



Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language xxx (2014) xxx–xxx

COMPUTER
SPEECH AND
LANGUAGE

www.elsevier.com/locate/csl

Review

A survey on the application of recurrent neural networks to statistical language modeling[☆]

Wim De Mulder^{a,*}, Steven Bethard^b, Marie-Francine Moens^{a,**}

^a Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium

^b University of Alabama at Birmingham, Birmingham, AL 35294, USA

Received 7 September 2013; received in revised form 10 July 2014; accepted 17 September 2014

Abstract

In this paper, we present a survey on the application of recurrent neural networks to the task of statistical language modeling. Although it has been shown that these models obtain good performance on this task, often superior to other state-of-the-art techniques, they suffer from some important drawbacks, including a very long training time and limitations on the number of context words that can be taken into account in practice. Recent extensions to recurrent neural network models have been developed in an attempt to address these drawbacks. This paper gives an overview of the most important extensions. Each technique is described and its performance on statistical language modeling, as described in the existing literature, is discussed. Our structured overview makes it possible to detect the most promising techniques in the field of recurrent neural networks, applied to language modeling, but it also highlights the techniques for which further research is required.

© 2014 Published by Elsevier Ltd.

MSC: 68T05; 68T10; 68T50

Keywords: Recurrent neural networks; Natural language processing; Language modeling; Speech recognition; Machine translation

Contents

1.	Introduction	00
1.1.	Statistical language modeling	00
1.2.	Recurrent neural networks	00
2.	Outline of the paper	00
3.	Evaluation of language models	00
4.	General description of basic recurrent neural networks	00
5.	Application of basic recurrent neural networks to statistical language modeling	00

[☆] This paper has been recommended for acceptance by E. Briscoe.

* Corresponding author.

** Corresponding author. Tel.: +32 16 32 53 83; fax: +32 16 32 79 96.

E-mail addresses: wim.demulder@cs.kuleuven.be (W. De Mulder), bethard@cis.uab.edu (S. Bethard), sien.moens@cs.kuleuven.be, Marie-Francine.moens@cs.kuleuven.be (M.-F. Moens).

<http://dx.doi.org/10.1016/j.csl.2014.09.005>

0885-2308/© 2014 Published by Elsevier Ltd.

29	5.1.	Corpora	00
30	5.2.	Vocabulary	00
31	5.3.	Training data set	00
32	5.4.	Batch	00
33	5.5.	Description of basic recurrent neural networks applied to the basic SLM task	00
34	5.6.	Parameters	00
35	5.7.	Initial value for $h_i(k, 0)$	00
36	5.8.	Interpretation of the produced output	00
37	5.9.	Training error and validation error	00
38	5.10.	Updating the parameters	00
39	5.11.	Stopping criterion	00
40	5.12.	Detailed training algorithm in pseudocode	00
41	5.13.	Empirical evaluation	00
42	5.14.	Limitations	00
43	5.15.	Recurrent versus feedforward neural networks	00
44	6.	Methods to reduce training time	00
45	6.1.	Limiting the size of the vocabulary	00
46	6.1.1.	Description	00
47	6.1.2.	Empirical evaluation	00
48	6.2.	Hierarchical probabilistic networks	00
49	6.2.1.	Description	00
50	6.2.2.	Empirical evaluation	00
51	6.3.	SOUL network	00
52	6.3.1.	Description	00
53	6.3.2.	Empirical evaluation	00
54	6.4.	Hierarchical subsampling networks	00
55	6.4.1.	Description	00
56	6.4.2.	Empirical evaluation	00
57	6.5.	Higher order gradient descent techniques	00
58	6.5.1.	Description	00
59	6.5.2.	Empirical evaluation	00
60	6.6.	Reservoir computing	00
61	6.6.1.	Description	00
62	6.6.2.	Empirical evaluation	00
63	7.	Methods to dynamically adapt the number of hidden neurons	00
64	7.1.	Growing and pruning neural networks	00
65	7.1.1.	Description	00
66	7.1.2.	Empirical evaluation	00
67	7.2.	Genetic algorithms	00
68	7.2.1.	Description	00
69	7.2.2.	Empirical evaluation	00
70	8.	Methods to increase the context size	00
71	8.1.	Bidirectional neural networks	00
72	8.1.1.	Description	00
73	8.1.2.	Empirical evaluation	00
74	8.2.	Long short-term memory	00
75	8.2.1.	Empirical evaluation	00
76	8.3.	Hierarchical subsampling networks	00
77	9.	Methods to reduce decoding time	00
78	9.1.	Multi-pass strategies	00
79	9.1.1.	Description	00
80	9.1.2.	Empirical evaluation	00
81	9.2.	Variational approximation	00
82	9.2.1.	Description	00
83	9.2.2.	Empirical evaluation	00
84	9.3.	Reduction to simpler models	00

Download English Version:

<https://daneshyari.com/en/article/6951576>

Download Persian Version:

<https://daneshyari.com/article/6951576>

[Daneshyari.com](https://daneshyari.com)