Predicting Clinical Scores Using Semi-supervised Multimodal Relevance Vector Regression

Bo Cheng[1], Daoqiang Zhang[1,2], Songcan Chen[1], and Dinggang Shen[2]

[1] Dept. of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[2] Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599
{cb729,dqzhang,s.chen}@nuaa.edu.cn, dgshen@med.unc.edu

Abstract. We present a novel semi-supervised multimodal relevance vector regression (SM-RVR) method for predicting clinical scores of neurological diseases from multimodal brain images, to help evaluate pathological stage and predict future progression of diseases, e.g., Alzheimer's diseases (AD). Different from most existing methods, we predict clinical scores from multimodal (imaging and biological) biomarkers, including MRI, FDG-PET, and CSF. Also, since mild cognitive impairment (MCI) subjects generally contain more noises in their clinical scores compared to AD and healthy control (HC) subjects, we use only their multimodal data (i.e., MRI, FDG-PET and CSF), not their clinical scores, to train a semi-supervised model for enhancing the estimation of clinical scores for AD and healthy control (HC). Experimental results on ADNI dataset validate the efficacy of the proposed method.

1 Introduction

Many pattern classification methods have been proposed for the diagnosis of Alzheimer's disease (AD) or its prodromal stage, i.e., mild cognitive impairment (MCI). Recently, people have started to investigate applying the pattern regression methods for estimating the continuous clinical scores of subjects from their respective brain images [1,2,4]. This kind of study is important because it can help evaluate the pathological stage and predict the future progression of neurological diseases. It is known that many diseases present a continuous spectrum of structural and functional changes. For example, AD pathology is known to progress gradually over many years, sometimes starting decades before a final clinical stage [2]. Thus, pattern regression methods can be used to help estimate the continuous clinical scores to evaluate the disease stage of MCI or AD, rather than simple categorical classification.

Recent studies have also demonstrated that the biomarkers from different modalities can provide complementary information for diagnosis of AD [5,8,9]. Accordingly, two or more modalities of biomarkers have been combined for multimodal classification [3,6,7,14]. However, to the best of our knowledge, there exist few related works which combine two or more biomarkers from multimodal data for regression. Instead, most existing methods on estimating clinical scores use only single modality of data [1,2,4]. In this paper, we will employ multiple-kernel combination method to combine multimodality data, e.g., MRI, PET, and CSF, for multimodal regression.

K. Suzuki et al. (Eds.): MLMI 2011, LNCS 7009, pp. 241–248, 2011.

On the other hand, at present, several works have adopted the supervised relevance vector machine regression (RVR) method to estimate continuous clinical scores [1,2,4]. It is known that the training of a supervised model often requires many well-labeled data in order to achieve a good performance. However, the clinical scores in cognitive tests such as Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale-Cognitive subtest (ADAS-Cog) are usually very noisy, especially for the MCI subjects who may or may not convert to AD within a period of follow-up time. To partially alleviate this problem, we will not exploit the clinical scores of the MCI subjects and instead use only their corresponding multimodal data (i.e., MRI, FDG-PET, and CSF) to help train a semi-supervised regression model. It is worth noting that similar idea, which treats MCI subjects as unlabeled data to train a semi-supervised *classification* model, has also been used for classification of AD [13,15]. However, to our knowledge, no previous studies have ever investigated the semi-supervised *regression* in AD research.

In this paper, we propose a semi-supervised multimodal relevance vector regression (SM-RVR) model to predict clinical scores based on both imaging and biological biomarkers. We further construct a graph Laplacian matrix based on the manifold learning theory [11] to select the most informative MCI subjects for better helping semi-supervised regression.

2 Method

In this section, we will first extend the standard relevance vector regression (RVR) method to the multimodal RVR (M-RVR), and then introduce our proposed semi-supervised multimodal RVR (SM-RVR) method.

2.1 Multimodal RVR (M-RVR)

We first briefly review the standard RVR algorithm. The main idea of RVR is summarized as follows. Specifically, RVR is a sparse kernel method formulated in a Bayesian framework [12]. Given a training set with its corresponding target values, such as $\{x_n, t_n\}_{n=1}^{N}$, RVR aims to find out the relationship between the input feature vector x_n and its corresponding target value t_n:

$$t_n = f(x_n, w) + \varepsilon_n \tag{1}$$

where ε_n is the measurement noise (assumed independent and following a zero-mean Gaussian distribution, $\varepsilon_n \sim N(0, \sigma^2)$, and $f(x_n, w)$ is a linear combination of basis functions $k(x, x_n)$ with the following form:

$$f(x, w) = \sum_{n=1}^{N} w_n k(x, x_n) + w_0 \tag{2}$$

Where $w = (w_0, w_1, \ldots, w_N)^T$ is a weight vector, $K_{N\times(N+1)}$ is the 'design' matrix with $K_{ij} = k(x_i, x_j), i, j = 1, \ldots, N$, and $k(x_i, x_0) = 1$. According to [12], we can obtain a sparse kernel regression model based on the weight vector w.

Now we can extend RVR to multimodal RVR (M-RVR) for multimodal regression, by defining a new integrated kernel function for comparison of two multimodal data x and x_n as below:

$$k(x, x_n) = \sum_{m=1}^{M} c_m k^{(m)}\left(x^{(m)}, x_n^m\right) \tag{3}$$

Where $k^{(m)}$ denotes the kernel matrix over the *m*-th modality, similar to the definition given for the single modality case. This new integrated multiple-kernel can be expediently embedded into the conventional single-kernel RVR, and thus solved by the programs developed for the conventional single-kernel RVR. Here, we constrain the sum of c_m to be 1 and adopt a coarse-grid search through cross-validation on the training samples to find their optimal values.

2.2 Semi-supervised Multimodal RVR (SM-RVR)

Fig. 1 shows the flowchart of our proposed semi-supervised multimodal RVR (SM-RVR) method for multimodal regression. Here, the main idea is to select the most informative MCI subjects as unlabeled samples to aid the regression of AD and healthy controls (which are used as labeled samples). The algorithmic procedure of SM-RVR is detailed as below:

Step 1: Given a labeled sample set $\boldsymbol{L}$ and an unlabeled sample set $\boldsymbol{U}$, initialize various parameters, including the maximum number of iterations T, RVR kernel function type, the kernel width parameter σ, and the number of the nearest neighbors k used in KNN algorithm (as detailed below);

Step 2: Randomly select n unlabeled samples from the unlabeled sample set $\boldsymbol{U}$ to constitute a new unlabeled sample set $\boldsymbol{U'}$, and then combine $\boldsymbol{U'}$ with $\boldsymbol{L}$ to constitute a new sample set $\boldsymbol{[L; U']}$ and further compute the respective graph Laplacian matrix $\boldsymbol{l}$, where according to manifold learning theory [10,11], local smoothness is assumed between all (labeled and unlabeled) samples. From the graph Laplacian matrix $\boldsymbol{l}$, select the top samples (from $\boldsymbol{U'}$) with the minimum distances to the labeled samples in $\boldsymbol{L}$, as the candidate sample set $\boldsymbol{W}$;

Step 3: For each sample x_j in $\boldsymbol{W}$, find its k-nearest neighbors in $\boldsymbol{L}$, compute the mean of clinical values of those k neighbors as its estimated clinical score y_j, and train M-RVR on $\boldsymbol{L}$ plus $\{(x_j, y_j)\}$. Then, we compute the value of square root of mean square error (RMSE) for each sample x_j in $\boldsymbol{W}$. Finally, a sample with the top confidence (i.e., minimum RMSE value) in $\boldsymbol{W}$ is selected and added into $\boldsymbol{L}$, and further deleted from $\boldsymbol{U}$.

Step 4: Go to Step 2 for running the next iteration;

Step 5: After reaching the total number of iterations T, train M-RVR on the latest $\boldsymbol{L}$ to build a final regression model.

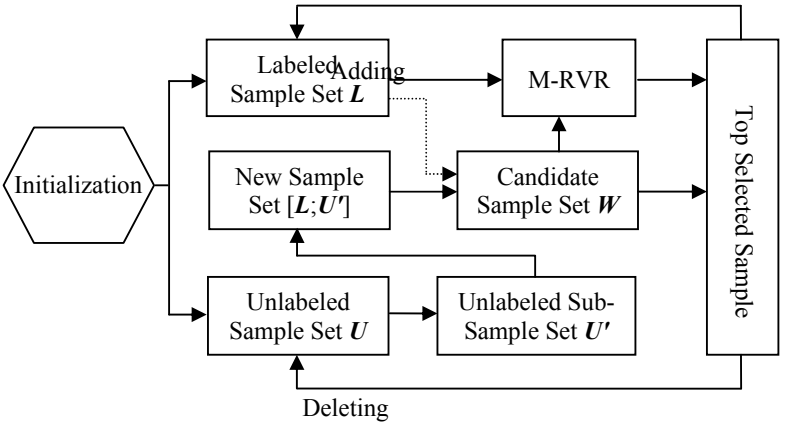

Fig. 1. The flowchart of the proposed SM-RVR method

3 Results

3.1 Subjects

In this paper, the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset is used to test our semi-supervised regression method. Only the baseline ADNI subjects with all corresponding MRI, PET, and CSF data are included, thus leading to a total of 202 subjects (including 51 AD patients, 99 MCI patients, and 52 healthy controls (HC)). Table 1 lists the demographics of these subjects.

The same image pre-processing as used in [3] is adopted here. First, for all structural MR images, we correct their intensity inhomogeneity by the N3 algorithm, do skull-stripping, and remove cerebellum. Then, we use the FSL package to segment each structural MR image into three different tissues: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). We further use an atlas warping algorithm [16] to partition each structural MR brain image into 93 ROIs. For each of the 93 ROIs, we compute GM volume in that ROI as a feature; For PET image, we use a rigid transformation to align it to its respective structural MR image of the same subject, and then compute the average PET value of each ROI as a feature. Accordingly, for each subject, we can acquire 93 features from the structural MR image, another 93 features from the PET image, and 3 features from the CSF biomarkers.

Table 1. Subject information (mean± std)

	AD	MCI	HC
Age	75.2±7.4	75.3±7.0	75.3±5.2
Education	14.7±3.6	15.9±2.9	15.8±3.2
MMSE	23.8±1.9	27.1±1.7	29.0±1.2
ADAS-Cog	18.3±6.0	11.4±4.4	7.4±3.2

Table 2. Comparison of the regression performance of SM-RVR with respect to different combinations of MRI, PET and CSF modalities

Modality	MMSE		ADAS-Cog	
	RMSE	CORR	RMSE	CORR
MRI	2.1711	0.7306	5.1565	0.7004
PET	2.4613	0.6178	5.0406	0.7055
CSF	2.4493	0.6001	5.6168	0.6412
MRI+PET	2.0948	0.7551	4.7315	0.7621
MRI+CSF	2.0324	0.7709	4.9821	0.7376
PET+CSF	2.3828	0.6631	4.8907	0.7344
MRI+PET+CSF	1.9187	0.8013	4.4482	0.7823

3.2 Experimental Setup

To evaluate the performance of regression methods, we use both RMSE and correlation coefficient (CORR) [4] as performance measures. The number of the nearest neighbor k in KNN algorithm and the maximum number of iterations T are both learned from the training samples, through an enumeration search using the range from 1 to 50 and 1 to 99, respectively. For the performance evaluation of regression methods, we use a 10-fold cross-validation strategy to compute the average RMSE and CORR measures. The RVM regression learning machine is implemented using Sparse Bayesian toolbox[1], with Gauss kernel and default kernel-width σ. The weights in the M-RVR are learned based on the training samples, through a grid search using the range from 0 to 1 at a step size of 0.1. Also, for each modality feature f_i in labeled samples and unlabeled samples, the same feature normalization scheme as used in [3] is adopted here.

3.3 Experimental Results

Table 2 shows the performance measures (including RMSE and CORR) of our SM-RVR method, using different combinations of MRI, PET and CSF modalities. As we can see from Table 2, the combination of MRI, PET, and CSF can consistently achieve better results than any other methods. Specifically, SM-RVR using all three modalities can achieve a RMSE of 1.9187 and a CORR of 0.8013 for MMSE scores, and a RMSE of 4.4482 and a CORR of 0.7823 for ADAS-Cog scores, as shown in Fig. 2 which gives the scatter plots of actual clinical scores vs. estimated scores. On the other hand, Table 2 also indicates that the use of two modalities can improve the regression performance, although they are inferior to the use of all three modalities together. These results validate the advantage of multimodal regression over the conventional single-modal regression in estimation of clinical scores.

Table 3 shows the comparison of SM-RVR with supervised M-RVR. It is worth noting that, for fair comparison, we implement two versions of M-RVR, i.e., one using only AD and HC subjects as training sample and another using all (AD, HC and MCI) subjects as training samples. As can be seen from Table 3, SM-RVR consistently

[1] `http://www.miketipping.com/index.php?page=rvm`

outperforms M-RVR (including both versions) on each performance measure, which validates the efficacy of our SM-RVR method that uses MCI subjects only as unlabeled samples in a semi-supervised regression framework. Also, from Table 3, it is interesting to note that M-RVR using all subjects achieves slightly better performance in terms of RMSE, but much worse performance in terms of CORR, compared with M-RVR using only AD and HC subjects. This implies that the clinical scores of MCI subjects may contain more noises than those of AD or HC subjects.

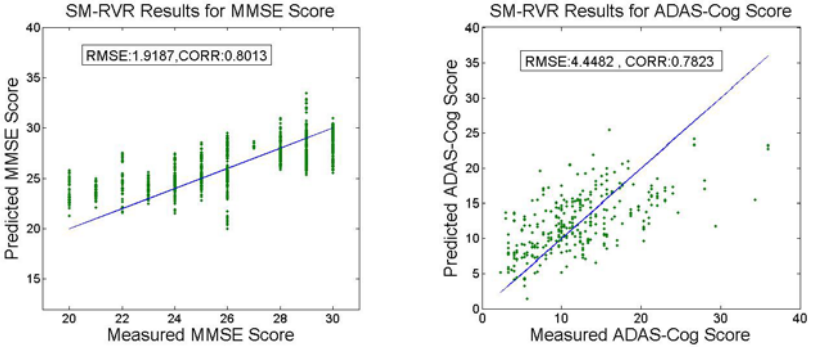

Fig. 2. Scatter plots of actual clinical scores vs. estimated scores for MMSE (left) and ADAS-Cog (right)

Table 3. Comparison of regression performance of SM-RVR and M-RVR

Methods	MMSE		ADAS-Cog	
	RMSE	CORR	RMSE	CORR
M-RVR (51AD+52HC)	2.2159	0.7285	4.9174	0.7325
M-RVR (51AD+52HC+99MCI)	2.1701	0.5261	4.6909	0.6404
SM-RVR(51AD+52HC+99MCI)	1.9187	0.8013	4.4482	0.7823

Finally, in Fig.3, we plot the curves of regression performance measures (of RMSE and CORR) with respect to the different number of unlabeled MCI samples when using different number of nearest neighbors (i.e., k=1, 3, 5) in SM-RVR. As can be seen from Fig. 3, the regression performance of SM-RVR is first steadily improved as the number of unlabeled MCI samples increases and is significantly better than that of M-RVR in most cases, but it declines after reaching a certain value. This implies that using selected MCI subjects as unlabeled samples is superior to using all MCI subjects as unlabeled samples in clinical score estimation.

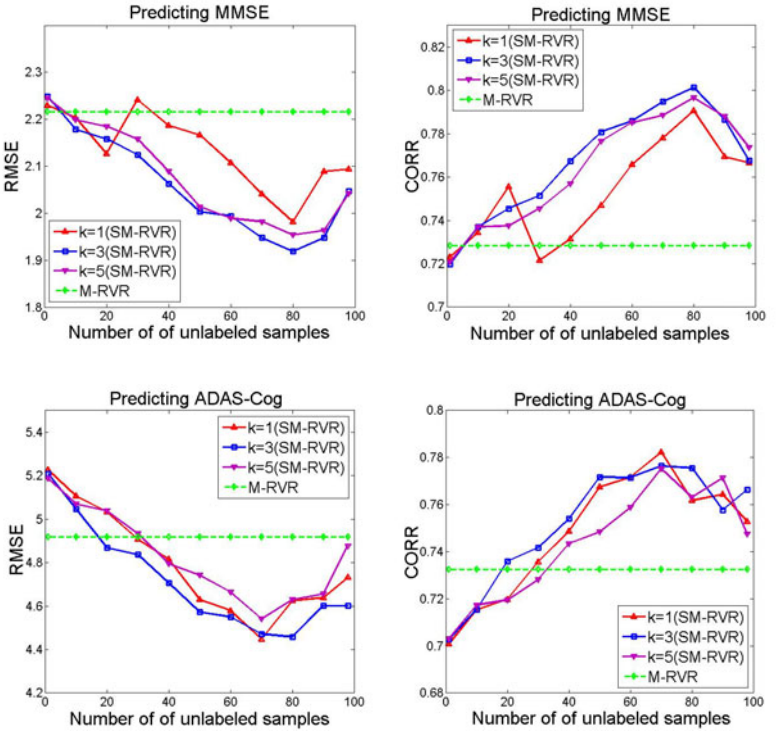

Fig. 3. Plots of regression performance (RMSE and CORR) vs. different number of unlabeled samples when using different number of nearest neighbors (i.e., k=1, 3, 5) in SM-RVR. The number of unlabeled MCI samples is equal to the total number of total iterations T.

4 Conclusion

This paper proposes a novel semi-supervised multimodal regression method, namely SM-RVR, to predict clinical scores of subjects (including AD, HC or MCI) from both imaging and biological biomarkers, i.e., MRI, PET and CSF. Our method assumes that the clinical scores obtained from MCI subjects may contain more noises than AD and HC subjects, and thus should be used in a semi-supervised regression framework as unlabeled data, rather than in a conventional supervised regression framework as labeled data. Furthermore, a scheme for selecting the most informative MCI subjects for helping training regression model is also derived. The experimental results on the ADNI dataset show the efficacy of our proposed method.

Acknowledgments. This work was supported in part by National Science Foundation of China under grant Nos. 60875030, 60973097 and 61035003, and also by NIH grants EB006733 and EB009634.

References

1. Stonnington, C.M., Chu, C., Klöppel, S., Jack, C.R., Ashburner, J., Frackowiak, R.S.J.: ADNI: Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. NeuroImage 51(4), 1405–1413 (2010)
2. Wang, Y., Fan, Y., Bhatt, P., Davatzikos, C.: High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. NeuroImage 50(4), 1519–1535 (2010)
3. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: ADNI: Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55(3), 856–867 (2011)
4. Fan, Y., Kaufer, D., Shen, D.: Joint estimation of multiple clinical variables of neurological diseases from imaging patterns. In: ISBI, pp. 852–855 (2010)
5. Walhovd, K.B., Fjell, A.M., Amlien, I., Grambaite, R., Stenset, V., Bjørnerud, A., Reinvang, I., Gjerstad, L., Cappelen, T., Due-Tønnessen, P., Fladby, T.: Multimodal imaging in mild cognitive impairment: Metabolism, morphometry and diffusion of the temporal–parietal memory network. NeuroImage 45(1), 215–223 (2009)
6. Ye, J., Chen, K., Wu, T., Li, J., Zhao, Z., Patel, R., Bae, M., Janardan, R., Liu, H., Alexander, G., Reiman, E.M.: Heterogeneous data fusion for Alzheimer's disease study. In: KDD 2008, pp. 1025–1033 (2008)
7. Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C.: Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage 41, 277–285 (2008)
8. Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M.: CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. J. Neurosci. 30, 2088–2101 (2010)
9. Mosconi, L., Brys, M., Glodzik-Sobanska, L., De Santi, S., Rusinek, H., de Leon, M.J.: Early detection of Alzheimer's disease using neuroimaging. Exp. Gerontol. 42, 129–138 (2007)
10. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. Journal of Machine Learning Research 7, 2399–2434 (2006)
11. Belkin, M., Niyogi, P.: Semi-Supervised Learning on Riemannian Manifolds. Machine Learning 56, 209–239 (2004)
12. Tipping, M.: Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244 (2001)
13. Filipovych, R., Davatzikos, C.: Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). NeuroImage 55(3), 1109–1119 (2011)
14. Hinrichs, C., Singh, V., Xu, G., Johnson, S.: MKL for robust multi-modality AD classification. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 786–794. Springer, Heidelberg (2009)
15. Zhang, D., Shen, D.: Semi-supervised multimodal classification of Alzheimer's disease. In: ISBI, 1628–1631 (2011)
16. Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging 21, 1421–1439 (2002)