Handwritting Recognition with Multi-gram language models

Wassim Swaileh[1] **and Thierry Paquet**[1]

[1]LITIS Laboratory - EA 4108 , Normandie University - University of Rouen, Rouen, France,
[1]wassim.swaileh2@univ-rouen.fr, thierry.paquet@univ-rouen.fr

ABSTRACT

We introduce a novel handwriting recognition system based on a sub-word model known as a multi-gram model. The idea is to split a word into a set of variable lengths character's sequences, called multi-grams. A hidden semi Markov model is used to model the multi-grams occurrences within the target language corpus. In decoding, the Markovian model provides multi-grams representations of word samples. We build language corpora of multi-grams that are used to train statistical n-gram language models of multi-grams. The multi-grams lexicon size is significantly lower than the one of a word lexicon which leads to a less complex language model. Our handwriting recognition system is composed of two components, an optical model and statistical n-grams language models. The two parts are combined together during the recognition process using a decoding technique based on weighted finite state transducers (WFST). We experiment our proposal on two Latin language datasets (the French RIMES and English IAM datasets) and we show that it overcomes word and character models when the rate of Out Of Vocabulary words (OOV) is high and that it gets similar performance with a low OOV rate, with the advantage of a reduced complexity.
Keywords: Multigram, Syllable, Handwriting recognition, Recurrent neural network, Weighted finite state transducer

Introduction

The general idea of handwriting recognition systems is to represent the image properties by probabilistic models that are optical models of characters (OCR) of the language to be recognized. Through a learning process, optical models are optimized on a set of ground truth examples in order to achieve the best possible transcription. Recognizing sequences of words within a line of text requires additional linguistic knowledge in addition to the OCR such as a lexicon or a language model. A lexicon provides some rules to provide admissible words[1], while a language model introduces rules between words that constrain the recognition system to find the most probable sequence of words by considering both the optical observations and the language properties. We propose a data driven optimization approach based on Hidden Semi Markov Models (HSMM) which are defined to determine m-multigrams sub-lexical units for any language. A n/m multigram language model is then trained and used for decoding handwritten texts. Here, the OCR is a Bidirectional Long Short-Term Memory (BSLTM) and we use a Weighted Finite State Transducers (WFST) to combine the OCR optical model hypotheses with the multigram language models hypotheses in a decoding automaton framework.

Background

In the literature, we can classify the spoken and written language lexical components into three main categories: lexical units, sub-lexical units and primary units. While lexical units are defined with no ambiguity as the words of the language, sub-lexical units have at least four main definitions in the literature. They are: 1-morphemes, i.e. the decomposition of a word according to its morphological structure, 2- syllables which are defined based on phonological structure of spoken words, 3- graphemic syllables, derived from the phonological syllable for written languages, 4- graphones for spoken languages (or hyphens for written languages) which are language-independent. Some predefined decomposition rules are used to produce these type of sub-lexical units. Besides, both spoken and written languages are based on unambiguous primary units which are either phonemes of spoken languages or characters and alphanumeric symbols of written languages. It is shown in[2] that using hybrid sub-lexical units for German spoken language modeling performs better than using one single type of sub-lexical unit. The main question raised by the literature concerns the feasibility of getting optimal sub-lexical units that may be of hybrid types. From an information theory point of view, sub-lexical units can be viewed as variable-length regularities constructed by streams of primary units[3]. Such components are named multi-grams in the literature[4], as opposed to fixed length n-gram units traditionally used for language modeling.

Method

In the n-multigram model, a word is considered as the concatenation of independent (zero order Markov source) sequences of characters of length k (with $1 \leqslant k \leqslant K$; $K \in \{2,3,4,5\}$). We introduce Hidden Semi-Markov Models (HSMM)[5] for decomposing words into multigrams. The HSMM model is trained to segment a stream of observations into consecutive segments of variable durations by choosing one model state s_k for beeing responsible for generating every multigram of length k. Once trained, looking for the optimal segmentation of a given observation stream can be achieved through Viterbi decoding of the HSMM that look for the most probable HSMM model state sequence. Learning multigrams from data has to be performed considering segmentation is missing. This issue is solved by the Expectation Maximization (EM) algorithm using the computation of Forward and Backward variables.

Our handwriting recognition system consists of two main parts; first part includes four layers recurrent neural network (RNN) optical models that consume the pixel values observed in the input image column normalized to 100 pixels height. At first, skew correction process is applied to the input image. The language models and their associated lexicons represent the second part of our system where we introduce the benefits of the multigram language models to account for OOV. We tested the performance of the system using the conventional words and characters language models and compared them with the performance of four types of multigram language models which are denoted in the tables of results by *mKgram*, where *K* is the maximum number of characters per multigram. State of the art performance was achieved with the proposed multigram language models as well as with the conventional language model of words on both targeted RIMES and IAM data sets. We examine the system performance when training the language models on the French RIMES and English IAM training data sets only, and when training the language models on the French Wikipedia + RIMES training dataset and the English LOB, Brown and Wellington corpus respectively. All the language models are trained with 9-gram order unless the character one of 10-gram order. The state of the art results reported by[6] on RIMES and IAM data sets are 9.6% and 9.3% WER respectively.

Training data sets	Measure	French language models					
		words @ 9gram	m5gram @ 9gram	m4gram @ 9gram	m3gram @ 9gram	m2gram @ 9gram	characters @ 10gram
RIMES training data set	WER%	12.01	11.51	11.12	11.44	11.44	11.51
	OOV%	3.10	1.61	1.16	0.53	0.16	0
	Lex	5.5k	5.2k	4.4k	3k	1.1k	100
Wikipedia + RIMES training dataset	WER%	18.68	13.88	12.29	12.37	10.78	13.47
	OOV%	8.3	4.1	2.3	1.4	0.22	0
	Lex	29k	19.3k	13.8k	6.3k	1.4k	100

Training data sets	Measure	English language models					
		words @ 9gram	m5gram @ 9gram	m4gram @ 9gram	m3gram @ 9gram	m2gram @ 9gram	characters @ 10gram
IAM training data set	WER%	31.88	29.13	28.8	28.59	28.18	30.6
	OOV%	14.1	10.95	7.3	2.2	0.16	0
	Lex	7.4k	6.9k	5.6k	2.9k	0.8k	80
LOB+BROWN+Welling.	WER%	13.66	13.47	14.41	14.1	14.93	20.72
	OOV%	1.4	0.82	0.33	0.06	0	0
	Lex	87.5k	52.7k	37.5k	10.7k	1.7k	80

Table 1. Recognition Word Error Rate (WER) of Words, multigrams & characters language models measured on RIMES & IAM test datasets associated with their OOV rates and lexicons size

References

1. Plamondon, R. & Srihari, S. N. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on pattern analysis machine intelligence* **22**, 63–84 (2000).
2. Shaik, M. A. B., Mousa, A. E.-D., Schlüter, R. & Ney, H. Hybrid language models using mixed types of sub-lexical units for open vocabulary german lvcsr. In *INTERSPEECH*, 1441–1444 (2011).
3. Huang, X., Acero, A., Hon, H.-W. & Foreword By-Reddy, R. *Spoken language processing: A guide to theory, algorithm, and system development* (Prentice hall PTR, 2001).
4. Kuhn, T., Niemann, H. & Schukat-Talamazzini, E. G. Ergodic hidden markov models and polygrams for language modeling. In *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vol. 1, I–357 (IEEE, 1994).
5. Yu, S.-Z. Hidden semi-markov models. *Artif. Intell.* **174**, 215–243 (2010).
6. Voigtlaender, P., Doetsch, P. & Ney, H. Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. *Front. Handwrit. Recognit. (ICFHR), 2016 15th Int. Conf. on* 2167–6445 (2017).