Automatic Detection of Gender and Handedness from On-Line Handwriting

Marcus LIWICKI [a] Andreas SCHLAPBACH [a] Peter LORETAN [a] and Horst BUNKE [a]

[a] *Department of Computer Science, University of Bern*
Neubrückstrasse 10
CH-3012 Bern, Switzerland
{*liwicki, schlpbch, bunke*}*@iam.unibe.ch*

Abstract. In this paper we address the problem of classifying handwritten data with respect to gender and handedness. For the classification we apply state-of-the-art classification methods to distinguish between *male* and *female* handwriting, and *left-* and *right- handedness*. Two classification systems have been evaluated, the first being based on Support Vector Machines and the second being based on Gaussian Mixture Models. Both systems show an improved performance over human-based classification.

1. Introduction

A population of individuals can often be partitioned into sub-categories based on various criteria. Dividing a population into sub-categories is interesting for numerous reasons, for example, if a researcher is only interested in one specific sub-category, or if specifically processing each sub-category leads to improved results. For example, in the field of face recognition, much research has been conducted on classifying a face image according to gender (Wiskott et al., 1995; Wu, Ai and Huang, 2003). Classification results up to 94% have been reported for this two-class problem.

For handwriting there exist several criteria for sub-categories. Whereas in KANSEI the sub-categories are feelings for character patterns (Hattori et al., 2004), handwriting can also be divided into writer-specific sub-categories including gender, handedness, age and ethnicity (Scheidat, Wolf and Vielhauer, 2006). Correlations between these sub-categories and handwriting features have been presented in Huber (1999). Special interest has been focused on determining the gender of the writer. In Hamid and Loewenthal (1996) humans were asked to classify the writer's gender of a given handwriting example. A classification rate of about 68% has been reported. Further studies in Beech and Mackintosh (2005), which inlcude a detailed analysis of the raters background, reported results in the same range.

Beside being an interesting research topic of its own, automatically identifying sub-categories can be used to improve a handwriting recognition system. The variability within a certain category is smaller than within a complete population, which allows us to train specialized recognizers. Another application is demographic studies. A concrete example would be to study the handwriting available on the world wide web and to find out how many people from each category contributed to the data.

Especially the classification of gender from handwriting has been a research topic for many decades (Broom, Thompson and et al., 1929; Newhall, 1926; Tenwolde, 1934). However, there exist conflicting results ranging from slightly more than 50% to more than 90%. An overview of several manual approaches detecting gender from handwriting can be found in Hecker (1996). This thesis tries to semi-automatically classify the handwriting, while it is done automatically in this paper.

Little work exists on automatically identifying sub-categories, such as gender or handedness, from handwriting. In Cha and Srihari (2001) a system for classifying the handwriting based on images of individual letters is presented. Results of 70.2% for gender classification and 59.5% for handedness have been achieved. If longer texts are available and multiple classifier approaches are applied even better results are reported (Bandi and Srihari, 2005). However, these systems are restricted to the off-line case and either the transcription of the text has to be known or even identical texts have to be provided by all writers.

In this paper we present a system that classifies gender and handedness of on-line, Roman handwriting. Both problems are two class problems, i.e., *male/female* and *left-/right-handedness*. On-line handwriting means that temporal information about the handwriting is available. The handwriting is unconstrained, thus any text can be used for classification. Two classifiers are applied to the gender and the handedness classification problem. The first classifier uses Gaussian Mixture Models (GMMs) to model the classes, while the second approach is based on Support Vector Machines (SVMs). Both classifiers are trained using the same set of features extracted from the handwriting. For the purpose of comparison, also an experiment with humans classifying the same data set is performed.

The rest of the paper is organized as follows. Section 2 introduces the features extracted from the on-line data. Section 3 introduces the GMM and the SVM classifiers. Experiments and results are

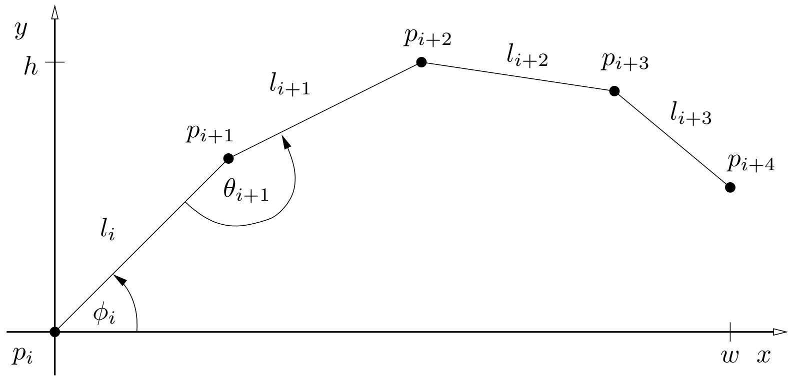

Figure 1. Features extracted from the on-line handwriting.

presented and discussed in Section 4, while Section 5 draws some conclusions and gives an outlook for future work.

2. Features

To acquire the handwritten data, the eBeam [1] interface is used. It outputs a sequence of (x, y)-coordinates representing the location of the tip of the pen together with a time stamp for each location. A series of normalization operations are applied before feature extraction. The operations intend to improve the quality of the features without removing writer specific, i.e. class specific information. The recorded on-line data contain noisy points and gaps within strokes, which are caused by loss of sampling data during acquisition. To recover from artifacts of this kind, two preprocessing steps are applied to the data (Liwicki and Bunke, 2005*a*). The cleaned text is then automatically divided into lines using a simple heuristic rule.

The next step is to divide each text line into sub-parts which can then be normalized independently of each other. For each sub-part the skew angle is corrected to horizontally align the text. Next each sub-part is divided into three regions: the upper area, which contains the ascenders of the letters; the median area with the corpus of the letters; and the lower area containing the descenders of the letters. These three areas are normalized to predefined heights. Finally, the width of each sub-part is normalized. The number of characters is estimated as a fraction of the number of strokes crossing the horizontal line between the base line and the corpus line. The text is then horizontally scaled according to this value.

The feature set used in this experiment contains on-line features as well as features extracted from an off-line representation of the on-line data. (In the remainder of this section, the number in round brackets behind the name of a feature indicates the number of individual features.) For a given stroke s consisting of points p_1 to p_n, the following 18 on-line features for each consecutive pair of points (p_i, p_{i+1}) are computed (see Fig. 1 for an illustration): *speed (1)*; *writing direction (2)*; *curvature (2)*; *normalized x- and y-coordinate (2)*; *speed in x- and y-direction (2)*; *overall acceleration (1)*; *acceleration in x- and y-direction (2)*; *log curvature radius (1)*, which is the length of the circle which best approximates the curvature at the point p_i (Richiardi, Ketabdar and Drygajlo, 2005); *vicinity aspect (1)*, i.e., the aspect of the trajectory in the vicinity of the point p_i; *vicinity curliness (1)*, i.e., the deviation from a straight line in the vicinity of the point p_i; *vicinity linearity (1)*, i.e., the average square distance between every point in the vicinity and the straight line linking the first and the last point in the vicinity; and *vicinity slope (2)*, i.e., the cosine and the sine of the angle of the straight line from the first to the last vicinity point (Jaeger et al., 2001).

The off-line features are computed using a two-dimensional matrix representing an off-line version of the data. The matrix is obtained by projecting the on-line strokes on the two-dimensional plane. The following features are used: *ascenders/descenders (2)*, i.e., the number of points above/below the corpus line whose x-coordinates are in the vicinity of the point and which have a minimal distance to the corpus/base line and *context map (9)*, i.e., the two-dimensional vicinity of the point is divided into three regions for each dimension. The number of black points in each region is taken as a feature value. Overall the feature set consists of 29 features (Liwicki et al., 2006).

3. Classifiers

Two approaches to address the two-class classification problem are taken. The first is a *discriminative approach* where a Support Vector Machine (SVM) directly attempts to maximize the discriminability

[1] eBeam System by Luidia, Inc. - www.e-Beam.com

between classes without modeling or estimating class conditional densities. The second approach is a *generative* one where one Gaussian Mixture Model (GMM) is used to model the distribution of the features of each class.

The idea of a SVM is to separate two different classes of patterns by a maximum margin hyperplane (Burges, 1998). The maximum margin hyperplane is the hyperplane for which the distance to the closest pattern of either class is maximal. Generally, two classes are not linearly separable without error because of outliers and noisy objects. In order to handle such errors, so called slack variables ξ are introduced. These slack variables measure the error in terms of distance to the class boundary. In order to control whether the maximization of the margin or the minimization of the error is more important, a weighting parameter C is defined.

If two classes are not separable in the original input space, the so-called *kernel trick* is applied. It maps the data into a high dimensional feature space and constructs a separating hyperplane with maximum margin there. This is equivalent to a nonlinear decision boundary in the original input space. Using a kernel function, it is possible to compute the separating hyperplane without explicitly carrying out the mapping into the feature space. A kernel function fulfills the condition:

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \sigma(\mathbf{x}), \sigma(\mathbf{y}) \rangle$$

where $\mathbf{x}$ and $\mathbf{y}$ are feature vectors and $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ is a function with $n, m \in \mathbb{N}$ and $m \geq n$, mapping objects from the input to a higher dimensional feature space. Different kernel functions exist, out of which the *linear*, the *radial basis*, the *sigmoid*, and the *polynominal* kernel functions were used in this work. The SVM is implemented using the LIBSVM library (Chang and Lin, 2001).

A GMM models the distribution of the feature vectors extracted from a person's handwriting by a Gaussian mixture density (Reynolds, Quatieri and Dunn, 2000). For a D-dimensional feature vector, $\mathbf{x}$, the Gaussian mixture density is a weighted linear combination of M uni-modal Gaussian densities, p_i, parametrized by a $D \times 1$ mean vector, μ_i, a $D \times D$ covariance matrix, C_i, and a mixture weight, w_i:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\mathbf{x}).$$

The weights w_i sum up to one. The parameters of a writer's density model are denoted as $\lambda = \{w_i, \mu_i, C_i\}$ with $i = 1, \ldots, M$. While the general model supports full covariance matrices, diagonal covariance matrices are used in this paper as they are computationally more efficient while achieving a performance equal to or even better than full covariance matrices (Reynolds, Quatieri and Dunn, 2000).

The GMM is trained using the Expectation-Maximization (EM) algorithm. The EM algorithm iteratively refines the GMM parameters so as to monotonically increase the likelihood of the estimated model for the observed feature vectors. We apply variance flooring to impose a lower bound on the variance parameters (Melin et al., 1998). The GMMs are implemented using the Torch library (Collobert, Bengio and Mariéthoz, 2002).

4. Experiments and Results

The experiments have been conducted on the IAM-OnDB, a large on-line handwriting database acquired from a whiteboard (Liwicki and Bunke, 2005*b*)[2]. This database consists of data from more than 200 writers with eight handwritten texts per writer. Each text consists of seven text lines on average. The classification task is to identify the correct gender or handedness for a given text line.

For the task of gender classification we randomly selected 50 male and 50 female writers for training the classifiers and 25 other writers of each group for testing. For the task of handedness classification we used less data for training and recognition because there are only 20 left-handed writers in the database. The forms of 15 randomly selected left-handed writers and 15 randomly selected right-handed writers have been used for training and the remaining 5 left-handed and 5 other right-handed writers have been used for testing. This assures that both classes are equally distributed in both the training and the test set.

For the SVM algorithm we optimized weighting parameter C and the kernel function with the corresponding parameters. For the GMM the number of Gaussian mixture components and the variance flooring factor have been optimized. The scores of the GMM system were normalized with respect to mean and variance.

Tables 1 and 2 show the classification results for both tasks. The radial basis kernel was the best choice for the SVM approach with the polynomial kernel achieving nearly the same performance. The GMM approach always achieved the highest performance, i.e., 67.06 % for the gender classification and

[2] http://www.iam.unibe.ch/˜fki/iamondb/

Table 1. Gender classification rates

Classifier	performance in %
SVM-linear	61.93
SVM-rbf	62.19
SVM-sigmoid	53.60
SVM-polynomial	62.19
GMM	**67.06**

Table 2. Handedness classification rates

Classifier	performance in %
SVM-linear	54.69
SVM-rbf	62.57
SVM-sigmoid	54.24
SVM-polynomial	61.46
GMM	**84.66**

84.66 % for the handedness classification. These results are statistically significantly higher than random classification (using a *z-test* with $\alpha = 0.05$).

To compare the performance of the classifiers to that of humans, we asked 30 persons to classify 20 images from different writers each. For these images we took five writers from each group, i.e. left-handed female, left-handed male, right-handed female and right-handed male. For "training" purposes, these humans also had classified images from other writers available. The classification rate of the humans is about 57 % for the gender and 62 % for the handedness.

5. Conclusions and Future Work

In this paper we have presented a system for the classification of gender and handedness of handwriting. The data is given in on-line format and we extract a set of 29 features. Two approaches for the classification have been applied, SVMs and GMMs.

In our experiments classification results higher than human classification have been achieved. The GMM results for gender classification are similar to results reported on another data set by Cha and Srihari (2001), but the handedness classification rate is higher. One possible explanation is that the on-line information is important for the handedness classification because some letters, such as the *o*, are often written in another way by left-handed persons than they are written by right-handed persons.

In future work we plan to extend our experiments to other datasets. We also have an off-line writer identification system available which could be used for the classification as well. In the eBeam system there is no information about the pressure of the pen available. From this specific information one could expect higher classification results, especially for the gender detection task (Broom, Thompson and et al., 1929; Hecker, 1996; Newhall, 1926; Tenwolde, 1934).

Acknowledgements

This work was supported by the Swiss National Science Foundation program "Interactive Multimodal Information Management (IM)2" in the Individual Project "Visual/Video Processing", as part of NCCR. Special thanks go to Petra and Michael Liwicki, who provided assistance in getting references from the Humboldt-University, Berlin.

References

Bandi, K. and S. N. Srihari. 2005. Writer Demographic Classification using Bagging and Boosting. In *Proc. 12th Int. Graphonomics Society Conference*. pp. 133–137.

Beech, John R. and Isla C. Mackintosh. 2005. "Do differences in sex hormones affect handwriting style? Evidence from digit ratio and sex role identity as determinants of the sex of handwriting." *Personality and individual differences* 39(2):459–468.

Broom, ME., B. Thompson and et al. 1929. "Sex differences in handwriting." *Journal of Applied Psychology* 13:159–166.

Burges, Christopher J. C. 1998. "A Tutorial on Support Vector Machines for Pattern Recognition." *Data Mining and Knowledge Discovery* 2(2):121–167.

Cha, Sung-Hyuk and Sargur N. Srihari. 2001. Apriori Algorithm for Sub-category Classification Analysis of Handwriting. In *Proc. 6th Int. Conf. on Document Analysis and Recognition*. pp. 1022–1025.

Chang, Chih-Chung and Chih-Jen Lin. 2001. *LIBSVM: a library for support vector machines*. Software available at: `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Collobert, R., S. Bengio and J. Mariéthoz. 2002. Torch: a modular machine learning software library. IDIAP-RR 46 IDIAP.

Hamid, S. and K. M. Loewenthal. 1996. "Inferring gender from handwriting in Urdu and English." *The Journal of social psychology* 136(6):778–782.

Hattori, Tetsuo, Tetsuya Izumi, Hiroyuki Kitajima and Toshinori Yamasaki. 2004. KANSEI Information Extraction from Character Patterns Using a Modified Fourier Transform. In *Proc. Sino-Japan Symposium on KANSEI & Artificial Life.* pp. 36–39.

Hecker, Mandred R. 1996. Die Untersuchung der Geschlechtsspezifitt der Handschrift mittels Rechnergesttzter Merkmalsextraktionsverfahren PhD thesis Humboldt-University, Berlin.

Huber, Roy A. 1999. *Handwriting Identification: Facts and Fundamentals.* CRC Press.

Jaeger, S., S. Manke, J. Reichert and A. Waibel. 2001. "Online handwriting recognition: the NPen++ recognizer." *Int. Journal on Document Analysis and Recognition* 3(3):169–180.

Liwicki, M., A Schlapbach, H. Bunke, S. Bengio, J. Mariéthoz and J. Richiardi. 2006. Writer Identification for Smart Meeting Room Systems. In *Proc. 7th IAPR Workshop on Document Analysis Systems.* Vol. 3872 of *LNCS* Springer pp. 186–195.

Liwicki, M. and H. Bunke. 2005*a*. Handwriting Recognition of Whiteboard Notes. In *Proc. 12th Conf. of the Int. Graphonomics Society.* pp. 118–122.

Liwicki, M. and H. Bunke. 2005*b*. IAM-OnDB - an On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard. In *Proc. 8th Int. Conf. on Document Analysis and Recognition.* Vol. 2 pp. 956–961.

Melin, Håkan, Johan Koolwaaij, Johan Lindberg and Frédéric Bimbot. 1998. A Comparative Evaluation of Variance Flooring Techniques in HMM-based Speaker Verification. In *Proc. 5th Int. Conf. on Spoken Language Processing.* pp. 2379–2382.

Newhall, SM. 1926. "Sex differences in handwriting." *Journal of Applied Psychology* 10:151–161.

Reynolds, D. A., T. F. Quatieri and R. B. Dunn. 2000. "Speaker verification using adapted Gaussian mixture models." *Digital Signal Processing* 10:19–41.

Richiardi, J., H. Ketabdar and A. Drygajlo. 2005. Local and Global Feature Selection for On-line Signature Verification. In *Proc. 8th Int. Conf. on Document Analysis and Recognition.* pp. 625–629.

Scheidat, T., F. Wolf and C. Vielhauer. 2006. Analyzing handwriting biometrics in metadata context. In *Proc. 8th SPIE conference at the Security, Steganography, and Watermarking of Multimedia Contents.* Vol. 6072 pp. 182–193.

Tenwolde, H. 1934. "More on sex differences in handwriting." *Journal of Applied Psychology* 18:705–710.

Wiskott, L., J.-M. Fellous, N. Krger and C. von der Malsburg. 1995. Face recognition and gender determination. In *Proc. Int. Workshop on Automatic Face- and Gesture-Recognition.* pp. 92–97.

Wu, Bo, Haizhou Ai and Chang Huang. 2003. *Audio- and Video-Based Biometric Person Authentication.* Vol. 2688 of *LNCS* Springer chapter *LUT-Based Adaboost for Gender Classification*, pp. 104–110.