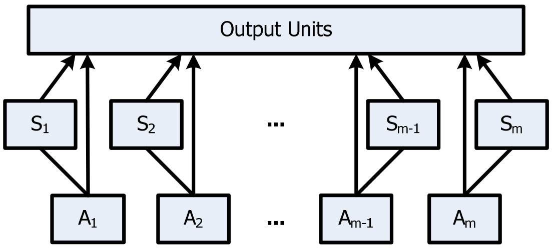

Fig. 1. The basic network structure of ITID.

B. *Ordered Feature Training in IAL*

According to the mechanism of ITID, it is manifest that features should be introduced to the system in some orders. Although feature ordering is seldom used in conventional methods where features are trained in one batch, it is believed that ordered features are necessary for improving final classification performance in pattern recognition based on IAL approaches [6,13]. In previous studies, feature ordering calculation has been developed by two different kinds of ways: ranking-based and contribution-based. Such an isolation of feature ordering approaches is similar to that in feature selection, where ranking-based and contribution-based approaches are called filter and wrapper, respectively. Different from feature selection where the purpose of which is to search a feature subset for obtaining the optimal results, feature ordering aims to arrange proper sequence of features for calculate the optimal results. That is, feature reduction like feature selection usually scraps useless features or reduces the weights of useless features, while feature ordering does nothing but give a sequence to features by discrimination ability. Therefore, apart from different objectives, feature selection and feature ordering are similar to each other. Hence in feature ordering studies, ranking-based approach also can be named filter-like method, while contribution-based approach can be called as wrapper-like method.

In previous studies, it has been validated that ranking-based feature ordering computing is better than contribution-based approaches usually at least in two different aspects: time and error rate [11]. Feature ordering is a unique and indispensable data preparation job of IAL. Once features are ranked or their contributions are calculated, datasets should be transformed according to their feature ordering. After that, patterns are randomly divided into three different datasets: training, validation and testing [13]. All vectors in these three datasets should be sorted according to the feature ordering and employed for classification by machine learning later.

III. Correlation among Inputs and Outputs

Usually, there are three types of correlations in the datasets of classification problems: correlation between input features, correlation between input features and output classes, and correlation between output classes. Because no patterns belong to two or more classes simultaneously, classification has an either-or situation. Consequently, it is only necessary to check the first two types of correlation for one dataset, and relations between input and output, and relations among each features should be studied in IAL.

Actually, the study on correlation in pattern recognition is not something new. Previous research on correlation in pattern recognition aims to develop feature selection approaches that can be used to alleviate the effect of the curse of dimensionality, enhance generalization capability, speed up learning process and improve model interpretability. Furthermore, most of the previous research in this area focused on feature selection. In order to achieve above objectives, feature selection approaches are divided into two categories: feature subset selection and feature ranking. The former searches a set of possible features for the optimal subset, while the latter ranks features by a metric and discards all features whose score is under the threshold according to some criteria. When correlation analysis is employed in the feature selection process, both feature ranking and subset searching, can be used for classification. Previous research confirmed that good feature subsets often contain high-correlated features with the classification, but uncorrelated to each other [10].

Therefore, in the process of feature selection, for feature ranking, we should select features which not only have high correlation with outputs, but also have low correlation with each other; for feature subset selection, we should search the optimal feature subset that has high correlations with classification outputs and low correlations among themselves. Thus, for feature subset selection and feature ranking, no matter which type is selected for feature selection, these two correlation analysis approaches for classification are the same in essence.

Moreover, in Incremental Attribute Learning, data preparation is quite different from conventional machine learning approaches, where features are trained by batch. Feature ordering, a new data preprocessing stage, is deemed as a requirement before training. Due to the fact that feature ranks have different values which can be employed as a measurement to arrange features in some order, feature ranking is more useful in data preprocessing phase than subset searching. Accordingly, features should be trained one by one according to the order derived by the fusion of correlations between input features and that with input and output together.

Correlation-based feature ordering can be calculated by

$$\mathrm{CorrelationIndex_i} = \frac{|\mathrm{Correlation}(\mathrm{Input_i}, \mathrm{Output})|}{\left(\sum_{j=1}^{n}\left|\mathrm{Correlation}(\mathrm{Input_i}, \mathrm{Input_j})\right|\right)/n} \tag{1}$$

where Correlation Index of *i*-th feature is presented, which can be calculated by the ratio between correlation of *i*-th input and all output, and the average correlation between *i*-th feature and all other input features. Furthermore, correlation can be calculated by Pearson Correlation Coefficient or Covariance Matrix. Similar to correlation-based feature selection, it is obvious that the greater the correlation index in (1), the earlier the feature should be trained.

IV. Benchmarks

In this study, proposed feature ordering based on correlation were tested with datasets from UCI Machine Learning Repository. There are three datasets used in this study: Diabetes Thyroid, and Cancer. The brief information about the datasets employed in the experiments has been shown in Table I.

TABLE I: BRIEF INFORMATION OF EXPERIMENTAL DATASETS

	Datasets	Input Number	Output Number	Pattern Number
1	Diabetes	8	2	768
2	Cancer	9	2	699
3	Thyroid	21	3	7200

In the experiments, all datasets were firstly normalized, and the covariance matrices of these normalized datasets were calculated in the next step. According to the covariance matrices, correlations can be obtained, and feature ordering also can be computed based on (1). Then Neural IAL approaches were employed for pattern classification. Patterns for training, validation and testing were divided by 50%, 25% and 25%, respectively.

In this study, all the experimental results were compared with results derived by other three approaches. Firstly, Contribution-based wrapper feature orderings, which have been presented in [12]; secondly, Original Orderings are based on feature's original order shown in datasets; and lastly, conventional method has no feature ordering, which trains all features in one batch and is not an IAL approach but a traditional one.

TABLE II: CORRELATION INDEX AND FEATURE ORDERING (THYROID)

	$\sum_{j=1}^{n}\left\vert\mathrm{Correlation}(\mathrm{Input}_i, \mathrm{Input}_j)\right\vert$	$\left\vert\mathrm{Correlation}(\mathrm{Input}_i, \mathrm{Output})\right\vert$	Feature No.	Correlation Index
1	1.3783	0.2938	17	0.010151
2	0.6549	0.0827	10	0.006013
3	2.4268	0.2449	21	0.004805
4	1.0809	0.086	3	0.003789
5	3.1643	0.2285	19	0.003439
6	0.3698	0.0251	13	0.003232
7	2.6242	0.1573	18	0.002854
8	0.415	0.0218	8	0.002501
9	0.6995	0.0285	16	0.00194
10	1.1841	0.0478	2	0.001922
11	0.6341	0.017	5	0.001277
12	1.4755	0.0332	7	0.001071
13	0.4228	0.0091	4	0.001025
14	0.3217	0.0046	15	0.000681
15	0.7091	0.0098	14	0.000658
16	0.4399	0.006	9	0.000649
17	0.5625	0.0075	6	0.000635
18	1.3091	0.0111	1	0.000404
19	2.3002	0.0151	20	0.000313
20	1.1851	0.0065	11	0.000261
21	0.2499	0.0012	12	0.000229

TABLE III: RESULT COMPARISON (THYROID)

	Feature Ordering	Classification Error
Correlation-Based	17-10-21-3-19-13-18-8-16-2-5-7-4-15-14-9-6-1-20-11-12	1.575%
Contribution-Based	17-21-19-18-1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-20	1.722222%
Original Ordering	1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21	1.591666%
Conventional Method		1.8638875%

TABLE IV: CORRELATION INDEX AND FEATURE ORDERING (DIABETES)

	$\sum_{j=1}^{n}\left\vert\mathrm{Correlation}(\mathrm{Input}_i, \mathrm{Input}_j)\right\vert$	$\left\vert\mathrm{Correlation}(\mathrm{Input}_i, \mathrm{Output})\right\vert$	Feature No.	Correlation Index
1	1.7701	0.448	2	0.031637
2	1.1088	0.2126	7	0.023967
3	1.8133	0.3363	6	0.023183
4	1.6392	0.2373	8	0.018096
5	1.3377	0.1931	1	0.018044
6	1.8752	0.1416	4	0.009439
7	1.7708	0.13	5	0.009177
8	1.2647	0.0691	3	0.00683

TABLE V : RESULT COMPARISON (DIABETES)

	Feature Ordering	Classification Error
Correlation-Based	2-7-6-8-1-4-5-3	21.32812%
Contribution-Based	2-6-1-7-3-8-5-4	22.96876%
Original Ordering	1-2-3-4-5-6-7-8	22.86458%
Conventional Method		23.93229%

TABLE VI: CORRELATION INDEX AND FEATURE ORDERING (CANCER)

	$\sum_{j=1}^{n}\lvert\text{Correlation}(\text{Input}_i, \text{Input}_j)\rvert$	$\lvert\text{Correlation}(\text{Input}_i, \text{Output})\rvert$	Feature No.	Correlation Index
1	4.8797	0.7043	1	0.01603696
2	5.4296	0.7771	6	0.015902542
3	5.2723	0.6953	8	0.014653103
4	5.6665	0.7417	7	0.014543565
5	6.3587	0.8135	3	0.014214995
6	6.2828	0.8026	2	0.014193955
7	5.4099	0.678	4	0.013925088
8	5.3797	0.6634	5	0.013701714
9	3.591	0.4275	9	0.013227513

TABLE VII: RESULT COMPARISON (CANCER)

	Feature Ordering	Classification Error
Correlation-Based	1-6-8-7-3-2-4-5-9	1.839082%
Contribution-Based	2-3-5-8-6-7-4-1-9	2.4999985%
Original Ordering	1-2-3-4-5-6-7-8-9	2.902299%
Conventional Method		1.867818%

According to Table II-VII, Classification Errors derived by Correlation-based approaches are quite lower than those calculated by contribution-based wrappers and approaches in original ordering and conventional methods which train all features in one batch. Therefore, the feature ordering criterion based on the metric of correlation index is useful to obtain feature orderings for IAL. Moreover, before training, validation and testing, all data should be reformed according to the descending ordering of correlation index of features. In the experiments of this study, Diabetes and Cancer are univariate classification problems, while Thyroid has three different output classes. Therefore, the correlation-based approach presented in this study cannot only solve univariate classification problems, but also cope with multi-category problems.

V. CONCLUSIONS

Correlations among inputs and outputs are crucial to rank features' significance for feature ordering in pattern classification based on IAL. The ratio between the average of input correlations and input-output correlations has been confirmed as a novel metric for feature ordering calculation. Experimental results on benchmarks denote that such a metric is able to be used not only in univariate classification problems, but also multi-category problems, and it can exhibit better performance in both of these two kinds of problems than contribution-based wrapper approaches and conventional batch feature training methods.

In the future, studies of correlation between inputs and outputs for feature ordering ranking will be continue, and the ranking criteria will be merged with some other metircs like mutual information, linear discriminant and so on. Whether such a proposal can bring higher accuracy in classification error rate and reduce the computational complexity in both preprocess and machine learning process will be an important issue in future research on IAL.

ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China under Grant 61070085.

REFERENCES

[1] H. Liu, “Evolving feature selection,” *IEEE Intelligent Systems*, vol. 20, no. 6,pp. 64-76, 2005.

[2] S. H. Weiss and N. Indurkhya, *Predictive data mining: apractical guide*, Morgan Kaufmann Publishers, CASan Francisco, 1998.

[3] S. Chao and F. Wong, “An incremental decision tree learning methodology regarding attributes in medical data mining,” in *Proc. of the 8th Int’l Conf. on Machine Learning and Cybernetics*, Baoding, 2009, pp.1694-1699.

[4] R. K. Agrawal and R. Bala,“Incremental Bayesian classification for multivariate normal distribution data,” *Pattern Recognition Letters*, vol.29, no. 13, pp.1873-1876, 2008.

[5] S. U. Guan and J. Liu, “Feature selection for modular networks based on incremental training,” *Journal of Intelligent Systems*, vol. 14, no. 4, pp. 353-383, 2005.

[6] F. Zhu, S. U.Guan, “Ordered incremental training for GA-based classifiers,” *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2135-2151, 2005.

[7] S. U. Guan and S. Li, “Incremental learning with respect to new incoming input attributes,” *Neural Processing Letters*, vol. 14, no. 3, pp. 241-260, 2001.

[8] S. U. Guan and S. Li,“Parallel growing and training of neural networks using output parallelism,” *IEEETrans. on Neural Networks*, vol. 13, no. 3, pp. 542 -550, 2002.

[9] Pablo Bermejo, Luis de la Ossa, José A. Gámez, and José M. Puerta, “Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking,” *Knowledge-Based Systems*, vol. 25, no.1, pp.35-44, 2012.

[10] Mark A. Hall. “Correlation-based Feature Selection for Machine Learning,” PhD Dissertation. University of Waikato, New Zealand, 1999.

[11] T. Wang, S. U. Guan, F. Liu, "Feature Discriminability for Pattern Classification Based on Neural Incremental Attribute Learning," in: *Foundations of Intelligent Systems*, The 2011 International Conference on Intelligent Systems and Knowledge Engineering, 2011, pp.275-280.
[12] S. U .Guan and J. Liu, "Incremental neural network training with an increasing input dimension," *Journal of Intelligent Systems*, vol. 13, no. 1, pp. 43-69, 2004.
[13] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge University Press, UK Cambridge, 1996.

Ting Wang, born in Wuxi, China, May 28th, 1981, is now a PhD Candidate in Computer Science of the University of Liverpool, the United Kingdom. He obtained his BSc degree in computer science at the China University of Mining and Technology, in 2003 and his MSc degree in computer science at the Guilin University of Technology, China, in 2008.

From July 2003 to July 2004, he was a system analyst at the Micro star International Inc. Before he started his PhD program in March 2009, he was a research and development engineer at the Jiangnan Institute of Computing Technology.

Mr. Wang is certificated as a Computer Science Engineer by China government, and a member of Wuxi Computer Federation, Jiangnan University.

Sheng-Uei Guan received his M.Sc. & Ph.D. from the University of North Carolina at Chapel Hill. He is currently a professor and head of the computer science and software engineering department at Xian Jiaotong-Liverpool University. Before joining XJTLU, he was a professor and chair in intelligent systems at Brunel University, UK.

Prof. Guan has worked in a prestigious R&D organization for several years, serving as a design engineer, project leader, and manager. After leaving the industry, he joined Yuan-Ze University in Taiwan for three and half years. He served as deputy director for the Computing Center and the chairman for the Department of Information & Communication Technology. Later he joined the Electrical & Computer Engineering Department at National University of Singapore as an associate professor.

Fei Liu completed her PhD from The Department of Computer Science & Computer Engineering, La Trobe University in 1998. Before joining the department as an academic staff member in 2002, she worked as a lecturer in the School of Computer & Information Science, University of South Australia, and the School of Computer Science & Information Technology, Royal Melbourne Institute of Technology. She also worked as a software engineer for Ericsson Australia. Her research interests include Logic Programming, Semantic Web and Security in Electronic Commerce.