Modeling User Consumption Sequences

October 12, 2017

1 Problem being addressed

Sequential Recommendation also needs to consider repeat consumption. Recognize increasing user "boredom" with an item leading to eventual abandonment. Explore-exploit conflict. Model microscopic properties (inter-arrival time, novelty and repeat item choices) and macroscopic properties (re-consumption, item lifetime, eventual abandonment). Personalize this model. Incorporate temporal features.

2 Key ideas and contributions

- **Macroscopic observations**: power law in item lifetime (index, temporal); increasing time gaps conditioned on last gap (index, temporal); existing idea of repeat consumption = recency + quality
- **3 stage model**:
  (a) *Temporal* model to capture inter-arrival times: A semi-Markov model to capture "within" and "between" session inter-arrival times. (2-state session model with 3 distributions - power law session length, double Pareto intra-session gap, power law inter-session gap)

  (b) *Novelty* model to determine if item is novel or repeat. Uses logistic regression

  (c) *Choice* model to pick an item, based on (b). If it's novel, pick one ensuring the expected number of users is the same. (chosen from a multinomial distribution) If it's repeat, use *recency* weights (or distance factor in terms of number of intervening items for an item "copied" from past), item *quality* scores, and *time* factors.
- **Optimization**: negative log-likelihood. Bin time using a piece wise constant. eg: 30 1.1^k second bins i.e. exponentially spaced bins.
- Recency accounts for most of the likelihood. Likelihood is improved more with time factor than item quality
- Proven that lifetime of an item is finite *almost surely.*
- Proven that boredom is a consequence of conditioning on the fact that user will stop consuming an item. Implicitly captured boredom: Conditioned on the fact that item will not be consumed again, the last gap becomes larger.
- Personalize recency weights using a double Pareto distribution.
- **Training data** - {music, video, check-in} LastFM, YouTube, YouTube Music, MapClicks, WikiClicks, BrightKite, Google+

3 Strong points

- Automatically identifies/models the properties during sequence generation
- Study both short-term and long-term effects by using larger bins.
- Compares the relative important of time factor over recency/distance and quality.

4 Weak Points

- Personalization: Only recency weights are personalized using a double Pareto distribution. The time factors and the item quality are not. This may not lead to significant improvement as recency correlates to most of the likelihood. However, the results did not show significant improvements using personalization of just recency weights. Perhaps the other two factors' personalization could have helped.
- Personalization experiments limited to heavy consumption users. Personalizing heavy consumption users has the advantage of having a lot of data. But with other users, the problem is more interesting to study about how to model the factors accordingly.
- Can use the model for prediction and study effectiveness. Would be interesting to see baselines.
- Can separate out or distinguish types of repeat consumption. Periodic repeat consumptions are far more simpler to model with appropriate binning.
- Does not take user preferences into account.

5 Questions

1. Figure 2 - conditioned on last gap, it doesn't look like there is a significant increase in the gap times.
2. Simplified parameter estimation using break time B of at least 20 minutes. Validity of this for say, YouTube? Also figure 4.
3. complex issue of engagement in YouTube.

6 Misc

Heavy read; Interesting; Moderate technical depth; Involves some theoretical analysis.
ACM Ref:

Austin R. Benson, Ravi Kumar, and Andrew Tomkins. 2016. Modeling User Consumption Sequences. In Proceedings of the 25th International Conference on World Wide Web (WWW '16). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 519-529. DOI: https://doi.org/10.1145/2872427.28