Using Additional Neighborhood Information in a Dynamic Ensemble Selection Method: Improving the KNORA Approach

Leila Maria Vriesmann
State University of Ponta Grossa
Ponta Grossa, PR, Brazil
Pontifical Catholic University of Parana
Curitiba, PR, Brazil
lmvriesmann@uepg.br

Alceu de Souza Britto Jr.
Pontifical Catholic University of Parana
Curitiba, PR, Brazil
State University of Ponta Grossa
Ponta Grossa, PR, Brazil
alceu@ppgia.pucpr.br

Luiz Eduardo Soares Oliveira
Federal University of Parana (UFPR)
Curitiba, PR, Brazil
lesoliveira@inf.ufpr.br

Robert Sabourin
École de Technologie Supérieure, ETS
Montreal, QC, Canada
robert.sabourin@etsmtl.ca

Albert Hung-Ren Ko
École de Technologie Supérieure, ETS
Montreal, QC, Canada
albert@livia.etsmtl.ca

Abstract—This work evaluates some strategies to approximate the performance of a dynamic ensemble selection method to the oracle performance of its pool of weak classifiers. For this purpose, different strategies are evaluated to combine the results of the KNORA dynamic ensemble selection method with the results of its built-in KNN used to define the neighborhood of a test pattern during the ensemble creation. The KNN results are considered as additional information which may be combined with the KNORA results to improve the recognition performance. A strong experimental protocol based on more than 60,000 samples of handwriting digits extracted from NIST-SD19 has shown that the fusion of the KNORA results with the results of its built-in KNN is very promising.

Keywords- dynamic ensemble selection; oracle; KNORA

I. INTRODUCTION

The dynamic ensemble selection explores the use of different classifiers in different test patterns [1, 2, 3, 4, 5, 6, 7]. Based on the different features or the different decision regions of each test pattern, a classifier is selected and assigned to the ensemble. This is an advantage when compared with the dynamic selection of single classifiers, since the choice of one individual classifier over the rest will depend on how much we trust the estimate of the generalization of the classifiers [6]. With an ensemble of classifiers, we distribute the risk of this over-generalization.

Some methods of dynamic selection have their performances compared with that of the oracle, which assigns the correct class label to a pattern when at least one individual classifier from an ensemble produces the correct class label. A Priori selection, A Posteriori selection, overall local accuracy (OLA) and local class accuracy (OCA) are popular examples of these methods [2, 3, 4, 7].

We consider the oracle the possible upper limit of classification accuracy which is defined as the ratio of samples that are classified correctly by at least one classifier in the pool to all samples. In this paper we focus in the investigation of some strategies to approximate the performance of a dynamic ensemble selection method to the oracle performance of its pool of week classifiers. Proposed in [8], the KNORA (*K*-nearest-oracles) uses a KNN to find the *K* nearest neighbors of a test pattern *X* (to be recognized) in a feature space, where we know the classifiers of the pool that correctly classify each sample. Then, different strategies are used to select these classifiers to compose an ensemble, that will be used to classify the test pattern.

In [8], the results were promising, and then our investigation considers the KNORA method applied in a handwriting numeral recognition problem. The main question is: May the use of the additional information provided by the KNN used to select the *K* nearest neighbors of the test pattern in the KNORA method be interesting to improve the KNORA final results?

To answer this question, we have evaluated different strategies to combine the KNN used to select the *K* nearest neighbors of the test pattern in the KNORA method with the final KNORA results. The paper is organized into 5 sections. Section II presents the

KNORA ensemble selection method. Section III describes the fusion strategies evaluated. The experimental results are presented in Section IV, while the conclusion and further works may be found in Section V.

II. KNORA Dynamic Ensemble Selection Method

The *K*-nearest-oracles (KNORA) method proposed by [8] considers the neighborhood of test patterns in the validation set to select the classifiers to compose the ensemble. For a given test pattern, it locates the *K* neighbors in the validation set. Since we know which classifiers in the pool can recognize each sample in the validation set, an ensemble of classifiers can be dynamically selected to label the given test pattern. Different schemes have been proposed:

- KNORA-Eliminate: Given *K* neighbors x_j, $1 \leq j \leq K$, of a test pattern *X*, and supposing $C(j)$, $1 \leq j \leq K$, a set of classifiers that correctly classifies all its *K*-nearest neighbors, then every classifier $c_i \in C(j)$ should submit a vote on the sample *X*. In the case where no classifier can correctly classify all the *K*-nearest neighbors of the test pattern, find the classifier that correctly classifies more neighbors in *K*. Then, only use the classifiers that hit this number of neighbors.
- KNORA-Union: Given *K* neighbors x_j, $1 \leq j \leq K$, of a test pattern *X*, and supposing that the j-nearest neighbor has been correctly classified by a set of classifiers $C(j)$, $1 \leq j \leq K$, then every classifier $c_i \in C(j)$ should submit a vote on the sample *X*. Note that, since all the *K*-nearest neighbors are considered, a classifier can have more than one vote if it correctly classifies more than one neighbor. The more neighbors a classifier classifies correctly, the more votes this classifier will have for a test pattern.

The other two schemes are KNORA-Eliminate-W and KNORA-Union-W, which are similar to KNORA-Eliminate and KNORA-Union, respectively, however, in these schemes each vote is weighted by the Euclidean distance between the neighbor pattern x_j and the test pattern *X*.

The KNORA method apparently gives better performances than static ensemble selection schemes such as GA with the MVE (majority voting error) as the objective function and also perform slightly better than other dynamic selection methods as reported in [8].

III. KNN and KNORA Fusion

As described in the previous section, all KNORA strategies take into account a built-in KNN. In this section we investigate different strategies to combine the KNN (already used to select the classifiers for the KNORA method) with the final KNORA results, in order to approximate the recognition performance of that estimated as the oracle of our pool of weak classifiers. In fact, five different schemes were implemented:

- KNORA CU (Conditional Use): execute the KNN for the test pattern. If less than *Y%* (*Y* is a predetermined value) of the all neighbors of the current test pattern belong to the same class, then execute KNORA (Eliminate or Union), and use only the KNORA final outputs. Otherwise, use the KNN outputs.
- KNN+KNORA CF (Conditional Fusion): execute the KNN for the test pattern. If less than *Y%* of the all neighbors of the current test pattern belong to the same class, then combine the KNN outputs with the outputs of the KNORA (Eliminate or Union). Otherwise, use the KNN outputs.
- KNN CU (Conditional Use): execute KNORA (Eliminate or Union). If less than *Y%* of the classifiers in the ensemble selected for the current test pattern have the same vote, then use only the KNN outputs. Otherwise, use the KNORA outputs.
- KNORA+KNN CF (Conditional Fusion): execute KNORA. If less than *Y%* of the classifiers of the current test pattern have the same vote, then combine the KNORA (Eliminate or Union) outputs with the KNN outputs. Otherwise, use the KNORA outputs.
- KNN+KNORA UF (Unconditional Fusion): combine the KNN outputs with the KNORA (Eliminate or Union) outputs.

The fusion or combination of outputs is always done by the majority voting scheme. The experimental results are reported in the next section.

IV. Results and Discussions

A. Database and Pool of Week Classifiers

The experiments undertaken to evaluate the proposed method are based on the same experimental protocol described in [8]. It was selected a large scale pattern recognition problem related to the recognition of handwritten numerals from NIST SD19. Three databases were used: the training set with 5000 samples (*hsf_{0 - 3}*), the validation set containing 10000 samples (*hsf_{0 - 3}*) and the test set containing 60089 samples (*hsf_{7}*). The accuracies were obtained evaluating the examples of the test set.

We need to address the fact that the pool of the KNORA method was composed of NN (*K*=1) classifiers generated with feature subsets having only 32 features out of 132. The same pool of weak classifiers proposed in [8] is used in our experiments. This pool contains 100 NN classifiers created based in the Random Subspaces scheme.

B. Benchmark Parameters

In [8], the authors have reported that with the 132-feature based KNN, with *K*=1, the performance on the testing set is 93.34%. The combination of all 32-feature based KNN classifiers available in our pool (100 elements) by simple majority voting gives 96.28% of classification accuracy. In addition, the best KNORA recognition rates for the same database were reported as: 97.25% for KNORA-Union (*K*=1) and 97.52% for KNORA-Eliminate (*K*=7 and *K*=8), as shown in Table I.

The *K* parameter of the KNORA method had been evaluated from 1 to 30.

TABLE I. BEST RECOGNITION RATES AND THE CORRESPONDING K VALUES IN [8]

Scheme	Recognition Rates (K)
KNN	93.34(1)
KNORA-Union	97.25(1)
KNORA-Eliminate	97.52(7,8)

The oracle for our pool of weak classifiers is 99.95% of recognition rate. This was obtained looking down at the pool of classifiers if there are some which settled the class of each sample test. If at least one hit, the value was incremented.

C. Evaluation of the Fusion Strategies

The *K* parameter of the KNORA method was evaluated from 1 to 30, and the *Y* parameter for the combination schemes proposed in this paper was evaluated from 10% to 100%. The results of each strategy are presented in accordance with the following sequence:

- KNORA CU (Conditional Use): In Table II, the two first lines show the best results obtained when using this particular scheme with KNORA-Eliminate and KNORA-Union, respectively. The best recognition rates are 97.66% (with *Y*=80% or more, and *K*=4) when using KNORA-Eliminate, and 97.54% (with *Y*=70% or more, and *K*=2) when using KNORA-Union. As we can observe, both cases provided better results than KNORA alone (Eliminate 97.52% and Union 97.25%) and than KNN.
- KNN+KNORA CF (Conditional Fusion): In this scheme we have also observed better results than use KNORA-Eliminate or KNORA-Union alone. In the third and fourth lines of the Table II, one can see the best recognition rates for this scheme by considering KNORA-Eliminate and KNORA-Union, respectively. The best recognition rates are 97.73% (with *Y*=80% or more) when using KNORA-Eliminate, and 97.54% (with *Y*=70% or more, and *K*=2) when using KNORA-Union.
- KNN CU (Conditional Use): The fifth and the sixth line of the Table II present the best results executing KNN (after KNORA-Eliminate) CU and KNN (after KNORA-Union) CU, respectively. The best recognition rate is 97.58% (with *Y*=50%, and *K*=7) when using KNN (after KNORA-Eliminate) CU, which is better than KNORA-Eliminate alone (97.52%). The best recognition rate is 97.48% (with *Y*=60%, and *K*=1) when using KNN (after KNORA-Union) CU, which is better than KNORA-Union alone (97.25%).
- KNORA+KNN CF (Conditional Fusion): The seventh line of the Table II presents the best results when after executing KNORA-Eliminate, if less than *Y*% of the classifiers of the current test pattern have the same vote, then we combine the KNORA-Eliminate outputs with the KNN outputs. Otherwise, we use the KNORA-Eliminate alone. A similar experiment was done for KNORA-Union, see the last line of Table II. The best recognition rates are 97.70% (with *Y*=70% or more, and *K*=7) when using KNORA-Eliminate, and 97.30% (with *Y*=60% or more, and *K*=1) when using KNORA-Union.
- KNN+KNORA UF (Unconditional Fusion): In this scheme, we always combine the KNN with KNORA (Eliminate or Union) for all samples to be classified. Thus, there is no *Y* parameter to evaluate. The best recognition rate (Table III) observed for KNN+KNORA (Eliminate) UF was 97.74% (*K*=6), while for the KNN+KNORA (Union) UF was 97.30% (*K*=1). The Figure 1 presents the recognition rate obtained for each value of the parameter *K* (of 1 up to 30) in this scheme, as well as the recognition rates obtained for KNORA-Eliminate and KNORA-Union in the work of [8]. It is observed that KNN+KNORA Union UF reached better results than KNORA-Union of [8], and that KNN+KNORA Eliminate UF reached better results than KNORA-Eliminate of [8].

As one can see, the best result of all experiments was 97.74% (*K*=6), achieved by KNN+KNORA (Eliminate) UF (Unconditional Fusion). This result is also better than that reported in [8].

The oracle performance (99.95%) was not reached. However, we have show that we can improve the KNORA results by considering additional information from the KNN used to select the ensembles.

TABLE II. RECOGNITION RATES AND THE CORRESPONDING *Y* AND *K* VALUES

Fusion Scheme	*Y%*									
	10%	*20%*	*30%*	*40%*	*50%*	*60%*	*70%*	*80%*	*90%*	*100%*
KNORA(E) CU	97.40(3)	97.40(3)	97.40(3)	97.42(3)	97.40(3)	97.64(2)	97.64(3)	97.66(4)	97.66(4)	97.66(4)
KNORA(U) CU	97.40(3)	97.40(3)	97.40(3)	97.42(3)	97.42(3)	97.48(4)	97.54(2)	97.54(2)	97.54(2)	97.54(2)
KNN+ KNORA(E) CF	97.40(3)	97.40(3)	97.40(3)	97.42(3)	97.42(3)	97.65(2)	97.70(3)	97.73(4)	97.73(6,7)	97.73(6,7)
KNN+ KNORA(U) CF	97.40(3)	97.40(3)	97.40(3)	97.42(3)	97.42(3)	97.48(4)	97.54(2)	97.54(2)	97.54(2)	97.54(2)
KNN (After KE) CU	97.52 (7,8)	97.52 (7,8)	97.53(8)	97.53(8)	97.58(7)	97.55(3,5)	97.46(3)	97.43(1)	97.40(1,3)	97.38(3)
KNN(After KU) CU	97.25(1)	97.25(1)	97.27(1)	97.34(1)	97.39(1)	97.48(1)	97.43(1)	97.43(3)	97.41(3)	97.40(3)
KNORA(E)+K NN CF	97.52 (7,8)	97.52 (7,8)	97.53(8)	97.56(4)	97.62(6)	97.69(5)	97.70(7)	97.70(7)	97.70(7)	97.70(7)
KNORA(U)+K NN CF	97.25(1)	97.25(1)	97.26(1)	97.28(1)	97.29(1)	97.30(1)	97.30(1)	97.30(1)	97.30(1)	97.30(1)

TABLE III. BEST RECOGNITION RATES AND THE CORRESPONDING K VALUES

Scheme	Recognition Rates (K)
KNN+KNORA (E) UF	97.74(6)
KNN+KNORA (U) UF	97.30(1)

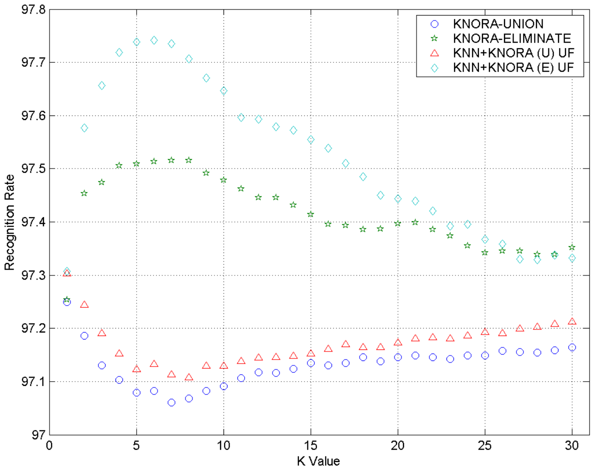

Figure 1. Recognition rates for KNORA-UNION, KNORA-ELIMINATE, KNN+KNORA (U) UF and KNN+KNORA (E) UF

V. CONCLUSION

As we can see in the experimental results the additional information provided by the KNN built in the KNORA method, originally used to define the neighborhood of the test pattern, allow us to improve the recognition performance in most of the strategies evaluated. The best results were achieved when an unconditional fusion were used. It means that the neighborhood additional information plays an important role in the classification process provided by the KNORA method. As future works, we plan to model the oracle properties, aiming to replace the KNN used in the KNORA process by a classifier whose the objective will be to define which classifiers will be part of the ensemble for a specific test pattern.

REFERENCES

[1] M. S. J. Cao, M. Ahmadi. "Recognition of handwritten numerals with multiple feature and multistage classifier," Pattern Recognition, vol. 28, n. 2, pp. 153-160, 1995.

[2] L. Didaci, G. Giacinto, F. Roli, and G. L. Marcialis. "A study on the performances of dynamic classifier selection based on local accuracy estimation," Pattern Recognition, vol. 38, n. 11, pp. 2188-2191, 2005.

[3] A. Didaci and G. Giacinto. "Dynamic classifier selection by adaptive k-nearest-neighbourhood rule," pp. 174-183, 2004.

[4] G. Giacinto and F. Roli. "Methods for dynamic classifier selection," In International Conference on Image Analysis and Processing (ICIAP 1999), pp. 659-664, 1999.

[5] P. L. S. P.-R. V. Gunes, M. Menard. "Combination, cooperation and selection of classifiers: a state of the art," In Int. J. Pattern Recognition Artif. Intell., vol. 17, pp. 1303-1324.

[6] L. Kuncheva. "Switching between selection and fusion in combining classifiers: an experiment," IEEE Trans. Syst. Man Cybern. Part B, vol. 32, n.2, pp.146-156, 2002.

[7] K. Woods, W. P. Kegelmeyer Jr, and K. Bowyer. "Combination of multiple classifiers using local accuracy estimates." IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, pp. 405-410, 1997.

[8] A. H. R. Ko, R. Sabourin, and A. S. Britto Jr. "From dynamic classifier selection to dynamic ensemble selection," Pattern Recognition, vol. 41, pp. 1718-1731, 2008.