Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Ensembles of Neural Networks for Language Modeling

A thesis presented in partial
fulfilment of the requirements

for the degree of

Master of Philosophy
in Information Technology

at Massey University, Auckland, New Zealand

Yujie Xiao

2018

Contents

1 Introduction **1**
1.1 Objective of Language Modeling 1
1.2 Foundations of the Language Model 3
1.2.1 Word Vector 4
1.2.2 Language Model 6
1.2.3 Assessment of Language Modeling 7
1.3 Scope of this research 8
1.4 Structure of this thesis 9

2 Traditional Approaches for Language Modeling **10**
2.1 The basic language model: N-gram model 10
2.2 Modified N-gram Language Models 11
2.3 Summary 12

3 Neural Network based Language Model **14**
3.1 Neural Networks based Language Model 14
3.2 Feed Forward Neural Networks based Language Models (FFNN-LMs) 15
3.2.1 FFNN-LM 15
3.2.2 Modified FFNN-LMs 17
3.3 Convolutional Neural Networks based Language Models (CNN-LMs) 19
3.4 Recurrent Neural Networks based Language Models 22
3.5 Training of Recurrent Neural Networks 23

4 Our Approach **26**
4.1 Long short term memory (LSTM) 27
4.2 Character-level Convolutional Neural Network 28
4.2.1 1D Convolution 29
4.2.2 Pooling 30
4.3 Highway network 32

4.4 GloVe.6B pre-trained word vector 32

5 Experiments and Results **34**

5.1 Datasets 34

5.2 Keras 34

5.3 Model Setup 34

5.4 Results 35

5.4.1 Perplexity and size 35

5.4.2 convergence speed 37

5.4.3 The dimensions of embedding vector 38

5.4.4 Future works 38

6 Conclusions **40**

Bibliography **41**

List of Figures

1.1 Two example encodings for one sentence "There is a heavy rain tomorrow". (a) sparse feature vector. Each dimension indicates a unique feature, all the values are binary, each value 1 represent one word. (b) Dense vector. Each item in the sentence is represented as a small dimensional vector, which is learned from the model (mapping function) 5

3.1 The framework for first neural network based probabilistic language model, Figure is referenced from [Bengio et al., 2006] 16
3.2 Overview of the enhanced FFNNLM model, referenced from [100] . 18
3.3 The framework for first neural network based probabilistic language model, figure is referenced from [4] 20
3.4 Enhanced framework for convolutional neural network based probabilistic language model, figure is referenced from [4] 21
3.5 Recurrent neutral network's architecture 23
3.6 The unfolded graph of recurrent neutral network 23
3.7 Architecture of a LSTM cell 25

4.1 Architecture of Character level recurrent neural network based language model(with a convolutional layer) 27
4.2 Image Reference : http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/ 29

5.1 Model architecture 36
5.2 Model architecture 38

List of Tables

5.1 Results on Penn Treebank dataset 37
5.2 compared with more models on PTB 37

Abstract

Language modeling has been widely used in the application of natural language processing, and therefore gained a significant amount of following in recent years. The objective of language modeling is to simulate the probability distribution for different linguistic units, e.g., characters, words, phrases and sentences etc, using traditional statistical methods or modern machine learning approach. In this thesis, we first systematically studied the language model, including traditional discrete space based language model and latest continuous space based neural network based language model. Then, we focus on the modern continuous space based language model, which embed elements of language into a continuous-space, aim at finding out a proper word presentation for the given dataset. Mapping the vocabulary space into a continuous space, the deep learning model can predict the possibility of the future words based on the historical presence of vocabulary efficiently than traditional models. However, they still suffer from various drawbacks, so we studied a series of variants of latest architecture of neural networks and proposed a modified recurrent neural network for language modeling. Experimental results show that our modified model can achieve competitive performance in comparison with existing state-of-the-art models with a significant reduction of the training time.

This thesis is organized as follows:
1) Language model has become one central component for various applications about artificial intelligence, therefore, we briefly introduced the objective and basic knowledge of language modeling in Chapter 1.

2) Secondly, we reviewed some closely related literature with our work in Chapter 2 and 3, also point out potential problems of existing models. Variants of Deep Neural Networks (DNNs) models for language modeling are analyzed in Chapter 3, merits and shortcomings are presented, some potential solutions to the shortcomings are also analyzed.

3) Latest popular framework for language modeling and our proposed model are described in Chapter 4. More details about convolutional neural networks and recurrent neural network are showed before we describing our proposed extension

framework based on Recurrent Neural Networks model.

4) Experiments and results are presented in Chapter 4. Overall, results with higher performance have been reported by our proposed framework. The experiments also shed some light for comprehending and interpreting the success of our proposed model for language modelling. We argue that our proposed model perform better than traditional models, due to the ensemble architecture that make it possible to discover the underlying statistical patterns and amplify the performance of RNN's model.

5) We conclude this work in Chapter 6, and predicted the future work. It shows that apart from the high training (computational) complexity, the extension of RNN models are much better than the standardized n-gram and simple neural network based language model in terms of perplexity.

Acknowledgements

Foremost, I would like to express my appreciation and indebtedness to my advisor Prof. Ruili Wang for the continuous support of my Master study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Master study.

Besides my advisor, I would like to thank the rest of my thesis reviewer, for their encouragement, insightful comments, and hard questions. I thank my labmates in Massey University for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the past years.

Last but not the least, I would like to thank my family for supporting me spiritually throughout my life. Also, I would like to thank the coffee machine sitting in my room, without whom it would be impossible to have finished my Thesis.