NN101

Introduction to Neural Net

Hitchhiker's Guide to Neural Network

Author:
Seungwoo Schin

Typeset by:
Seungwoo Schin

KAIST, School of Computing

December 8, 2017

CONTENTS

1. Introduction

This document contains theoretical aspects and actual implementation of various neural nets.

Structure of this document is organized as follows: in section 1, general introduction of machine learning and neural net will be given. From section 2, each neural net will be studied in detail. Implementation by python and its test results will be given for each neural net. For now[1], architectures below will be covered in this document[2]. The list below is sorted in timal order based on the publication date of original pdf, so that reader can follow the historic timeline of artificial neural net development.

- Perceptron/Feedforward Network [1]

- Kohonen Network(KN) [2]

- Boltzmann Machine(BM) [3]

- Restricted BM(RBM) [4]

- Radial Basis Network(RBF) [5]

- AutoEncoder(AE) [6]

- Hopfield Network(HN) [7]

- Recurrent Neural Network(RNN) [8]

- Support Vector Machine(SVM) [9]

- Long/Short Term Memory(LSTM) [10]

- Bidirectional RNN(Bi-RNN) [11]

- Deep Convolutioanl Network(DCNN) [12]

- Liquid State Machine(LSM) [13]

- Echo State Network(ESN) [14]

- Sparse AE(SAE) [15]

- Deep Belief Network(DBN) [16]

- Denoising AE(DAE) [17]

- Deconvolutional Network(DN) [18]

- Variational AE(VAE) [19]

- Markov Chain(MC) [20]

- Gated Recurrent Unit(GRU) [21]

- Generative Adversarial Network(GAN) [22]

- Neural Turing Machine(NTM) [23]

---

[1]2017.12.07
[2]List from wikipedia and Asimov Institute

- Deep Convolutional Inverse Graphics Network(DCIGN) [24]

- Extreme Learning Machine(ELM) [25]

- Deep Residual Network(DRN) [26]

Implementation for each structures will be uploaded on github[3].

After all implementations are covered, optimization of neural networks will be discussed in two perspectives: application of evolutionary algorithms on topology of neural net architecture and evolutionary algorithm on generatlization of activation function.

---

[3]https://github.com/principia12/NN101

2. Background for Machine Learning and Neural Network

I.   Introduction to Machine Learning

II.   Introduction to Neural Network

3.   VARIOUS NEURAL NETWORKS

I.   Perceptron/Feedforward Network [1]

II.   Kohonen Network(KN) [2]

III.   Boltzmann Machine(BM) [3]

IV.   Restricted BM(RBM) [4]

V.   Radial Basis Network(RBF) [5]

VI.   AutoEncoder(AE) [6]

VII.   Hopfield Network(HN) [7]

VIII.   Recurrent Neural Network(RNN) [8]

IX.   Support Vector Machine(SVM) [9]

X.   Long/Short Term Memory(LSTM) [10]

XI.   Bidirectional RNN(Bi-RNN) [11]

XII.   Deep Convolutioanl Network(DCNN) [12]

XIII.   Liquid State Machine(LSM) [13]

XIV.   Echo State Network(ESN) [14]

XV.   Sparse AE(SAE) [15]

XVI.   Deep Belief Network(DBN) [16]

XVII.   Denoising AE(DAE) [17]

XVIII.   Deconvolutional Network(DN) [18]

XIX.   Variational AE(VAE) [19]

XX.   Markov Chain(MC) [20]

XXI.   Gated Recurrent Unit(GRU) [21]

XXII.   Generative Adversarial Network(GAN) [22]

XXIII.   Neural Turing Machine(NTM) [23]

XXIV.   Deep Convolutional Inverse Graphics Network(DCIGN) [24]

XXV.   Extreme Learning Machine(ELM) [25]

XXVI.   Deep Residual Network(DRN) [26]

4. APPLICATION OF EVOLUTION ALGORITHM ON NEURAL NETWORKS

I. Neural Network Architecture Generation via Evolution Algorithm

II. Activation Function Optimization via Evolution Algorithm

References

[1] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[2] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

[3] G E. Hinton and Terrence Sejnowski. Learning and relearning in boltzmann machines. 1, 01 1986.

[4] P. Smolensky. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Information Processing in Dynamical Systems: Foundations of Harmony Theory, pages 194–281. MIT Press, Cambridge, MA, USA, 1986.

[5] M. J. D. Powell. Algorithms for approximation. chapter Radial Basis Functions for Multivariable Interpolation: A Review, pages 143–167. Clarendon Press, New York, NY, USA, 1987.

[6] H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. Manuscript M217, Philips Research Laboratory, Brussels, Belgium, 1987.

[7] J. J. Hopfield. Neurocomputing: Foundations of research. chapter Neural Networks and Physical Systems with Emergent Collective Computational Abilities, pages 457–464. MIT Press, Cambridge, MA, USA, 1988.

[8] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

[9] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.

[10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[11] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, November 1997.

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, volume 86, pages 2278–2324, 1998.

[13] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.

[14] H. Jaeger and H. Haas. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication, 2004.

[15] Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann L. Cun. Efficient learning of sparse representations with an energy-based model. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1137–1144. MIT Press, 2007.

[16] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 153–160. MIT Press, Cambridge, MA, 2007.

[17] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA, 2008. ACM.

[18] Matthew D. Zeiler, Dilip Krishnan, Graham W. Taylor, and Rob Fergus. *Deconvolutional networks*, pages 2528–2535. 2010.

[19] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013.

[20] Brian Hayes. First links in the markov chain. *American Scientists*, 101(2):92, 2013.

[21] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555, 2014.

[22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[23] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *CoRR*, abs/1410.5401, 2014.

[24] Tejas D. Kulkarni, William F. Whitney, Pushmeet Kohli, and Joshua B. Tenenbaum. Deep convolutional inverse graphics network. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, pages 2539–2547, Cambridge, MA, USA, 2015. MIT Press.

[25] Shifei Ding, Han Zhao, Yanan Zhang, Xinzheng Xu, and Ru Nie. Extreme learning machine: Algorithm, theory and applications. *Artif. Intell. Rev.*, 44(1):103–115, June 2015.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. cite arxiv:1512.03385Comment: Tech report.