

Lei XU, Yanda LI

Machine learning and intelligence science: Sino-foreign interchange workshop IScIDE2010 (A)

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2011

All the papers in this special issue are invited and compiled as Part A of the publication of the 2010 Sino-foreign-interchange workshop on Intelligence Science and Intelligent Data Engineering (IScIDE2010). This workshop was held on June 3-5, 2010 in Harbin, and was jointly held with 2010 academician forum on Information Science & Intelligence Science (ISIS). IScIDE2010 is scheduled as the first of the annual workshops that aim at a collective venue where we can introduce core achievements of Chinese colleagues to international communities and also bring worldwide frontier researches to the Chinese ISIS community. This community is featured by a huge population of Chinese researchers and students and a rapid development in past decades, and also by a long and memorable history.

Starting from the late 1970s and the early 1980s, the beginning period of China's Reform, Prof. T. Chang of Tsinghua University and his collaborator, Prof. M. T. Cheng of Peking University, both of whom were the senior academicians, have made great efforts on the foundation and developments of the Chinese ISIS community. Another senior academician, Prof. K. S. Fu of Purdue University (USA) has dedicated his enthusiastic and concrete supports. For instance, via a nonofficial Zhongguancun Golden Triangle ISIS forum that they founded, a large number of international pioneers from the fields of signal processing, pattern recognition and machine intelligence have visited China for academic exchanges, leaving significant impacts on the academic researches in the related Chinese communities. Their early contributions were instrumental to the establishments of several key national laboratories and the growth of a large number of high quality scientists and professors who have become the backbone of the current Chinese ISIS community. More than 30 years later, it is probably time for another start for promoting similar international exchanges. Still, Chinese scholars will learn from the world not just frontier researches but also dedicated spirits and serious attitudes for quality research. On the other hand, there also needs a collective window for international communities to observe and share the achievements of Chinese colleagues.

For this purpose, three special issues have been scheduled in *Frontiers of Electrical and Electronic Engineering in China* (FEE). One was called *Emerging Themes on Information Theory and Bayesian Approach* (FEE, Vol. 5, No. 3, 2010), which consisted of nine invited papers from international well known research pioneers and Chinese senior scientists. The other two special issues come from IScIDE2010, which was featured by 24 invited presentations by core researchers from abroad and Chinese major institutes. One of the two issues is scheduled to appear in the middle of 2011. The first authors of the papers in that special issue are all Ph.D students who are mostly studying in mainland China universities. The other one is the present special issue, consisting of 14 invited papers.

This special issue is bisected. The first half is motivated by a double purpose. One is to provide a helpful angle for viewing the core achievements by Chinese colleagues, with papers written by the professors mostly from major universities in mainland China and for whom one or two sentences of personal introduction may be added during introducing their papers. The other purpose is a focus of machine learning algorithms on the following

Received January 19, 2011

Lei XU

Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
E-mail: lxu@cse.cuhk.edu.hk

Yanda LI

Department of Automation, Tsinghua University, Beijing 100084, China
E-mail: daulyd@tsinghua.edu.cn

fundamental topics:

- Describing data samples with reduced dimensionality;
- Integrating multi-sources of partial information for a better performance.

As such an up-to-date status of advances on the topics is systematically outlined. Complementarily, the second half aims at a much wider scope of intelligence science (vision system, neural system, human behavior, and generic issues) for broadening the readers' knowledge of this journal, with papers written by professors outside the mainland of China.

We start with the first half from the above first topic. As further developments of the classic topic of feature extraction, up-to-date efforts cover various studies made under the name of manifold learning or topology preserved dimensionality reduction. In this special issue, He and Lin reviewed canonical manifold learning techniques and discussed two fundamental problems in tangent space learning. With the help of the statistics on Grassmann manifold, they proposed a novel method called persistent tangent space learning, which estimates the tangent space at each local neighborhood while ensures that the tangent spaces vary smoothly on the manifold. Professor He, a young talent from Zhejiang University, is internationally known by his extensively cited publications with coauthors on graph Laplacian based locality preserving projections (LPP) and applications [1,2]. The paper by Zhang and Wang in this special issue also involves LPP manifold learning with a new method for constructing a graph with a multilevel scheme that makes their method scalable for large data sets. Precisely, their work is a systematical effort aiming at graph-based semi-supervised learning, which is left to our introduction on the second topic slightly later. Additionally, Sect. 3.2 of Xu's paper in this special issue also suggests a Bayesian Ying-Yang (BYE) harmony learning method for LPP in an alternative formulation and with a nature of automatic model selection on manifold dimension.

The paper by Prof. Yin from the University of Manchester provides an overview on a much wider scope not just from the aspect of geometry but also the aspects of statistics and adaptive neural networks. Various state-of-the-art algorithms are elaborated along three streams of developments. Yin's paper differs from a number of other overview papers in its third stream of adaptive neural network approaches along two closely related lines. One is featured by further extensions of self-organizing map (SOM) [3,4], including Yin's unique contributions on visualization induced SOM (ViSOM) [5,6]. The other line starts from adaptive PCA learning [7], nonlinear PCA [8,9], principal curve and surface [10,11], to kernel PCA and kernel based view of the dimensionality reduction by manifold learning [12,13]. A complementary overview on the early developments is referred to Ref. [14]. Continuing the kernel based studies, Yang's paper in this special issue introduces an idea of generating a kernel from an arbitrary function by embedding the training samples into the function, which provides a unified perspective to classify the existing kernel Fisher discriminant algorithms. Prof. Yang, another young talent from Nanjing University of Science and Technology, is internationally known by his extensively cited publications with coauthors on two-dimensional principal component analysis (2DPCA) [15]. In this special issue, Yang's paper also outlines the advances on 2DPCA and tensor-based feature extraction, as well as possible extensions to their generating kernel versions.

The efforts on the second topic can be traced back to the late 1980s through the middle 1990s. In the character recognition literature, various classifiers are available from different methodologies and different features, which motivate studies on combining multiple classifiers. A systematic effort on this early stage was made in Ref. [16] and an early overview is referred to Ref. [17]. In the neural networks literature, there were studies under the name of ensemble learning that combines an ensemble of diverse and complementary pre-trained networks [18-22]. Further progresses make training and combining jointly by the mixture-of-experts [23-28], with the help of the expectation-maximization (EM) algorithm [29,30]. Readers are referred to an overview on these studies and extensions [31]. Complementarily, an alternative modular neural network was proposed based on task decomposition and min-max module combination by Lu and Ito [32]. In this special issue, Lu and the coauthors further develop this well cited previous work for learning imbalanced data sets, featured by decomposing the training sets of the majority classes into subsets of similar size and pairing them into balanced two-class classification sub-problems that are tackled by support vector machines with the help of min-max combination.

During the last decade and more, further efforts have been paid on ensemble learning, especially on theoretical analyses toward achieving strong generalization. Meanwhile, studies under the name of semi-supervised learning grow rapidly and become a popular theme in the current machine learning literature. Semi-supervised learning has some relationship with the above mentioned studies. For example, similar studies were made by Refs. [33] and [34] both in 1997 on Gaussian mixture with labeled and unlabeled data, though it was named semi-unsupervised learning in Ref. [34], (see Eq. (7.14)). However, efforts have seldom made on connections with different semi-supervised learning categories and with the studies on classifier combination, ensemble learning, and mixture-of-experts.

In this special issue, Zhou's paper makes an overview on the advances of semi-supervised learning and ensemble

learning in four categories, i.e., generative methods, semi-supervised support vector machines, graph-based methods, and disagreement based methods. With empirical and theoretical evidences obtained from disagreement-based semi-supervised learning, Zhou's paper shows that semi-supervised learning and ensemble learning are beneficial to each other, and stronger learning machines can be generated by leveraging unlabeled data and classifier combination. Prof. Zhou is not only among a few leading mainland researchers in machine learning and pattern recognition, but is also internationally known by his highly cited papers on ensemble learning and disagreement-based semi-supervised learning [35,36]. Based on his expertise, Zhou further suggests that disagreement-based methods provide a good vessel to accommodate advantages not just from all semi-supervised learning categories but also from classifier combination, ensemble learning, and mixture-of-experts, as illustrated in its Fig. 3.

Besides the disagreement-based semi-supervised learning, another category has witnessed a surge of interests, which is graph-based or manifold-based semi-supervised learning. Based on the assumption that data points (nearly) reside on a low-dimensional manifold with nearby points likely in the same label (local consistency) and points of same cluster or a sub-manifold likely in the same label (global consistency), the problem turns out to be finding such manifold by preserving the local structure of the data set. Graphs are obtained as discretization of manifolds. Given a graph, the label of an unlabeled vertex is predicted from its neighboring labels, via a constraint relaxation. In this special issue, Zhang and Wang introduce the works made by the group of Prof. Zhang, another leading mainland researcher who has published many papers at international top conferences and on journals. One unique contribution is called the generalized point charge scheme, with labeled data points as point charges while unlabeled data points placed in the electrostatic fields of these charges. The labels of unlabeled points are regarded as the electric potentials in the field, which is efficiently solved by the discrete Green's function. Another of their work is generalizing the classic LPP for multi-label problems with one additional regularization term via the Laplacian of a category graph.

It is interesting to note the two related terms of blind deconvolution and semi-blind deconvolution, used in the signal processing literature one decade earlier than the appearance of the term semi-supervised learning. Blind deconvolution [37,38] refers to the tasks of estimating a system only from its output observations, while semi-blind deconvolution [39] refers to the cases that we know partially either or both the system and its input. The term deconvolution refers to seeking the inverse of a convolution system. Regarding labels as inputs and data points as outputs while class structures as a system, we can observe that semi-blind deconvolution and semi-supervised learning share a similar concept but differ in a specific system and specific types of input and output. Probably, semi-blind learning is a better name for efforts that put attention on the cases of knowing partially either or both the system and its input, instead of just knowing partially the inputs by semi-supervised learning.

In Sect. 4.4 of Ref. [40], the BYY system is shown to provide a general framework to accommodate several cases of semi-supervised learning. In a complementary role, Xu's paper in this special issue provides another perspective of the BYY harmony learning, featured with learning multiple layers of inner representations in co-dimensional matrix pairs that coordinately optimize the performances of parameter learning and model selection.

In a complementary role, the second half of this special issue broadens the readers' scope with a number of mathematical modeling problems in intelligence science, involving vision system, neural system, cognitive behavior, and generic issues. We start with the paper "Processing of real-world imagery with facade-based approaches" by the group of Dwen Hu, Professor of Education Ministry, Chang Jiang Scholar and one of Chinese National Science Foundation Distinguished Young Scholars (two prestigious academic titles in mainland). Their work uses Grossberg's FACADE framework in modeling the human vision system (HVS). To address the inefficient and non-robust issues suffered by the original FACADE framework, they use the framework as an infrastructure and incorporate it with the anisotropic spatial mechanism and diffusive orientational-competition mechanism, which makes the HVS featured model efficient and robust. Instead of modeling the HVS in a high level, Prof. Fiori from University of Politecnica delle Marche (Italy) considers computing the average character of a set of eyes by linear optical models. He shows that a Lie group is a mathematical description of linear optical models, based on which a Lie-group averaging iteration algorithm is examined in comparison with Harris' non-iterative averaging rule.

We are then led to neural system by the paper of Prof. Ikeda and coauthors, from Nara Institute of Science and Technology. Neurons in cortical areas are characterized by the inter-spike intervals of spike sequences of neurons, which is modeled as a gamma process with a time-variant spike rate, a fixed shape parameter and a fixed absolute refractory period. Their paper formulates the problem of estimating the fixed parameters as semiparametric estimation and applies an information geometrical method to deriving the optimal estimators. Prof. Siegelmann from UMass Amherst (USA) further brings us to human behavior. Her paper models human addiction by two subsystems in parallel that tip toward each other based on a continuous rationality factor, resulting in a dynamical system that qualitatively emulates the behaviors of seeking, cessation, and relapse—enabling. Also, the paper by

Chen and Yu emulates human behavior in double auction markets by autonomous agents with the help of genetic programming, from which an autonomous-agent-inspired economic theory with regard to the optimal procrastination is derived.

Finally, we have reached two generic issues that are not limited but applicable to the intelligence science. In general, mathematical modeling is made with representations by real-valued numbers, and the paper by Prof. Hirose from the University of Tokyo shows that complex-valued representations have unique advantages in modeling those problems featured with waves such as electromagnetic wave and lightwave or generally treated in frequency-domain through Fourier transform. Next, the paper by Prof. Liou and coauthors further shows how to generate the so-called hierarchical modular system (HMS) [41] by using the principle of extreme physical information. The HMS is devised by Caianiello on how a self-organized system is constructed naturally with levels, with the help of the hierarchical thermodynamic approach. It has been used to model the discrete denomination values of monetary system, the distribution rule of coins, and the distribution of populations on a territory. Liou's paper also discusses that the equal temperament music scale shares many common properties with the HMS too.

Last but not the least, we are grateful to all the invited authors for the valuable supports to this special issue. We fully understand certain reluctance, from you and especially those from mainland or Asian regions (probably including the guest editor himself), on publishing a long time written paper to a so-called 'non-SCI journal'. We believe that such a reluctance will disappear as time elapses, probably because we will be in an academic environment that has no longer a superficial preference for SCI journals or probably because time will prove that publishing on this journal is worth it. Toward this end we need your continued support to both this journal and the future IScIDE.

Acknowledgements The guest editor was partially supported by the National Key Basic Research & Development Program of China (973 Program) (No. 2009CB825404). The guest editor is a member of the Steering Committee for Information Sciences under CUHK focused investments scheme, to which this work is partially dedicated.

References

1. He X, Niyogi P. Locality preserving projections. In: *Advances in Neural Information Processing Systems 16*. Cambridge, MA: MIT Press, 2003, 152–160
2. He X, Cai D, Niyogi P. Tensor subspace analysis. In: *Proceedings of Advances in Neural Information Processing Systems*. 2005, 18: 499–506
3. Kohonen T. Self-organized formation of topologically correct feature map. *Biological Cybernetics*, 1982, 43(1): 59–69
4. Kohonen T. *Self-Organizing Maps*. 2nd ed. Berlin: Springer, 1997
5. Yin H. Data visualisation and manifold mapping using the ViSOM. *Neural Networks*, 2002, 15(8-9): 1005–1016
6. Yin H. On multidimensional scaling and the embedding of self-organising maps. *Neural Networks*, 2008, 21(2-3): 160–169
7. Oja E. Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1989, 1(1): 61–68
8. Oja E, Ogawa H, Wangviwattana J. Learning in nonlinear constrained Hebbian networks. In: *Proceedings of ICANN'91*. 1991, 385–390
9. Xu L. Least MSE reconstruction for self-organization. In: *Proceedings of IJCNN91-Singapore*. 1991, 3: 2363–2373
10. Hastie T, Stuetzle W. Principal curves. *Journal of the American Statistical Association*, 1989, 84(406): 502–516
11. LeBlanc M, Tibshirani R J. Adaptive principal surfaces. *Journal of the American Statistical Association*, 1994, 89(425): 53–64
12. Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998, 10(5): 1299–1319
13. Ham J, Lee D D, Mika S, Scholkopf B. A kernel view of the dimensionality reduction of manifolds. In: *Proceedings of the 21st International Conference on Machine Learning*. 2004, 369–376
14. Xu L. Independent Subspaces. In: Ramón J, Dopico R, Dorado J, Pazos A, eds. *Encyclopedia of Artificial Intelligence*. Hershey(PA): IGI Global, 2008, 903–912
15. Yang J, Zhang D, Frangi A F, Yang J Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(1): 131–137
16. Xu L, Krzyzak A, Suen C Y. Several methods for combining multiple classifiers and their applications in handwritten character recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1992, 22: 418–435
17. Kittler J, Hatef M, Duin R P W, Matas J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(3): 226–239
18. Hansen L K, Salamon P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(10): 993–1001

19. Xu L, Krzyzak A, Sun C Y. Associative switch for combining multiple classifiers. In: Proceedings of IJCNN91, Seattle, WA. 1991, (I): 43–48
20. Wolpert D H. Stacked generalization. *Neural Networks*, 1992, 5(2): 241–259
21. Baxt W G. Improving the accuracy of an artificial neural network using multiple differently trained networks. *Neural Computation*, 1992, 4(5): 772–780
22. Breiman L. Stacked Regression. Department of Statistics, Berkeley. 1992, TR-367
23. Jacobs R A, Jordan M I, Nowlan S J, Hinton G E. Adaptive mixtures of local experts. *Neural Computation*, 1991, 3(1): 79–87
24. Jordan M I, Jacobs R A. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 1994, 6(2): 181–214
25. Jordan M I, Xu L. Convergence Results for The EM Approach to Mixtures of Experts Architectures. *Neural Networks*, 1995, 8(9): 1409–1431
26. Xu L, Jordan M I, Hinton G E. An Alternative Model for Mixtures of Experts. In: Cowan, Tesauro, Alspector, eds. *Advances in Neural Information Processing Systems 7*. MIT Press, 1995, 633–640
27. Xu L, Jordan M I, Hinton G E. A modified gating network for the mixtures of experts architecture, Proc. In: Proceedings of WCNN'94, San Diego, CA. 1994, (2): 405–410
28. Xu L, Jordan M I. EM learning on a generalized finite mixture model for combining multiple classifiers. In: Proceedings of WCNN'93. 1993, (IV): 227–230
29. Dempster A P, Laird N M, Rubin D B. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological*, 1977, 39(1): 1–38
30. Redner R A, Walker H F. Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Review*, 1984, 26(2): 195–239
31. Xu L, Amari S. Combining classifiers and learning mixture-of-experts. In: Dopico J R R, Dorado J, Pazos A, eds. *Encyclopedia of Artificial Intelligence*. 2009, 318–326
32. Lu B L, Ito M. Task decomposition and module combination based on class relations: a modular neural network for pattern classification. *IEEE Transactions on Neural Networks*, 1999, 10(5): 1244–1256
33. Miller D J, Uyar H S. A mixture of experts classifier with learning based on both labelled and unlabelled data. In: Mozer M, Jordan M I, Petsche T, eds. *Advances in Neural Information Processing Systems 9*. Cambridge: MIT Press, 1997, 571–577
34. Xu L. Bayesian Ying Yang system and theory as a unified statistical learning approach: (I) Unsupervised and semi-supervised learning. In: Amari S, Kassabov N, eds. *Brain-like Computing and Intelligent Information Systems*. Springer-Verlag, 1997, 241–274
35. Zhou Z H, Wu J, Tang W. Ensembling neural networks: many could be better than all. *Artificial Intelligence*, 2002, 137(1–2): 239–263
36. Zhou Z H, Li M. Tri-training: exploiting unlabeled data using three classifiers. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(11): 1529–1541
37. Stockham T G, Cannon T M, Ingebretsen R B. Blind deconvolution through digital signal processing. *Proceedings of the IEEE*, 1975, 63(4): 678–692
38. Kundur D, Hatzinakos D. Blind image deconvolution revisited. *IEEE Signal Processing Magazine*, 1996, 13(6): 61–63
39. Xu L, Yan P F, Chang T. Semi-blind deconvolution of finite length sequence: (I) linear problem & (II) nonlinear problem. *SCIENTIA SINICA, Series A*, 1987, (12): 1318–1344
40. Xu L. Bayesian Ying-Yang system, best harmony learning, and five action circling. *Frontiers of Electrical and Electronic Engineering in China*, 2010, 5(3): 281–328
41. Caianiello E R. Some remarks on organization and structure. *Biological Cybernetics*, 1977, 26(3): 151–158

Lei XU

Guest Editor of the special issue

Yanda LI

Editor-in-Chief of *Frontiers of Electrical and Electronic Engineering in China*