Recommender System

Applications

e-commerce (Amazon, Alibaba), entertainment (Netflix, YouTube), social networking (Twitter, Facebook)

Challenges

- **large scale/high volume**: e.g. 2nd quarter active users on Alibaba reached 828M [2], over 75.1M products are being sold on Amazon [3] => parallelizable/scalable algorithm
- **real-time**: fast inference
- **non-static**:
  - data are being generated every moment, online algorithm
  - trends and user preferences can vary over time
- **cold start**: insufficient data for newer user

Methods

Content-based

create a profile for each user or item based on attributes such as demographic, age, gender, genre, box office, etc.

pro: fully trained offline => fast inference

cons:

- poor accuracy
- relies heavily on domain knowledge

Collaborative filtering

recommendation is based on the similarity score (cosine measure)

- **user-user**: each user is characterized by their past behavior w.r.t. all items (purchase/transaction history, rating records)
- (Amazon) **item-item** [4]: each item is characterized by all users' actions (e.g. like/dislike, positive/negative rating)

pros:

- more accurate
- partially trained offline

cons:

- computation intensive: often requires pairwise distances for ranking
- suffers from cold start

Latent factor model

- (Netflix) SVD++ [5]: an extension of SVD which can be applied on sparse interaction data between user and item; considers user-specific and item-specific bias
  - optimization: SGD or iterative least square
- Factorization Machine (FM) [6]: concatenates all the factors (user id, item id, timestamp, last rating, etc.) into one vector; one latent representation per input (usua. transaction)

pros:

- work with data with high sparsity
- easy to optimize

cons:

- dense embedding can sometimes be bad
- shallow encoder

Machine learning

- (Facebook 2014) GBDT+LR [7]: uses gradient boosting decision trees to transform contextual and historical features and then inputs the transformed features into a linear classifier (logistic regression)
- (Google 2016) Wide & Deep [8]: encodes the input data by a feed-forward neural network (deep component, generalization) and cross-product transformation (wide component, memorization); jointly trains the model
- (Huawei 2017) DeepFM [9]: replace the wide component in Wide & Deep by FM
- (Alibaba 2018) DIN [10]: introduces attention mechanism to obtain ad-aware embeddings; develops two techniques (mini-batch aware regularization and data adaptive activation) that can be useful in practical industrial setting

Datasets

- Recommender Systems and Personalization Datasets - Julian McAuley, UCSD
- MovieLens Latest Datasets - last updated 9/2018
- TMDb 5000 Movie Dataset

References

[1] Awesome Deep Learning papers for industrial Search, Recommendation and Advertising

[2] Number of annual active consumers across Alibaba's online shopping properties from 2nd quarter 2016 to 2nd quarter 2021

[3] How Many Products Does Amazon Sell? – March 2021

[4] Linden, G., Smith, B., & York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Comput.*, *7*, 76-80.

[5] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, *42*(8), 30–37. doi:10.1109/MC.2009.263

[6] Rendle, S. (2010). Factorization Machines. *2010 IEEE International Conference on Data Mining*, 995–1000. doi:10.1109/ICDM.2010.127

[7] He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., ... Candela, J. Q. (2014). Practical Lessons from Predicting Clicks on Ads at Facebook. *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, 1–9. Presented at the New York, NY, USA. doi:10.1145/2648584.2648589

[8] Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Shah, H. (2016, September). Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems* (pp. 7-10).

[9] Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247*.

[10] Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., ... & Gai, K. (2018, July). Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1059-1068).