Pattern Recognition 51 (2016) 112–124

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Combination of global and local contexts for text/non-text classification in heterogeneous online handwritten documents

Truyen Van Phan, Masaki Nakagawa*

Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho Koganei-shi, Tokyo 184-8588, Japan

ARTICLE INFO

Article history:
Received 26 September 2014
Received in revised form
14 April 2015
Accepted 24 July 2015
Available online 4 August 2015

Keywords:
Text/non-text classification
Ink stroke classification
Online handwritten documents
Heterogeneous documents
Recurrent neural networks
Long short-term memory

ABSTRACT

The task of text/non-text classification in online handwritten documents is crucially important to text recognition, text search, and diagram interpretation. It, however, is a challenging problem because of the large amount of variation and lack of prior knowledge. In order to solve this problem, we propose to use global and local contexts to build a high-performance classifier. The classifier assigns a text or non-text label to each stroke in a stroke sequence of a digital ink document. First, a neural network architecture is used to acquire the complete global context of the sequence of strokes. Then, a simple but effective model based on a marginal distribution is used for the local temporal context of adjacent strokes in order to improve the sequence labeling result. The results of experiments on available heterogeneous online handwritten document databases demonstrate the superiority and effectiveness of our context combination approach. Our method achieved classification rates of 99.04% and 98.30% on the Kondate (written in Japanese) and IAMonDo (written in English) heterogeneous document databases. These results are significantly better than others reported in the literature.

1. Introduction

The task of text/non-text classification in online handwritten documents is to classify handwritten strokes into two categories: text and non-text, where a stroke is a time sequence of pen-tip or finger-tip points recorded from pen-down to pen-up. This task can be used as a preprocessing step for text recognition, text search, or diagram interpretation. It is also a prerequisite to the selection of an appropriate engine for processing the handwritten objects further. The classification results are used to decide whether the text strokes should be sent to a handwriting recognizer or an ink search engine. On the other hand, non-text strokes can be grouped together and recognized as higher level graphical entities like flow-chart, finite automata, etc., by a diagram interpreter. Text and non-text classification can be extended for multi-class non-text classification as [1,2], but this paper focuses on text and non-text classification since it is most basic and generic for many applications.

In recent years, smart phones, tablets, tablet PCs, electronic whiteboards equipped with pen-based and touch-based handwriting interfaces have become popular with a vast number of people. Moreover, in the near future, electronic paper and paper-like PCs will become available. People are now able to take notes, draw sketches, and create diagrams on their mobile devices. Online pen-tip traces or finger-tip traces, which are also called digital ink, are a natural and efficient way to express ideas, draw up concepts or summarize knowledge without requiring people to pay any attention to the mode of input. Due to the heterogeneous mixture of text and graphics, however, the advent of digital ink has brought new challenges to document analysis and recognition systems. Many researchers have proposed methods to solve these problems [1,3–9] and several heterogeneous digital ink databases have been collected to evaluate their methods [9–12]. Fig. 1 shows examples of heterogeneous digital ink from commonly used databases: Kondate which is in Japanese [9] and IAMonDo which is in English [10].

In this paper, we propose a novel method to combine global and local contexts of a stroke sequence for the purpose of text/non-text classification in online handwritten documents. Global context refers to the feature vector sequence of an entire document and local context refers to the prediction of directly adjacent strokes. The method is simple but effective since the global context is determined from the output of a neural network and the local context is obtained from the relationship between temporally adjacent strokes. It also establishes a new level of performance for text/non-text classification in heterogeneous online handwritten documents.

The rest of this paper is structured as follows. Section 2 reviews related work on the classification of text and non-text contents in

* Corresponding author. Tel./fax: +81 423 88 7144.
E-mail address: nakagawa@cc.tuat.ac.jp (M. Nakagawa).

http://dx.doi.org/10.1016/j.patcog.2015.07.012
0031-3203/

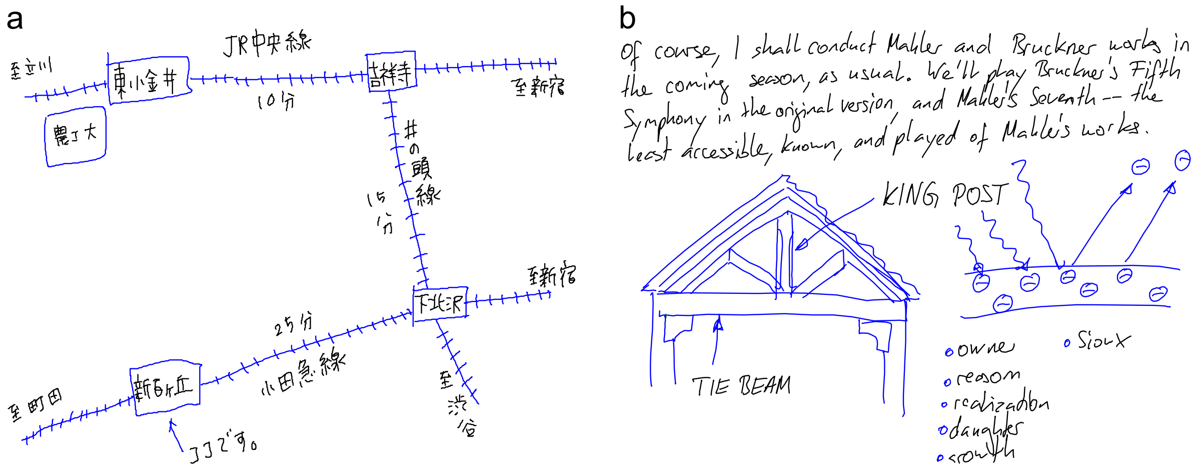

Fig. 1. Examples of heterogeneous digital ink from Kondate (a) and IAMonDo (b). Text strokes are shown in black, non-text strokes are shown in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Performance of various classifiers.

Database		*SVM*	*Our SVM*	*MLP*	*RNN*	*LSTM*	*BRNN*	*BLSTM*
Kondate (*11 features*)	*Rate* (%)	92.58 [4]	92.63	92.78	94.88	95.44	96.34	96.57
	Total time (s)	–	5.31	0.32	0.23	0.83	0.44	1.76
IAMonDo (*19 features*)	*Rate* (%)	94.44 [7]	93.09	93.74	96.57	97.25	97.47	97.72
	Total time (s)	–	20.55	0.31	0.25	0.83	0.44	1.78

heterogeneous online handwritten documents. This survey presents the requirements and ways of using contextual information in the text/non-text classification task. Section 3 overviews our approach, and Section 4 describes the method in detail and introduces recurrent neural network architectures that are used as classifiers. Section 5 evaluates the effectiveness of our approach on the available databases. Section 6 draws the conclusion.

2. Related work

The text/non-text classification task for digital ink is to classify online handwritten strokes into two categories: text and non-text. It is also called text/drawing segmentation [13], text/graphics separation [3,14], text/shape division [15–17], and so on. It is basically a two-category problem but text may be further divided into text and formulas [13], and non-text strokes may be further classified into several categories of graphics [18–21]. Another very similar task is mode detection [22–24]. It allows a user to write text and graphics without specifying or changing the mode in which the user is writing so that each graphic object is assumed to be written without switching to text. The difference between them lies in the input and output of the classification task. The input of text/non-text classification can be any mixed sequence of text and non-text strokes and the output is a sequence of labels for all the strokes, whereas the input of mode detection is a sequence of strokes for text or non-text without a specified mode and the output is the estimate of the mode. Although these tasks or problems are different, they share a number of technologies so both will be reviewed in this section.

2.1. Text/non-text classification

Many methods have been proposed for solving the text/non-text classification problem. They can be divided into three groups: isolated classification, context-integrated classification, and sequence classification according to how contextual information is made use of.

Isolated classification uses descriptions to classify strokes rather than exploiting any contextual information. Jain et al. [1] proposed a linear classifier to distinguish between text and non-text strokes represented by only two features: length and curvature. Isolated classification can also be employed before context-integrated classification [3–9].

Context-integrated classification makes use of several sources of contextual information (local, spatial and temporal) to improve classification accuracy. Mochida et al.'s study [9] was an early attempt. They used the stroke size feature to classify the digital ink and then modified the classification by taking into account stroke crossings. They further classified text into Japanese text and formulas based on the stroke density as well as on text and formula recognition scores. Bishop et al. [3], Zhou and Liu [4], and Delaye et al. [5–7] proposed probabilistic graphical models: Hidden Markov Models (HMMs), Markov Random Fields (MRFs), and Condition Random Fields (CRFs) for better integrating interactions between neighboring strokes. Bishop et al. [3] used a multilayer perceptrons (MLPs) for isolated classification to acquire the probability of a stroke being text or non-text. Then, they used a HMM model to incorporate temporal contexts between adjacent strokes. On the other hand, Zhou and Liu [4] and Delaye et al. [5–7] used support vector machines (SVMs) for isolated classification of single strokes and classification of stroke pairs before the context-integrated classification. The probabilities of single strokes and stroke pairs were created by fitting the SVM outputs to sigmoid functions. Zhou and Liu [4] incorporated spatial interactions between neighboring strokes in their MRF model. Furthermore, Delaye et al. [5–7] integrated multiple sources of context. They used a CRF framework to present the interactions between strokes in terms of neighboring systems and clique potentials. They proposed five neighboring systems describing different sources of contextual information for stroke classification: spatial system, temporal system, intersecting system, lateral system and stroke continuation system. The combination of all these systems performed the best. These methods demonstrate the superiority of