

A New Inverse N^{th} Gravitation Based Clustering Method for Data Classification

Huarong Xu^(✉), Li Hao, Chengjie Jianag, and Ejaz Ul Haq

Department of Computer Science and Technology, Xiamen University of Technology,
Xiamen 361024, China

hrxu@xmut.edu.cn, {923752195, 1181502446, ejaz4616}@qq.com

Abstract. Data classification is one of the core technologies in the field of pattern recognition and machine learning, which is of great theoretical significance and application value. With the increasing improvement of data acquisition, storage, transmission means and the amount of data, how to extract the essential attribute data from massive data, data accurate classification has become an important research topic. Inverse n^{th} n order gravitational field is essentially a generalization of the n order in the physics, which can effectively describe the interaction between all the particles in the gravitational field. This paper proposes a new inverse n^{th} power gravitation (I-n-PG) based clustering method is proposed for data classification. Some randomly generated data samples as well as some well-known classification data sets are used for the verification of the proposed I-n-PG classifier. The experiments show that our proposed I-n-PG classifier performs very well on both of these two test sets.

Keywords: Data classification · Inverse n^{th} power gravitation · Clustering algorithm

1 Introduction

1.1 Background

With the rapid development of data acquisition, storage and transmission technology, the number of digital data storage is growing, which provides a wealth of information for the production and life of human. But at the same time, with the sharp increase in the amount of data, the human has not in limited time interested data were analyzed one by one, which gives the mankind put forward a new challenge, how to automatically achieve massive data classification by computer?

Data classification is one of the core technologies in pattern recognition and machine learning. It has been widely used in many aspects of people's production and life. The goal of data classification is to automatically extract the model from the known data set, and then judge the classification of unknown data. In recent years, researchers have carried out extensive research in data classification, put forward the classification method of large amounts of data, including decision tree classification method, Bayes method, classification method, classification based on neural network based

classification method, subspace learning and manifold learning based on kernel learning classification method and so on. However, these methods still cannot fully reveal the mass data concentration the relationship between the data, so how to effectively describe the correlation between the data structure and has strong discriminative ability of the classifier, one of the problems to be solved is still facing researchers.

The inverse n^{th} [1–3] is a generalized n order generalized field in physics, which can effectively describe the interaction of all the particles in the gravitational field. In the gravitational field, each particle is subjected to the gravitational force of the other particles and is maintained in a state of dynamic equilibrium. Once a new particle to enter or exit the gravitational field, the dynamic balance of the original state will be broken, all the particles in the gravitational field will be in accordance with their respective force movement adjustment, finally realize the new dynamic balance. If each data set as an inverse n^{th} particle in gravitational field, and give it a certain quality in accordance with certain rules, then the data can be mapped to an inverse n^{th} gravitational field, the classification of the data will be transformed into the original particle in the field of inverse n^{th} gravity classification problem. The interaction between particles in the field through the research of inverse n^{th} order to construct particle classification model, will be more effective to describe the intrinsic attributes of the original data, realizes the automatic classification of massive data accurately.

2 Related Works

As the core problem in the field of pattern recognition and machine learning, data classification has been studied in recent decades. Through continuous exploration and research, researchers have made many important achievements in the research of data classification.

The basic principle of decision tree classification is to use the information gain search data set with the maximum amount of information in the field of information, and then set up a decision tree node [4–8]. Then a branch of the tree is established according to the different values of the attribute field, and the next node and the corresponding branch of each branch are continued to build the tree. Finally, each path in the decision tree corresponds to a rule, and the whole decision tree corresponds to a set of expression rules. For example, Murth gives a general procedure for constructing decision trees based on a given data set [5]. Zheng [7] presents a multi variable decision tree algorithm, which improves the classification accuracy of decision trees by constructing new binary features. Brazdil and Gama [8] fusion decision tree and linear discriminant model are used to construct multi variable decision tree. In this model, the new feature is obtained by the linear combination of the previous features.

The basic principle Bayes classification is to use probability to describe the data samples of the category of uncertainty, in the case of a priori probability and conditional probability is known, based on the Bayes theorem to predict the type of test samples [9–11]. In the literature, the existing Bayes classification methods can be divided into Naive Bayes classifiers [9] and Bayesian Networks [10, 11] two categories. Naive Bayes classifiers assume that each feature in a given feature is independent from each other. This assumption limits

the scope of application of Naive Bayes classifiers, but its expression is simple and can effectively deal with the classification problem of independent (or approximately independent) features of some features. However, in many cases, the characteristics of the feature sets are often not independent, but have a strong correlation, the relationship between these features cannot be trained through the simple Bayes classifier. Bayes network is a graphical network based on probabilistic reasoning, which allows the definition of class conditional independence between subsets of variables, so it can effectively deal with the problem of data classification in feature correlation.

Artificial neural network classification [12–15]: artificial neural network is mathematical model that simulate neural network behavior characteristics and process distributed parallel information, has a wide range of applications in the pattern classification. Lain and Lee [15] analyses the use of artificial neural network to construct text classifier and dimensionality reduction method. Kon and Plaskota [13] study the necessary minimum number of neurons in the feedforward neural network for a given task. Siddique and Tokhi [14] use genetic algorithm to train the weights of artificial neural network.

The classification based on spatial learning and manifold learning: In linear discriminant analysis (LDA) [31], the original high dimensional samples are projected into the low dimensional subspace, while ensuring the original sample with the maximum and minimum distance between class distance in low dimensional subspace. Dick et al. proposed a supervised locally linear embedding algorithm based on the original local linear embedding algorithm [27, 28], which can maintain the local topological relations of the original data in the low dimensional data. Tenenbaum et al. propose a manifold learning method (Isomap [29]) based on geodesic distance based on MDS (Scaling Multidimensional). Belkin and Niyogi proposed Laplacian Eigenmaps [30], whose basic idea is to use an undirected weighted graph to describe a manifold, and the low dimensional expression of the original sample is embedded in the graph. He et al. propose the Linearization of the method in the literature [30], and the Laplacian face [37] is proposed. The method is to keep the neighbor relationship between samples before and after dimensionality reduction. Yan et al. [38] proposed a unified framework for graph embedding based on the sample neighbor relationship, and an interval Fisher analysis method (MFA) was proposed

Kernel based classification [39, 40, 17–26]: the basic idea is to map the samples from low dimensional space to high dimensional feature space by kernel function. The linear non separable data samples in the input space can be linearly separable in the feature space with higher dimension. For example, Vapnik [25] proposed is to establish a learning kernel method based on statistical theory and support vector machine (Support Vector Machine, SVM), its basic principle is to find a optimal classification requirements of the hyperplane, the hyperplane in ensuring the precision of classification at the same time, the maximum over both sides of the plane interval. In theory, the support vector machine is able to achieve the optimal classification of linear separable data. Tipping [41] put forward the relevance vector machine (Relevance Vector Machine, RVM), its basic principle is the point of decision theory to remove irrelevant structure in active

correlation based on priori parameters, to obtain the sparse model. How to select a suitable kernel function based on the kernel function classification method is not suitable for a given data set.

Classification based on sparse representation: a face recognition method based on sparse representation is proposed by [34]. This method assumes that the face image can be represented as a linear combination of a small number of samples in the training data set. Based on the same assumption, the paper proposed a method of facial image restoration and recognition based on sparse representation [35]. Using the sparsity of wavelet coefficients, a new method of wavelet denoising based on sparse representation is proposed in the paper, which is based on the [36]. Although the sparse representations in many classification applications show excellent performance, however, recently, there are some references [31–33] further discuss the role of sparse representation plays in the image classification problem of face recognition.

3 The New Gravitation Based Clustering Algorithm

A new data clustering approach based on inverse n th power gravitation for large scale data classification is present in this section. Before the discussion of the new algorithm, we firstly discuss some knowledge about the universal gravitation.

The traditional law of gravitation was proposed by Newton in 1687, with the equation shown below in Eq. 1, where F is the force, G is the gravitational constant ($G = 6.67 \times 10^{-11}$), m_1 and m_2 are the masses of the object interacting, r is the distance between the centers of the masses.

$$F = G \times \frac{m_1 \times m_2}{r^2} \quad (1)$$

This gravitation is everywhere, and gravitation also shows effect on space-time such as distortion generated by the mass of an object in general relativity. A figure of this distortion is shown below in Fig. 1.

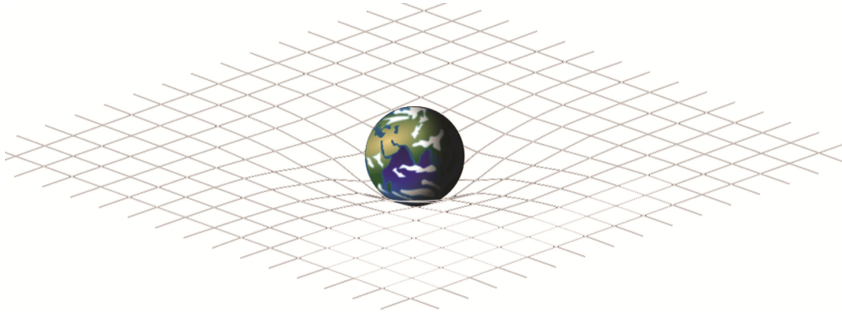


Fig. 1. Space-time distortion affected by gravitation

Gravitation not only can make distortion over space-time, but also can be utilized for data clustering. Samples in each training set can be considered as particles, and if each particle is assigned a mass value, then we can form a new attraction equation. We make an extension of the traditional equation of gravitation shown in Eq. 1 to a new one, shown in Eq. 2

$$F(x_i, x_j) = \frac{m(x_i)m(x_j)}{\|x_i - x_j\|^{n+1}}(x_i - x_j), n \geq 2 \quad (2)$$

X_i and X_j are two D-dimensional samples/particles in the training set with the mass values $m(X_i)$ and $m(X_j)$. We denote the equation as the inverse n^{th} power gravitation (I-n-PG) for data classification. The mass values of X_i and X_j incorporate the gravitational constant G , so there is no G in Eq. 2.

For the training set $S, S = \{x_1, x_2, \dots, x_N\} \subset R^D$, there are N samples in it. As to each sample X of set S in the gravitation field, the attraction force can be calculated by Eq. 3.

$$F(S, x) = m(x) \sum_{x_i \in S} \frac{m(x_i)}{\|x_i - x\|^{n+1}}(x_i - x) \quad (3)$$

Accordingly, modulo of $F(S, x)$ can be calculated according to Eq. 4.

$$F_m(S, x) = \|F(S, x)\| = m(x) \left\| \sum_{x_i \in S} \frac{m(x_i)}{\|x_i - x\|^{n+1}}(x_i - x) \right\| \quad (4)$$

The flowchart of the new proposed I-n-PG large scale data classification algorithm can be illustrated as follows (Fig. 2):

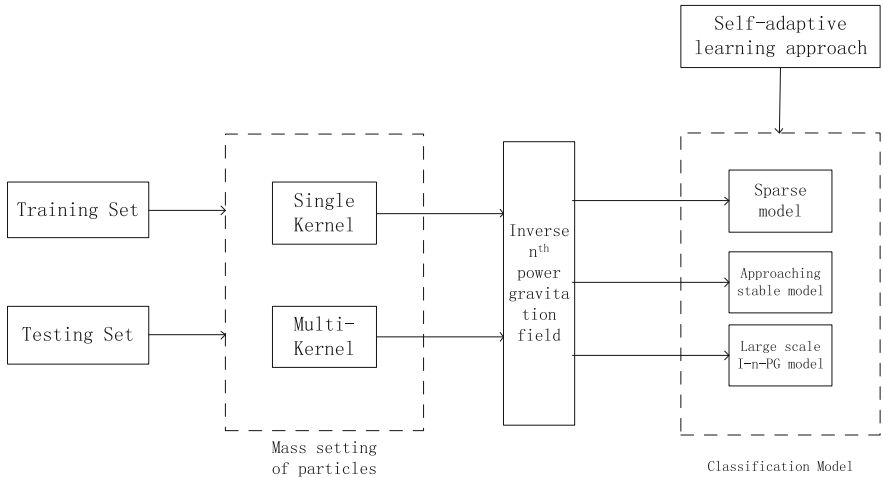


Fig. 2. The flowchart of the I-n-PG model

For the assignment of particle mass value of X_i , a kernel learning based approach is proposed. There are mainly two approaches, one is based on single kernel, and the other is on multi-kernel. In the single kernel approach, samples are projected into high dimensional space by kernel function, the mass value of a certain particle X_i in inverse n^{th} gravitation field can be calculated by the neighborhood distance of the particle in the kernel space. The nearer the distance toward other samples in the same class is, the heavier the particle mass is. Training samples and proper kernel function are used for the parameters adaption, and then we can calculate the mass value of particles. In the multi-kernel approach, it is an enhancement of the single kernel approaching, and it is used for tackling the cases that the training samples are in large scales, and usually are heterogeneous, that the single kernel approach cannot tackle. The multi-kernel model are constructed as follows in Eq. 5.

$$K = \sum_{i=1}^M \alpha_i K_i \quad (5)$$

Where, $\alpha_i \geq 0$, $\sum_{i=1}^M \alpha_i = 1$, K_i denotes the i^{th} kernel function, M is the number of kernels. Then Mass value of particles are calculated by the neighborhood distance in the multi-kernel space. Figure 3 gives the illustration of the calculation of mass value of certain particle in multi-kernel space.

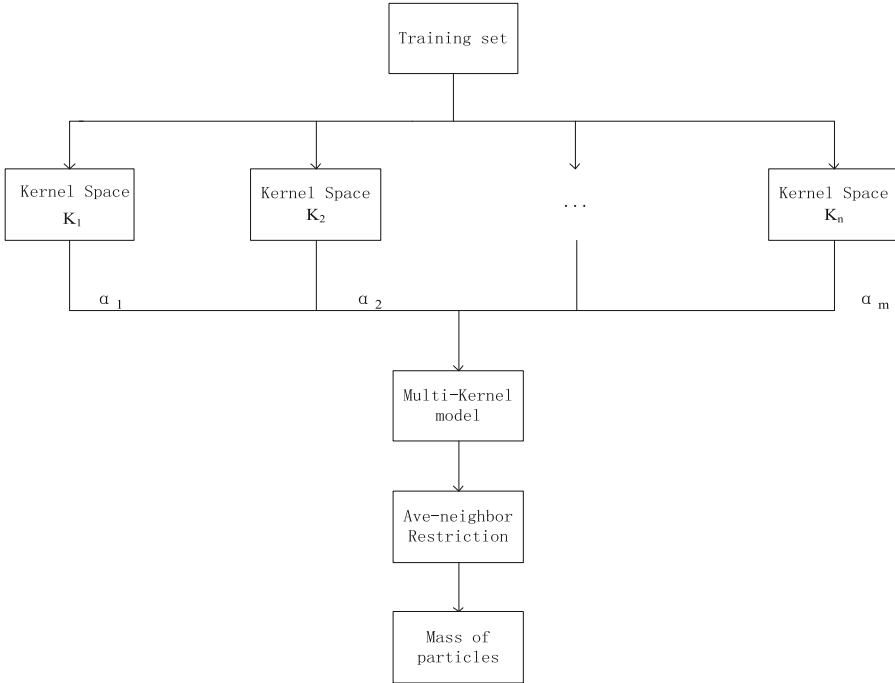


Fig. 3. The Mass value calculation in multi-kernel space

For the steady state calculation in the gravitational field, if a particle in the field has a relative small attraction force, it is considered as in a steady state, and when 90% of the particles in the training set are in steady state, we supposed this case a steady state of the particles. The new I-n-PG classifier can be illustrated as follows in Fig. 4.

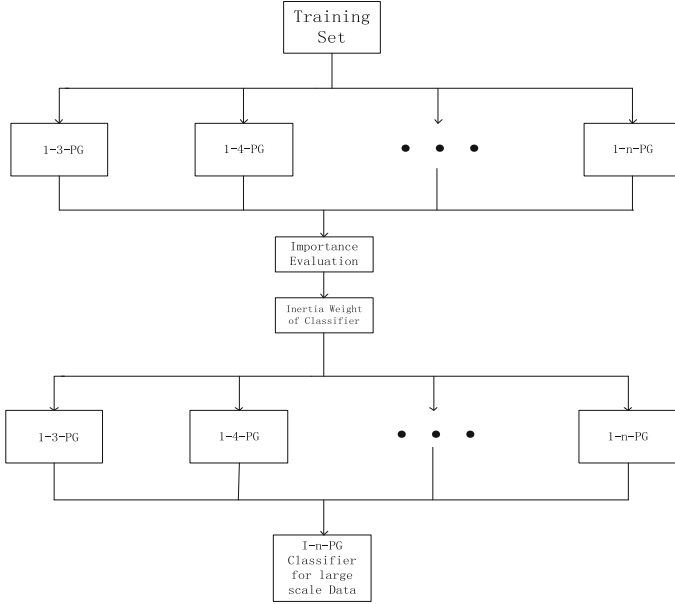


Fig. 4. The new I-N-PG classifier for large scale data

4 Experiment Result Analysis

Randomly generated data set as well as a well-known classification data set, the IRIS data are used here for the verification of the proposed inverse n^{th} gravitation based classifier. Several well-known clustering approaches such as 1NN, SVM SGF network are also used for the comparison here. For the randomly generated data of 2dimension samples, figures are given for the classification of different data samples.

Figure 5 gives the classification results by the I-n-PG classifier in a 2-D data set. The classification process begins the mass values initialization of samples in the n^{th} gravitation field, and then the calculation of steady state are done over all these samples. Then, the I-n-PG classifier are constructed based single kernel approach or multi-kernel approach. Finally, samples are classified by the I-n-PG classifier. In Fig. 5.

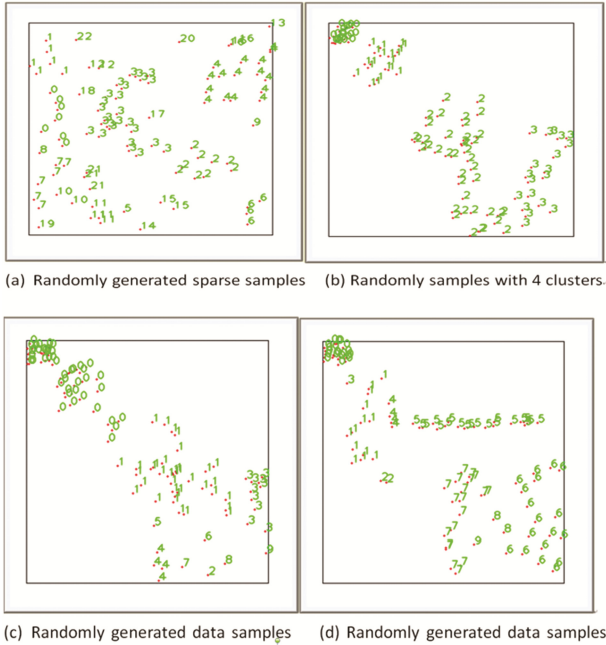


Fig. 5. Illustration of randomly generated data set under I-n-PG classifier

(a) samples are randomly generated in a square space without evident clusters. Samples in Fig. 5(b) can be easily seen that there are four clusters. Figure 5(c) gives samples with two evident clusters while other samples are in sparse and randomly locations. Figure 5(d) shows some lines cluster cases. Table 1 gives the comparison between the proposed I-n-PG classifier and some well-known classifiers, 1NN, SVM and SGF network under IRIS data sets. We can see form the table that the classic 1NN classifier achieves 94.2% success, classic SVM achieves 97.4% success, SGF network secure same performance as SVM classifier, and the proposed I-n-PG classifier outperforms all these contrasted ones.

Table 1. Comparison of best result for IRIS data set

I-n-PG	1NN	SVM	SGF network
98.5%	94.2%	97.4%	97.4%

5 Conclusion

In this paper, new inverse n^{th} power gravitation (I-n-PG) based clustering method is proposed for data classification. Some randomly generated data samples as well as some well-known classification data sets are used for the verification of the proposed I-n-PG classifier. The experiments show that our proposed I-n-PG classifier performs very well on both of these two test sets. A future work is to reduce the time cost during the

classification process as the calculation of n^{th} distance and the adaptive parameter for the steady state consumes lots of time.

Acknowledgment. This work was supported by National Nature Science Foundation of China under the research project 61273290.

References

1. Chandrasekhar, S.: *Newton's Principia for the Common Reader*. Oxford University Press, New York (1995)
2. Heintz, W.H.: Runge-Lenz vector for non- relativistic Kepler motion modified by an inverse cube force. *Am. J. Phys.* **44**(7), 687–694 (1976)
3. Roseveare, N.: *Mercury's Perihelion from Le verrier to Einstein*. Oxford University Press, New York (1982)
4. Paul, E.: Incremental Induction of Decision Trees. *Mach. Learn.* **4**(2), 161–186 (1997)
5. Murthy, S.K.: Automatic construction of decision trees from data: a multi-disciplinary survey. *Data Min. Knowl. Disc.* **2**, 345–389 (1998)
6. Mugambi, E.M.: Polynomial—fuzzy data knowledge-based systems: decision tree structures for classifying medical data. *Knowl.-Based Syst.* **17**(2), 81–87 (2004)
7. Zheng, Z.: Constructing conjunctions using systematic search on decision trees. *Knowl. Based Syst. J.* **10**, 421–430 (1998)
8. Gama, J., Brazdil, P.: Linear Tree. *Intell. Data Anal.* **3**, 1–22 (1999)
9. Meretakakis, D., Wuthrieh, B.: Extending Naive Bayes classifiers using long itemsets. In: *Proceeding 1999 International Conference Knowledge Discovery and Data Mining (KDD 1999)*, San Diego, pp. 165–174, August 1999
10. Cheng, J., Bell, D., Liu, W.: *Learning Bayesian Networks from Data: An Efficient Approach Based on Information Theory* (1999)
11. Acid, S., de Campos, L.M.: Searching for Bayesian Network structures in the space of restricted acyclic partially directed graphs. *J. Artif. Intell. Res.* **18**, 445–490 (2003)
12. Zhang, G.: Neural networks for classification: a survey. *IEEE Trans. Syst. Man Cybern. Part C* **30**(4), 451–462 (2000)
13. Kon, M., Plaskota, L.: Information complexity of neural networks. *Neural Netw.* **13**, 365–375 (2000)
14. Siddique, M.N.H., Tokhi, M.O.: Training neural networks: backpropagation vs. genetic algorithms. In: *IEEE International Joint Conference on Neural Networks*, vol. 4, pp. 2673–2678 (2011)
15. Lam Savio, L.Y., Lee, D.L.: Feature reduction for neural network based text categorization. In: *Digital Symposium Collection of 6th International Conference on Database System for Advanced Application* (1999)
16. Ji, S.W., Xu, W., Yang, M., Yu, K.: 3D convolutional neural networks for human action recognition. In: *The 27th International Conference on Machine Learning*, pp. 495–502 (2010)
17. Damoulas, T., Girolami, M.A.: *Pattern Recognition with a Bayesian Kernel Combination Machine* (2009)
18. Kingsbury, N., Tay, D.B.H., Palaniswami, M.: *Multi-Scale Kernel Methods for Classification* (2005)
19. Li, B., Zheng, D., Sun, L.: *Exploiting Multi-Scale Support Vector Regression for Image Compression* (2007)

20. Pozdnoukhov, A., Kanevski, M.: Multi-Scale Support Vector Algorithms for Hot Spot Detection and Modeling (2007)
21. Lin, Y.Y., Liu, T.L., Fuh, C.S.: Local Ensemble Kernel Learning for Object Category Recognition (2007)
22. Sonnenburg, S., Rlitsch, G., Schafer, C.: A General and Efficient Multiple Kernel Learning Algorithm (2005)
23. Smola, A.J., Schölkopf, B.: A Tutorial on Support Vector Regression (2004)
24. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
25. Karampatziakis, N.: Static analysis of binary executables using structural SVMs. *Adv. Neural. Inf. Process. Syst.* **23**, 1063–1071 (2010)
26. Ridder, D., Kouropteva, O., Okun, O., Pietikäinen, M., Duin, Robert P.W.: Supervised locally linear embedding. In: Kaynak, O., Alpaydin, E., Oja, E., Xu, L. (eds.) *ICANN/ICONIP -2003*. LNCS, vol. 2714, pp. 333–341. Springer, Heidelberg (2003). doi:[10.1007/3-540-44989-2_40](https://doi.org/10.1007/3-540-44989-2_40)
27. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000)
28. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000)
29. Belkin, M., Niyogi, P.: Laplacian Eigen maps for dimensionality reduction and data representation. *Neural Comput.* **15**(6), 1373–1396 (2003)
30. Rigamonti, R., Brown, M., Lepetit, V.: Are sparse representation really relevant for image classification? In: *Proceeding of International Conference on Computer Vision and Pattern Recognition* (2011)
31. Shi, Q., Eriksson, A., Van Den Hengel, A., Shen, C.: Is face recognition really a Compressive Sensing problem? In: *Proceeding of International Conference on Computer Vision and Pattern Recognition* (2011)
32. Zhang, L., Yang, M., Feng, X.: Sparse representation or collaborative representation: which helps face recognition? In: *Proceeding of International Conference on Computer Vision* (2011)
33. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *PAMI* **31**(2), 210–227 (2009)
34. Zhang, H., Yang, J., Zhang, Y., Huang, T.: Close the loop: joint blind image restoration and recognition with sparse representation prior. In: *ICCV* (2011)
35. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using Laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328–340 (2005)
36. Yan, S., Dong, X., Zhang, B., Zhang, H., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 40–51 (2007)
37. Lanckriet, G.R.G., Cristianini, N., Bartlett, P., Ghaoui, L.E., Jordan, M.I.: Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.* **5**(1), 27–72 (2004)
38. Lee, W.-J., Verzakov, S., Duin, R.P.W.: Kernel combination versus classifier combination. In: Haindl, M., Kittler, J., Roli, F. (eds.) *MCS 2007*. LNCS, vol. 4472, pp. 22–31. Springer, Heidelberg (2007). doi:[10.1007/978-3-540-72523-7_3](https://doi.org/10.1007/978-3-540-72523-7_3)
39. Tipping, M.E.: The relevance vector machine. In: *Proceeding of Advances in Neural Information Processing Systems*, vol. 12, pp. 652–658 (2000)

Smart Health

International Conference, ICSH 2016, Haikou, China,

December 24-25, 2016, Revised Selected Papers

Li, X.; Zhang, Y.; Liang, Y. (Eds.)

2017, XI, 252 p. 123 illus., Softcover

ISBN: 978-3-319-59857-4