Research Statement

Seyed Abbas Hosseini

June 2021

My background and research interest lies in the intersection of Machine Learning, Scalable Data mining and Machine Learning Systems and my main application domain is recommendation and personalization. Based on these background and interest, I have the following plan for research and teaching in next three years. First, I discuss interesting research topics in each of these areas and then discuss the courses that I have plan to offer in next years.

1 Research Plan

My main research interest lies in designing end-to-end machine learning systems that can efficiently work in real world contexts. My current focus is on search, recommendation and personalization application domains and my goal for next three years is to develop a large-scale recommendation system that can recommend to online users based on their history and their context or search query. However, designing such system, requires innovation and technical breakthrough in two main directions:

1. Accurate modeling of online user behavior using efficient statistical and machine learning models and designing scalable data mining algorithms that enable us to apply the proposed ML models on real world scenarios.

2. Developing large scale end-to-end machine learning systems that can train the proposed statistical models on massive scale datasets and streaming data and evaluate and monitor the model performance continuously and provide real time prediction

1.1 Statistical Machine Learning

During my PhD, my research focused on user behavior modeling by continuous time modeling of events using Bayesian methods. This led to a series of papers [1, 2, 3, 4] which significantly advanced the state of the art in the time-sensitive recommendation and churn prediction . Bayesian methods provide us with a very flexible framework to manage uncertainty in complex systems and hence is an appropriate approach for modeling complex systems. However, there are many direction to improve the proposed methods for user behavior modeling and prediction:

- In many real world application such as recommendation, the ML method gives feedback and can be retrained to adapt itself to changes in the environment. However, in this case, the data that we have is affected by our previous recommendations and hence a biased model may lead us to a very non-optimize situation. The exploration-exploitation tradeoff is an inherent problem in such situations and we always need a policy to resolve the tradeoff efficiently. Bayesian is a very appropriate approach for such problems since it provide the uncertainty that the model has beside the latent variable estimations. Contextual multi-armed bandit [5] and (deep) reinforcement learning are established frameworks for this problem and combining them with Bayesian approach can result in very efficient policies to resolve exploration-exploitation tradeoff.

- Deep neural networks have shown superior performance in many applications. Although these models have so many parameters, but they can generalize on unseen data very well and estimate their labels efficiently. However, one of the limitations of these models, is that they don't provide the uncertainty about their estimations. It can mainly cause problems in scenarios such as recommendation where the dataset is affected by its previous estimations. Bayesian deep neural networks [6] is a category of DNNs that are able to solve this problem. However, there are still many challenges with applying such models in large scale or real time problems which make it a very interesting research topic.

1.2 Machine Learning Systems

In traditional ML research, it usually assumed that there is a dataset that has been previously cleaned and we should propose a model with a very good generalization power so that it works for test data when it is trained on training set. However, in real world scenarios, there is nothing as clean data and usually the model should be able to estimate or predict the desired information very fast and with a very high throughput. Moreover, usually there are many concept drifts that make a model inefficient through time and hence it should be able to adapt itself by continuously training on new provided data. Therefore, we need so many components such as data cleaner, feature store, infrastructure to continuously training, evaluating, monitoring, integrating and serving the model [7]. Also, in many contexts there are privacy concerns that we may not be able to even collect the data and we need to train the model without having the data collected in one place!

Moreover, as it was mentioned in section 1.1, Bayesian is one our main approaches for modeling complex real systems. Although inference in Bayesian models are based on marginalization and conditioning, but exact inference is intractable in almost all complex models and we need approximate inference methods such as variational approximation [8] or sampling based methods such as MCMC or SMC [9] and hence for each model we need to derive a customized inference algorithm which is very time consuming. Emerging programming frameworks such as Pytorch and Tensorflow immensely reduce time to market of neural models by automating the inference on such models. Although there have been some endeavor for automating the inference on Bayesian methods, but there is a long way to standardize and stabilize these libraries [10] and hence we need software systems to make the inference on Bayesian models automatic so that we can test different models on our problems easier and faster.

All these interesting topics are covered under the name of ML systems which is one of our main topics of focus.

References

[1] Seyed Abbas Hosseini, Ali Khodadadi, Ali Arabzadeh, and Hamid R Rabiee. Hnp3: A hierarchical nonparametric point process for modeling content diffusion over social media. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 943–948. IEEE, 2016.

[2] Seyed Abbas Hosseini, Keivan Alizadeh, Ali Khodadadi, Ali Arabzadeh, Mehrdad Farajtabar, Hongyuan Zha, and Hamid R Rabiee. Recurrent poisson factorization for temporal recommendation. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 847–855. ACM, 2017.

[3] Ali Khodadadi, Seyed Abbas Hosseini, Erfan Tavakoli, and Hamid R Rabiee. Continuous-time user modeling in presence of badges: A probabilistic approach. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(3):37, 2018.

[4] Ali Khodadadi, Seyedabbas Hosseini, Ehsan Pajouheshgar, Farnam Mansouri, and Hamid R Rabiee. Choracle: A unified statistical framework for churn prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[5] Mark Collier and Hector Urdiales Llorens. Deep contextual multi-armed bandits. *arXiv preprint arXiv:1807.09809*, 2018.

[6] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[7] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. *Advances in neural information processing systems*, 28:2503–2511, 2015.

[8] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[9] Adrian Smith, Arnaud Doucet, Nando de Freitas, and Neil Gordon. *Sequential Monte Carlo methods in practice*. Springer Science & Business Media, 2013.

[10] Dustin Tran, Matthew D. Hoffman, Rif A. Saurous, Eugene Brevdo, Kevin Murphy, and David M. Blei. Deep probabilistic programming. In *International Conference on Learning Representations*, 2017.