Active-DTW : A Generative Classifier that combines Elastic Matching with Active Shape Modeling for Online Handwritten Character Recognition

Muralikrishna Sridhar, Dinesh Mandalapu, Mehul Patel
Hewlett-Packard Labs, Bangalore, India
{*mkrishna, dinesh.mandalapu*}*@hp.com*

Abstract

Developing handwriting recognition systems that are fast and highly reliable is a challenging problem that has become increasingly relevant following the broad acceptance of hand-held devices that use pen based handwriting inputs. Generative classifiers offer a promising solution in combination with discriminative classifiers for addressing this challenge. This paper describes a novel generative classifier - Active-DTW, that combines Active Shape Models with Elastic Matching. Experimental results show that the Active-DTW classifier shows substantial promise as a generative classifier.

Keywords: Active Shape Model, DTW, Eigen-Deformations, generative classifiers

1. Introduction

Research on online handwriting recognition has been receiving renewed attention with the broader acceptance of tablet computers and stylus enabled hand-held devices such as PDAs and cellular phones. Development of Online Character Recognition (OLCR) for these devices has been perceived to be particularly relevant for scripts with large character sets, as keyboard entry for these scripts is often not user friendly. However, creating a robust technological solution that would enable widespread use of handwriting recognition still remains a challenging problem for OLCR researchers.

While each script poses unique challenges with regard to recognition, one may isolate a few important factors that make OLCR a difficult problem for most scripts. Firstly, it may be said that the complexity of recognition often arises from the fact that many character sets tend to share pairs of similar characters with subtle distinguishing characteristics. Another significant reason is that even for the same character, it is often not trivial to capture every possible variation in writing style using the available data. Other factors which complicate the task of recognition are the presence of ambiguous samples, mislabeled data and so on.

In response to these issues, both discriminative and generative approaches have been explored with reasonable success in the last decade of handwriting research. Their complimentary nature has aroused considerable attention in the recent years leading to many hybrid classifiers[2][7].

Discriminative classifiers[1] address the first issue of *between class variations* and try to construct decision boundaries to distinguish classes. On the other hand generative classifiers[3][8] attempt to model the *variations within a class*, so that these variations may be de-emphasized while computing the similarity of the test sample to that class. Generative classifiers are also known to perform better in the face of missing data and outliers as they assume distributions a prior.

This paper introduces a generative classifier called Active-DTW that combines Elastic Matching with Active Shape Models in a novel way. We compare Active-DTW classifier with our implementation of the Eigen-Deformations classifier that is described in the paper by Mitomo et. al.[8], and a DTW based Nearest Neighbor classifier that uses LVQ for prototype selection.

The next section deals with related work on generative classifiers. Section 3 introduces Active Shape Models and Section 4 describes the Active-DTW classifier. Section 5 describes an experimental comparison of the generative classifiers under consideration. Conclusions and possible future explorations are discussed in sections 6 and 7 respectively.

2. Related Work

Generative models have been explored extensively in handwriting recognition research in the last decade. One of the main motivations for their use is that it is possible to perceive handwriting data as having been generated by deforming prototypes according to some distribution. The most popular distribution has been the Gaussian distribution, because of its well known analytical properties for modeling real life data.

Bahlmann and Burkhardt [3] describe a CSDTW classifier in which they modify the distance measure during the DTW matrix computation. The cost measure of dissimilarity between two features is a statistical distance instead of the traditional Euclidean distance, and is computed using the covariance matrix for each feature.

As opposed to calculating the local covariance matrix for each feature in the CSDTW classifier, subspaces are derived for each character class using Principal Component Analysis in the work of Deepu et.al.[4]. One drawback with this technique is that it assumes point to point correspondence between two character samples.

Mitoma et. al. [8] overcomes the above problem by using DTW[5] to align the characters after which they model deformations using PCA.

Active Shape Models are another class of generative classifiers that use PCA to model the variations within a class. Active Shape Models has been used with off-line handwriting data. Gunn et. al [6] use Kernel PCA to construct Non-Linear Active Shape Models in the context of off-line Chinese radical recognition.

3. Active Shape Models

Active Shape Models essentially aim to capture the principal variations of a given set of samples from the mean prototype for that class, so that it becomes possible to generate a new sample by applying these variations to the prototype. Active Shape Models use Principal Component Analysis to statistically model the variations in a low dimensional subspace of the original feature space. The parameters of the Active Shape Model are varied within a range in this subspace to generate new samples.

PCA is a well known statistical technique that constructs a subspace that captures the principal variations in the data. The principal variations are represented using a orthogonal set of eigenvectors. Each eigenvector is a linear combination of features, along the direction in which these features are highly correlated. The first eigenvector represents the direction along which the features are maximally correlated. The second eigenvector representing the direction of second maximum variance is uncorrelated to the first, as it is orthogonal to the first. In this manner PCA forms a subspace using an independent set of eigenvectors.

By projecting the original data onto this subspace, the essential structure of the data can be captured in lower dimensions. Assuming that the data in the original space may be modeled by a Gaussian distribution, the data tends to form a similar distribution in the low dimensional space. The projected data may be visualized as residing in a low dimensional hyper-ellipsoid in this subspace. This hyper-ellipsoid can be described by the eigenvectors which form the principal axes and whose boundaries are given using the corresponding eigenvalues.

Such a model allows us to generate new samples that are valid according to the model and we call these samples valid deformations. Valid deformations may be obtained by applying a weighted combination of the principal variations to the mean prototype, within the limits given by the ellipsoid. These limits are usually of the order of thrice the square root of the eigenvalues. Using these ideas, the Active-DTW Classifier can be derived as follows.

Let n be the dimension of the feature vector. Let $D_c = \{d_{i=1:n}\}$ be a valid deformation vector for class c that can be generated using some parameter vector β. Let $V_n = \{v_{i=1:n}\}$ be the matrix of all eigenvectors and $\Lambda_n = \{\lambda_{i=1:n}\}$ be the vector of eigenvalues . Finally, let μ_c be the mean that corresponds to class c. Given D_c, one can compute the corresponding parameters $\beta_n = \{b_{i=1:n}\}$, by projecting D_c onto the eigenvectors V_n as follows:

$$(D_c - \mu_c).V_n = \beta_n \tag{1}$$

Since D_c is a valid deformation, the following inequality holds uniformly for all the elements of the vectors β_n and $\sqrt{\Lambda_n} = \{\sqrt{\lambda_{1:n}}\}$ respectively.

$$-3\sqrt{\Lambda_n} \le \beta_n \le 3\sqrt{\Lambda_n} \tag{2}$$

We multiply V_n' to Eq. 1. It may be observed that $V_n V_n' = I$, since V_n is an orthogonal matrix. We then take μ_c to the right hand side resulting in the following expression.

$$D_c = \mu_c + \beta_n V_n' \text{ where } -3\sqrt{\Lambda_n} \le \beta_n \le 3\sqrt{\Lambda_n} \tag{3}$$

D_c can be approximated as $\hat{D}_c$ using the first m significant eigenvectors $V = \{v_{1:m}\}$ and eigenvalues $\Lambda = \{\lambda_{1:m}\}$ and corresponding parameters $\beta = \{b_{1:m}\}$ as follows:

$$\hat{D}_c = \mu_c + \beta V' \text{ where } -3\sqrt{\Lambda} \le \beta \le 3\sqrt{\Lambda} \tag{4}$$

This expression defines the Active Shape Model for class c and it allows us to generate any valid deformation for class c by varying the parameters β within the given limits.

4 Active-DTW Classifier

In the Active-DTW classifier, we find the minimum distance of the test sample to each of the N classes. The distance to a class is the DTW distance to the closest deformation that can be obtained from each of the Active Shape models in that class. The closest or optimal deformation D_c^{opt} for each class c is *that* valid deformation whose Euclidean distance to the test sample is minimum.

Let $\underline{Dc}$ be the set of valid deformations $\hat{D}_c$ for class c. If T is the test sample then we have to find the D_c^{opt} that minimizes the following distance:

$$\text{Dist}(c, T) = \min_{\hat{D}_c \in \underline{D_c}} \|(T - \hat{D}_c)\| \tag{5}$$

Using Eq. 3, if β_{opt} are the parameters corresponding to the optimal deformation D_c^{opt}, we have

$$\min_{\hat{D}_c \in \underline{D_c}} \|(T - \hat{D}_c)\| = \|T - (\mu_c + \beta_{opt} V')\| \tag{6}$$

$$\text{where } \beta_{opt} = \arg\min_{-3\sqrt{\Lambda} \le \beta \le 3\sqrt{\Lambda}} \|T - (\mu_c + \beta V')\| \tag{7}$$

The modulo can be expanded as follows:

$$
\begin{aligned}
\|T - (\mu_c + \beta V)\|^2 & \\
&= ((T-\mu_c) - \beta V)((T-\mu_c) - \beta V)' \\
&= \|T-\mu_c\|^2 - 2\beta V(T-\mu_c)' + \|\beta\|^2
\end{aligned}
$$

Therefore the optimization problem can be posed as:

$$\beta_{opt} = \underset{\beta}{arg\ min}(\|\beta\|^2 - 2\beta V(T-\mu_c)' + \|T-\mu_c\|^2) \quad (8)$$

$$\text{subject to} - 3\sqrt{\Lambda} \leq \beta \leq 3\sqrt{\Lambda} \quad (9)$$

By minimizing this expression using constrained quadratic optimization, we can arrive at the parameter β_{opt} that can be used to reconstruct the closest shape to the test sample according to the model of class c. Once β_{opt} is found, the best fitting deformation is found.

$$D_c^{opt} = \mu_c + \beta_{opt} V' \quad (10)$$

At this point, elastic matching is used to compute the DTW-Distance between the test sample and the optimal deformation. This distance which we call as the Active-DTW-Distance would be the minimum distance of the class c to the test sample T, and can be expressed as follows:

$$\text{Active-DTW distance}(T, c) = \text{DTW-Distance }(T, D_c^{opt})$$

The recognized class according to this Active-DTW classifier would be the class with the minimum Active-DTW-Distance to the test sample, i.e. Recognized Class,

$$C = \underset{c=1..N}{arg\ min}(\text{Active-DTW-Distance}(T, c))$$

This formulation thus combines the inherent strengths of generative modeling with elastic matching to yield a better distance measure. This is illustrated in the figure below, using a two dimensional active shape model.

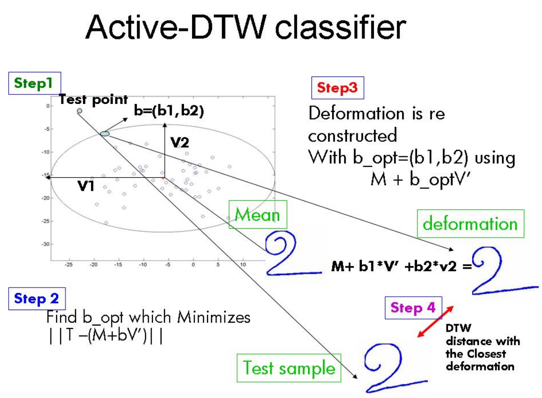

Figure 1. Illustration of the recognition process for the Active-DTW classifier

5 Experiments

In the experiments we have used the IRONOFF[10] database, which contains both online and off-line handwriting information collected by IRCCyN in Nantes, France. The IRONOFF database consists of 4086 isolated digits, 10,685 isolated lower case English characters and 10,679 isolated upper case English characters.

We used the above dataset to compare the Active-DTW classifier with the Eigen-Deformations based classifier. The evaluation on the above dataset is performed using three fold cross-validation.

In order to capture all the variations in style, number of strokes and shape, it is required that each class be represented using more than one model. Hence, samples of each class are clustered using hierarchical clustering[9].

In the case of the Eigen-Deformations classifier, the deformations were estimated for each cluster using the median of the cluster as the reference pattern. For the Active-DTW classifier, Active Shape Models were constructed from each cluster. In order to build subspaces with a certain number of dimensions, we require that the cluster size at least equal to that many dimensions. Therefore, we built models only on those clusters whose sizes were larger than a threshold.

In order to compare the Eigen-Deformations classifier and the Active-DTW classifier, in the first experiment, the samples from the smaller clusters were not used. It may also be noted that while the authors of the Eigen-Deformations classifier[8] have selected the cluster representatives manually and then grouped the samples of a class using the nearest neighbor rule, we have used the hierarchical clustering algorithm to automatically find the clusters. In a second experiment we included samples from the smaller clusters (“free samples”).

While testing, the distance to the nearest cluster of a class c is taken as the distance to the class. The distance of class c to the test sample is computed as the DTW-distance to the closest deformation D_c for the Active-DTW classifier. Fig. 1 uses two dimensions to illustrate how the Active-DTW classifier performs recognition.

The distance measure for the Eigen-Deformations classifier is a combination of the DTW distance to the median of the nearest large cluster and the modified Mahalanobis distance as described in [8].

In addition to the Eigen-Deformations classifier, we also evaluated a k-Nearest Neighbor classifier with DTW as the distance measure. Prototypes for the k-NN classifier were obtained using Learning Vector Quantization. Therefore we refer to this as the LVQ-DTW classifier. The results from the experiments are tabulated in Table 1.

It can be observed from Table 1 that the Active-DTW classifier results in greater accuracy in comparison with the Eigen-Deformations classifier. It can be seen that incorporating free samples does result in an increase in accuracy. This may be explained by noting that while these samples may not form natural hyper-ellipsoidal clusters, they do play an important role in definition of the decision boundaries.

Table 1. Table of accuracies on the IRONOFF data set, comparing Active-DTW with Eigen-Deformations and the LVQ-DTW classifier

Classifiers	Accuracies		
	Digits	**Lower**	**Upper**
EigenDeformations	96.8	85.1	90.0
LVQ-DTW	96.9	84.8	91.6
Active-DTW	97.1	86.9	94.1
Active-DTW (free samples)	97.8	87.7	94.3

It is significant to note that, by being able to incorporate these samples, Active-DTW is able to provide a framework for integrating discriminative classification using samples with high discriminative information.

This is due to the fact that the distance to the model in the case of Active-DTW classifier is the DTW distance to the closest deformation. This distance measure makes it comparable to the DTW-distance between the test sample and any other sample in the training set. It may be noted that this is not possible in the Eigen-deformations based classifier as the distance to the model is not comparable to the distance with other training samples.

In this fashion, the Active-DTW classifier makes it possible to integrate samples with high discriminative power such as the support vectors into its framework.

6 Conclusions

A new classifier called Active-DTW has been proposed in this work. Experimental results have shown that this classifier shows promise as a generative classifier. Being a model based classifier, it has lower space and time complexity compared to a nearest neighbor classifier based on DTW.

In addition, Active-DTW offers the advantage of making it possible to incorporate samples that do not naturally form clusters. As these samples can be chosen to be discriminative(e.g. by choosing support vectors), it becomes possible to incorporate discriminative classification into the Active-DTW framework.

We believe that this is only a preliminary exploration using the Active-DTW classifier, and that there are many promising extensions to improve this work as described below.

7 Future Work

In the future, we would like to further improve the performance of the Active-DTW classifier by exploring several threads. The Active-DTW classifier heavily depends on the clustering scheme as it builds statistical models using these clusters. If the resulting clusters significantly deviate from Gaussian distributions, the PCA based statistical model would not be able to model the distributions . We believe that there are two possible approaches to address this issue. First, algorithms such as the EM algorithm, may be used to obtain better clusters. Second, Kernel PCA[6] may be used to model non-Gaussian distributions.

Another potential improvement to the Active-DTW classifier would be to incorporate DTW mapping into the computation of the covariance matrix and the optimization expression in Eq. 7. This is a promising approach as DTW has shown to obtain a superior distance measure than the Euclidean measure for handwriting data and therefore DTW has been incorporated into well known classifiers such as the Eigen-Deformation classifier[8] and the GDTW classifier[1].

Finally, the Active-DTW classifier may be extended to accommodate discriminative classification by incorporating support vectors into the framework.

Acknowledgements

We are grateful to Dr. Sriganesh Madhvanath and Dr. K.S.R. Anjaneyulu for many valuable discussions. We would like to thank Vijayasenan Deepu and Bharath A for their valuable help.

References

[1] Bahlmann C., Haasdonk B., Burkhardt H., "On-line Handwriting Recognition with Support Vector Machines A Kernel Approch", *International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, 2002,pp. 49-54.

[2] Bellili, A., Gilloux, M. Gallinari, P. "An MLP-SVM Combination Architecture for Offline Handwritten Digit Recognition", *International Journal on Document Analysis and Recognition*, 2003, pp. 244-252.

[3] Claus Bahlman, Hans Burkhardt, "The Writer Independent Online Handwriting Recognition System Frog on Hand and Cluster Generative Statistical Dynamic Time Warping", *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2004, vol. 26, pp. 3.

[4] Deepu Vijayasenan, Sriganesh Madhvanath, "Principal Component Analysis for Online Handwritten Character Recognition", *17th Intl Conf. Pattern Recognition (ICPR 2004)*, Cambridge, United Kingdom, August 23-26, 2004.

[5] Introduction to Dynamic Timewarping, *http://web1.mtnl.net.in/ nilami/dtw.html*.

[6] D. Shi S. R. Gunn, R. I. Damper, "A Radical Approach to Handwritten Chinese Character Recognition Using Active Handwriting Models", *IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pp. 670-675.

[7] Prevost, L., Michel-Sendis, C., Moises, A., Oudot, L., Milgram, M. "Combining Model-based and Discriminative Classifiers: Application to Handwritten Character Recognition", *International Conference on Document Analysis and Recognition*, 2003, pp. 31-35.

[8] Hirito Mitoma, Seiichi Uchida and Hiroaki Sakoe "Online Character Recognition Using Eigen-Deformation", *9th International Workshop on Frontiers in Handwriting Recognition*, 2004.

[9] Vuori, Vuokko, "A Comparison of Techniques for Automatic Clustering of Handwritten Characters", *16th International Conference on Pattern Recognition*, Quebec, Canada, August 2002,volume 3, pp. 168-171.

[10] C. Viard-Gaudin, P. M. Lallican, S. Knerr, and P. Binter. "The IRESTE on/off (IRONOFF) dual handwriting database". *5th International Conference on Document Analysis and Recognition*, 1999, pp. 455–458.