Latent Explanation and Critiquing for Conversational Recommender Systems

ESC499 Thesis Proposal

Hanze Li
1002526493

Supervised by: Professor Scott Sanner

An undergraduate thesis submitted in fulfillment of the requirements
for the degree of Bachelor of Applied Science

In the
Data Driven Decision Making Lab
Department of Mechanical and Industrial Engineering

Engineering Science (Robotics)
University of Toronto

Introduction

Critiquing is a method for conversational recommendation that adapts recommendations in response to user preference feedback. In this setting, user will be iteratively provided with item recommendation and attribute description for that item: a user may either accept or critique the attributes to generate new recommendation.

Explainable recommendation refers to personalized recommendation algorithms that provide recommendation results and explanations, offering interpretability and transparency to help understand, diagnose and refine recommendation algorithm [1]. Explanation algorithms are developed and evaluated together with critiquing to improve recommendation performance.

Objective

In the above settings, we would like to build iterative critiquing algorithms that (1) minimize the number of iterations before reaching users' satisfactory (i.e. recommend users with their target item), (2) understand user/item's intrinsic characteristic as explanation to our recommendation.

Related Works

Historical critiquing works were largely based on constraint- and utility-based methods for modifying recommendations w.r.t. these critiqued attributes, including incremental critiquing that consider the cumulative effect of iterated critiquing interactions[2], and experience-based methods that attempts to collaboratively leverage critiquing interactions from multiple users[3] etc. Some recent works also explored speech- and dialog-based interfaces for critiquing-style frameworks [4]. Latent embeddings as critiquing method are also revisited and built upon in the thesis [8, 9, 10, 11]

Approach

Existing algorithms that leverage latent user preference representation to update recommendations are often nontrivial to interpret and debug given the nature of learned representations from linear regressions [6] or auto-encoders [10]. Considering the following two scenarios where

1. a user with some preferences showed in history, can we understand this user's preference with a personalized keyphrases, and interpret why we made this explanation from latent representations?
2. a user might show preference to Japanese restaurants from user preference history and made a critique against spicy cuisines, would the updates from the latent representation show that this user appears more similar to users who likes non-spicy cuisines including possibly Japanese food?

Given the above questions, we are interested in looking into how *simple*, *interpretable*, *yet empirically comparable (against deep-learning based formulation)* algorithms such as matrix-based similarity framework with Nearest-Neighbour can trace both the source of explanations and the direct impact of critiquing. We would also like to further explore the Attention based algorithms to properly weight multi-step critique feedback that are not necessarily independent, nor of equal weights.

We would like to build and test the above-mentioned ideas based on three publicly available dataset: Yelp academic review, BeerAdvocate, and Amazon CDs&Vinyl, all contains more than 100,000 reviews and product rating records.

Reference

1. Yongfeng Zhang and Xu Chen. 2018. Explainable recommendation: A survey and new perspectives. arXiv preprint arXiv:1804.11192 (2018).
2. James Reilly, Kevin McCarthy, Lorraine McGinty, and Barry Smyth. 2004. Incremental critiquing. In International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, 101–114.
3. Kevin McCarthy, Yasser Salem, and Barry Smyth. 2010. Experience-based critiquing: Reusing critiquing experiences to improve conversational recommendation. In International Conference on Case-Based Reasoning. Springer, 480–494.
4. Ga Wu, Kai Luo, Scott Sanner, and Harold Soh. 2019. Deep Language-based Critiquing for Recommender Systems. In Proceedings of the 13th ACM Conference on Recommender Systems (RecSys-19). Copenhagen, Denmark.
5. Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Darius Braziunas. 2016. On the effectiveness of linear models for one-class collaborative filtering. In Thirtieth AAAI Conference on Artificial Intelligence.
6. Xia Ning and George Karypis. 2011. Slim: Sparse linear methods for top-n recommender systems. In 2011 IEEE 11th International Conference on Data Mining. IEEE, 497–506.
7. D. Bahdanau, K. Cho and Y. Bengio. 2019. Neural Machine Translation by Jointly Learning to Align and Translate. CoRR, abs/1409.0473
8. Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 173–182.
9. Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In Proceedings of the 2018 World WideWeb Conference (WWW'18). Republic and Canton of Geneva, Switzerland, 689–698.
10. S. Sedhain, A. Menon, S. Sanner, and L. Xie. 2015. AutoRec: Autoencoders Meet Collaborative Filtering. In Proceedings of the 24th International Conference on the World Wide Web (WWW-15). Florence, Italy.
11. Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. 2016. Collaborative denoising auto-encoders for top-n recommender systems. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. ACM, 153–162.