Vista: A Visually, Socially, and Temporally-aware Model for Artistic Recommendation

October 14, 2017

1 Problem being addressed

How do you mash up visual features, temporal dynamics and social aspects into recommendation of highly subjective content like artistic images? How do you deal with large scale sparse data? Also model both long term user preferences and short term, session level dynamics. In this case, the social aspect cannot be simply modelled as trust relations, rather we have users who are both creators and evaluators. The focus is an *ownership* signal.

2 Key ideas and contributions

- Introduces *Behance* dataset as a benchmark for modelling visual and artistic preferences. 2 datasets - clicks (implicit) and appreciate (explicit)
- Introduces a 1st order Markov Chain model where the transition from item S^u_{t-1} to item i is dependent on long term dynamics - owner appreciation OA, item appreciation IA, and short term dynamics - owner similarity OS and item similarity IS.
  $P(X^u_t = i|S^u_{t-1})$ = OA+IA + w_u. (OS + IS)
- 3 latent spaces - user-item affinity, user-user similarity, item-item similarity.

  Modelling of ownership data and user interactions helps address both cold user and item problems.
- Address higher order personalized MCs using personalized decaying scheme. Idea: "recent items should be more correlated with future actions for each user".
- Incorporate visual content features using embedding matrices that project item i into the latent item spaces. So the item latent vectors have a residue and the base part from the embedding.
- Sessions created by dividing the data with gaps larger than 1 hr.
- **Training data** - Behance
  Optimization - S-BPR, Asynchronous SGD, 5% Sampling for global parameters of the embedding matrix.
  Evaluation metric - AUC
  Baselines Most popular, BPR-MF, VBPR, MC, FPMC

3 Strong points

- Successfully demonstrate the power of visual content features and social dynamics in the case of sparse datasets in improving the recommendation quality
- Captures both long term and short term temporal dynamics.
- Creatively captures the social interaction between users.

4 Weak Points

- The authors do not describe the design decisions for the choice of decay scheme. Granted they do state that this should be a light-weight, expressive model, the choice of the inverse exponential function seems arbitrary. It would be interesting to explore the choice of decay schemes here. See *Dynamic Item-Based Recommendation Algorithm with Time Decay* (reference below).
- Not using session-level data and using a global binning mechanism may not be a reasonable assumption. On the other hand, how much more efficiency can be squeezed with session meta data?

5 Questions

1. Better way to incorporate sessions?
2. Can we identify sessions as a hidden variable? Possibly make it as a HMM?
3. Is the 3rd order MC test adequate? Within a session, a user could have traced multiple items. Could we do more?
4. The authors mention the metric spaces and probabilistic metric embeddings are a different line of work which have no known results that scale well. It would be interesting to repeat these experiments considering the same authors showed remarkable results on Google Local dataset in a 2017 paper (Transrec)
5. Design decision of 5% Sampling probability to update the two embedding matrices.

6 Misc

Easy, quick read. Creative. Gives a nice summary of existing work!
ACM ref:
(1). Ruining He, Chen Fang, Zhaowen Wang, and Julian McAuley. 2016. Vista: A Visually, Socially, and Temporally-aware Model for Artistic Recommendation. In Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16). ACM, New York, NY, USA, 309-316
(2). C. Xia, X. Jiang, Sen Liu, Zhaobo Luo and Zhang Yu, "Dynamic item-based recommendation algorithm with time decay," 2010 Sixth International Conference on Natural Computation, Yantai, 2010, pp. 242-247.