Combining Neural Networks to Improve Performance of Handwritten Keyword Spotting

Volkmar Frinken, Andreas Fischer, and Horst Bunke

Institute of Computer Science and Applied Mathematics, University of Bern,
Neubrückstrasse 10, CH-3012 Bern, Switzerland
{frinken, afischer, bunke}@iam.unibe.ch

Abstract. Keyword spotting refers to the process of retrieving all instances of a given word in a document. It has received significant amounts of attention recently as an attractive alternative to full text transcription, and is particularly suited for tasks such as document searching and browsing. In the present paper we propose a combination of several keyword spotting systems for unconstrained handwritten text. The individual systems are based on a novel type of neural network. Due to their random initialization, a great variety in performance is observed among the neural networks. We demonstrate that by using a combination of several networks the best individual system can be outperformed.

1 Introduction

The automatic recognition of handwritten text – such as letters, manuscripts or entire books – has been a focus of intensive research for several decades [1, 2]. Yet the problem is far from being solved. Particularly in the field of unconstrained, writer independent handwriting recognition where the writing styles of various writers must be dealt with, severe difficulties are encountered.

Making handwritten texts available for searching and browsing is of tremendous value. For example, one might be interested in finding all occurrences of the word “complain” in the letters sent to a company. As another example, libraries all over the world store huge numbers of handwritten books that are of crucial importance for preserving the world’s cultural heritage. Making these books available for searching and browsing would greatly help researchers and the public alike. Finally, it is worth mentioning that Google and Yahoo have announced to make handwritten books accessible through their search engines [3].

Transcribing the entire text of a handwritten document for searching is not only inefficient as far as computational costs are concerned, but it may also result in poor performance, since misrecognized words cannot be found. Therefore, techniques especially designed for the task of keyword spotting have been developed.

Current approaches to word spotting can be split into two categories, viz. query-by-example (QBE) and query-by-string (QBS). With the former approach, all instances of the search word in the training set are compared with all word images in the test set. Among the most popular approaches in this category
are dynamic time warping (DTW) \cite{4,5,6} and classification using global features \cite{7,8}. Word shape methods using Gradient, Structural and Concavity features (GSC) have been shown to outperform DTW in \cite{9,10}. Algorithms based on QBE suffer from the drawback that they can only find words appearing in the training set. The latter approach of QBS models the key words according to single characters in the training set and searches for sequences of these characters in the test set \cite{11,12}. Recently, keyword spotting systems that are modified versions of handwriting recognition systems have received increasing attention. In \cite{11,13,14}, hidden Markov models are used to find the words to be searched.

To the knowledge of the authors, only single stand-alone keyword spotting systems have been proposed, but no attempt for a combination has been published yet. In this paper we propose to combine several systems using well-known multiple classifier combination techniques. We demonstrate that combining several keyword spotting systems is a convenient method to substantially increase the performance.

In \cite{15} a novel neural network based keyword spotting system is proposed that performs very well. However, a potential problem with this system is the fact that different neural networks vary in their performance due to their random initialization. But an ensemble of well performing keyword spotting systems can be created easily. This renders this technique ideal as a basis to investigate the combination of keyword spotting systems.

The rest of the paper is structured as follows. In Section 2, the underlying neural network based keyword spotting system is introduced. In Section 3, we present the combination methods used for experimental evaluation. The experiments and their results are given in Section 4 and conclusions are drawn in Section 5.

2 The Base Keyword Spotting System

Keyword spotting refers to the process of retrieving all instances of a given word in a document. In this paper, we focus on handwritten documents, such as letters, memos, or manuscripts. Without transcribing the data, a user should still be able to search for any possible word, just like using a search engine. How the results of such a search may look like can be seen in Fig. 1. Note that the base system just returns a likelihood of the word being found. Afterwards, this likelihood can be compared to a threshold to decide whether or not this is a true match.

2.1 Preprocessing

We consider complete text lines as input units for our keyword spotting system. The texts used in the experiments come from the IAM database \cite{16}. They are extracted from pages of handwritten texts, which were scanned and segmented into individual text lines. After binarizing the image with a threshold on the grey scale value, the slant and skew of each textline are corrected and the width and