Using Non-negative Matrix Factorization for Bankruptcy Analysis

NING CHEN[1]
BERNARDETE RIBEIRO[2]
AN CHEN[1,3]

[1]GECAD, Instituto Superior de Engenharia do Porto, Instituto Politecnico do Porto
Rua Dr. Antonio Bernardino de Almeida, 431, 4200-072 Porto, Portugal
ningchen74@gmail.com
[2]CISUC- Department of Informatics Engineering, University of Coimbra, Portugal
bribeiro@dei.uc.pt
[3] Institute of Policy and Management, Chinese Academy of Sciences, P.R. China
acn@isep.ipp.pt

Abstract. Dimensionality reduction is demonstrated crucial to improve the predictive capability of models by means of linear or nonlinear projections. Non-negative matrix factorization (NMF) is a popular multivariate analysis technique for part-based data representation. It attempts to find an approximation of a high dimensional matrix as the product of two low dimensional matrices under the non-negative constraint. Recently a graph regularized non-negative matrix factorization (GNMF) provides a formal way to incorporate the geometrical structure into the NMF decomposition, particularly applicable to the data embedded in submanifolds of the Euclidean space. In this paper, the usage of GNMF in financial analysis is discussed from the perspectives of unsupervised clustering and supervised classification. Experimental results on a French bankruptcy data set show the potential of GNMF on data representation.

Keywords: bankruptcy analysis, clustering, classification, non-negative matrix factorization, manifold.

(Received November 16th, 2011 / Accepted December 29th, 2011)

1 Introduction

Bankruptcy prediction aims to predict the probability that the company may become bankrupt in the following years given a set of financial ratios that describe the situation of a company over a given period. In the world faced with the global economic and financial crisis, there has been a raising interest in seeking for more accurate predictive models able to better understand the financial data and prevent the sudden distress of companies. So far a large number of methods have been proposed following the research direction of statistical or intelligent approaches [2]. Statistical techniques aim to find an optimal linear combination of explanatory input variables in order to model, analyze and predict corporate default (bankrupt) risk. The pioneer statistical techniques include univariate, multivariate discriminant analysis, risk index models, and conditional probability model. Due to the criticism on traditional statistical models, many recent efforts are devoted into state-of-the-art intelligent approaches, which offer theories about how financial crises could be predicted. Various prediction models have been proposed using a wide range of intelligent methods including neural network (NN), fuzzy set theory (FS), decision tree (DT), case-based reasoning (CBR), support vector machine (SVM), and soft computing [16]. Neural Networks (NNs) have been actively used in bankruptcy prediction yielding reasonably accurate models [1, 6]. A multi-layer perceptron (MLP) obtains desirable outcome on Taiwan and United States markets [9], and Iranian companies [15]. In [8] a stable credit rating model based on Learning Vector Quantization (LVQ) is applied to corporate failure prediction and credit risk analysis. Likewise, SVM has been proven to yield sound predictive performance with a relatively small amount of data [13, 23].

Dimensionality reduction is an effective information visualization approach capable to project the high-dimensional input data to a low-dimensional output space so that the intrinsic relationship can be captured and graphically represented. A large variety of dimensionality reduction methods have been proposed to provide appealing solution to the task, including t-test, correlation matrix, factor analysis, principle component analysis (PCA) [17], independent component analysis (ICA) [7], etc. Regarding the financial data, some variables have small discriminatory capabilities for default (bankrupt) prediction with linear statistical models, whereas non-linear approaches can extract relevant (and discriminatory) information improving the visualization and classification. Manifold learning includes a number of non-linear approaches to data analysis that exploit the geometric properties of the manifold on which the data is supposed to lie. Their properties make them often used in financial applications due to their excellent capability to treat non-linear data. Manifold learning methods such as ISOMAP, Supervised ISOMAP algorithm (ES-ISOMAP), local linear embedding (LLE), Laplacian Eigenmaps have been used successfully in this task [20, 21].

Non-negative matrix factorization (NMF) [10] is a multivariate analysis technique which factorizes a matrix X into two matrices U and V so that $X \approx UV^T$. Different from other factorization methods, such as principal component analysis (PCA) and singular value decomposition (SVD), it enforces the non-negative constraint on the decomposition, in other words, the values of U and V must be equal to or greater than zero. Different objective functions have been proposed leading to a number of variants of NMF algorithms, in which two commonly used are the squared error [14] and the divergence [11]. Accordingly, optimization schemes are derived to minimize the underlying objective functions by iterative update rules. NMF receives considerable attentions in diverse domains including pattern recognition, computer vision and information retrieval. It is demonstrated that NMF outperforms other matrix factorization techniques in document clustering and face recognition [12]. In the field of financial analysis, NMF is used to extract the discriminative features by the embedded learning process to which classification and prediction algorithms can be easily applied [19].

In a recent study [5], a graph regularized non-negative matrix factorization (GNMF) uses a new objective function which incorporates a nearest neighbor graph structure, to reflect the geometrical property of the data in the resulted part-based representation. The new optimization scheme makes GNMF particularly applicable to detect the manifold structure embedded in the Euclidean space. Among various dimensionality reduction techniques, GNMF is beneficial to discriminative feature extraction by encoding the intrinsic geometric information into matrix factorization, and hence helpful to the subsequent classification. It is shown that GNMF combined with SVM is very effective for both bankruptcy prediction (supervised) and visualization (unsupervised) [18]. In this paper, GNMF is used to get more compact representation of financial data through a clustering process, and to construct hybrid classification models combined with advanced learning methods. The contribution of this study is twofold: to investigate the the potential of GNMF on predictive performance enhancement, and to compare some hybrid prediction models based on GNMF.

The remainder of this paper is organized as follows. The next section introduces the principle and optimization scheme of graph regularized non-negative matrix factorization. Section 3 presents the experimental results on a real world financial data. Lastly, the conclusions and future directions are discussed in section 4.

2 A Brief Introduction of GNMF

The problem of non-negative matrix factorization (NMF) is stated as follows: given a matrix $X = \{x_{ij}\} \in R^{m\times n}$, decompose X into two non-negative matrices $U = \{u_{ij}\} \in R^{m\times k}$ and $V = \{v_{ij}\} \in R^{n\times k}$ so that $X \approx U * V^T$. The squared Euclidean distance (F-norm) is the commonly used objective function, thereby NMF can be formulated as an optimization problem:

$$\begin{array}{ll} Min_{U,V} & ||X - UV^T||^2 \\ st. & u_{ij} \geq 0, v_{ij} \geq 0 \end{array} \quad (1)$$

In real applications, usually $k \ll m$ and $k \ll n$, thus each data vector $x_j (j = 1, ..., n)$ can be approximated by a linear combination of the columns of U, weighted by the components of V.

The rational behind GNMF is that the intrinsic geometrical structure of the data is approximated by manifold rather than the Euclidean space. The local geometric structure can be modeled by a nearest graph to which a weight matrix W is specified by the p-nearest neighbors manner. The weighting schemes include binary weighting, heat kernel weighting and dot-product weighting [5].

We denote $z_j = \{v_{j1}, ..., v_{jk}\}, j = 1, ..., n$ as the row vector of V, and W_{jl} as the weight of the edge between point x_j and x_l on the graph. The smoothness of low dimensional representation is measured as:

$$R = \frac{1}{2}\sum_{j,l=1}^{n} ||z_j - z_l||^2 W_{jl} \quad (2)$$

GNMF intends to combine the geometrically based regularizer with the traditional objective function of NMF in a more sophisticated objective function, defined as:

$$||X - UV^T||^2 + \frac{\lambda}{2}\sum_{j,l=1}^{n} ||z_j - z_l||^2 W_{jl} \quad (3)$$

In the above function, $\lambda \geq 0$ is the regularization parameter, in particular when $\lambda = 0$, GNMF becomes traditional NMF. Normally, the optimization is solved by updating U and V alternatively through an iterative process:

1. Construct a p-nearest neighbor graph, taking each column vector of matrix X as a point;
2. Assign the weight matrix W of the graph;
3. Initialize the matrix U and V randomly;
4. Fix U, optimize V to minimize the objective function:
$$u_{ik} = u_{ik}\frac{(XV)_{ik}}{(UV^TV)_{ik}} \quad (4)$$
5. Fix V, optimize U to minimize the objective function:
$$v_{jk} = v_{jk}\frac{(X^TU + \lambda WV)_{jk}}{(VU^TU + \lambda DV)_{jk}} \quad (5)$$
6. Repeat from 4 until it converges.

3 Experiment and Discussion

As was known, GNMF can be employed for different usages. Firstly, the results of GNMF forms a natural clustering of original data in the way that each instance can be regarded as the additive mixture of the column vectors of matrix U, and the matrix V implies the membership values. Secondly, GNMF can be used as a dimensionality reduction method before the classification process. In this section, we apply GNMF to a French data set and explore its usage in both clustering and classification. We use the Matlab implementation of GNMF provided in [3].

3.1 Data Description

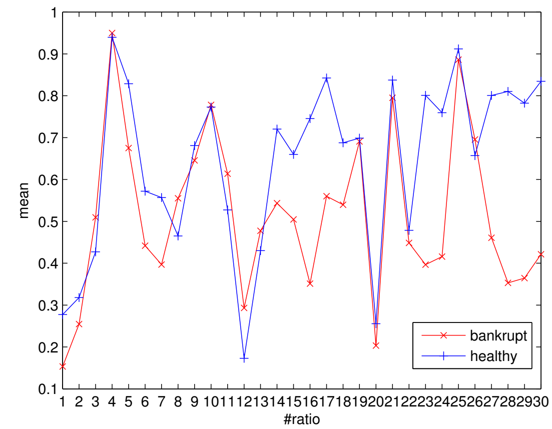
The French data set contains financial ratios of small or medium sized companies. In the total 1200 instances, 600 examples are distressed in 2007. The financial ratios used to describe the statement of companies in the year 2006 include Number of Employees (x_1), Capital Employed / Fixed Assets (x_2), Financial Debt / Capital Employed (x_3), Depreciation of Tangible Assets (x_4), Working Capital / Current Assets (x_5), Current Ratio (x_6), Liquidity Ratio (x_7), Stock Turnover days (x_8), Collection Period days (x_9), Credit Period days (x_{10}), Turnover per Employee (x_{11}), Interest / Turnover (x_{12}), Debt Period days (x_{13}), Financial Debt / Equity (x_{14}), Financial Debt / Cashflow (x_{15}), Cashflow / Turnover (x_{16}), Working Capital / Turnover days (x_{17}), Net Current Assets/Turnover days (x_{18}), Working Capital Needs / Turnover (x_{19}), Export (x_{20}), Added Value per Employee ($x_1 21$), Total Assets Turnover (x_{22}), Operating Profit Margin (x_{23}), Net Profit Margin (x_{24}), Added Value Margin (x_{25}), Part of Employees (x_{26}), Return on Capital Employed (x_{27}), Return on Total Assets (x_{28}), EBIT Margin (x_{29}), and EBITDA Margin (x_{30}). In the preprocessing phase, the data is normalized to unity range. Figure 1(a) illustrates the mean values of the financial ratios with respect to bankrupt companies and healthy companies. The significance test results (p-value) are shown for each ratio in Figure 1(b). As can been seen, all financial ratios except x_{10} (p=0.148624) and x_{19} (p=0.113525) are significantly different between bankrupt and healthy companies at the level 5%.

3.2 Clustering

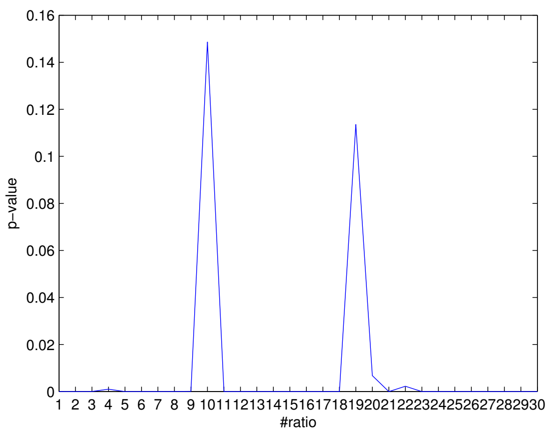
In this experiment, GNMF is used for clustering and compared with two popular clustering algorithms, namely PCA and K-means.

PCA is the simplest multivariate analysis method for dimensionality reduction and visualization by computing the eigenvectors of covariance matrix. PCA is a linear orthogonal transformation in the sense that the new features (components) are orthogonal with each other and represented as a linear combination of the original features. By extracting a small number of components which account for the most variability of the data, the high dimensional data is projected to a low dimensional space. For comparison, PCA projects the original data into a 2-d space, thereby the instances are separated into two clusters according to the dominant component.

K-means is one of the most widely used cluster analysis method which partitions the data into k clusters



(a) Mean values of financial ratios



(b) Significance test results

Figure 1: Data analysis of distressed companies and healthy companies

by the nearest neighbor principle. In the setting of K-means algorithm, the number of clusters (k) is 2, and the number of runs with different initial centers is 5.

As a part-based representation, GNMF decomposes an instance into a combination of k ranks with the coefficients indicating the membership of the clusters, thus the instances can be assigned to the cluster with the maximal membership. Regarding GNMF, the binary weighting assignment is used to construct the weight matrix, that is to say $W_{ij} = 1$ if and only if x_i and x_j are among the 5-nearest neighbors with each other, otherwise $W_{ij} = 0$. The value of λ is set as $0, 1, 10, 100, 1000, 10000$ respectively.

As is conventional in the literature, we evaluate the clustering results in terms of three measures, namely purity, rand index and normalized mutual information. We denote the real clustering as C and the resulted clustering as C'.

To compute purity (Pu), each resulted cluster is assigned by the majority label of the members, and the purity is the percent of the correctly assigned instances. In the following definition, $c(x_i)$ is the real label of x_i, $c'(x_i)$ is the clustered label, and δ is the indicator function taking the value 1 when the condition satisfies, otherwise 0.

$$Pu = \frac{\sum_{1 \le i \le n} \delta(c(x_i) = c'(x_i))}{n} \quad (6)$$

Rand index (RI) is the percent of the pairs that belong to the same or different classes simultaneously in the real and resulted clustering respectively.

$$RI = \frac{\sum_{i \neq j} \delta(c(x_i), c'(x_i))}{n(n-1)} \quad (7)$$

where $\delta(c(x_i), c(x_j)) = 1$ if $c(x_i) = c(x_j) \bigcap c'(x_i) = c'(x_j)$ or $c(x_i) \neq c(x_j) \bigcap c'(x_i) \neq c'(x_j)$, otherwise 0.

Mutual information (MI) measures the correlation between two clustering using entropy. Let $P(c_i)$ be the probability of instances selected from the i^{th} cluster of C, $P(c'_i)$ be the probability selected from the i^{th} cluster of C', $P(c_i \bigcap c'_j)$ be the probability selected from i^{th} cluster of C and j^{th} cluster of C' simultaneously, the mutual information is defined as:

$$MI = \sum_{c_i \in C} \sum_{c_j \in C'} P(c_i \bigcap c'_j)\ log_2 \frac{P(c_i \bigcap c'_j)}{P(c_i)P(c'_j)} \quad (8)$$

The normalized mutual information (NMI) is defined in several forms. Here we use the one defined in [4], where E_c and $E_{c'}$ are the entropy of real and resulted clustering respectively.

$$NMI = \frac{MI}{max(E_c, E_{c'})} \quad (9)$$

Table 1 shows the clustering results of PCA, K-means and GNMF. As was seen from this table, GNMF always outperforms PCA and K-means in terms of purity and rand index regardless of the λ values. When using NMI, GNMF performs comparatively with K-means but still better than PCA in all cases. This suggests that GNMF has a good clustering capability by leveraging the power of part-based representation and geometrical structure.

3.3 Classification

Given the class label, the problem of the classification task is to predict whether a company may go bankrupt

Table 1: Clustering results of 3 methods (Pu: Purity, RI: rand index, NMI: normalized mutual information).

Method	Pu	RI	NMI
PCA	0.823	0.708	0.351
K-means	0.823	0.709	0.388
GNMF($\lambda = 0$)	0.836	0.726	0.382
GNMF($\lambda = 1$)	0.837	0.727	0.383
GNMF($\lambda = 10$)	0.832	0.720	0.372
GNMF($\lambda = 100$)	0.833	0.721	0.376
GNMF($\lambda = 1000$)	0.831	0.719	0.375
GNMF($\lambda = 10000$)	0.837	0.727	0.387

according to the historical situation. In the hybrid approaches, GNMF serves as a dimensionality reduction method, followed by the classification methods to construct the prediction model. In the present study, we used 5 advanced learning models, namely Decision Table, BayesNet, Logistic, SVM, and Multilayer Perceptron (MLP) for the complicated non-linear separation problem. The experiments are performed in Weka, an open-source data mining tool [22]. For all models, we use the default parameters and test different values of dimension (rank). The experiments are performed as follows (illustrated in Figure 2):

1. Set the number of rank k;
2. The entire data set is divided randomly into 10 folds for cross-validation;
3. For each training data set D_{train}, GNMF is performed to calculate the decomposition matrices: U and V_{train} so that $D_{train} \approx U * V_{train}^T$;
4. For the test data set D_{test}, calculate the transformed matrix $V_{test}^T = U' * D_{test}$ where U' is the Moore-Penrose pseudoinverse of matrix U obtained in the training stage.
5. Construct classification models using the data set V_{train};
6. For validation, feed the resultant V_{test} to the model and predict the class.
7. After the experiment is repeated 10 times, evaluate the performance by averaging the results of each run.

The experiment is done with 10-fold validation, and we average the results by running each model 10 times. The classification accuracy is shown in Table 2, giving the accuracy and standard deviation. Due to the fact that the performance on default company (true positive rate) is more important than that on non-default company (true negative rate), we also list the results of true positive rate in Table 3. The significance test is performed for each row taken Decision Table as the baseline model. The first row shows the performance of models without GNMF, and the following rows show the performance of hybrid models using different ranks (k = 2, 4, 9, 16, 25, 36, 64, 81). As can be seen, all hybrid models for specific k (ranks) yield better performance in terms of both overall accuracy and true positive rate compared to the stand-alone models without GNMF. Indeed, we observe the accuracy improved by 0.82% (Decision Table), 0.23%(BayesNet), 0.43% (Logistic), 0.6% (SVM), and 1.15% (MLP), while the true positive rate improved by 2% (Decision Table), 0.2% (BayesNet), 0.21% (Logistic), 1.23% (SVM), and 1.27% (MLP). It suggests that GNMF is effective to improve the predictive accuracy.

The statistical significance of the difference among five models using a t-test is summarized in Table 4 and Table 5. The significance level is set as 5%, so that the p-value less than 5% indicates that the two underlying methods are significantly different in the mean. As was observed, method (d) and (c) significantly outperform the other three methods in terms of prediction accuracy as shown by the p-values less than 5%. The p-values 0.001181 and 0.001681 show that model (e) is significantly better than (a) and (b). Meantime, in terms of true positive rate method (c) and (e) perform significantly better than (a) and (d), followed by (b). It can be concluded that Logistic, MLP and SVM have higher prediction power than Decision Table and BayesNet.

4 Conclusions

Nonlinear projection is demonstrated effective in bankruptcy prediction problem reducing the dimensionality of data and improving the prediction performance. In this paper, we discuss the usage of graph regularized non-negative matrix factorization (GNMF) in financial analysis as a tool of clustering and classification. Experimental results show GNMF is applicable to explore the intrinsic structure of the high dimensional bankrupt data and enhance the predictive ability in corporation with classification models.

Future work will extend this study in several research directions. There are several parameters of GNMF algorithm, such as λ, weighting scheme, and number of nearest neighbors. We intend to optimize the the parameters through a comparative study. In the presented work, we use the random initialization for the decomposition matrices, however, other initialization methods which are reported to achieve a desired

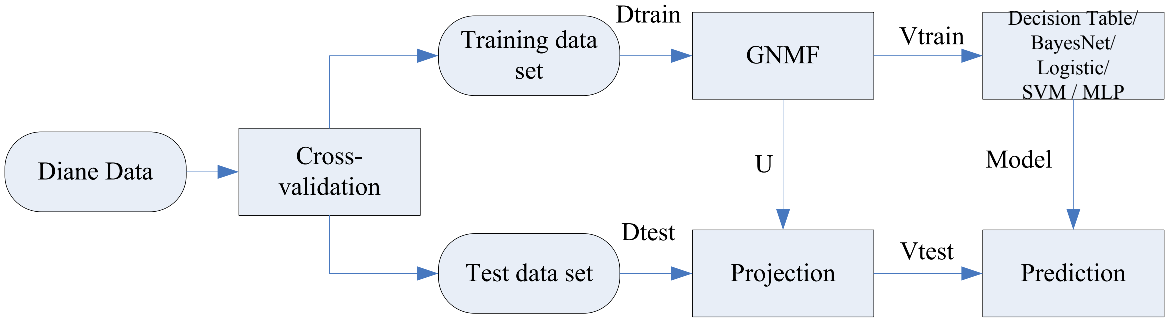


Figure 2: The classification experiments

Table 2: Accuracy results of classification using (a) Decision Table, (b) BayesNet, (c) Logistic, (d) SVM, (e) MLP

Dataset	(a)	(b)	(c)	(d)	(e)
D	86.92±3.05	86.60±2.96	91.09±2.51 ∘	90.57±2.41 ∘	88.95±2.72 ∘
D2	84.78±3.11	83.96±3.22 •	84.88±3.01	83.90±2.94 •	84.52±3.20
D4	82.61±3.31	83.28±3.25 ∘	84.75±2.99 ∘	84.22±2.82 ∘	84.18±2.92 ∘
D9	84.22±3.33	84.22±2.82	85.40±3.24 ∘	85.32±3.28 ∘	84.50±3.34
D16	85.40±3.01	85.56±2.74	89.92±2.57 ∘	89.32±2.59 ∘	87.17±3.01 ∘
D25	85.14±3.11	86.83±2.95 ∘	91.52±2.55 ∘	91.17±2.40 ∘	88.67±2.81 ∘
D36	86.21±2.86	84.54±3.09 •	91.11±2.25 ∘	91.06±2.47 ∘	90.05±2.65 ∘
D49	84.95±2.91	85.62±2.72 ∘	90.34±2.68 ∘	90.96±2.53 ∘	89.41±2.75 ∘
D64	87.74±2.70	85.78±2.91 •	90.19±2.80 ∘	91.12±2.49 ∘	90.00±2.79 ∘
D81	86.19±2.81	85.60±2.94 •	89.56±2.88 ∘	91.43±2.16 ∘	90.10±2.67 ∘
Average	85.42	85.20	88.87	88.91	87.75

∘, • statistically significant improvement or degradation

Table 3: True positive rate results of classification using (a) Decision Table, (b) BayesNet, (c) Logistic, (d) SVM, (e) MLP.

Dataset	(a)	(b)	(c)	(d)	(e)
D	82.10±5.33	79.35±4.83 •	88.82±3.77 ∘	85.47±4.14 ∘	87.95±3.89 ∘
D2	78.95±4.93	77.90±5.00 •	77.90±4.75 •	74.00±5.09 •	79.45±5.73
D4	76.88±6.04	77.53±5.18	77.90±4.71 ∘	74.27±4.73 •	79.27±5.17 ∘
D9	77.03±5.74	73.57±5.16 •	79.42±4.97 ∘	76.90±5.00	81.13±5.92 ∘
D16	79.70±5.06	77.13±4.85 •	86.67±4.36 ∘	84.22±4.50 ∘	85.62±4.30 ∘
D25	81.28±5.65	79.55±4.82 •	89.03±3.69 ∘	86.70±4.23 ∘	87.52±4.09 ∘
D36	82.97±5.07	79.18±4.99 •	88.63±3.51 ∘	86.25±4.16 ∘	88.77±4.07 ∘
D49	80.35±5.53	78.17±4.50 •	88.70±3.49 ∘	86.60±4.38 ∘	88.30±3.72 ∘
D64	84.10±4.79	78.28±4.79 •	88.32±3.49 ∘	86.65±4.28 ∘	88.82±4.21 ∘
D81	82.82±4.91	76.83±4.89 •	87.77±3.72 ∘	86.48±3.96 ∘	89.22±3.85 ∘
Average	80.62	77.75	85.32	82.75	85.60

∘, • statistically significant improvement or degradation

solution (e.g., the cut-weighted NMF [5] and K-means [19]) will be included in the classification method.

ACKNOWLEDGEMENTS

This work was partly supported by Natural Science Foundation of China (Contract No. 91024004).

References

[1] Baek, J. and Cho, S. Bankruptcy prediction for credit risk using an auto-associative neural network in korean firms. In *Proceedings of IEEE International Conference on Computational Intelligence for Financial Engineering*, pages 25–29, 2003.

[2] Bellovary, J., Giacomino, D., and Akers, M. A

Table 4: Significance test results on accuracy: (a) Decision Table, (b) BayesNet, (c) Logistic, (d) SVM, (e) MLP, significance is at 5% level.

	b	c	d	e
a	0.551285	0.000365 ○	0.000886 ○	0.001181 ○
b		0.000146 ○	0.000613 ○	0.001681 ○
c			0.912293	0.011428 •
d				0.003454 •

○, • statistically significant improvement or degradation

Table 5: Significance test results on true positive rate: (a) Decision Table, (b) BayesNet, (c) Logistic, (d) SVM, (e) MLP, significance is at 5% level.

	b	c	d	e
a	0.001573 •	0.000929 ○	0.093297	0.000047 ○
b		0.000258 ○	0.009522 ○	0.000065 ○
c			0.000005 •	0.467125
d				0.000251 ○

○, • statistically significant improvement or degradation

review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, 33:1–43, 2007.

[3] Cai, D. Non-negative matrix factoriaztion on manifold. http://www.zjucadcg.cn/dengcai/GNMF/index.html.

[4] Cai, D., He, X., and Han, J. Document clustering using locality preserving indexing. *IEEE Transactions on Knowledge and Data Engineering*, 17(12):1624–1637, 2005.

[5] Cai, D., He, X., Han, J., and Huang, T. Graph regularized non-negative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1548–1560, 2011.

[6] Charalambous, C., Charitou, A., and Kaourou, F. Application of feature extractive algorithm to bankruptcy prediction. In *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, volume 5, pages 303–308, 2000.

[7] Chen, N. and Vieira, A. Bankruptcy prediction based on independent component analysis. In *1st International Conference on Agents and Artificial Intelligence (ICAART09)*, pages 150–155, 2009.

[8] Chen, N., Vieira, A., Ribeiro, B., Duarte, J., and Neves, J. A stable credit rating model based on learning vector quantization. *Intelligent Data Analysis*, 15(2):237–250, 2011.

[9] Huang, Z., Chen, H., Hsu, C., Chen, W., and Wu, S. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, 37:543–558, 2004.

[10] Lee, D. and Seung, H. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[11] Lee, D. and Seung, H. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 13:556–562, 2001.

[12] Lee, J.-H., Park, S., Ahn, C.-M., and Kim, D. Automatic generic document summarization based on non-negative matrix factorization. *Information Processing & Management*, 45(1):20–34, 2009.

[13] Min, J. H. and Lee, Y. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4):603–614, 2005.

[14] Paatero, P. and Tapper, U. Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics.*, 5(2):111–126, 1994.

[15] Rafiei, M., Manzari, S., and Bostanian, S. Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications*, 38(8):10210–10217, 2011.

[16] Ravi Kumar, P. and Ravi, V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques - a review. *European Journal of Operational Research*, 180(1):1–28, July 2007.

[17] Rekba Pai, G., Annapoorani, R., and Vijayalakshmi Pai, G. Performance analysis of a statistical and an evolutionary neural network based classifier for the prediction of industrial bankruptcy. In *IEEE Conference on Cybernetics and Intelligent Systems*, pages 1033–1038, 2004.

[18] Ribeiro, B. and Chen, N. Graph weighted subspace learning models in bankruptcy. In *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 2055–2061, 2011.

[19] Ribeiro, B., Silva, C., Vieira, A., and Neves, J. Extracting discriminative features using non-negative matrix factorization in financial distress data. In Kolehmainen, M. and et al., editors, *Int Conf on Adaptive and Natural Computing Algorithms*, volume 4432, pages 537–547. Lecture Notes in Computer Science (LNCS), Springer, 2009.

[20] Ribeiro, B., Vieira, A., Duarte, J., Silva, C., Neves, J., Liu, Q., and Sung, A. Learning manifolds for bankruptcy analysis. In Köppen, M. and et al., editors, *International Conference on Neural Information Processing*, volume 5506, pages 722–729. Lecture Notes in Computer Science (LNCS), Springer, 2009.

[21] Ribeiro, B., Vieira, A., and Neves, J. Supervised isomap with dissimilarity measures in embedding learning. volume 5197, pages 389–396. Lecture Notes in Computer Science (LNCS), Springer, 2008.

[22] Witten, I. and Frank, E. *Data mining: Practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann Publishers Inc., San Francisco, CA, 2000.

[23] Yang, Z., You, W., and Ji, G. Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 38(7):8336–8342, 2011.