FINAL PUBLISHABLE SUMMARY REPORT

May 10, 2013

Abstract

The scientific objective of this project is to develop novel and effective ensemble classifier systems via the optimization of diversity and accuracy trade off. In this study the following key objectives have been addressed:

- Ensemble classifier systems are improved considering accuracy diversity trade off via novel optimization schemes in Task 1
- Time complexity of model selection is reduced in Task 2 and 3
- Overall model for heterogeneous data is generalized and experimental comparison for classification of facial expressions is provided in Task 4 and 5.

The proposed model is applied to multiclass classification problems from UCI data base and facial expression classification from Cohn Kanade Database.

1 Task 1: Modelling the weighted ensemble classifier system via optimization of diversity and accuracy trade off

Task 1 considers the important problem of finding an optimization model for ensemble classifier systems that manages the diversity and accuracy trade off. A continuous unconstrained optimization model is developed which is implemented on well known UCI machine learning repository [2] and the facial action unit data provided by the University of Surrey CVSSP lab. The objective function involves a quadratic term $x^T Gx$ where x_i represents whether the i^{th} classifier is included in the ensemble (binary variable), G is a square error matrix with diagonal term G_{ii} being the total error of classifier i, and with off diagonals G_{ij} representing the common errors of classifier i and j so that off diagonals correspond to the measure for diversity. [1]

2 Task2: Model selection of base classifiers

Task 2 integrates the new weighted ensemble classifier obtained from Task 1 with a model selection strategy. A novel optimization model is invented and applied to Error Correcting Output Codes (ECOC) which prunes the number of base classifiers of the ECOC matrix. Although techniques are known for creating efficient ECOC matrices, pruning base classifiers of the ECOC matrix has not previously been studied. Regarding the objective function given in Task 1, a term $\lambda \|x\|_0$ is added, where λ is a regularization parameter. This additional term allows us to find a subset of classifiers that constitutes the most accurate and diverse ensemble. We approximated $\|\|_0$ the zero norm by

$$\|x\|_0 := \sum_{i=1}^{n} \mathbf{1}_{\mathbf{x_i} \neq \mathbf{0}} = \lim_{\epsilon \to \mathbf{0}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \frac{\log(\mathbf{1} + |\mathbf{x}|_{\mathbf{i}}/\epsilon)}{\log(\mathbf{1} + \mathbf{1}/\epsilon)}$$

Our model becomes a nonconvex unconstrained optimization problem since the matrix G is symmetric but not positive definite. We transformed the nonconvex problem into a convex problem by using difference of convex functions (DC Programming) and solving by nonlinear optimization method Sequential Quadratic Programming (SQP). More detail can be found in the attached paper.

3 Task 3: Regularization of model complexity

In task 3, we aim to add a regularization term to penalize the complexity. As the number of classifiers increases, the size of the weight vector increases accordingly and hence the model complexity. Therefore, including the ensemble size constraint given in [5] into the objective function with a parameter λ in our new formulation regularizes the model complexity directly. In this way, taking the zero norm approximation of the model complexity allows us to get better approximation then Tikhonov regularization as proposed in the Task 2. The proposed model is compared with other pruning methods including Reduced Error pruning, Kappa Pruning and Random Guessing. We reported the results in a paper (attached) for the Machine Learning journal which is under review.

4 Task 4: Generalization of overall model for heterogeneous multiclass data

Genetic algorithm is proposed in task 4 to learn the kernel weights for each subproblem in ECOC but the computational cost is increased. To improve the slowness of the algorithm we implemented Generalized MKL [4] adapted to our methodology with ECOC, but obtained similar results with the single kernel approach. The reason may lie with the nature of Error Correcting Output Coding since it corrects the errors regardless of single or multiple kernels. In other words, trying to learn by MKL may be redundant and more costly than single kernel learning. ECOC already corrects the output of single kernel binary classifiers (i.e. base classifiers), so we conclude that single kernel will be sufficient for heterogeneous data in ECOC.

5 Task 5: Application

We applied all the above tasks both to UCI data sets and Cohn Kanade facial expression data sets (http://vasc.ri.cmu.edu/idb/html/face/facial expression/). Image processing, feature extraction, dimensionality reduction are performed as in [3]. The method is tested on different datasets such as ecoli, glass, dermatalogy, yeast and wine data sets from UCI, and facial expression classification problem as stated above. We compared the accuracy and running time of our algorithm with different pruning methods such as Reduced Error Pruning, Kappa pruning and Random Guessing. All experiments are implemented for 10 fold training and testing data and accuracy is computed by calculating the mean of 10 fold accuracy. The results show that eliminating the parameter of ensemble size from the constraints but including it in the objective function by approximation algorithm reduces the time complexity and improves the accuracy for most of the data sets. More detail can be found in the attached paper.

In this project, we proposed a novel algorithm which prunes the Error Correcting Output Code(ECOC) method in an optimization framework. The proposed model enables us to eliminate the ensemble size parameter by relaxing the binary optimization model to a continuous model. Application of the project to the facial expression classification problem motivated the forthcoming collaborative project on recognition of heart attack risk of people exercising on a treadmill by facial expressions from online video images. The second illustration can be on recognition of student's understanding capacity for online learning courses through the video images. As a researcher, being part of this project and hosted by the CVSSP Department has developed Dr. Akyuz's career on theoretical machine learning as well as on an application. Discussion with other researchers in different application areas in the CVSSP labaratory has motivated her thinking in engineering applications in the Computer Vision field.

References

[1] C. Bishop. Neural networks;a principled perspective. Technical report, NCRG, 1995.

[2] C.L. Blake and C.J. Merz. UCI repository of machine learning databases. University of California, Irvine, Dept. of Information and Computer Sciences, 1998. http://www.ics.uci.edu/∼mlearn/MLRepository.html.

[3] Raymond Smith and Terry Windeatt. Facial action unit recognition using filtered local binary pattern features with bootstrapped and weighted ecoc classifiers. In Oleg Okun, Giorgio Valentini, and Matteo Re, editors, *Ensembles in Machine Learning Applications*, volume 373 of *Studies in Computational Intelligence*, pages 1–20. Springer Berlin / Heidelberg, 2011. 10.1007/978-3-642-22910-7_1.

[4] M. Varma and B. R. Babu. More generality in efficient multiple kernel learning. In *In Proceedings of the International Conference on Machine Learning, Montreal, Canada*, 2009.

[5] Y. Zhang, S. Burer, and W. N. Street. Ensemble pruning via semi-definite programming. *Journal of Machine Learning Research*, 7:1315 – 1338, 2006.