On the Correlation Between Genotype and Classifier Diversity

Jean-François Connolly, Eric Granger and Robert Sabourin

Laboratoire d'imagerie, de vision et d'intelligence artificielle

École de technologie supérieure, Univertié de Québec, Montreal, Canada

jfconnolly@livia.etsmtl.ca, eric.granger@etsmtl.ca, and robert.sabourin@etsmtl.ca

Abstract

Diversity is a key element in the success of classifier ensembles, and has attracted much recent attention. It is typically measured by directly computing the amount of disagreement between ensemble classifiers at the decision level. This costly process usually involves evaluating output predictions of each classifier over some validation data set. Since most statistical and neural network classifiers can adjust internal learning dynamics by varying their hyperparameter values (corresponding to genotype values), this information can also provide an estimate of diversity. This paper measures the correlation between genotype and classifier diversity among an ensemble of fuzzy ARTMAP neural network classifiers applied to video face recognition. It is empirically shown that as genotype diversity increases within an ensemble, classifier diversity also significantly increases. This correlation can then be exploited to measure the diversity among base classifiers during ensemble design with a significantly lower computational cost.

1. Introduction

In pattern recognition systems, neural or statistical classifiers define class models using data samples defined in a $\mathbb{R}^I$ input feature space, and map those models to a decision space to perform predictions (see Figure 1). In this context, exploiting several different views of a same problem with classifier ensembles has been shown to improve the overall accuracy and reliability for a wide range of applications. Recently, several authors have attempted to incorporate *classifier diversity* measures during the design of classifier ensembles, e.g., [6, 8, 9]. Indeed, through bias-variance error decomposition, it has been empirically shown that considering diversity for ensemble selection improves the generalization capabilities of multiple classifiers systems [1].

In previous work on video-based face recognition, the authors have used representation space traversal [1] to maintain diversity when generating and adapting a pool of incremental-learning classifiers over time [3]. As illustrated in Figure 1, the adaptive multi-classifier system is then defined according to two environments. A *classification environment* that maps an $\mathbb{R}^I$ input feature space (also referred to as hypothesis space in [1]) to a decision space, respectively defined by feature vectors $\mathbf{a}$, and a set of class labels C_k. Since the learning dynamics of a base classifier is governed by a genotype value, a vector $\mathbf{h}$ of hyperparameters (with components like, *e.g.*, the learning rate), changing the latter when learning the some data set (circles and triangles) leads to different decision boundaries in the feature space (limit between shades of gray) and decisions. When adapting hyperparameters to maximize some performance measure, such as accuracy in [3], the classification environment interacts with an *optimization environment*, where each value of $\mathbf{h}$ indicates a position on an objective function defined in a search space.

A diversified ensemble is composed of base classifiers that present different inter-class decision boundaries in the feature space. Since these bounds are either explicitly defined functions (as with SVMs, MLPs, etc.) or implicitly defined (as with kNNs, ARTMAPs, etc.), diversity in the classification environment is mostly measured by computing disagreement between classifiers in the decision space, over several predictions. Assuming a correlation between the two environments, this costly process could be avoided by focusing instead on computing diversity in the search space.

In this paper, the relationship between classifier diversity (in the classification environment) and genotype diversity (in the optimization environment) is studied. Using the fuzzy ARTMAP neural network classifier and different diversity indicators, this paper verifies the hypothesis under which these two types of diversity are correlated. More precisely, it is shown that, as genotype
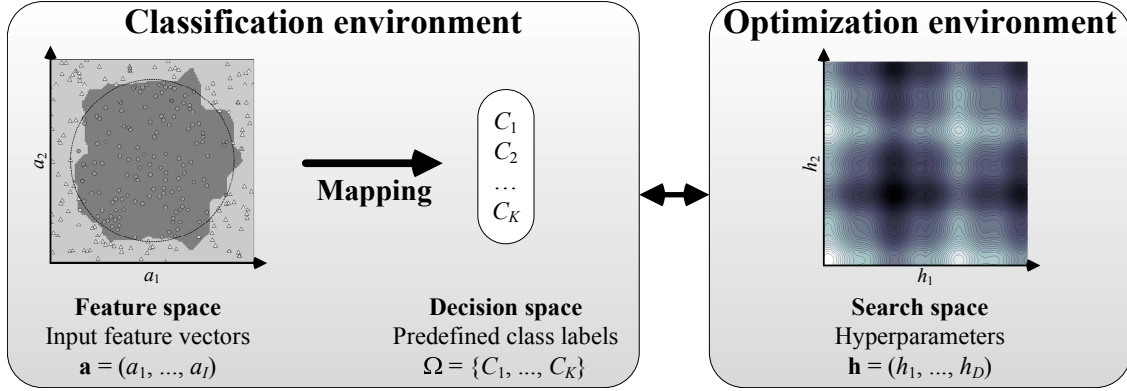
Figure 1. Environments involved in a pattern recognition system.

diversity in the search space increases, diversity in the decision space also increases significantly. The experimental methodology is described in the next section, followed in Section 3 by the results and a discussion on how this correlation can be applied to adaptive ensemble methods.

2. Methodology

2.1 Classifier and data base

This study is part of a project that aims to evolve ensembles of incremental-learning classifiers in response to new reference data, in order to adapt systems for video-based face recognition [3]. In this project, classification is performed using the fuzzy ARTMAP neural network [2]. This self-organizing neural network is capable of stable, on-line, unsupervised or supervised incremental learning, and provides a unique solution to the stability-plasticity dilemma [2]. The internal learning dynamics of fuzzy ARTMAP is set using four hyperparameters in vector $\mathbf{h} = (\alpha, \beta, \epsilon, \bar{\rho})$.

The data base used for this empirical study was collected by the Institute for Information Technology of the Canadian National Research Council (IIT-NRC) [5] for user identification in secured computer login. It is composed of 22 video sequences captured from eleven individuals positioned in front of a computer. For each individual, two color video sequences of about fifteen seconds are captured with a web cam at a rate of 20 frames per seconds – one dedicated to classifier design and the other for testing. The number of facial regions detected with the Viola-Jones algorithm varies from person to person, ranging from 40 to 190 per video sequence, for a total of 1527 design and 1585 test samples.

2.2 Experimental protocol

The hypothesis under which diversity in the optimization and classification environments are correlated is verified with the experiment depicted in Figure 2. It is performed in two steps: (1) optimization during supervised batch learning of the whole IIT-NRC data base with a learning strategy based on particle swarm optimization presented in [3], and (2) particles expansion.

As shown in Figure 2, prior to optimization of hyperparameters, the search space is normalized and bound by a constraint of 0.2. Once the hyperparameter vector $\mathbf{h}$ that maximizes accuracy is found, an ensemble is formed with 17 classifiers. They are each associated with a solution $\mathbf{h}$ and organized into a hypercube centered around the global optimum in the $\mathbb{R}^4$ (fuzzy ARTMAP hyperparameters) search space. To vary diversity in the optimization environment, all solutions are initially situated on the global optimum. The size of the hypercube is then linearly expanded up to the value of the constraint to form different ensembles. During this expansion, genotype diversity varies linearly, affecting the corresponding ensemble of classifiers.

In total, ten trials are performed using ten-fold cross-validation. Out of the ten folds, eight are dedicated to training, one fold is used to validate and determine the number of training epochs, and the remaining fold used to estimate the accuracy of each solution during optimization [3]. Between successive epochs, the presentation order of training patterns is changed randomly.

2.3 Diversity indicators

Diversity in both environments is measured with pairwise indicators between ensemble members e_1 and e_2. In the optimization environment, the pairwise indicator for two classifiers is defined as the the Euclidean
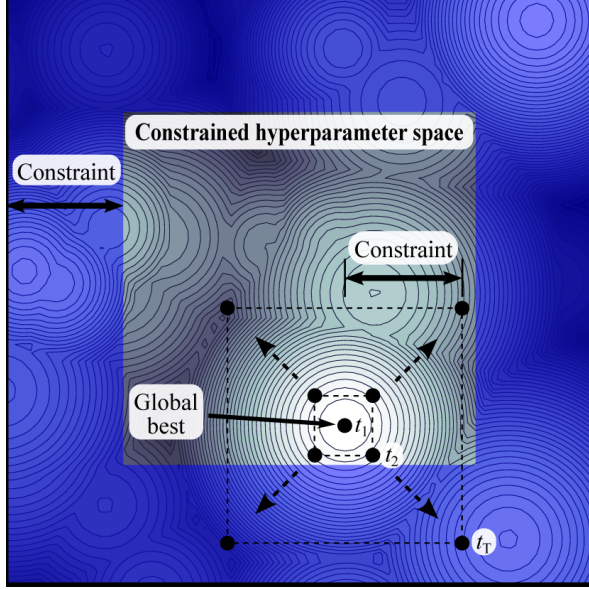
Figure 2. An example of the position of solutions for an objective function in a $\mathbb{R}^2$ hyperparameter search space.

distance between their hyperparameter values in the normalized search space [4, 7]. genotype diversity for the entire ensemble is the average value of all pairwise combination such as:

$$\overline{\delta_{e_1 e_2}} = \frac{2}{E(E-1)} \sum_{e_1=1}^{E-1} \sum_{e_2=e_1+1}^{E} \delta_{e_1 e_2}, \quad (1)$$

where $\delta_{e_1 e_2}$ is the Euclidean distance between two solutions e_1 and e_2 in the search space and E is the number of networks in the ensemble. Although this genotype diversity indicator has a time complexity of $O(N^2)$, it was revealed to be the most accurate [7].

On the other hand, diversity in the classification environment is measured in through correlation indicators: the Q statistic and the correlation coefficient [9]. As with most measures present in literature, they rely on classifier disagreement (in the decision space) to compute correlation among classifiers. For two ensemble classifiers e_1 and e_2, and a given data set (in our case the test data set), each indicator is computed as followed:

The Q statistic:

$$Q_{e_1 e_2} \in [0, 1] = \frac{N_{11}N_{00} - N_{10}N_{01}}{N_{11}N_{00} + N_{10}N_{01}}, \quad (2)$$

where N_{11}, N_{00}, N_{10}, and N_{01} are, for each combination, the number of correct and incorrect predictions by classifiers e_1 and e_2 on the test data set (see Table 1).

Table 1. Contingency table used to compute pairwise diversity between two ensemble classifiers e_1 and e_2 with the Q statistic and correlation coefficient.

	e_1 correct	e_2 incorrect
e_1 correct	N_{11}	N_{10}
e_2 incorrect	N_{01}	N_{00}

Correlation coefficient:

$$\rho_{e_1 e_2} \in [0, 1] =$$
$$\frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{(N_{11}+N_{10})(N_{01}+N_{00})(N_{11}+N_{01})(N_{10}+N_{00})}}, \quad (3)$$

As with genotype diversity, overall ensemble classifier diversity (or correlation) is the average value of all pairwise combination [9]. They are respectively noted $\overline{Q_{e_1 e_2}}$ and $\overline{\rho_{e_1 e_2}}$.

3. Results and discussion

Figure 3 presents the value of the two ensemble correlation indicators as a function of particle diversity in the search space when training on the IIT-NRC data base. As mentioned earlier, fuzzy ARTMAP performs on-line learning. That is, it defines its decision boundaries by learning patterns sequentially. Although all solutions are initially trained with the same hyperparameter values, they are trained with different random pattern presentation orders, leading to a minimal amount of diversity (or correlation). When all particles are initially positioned at the global best position, this yields correlation indicators that are lower than one: $\overline{Q_{e_1 e_2}} = 0.80 \pm 0.01$ and $\overline{\rho_{e_1 e_2}} = 0.47 \pm 0.01$.

As the hypercube expands, genotype diversity increases linearly. By computing diversity in the decision space with indicators based on ensemble disagreement, classifier diversity (correlation) follows the same trend by increasing (decreasing) constantly. Depending the indicator used, diversity in the classification environment changes significantly for different levels of genotype diversity: the Q statistic and correlation coefficient differ for respective genotype diversities 0.26 ($\overline{Q_{e_1 e_2}} = 0.7 \pm 0.1$) and 0.25 ($\overline{\rho_{e_1 e_2}} = 0.41 \pm 0.06$).

These results thus confirm the initial hypothesis that genotype diversity in the optimization environment does indeed translates to diversity among classifiers in the classification environment. Measuring classifier diversity would then necessitate only computing simple Euclidean distances, rather than relying on costly classifier diversity indicators. When adapting multiclassi-
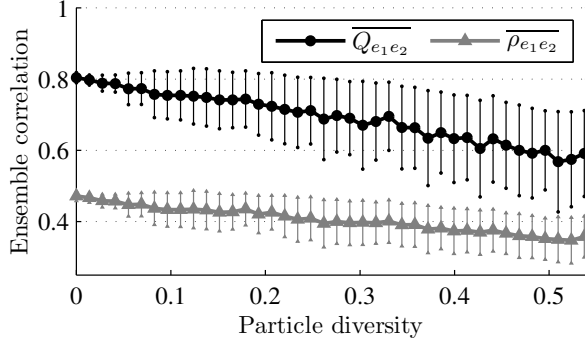
Figure 3. Ensemble diversity in the classification environment as a function of genotype diversity in the optimization environment. Diversity is shown using the classifier correlation indicators $\overline{Q_{e_1 e_2}}$ and $\overline{\rho_{e_1 e_2}}$. A decrease in correlation signifies an increase in diversity; each indicator is shown with its 90% confidence interval.

fier systems, diversity among a pool of classifiers can then be maintained simply by controlling solutions in the optimization space through their genotype values. Moreover, with an overproduce and choose ensemble design strategy, genotype information can also be used to efficiently select a diversified set of classifiers among the most accurate in the pool. For instance, accuracy *and* genotype diversity of solutions in a pool of classifiers has been successfully used for selection of robust ensembles in [3].

When guiding several classifiers according to performance with a population-based optimization algorithms, properties of the latter to preserved genotype diversity should be considered so that classifier diversity is maintained. In the search space, this means being able to detect multiple local minima that would yield accurate, but diverse classifiers. An optimization algorithm used to generate ensembles should also be able to maintain diversity among each local minimum such as diversity is still maintained if there are but few local optimum in the search space.

4. Conclusion

Using the fuzzy ARTMAP neural classifier and different diversity indicators, empirical results presented in this paper indicate that there is indeed a correlation between diversity in the classification and optimization environments. Diversity in hyperparameter space can therefore be used during generation of a pool of clas-

sifiers and during selection of ensembles among that pool, instead of evaluating output predictions of each classifier over some validation data set, as required to compute classifier diversity measures. While this study only considers overall ensemble diversity, future work should focus on local diversity around each optima in the search space, to provide insight on properties of an optimization algorithm needed when adapting classifiers. Moreover, a comparison of different genotype diversity indicators [4, 7] should also be considered to isolate ones most correlated with classifier diversity.

5   Acknowledgments

References

[1] G. Brown, J. Wyatt, R. Harris, and X. Yao. Diversity creation methods: a survey and categorization. *Information Fusion*, 29(6):5–20, 2005.

[2] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen. Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Systems, Man, and Cybernetics C*, 3(5):698–713, 1992.

[3] J.-F. Connolly, E. Granger, and R. Sabourin. Evolution of heterogeneous ensembles through dynamic particle swarm optimization for video-based face recognition. *Pattern Recognition*, 45(7):2460–2477, 2012.

[4] G. Corriveau, R. Guilbault, A. Tahan, and R. Sabourin. Review and study of genotype diversity measures for real-coded representations. *IEEE Transaction on Evolutionary Computation*, 2012, doi: 10.1109/TEVC.2011.2170075.

[5] D. O. Gorodnichy. Video-based framework for face recognition in video. In *Second Workshop on Face Processing in Video in Proceedings of the Conference on Computer and Robot Vision*, pages 325–344, Victoria, Canada, May 2005.

[6] S. T. Hadjitodorov, L. I. Kuncheva, and L. P. Todorova. Moderate diversity for better cluster ensembles. *Information Fusion*, 7(3):264–275, 2006.

[7] O. Olorunda and A. P. Engelbrecht. Measuring exploration/exploitation in particle swarms using swarm diversity. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 1128–1134, Hong Kong, China, May 2008.

[8] K. Sirlantzis, S. Hoque, and M. Fairhurst. Diversity in multiple classifier ensembles based on binary feature quantisation with application to face recognition. *Journal of Visual Language and Computing*, 8:437–445, 2008.

[9] A. Ulaş, M. Semerci, O. T. Yildiz, and E. Alpaydin. Incremental construction of classifier and discriminant ensembles. *Information Sciences*, 179(9):1298–1318, 2009.