Application of adaptive committee classifiers in on-line character recognition

Matti Aksela[1], Jorma Laaksonen[1], Erkki Oja[1], and Jari Kangas[2]

[1] Helsinki University of Technology, Neural Networks Research Centre, P.O.Box 5400, Fin-02015 HUT, Finland
[2] Nokia Research Center, P.O.Box 100, Fin-33721 Tampere, Finland

Summary. There are two main approaches to classifier adaptation. A single adaptive classifier can be used, or an adaptive committee of classifiers whose members can be either adaptive or non-adaptive. We have experimented with some approaches to adaptive committee operations, including the Dynamically Expanding Context (DEC) and the Modified Current-Best-Learning (MCBL) approaches.

In the experiments of this paper the feasibility of using an adaptive committee classifier is explored and tested with on-line character recognition. The results clearly show that the use of adaptive committees can improve on the recognition results, both in comparison to the individual member classifiers and the non-adaptive reference committee.

Keywords. adaptive, committee, classifier combining, character recognition

1 Introduction

A common approach to any classification task is to use a set of reference samples, stored as prototypes or model coefficients, and match the input sample with them. In order to improve the classification performance in situations where a significant amount of variation in the input samples exists, classifier adaptation is an effective method.

Since the primary objective of any recognition system is to achieve the best attainable performance, it is viable to combine different classifiers in a committee formation to enhance overall performance. This is possible because in the outputs of several classifiers the errors are not necessarily overlapping and thus the committee can improve on its members' results [1].

Although the most common way of adaptation is to adapt a single recognizer to the given training data, it is also possible to construct a committee that as a whole is adaptive. The members of such a committee can be adaptive or non-adaptive themselves.

In on-line handwriting recognition the classifier or classifiers must be capable of dealing with natural handwriting. Because of the intrinsic variation

[1] Acknowledgement: This research was partly financed by the project New Information Processing Principles, Finnish Centre of Excellence Programme 2000-2005, Academy of Finland.

in writing styles adaptation is necessary for a user-dependent handwriting recognition application, as adopting the vast amount of variation into the initial models is usually impossible. With the continuous increase in computational power, the use of committee methods generally requiring more than one member classifier to recognize the input is no longer computationally too complex for even the smallest platforms performing on-line handwriting recognition, Personal Digital Assistants (PDAs).

In our research group, very positive results have been obtained with the Dynamic Time Warping (DTW) -based recognizer using single classifier adaptation [2–4]. Still, the question as to how these results could be improved further was left open. When searching for a suitable method of committee adaptation the idea of using the Dynamically Expanding Context (DEC) principle, previously mainly used for speech recognition [5,6], arose. The principle was modified somewhat to suit application in handwriting recognition [7].

Even though committee classification has been extensively researched, the use of adaptive committee classifiers is a much more novel approach. In this paper we present two examples of adaptive committee classifiers. In addition to the DEC committee, also a modification of the Current-Best-Learning (CBL) algorithm [8] will be examined and are explained below. We show that they outperform both a non-adaptive method and a simpler adaptive structure.

In Section 2 the principles for adaptive committee recognition are explored and the adaptive committees used later in the experiments are described. Section 3 explains the data sets and member classifiers used in our experiments and in Section 4 the obtained results are shown. Finally in Section 5 conclusions on the results are drawn and some future directions elaborated on.

2 Committee adaptation methods

The basic operation of a committee classifier is to take the results of the member classifiers and attempt to combine them in a way that improves performance. The member classifiers have a significant impact on the final performance of the committee. It can generally be said that the less the errors of the member classifiers are correlated, the more effective the committee can be in improving recognition accuracy.

Numerous committee structures have recently gained attention. Arguably the most widely known method of classifier combining, majority voting, has in spite of its simplicity been shown to be very effective [9]. Also Bayesian combination methods [10], multistage combinators [11], group-wise classification [12] and critic-driven combining [13] have been studied.

An adaptive committee can be thought of as consisting of two parts. First, every committee must have a base decision rule, which can be used when no adaptation has been performed. Then, some rule or set of rules for the adaptation must be included. The type of rules can vary from very simple weighting or preference adjusting schemes to the creation of complex lists of rules to determine the committee's behavior. Adaptive committee recognition methods found in the literature include, for instance, the Adaptive Integration of Multiple Experts (AIME) system [14].

2.1 Dynamically Expanding Context

The most effective adaptive committee used in our work in on-line handwritten character recognition is based on the Dynamically Expanding Context (DEC) algorithm. The algorithm was originally developed to create transformation rules that would correct typical coarticulation effects in phonemic speech recognition [5]. The notation for a DEC rule stands as $l(A)r \rightarrow B$, where A is a segment of the source string S, B is the corresponding segment in the transformed string T, and $l(\cdot)r$ is the context in string S where A occurs. In other words, A is replaced by B under the condition $l(\cdot)r$.

The main philosophy behind the approach is to determine just a sufficient amount of context for each individual segment A so that all conflicts in the set of training samples will be resolved [5]. Thus an optimal compromise between accuracy and generality is expected to be obtained. The central idea of the method is to always first try to find a production of the lowest contextual level to sufficiently separate contradictory cases. Starting with context level 0, or the context-free level, contexts of successively higher levels will be utilized until all conflicts are resolved.

The DEC principle has to be slightly modified to suit the setting of isolated handwritten character recognition [7]. The DEC committee consists of a number of classifiers, that are first initialized and then tested and ranked in the order of decreasing performance. The primary outputs and the second-ranking results of the member classifiers are used as a one-sided context for the creation of the DEC rules. The primary outputs and the second-ranking results of every member classifier are always different character classes. A schematic diagram of the DEC-based adaptive committee classifier is shown in Figure 1. In this example there are three member classifiers. The first-ranking results are denoted symbolically as a, b and c, and the second-ranking ones as d, e and f. For instance the rule "$abcd \rightarrow s$" means that if the first-ranking results for classifiers 1, 2 and 3 are a, b and c and the second-ranking result for classifier 1 is d, then the input character is classified in class s.

When training the DEC committee, characters of known classification are input one by one. Each time a character is input to the system, the member classifiers give the first- and second-ranking class. Then the existing rules are searched through and the first applicable rule gives the classification result. If no applicable rule is found, the default decision is applied. The classification result is compared to the correct class. If the recognition was incorrect, a new rule is created. Every new rule that is created employs more contextual knowledge, if at all possible, than the rule causing the conflict. Eventually the entire context available will be used and more precise rules can no longer be

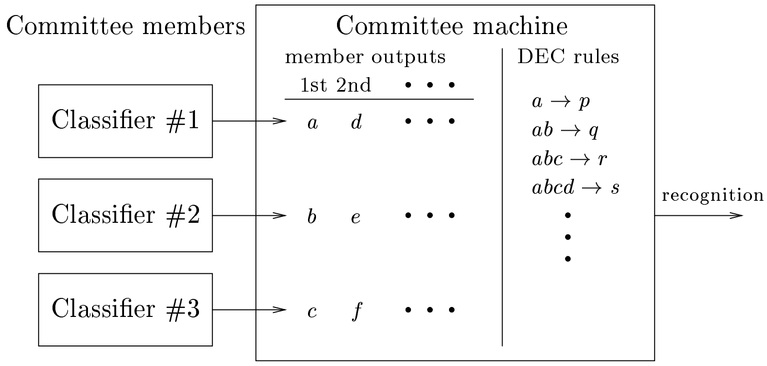

Fig. 1. A block diagram of the DEC-based adaptive committee classifier

written. For this situation a method for tracking the correctness of the rules can be used and the highest level rule most likely to be correct is applied.

The introduction of a new writer always results in the re-initialization of the rule base, as the adaptation is aimed to be user-dependent. With off-line training the training set could be reiterated until rule consistency is ensured. But with an on-line system storing all previous input samples and using them in an iterative manner would be too expensive in terms of both performance and storage space. Thus it is assumed that prior samples will not be available afterwards.

Several options were explored in the search for the best achievable recognition result using the DEC committee. These options included the following.

Default decision: The system's default decision rule is needed when no character-specific rules yet exist. Two methods for producing the default decision were experimented with. The first is to simply use the output of the best-ranked classifier. The alternative is to perform majority voting on the results obtained from the classifiers to make the default decision.

Requiring the inclusion of the output: Another variation implemented was the possibility to require that the output symbol B for a rule of the form $(A)r \rightarrow B$ must be included in the context $(A)r$. In other words, one of the classifiers must produce the result for it to be the output of the committee.

Use of second-ranking results: The committee can function either by using just the first-ranking results from its member classifiers or by also including the second-ranking results. The second-ranking results can be used in two ways, either horizontally or vertically.

The horizontal inclusion of the second-ranking results means that the first and second-ranking results from the best-performing member are used first. Then the two results from the second-best performing classifier are used in the same order, then the third classifier and so on. In Figure 1, this corresponds to the order 'a', 'd', 'b', 'e', 'c' and 'f'.

The vertical approach uses all first-ranked results prior to any second-ranked results from any classifier. So the first-ranked result of the best classifier is followed by the first-ranked results from the other classifiers until all primary outputs have been used. Then the second-ranked results are used in a similar fashion. This approach corresponds to the order 'a', 'b', 'c', 'd', 'e' and 'f' in Figure 1.

Conflict resolution: The initial version of the DEC implementation simply discarded rules as they resulted in an incorrect answer but this was quickly seen to be suboptimal. Hence three options were implemented to discriminate between conflicting high-level rules. These are 1) inactivation of the latest incorrect rule, 2) counting the correct applications and using the one with most correct results, or 3) counting both the correct and incorrect applications and making the decision based on their difference.

2.2 Modified Current-Best-Learning

The Current-Best-Learning (CBL) algorithm [8] strives for a consistent hypothesis for the entire set of samples by generalizing or specializing an initial hypothesis. The original algorithm uses backtracking to ensure that the hypothesis is also consistent with all prior samples. The specialization operation indicates that a unit, a location within the hypothesis space, that was previously positive must be deemed negative, and the generalization then refers to setting a previous negative to positive.

The algorithm used here has deviated quite far from that initial idea, but as the resemblance is still evident, it is here called Modified Current-Best-Learning (MCBL). As in the original version, the data space is a two-dimensional grid. The use of just a positive and negative value would require a separate class for each sample, which would not be practical. So the values used here are in a way estimates of the confidence in a particular decision, and are defined as

$$c_j(\overline{x}) = 1 - \frac{d_j(\overline{x})}{d_1(\overline{x}) + d_2(\overline{x})}, \tag{1}$$

where $c_j(\overline{x})$ is the confidence output for the sample $\overline{x}$. $j \in \{1, 2\}$ is the index indicating whether the confidence value is being calculated for the first or second-ranking result, and $d_1(\overline{x})$ and $d_2(\overline{x})$ are the distances to the first and second-ranked prototypes, respectively.

By collecting the values and combining them into class-wise confidence values $p_k(\omega_j)$, where k is the number of the classifier and ω_j the class, a table containing the confidences of each classifier in the result for a particular class can be formed. The decision of the committee is simply that member classifier's result which has the largest confidence value. To modify the hypothesis, the values $p_k(\omega_j)$ are adjusted when the committee as a whole is incorrect. So when an individual classifier k is correct, the confidence of the

Table 1. Summary of the databases used in the experiments

Database	Subjects	Left-handed	Females	Characters	(a-z,0-9)
DB1	22	1	1	$\sim$ 10 400	8461
DB2	8	0	5	$\sim$ 8 100	4643

result for that classifier is added to the overall confidence of the class for that classifier. On the other hand, when a classifier produces an incorrect result, its total confidence is reduced by the corresponding amount, but not below zero. When the committee produces a correct result, no changes are made. The confidence values were initialized as the inverse of the ordering of the classifiers according to their decreasing recognition performance, ie. $p_k(\omega_j) = \frac{1}{k}$ for all k and j.

2.3 Selecting the currently best classifier

For the sake of comparison a very simple form of committee adaptation was also implemented. The main idea is to select the best classifier for each individual writer by evaluating each classifier's performance during operation and use the result from the classifier that has performed the best up to that point.

3 Experiments

All the committee experiments were run in batch mode simulating on-line operation by taking data in its original order and disallowing reiteration.

3.1 Description of the data sets

The data used in the experiments were isolated on-line characters collected on a Silicon Graphics workstation using a Wacom Artpad II tablet. The data was stored in UNIPEN format [15]. The preprocessing is covered in detail in [2]. The databases are summarized in Table 1, giving the total amount of writers and how many of them were female and left-handed, respectfully, as well as the total amount of characters and characters in the classes used for testing (a-z,0-9).

Database 1 consists of characters which were written without any visual feedback. The pressure level thresholding of the measured data into pen up and pen down movements was set individually for each writer. The distributions of the classes were according to their frequency in the Finnish language.

Database 2 was collected with a program that showed the pen trace on the screen and recognized the characters on-line. The minimum writing pressure for detecting pen down movements was the same for all writers. The distribution of the character classes was approximately even.

Table 2. Recognition error rates of the four committee member classifiers

Classifier	Distance measure	BBC	MC	Error %	Tail error %
1	PL		•	14.9	16.4
2	NPP		•	15.1	15.8
3	NPP	•		18.2	19.1
4	PL	•		19.6	20.9

The databases consisted of different writers. Only lower case letters and digits, a total of approximately 580 characters per writer, were used in the experiments. Database 1 was used for forming the initial user-independent prototype set which consisted of 7 prototypes per class and Database 2 was used as a test set.

3.2 Member classifiers

The experiments were performed using a committee consisting of four individual classifiers. All member classifiers are based on stroke-wise matching between the given character and prototypes. Dynamic Time Warping (DTW) was used to compute both the normalized point-to-point (NPP) and point-to-line (PL) distances [3], one of which was used by each classifier. The NPP distance simply uses the squared Euclidean distance between two data points as the cost function and the total sum is divided by the number of matchings performed. In the PL distance the points of a stroke are matched to lines interpolated between the successive points of the opposite stroke [16]. All samples were scaled so that the longer side of their bounding box was 1000 and the aspect ratio kept unchanged [3]. Also the centers of the character, defined by either the 'Mass center' as the input sample's mass center (MC) or by 'Bounding box' as the center of the sample's bounding box (BBC), is moved to the origin [3]. The configurations and error rates of the member classifiers are shown in Table 2.

In general a committee can be expected to perform the better the less the errors made by its members are correlated. Unfortunately uncorrelatedness is not the case here. As the DTW-based classifier was the only one capable of acceptable recognition performance, all the member classifiers are rather similar. This was confirmed by experiments. For all pair-wise combinations of the four classifiers, the occurrence of the same error is much more common (from 8.1% to 11.7%) than different errors (from 2.2% to 3.3%).

4 Results

Some averages of the effects of the different options on the DEC committee performance have been collected into Table 3. The tail error percentage in the

Table 3. Estimation of the effect of various individual options alone

Parameter	Total error %	Tail error %
default decision: best	12.8	13.2
default decision: majority	13.5	13.6
inclusion required	12.4	12.6
inclusion not required	14.0	14.2
vertical 2^{nd} results	12.1	12.0
horizontal 2^{nd} results	13.5	13.4
no 2^{nd} results	14.1	14.7
just correct conflict resolution	12.9	12.9
correct and wrong conflict resolution	13.0	13.0
inactivate rule conflict resolution	13.8	14.3

Table 4. Comparison with reference classifiers

Combination method	Error %	Tail error %
DEC	11.1	11.3
MCBL	13.0	14.3
Selecting the currently best classifier	14.5	15.0
Non-adaptive Majority Voting	14.6	15.9
Best individual member classifier	14.9	16.4

tables corresponds to the error percentage calculated for the last 200 samples for each writer.

As a conclusion from Table 3, the following can be seen: 1) the default rule of the best classifier outperformed majority voting; 2) requiring the output symbol to be included in the input was in general preferable; 3) second-ranking results should be used in the vertical ordering; 4) the best conflict resolution of rules was based on correct results only.

The results of the adaptive committee classifiers and the non-adaptive majority voting reference as well as the result from the best member classifier are compared in Table 4. The DEC committee employed the best individual classifier base decision rule, vertical second results use and just correct tracking for conflict resolution to obtain this best result. All of the combination methods outperform the best member classifier. The DEC committee clearly outperforms all the other methods used. Also the MCBL committee provides a notable improvement and performs better than the two simpler committee classifiers. Selecting the currently best classifier provides an improvement especially in the tail error percentage in comparison with the majority voting approach.

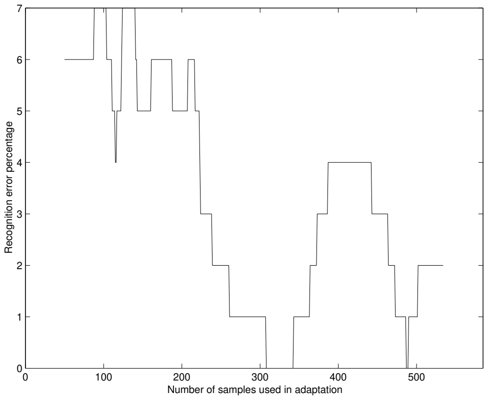

Fig. 2. The evolution of the recognition error rate for one writer from the DEC committee

The evolution of the recognition error rate, calculated within a sliding window of 100 characters, from the DEC committee for an example writer is shown in Figure 2. The average error rate for the writer was 3.2%, but the initial error rate is around 6-7%, and the final level is below 2%.

5 Conclusions

The experiments regarding adaptive committees have shown notable improvements in performance over any of the individual members for both the DEC and MCBL based committee combiners. The most effective combination for the DEC committee was to use the best individual classifier's result as the default rule, use the second results in the vertical manner and use either just the correct results or both correct and incorrect results for conflict resolution.

The next logical stage in the experiments with committee classifiers will be combining the adaptive committee with adaptive member classifiers. Perhaps the simplest way to combine member classifier adaptation and committee adaptation would be to simply first adapt the individual classifiers. The committee adaptation could be started when for example a certain accuracy level has been reached.

A notable problem with on-line adaptation in general is the difficulty of obtaining the correct labels for input samples. As in any real-world application the labeling will ultimately depend on how carefully the user corrects recognition mistakes. Labels can probably never be obtained with 100% correctness. So also the possibility of recovering from errors is something that must be taken into consideration when developing any adaptive on-line recognition system. Adaptive committees may be able to provide more effective error handling mechanisms for such situations and prove beneficial also in this respect.

References

1. Holmström L., Koistinen P., Laaksonen J., Oja E. Neural and statistical classifiers - taxonomy and two case studies. *IEEE Transactions on Neural Networks*, 8:5–17, 1997.
2. Vuori V. Adaptation in on-line recognition of handwriting. Master's thesis, Helsinki University of Technology, 1999.
3. Vuori V., Laaksonen J., Oja E., Kangas J. On-line adaptation in recognition of handwritten alphanumeric characters. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 792–795, 1999.
4. Vuori V., Laaksonen J., Oja E., Kangas J. Experiments with adaptation strategies for a prototype-based recognition system of isolated handwritten characters. *International Journal of Document Analysis and Recognition*, Accepted to be published.
5. Kohonen T. Dynamically expanding context. *Journal of Intelligent Systems*, 1(1):79–95, 1987.
6. Torkkola K., Kohonen T. Correction of quasiphoneme strings by the dynamically expanding context. In *International Conference on Pattern Recognition*, volume 1, pages 487–489, 1988.
7. Laaksonen J., Aksela M., Oja E., Kangas J. Dynamically Expanding Context as committee adaptation method in on-line recognition of handwritten latin characters. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 796–799, 1999.
8. Russell S. J., Norvig P. *Artificial Intelligence: A Modern Approach.* Prentice Hall, 1995.
9. Lam L., Suen C. Y. A theoretical analysis of the application of majority voting to pattern recognition. In *Proceedings of 12th International Conference on Pattern Recognition*, volume II, pages 418–420, Jerusalem, October 1994. IAPR.
10. Bouchaffra D., Govindaraju V. A methodology for mapping scores to probabilities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):923–927, 1999.
11. Paik J., Cho S.b. , Lee K., Lee Y. Multiple recognizers system using two-stage combination. In *Proceedings of International Conference on Pattern Recognition*, pages 581–585. IEEE, 1996.
12. Rahman A., Fairhurst M. A comparative study of decision combination strategies for a novel multiple-expert classifier. In *International Conference on Image Processing and Its Applications*, volume 1, pages 131–135, 1997.
13. Miller D., Yan L. Ensemble classification by critic-driven combining. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1029–1032. IEEE, 1999.
14. Teow L.-N., Tan A.-H. Adaptive integration of multiple experts. In *Proceedings of International Conference on Neural Networks*, volume 3, pages 1215–1220, 1995.
15. Guyon I., Schomaker L., Plamondon R., Liberman M., Janet S. Unipen project of on-line data exchange and recognizer benchmark. In *Proceedings of International Conference on Pattern Recognition*, pages 29–33, 1994.
16. Sankoff D., Kruskal J. B. *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison.* Addison-Wesley, 1983.