Pattern Recognition Letters 37 (2014) 135–142

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Effective balancing error and user effort in interactive handwriting recognition

N. Serrano [*], J. Civera, A. Sanchis, A. Juan

DSIC, Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain

ARTICLE INFO

Article history:
Available online 23 March 2013

Keywords:
Handwriting recognition
Computer-assisted annotation
Accuracy prediction

ABSTRACT

Transcription of handwritten text documents is an expensive and time-consuming task. Unfortunately, the accuracy of current state-of-the-art handwriting recognition systems cannot guarantee fully-automatic high quality transcriptions, so we need to revert to the computer assisted approach. Although this approach reduces the user effort needed to transcribe a given document, the transcription of handwriting text documents still requires complete manual supervision. An especially appealing scenario is the interactive transcription of handwriting documents, in which the user defines the amount of errors that can be tolerated in the final transcribed document. Under this scenario, the transcription of a handwriting text document could be obtained efficiently, supervising only a certain number of incorrectly recognised words. In this work, we develop a new method for predicting the error rate in a block of automatically recognised words, and estimate how much effort is required to correct a transcription to a certain user-defined error rate. The proposed method is included in an interactive approach to transcribing handwritten text documents, which efficiently employs user interactions by means of active and semi-supervised learning techniques, along with a hypothesis recomputation algorithm based on constrained Viterbi search. Transcription results, in terms of trade-off between user effort and transcription accuracy, are reported for two real handwritten documents, and prove the effectiveness of the proposed approach.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Information has been stored for posterity for centuries. The arrival of the digital era has led to efficient storage and access to this information, but in some cases its latter digestion and analysis present challenging problems. This is the case of handwritten text recognition (HTR). Nowadays, there is a great interest in the study of information stored in manuscripts in libraries all over the world. However, these manuscripts cannot be fully exploited by natural language processing (NLP) tools if transcriptions are not available in an electronic format. Furthermore, transcription of handwritten text documents is an expensive and time-consuming task, which in most cases has to be carried out by paleographic experts. Despite the fact that HTR has been studied since the beginning of Pattern Recognition (PR), current state-of-the-art systems (Graves et al., 2009) still cannot produce fully-automatic high quality transcriptions. This has led to the integration of automatic HTR systems as an assistive tool in the transcription process by experts. The idea behind this integration is to reduce the effort required to generate transcriptions while guaranteeing high levels of accuracy. This approach is commonly referred as computer assisted transcription (CAT).

CAT systems deal with the interactive transcription of a handwritten text document, where the user is continuously aided by a system. The main problem with this approach is that user supervisions have to be efficiently employed, as their overuse may cause the user to ignore the system and transcribe the document manually. In previous works, we have focused on developing techniques to reduce user effort and maximise its utility. For instance, in (Serrano et al., 2009), active learning is used together with semi-supervised learning techniques to adapt (and improve) the system from partially-supervised transcription. Alternatively, in (Serrano et al., 2010a), we developed a technique to improve the current system hypothesis when a user interaction is performed, and thus improve the final transcription. These techniques were implemented on top of an open source interactive prototype called GIDOC (Serrano et al., 2010c).

Although the aim of CAT tools is to save on user effort when transcribing a document, its complete annotation still requires the manual revision of the whole document. It is therefore difficult to measure how much user effort is actually saved when transcribing a document with a CAT tool. In contrast, an alternative approach to CAT is to predefine the desired transcription accuracy after the transcription process. This means that we are accepting

[*] Corresponding author. Tel.: +34 963877350/73533; fax: +34 963877359.
E-mail addresses: nserrano@dsic.upv.es (N. Serrano), jcivera@dsic.upv.es (J. Civera), josanna@dsic.upv.es (A. Sanchis), ajuan@dsic.upv.es (A. Juan).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.03.010

an amount of residual error in our transcriptions in order to save on user effort. For instance, an automatically transcribed document that has been partially supervised by a user may contain a small number of errors but still it can be sufficient to convey the meaning. Similarly, there are many applications dealing with tasks that tolerate erroneous input. For example, the output of an Automatic Speech Recognition (ASR) system can be successfully used as input for well-known tasks such as dialogue act annotation (Stolcke et al., 2000), information retrieval (Grangier et al., 2003), or speech-to-speech translation (Matusov et al., 2006). All these applications may not require perfect annotation of the data, but only a sufficiently good annotation that guarantees the desired accuracy at lower user effort. In this scenario, the ideal CAT tool achieves the required transcription accuracy in exchange of the minimum user effort.

We have studied this latter scenario in the transcription of handwritten text documents (Serrano et al., 2010b) and, more recently, the transcription of speech (Sánchez-Cortina et al., 2012). In these works, we developed a simple yet effective algorithm for estimating the expected error of recognised words that have not been supervised yet. This algorithm was used to adjust the error of transcriptions produced by a CAT system to a given user-defined error threshold. However, even though the described approach guaranteed that the error on the final transcriptions was below the user-defined threshold, it was far too pessimistic and required from the user more effort than was actually needed. In this work, we proposed a new algorithm for predicting the error-rate of recognised words of a HTR system, which outperforms our previous algorithm. This improvement is mainly due to two factors. First, a more precise estimation of the error for each word. Second, the estimation of the error is now performed for a whole block of words, which is more accurate that the previous biased, line by line estimation. This new algorithm will be combined with the best-performing techniques presented in previous works. Our CAT system was evaluated on two real handwritten text documents showing that user effort was closely estimated by the proposed algorithm.

The rest of this paper is organised as follows. First, a brief description of related work is provided in Section 2. In Section 3 we present our new error estimation algorithm. Section 4 shows the empirical results of the proposed approach. Finally, conclusions are drawn and future work is envisioned in Section 5.

2. Related work

The present work deals with the interactive transcription of handwritten text documents, in which a defined quantity of errors in the transcriptions produced can be tolerated in exchange for a substantial savings of manual effort in the annotation process. This approach deals with multiple techniques to successfully complete the task, such as active learning, semi-supervised learning or error-rate prediction. In the following section, we describe the similarity between the diverse components of our approach and previous works, because to our knowledge there are not previous works integrating all the techniques in the same system.

User supervision is typically the most expensive and time-consuming resource in the transcription process. In our case, we deal with the correction of machine-generated output, in which user supervision is only employed to supervise recognised words. Consequently, two problems are tackled in our CAT system. First, the user effort available must be intelligently employed in supervising incorrectly-recognised words, and secondly, unsupervised correctly-recognised words should be identified to be incorporated as training data. The first problem is solved by applying active learning algorithms (Settles, 2009), while the second is solved using semi-supervised learning techniques (Zhu, 2006).

It is worth noting that the combination of active and semi-supervised learning is really necessary for our CAT system to achieve a maximum improvement of transcription accuracy with minimum user effort. Active and semi-supervised learning are used to select the most suitable unannotated samples for user supervision and system adaptation respectively. They can be applied separately or, for better results, in combination, so as to boost their complementary beneficial effect. Indeed, their combination has recently been studied in areas other than HTR, such as ASR (Tur et al., 2005), image retrieval (Zhou et al., 2006) and other fields (Wang and Zhou, 2008). Usually, the key idea behind these learning techniques is the use of confidence measures (CMs) (Wessel et al., 2001; Sanchis et al., 2012) to measure the uncertainty of each hypothesis. In our HTR case, a recognised word with a low confidence value is likely to be an error, whereas a high confidence word is expected to be correctly recognised. Therefore, low confidence words are candidates for supervision, while high confidence words are likely to be useful for system adaptation (re-training).

CAT approaches exploit the impact of user supervision beyond the simplistic idea of correcting incorrectly-recognised words. An incorrectly-recognised word in a given text line, typically affects the surrounding words, generating more errors. When the user supervises a recognised word, the uncertainty of the system around that word is reduced. In this regard, one of the most successful approaches is the prefix-based approach. The main idea of this approach is to improve the system hypothesis on a sample by recomputing the best system hypothesis constrained to a correct prefix. Specifically, first, the user validates the prefix of a system hypothesis up to the first incorrect word, which is corrected. Next, the validated prefix and the user corrected word are employed to predict the remaining suffix by constraining the search process. This process is repeated until the whole transcription has been revised. This approach has been the base of many works dealing with very different applications, such as HTR (Toselli et al., 2007), ASR (Revuelta-Martínez et al., 2012) or syntactic tree annotation (Sánchez-Sáez et al., 2010). All these approaches successfully reduce the effort needed to obtain the required output. However, as mentioned above, the whole machine-generated transcription still has to be revised by a user. Although our approach also follows the idea of constrained search, it must not be confused with the described prefix-based approach. As explained above, in our case we consider a limited amount of user effort, which keep us from supervising the complete output, but only those words that are likely to be wrong. This leads to the supervision of individual words in the output transcription rather than complete prefixes or suffixes. Supervision of individual words saves a significant amount of user effort by focusing user attention on those parts most likely to need correcting. In order to perform a search process constrained to those isolated words supervised by the user, we extrapolated the constrained-Viterbi search proposed by Kristjannson et al. (2004) for information retrieval to HTR.

So far, we have described some techniques to efficiently exploit a limited amount of user supervision. Nevertheless, in our approach, we must first estimate the error-rate of a set of recognised words, to then decide on the supervision effort to achieve the error rate desired by the user. This problem is typically known in the literature as accuracy or error-rate prediction. In the following, we speak in terms of error-rate prediction (EP), as our results are reported in error rate. EP has been typically used on practical applications. In these applications, EP estimation typically employs CMs to validate system performance on a given task. For instance, Schlapbach et al. (2008b) used a EP system based on support vector regression in HTR, in which the estimation is employed to decide if a recognised text is readable enough. Similarly, Yoon et al. (2010) proposed a linear regression of multiple speech features to determine the quality of the English in real oral exams. Another

Download English Version:

https://daneshyari.com/en/article/534551

Download Persian Version:

https://daneshyari.com/article/534551

Daneshyari.com