Ensemble Universum SVM Learning for Multimodal Classification of Alzheimer's Disease

Xiaoke Hao and Daoqiang Zhang*

Dept. of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
dqzhang@nuaa.edu.cn

Abstract. Recently, machine learning methods (e.g., support vector machine (SVM)) have received increasing attentions in neuroimaging-based Alzheimer's disease (AD) classification studies. For classifying AD patients from normal controls (NC), standard SVM trains a classification model from only AD and NC subjects. However, in practice besides AD and NC subjects, there may also exist other subjects such as those with mild cognitive impairment (MCI). In this paper, we investigate the potential of using MCI subjects to aid the identification of AD from NC subjects. Specifically, we propose to use the universum support vector machine (U-SVM) learning by treating MCI subjects as the universum examples that do not belong to either of the classes (i.e., AD and NC) of interest. The idea of U-SVM learning is to separate AD from NC subjects through large margin hyperplane with the universum MCI subjects laying inside the margin borders, which is in accordance with our domain knowledge that MCI is a prodromal stage of AD with cognitive status between NC and AD. Furthermore, we propose ensemble universum SVM learning for multimodal classification by training an individual U-SVM classifier for each modality. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database demonstrate the efficacy of our proposed method.

1 Introduction

Alzheimer's disease (AD) is one of the most common forms of dementia in elderly people worldwide. Early diagnosis of AD is very important for possible delay of the disease. Over the past decades many machine learning methods (e.g., support vector machine (SVM)) have been developed for classification of AD or its prodromal stage, i.e., mild cognitive impairment (MCI), based on either single or multiple modalities of biomarkers [1-3]. However, one challenge in AD classification is that the number of AD patients and normal controls (NC) is usually very limited while the dimensionality of imaging data is quite high, which makes it very difficult to train a robust and powerful AD classifier under the standard SVM framework. On the other hand, besides AD and NC subjects, we may have other domain-related subjects such as those with MCI, a prodromal stage of AD. These subjects may help to build a powerful AD classifier, although their cognitive status may not belong to AD or NC.

* Corresponding author.

G. Wu et al. (Eds.): MLMI 2013, LNCS 8184, pp. 227–234, 2013.

To address that problem, in this paper we propose to use a new learning technique called universum support vector machine (U-SVM) which classifies AD from NC subjects with the extra help from MCI subjects. It's noteworthy that in some recent neuroimaging-based classification works, some other machine learning techniques including semi-supervised learning and transfer learning have also been used for enhancing the neuroimaging-based classification with subjects from other domains [4, 5]. However, different from semi-supervised learning and transfer learning which use the auxiliary knowledge from the related subjects or domains, universum learning (including U-SVM) aims at employing universum data as contradiction samples that do not belong to either of the classes (i.e., AD and NC) of interest to boost the generalization performance [6]. Although U-SVM has been successful applied for many other applications such as handwritten digits recognition [7], however, to the best of our knowledge it has not been introduced to neuroimaging-based brain disease classification.

In this paper, we investigate the potential of using U-SVM learning to aid identification of AD from NC subjects with MCI subjects as universum data. The objective of exploiting U-SVM learning is to separate AD from NC subjects through large margin hyperplane with the universum MCI subjects laying inside the margin borders, which is in accordance with our domain knowledge that MCI is a prodromal stage of AD with cognitive status between NC and AD. Furthermore, the ensemble learning classifier is adopted to combine individual U-SVM classifiers trained from each modality of biomarkers, including magnetic resonance imaging (MRI), fluorodeoxyglucose positron emission tomography (FDG-PET) and cerebrospinal fluid (CSF), for multimodal classification [8, 9]. The proposed methods are validated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

2 Method

The U-SVM learning aims at employing a portion of the universum data as a priori knowledge to boost generalization performance [6]. To exploit the potential of using MCI subjects to aid classification between AD and NC subjects, we treat MCI subjects as universum data, and the then adopt U-SVM learning to solve the classification problem. In the following sections, we will first introduce the U-SVM method for single-modality classification, and then present the ensemble U-SVM method for multimodal classification.

2.1 Universum Support Vector Machine (U-SVM)

We first compare U-SVM with SVM through an illustration, as shown in Fig. 1. As can be seen from Fig. 1, compared to the standard SVM with solid lines, in U-SVM the universum samples are constrained to fall inside the margin borders with dashed lines. Here in U-SVM we require the universum samples lie inside the margin borders (dashed lines in Fig. 1), because these samples do not belong to either class. In this way, the universum learning achieves a trade-off between explaining training samples using large margin hyperplanes and maximizing the number of contradictions on the universum [7] [10].

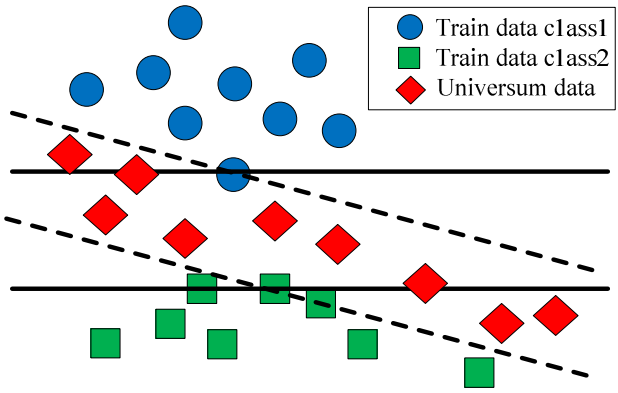

Fig. 1. Two large margin separating hyperplanes on training data

Assume we have n samples with corresponding class labels as $\{x_i, y_i\}_{i=1}^{n}$, where $x_i \in R^d$ is a sample and $y_i \in \{+1, -1\}$ is the class label (i.e., AD as 1 and NC as -1). Also, assume we have m universum data without corresponding class labels denoted as $\{x_j^*\}_{j=1}^{m}$, where $x_j^* \in R^d$ is a universum sample without class label (i.e., MCI in this paper). U-SVM tends to make the normal vector orthogonal to the principal direction of the universum data [11], with the following objective function

$$\begin{aligned} &\min_{w,b} R(w,b) = \tfrac{1}{2}(w \cdot w) + C\sum_{i=1}^{n}\xi_i + C^*\sum_{j=1}^{m}\xi_j^* \\ &\text{s.t.} \quad y_i[(w \cdot x_i) + b] \geq 1 - \xi_i \quad \xi_i \geq 0, i = 1, \ldots, n \\ &\quad |(\mathrm{w} \cdot x_j^*) + \mathrm{b}| \leq \varepsilon + \xi_j^* \quad \xi_j^* \geq 0, j = 1, \ldots, m \end{aligned} \tag{1}$$

where C, $C^* \geq 0$, $\varepsilon \geq 0$, W is the parameter vector of the classifier and b is the bias term. U-SVM augments the objective term $C^*\sum_{j=1}^{m}\xi_j^*$ and constraint $|(\mathrm{w} \cdot x_j^*) + \mathrm{b}| \leq \varepsilon + \xi_j^*$ for universum data. The parameters C and C^* control the trade-off between the minimization of errors and the maximization of the number of contradictions. When $C^* = 0$, the U-SVM formulation is reduced to standard SVM. For labeled training data (i.e., AD and NC), we use standard SVM soft-margin loss with slack variables ξ_i. Also, for the universum samples (i.e., MCI) we adopt the ε-insensitive loss as in standard support vector regression with ξ_j^* denoting slack variables for samples from the universum [7].

The solution to the optimization formulation (1) defines the large margin classification hyperplane between AD and NC that incorporates the priori knowledge with MCI into the final model. The decision function in the dual space is constructed by using a kernel matrix (K(*,*)) of both the labeled samples and the universum samples [10]. As this optimization problem is convex, the solution can also be computed through the corresponding dual optimization problem with kernel version as:

$$max_{\alpha,\mu,\gamma} \sum_{i=1}^{n} \alpha_i - \varepsilon \sum_{j=1}^{m} (\mu_j + \gamma_j) - \frac{1}{2} \sum_{i,k=1}^{n} \alpha_i \alpha_k y_i y_k K(x_i, x_k) - \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i y_i (\mu_j - \gamma_j) K(x_i, x_j^*) - \frac{1}{2} \sum_{j,l=1}^{m} (\mu_j - \gamma_j)(\mu_l - \gamma_l) K(x_j^*, x_l^*)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i + \sum_{j=1}^{m} (\mu_j - \gamma_j) = 0$$

$$0 \le \alpha_i \le C, \ i = 1, \ldots, n$$

$$0 \le \mu_j, \gamma_j \le C^*, \ j = 1, \ldots, m \tag{2}$$

where μ_j, γ_j and α_i are Lagrangians for the dual problem. Also μ_j^0, γ_j^0 and α_i^0 are the solutions of the dual problem, and then the bias b_0 can be calculated. The decision function is formulated as:

$$\mathrm{f}(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i^0 y_i K(x_i, x) + \sum_{j=1}^{m} (\mu_j^0 - \gamma_j^0) K(x_j^*, x) + b_0\right) \tag{3}$$

2.2 Ensemble Universum SVM

A lot of studies have shown that biomarkers from different modalities may have complementary information for discrimination [8, 9]. A common practice in combining different modalities is the concatenation of all features into a longer feature vector.

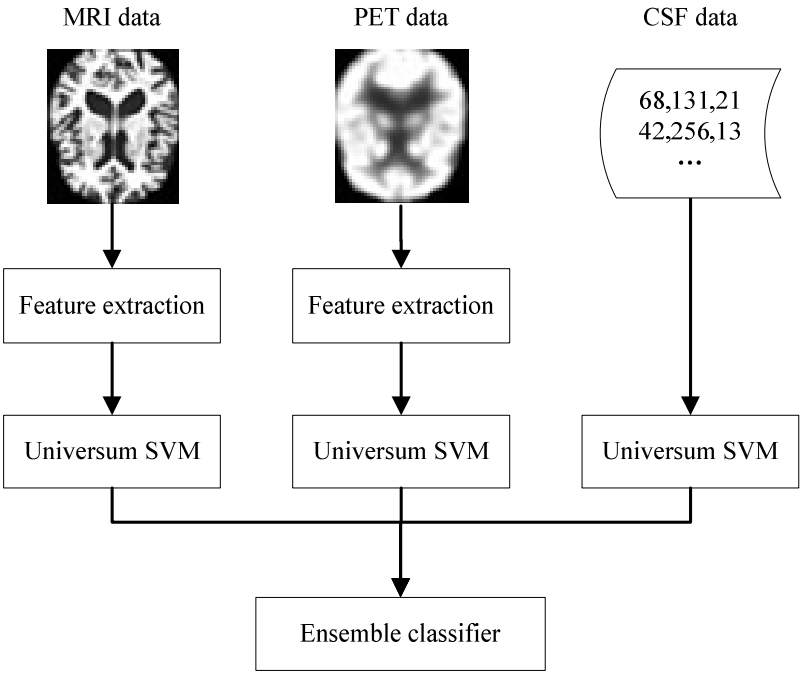

Fig. 2. Flow chart of the proposed Ensemble U-SVM classification method

However, this may be not enough for effective combination of features from different modalities. In this paper, we propose to use ensemble learning [12] to combine complementary information from heterogeneous data.

We give the flow chart of our proposed ensemble U-SVM method for multimodal AD classification in Fig. 2. As can be seen from Fig.2, firstly for each subject we extract features from raw MRI, PET respectively. And then three individual U-SVM classifiers are trained from each different modality using U-SVM. In the procedure of decision-making, for a new unseen testing sample, each of these trained models will have a predication on it, and finally we aggregate all predictions using majority voting strategy to get the final decision.

3 Experiments

In this section, we evaluate the effectiveness of the proposed methods for AD vs. NC classification on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). In our experiments, we use three modalities of data including MRI, PET and CSF data.

3.1 Subjects

The ADNI database contains approximately 200 cognitively normal elderly subjects to be followed for 3 years, 400 subjects with MCI to be followed for 3 years, and 200 subjects with early AD to be followed for 2 years. In this paper, all ADNI baseline subjects with the corresponding MRI, PET, and CSF data are included. This yields a total of 202 subjects, including 51 AD patients, 99 MCI patients, and 52 healthy controls.

Standard image pre-processing is performed for all MRI and PET images, including anterior commissure (AC) - posterior commissure (PC) correction, skull-stripping, removal of cerebellum, and segmentation of structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). With atlas warping, we can partition each subject image into 93 regions of interests (ROIs). For each of the 93 ROIs, we compute the GM tissue volume from the subject's MRI image. For PET image, we first rigidly align it with its respective MRI image of the same subject, and then compute the average value of PET signals in each ROI. Therefore, for each subject, we can finally obtain totally 93 features from MRI image, other 93 features from PET image, and 3 features (Aβ42, t-tau, and p-tau) from CSF biomarkers. A detailed description on acquiring MRI, PET and CSF data from ADNI as used in this paper can be found at [1].

3.2 Experiment Settings

In the experiment settings, the whole set of subject samples are equally partitioned into 10 subsets, and each time the subject samples within one subset are selected as the testing samples and all remaining subject samples in the other 9 subsets are used for training the models. This process is repeated for 10 times.

Linear kernel is used in SVM and U-SVM after performing a common feature normalization step. For normalization, we first perform a feature-level normalization (i.e., across samples) to make each feature have zero mean and unit standard deviation. Then, we perform a sample-level normalization (i.e., across features) to make each sample have unit L_2 norm.

During U-SVM model parameters selection, note small values of parameter C^* suggest that universum samples have little effect on the final model. So the effectiveness of the universum samples is mainly determined by the values of C and C^* (or their ratio) [7]. To investigate the effect of performance of our multimodality classification method, we test all of their possible parameter values given above by a coarse-grid search through cross validation on training samples.

3.3 Results

We compare U-SVM with standard SVM for both single modality and multimodal cases. Table 1 shows the comparison results achieved by standard SVM and U-SVM methods on different modalities. Note that Table 1 shows the averaged results of 10 independent experiments, given the mean and standard deviation of classification accuracies, sensitivities and specificities. As we can see from Table 1, U-SVM can consistently achieve better accuracy than SVM in all cases, which validates the efficacy of our U-SVM method on using MCI subjects as universum data for helping the classification between AD and NC. Specifically, for multimodal case, our proposed ensemble U-SVM can achieve a classification accuracy of 92.76%, which is better than ensemble SVM that achieve only 92.06%. For other performance measures such as sensitivity and specificity, U-SVM also outperforms SVM in most cases.

The results in Table 1 show the advantages of the proposed ensemble U-SVM in improving classification accuracies. We also perform the significance test using paired t-test at the significance level of 0.95. The corresponding results show that our proposed ensemble U-SVM method is significantly better than all other methods including standard SVM and ensemble SVM. This again validates the efficacy of our proposed method.

Table 1. Comparison of performance measures of SVM and U-SVM for AD vs. NC classification using different modalities. (ACC= Accuracy, SEN=Sensitivity, SPE=Specificity).

Methods	SVM			U-SVM		
	ACC%	SEN%	SPE%	ACC%	SEN%	SPE%
MRI	87.11 ±1.24	82.40 ±1.73	91.77 ±1.41	87.98 ±0.87	84.00 ±1.36	91.91 ±1.22
PET	86.25 ±1.12	88.20 ±1.46	84.29 ±1.50	86.55 ±0.79	87.60 ±1.12	85.49 ±1.29
CSF	82.26 ±0.57	82.60 ±0.78	82.00 ±1.64	83.36 ±0.84	83.97 ±0.74	82.86 ±1.51
Ensemble	92.06 ±0.59	91.37 ±1.06	92.71 ±1.32	92.76 ±0.70	91.57 ±1.03	93.91 ±0.88

Finally, in Fig. 3, we compare the classification performance of the proposed ensemble U-SVM method for multimodal classification, with respect to different number of MCI subjects as universum data. As we can see from Fig. 3, as the number of universum data (i.e., MCI subjects) increases, the performance of ensemble U-SVM also increases steadily, showing the usefulness of the universum data in improving the performance of AD classification. This further validates the effectiveness of adopting U-SVM learning by considering MCI as universum data, compared with standard SVM learning.

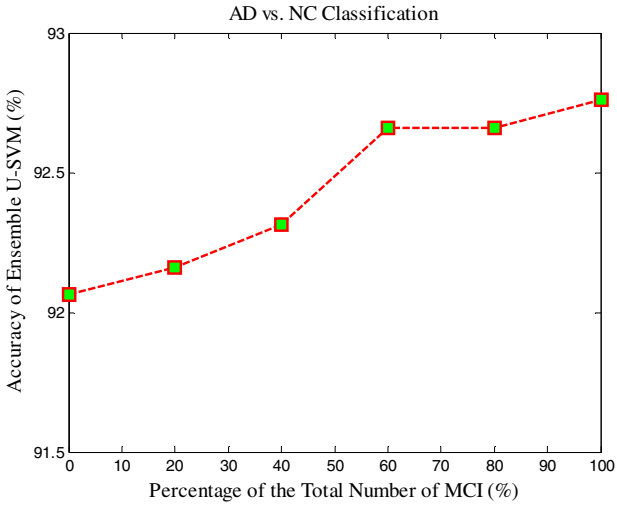

Fig. 3. Average accuracy of Ensemble U-SVM on different numbers of MCI subjects

4 Conclusion

In this paper, we investigate the potential of exploiting MCI subjects that do not belong to either of the classes (AD and NC) of interest to aid the classification between AD and NC subjects. Specifically, we propose to use the universum support vector machine (U-SVM) learning to separate AD from NC subjects through large margin hyperplane with the universum data (i.e., MCI subjects) laying inside the margin borders. Furthermore, we adopt ensemble learning to combine individual U-SVM classifiers trained from each modality. Experimental results on the multimodal imaging data and biological biomarkers from the ADNI database have validated the efficacy of our proposed method.

Acknowledgements. This work was supported in part by SRFDP grant (No. 20123218110009), NUAAFRF grant (No. NE2013105), and JiangsuSF for Distinguished Young Scholar.

References

1. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55, 856–867 (2011)
2. Davatzikos, C., Bhatt, P., Shaw, L., Batmanghelich, K., Trojanowski, J.: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging 32, e2322.e19–e2322.e27 (2011)
3. Cho, Y., Seong, J., Jeong, Y., Shin, S.: Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage 59, 2217–2230 (2012)
4. Zhang, D., Shen, D.: Semi-supervised multimodal classification of Alzheimer's Disease. In: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1628–1631 (2011)
5. Cheng, B., Zhang, D., Shen, D.: Domain Transfer Learning for MCI Conversion Prediction. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 82–90. Springer, Heidelberg (2012)
6. Weston, J., Collobert, R., Sinz, F., Bottou, L., Vapnik, V.: Inference with the Universum. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 1009–1016 (2006)
7. Cherkassky, V., Dhar, S., Dai, W.: Practical conditions for effectiveness of the Universum learning. IEEE Trans. Neural Networks 22, 1241–1255 (2011)
8. Westman, E., Muehlboeck, J., Simmons, A.: Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage 62, 229–238 (2012)
9. Walhovd, K., Fjell, A., Brewer, J., McEvoy, L., Fennema-Notestine, C., Hagler, D., Jennings, R., Karow, D., Dale, A.: Combining MR Imaging, Positron-Emission Tomography, and CSF Biomarkers in the Diagnosis and Prognosis of Alzheimer Disease. Am. J. Neuroradiol. 31, 347–354 (2010)
10. Vapnik, V.: Estimation of dependences based on empirical data. Springer, New York (2006)
11. Sinz, F., Chapelle, O., Agarwal, A., Scholkopf, B.: An Analysis of Inference with the Universum. In: Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS), pp. 1–8 (2008)
12. Tan, A., Gilbert, D.: Ensemble machine learning on gene expression data for cancer classification. Appl. Bioinformatics 2, S75–S83 (2003)