In **Proc. 10th Int. Workshop on Frontiers in Handwriting Recognition, La Baule, France, 2006**

Unsupervised Estimation of Writing Style Models for Improved Unconstrained Off-line Handwriting Recognition

Gernot A. Fink

University of Dortmund
Robotics Research Institute
Intelligent Systems Group
Gernot.Fink@udo.edu

Thomas Plötz[*]

Bielefeld University
Faculty of Technology
Applied Computer Science Group
tploetz@techfak.uni-bielefeld.de

Abstract

The performance of writer-independent unconstrained handwriting recognition is severely affected by variations in writing style. In a segmentation-free approach based on Hidden-Markov models we, therefore, use multiple recognition models specialized to specific writing styles in order to improve recognition performance. As the explicit definition of writing styles is not obvious we propose an unsupervised clustering procedure that estimates Gaussian mixture models for writing styles in a completely data-driven manner and thus implicitly establishes classes of writing styles. On a challenging writer-independent unconstrained handwriting recognition task our two stage recognition approach – first performing a writing style classification and then using a style-specific writing model for decoding – achieves superior performance compared to a single style-independent baseline system.

Keywords: unconstrained handwriting, segmentation-free recognition, writing-style model.

1. Introduction

The writer independent recognition of unconstrained handwriting is still an extremely challenging task. In contrast to easier recognition problems, dealing with unconstrained handwriting means that no restrictions on the writing style are imposed. Therefore, data might contain hand printed or cursively written words or a mixture of those two basic writing styles. However, "hand printed" or "cursive" are neither well defined nor can those prototypic styles be considered extremes between which all style variation takes place. Furthermore, it is hard to draw a boundary between writing style variation and the peculiarities in appearance introduced by idiosyncrasies of a specific writer. A few examples of writing styles found in the IAM database [5] are shown in Figure 1

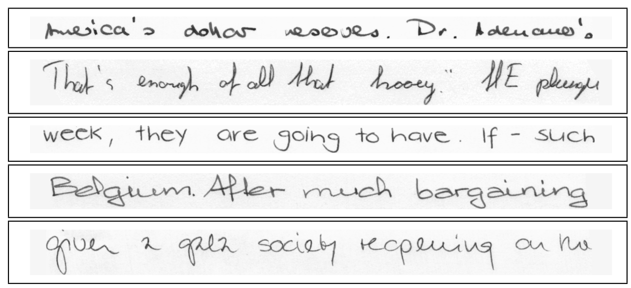

Figure 1. Examples of writing styles

For capturing the large degree of variation statistical approaches to unconstrained handwriting recognition proved very effective. These so-called "segmentation-free" methods first convert words or text lines into linear sequences of feature vectors by applying a local analysis scheme. Subsequently, the feature stream is analyzed by a statistical writing model based on Hidden-Markov Models (HMMs).

Though these approaches are quite successful there is no doubt that a statistical model for unconstrained writer independent data has to account for a substantial variation in the feature representation of handwritten texts, as no effective style normalization procedures are known. The methods for baseline, slant, or size normalization usually applied during preprocessing can only reduce some aspects of style variation. However, the more variation a statistical model has to account for the more its overall modeling quality will be adversely affected.

A very promising method for improving the performance is to use multiple models for describing subsets of the data exhibiting similar types of variation[1]. A well known example from automatic speech recognition is the separate modeling of male and female speech. For handwriting recognition a starting point could be to estimate separate models for hand printed and cursively written

[*]Thomas Plötz is now with University of Dortmund, Robotics Research Institute, Intelligent Systems Group. Contact him at Thomas.Ploetz@udo.edu.

[1]An extreme version of this method would be to train writer-dependent models which is rarely feasible because of insufficient training data and the fact that the identity of writers can not be known exhaustively beforehand in most applications.

text. However, how should mixtures of those basic styles and data not easily associated with a specific writing style be handled?

In order to circumvent the problem of defining writing styles explicitly, we propose an unsupervised procedure for estimating style models automatically from training data. For every writing style identified by this clustering approach a separate writing model is established. In order to avoid problems of data-sparseness the style-specific writing models are adapted versions of a general style-independent baseline model. In the recognition phase first the style models are used to select the appropriate writing model which then is applied to segmenting the data.

In the following section we will first review some relevant related work. Then we will describe the overall approach of writing style specific modeling of unconstrained handwriting. The unsupervised estimation of style models and the creation of style-adapted writing models, which are the two major processing steps involved, are described in detail in section 5 and 6, respectively. Results on a challenging recognition task that demonstrate the effectiveness of our approach are presented in section 7.

2. Related Work

In on-line handwriting recognition many systems use template-like representations of stroke sequences for isolated characters as the writing model. For such a representation it is crucial to capture all relevant variations in appearance of writing – which can be considered to be the style variations – with the prototypical character shapes stored – the so-called allographs. As the manual definition of such allograph models is error prone and requires substantial effort several approaches exist for deriving allograph models automatically by applying a clustering procedure to the training samples available for every character class [1, 2, 10, 9].

In a segmentation-free approach to off-line handwriting recognition a method for deriving allograph HMMs by an unsupervised clustering procedure was proposed in [6]. However, to our knowledge no statistical off-line recognition approaches using an explicit style model are reported in the literature. This is most likely due to the fact that it is not obvious, how such a model could be defined in a segmentation-free framework.

A style model used in a statistical handwriting recognition system would be required to deliver a probability estimate for a given feature vector sequence belonging to some style class. This behaviour can effectively be described by Gaussian mixture models (GMMs). Such models are very popular in automatic speaker recognition where the task is quite similar: Given the representation of speech as feature streams the model is used to decide for one of a set of known speakers (cf. [8]). The GMM-decisions for some broad model category can rather easily be combined with a subsequent segmentation process based on HMMs (cf. e.g. [7]).

3. Baseline Recognition System

Our baseline system for unconstrained handwritten text recognition is a state-of-the-art segmentation-free recognizer. It is based on HMMs and was successfully applied to challenging writer-independent recognition tasks [11, 12].

After text lines are extracted from the document image the handwriting is normalized with respect to skew, baseline orientation, and slant. Additionally, a re-sizing of the line images is performed trying to normalize the character width by scaling the image such that the average distance between local minima of the text contour equals a certain parameter (25 pixels).

After binarization of the normalized text lines frames of constant width (4 pixels) and of the height of the text line are extracted with some overlap (2 pixels). Per frame 9 geometric features (cf. [11]) together with a discrete approximation of their first order derivatives are computed.

The writing model consists of semi-continuous HMMs with Bakis-topology and a varying number of states for context independent characters (both upper and lower case), numerals, punctuation symbols, and white space (75 models in total). The models' emissions in the 18-dimensional feature space are described by state-specific continuous mixtures based on a shared set of component densities (Gaussians with diagonal covariance matrices).

4. Style-based Handwriting Recognition

In order to reduce the variability in handwriting appearance to be captured by a writer-independent recognition system we propose to use multiple models for different writing styles. As the models are then able to specialize on a specific "type" of variability the modeling quality will be increased and, consequently, the recognition error rate will be reduced.

The use of style specific writing models requires two main problems to be solved. First, styles have to be defined and style models have to be estimated robustly. As the explicit definition of writing styles is far from obvious and would also require substantial effort of an expert we decided to solve both aspects of the first problem jointly and in a completely data-driven manner. Second, writing models for specific styles need to be trained on style-specific data. However, the focusing of the parameter estimation process on data from one style class only, necessarily reduces the total number of samples available for training. Therefore, care has to be taken when training style-specific writing models. In contrast to style-independent models that can be trained from scratch on the complete training data available the specific models are derived from the baseline model by applying an adaptation scheme on the style-specific data.

Given a set of style models and corresponding style-specific writing models its application for the recognition of an unknown piece of handwriting is quite obvious. First, the data to be recognized is classified into one of the established style classes by evaluating all available style models in parallel on the feature vector stream. The style model which delivers the optimum score for the test data determines its putative writing style. The actual segmentation of the text in question into words and characters is then computed by using the style-specific writing model identified in the previous style-classification step.

5. Learning Style Models

Reconsidering the baseline system for handwriting recognition including the geometric features used (see section 3) it becomes clear that the majority of writing style variability can be observed at the level of feature vectors. Hardly any normalization (except regarding slope, slant, and size) is performed which stands in contrast to alternative application domains of HMMs, e.g. automatic speech recognition, where more abstract features are used. Thus, the identification of distinct writing styles needs to address the feature data directly.

In our approach we focus on learning explicit style models, namely Gaussian Mixture Models (GMMs), which will be used for partitioning the original set of unconstrained handwritten text with respect to certain writing styles. Thus, learning the models corresponds to a clustering task. Since no prior knowledge regarding these styles is available the learning technique is completely unsupervised. As our data is organized in documents that are split into text lines prior to recognition we analyze it at the document level. Consequently, the aim of the learning process is the detection of clusters within the corresponding "document space".[2]

The estimation of the GMMs is straightforward if the style specific partitioning of the sample data is known in advance. In this case standard approaches for mixture density training, e.g. Expectation Maximization (EM), are applied to the style specific training sets. However, this procedure can not be performed if neither the necessary sample set annotation nor the models required for obtaining this annotation are known beforehand.

In [6] an HMM based, Lloyd like clustering procedure was proposed which simultaneously determines the optimal partition of a training set and, based on this, estimates optimal models, in this case HMMs. Basically, model estimation and evaluation is alternated until no substantial differences between two succeeding annotations can be encountered. Generally, this procedure is suitable for GMMs, too. However, we found that the proposed initialization of cluster models by randomly partitioning the training set is, in our case, not suitable. In fact only two clusters could be established when using the abovementioned procedure which, certainly, does not properly cover the actual writing style variability.

Generally, the initialization of the clustering process is very crucial with respect to the overall quality of the style models. If the partition of the training data used for updating the style model is too general no suitable clusters can be identified because distinct models tend to capture all style variation. In this case most of the training samples will be assigned to a single (generic) style model which is counterproductive for further writing style detection. In order to circumvent this erroneous generalization of particular style models we apply a bootstrapping procedure which establishes reasonable partitionings of the training samples using agglomerative clustering of more abstract GMMs. Given this partitioning the final style GMMs are estimated as in the straightforward approach.

Initially, for all documents of the training set specific GMMs are estimated. Since only limited quantities of training samples are available per "document GMM" the number of individual Gaussians estimated is rather moderate (a few dozens per model). In order to obtain these document GMMs different approaches can be used. As one example we applied the K-means algorithm document specifically resulting in the desired initial models. Alternatively, a global GMM covering all training samples was individually specialized applying Maximum A-Posteriori (MAP) adaptation to every training document.

Given the set of document GMMs, in the next step all training data is re-classified. The newly obtained annotation of the training set can be used for re-initialization or updating the particular models. Both steps are alternated until convergence, i.e. until no substantial differences between two succeeding annotations can be observed. Note that the number of iterations necessary until convergence is usually very small. During clustering similar documents, i.e. those feature sets belonging to similar styles, are captured by identical GMMs. Models that are not assigned to any document during re-classification are discarded. Thus, the initial set of clusters (one per document) is stepwise agglomerated.

The final partition of the training set is the baseline for actually estimating style models. For every cluster of reasonable size (in our case a minimum number of 10 documents per cluster) GMMs including either 128 or 256 Gaussians are estimated using the K-means algorithm.

Based on the final partition of the training samples style specific writing models are created (see next section). In order to use these specialized models reasonably for unknown test data, generally, two application variants are possible. Either style models, i.e. the GMMs, are applied beforehand explicitly subdividing the test set, or the actual writing style decision is performed indirectly by competitively evaluating the style specific writing HMMs.

6. Estimating Style Specific Writing Models

Given the partitioning of the training set obtained from the clustering process as described in the previous section style specific writing models are estimated. The straightforward approach for this is to initialize and train specific writing models, i.e. HMMs, from scratch by individually exploiting the appropriate sample data assigned to the particular styles during clustering.

However, for most practical applications the amount of style specific training data is far from sufficient. Conse-

[2] Note that this specialization is not a prerequisite for our method of learning style models.

quently, robust model estimation can hardly be performed successfully de novo. The beforementioned clustering approach is motivated by the fact that differences between and specialties of certain writing styles become manifest within the feature space. Thus, instead of exploiting style specific data for complete re-estimation writing model specialization can be achieved by modifying the underlying mixture density model of a style independent recognition system using *adaptation* techniques.

One of the most promising approaches for the adaptation of mixture density based feature space representations for (semi-)continuous HMMs is the Maximum Likelihood Linear Regression (MLLR) technique [4]. Originally developed for speaker adaptation of automatic speech recognition systems the modification of the mixtures' mean vectors is achieved using affine transformations. These transformations represent rotations and translations of the feature space estimated on small adaptation sets. They can be generalized to groups of Gaussians including densities not covered by the adaptation set via linear regression.

For style specialization we apply a single regression class MLLR procedure (cf. [3]) to the reference recognition system (see section 3) exploiting the set of appropriate style specific cluster data. This procedure is performed for all clusters obtained in the previous step which, finally, results in a set of style adapted writing models.

After adapting the baseline model towards specialized writing HMMs using MLLR these models can be used directly as recognizers. However, we found that given the adapted systems a complete re-initialization of the particular models followed by further Baum Welch training with integrated MAP adaptation of the mixture densities is more favorable. Compared to the results achieved when directly using the adapted writing models the recognition performance of those adapted models which were further specialized is significantly better.

Intuitively the additional re-initialization and training procedure seems needless since the actual writing style specialization has already been performed by applying the MLLR procedure using style specific training data. However, HMM and mixture density optimization severely depend on proper initialization. Since the style related adaptation "pushes" the models towards "the right direction", i.e. towards the style specialties which are intended to be specifically represented by the particular writing model, further training is reasonable.

7. Experimental Evaluation

In order to evaluate our procedure for automatic writing style estimation with respect to improved handwriting recognition we performed a series of writer-independent recognition experiments on the IAM database [5].

The database consists of several hundred documents of handwritten text scanned at 300 dpi which were generated by having subjects write short paragraphs of text from several different text categories. The documents collected represent truly unconstrained handwriting as no instructions concerning the writing style were given.

We used all documents from text categories A to D (485 documents, 4222 extracted text lines) for training and the documents from categories E and F (129 documents, 1076 extracted text lines) for testing.

During recognition for all experiments performed the use of a lexicon or a statistical language model was deliberately avoided. The reason for this was to judge the effectiveness of the writing style adapted system without possible bias resulting from higher order models. Thus, no restrictions were imposed on the hypothesized character sequences. The performance was measured using the Character Error Rate (CER) of the recognition results with respect to the reference transcription of the data.

In table 1 the results of the particular experiments with special focus on the effectiveness of different configurations of the clustering process are summarized. Compared to the CER of 26.8% measured for the reference system (first row) all systems corresponding to writing style specialization perform significantly better.

Table 1. Results of recognition experiments

# writing styles	Clustering Configuration (Seed Points GMM Estimation)	CER [%]
1	Reference System (no style models)	26.8
	HMM based style decision	
4	K-means estimated GMMs (128 Gaussians)	25.3
5	K-means estimated GMMs (256 Gaussians)	25.1
5	MAP adapted from global GMM	25.8
	GMM based style decision	
4	**K-means estimated GMMs** (128 mixtures)	**24.9**
5	K-means estimated GMMs (256 mixtures)	25.2
5	MAP adapted from global GMM	26.2
2	Lloyd optimized from random (K-means initialization)	25.9

As described in section 5 two variants of exploiting the writing style differentiation can be used for the actual recognition phase ("HMM based style decision" and "GMM based style decision", respectively). Furthermore, different initialization methods of the agglomerative document clustering process aiming at GMM based style models were evaluated (second column). Depending on the clustering method different numbers of writing styles could be extracted (first column). Additionally, in the last row the results for a Lloyd optimization of two randomly initialized writing-style models are given.

Analyzing the figures certain conclusions can be drawn. Compared to the indirect style classification using the specialized writing models the explicit application of the GMM based style models for sub-dividing the test data beforehand is favorable. The most crucial part of the overall style clustering procedure is its proper initialization. Based on document specific GMMs initialized using the K-means algorithm, the agglomerative clustering procedure terminates with 4 suitable clusters. Using this configuration and the GMM based writing style decision relative improvements of approximately 7% compared to the reference system can be achieved. Some randomly selected sample text lines from the four writing-style clusters obtained when using the best performing clustering configuration (bold face in table 1) are shown in Figure 2.

In order to demonstrate the effectiveness of our writing style specialization procedure including model re-initialization and further training certain alternative estimation procedures aiming at writing style models were performed. In table 2 we compare the corresponding figures (CER). For all variants shown the clustering configuration which performed best before (cf. table 1) was used, i.e. 4 writing models were applied.

Table 2. Evaluation results for writing model estimation variants (best clustering configuration used)

Estimation of Writing Models	CER [%]
1. MLLR adaptation of base system 2. Recognition	27.3
1. MLLR adaptation of base system 2. Re-Initialization 3. MAP based Baum-Welch Training 4. Recognition	24.9
1. MLLR adaptation of base system 2. MAP based Baum-Welch Training 3. Recognition	25.9

It can be seen that the application of the additional re-initialization and further model training steps are very effective for the improvement of unconstrained off-line handwriting recognition.

8. Conclusion

In this paper we proposed a new approach to unconstrained off-line handwriting recognition based on the explicit modeling of writing styles. Our method makes use of models for writing-style classification and a corresponding set of style-specific writing models. Style models are realized as GMMs which are estimated via an unsupervised clustering procedure. HMMs are used for modeling handwriting appearance in the feature space of our segmentation-free framework. In contrast to our state-of-the-art baseline system which uses only a single style- and writer-independent writing model, the modeling quality is improved by estimating style-specific writing models per writing style identified by the GMM clustering method. A significant reduction in character error rate of 7% relative on a challenging writer-independent unconstrained handwriting recognition task from the IAM database [5] clearly demonstrates the superior modeling quality achieved by our style-based approach.

References

[1] C. Bahlmann and H. Burkhardt. The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(3):299–310, 2004.

[2] A. Bharath, V. Deepu, and S. Madhvanath. An approach to identify unique styles in online handwriting recognition. In *Proc. Int. Conf. on Document Analysis and Recognition*, volume 2, pages 775–778, Seoul, 2005.

[3] A. Fischer and V. Stahl. Database and online adaptation for improved speech recognition in car environments. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, 1999.

[4] C. J. Leggetter and P. C. Woodland. Maximum Likelihood Linear Regression for speaker adaptation of continuous density Hidden Markov Models. *Computer Speech & Language*, 9:171–185, 1995.

[5] U.-V. Marti and H. Bunke. The IAM-database: An english sentence database for offline handwriting recognition. *Int. Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

[6] M. P. Perrone and S. D. Connell. K-means clustering for Hidden Markov Models. In *Proc. Int. Workshop on Frontiers in Handwriting Recognition*, pages 229–238, Amsterdam, 2000.

[7] T. Plötz and G. A. Fink. Robust time-synchronous environmental adaptation for continuous speech recognition systems. In *Proc. Int. Conf. on Spoken Language Processing*, volume 2, pages 1409–1412, Denver, 2002.

[8] D. A. Reynolds and R. C. Rose. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. on Speech and Audio Processing*, 3(1):72–83, 1995.

[9] V. Vuori. Clustering writing styles with a self-organizing map. In *Proc. Int. Workshop on Frontiers in Handwriting Recognition*, pages 345–350, Niagara on the Lake, Canada, 2002.

[10] L. Vuurpijl and L. Schomaker. Finding structure in diversity: A hierarchical clustering method for the categorization of allographs in handwriting. In *Proc. Int. Conf. on Document Analysis and Recognition*, pages 387–393, Ulm, Germany, 1997.

[11] M. Wienecke, G. A. Fink, and G. Sagerer. Experiments in unconstrained offline handwritten text recognition. In *Proc. 8th Int. Workshop on Frontiers in Handwriting Recognition*, Niagara on the Lake, Canada, August 2002.

[12] M. Wienecke, G. A. Fink, and G. Sagerer. Toward automatic video-based whiteboard reading. *Int. Journal on Document Analysis and Recognition*, 7(2–3):188–200, 2005.

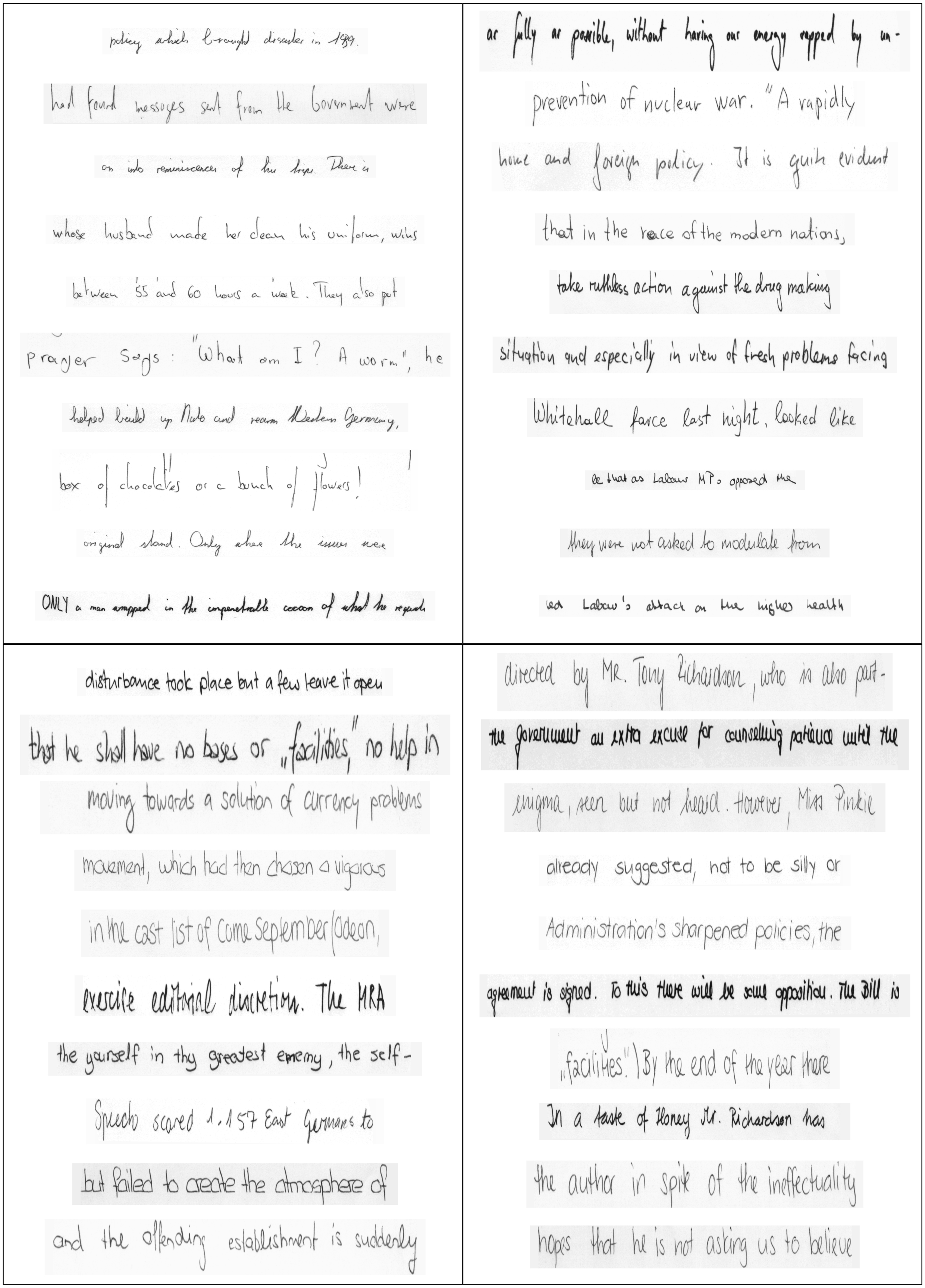

Figure 2. Samples from the 4 automatically generated writing-style clusters.