Handré: An Experimental Approach to Intelligent Character/Word Recognition using Support Vector Machines & Dynamic Time Warping

Sang Woo Jun
wjun@mit.edu

Chong-U Lim
culim@mit.edu

December 5, 2011

Abstract

Optical Character Recognition (OCR) is the term used to describe the process of recognizing scanned images of text from documents, which may be handwritten manuscripts or printed text. An extension to OCR is Intelligent Character Recognition (ICR), which involves additional processing and recognition techniques in order to improve the accuracy of translating such documents by performing recognition on the level of words as opposed to individual characters. In this paper, we present an experimental system to outline the process of performing recognition of an entire handwritten document. Our approach outlines the feasibility of performing word recognition without having to collect hand-written samples for training. We make use of a windowed time-series method and pixel analysis to perform segmentation of the document invididual words and characters respectively. To perform individual character recognition, we used 2 different approaches – a kernel-based support vector machine (SVM) classifier and a dynamic time-warping (DTW) method which were both trained using a database of TrueType fonts. We then perform a committee based process of combining results from both models together with a probablistic spelling checker in order to perform the word recognition.

1 Introduction

2 Segmentation

2.1 Word Segmentation

2.1.1 Time-series Segmentation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

2.1.2 Post-processing

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

2.2 Character Segmentation

2.2.1 Pre-processing

Hand-labeled Training Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

2.2.2 Pixel Analysis

Pixel Analysis Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

3 Recognition

In this section, we describe how we approached the problem of recoginition. In Section 3.1, we describe desribe support vector machines (SVM) and how they can be used for classification of the different alpha-numerical characer by training on a dataset in order to learn a model which generalizes well for all characters.. Section 3.2 covers dynamic time warping, and describes our approach to using it to classify alpha-numerical characters by minimising the distance of difference between the pixel-representations of the different characters. The method of which we combined both the classifers in order to make a committee or grouped decision on the final set of characters in a word is covered in Section 3.3. Finally, we describe the use of a spelling checker which uses a probabalistic modelling of words which was used to perform the finalized word recognition.

3.1 Support Vector Machines for Character Classification

Support Vector Machines (SVMs) are a form of supervized learning methods which can be used for classificaiton or regression problems. In a binary classification example, we would train the SVM on a labelled dataset and if they they are linearly separable, the SVM will find a unique separation boundary in the form of a hyperplane with points falling on each side having different classifications. The separation boundary would be one in which the margin is maximised.

In general, not all data points will be linearly separable often as a result of overlapping class-conditional probabilities. Also, there is a chance of overfitting on the training data points which might negatively affect the generalizability of the classifier for future points. As such, we adopt the use of *slack variables* which results in a 'soft margin', in which we allow some data points to be incorrectly misclassified with a certain penalty with the aim of overcome overfitting. The general formulation of SVMs as constained quadratic programming problem is as follows

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & C\sum_{i=0}^{n} \xi_i + \frac{1}{2} \parallel \theta \parallel^2 \\ \text{subject to} \quad & y_i(\theta \cdot \mathbf{x}_n + \theta_0) \geq 1 - \xi_i,\ i = 1, \ldots, m \end{aligned}$$

where x_i represents each training data point, with y_i being its corresponding target classification. θ is the model, or parameter, of the classifier with offset θ_0, while C and ξ represent the penalty and slack variables respectively.

3.1.1 Mutliple Classes

In our project, we are interested in the use of SVMs to classify individual characters based the a given vector input of pixel values. Extending SVMs to support multiple target classifications requires us to train multiple binary classifiers, one for each individual target character. Given an input vector, each classifier would give a possible classification, and we then employ **voting** as a way of deciding the best classification result for our data. The common voting strategies to decide on a classification described as follows.

One-versus-one In one-versus-one voting, the idea is to fit a classifier for each pair of classes, and when it comes to making the prediction, we select the class with the most number of votes.

One-versus-rest In one-versus-rest voting, each character has a classifier which is fit to it. When making the predicition, the class with the highest classification output is chosen in a *winner-takes-all* strategy.

3.2 K-means Clustering using Dynamic Time-Warping

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

3.3 Combining Approaches for Word Recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

3.4 Dictionary Spelling Corrector

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

4 Experimentation

In this section, we cover the experiements which we undertook in order to apply the segmentation and recognition techniques from Sections 2 and 3. In Section 4.1, we describe the prodedure in which we trained the multi-class SVMs on training data which consisted of different images of each alphabet character for a variety of fonts. Section 4.2 outlines the application of DTW onto a single font set and how it was used to perform its classification. How the results from both experiments were combined together to make committee-based decision is covered in Section 4.3. Finally, Section 4.4 explains the use of the dictionary spelling checker to improve the result and come up with a final decision.

4.1 Multi-class Support Vector Machines

4.1.1 Training Data

For the purpose of our project, we tried to locate a resource of training data for alphabets which was commonly available. Several repositories containing data were either unsuitable or was paid-only. More importantly, we wanted to show-case that system could make use of a generic database for its training data, and not have to resort in having a particular target domain to have to first provide us with both a vast number of examples, but also go through the tedious task of hand-labelling them though it given the resources and time, it might be a possible way to improve the system (Section 6). As such, we decided to using readily available TrueType fonts as our training data.

Training on Fonts We made use of fonts for our training data by randomly picking a set of handwritten fonts which are freely available on the web (see Appendix A.1 for the list and samples.) In order to normalize the characters, we resized each character to fit within a 28x28 square image[1]. The character is then set to black-and-white mode, in which the pixel values are either 0 (black) or 255 (white). For each character in the font, we also created additional samples of the character image by rotating the characters by an angle between [-16°, +16°].

SVM Kernels and Parameters

[1]This is the same dimensions used by the MNIST database, and serves as a reference point for common handwritten character sizes.

4.1.2 Testing on Handwritten Text

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

4.2 K-means Clustering using DTW

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

4.3 Combined Character recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

4.4 Combined Word recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

5 Analysis

5.1 Multi-class Support Vector Machines

5.1.1 Optimal Tuning Parameters

5.2 K-means Clustering using DTW

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

5.3 Combined Character recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

5.4 Combined Word recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

6 Conclusions & Future Work

6.1 Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

6.2 Future Work

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque posuere molestie metus. Suspendisse tellus urna, porta sit amet rutrum eu, tristique quis urna. Donec varius pharetra purus, eget mollis tortor ornare vel. Nullam sagittis tellus id dui placerat eget congue libero facilisis. Donec mattis sagittis lectus, eget porta quam facilisis vel. Vestibulum non urna ante, nec mattis mauris. Nulla sit amet interdum eros. Nam congue lacinia nulla, vitae aliquet nisl tincidunt vel. Morbi gravida bibendum ipsum, at accumsan nisl suscipit ac. Sed accumsan cursus tortor a faucibus. Phasellus tempus, orci ac lacinia hendrerit, dui justo accumsan mi, congue dapibus massa turpis at lectus. Cras a tellus nisi. Aliquam vitae dolor id nunc lacinia fermentum et sit amet metus. Nullam viverra ante eu mauris ultrices nec adipiscing lectus dapibus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam mollis commodo lacus, eget bibendum risus lobortis nec.

7 Acknowledgements

We would like to thank Dahua Lin for reviewing our initial project proposal, and subsequently taking time out of his busy schedule to meet with us and provide guidance, ideas and tips on approaches tp the project. Also, we thank Prof. Leslie Kaelbling for imparting her wisdom and stimulating our ethusiasm in the topics of machine learning.

A Data & Results

A.1 Font Samples

B Project Timeline

Week 1: 10/28 – 11/5

- Submit project proposal
- Narrow down choices to one for the project
- Read up on related papers/books

Week 2: 11/6 – 11/12

- Continue reading related works

- Word Segmentation investigation
- Numerical digit classification investigation

Week 3: 11/13 – 11/19

- Character segmentation
- Dynamic time-warping experiments
- Alphabet characters classification investigation

Week 4: 11/20 – 11/26

- Font database
- Dynamic time-warping for classification
- SVMs for training on fonts and classification

Week 5: 11/27 – 12/5

- Improving individual results
- Heuristics for combining both predictors
- Dictionary spell-correction
- Report writing

C Division of Labor