Composite Kernel Machines on Kernel Locally Consistent Concept Factorization Space for Data Mining

Shian-Chang Huang National Changhua University of Education, Department of Business Administration Email: shhuang@cc.ncue.edu.tw

> Lung-Fu Chang National Taipei College of Business, Department of Finance Email: lfchang@webmail.ntcb.edu.tw

Tung-Kuang Wu National Changhua University of Education, Department of Information Management Email: tkwu@im.ncue.edu.tw

Abstract—This paper proposes a novel approach to overcome the main problems in high-dimensional data mining. We construct a composite kernel machine (CKM) on a special space (the kernel locally consistent concept factorization (KLCCF) space) to solve three problems in high-dimensional data mining: the curse of dimensionality, data complexity and nonlinearity. CKM exploits multiple data sources with strong capability to identify the relevant ones and their apposite kernel representation. KLCCF finds a compact representation of data, which uncovers the hidden information and simultaneously respects the intrinsic geometric structure of data manifold. Our new system robustly overcomes the weakness of CKM, it outperforms many traditional classification systems.

Index Terms—data mining, multiple kernel learning, kernel locally consistent concept factorization, manifold learning, support vector machine

I. INTRODUCTION

Data sets of high dimensionality pose great challenges on efficient processing to most existing data mining algorithms (Witten and Frank [1]). Mining highdimensional heterogeneous data is a crucial component in many information applications. Financial data mining becomes a popular topic owing to the late-2000s financial crisis. Many techniques have been developed for bankruptcy predictions. Popular methods include regression, discriminant analysis, logistic models, factor analysis, decision trees, neural networks, fuzzy logic, genetic algorithms, etc. However, their performance is usually not satisfactory.

A reliable high-dimensional data mining system for financial distress predictions is urgently demanded by all banking and investment institutes to control their financial risk. Previously, they invested heavily on establishing an automatic decision support system for evaluating the credit quality of their borrowers. The objective of this paper is to develop such a system to prevent banking institutes from investing a distressed company.

Reviewing recent literature, many advanced approaches from data mining or artificial intelligence were developed to solve the problems as mentioned above. These methods (Witten and Frank [1]) include inductive learning, case-based reasoning, neural networks, rough set theory (Ahn et al. [2]), and support vector machines (SVM) (Wu et al. [3]; Hua et al. [4]). SVM, a special form of kernel classifiers, has become increasingly popular. SVM considers the structural risk in system modeling, and regularizes the model for good generalization and sparse representation. SVMs are successful in many applications. They outperform typical methods in classifications. However, the success of SVM depends on the good choice of model parameters and the kernel function, (namely, the data representation). In kernel methods, the data representation is implicitly chosen through the so-called kernel. This kernel actually plays two important roles: it defines the similarity between two examples, while defining an appropriate regularization term for the learning problem.

The choice of kernel and features are typically handcrafted and fixed in advance. However, hand-tuning kernel parameters can be difficult as can selecting and combining appropriate sets of features. Recent applications have also shown that using multiple kernels instead of a single one can enhance the interpretability of the decision function and improve performances (Lanckriet *et al.* [5]). Multiple Kernel Learning (MKL) seeks to address this issue by learning the kernel from training data. In particular, it focuses on how the kernel can be learnt as a linear combination of given base kernels.

Manuscript received January 24, 2014; revised April 21, 2014.

The flat combination of kernels in MKL does not include any mechanism to cluster the kernels related to each source. In order to favor the selection/removal of kernels between or within predefined groups, one has to define a structure among kernels, which will guide the selection process. Composite kernel machines (CKM, Szafranski *et al.* [6]) addresses the problem by defining a structure among kernels, which is particularly well suited to the problem of learning from multiple sources. Then, each source can be represented by a group of kernels, and the algorithm aims at identifying the relevant sources and their apposite kernel representation.

In financial data mining, high dimensional data from public financial statements and stock markets can be used for bankruptcy predictions. However, the high dimensional data make kernel classifiers infeasible due to the curse of dimensionality (Bellman [7]). Regarding dimensionality reduction, linear algorithms such as principal component analysis (PCA, Fukunaga [8]) and discriminant analysis (LDA, Fukunaga [8]) are the two most widely used methods due to their relative simplicity and effectiveness. However, classical techniques for manifold learning are designed to operate when the submanifold is embedded linearly, or almost linearly, in the observation space. Such algorithms often fail when nonlinear data structure cannot simply be regarded as a perturbation from a linear approximation. The task of nonlinear dimensionality reduction (NLDR) is to recover meaningful low-dimensional structures hidden in high dimensional data.

Recently, matrix factorization based techniques, such as Non-negative Matrix Factorization (NMF, Lee and Seung [9]) and Concept Factorization (CF, Xu and Gong [10]), have yielded impressive results in dimensionality reduction. The non-negative constraints of NMF only allow additive combinations among different basis vectors which can learn a parts-based representation (Lee and Seung [9]).

Financial data are probably sampled from a submanifold of the ambient Euclidean space. In fact, the financial data cannot fill up the high dimensional Euclidean space uniformly. Therefore, the intrinsic manifold structure needs to be considered while performing the matrix factorization. The major limitation of NMF is that it is unclear how to effectively perform NMF in the transformed data space, e.g. reproducing kernel Hilbert space (RKHS). To get rid of the limitations of NMF while inheriting all its strengths, Xu and Gong [10] proposed Concept Factorization (CF). Li and Ding [11] also proposed several interesting variations of NMF. The major advantage of CF over NMF is that it can be performed on any data representations, either in the original space or RKHS.

However, NMF and CF only concern the global Euclidean geometry, whereas the local manifold geometry is not fully considered. Cai *et al.* [12] proposed a new version of CF called locally consistent concept factorization (LCCF) to extract the basis vectors which is consistent with the manifold geometry. Central to the approach of LCCF is a graph model which captures the

local geometry of the data manifold. By using the graph Laplacian to smooth the data mapping, LCCF can extract features with respect to the intrinsic manifold structure. This study also employs a kernel version of LCCF (KLCCF) to mining underlying key features in high dimensional financial data, and constructs CMKMs on the submanifold created by KLCCF. Moreover, we incorporate the label information to the graph model used in KLCCF to improve our system performance.

The remainder of this paper is organized as follows: Section 2 describes the CKM classifiers and KLCCF. Subsequently, Section 3 describes the study data and discusses the empirical findings. Conclusions are given in Section 4.

II. THE PROPOSED METHODOLOGY

To reduce the computational loading of kernel machines and simultaneously enhance their performance. This study constructs CKMs on a non-linear graph-based KLCCF.

A. Composite Multiple Kernel Machines

In multiple kernel learning (MKL), we are provided with M candidate kernels, K_1, \ldots, K_M , and wish to estimate the parameters of the SVM classifier together with the weights of a convex combination of kernels K_1, \ldots, K_M , that defines the effective kernel K_q .

$$\Box = \{K_{\sigma} = \sum_{m=1}^{M} \sigma_m K_m, \sigma_m \ge 0, \sum_{m=1}^{M} \sigma_m = 1\}$$
(1)

Each kernel K_m is associated to a RKHS H_m whose elements will be denoted f_m , and $\sigma_1, \ldots, \sigma_M$ is the weighting vector to be learned under the convex combination constraints.

In order to favor the selection/removal of kernels between or within predefined groups. Szafranski *et al.* [6] improved traditional MKL by considering a tree structure among kernels. Szafranski *et al.* [6] indexes the tree depth by h, with h=0 for the root, and h=2 for the leaves. The leaf nodes represent the kernels at hand for the classification task; the nodes at depth 1 stand for the group-kernels formed by combining the kernels within each group; the root represents the global effective kernel merging the group-kernels.

In the learning process, one would like to suppress the kernels and/or the groups that are irrelevant for the classification task. In the tree representation, this removal process consists in pruning the tree. When a branch is pruned at the leaf level, a single kernel is removed from the combination. When a subtree is pruned, a group-kernel is removed from the combination, and the corresponding group of kernels has no influence on the classifier.

The *M* kernels situated at the leaves are indexed by $\{1, ..., m, ..., M\}$, and the group-kernels (at depth 1) are indexed by $\{1, ..., l, ..., L\}$. The set G_l of cardinality d_l indexes the leaf-kernels belonging to group-kernel *l*, that is, the children of node *l*. The groups form a partition of the leaf-kernels, that is, $\cup_l G_l = \{1, ..., m, ..., M\}$ and

 $\sum_{l} d_{l} = M$. The CKM of Szafranski *et al.* [6] is formulated as follows:

$$\min_{f_1,\dots,f_M,b,\xi,\sigma_1,\sigma_2} \frac{1}{2} \sum_m \|f_m\|_{\mathbf{H}_m}^2 + C \sum_i \xi_i$$
(2)

$$st.y_{i}\left(\sum_{l}\sigma_{1,l}\sum_{m\in G_{l}}\sigma_{2,m}f_{m}(x_{i})+b\right)\geq 1-\xi_{i}\ i=1,...,n$$
$$\xi_{i}\geq 0,\ i=1,...,n$$
$$\sum_{l}d_{l}\sigma_{1,l}^{2/p}\leq 1,\ \sigma_{1,l}\geq 0,\ l=1,...,L$$
$$\sum_{m}\sigma_{2,m}^{2/q}\leq 1,\ \sigma_{2,m}\geq 0,\ m=1,...,M.$$

where $\sigma_1 = (\sigma_{1,1} \dots \sigma_{1,L})$ and $\sigma_2 = (\sigma_{2,1} \dots \sigma_{2,M})$ are weighting vectors; hyper-parameters *p* and *q* control the sparsity within or between groups. For details of the model, we refer to Szafranski *et al.* [6].

B. Kernel Locally Consistent Concept Factoriaztion

Recently, Non-negative Matrix Factorization (NMF) has yielded impressive results in dimensionality reduction. In general given a nonnegative data matrix X, NMF tries to find reduced rank nonnegative matrices U and V so that $X \approx UV^{T}$. The column vectors of U can be thought of as basis vectors and V contains the coordinates.

NMF can only be performed in the original feature space of the data points. In the case that the data are highly non-linear distributed, it is desirable that we can kernelize NMF and apply the powerful idea of the kernel method. To achieve this goal, Xu and Gong [10] proposed an extension of NMF, which is called Concept Factorization (CF). In CF, each basis u_k is required to be a non-negative linear combination of the sample vectors x_{j} .

$$u_k = \sum_{j=1}^N x_j h_{jk} \tag{3}$$

where $h_{jk} \ge 0$. Let $H = [h_{jk}]$, CF essentially tries to find the following approximation, $X \approx XHV^T$, through minimization of $O = || X - XHV^T ||^2$.

CF tries to find a basis that is optimized for the linear approximation of the data. Let z_j^T denote the j-th row of V, $z_j^T = [v_{j1}, ..., v_{jk}]$, can be regarded as the new representation of each data point in the new basis. Cai *et al.* [12, 13] indicated that knowledge of the geometric structure of the data can be exploited for better discovery of this basis. A natural assumption here could be that if two data points \mathbf{x}_i , \mathbf{x}_j are close in the intrinsic geometry of the data distribution, then z_i and z_j , the representations of this two points in the new basis, are also close to each other. Actually z_i and z_j are equivalent to V_i and V_j in our formulation. This assumption is usually referred to as local consistency assumption, which plays an essential

role in developing various kinds of algorithms including dimensionality reduction algorithms (Belkin and Niyogi [14]) and semi-supervised learning algorithms (Belkin *et al.* [15]).

Local geometric structure can be effectively modeled through a nearest neighbor graph on a scatter of data points. Consider a graph with N vertices where each vertex corresponds to a data point. Define the edge weight matrix W as follows:

$$\mathbf{W}_{ij} = \begin{cases} \frac{\mathbf{x}_{i}^{T} \mathbf{x}_{j}}{\|\mathbf{x}_{i}\|\| \|\mathbf{x}_{j}\|} & \text{if } x_{i} \text{ and } x_{j} \text{ belong to the same class;} \\ 0 & \text{otherwise,} \end{cases}$$

Here, prior class-label information is also used to define the *W*. That is, the within-class geometric information is emphasized, and the similarity between two samples is set to zero if they belong to different classes.

The optimal z needs to minimize the following objective:

$$\sum_{i,j} (z_i - z_j)^2 W_{i,j}.$$
 (4)

This objective function incurs a heavy penalty if neighboring vertices i and j are mapped far apart. With some simple algebraic formulations, we have

$$\sum_{i,j} (z_i - z_j)^2 W_{i,j} = 2\mathbf{z}^T L \mathbf{z} = 2\mathbf{V}^T L \mathbf{V},$$
(5)

where L=D-W is the graph Laplacian (Chung [16]) and D is a diagonal matrix whose entries are column (or row, since W is symmetric) sums of W, $D_{ii} = \sum_{j} W_{ji}$. Finally, the minimization problem reduces to find the minimum of

$$O = ||X - XHV^{T}||^{2} + \lambda Tr(V^{T}LV)$$
(6)

Define $K = X^T X$. We can rewrite the objective function:

$$O = Tr((X - XHV^{T})^{T}(X - XHV^{T})) + \lambda Tr(V^{T}LV)$$

= Tr((I - HV^{T})^{T}K(I - HV^{T}) + \lambda Tr(V^{T}LV)
= Tr(K) - 2Tr(VH^{T}K) + Tr(VH^{T}KHV^{T}) + \lambda Tr(\mathbf{V}^{T}L\mathbf{V})

Next, we nonlinearly extend the formulation to highdimensional RKHS:

$$\mathbf{u}_{k} = \sum_{j=1}^{N} \phi(\mathbf{x}_{j}) h_{jk}, \qquad (7)$$

and $K = \varphi(X)^T \varphi(X)$. Similarly, KLCCF essentially tries to find the following approximation, $X \approx \phi(X) H V^T$, through the minimization of

$$O = ||X - \varphi(X)HV^{T}||^{2} + \lambda Tr(V^{T}LV)$$
(8)

III. EXPERIMENTAL RESULTS AND ANALYSIS

This study used bankrupt companies listed in the Taiwan Stock Exchange (TSE) for analysis. Their public financial information is used for the model input. These bankrupt companies were matched with normal companies for comparison. The sample data covers the period from 2000 to 2007.

For the balance of positive and negative samples, one company in financial crisis should be matched with one or two normal companies in the same year, in the same industry, running similar business items. Namely, they should produce the same products with the failed company and have similar scale of operation. Additionally, the normal company whose total asset or the scale of operation income should be close to the failed company. In our samples, 50 failed firms and 100 nonfailed firms were selected. The study traced the data up to 5 years, which started from the day a respective company falls into financial distress backward up to a period of 5 years. The financial reports of the non-failed companies will be matched (pooled together) with the failed company in the same year. For example, company A failed in 2005 and company B failed in 2007. We will pool them and their matched companies A', B' in the same file labeled C000 representing their financial status in the year of bankruptcy. Companies A and A' (or company B and B') will be traced backward up to five years. These data were put in separate files labeled C000, C111, C222, C333, and C444 respectively for classification.

The variables of this research are selected from the TEJ (Taiwan Economic Journal) financial database, which contains the following five financial indexes: profitability index, per share rates index, growth rates index, debt-paying ability index, management ability index. Altogether, there are 54 financial ratios covered by the five indexes. If some values of a ratio lost on some firms, this ratio was deleted. As a result, overall 48 financial ratios were obtained for analysis.

This study tested five conventional classifiers and a kernel classifier (SVM) for bankruptcy predictions,

including decision tree (J48), nearest neighbors with three neighbors (KNN), logistic regressions, Bayesian networks (BayesianNet), radial basis neural network (RBFNetwork), and SVM. For kernel classifiers, this study selected the polynomial kernel of two degrees for input owing to its good performance compared with other types of kernels. The data set was randomly divided into ten parts, and ten-folds cross validation was applied to evaluate the model performance.

Table I shows that SVM outperforms other classifiers. Namely, kernel classifiers outperform traditional classifiers due to their flexibility in dealing nonlinear and high-dimensional data. Consequently, this study implemented an advanced kernel classifier, the CKM, for subsequent classifications.

In high-dimensional classification problems, some input variables or features may be irrelevant. Avoiding irrelevant features is important, because they generally deteriorate the performance of a classifier. There are two approaches to address the problem: feature subset selection and dimensionality reduction. First, we try the two means of feature selection: Chi-Squared Statistics (x^2 , Witten and Frank [1]) and Information Gain (IG, Witten and Frank [1]). After determination of the optimal feature subset, the selected input variables were fed into six classification algorithms (J48, KNN, BayesianNet, Logistic, RBFNetwork, SVM) for distress prediction.

Table II and III show that x^2 and IG could slightly improve the performance of all classifiers. However, they deteriorate the performance of RBFNetwork significantly. RBFNetwork is a strong classifiers capable to re-scale feature weighting (importance) internally. Outside feature selection schemes using different criterions to select feature subsets do not always match its need.

| | Sample C000 | Sample C111 | Sample C222 | Sample C333 | Sample C444 |
|-------------|-------------|-------------|-------------|-------------|-------------|
| J48 | 90.7007 | 87.5912 | 84.6715 | 81.3433 | 71.8750 |
| KNN | 88.3212 | 90.5109 | 81.7518 | 76.1194 | 73.4375 |
| BayesianNet | 90.2409 | 90.5109 | 85.4015 | 84.3284 | 80.0313 |
| Logistic | 88.4058 | 84.0580 | 75.3623 | 71.1111 | 70.5426 |
| RBFNetwork | 89.1304 | 90.5797 | 85.5072 | 74.8148 | 78.2946 |
| SVM | 91.2409 | 91.3043 | 86.2319 | 83.5821 | 80.6875 |

TABLE I. PERFORMANCE COMPARISON ON BASIC PREDICTION MODELS (ACCURACY %)

TABLE II. PERFORMANCE ENHANCEMENTS BY CHI-SQUARED (x^2) STATISTICS

| | Sample C000 | Sample C111 | Sample C222 | Sample C333 | Sample C444 |
|------------------------------------|-------------|-------------|-------------|-------------|-------------|
| x ² + J48 | 92.2007 | 90.5109 | 88.3212 | 84.3284 | 78.1250 |
| x^2 + KNN | 91.2409 | 87.5912 | 84.6415 | 74.6269 | 78.1250 |
| x ² +BayesianNet | 92.2409 | 92.1606 | 89.0511 | 86.5672 | 79.6875 |
| x^2 +Logistic | 92.7536 | 93.4783 | 87.6812 | 84.4444 | 82.1705 |
| x ² + RBFNetwork | 76.8116 | 73.1884 | 73.1884 | 75.5556 | 82.1705 |
| x^2 +SVM | 92.5797 | 93.4783 | 88.4058 | 83.7037 | 82.1705 |

TABLE III. PERFORMANCE ENHANCEMENTS BY INFORMATION-GAIN (IG)

| | Sample C000 | Sample C111 | Sample C222 | Sample C333 | Sample C444 |
|----------------|-------------|-------------|-------------|-------------|-------------|
| IG+J48 | 93.2007 | 92.1406 | 88.3212 | 84.3284 | 78.1250 |
| IG+KNN | 91.2409 | 91.2409 | 84.6715 | 74.6269 | 78.1250 |
| IG+BayesianNet | 92.8909 | 92.1606 | 89.0511 | 86.5672 | 79.6875 |
| IG+Logistic | 92.7536 | 92.7536 | 87.6812 | 84.4444 | 82.1705 |
| IG+RBFNetwork | 76.8116 | 73.1884 | 73.1884 | 75.5556 | 79.8450 |
| IG+SVM | 92.0290 | 92.7536 | 88.4058 | 82.9630 | 82.1705 |

| | Sample C000 | Sample C111 | Sample C222 | Sample C333 | Sample C444 |
|------------|-------------|-------------|-------------|-------------|-------------|
| ICA+SVM | 74.8350 | 65.2200 | 72.1980 | 65.2750 | 65.3210 |
| PCA+SVM | 82.6400 | 84.0700 | 76.0990 | 78.6810 | 78.4620 |
| LDA+SVM | 86.2091 | 81.0989 | 81.0990 | 80.0550 | 77.5640 |
| KPCA+SVM | 88.8680 | 87.0330 | 80.9340 | 81.5930 | 79.2310 |
| Isomap+SVM | 83.2970 | 86.2641 | 82.6919 | 83.0770 | 76.0260 |
| KLCCF+CKM | 100.0000 | 92.8600 | 85.7100 | 84.6200 | 79.4900 |

TABLE IV. PERFORMANCE IMPROVEMENTS BY DIMENSIONALITY REDUCTIONS

| TABLE V. AVER | AGE PERFORMANCE OF | EACH CLASSIFIER |
|---------------|--------------------|-----------------|
| | | |

| J48 | KNN | BNet | Log | RBFNet | SVM |
|-------------|------------|-------------|-------------|---------------|------------|
| 83.2363 | 82.0282 | 86.1026 | 77.8960 | 83.6653 | 86.6093 |
| $x^{2}+J48$ | x^2 +KNN | x^2 +BNet | x^2 +Log | x^2 +RBFNet | x^2 +SVM |
| 86.6972 | 83.2451 | 87.9415 | 88.1056 | 76.1829 | 88.0676 |
| IG +J48 | IG +KNN | IG +BNet | IG +Log | IG +RBFNet | IG +SVM |
| 87.2232 | 83.9810 | 88.0715 | 87.9607 | 75.7178 | 87.6644 |
| ICA +SVM | PCA +SVM | LDA +SVM | Isomap +SVM | KPCA +SVM | KLCCF+CKM |
| 68.5698 | 79.9904 | 81.2052 | 82.2712 | 83.5318 | 88.5360 |

Note: "Bnet" is the abbreviation of "Bayesian Network"; "Log" is the abbreviation of "Logistic".

Next, we compare our method (CKM on KLCCF) with other dimensionality reduction methods. We compared our system with other famous subspace or manifold learning algorithms such as the PCA, ICA (Independent Component Analysis, Hyvärinen *et al.* [17]), LDA, kernel PCA (KPCA), and Isomap (Tenenbaum *et al.* [18]). The dimension of subspace was set to five for all algorithms.

Table IV shows that CKM on KLCCF significantly outperform other classifiers. It achieved the highest accuracy. This results fully demonstrate that financial data are not sampled from a linear manifold. Hence, linear algorithms such as PCA ICA, and LDA fail to extract discriminative information from data manifold. Considering graph-based nonlinear manifold learning algorithms (KLCCF) are more effective. On the other hand, our data come from diverse sources, only multiple kernel machines such as CKM are powerful enough to handle the complex structure in data. We also find in Table 4 that nonlinear dimensionality reduction methods (such as kernel PCA) is not always better than linear algorithms (PCA ICA, and LDA), since KPCA works in an unsupervised manner which lacks information to guide the mapping learning that could maintain most discriminant power. However, KLCCF is a supervised algorithm which nonlinearly forms a manifold not only preserving local geometry of the data samples, but also contains label information to discriminate the data.

Table V displays average performance for each classifier. Table V clearly demonstrate the superiority of the our new classifier. The new classifier substantially outperforms other dimensionality reduction based classifiers. Moreover, it also outperforms typical SVM classifiers.

IV. CONCLUSIONS

From geometric perspective, data is usually sampled from a low dimensional manifold embedded in high dimensional ambient space. KLCCF finds a compact representation which uncovers the hidden information and simultaneously respects the intrinsic geometric structure. This study constructed a CKM on KLCCF to create a novel system for bankruptcy predictions. In KLCCF, an affinity graph is constructed to encode the geometrical information and KLCCF seeks a matrix factorization which respects the graph structure. CKM is an excellent framework to exploit multiple data sources with strong capability to identify the relevant sources and their apposite kernel representation. Combining the above two techniques make our hybrid classifier powerful and robust. The empirical results confirmed the superiority of the proposed system. CKM on KLCCF is a robust and reliable framework for high dimensional data mining.

Future research may consider semi-supervised subspace or manifold learning algorithms to enhance system performance, or to include more variables such as non-financial and macroeconomic variables to improve accuracy.

REFERENCES

- [1] H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.
- [2] B. S. Ahn, S. S. Cho, and C. Y. Kim, "The integrated methodology of rough set theory and artificial neural network for business failure prediction," *Expert Systems with Applications*, vol. 18, no. 2, pp. 65-74, 2000.
- [3] C. H. Wu, W. C. Fang, and Y. J. Goo, "Variable selection method affects SVM-based models in bankruptcy prediction," in *Proc.* 9th Joint International Conference on Information Sciences, 2006.
- [4] Z. Hua, Y. Wang, X. Xu, B. Zhang, and L. Liang, "Predicting corporate financial distress based on integration of support vector machine and logistic regression," *Expert Systems with Applications*, vol. 33, no. 2, pp. 434-440, 2007.
- [5] G. R. G. Lanckriet, N. Cristianini, L. E. Ghaoui, P. Bartlett, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research*, vol. 5, pp. 27-72, 2004.
- [6] M. Szafranski, Y. Grandvalet, and A. Rakotomamonjy, "Composite kernel learning," *Machine Learning Journal*, vol. 79, pp. 73-103, 2010.
- [7] R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, 1961.
- [8] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, 1990.
- [9] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, pp. 788-791, 1999.
- [10] W. Xu and Y. Gong, "Document clustering by concept factorization," in Proc. 2004 Int. Conf. on Research and Development in Information Retrieval (SIGIR' 04), Jul. 2004, pp. 202-209.

- [11] T. Li and C. Ding, "The relationships among various nonnegative matrix factorization methods for clustering," in *Proc. IEEE International Conference on Data Mining*, 2006, pp. 362-371.
- [12] D. Cai, X. He, and J. Han, "Locally consistent concept factorization for document clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 6, pp. 902-913, 2011.
- [13] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. International Joint Conference* on Artificial Intelligence (IJCAI), Jan. 2007, pp. 708-713.
- [14] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, pp. 585-591, 2001.
- [15] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from examples," *Journal of Machine Learning Research*, vol. 7, pp. 2399-2434, 2006.
- [16] F. R. K. Chung, Spectral Graph Theory, American Mathematics Soc., 1997.
- [17] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley Interscience, 2001.
- [18] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319-2323, 2000.

Shian-Chang Huang received his MS degree in electric engineering from National Tsing Hwa University, and his PhD degree in financial engineering from National Taiwan University. He is currently a professor at the Department of Business Administration, National Changhua University of Education, Taiwan. His research interests include machine learning, soft computing, signal processing, data mining, computational intelligence, and financial engineering.

Lung-fu Chang received his PhD degree in financial engineering from National Taiwan University. He is currently an assistant professor at the Department of Finance, National Taipei College of Business, Taiwan. His research interests include financial engineering, risk management, asset pricing.

Tung-Kuang Wu received his PhD degree in computer engineering from the Department of Computer Science & Engineering at Pennsylvania State University in 1995. He is currently a professor at the Department of Information Management of National Changhua University of Education, Changhua, Taiwan. His current research interests include parallel processing, wireless networks, special education technologies, and e-Learning.