Deep Bayesian Mining, Learning and Understanding

Jen-Tzung Chien
National Chiao Tung University
Hsinchu, Taiwan
jtchien@nctu.edu.tw

ABSTRACT

This tutorial addresses the advances in deep Bayesian mining and learning for natural language with ubiquitous applications ranging from speech recognition to document summarization, text classification, text segmentation, information extraction, image caption generation, sentence generation, dialogue control, sentiment classification, recommendation system, question answering and machine translation, to name a few. Traditionally, "deep learning" is taken to be a learning process where the inference or optimization is based on the real-valued deterministic model. The "semantic structure" in words, sentences, entities, actions and documents drawn from a large vocabulary may not be well expressed or correctly optimized in mathematical logic or computer programs. The "distribution function" in discrete or continuous latent variable model for natural language may not be properly decomposed or estimated. This tutorial addresses the fundamentals of statistical models and neural networks, and focus on a series of advanced Bayesian models and deep models including hierarchical Dirichlet process, Chinese restaurant process, hierarchical Pitman-Yor process, Indian buffet process, recurrent neural network (RNN), long short-term memory, sequence-to-sequence model, variational auto-encoder (VAE), generative adversarial network (GAN), attention mechanism, memory-augmented neural network, skip neural network, stochastic neural network, predictive state neural network, policy neural network. We present how these models are connected and why they work for a variety of applications on symbolic and complex patterns in natural language. The variational inference and sampling method are formulated to tackle the optimization for complicated models. The word and sentence embeddings, clustering and co-clustering are merged with linguistic and semantic constraints. A series of case studies are presented to tackle different issues in deep Bayesian mining, learning and understanding. At last, we will point out a number of directions and outlooks for future studies.

CCS CONCEPTS

• **Mathematics of computing → Bayesian computation**; • **Computing methodologies → Natural language processing**; **Neural networks**.

KEYWORDS

deep learning; Bayesian learning; natural language processing

KDD '19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08.
https://doi.org/10.1145/3292500.3332267

ACM Reference Format:
Jen-Tzung Chien. 2019. Deep Bayesian Mining, Learning and Understanding. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3292500.3332267

1 INTRODUCTION

Given the current growth in research and related emerging technologies in machine learning and deep learning [35], it is timely to introduce this tutorial to a large number of researchers and practitioners who are attending KDD 2019 and working on statistical models, deep neural networks, sequential learning and natural language understanding. This half-day conventional tutorial concentrates on a wide range of theories and applications and systematically present the recent advances in deep Bayesian learning which are impacting the communities of machine learning, data mining, natural language processing and human language technology. This tutorial is useful to the graduate students who work in natural language processing and understanding, and the research scientists who would like to explore statistical data mining, machine learning and deep learning. The prerequisite knowledge includes calculus, linear algebra, probability and statistics.

2 TUTORIAL DESCRIPTION

The presentation of this tutorial is arranged into five parts. First of all, we share the current status of researches on natural language processing, statistical modeling and deep neural network and explain the key issues in deep Bayesian learning for discrete-valued observation data and latent semantics. Modern natural language models are introduced to address how data analysis is performed from language processing to semantic learning and memory networking. Secondly, we address a number of Bayesian models ranging from latent variable model to variational Bayesian inference [2, 5–7, 33] and Bayesian nonparametric learning [1, 3, 4] for hierarchical, thematic and sparse topics from natural language. In the third part, a series of deep models including deep unfolding [10], GAN [19, 28], memory network [11, 29], sequence-to-sequence learning [18, 21], convolutional neural network [14, 23, 34] and attention network with transformer [15, 31] are introduced. The fourth part focuses on a variety of advanced studies which illustrate how deep Bayesian learning is developed to infer the sophisticated recurrent models for natural language understanding. In particular, the Bayesian RNN [8, 17], VAE [12], neural variational learning [13, 27], neural discrete representation [22, 30], recurrent ladder network [25, 26], stochastic neural network [9, 16, 20], Markov recurrent neural network [24, 32], reinforcement learning and sequence GAN [36] are introduced in various deep models which open a window to more practical tasks, e.g. reading comprehension, sentence generation, dialogue system, question answering and machine translation. In the final part, we spotlight on some future directions for deep language understanding which can handle the challenges of big data, heterogeneous condition and dynamic system. In particular, deep learning, structural learning, sequential learning and stochastic learning are emphasized.

3 INSTRUCTOR

Jen-Tzung Chien is now with the Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan, where he is currently the University Chair Professor. He held the visiting researcher position with the IBM T. J. Watson Research Center, Yorktown Heights, NY, in 2010. His research interests include machine learning, deep learning, natural language processing and computer vision. He served as the associate editor of the IEEE Signal Processing Letters in 2008-2011, the guest editor of the IEEE Transactions on Audio, Speech and Language Processing in 2012, the organization committee member of ICASSP 2009, the area coordinator of Interspeech 2012, EUSIPCO 2017, 2018, 2019, the program chair of ISCSLP 2018, the general chair of MLSP 2017, and currently serves as an elected member of the IEEE Machine Learning for Signal Processing (MLSP) Technical Committee. He received the Best Paper Award of IEEE Automatic Speech Recognition and Understanding Workshop in 2011 and the AAPM Farrington Daniels Award in 2018. Dr. Chien has published extensively including the books “Bayesian Speech and Language Processing”, Cambridge University Press, in 2015, and “Source Separation and Machine Learning”, Academic Press, in 2018. He has served as the Tutorial Speaker for ICASSP 2012, 2015, 2017, Interspeech 2013, 2016, APSIPA 2013, ISCSLP 2014, COLING 2018, AAAI 2019, ACL 2019, and IJCAI 2019.

ACKNOWLEDGMENTS

This work was partially supported by the Ministry of Science and Technology, Taiwan, under MOST 108-2634-F-009-003.

REFERENCES

[1] Jen-Tzung Chien. 2015. Hierarchical Pitman-Yor-Dirichlet Language Model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 8 (2015), 1259–1272.
[2] Jen-Tzung Chien. 2015. Laplace Group Sensing for Acoustic Models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 5 (2015), 909–922.
[3] Jen-Tzung Chien. 2016. Hierarchical Theme and Topic Modeling. *IEEE Transactions on Neural Networks and Learning Systems* 27, 3 (2016), 565–578.
[4] Jen-Tzung Chien. 2018. Bayesian Nonparametric Learning for Hierarchical and Sparse Topics. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26, 2 (2018), 422–435.
[5] Jen-Tzung Chien and Ying-Lan Chang. 2014. Bayesian Sparse Topic Model. *Journal of Signal Processing Systems* 74, 3 (2014), 375–389.
[6] Jen-Tzung Chien and Chuang-Hua Chueh. 2011. Dirichlet Class Language Models for Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 3 (2011), 482–495.
[7] Jen-Tzung Chien and Chuang-Hua Chueh. 2012. Topic-Based Hierarchical Segmentation. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (2012), 55–66.
[8] Jen-Tzung Chien and Yuan-Chu Ku. 2016. Bayesian Recurrent Neural Network for Language Modeling. *IEEE Transactions on Neural Networks and Learning Systems* 27, 2 (2016), 361–374.
[9] Jen-Tzung Chien and Kuan-Ting Kuo. 2017. Variational Recurrent Neural Networks for Speech Separation. In *Proc. of Annual Conference of International Speech Communication Association*. 1193–1197.
[10] Jen-Tzung Chien and Chao-Hsi Lee. 2018. Deep Unfolding for Topic Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 2 (2018), 318–331.
[11] Jen-Tzung Chien and Ting-An Lin. 2018. Supportive Attention in End-To-End Memory Networks. In *Proc. of IEEE International Workshop on Machine Learning for Signal Processing*. 1–6.
[12] Jen-Tzung Chien and Chun-Wei Wang. 2019. Variational and Hierarchical Recurrent Autoencoder. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*. 3202–3206.
[13] Junyoung Chung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C Courville, and Yoshua Bengio. 2015. A Recurrent Latent Variable Model for Sequential Data. In *Advances in Neural Information Processing Systems*. 2980–2988.
[14] Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. 2017. Language Modeling with Gated Convolutional Networks. In *Proc. of International Conference on Machine Learning*. 933–941.
[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* (2018).
[16] Marco Fraccaro, Søren Kaae Sø nderby, Ulrich Paquet, and Ole Winther. 2016. Sequential Neural Models with Stochastic Layers. In *Advances in Neural Information Processing Systems*. 2199–2207.
[17] Yarin Gal and Zoubin Ghahramani. 2016. A Theoretically Grounded Application of Dropout in Recurrent Neural Networks. In *Advances in Neural Information Processing Systems*. 1019–1027.
[18] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional Sequence to Sequence Learning. In *Proc. of International Conference on Machine Learning*. 1243–1252.
[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*. 2672–2680.
[20] Anirudh Goyal, Alessandro Sordoni, Marc-Alexandre Côté, Nan Ke, and Yoshua Bengio. 2017. Z-Forcing: Training Stochastic Recurrent Networks. In *Advances in Neural Information Processing Systems 30*. 6713–6723.
[21] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. In *Proc. of International Conference on Machine Learning*. 369–376.
[22] E. Jang, S. Gu, and B. Poole. 2017. Categorical Reparameterization with Gumbel-Softmax. In *Proc. of International Conference on Learning Representations*.
[23] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. In *Proc. of Annual Meeting of the Association for Computational Linguistics*. 655–665.
[24] Che-Yu Kuo and Jen-Tzung Chien. 2018. Markov Recurrent Neural Networks. In *Proc. of IEEE International Workshop on Machine Learning for Signal Processing*. 1–6.
[25] Isabeau Prémont-Schwarz, Alexander Ilin, Tele Hao, Antti Rasmus, Rinu Boney, and Harri Valpola. 2017. Recurrent Ladder Networks. In *Advances in Neural Information Processing Systems*. 6011–6021.
[26] Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. 2015. Semi-Supervised Learning with Ladder Networks. In *Advances in Neural Information Processing Systems*. 3546–3554.
[27] Iulian V. Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2017. A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues. In *Proc. of AAAI Conference on Artificial Intelligence*. 3295–3301.
[28] Jen-Chieh Tsai and Jen-Tzung Chien. 2017. Adversarial Domain Separation and Adaptation. In *Proc. of IEEE International Workshop on Machine Learning for Signal Processing*. 1–6.
[29] Kai-Wei Tsou and Jen-Tzung Chien. 2017. Memory Augmented Neural Network for Source Separation. In *Proc. of IEEE International Workshop on Machine Learning for Signal Processing*. 1–6.
[30] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. 2017. Neural Discrete Representation Learning. In *Advances in Neural Information Processing Systems*. 6309–6318.
[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*. 5998–6008.
[32] Arun Venkatraman, Nicholas Rhinehart, Wen Sun, Lerrel Pinto, Martial Hebert, Byron Boots, Kris Kitani, and J Bagnell. 2017. Predictive-State Decoders: Encoding the Future into Recurrent Networks. In *Advances in Neural Information Processing Systems*. 1172–1183.
[33] Shinji Watanabe and Jen-Tzung Chien. 2015. *Bayesian Speech and Language Processing*. Cambridge University Press.
[34] Shi Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*. 802–810.
[35] Dong Yu, Geoffrey Hinton, Nelson Morgan, Jen-Tzung Chien, and Shigeki Sagayama. 2011. Introduction to the Special Section on Deep Learning for Speech and Language Processing. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (2011), 4–6.
[36] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient. In *Prof. of AAAI Conference on Artificial Intelligence*, Vol. 31. 2852–2858.