Regularized Multi-source Matrix Factorization for Diagnosis of Alzheimer's Disease

Xiaofan Que[1], Yazhou Ren[1], Jiayu Zhou[2], and Zenglin Xu[1(✉)]

[1] SMILE Lab, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
xfque@std.uestc.edu.cn, {yazhou.ren,zlxu}@uestc.edu.cn
[2] Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
jiayuz@msu.edu

Abstract. In many real-world systems with multiple sources of data, data are often missing in a block-wise way. For example, in the diagnosis of Alzheimer's disease, doctors may collect patients data from MRI images, PET images and CSF tests, while some patients may have done the MRI scan and the PET scan only, while other patients may have done the MRI scan and the CSF test only. Despite various data imputation technologies exist, in general, they neglect the correlation among multi-sources of data and thus may lead to sub-optimal performances. In this paper, we propose a model called regularized multi-source matrix factorization (RMSMF) to alleviate this problem. Specifically, to model the correlation among data sources, RMSMF firstly uses non-negative matrix factorization to factorize the observed multi-source data into the product of subject factors and feature factors. In this process, we assume different subjects from the same data source share the same feature factors. Furthermore, similarity constraints are forced on different subject factors by assuming for the same subject, the subject factors are similar among all sources. Moreover, self-paced learning with soft weighting strategy is applied to reduce the negative influence of noise data and to further enhance the performance of RMSMF. We apply our model on the diagnosis of the Alzheimer's disease. Experimental results on the ADNI data set have demonstrated its effectiveness.

Keywords: Multi-source neuroimage data · Matrix factorization · Alzheimer's disease · Self-paced learning

1 Introduction

In many real-world systems with multiple sources of data, data are often missing in a block-wise way. For example, in the diagnosis in Alzheimer's disease, doctors may collect patients data from MRI images, PET images and CSF tests. According to the study in Alzheimer's Disease Neuroimaging Initiative (ADNI)[1], over

[1] http://adni.loni.usc.edu/.

D. Liu et al. (Eds.): ICONIP 2017, Part I, LNCS 10634, pp. 463–473, 2017.
https://doi.org/10.1007/978-3-319-70087-8_49

half of the subjects do not have CSF measurement, another half of the subjects' PET scan are absent [20]. The reasons of data missing are diverse, including financial factors, data quality problems, subject's personal privacy and data lost during collection. An illustration can be found in Fig. 1. Here the row means subjects and the column means features. Three different colours represent three data source respectively: CSF, MRI, and PET. The blank space means the subjects' corresponding data source is missing.

In the past few decades, missing data completion methods have been widely used in many scientific research areas. For instances, in biological researches, missing value estimation methods such as weighted K-nearest neighbors (KNN) and singular value decomposition (SVD) are carried out to obtain a complete matrix of gene expression microarrays for further analysis [17]. In recommender systems, matrix factorization technologies are used to generate product recommendations which recommend new items to strange users [9]. The main idea of SVD method is that it assumes the real value matrix can be described by the inner product of the following three matrices: a diagonal matrix composed by a certain number of original matrix's largest singular values and two orthogonal matrices corresponding to its right and left singular vectors. To approximate the original matrix, firstly, row average should be assigned to missing values, then use expectation maximization method to obtain estimation [17]. The paradigm of KNN method is to select the k closest values and compute their weighted average as the estimation of the specific missing value [17]. Matrix factorization methods typically map subject and data source to a latent space demonstrated by factors with a certain dimension, then use the inner product of those two latent space vectors as the reconstructed matrix [1].

In order to cope with the problem of data block-wise missing, Yuan et al. proposed an incomplete Multi-Source Feature learning method (iMSF) which uses a multi-task method to avoid the direct missing data completion that may

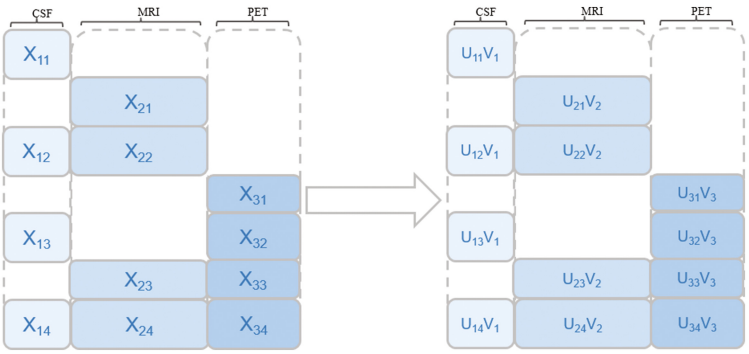

Fig. 1. Data reconstruction by MSMF. Here X_{ij} denotes the i-th data source/view of the j-th group of patients, V_i denotes the factor matrix of the i-th view, and U_{ij} denotes the patient factors of the j-th group of patients under the i-th view.

involve unnecessary noise [20]. After that, Xiang et al. proposed the incomplete source-feature selection method (iSFS) which uses the same strategy to group different data sources but with data overlap in different combinations. Upon different combinations of data source, the iSFS framework learns independent models which are integrated by regularization term [18].

Nonetheless, the original data used in iMSF may contain some noise which is inevitable during data collection process. Inspired by the data fusion idea, we propose to use matrix factorization method to reconstruct data and reduce noise appropriately. In addition, iMSF treats all the subjects equally, which clearly does not take subject diversity into consideration. Therefore, we introduce the self-paced learning (SPL) method [10] with soft weighting to do sample re-weighting. The SPL framework can simultaneously select relatively easy samples and learn a new parameter vector at each iteration. Furthermore, instead of hard weighting technique used in the original SPL model [10,16], soft weighting strategy is used to re-weight subjects using real values rather than just 0 and 1 [4,5,15].

This paper is organized as the following. In Sect. 2, we provide a brief overview to the Alzheimer's disease. In Sect. 3, we present the multi-source matrix factorization approach, followed by the self-paced version in Sect. 4. Finally, we present the real-world study in AD followed by the conclusion in Sect. 6.

2 Related Work on AD Study

AD is a gradually progressing brain disorder, whose exact cause remains undiscovered. The typical symptoms of AD include memory loss, personality and behavior changes, and cognition impairment. According to the National Institute on Aging (NIA), the morbidity of people over 65 years old increases along with aging, which results in nearly half of over age 85 people's illness. What's worse, an AD patient can only live eight years after the occurrence of symptoms in average [2].

Mild Cognitive Impairment (MCI) also causes a decline in cognitive ability and memory loss [3]. Although such degeneracy is not serious enough to influence daily activities, a person with MCI might develop into AD within a few years. Consequently, to distinguish between AD and MCI patients is very important for physicians to make correct decisions of medical treatments. Despite the accurate cause of AD has not been discovered, there are some anomalies of AD patients worth noticing.

Firstly, researchers found that an excess of two abnormal structures in the brain related to AD: plaques and tangles. Their main components beta-amyloid and tau proteins can be detected by cerebrospinal fluid.

Secondly, advanced neuroimaging technologies, including Magnetic Resonance Imaging (MRI) and Positron-Emission Tomography (PET) are able to detect patients's brain structure and function deviations. Specifically, MRI measures brain structures changes related to AD, such as the reduction in hippocampal volumes, the enlargement of the temporal horns and the third and lateral

ventricles [6,7]. The PET scanning is able to examine brain functions including cerebral blood flow, metabolism, and receptor binding. Besides, these information collected from different sources including brain structures, brain functions and protein level variations provide researchers diversified views in classification among the AD patients, MCI patients and normal controls (NC).

Thirdly, the association study between genetic variations and brain imaging biomarks has been conducted to reveal the genetic influences of some specific genes to regions of brains [19,22,23].

3 Regularized Multi-source Matrix Factorization

The multi-source AD data provides multiple views to analyze a patient's condition, and the shared information is of great benefit for disease diagnosis. To deal with the data block-wise missing problem, iMSF method first groups the block-wise missing data into different tasks by subjects' existence of different data sources. In Fig. 1, the three different data sources can be divided into seven different tasks, each task has one of the following data sources combinations: (1) CSF, (2) MRI, (3) CSF and MRI, (4) PET, (5) CSF and PET, (6) MRI and PET, (7) CSF, MRI and PET. Then iMSF uses multi-task method and the regularization term l_{21}-norm to achieve the goal of information sharing. However, data noise is inevitable during data collection, which may lead to data distortion and inaccurate results. For this reason, we propose the multi-source matrix factorization method to do data fusion and reduce the noise to some extent.

Assuming the number of data sources is S, each data source has B_s blocks, $s = 1, 2, ..., S$. We suggest that every block of same data source should have more strong constraints. According to the Non-negative Matrix Factorization (NMF) method [11,12], a non-negative matrix X can be approximated by the inner product of two non-negative matrices U and V:

$$X \approx UV, \tag{1}$$

where $X \in R^{n \times m}$, $U \in R^{n \times r}$, $V \in R^{r \times m}$.

If every data block X_{si} (the s-th data source's i-th block) in the matrix (Fig. 1) can be approximated by two non-negative matrices: U_{si}, V_{si}, that is, $X_{si} = U_{si}V_{si}$, for $s = 1, ..., S$, $i = 1, ..., B_s$, it is quite obvious that we can use U_{si} and V_{si} to describe the subject factors and feature factors of X_{si} respectively. To force more strong constraint on share information, we assume that the same data source's different blocks' V matrices are exactly identical. That is, for every data block belongs to the s-th data source shares the same V_s. Under this assumption, we propose to do multi-source matrix factorization as follows:

$$\min_{U,V} \sum_{s=1}^{S} \sum_{i=1}^{B_s} ||X_{si} - U_{si}V_s||_F^2. \tag{2}$$

In order to simplify our calculating process, we update U_{si}, V_s by the following equations:

$$U_{si(k,l)}^{t+1} = U_{si(k,l)}^{t} \frac{(X_{si}V_s^T)_{(k,l)}}{(U_{si}^t V_s V_s^T)_{(k,l)}}, \tag{3}$$

$$V_{s(l,\mu)}^{t+1} = V_{s(l,\mu)}^{t} \frac{(U_{si}^T X_{si})_{(l,\mu)}}{(U_{si}^T U_{si} V_s^t)_{(l,\mu)}}. \tag{4}$$

where $X \in R^{n\times m}$, $U \in R^{n\times r}$, $V \in R^{r\times m}$, the subscript (k,l) means the k-th row and l-th column of U_{si}, (l,μ) means the l-th row and μ-th column of V_s.

The pseudo-code of MSMF is shown in Algorithm 1. In our experiments, ϵ is set to 10^{-3} and MaxIter (the maximum number of iterations) is set to 20000. The process of MSMF is demonstrated in Fig. 1 where every block in the matrix is approximated by the method.

Algorithm 1. Multi-source matrix factorization (MSMF)

Require: $X \geq 0$
Ensure: U_{si}, V_s, for $i = 1, 2, ..., B_s$, $s = 1, 2, ..., S$
  Initialize $U_{si} \geq 0, V_s \geq 0$, for $i = 1, 2, ..., B_s$, $s = 1, 2, ..., S$
  while $obj \geq \epsilon$ or iter $\leq$ MaxIter **do**
    for $s = 1 : S$ **do**
      for $i = 1 : B_s$ **do**
        Fix V_s, update U_{si} according to (3)
        Fix U_{si}, update V_s according to (4)
      end for
    end for
  end while

Furthermore, we introduce similarity constraints between the same subjects' different data sources. As illustrated in Fig. 1, we assume that U_{12} is similar to U_{22}, U_{13} is similar to U_{32}, and so on. The assumption is based on the fact that X_{12} and X_{22} are two different data sources of the same group of patients. Therefore, our objective function becomes:

$$\min_{U,V} \sum_{s=1}^{S} \sum_{i=1}^{B_s} ||X_{si} - U_{si} V_s||_F^2, \tag{5}$$

$$\text{s.t.} \begin{cases} U_{12} \sim U_{22}, \\ U_{13} \sim U_{32}, \\ U_{23} \sim U_{33}, \\ U_{14} \sim U_{24}, \\ U_{24} \sim U_{34}, \\ U_{si} \geq 0, V_s \geq 0. \end{cases}$$

The symbol "$\sim$" means the two matrices are similar. For clarification, the definition of similar is not the same as the similar matrix in mathematics, and actually, it is measured by the Frobenius norm of matrices. Assume that the number of similarity constraints is C, and the similar matrix of $U_j^{(1)}$ is $U_j^{(2)}$, for instance,

in the constraint $U_{12} \sim U_{22}$, $U_j^{(1)} = U_{12}$, $U_j^{(2)} = U_{22}$. As a result, the regularized multi-source matrix factorization (RMSMF) can be formulated as follows:

$$\min_{U,V} \sum_{s=1}^{S} \sum_{i=1}^{B_s} ||X_{si} - U_{si}V_s||_F^2 + \lambda \sum_{j=1}^{C} ||U_j^{(1)} - U_j^{(2)}||_F^2,$$
$$\text{s.t. } U_{si} \geq 0, V_s \geq 0. \tag{6}$$

where the parameter λ controls the significance of similarity constraints.

Similarly, we can compute the update rules for U_{si}, V_s as follows [11,12]:

$$U_{si(k,l)}^{(1)t+1} = U_{si(k,l)}^{(1)t} \frac{(X_{si}V_s^T + \lambda U_{si}^{(2)})_{(k,l)}}{(U_{si}^{(1)t}V_sV_s^T + \lambda U_{si}^{(1)t})_{(k,l)}}, \tag{7}$$

$$V_{s(l,\mu)}^{t+1} = V_{s(l,\mu)}^{t} \frac{(U_{si}^T X_{si})_{(l,\mu)}}{(U_{si}^T U_{si} V_s^t)_{(l,\mu)}}. \tag{8}$$

4 Self-paced Classification

In order to demonstrate the advantage of our reconstructed data, we propose to use the identical classification model on both the reconstructed data and the observed data. As we mentioned before, iMSF uses a multi-task method to learn a model parameter. First of all, we should briefly go through the structure of iMSF. Assuming that the data set is divided into m tasks, the i-th task with totally N_i subjects is described as: $T^i = \{x_j^i, y_j^i\}$, $j = 1...N_i$, in which $\{x_j^i, y_j^i\}$ is the corresponding feature matrix and label of the j-th subject in i-th task. The iMSF framework is formulated as follows:

$$\min_{\beta} \frac{1}{m} \sum_{i=1}^{m} \frac{1}{N_i} \sum_{j=1}^{N_i} L(x_j^i, y_j^i, \beta^i) + \lambda \sum_{s=1}^{S} \sum_{k=1}^{p_s} \|\beta_{I(s,k)}\|_2. \tag{9}$$

where $L(\cdot)$ is the loss function and logistic loss is employed in this study, β^i is the model parameter of i-th task. The second part of this formulation is the $l_{2,1}$-norm regularization [21]. S is the total number of data source and P_s is the total number of s-th data source's feature dimension. $I(s,k)$ is a index function, $\beta_{I(s,k)}$ indicates the parameter of k-th feature in s-th data source. Please refer to [13,14,20] for more details of iMSF model and the optimization method Accelerated Gradient Descent (AGD).

As we mentioned above, iMSF treats all the subjects fairly. But, the data of some subjects might be inaccurate or with noise. This can negatively affect its performance. In this work, we use SPL with soft weighing strategy to address this issue. Concretely, according to [10], the original SPL framework with hard weighting simply assigns 0 or 1 to a subject which means the subject is not selected or selected. In order to treat different subjects accordingly, we propose to use the SPL framework with soft weighting [4] to assign real-valued weights to different subjects. In this paper, we adopt the linear soft weighting.

Algorithm 2. Self-paced learning framework with soft weighting strategy

Require: initial value: $\beta^i,\ K^i$, for $i = 1, ..., m$, learning rate μ
Ensure: β^i, for $i = 1, 2, ..., m$
  Initialize $K^i \leftarrow K_0^i$, for $i = 1, 2, ..., m$
  while $\exists v_j^i = 0, \forall i,\ j$ **do**
    Fix v_j^i, update β^i, using AGD
    Fix β^i, compute loss L_j^i, update v_j^i by (11)
    $K^i \leftarrow \frac{K^i}{\mu}$, for $i = 1, ..., m$
  end while

Linear Soft Weighting:

$$\min_{\beta,v} \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i} v_j^i L(x_j^i, y_j^i, \beta^i) + \lambda\sum_{s=1}^{S}\sum_{k=1}^{P_s}\|\beta_{I(s,k)}\|_2 + \sum_{i=1}^{m} K^i(\frac{1}{2}\|v^i\|_1^2 - \sum_{j=1}^{N_i} v_j^i), \tag{10}$$

where K^i is the SPL parameter for the i-th task. For simplicity, we use L_j^i to denote $L(x_j^i, y_j^i, \beta^i)$. Thus, for fixed β^i,

$$v_j^i = \begin{cases} -\frac{L_j^i}{K^i} + 1, & L_j^i < K^i, \\ 0, & otherwise. \end{cases} \tag{11}$$

This solution means that in a specific task i, if a subject j's loss is less than the threshold K^i, this subject is defined to be easy and assigned a real-valued weight, otherwise, it will be neglected until next iteration. As K^i grows, more and more subjects' loss would be lower than the threshold K^i, so that more subjects will be selected to train a model with better performance. Apparently, according to the linear soft weighting strategy, the noisy subjects which are typically with large loss will be assigned small weight. In this way, the negative influence of noisy data can be reduced to some extent. The pseudo-code of self-paced learning framework is described in Algorithm 2.

5 Experimental Results

In this section, we are going to present the data used in the experiments to validate our algorithm and give convincing suggestions to AD diagnosis. Moreover, adequate experiments are conducted to evaluate the performances of our algorithm and the comparing methods.

5.1 Data

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database provides the data used in our experiments [20]. This data set is consisted of 742 subjects with three data sources: MRI, PET, and CSF. The labels of subjects are diagnosed by physicians seven times within four years by three types: AD, MCI and

Table 1. Description of the ADNI data set

Data source	#features	#AD subject	#Non subject	#Con subject	#NC subject
CSF	3	103	122	85	105
MRI	305	392	189	142	178
PET	116	77	105	70	75

NC (some diagnosis results are missing). According to the change over diagnosis results, we group all the subjects into four different classes: AD, Converter (Con), Non-converter (Non), and Normal Control (NC). The labels of AD and NC are quite explicit. Those who are firstly diagnosed as MCI and then gradually transform into AD are defined as converters, while those subjects are defined as non-converters if they remain MCI till the diagnosis period ends. According to this definition, the detailed description of the data set is given by Table 1. For classification and practical usages, we focus on subjects that are difficult to distinguish: AD vs Non, Con vs Non, and Non vs NC. We believe that the challenging subject classification can provide more valuable suggestions on clinical diagnosis and help physicians to establish distinctive therapeutic schedule on different patients as soon as possible and avoid delaying optimal treatment period.

5.2 Results and Analysis

In order to demonstrate the effectiveness of our methods, we use two baseline methods: the original iMSF method and the traditional matrix completion method SVD. The comparing methods used in the experiments are stated as follows.

- **iMSF-obs** [20,24]: as mentioned above, iMSF is a novel multi-task method that adopts an innovative task construct method to avoid incomplete data. We are going to perform the original iMSF model on observed data.
- **SVD** [8]: singular value decomposition (SVD) is presented as an representative of traditional matrix completion method. SVD is a low rank approximation matrix completion method. We firstly initialize the missing data entries as zeros and then apply SVD to obtain a complete matrix.
- **MSMF**: we firstly use the proposed MSMF method to reconstruct data and then perform iMSF on the reconstructed data.
- **MSMF-S**: in this method, we first use MSMF to reconstruct data, then use iMSF model with SPL soft weighting strategy to do classification.
- **RMSMF**: this method is similar to MSMF method except that the RMSMF method is used to reconstruct data.
- **RMSMF-S**: this method is similar to MSMF-S method except that the RMSMF method is used to reconstruct data.

The experiments of all methods adopt a 10-fold cross validation scheme. The searching range of λ is set to [0.01 0.03 0.1 0.3 1] for all iMSF relevant methods.

Table 2. AUC results of the comparing methods

Method	iMSF-obs	SVD	MSMF	MSMF-S	RMSMF	RMSMF-S
AD vs Non	0.8052	0.7152	0.8118	0.8287	0.8192	**0.8314**
Con vs Non	0.7253	0.6211	0.7279	**0.7415**	0.7319	0.7323
Non vs NC	0.6453	0.6422	0.6765	**0.6818**	0.6503	0.6503

The initial SPL parameter K is set by the median of initial loss and the learning pace is set by limiting the iteration times within 5. The AUC results of all the aforementioned methods are presented in Table 2.

As reported by Table 2, it can be observed that the iMSF method defeats the SVD method, which suggests that the simple matrix completion has poor performance dealing with data block-wise missing. Accordingly, to compare our method with the iMSF method is fairly reasonable. Specifically, we can find that MSMF always outperforms iMSF-obs in the three classification tasks, indicating the effectiveness of reconstructed data. Moreover, RMSMF performs better than MSMF in the classifications of AD vs Non, Con vs Non. Hence, we can draw a conclusion that the MSMF model with constraints do improve the performance of MSMF data integration to some extent. In the results of SPL soft weighting models, on one hand, the increments between MSMF, RMSMF and MSMF-S, RMSMF-S are stable, which means that the soft weighting strategy have reduced the influence of the noisy data in the reconstructed MSMF and RMSMF data effectively. On the other hand, although MSMF-S have come out with the best auc results in two classifications, RMSMF-S still performed excellent in the classification of AD vs Non, which is of great help in diagnosis of AD, since it is crucial to take different medical treatments for AD and Non patients.

6 Conclusions

To alleviate the block-wise missing data problem in the diagnosis of Alzheimer's disease, we propose a novel multi-source matrix factorization method. To further improve the performance, we adopt self-paced learning with soft weighting strategy to the factorization model. The integrated model can not only effectively utilize multiple sources of data, but also reduce the influence of noisy data. The effectiveness of the proposed model is empirically verified on the ADNI data set.

Acknowledgments. This paper was in part supported by Grants from the Natural Science Foundation of China (No. 61572111), the National High Technology Research and Development Program of China (863 Program) (No. 2015AA015408), a 985 Project of UESTC (No. A1098531023601041), a Project funded by China Postdoctoral Science Foundation (No. 2016M602674), and two Fundamental Research Funds for the Central Universities of China (Nos. ZYGX2016J078 and ZYGX2016Z003).

References

1. Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., Plemmons, R.J.: Algorithms and applications for approximate nonnegative matrix factorization. Comput. Stat. Data Anal. **52**(1), 155–173 (2007)
2. Bren, L.: Alzheimer's: searching for a cure. http://www.mamashealth.com/senior/alz4.asp
3. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al.: Mild cognitive impairment. Lancet **367**(9518), 1262–1270 (2006)
4. Jiang, L., Meng, D., Mitamura, T., Hauptmann, A.G.: Easy samples first: self-paced reranking for zero-example multimedia search. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 547–556. ACM (2014)
5. Jiang, L., Meng, D., Zhao, Q., Shan, S., Hauptmann, A.G.: Self-paced curriculum learning. In: Association for the Advancement of Artificial Intelligence, vol. 2, p. 6 (2015)
6. Killiany, R., Hyman, B., Gomez-Isla, T., Moss, M., Kikinis, R., Jolesz, F., Tanzi, R., Jones, K., Albert, M.: MRI measures of entorhinal cortex vs hippocampus in preclinical ad. Neurology **58**(8), 1188–1196 (2002)
7. Killiany, R.J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., Tanzi, R., Jones, K., Hyman, B.T., Albert, M.S.: Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. Ann. Neurol. **47**(4), 430–439 (2000)
8. Klema, V., Laub, A.: The singular value decomposition: its computation and some applications. IEEE Trans. Autom. Control **25**(2), 164–176 (1980)
9. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
10. Kumar, M.P., Packer, B., Koller, D.: Self-paced learning for latent variable models. In: Advances in Neural Information Processing Systems, pp. 1189–1197 (2010)
11. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788 (1999)
12. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems, pp. 556–562 (2001)
13. Liu, J., Ji, S., Ye, J.: Multi-task feature learning via efficient l_2, 1-norm minimization. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, pp. 339–348. AUAI Press (2009)
14. Nesterov, Y., et al.: Gradient methods for minimizing composite objective function (2007)
15. Ren, Y., Zhao, P., Sheng, Y., Yao, D., Xu, Z.: Robust softmax regression for multi-class classification with self-paced learning. In: International Joint Conference on Artificial Intelligence, pp. 2641–2647 (2009)
16. Ren, Y., Zhao, P., Xu, Z., Yao, D.: Balanced self-paced learning with feature corruption. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 2064–2071. IEEE (2017)
17. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B.: Missing value estimation methods for dna microarrays. Bioinformatics **17**(6), 520–525 (2001)
18. Xiang, S., Yuan, L., Fan, W., Wang, Y., Thompson, P.M., Ye, J.: Multi-source learning with block-wise missing data for Alzheimer's disease prediction. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 185–193. ACM (2013)

19. Xu, Z., Zhe, S., Qi, Y., Yu, P.: Association discovery and diagnosis of Alzheimer's disease with Bayesian multiview learning. J. Artif. Intell. Res. (JAIR) **56**, 247–268 (2016)
20. Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J., Initiative, A.D.N., et al.: Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. NeuroImage **61**(3), 622–632 (2012)
21. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **68**(1), 49–67 (2006)
22. Zhe, S., Xu, Z., Qi, Y., Yu, P.: Joint association discovery and diagnosis of Alzheimer's disease by supervised heterogeneous multiview learning. In: Pacific Symposium on Biocomputing, pp. 300–311 (2013)
23. Zhe, S., Xu, Z., Qi, Y., Yu, P.: Sparse Bayesian multiview learning for simultaneous association discovery and diagnosis of Alzheimer's disease. In: Association for the Advancement of Artificial Intelligence, pp. 1966–1972 (2015)
24. Zhou, J., Chen, J., Ye, J.: Malsar: multi-task learning via structural regularization. Arizona State University 21 (2011)