UNCERTAINTY IN RECURRENT NEURAL NETWORKS

ALIREZA SAMAR

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

Advanced Informatics School
Universiti Teknologi Malaysia

APRIL 2017

ABSTRACT

Deep learning has outperformed in various fields from computer vision, and language processing to physics, biology, and manufacturing. This means the deep or multi-layer architecture of neural networks are being extensively used in these fields; for instance convolutional neural networks (CNN) as image processing tools, and recurrent neural networks (RNN) as sequence processing model.

However, in traditional sciences fields such as physics and biology, model uncertainty is crucial, especially in time series models where delay cant be tolerated. In this work, I aim to propose a novel theoretical framework and develop tools to measure uncertainty estimates, especially in deep recurrent neural networks.

This work also tackles a widely known difficulty of training recurrent neural networks, vanishing gradient by proposing a novel architecture of RNN that compute weighted average unit on past iteration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
RNN	-	Recurrent Neural Network

LIST OF SYMBOLS

γ	-	Whatever
σ	-	Whatever
ε	-	Whatever

CHAPTER 1

INTRODUCTION

1.1 Introduction: The Importance of Uncertainty

The Bayesian approach to machine learning is based on using probability to represent all forms of uncertainty. There are different models like the Gaussian process to understand possible likely and less likely options to generalize the observed data by defining the probability of distributions over functions. This observation and probabilistic models provides the confidence bounds for understanding data and making the decision based on analysis. For instance, an autonomous vehicle would use the determination from confidence bounds to whether brake or not. The confidence bounds simply means ***how certain the model is about its output?***

Understanding whether the chosen model is the right one or not, or the data has enough signals or not is an active field of research [1] in *Bayesian machine learning*, especially in *deep learning models* where based on predictions result it's difficult to make sure about the certainty level of predictions.

1.1.1 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) achieve state-of-the-art performance on a wide range of sequence prediction tasks ([2]; [3]; [4]; [5]; [6]). In this work we shall examine how to add uncertainty and regularisation to RNNs by means of applying Bayesian methods to training. Bayesian methods give RNNs another way to express their uncertainty (via the parameters). At the same time, by using a prior to integrate out the parameters to average across many models during training, this gives a regularisation effect to the network. Recent approaches either attempt to justify dropout [7] and weight decay as a variational inference scheme [8], or apply Stochastic Gradient Langevin dynamics [9] to truncated backpropagation in time directly [10].

Interestingly, recent work has not explored further directly apply a variational Bayes inference scheme for RNNs as was done in practical. We derive a straightforward approach based upon Bayes by Backprop [11] that we show works well on large scale problems. Our approach is a simple alteration to truncated backpropagation through time that results in an estimate of the posterior distribution on the weights of the RNN. Applying Bayesian methods to successful deep learning models affords two advantages: explicit representations of uncertainty and regularisation. Our formulation explicitly leads to a cost function with an information theoretic justification by means of a bits-back argument [12] where a KL divergence acts as a regulariser.

The form of the posterior in variational inference shapes the quality of the uncertainty estimates and hence the overall performance of the model. We shall show how performance of the RNN can be improved by means of adapting ("sharpening") the posterior locally to a batch. This sharpening adapts the variational posterior to a batch of data using gradients based upon the batch. This can be viewed as a hierarchical distribution, where a local batch gradient is used to adapt a global posterior, forming a local approximation for each batch. This gives a more flexible form to the typical assumption of Gaussian posterior when variational inference is applied to neural networks, which reduces variance. This technique can be applied more widely across other variational Bayes models.

1.1.2 Model Confidence

1.1.3 Model Uncertainty and Safety

1.2 Problem Background

1.2.1 Applications of Model Uncertainty

1.3 Problem Statement

1.4 Project Aim

1.5 Project Questions

1.6 Objective and Scope

CHAPTER 2

LITERATURE REVIEW

2.1 State-of-the-Arts

2.2 Limitations

1. Mentor Graphics 2
   (a) item 3
2. item 4

2.3 Research Gaps

The processing at layer-5[1] is done ...

[1] In this thesis, OSI model is used.

CHAPTER 3

RESEARCH METHODOLOGY

CHAPTER 4

PROPOSED WORK

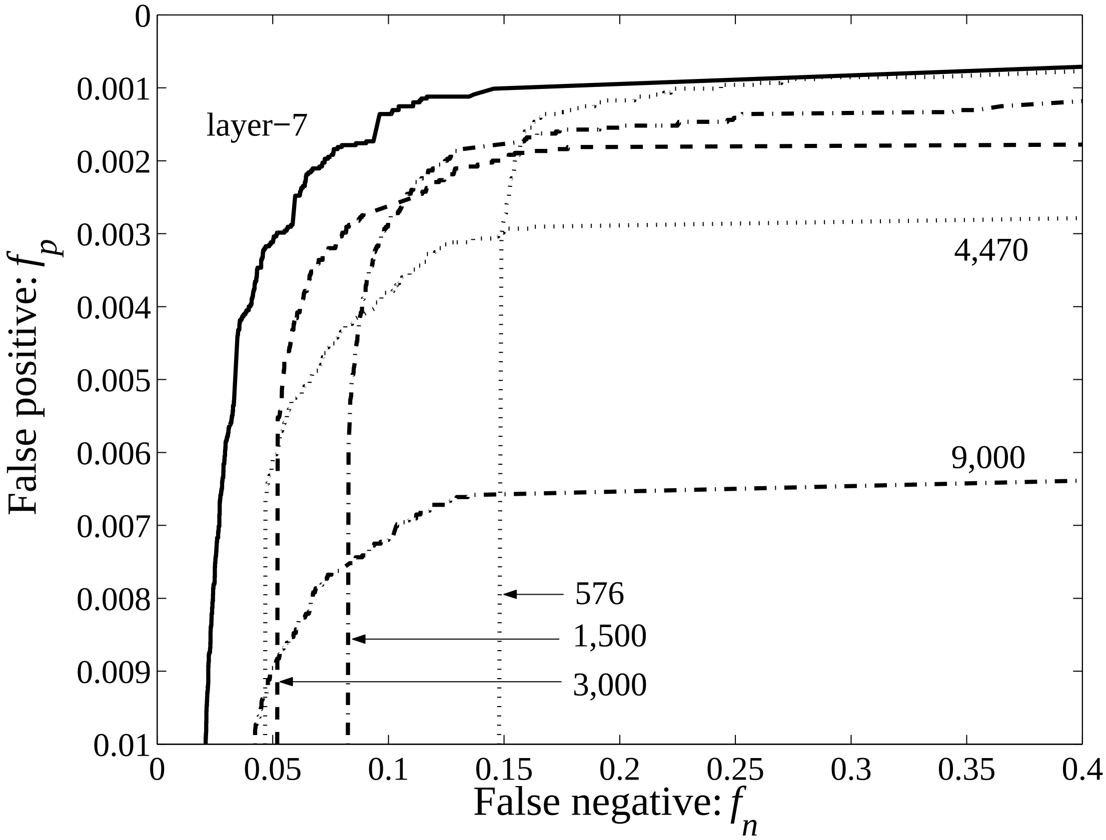

Figure 4.1: Example of a figure. This is a long, very long, long long, long caption. You can give a shorter caption for the "list of figures" using the square braket symbol.

Table 4.1: Example of a table. This is a long, very long, long long, long caption. You can give a shorter caption for the "list of table" using the square braket symbol.

Temperature	Resonant Frequency	Q factor
13 mK $\pm$ 1 mK	16.93	811
40 mK $\pm$ 1 mK	16.93	817
100 mK $\pm$ 1 mK	16.93	815
300 mK $\pm$ 1 mK	16.93	806
500 mK $\pm$ 1 mK	16.93	811
800 mK $\pm$ 5 mK	16.93	814
1000 mK $\pm$ 5 mK	16.93	806

REFERENCES

1. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature*, 2015. 521(7553): 452–459. ISSN 0028-0836. URL http://dx.doi.org/10.1038/nature14541http://10.0.4.14/nature14541.

2. Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M. and Dean, J. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *ArXiv e-prints*, 2016: 1–23. URL http://arxiv.org/abs/1609.08144.

3. Amodei, D., Anubhai, R., Battenberg, E., Carl, C., Casper, J., Catanzaro, B., Chen, J., Chrzanowski, M., Coates, A., Diamos, G., Elsen, E., Engel, J., Fan, L., Fougner, C., Han, T., Hannun, A., Jun, B., LeGresley, P., Lin, L., Narang, S., Ng, A., Ozair, S., Prenger, R., Raiman, J., Satheesh, S., Seetapun, D., Sengupta, S., Wang, Y., Wang, Z., Wang, C., Xiao, B., Yogatama, D., Zhan, J. and Zhu, Z. Deep-speech 2: End-to-end speech recognition in English and Mandarin. *Jmlr W&Cp*, 2015. 48: 28. ISSN 10987576. doi:10.1145/1143844.1143891.

4. Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N. and Wu, Y. Exploring the Limits of Language Modeling. *arXiv:1602.02410 [cs]*, 2016. URL http://arxiv.org/abs/1602.02410{%}5Cnhttp://www.arxiv.org/pdf/1602.02410.pdf.

5. Zaremba, W., Sutskever, I. and Vinyals, O. Recurrent Neural Network Regularization. *Iclr*, 2014. (2013): 1–8. ISSN 0157244X. doi:ng. URL http://arxiv.org/abs/1409.2329.

6. Lu, J., Xiong, C., Parikh, D. and Socher, R. Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning. *1612.01887V1*, 2016. URL http://arxiv.org/abs/1612.01887.

7. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.

Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 2014. 15: 1929–1958. ISSN 15337928. doi: 10.1214/12-AOS1000.

8. Gal, Y. and Ghahramani, Z. Dropout as a Bayesian Approximation : Representing Model Uncertainty in Deep Learning. *Icml*, 2015. 48: 1–10.
9. Welling, M. and Teh, Y.-W. Bayesian Learning via Stochastic Gradient Langevin Dynamics. *Proceedings of the 28th International Conference on Machine Learning*, 2011: 681–688.
10. Gan, Z., Li, C., Chen, C., Pu, Y., Su, Q. and Carin, L. Scalable Bayesian Learning of Recurrent Neural Networks for Language Modeling. *arXiv preprint*, 2016.
11. Blundell, C., Cornebise, J., Kavukcuoglu, K. and Wierstra, D. Weight Uncertainty in Neural Networks. *Icml*, 2015. 37: 1613–1622. URL `http://arxiv.org/abs/1505.05424{%}5Cnhttp://www.arxiv.org/pdf/1505.05424.pdf`.
12. Hinton, G. E., Hinton, G. E., van Camp, D. and van Camp, D. Keeping the neural networks simple by minimizing the description length of the weights. *Proceedings of the sixth annual conference on Computational learning theory - COLT '93*, 1993: 5–13. doi:10.1145/168304.168306. URL `http://portal.acm.org/citation.cfm?doid=168304.168306`.