Background
Our model
Results
Conclusion

Bounce Rate Optimization

with Markov traffic models

Steven Shao, James Xue

April 2018

Overview

1 Background

2 Our model

3 Results

4 Conclusion

Background

Problem (Optimal tagging)

In a network of items and tags, how to choose tags for a new item to maximize the probability of a user reaching that item?

Recommended paper "Optimal tagging with Markov chain optimization" [RG16]

- Model traffic using a Markov chain, and modify transitions in this chain to maximize traffic into a certain state of interest.
- Problem proven to be NP-hard, but with simple greedy approximation from [NWF78]

Their model

- Problem formulation:
  1. n states, plus a new σ state
  2. 2 constrained (details omitted) weight sets: q_{ij} and $\bar{q}_{ij}$ - connected and not connected to σ respectively
  3. S decides which weights to use: $\rho_{ij}(S) = \begin{cases} q_{ij} & i \in S \\ \bar{q}_{ij} & i \notin S \end{cases}$
  4. Objective: for a walk X, find $S \in [n], |S| \leq k$ maximizing

$$c(S) = \Pr_S[X_t = \sigma \text{ for some } t \geq 0]$$

- Note: only interesting if there are other 'absorbing' states
- Application: system (with ability to estimate network edge weights) recommending tags to user to maximize engagement with their new item

Theoretical results

- NP-hard - reduction to vertex cover, but...

Theorem (Monotonicity)

For all $S, S' \subset [n]$ *with* $S \subseteq S'$, $c(S) \leq c(S')$.

Theorem (Submodularity)

(Decreasing marginal value) For all $S \subset [n], z_1, z_2 \in [n] \setminus S$,

$$c(S \cup \{z_1\}) + c(S \cup \{z_2\}) \geq c(S \cup \{z_1, z_2\}) + c(S)$$

- Classic result [NWF78]: monotonic, submodular $\implies (1 - \frac{1}{e})$-approximation algorithm
- We devise simpler proofs of the two theorems above

A more realistic approach

- Our problem formulation:
  1. n states plus σ, S as before
  2. 2 *differently* constrained weight sets: q_{ij} and $\bar{q}_{ij}$
  3. Notably: user has ε_i or $\bar{\varepsilon}_i$ chance of leaving at each node i
  4. Objective: for a walk X, find S maximizing $\ell(S) = \mathbb{E}_S[|X|]$
- Application: given new item with tags, system decides whether or not to feature item in the page for each of the item's tags (decide: connect tag node to new item node?)
- More realistic: system incentivized to maximize *overall* engagement, not just with the new item

Our model

- More elusive theoretical guarantees
- In particular, under current conditions: not monotonic (what if new item is *really bad*?)
- We showed with a modified construction it's still NP-hard
- Promising empirical results

Experimental results

- Their model vs. our model: expected path length

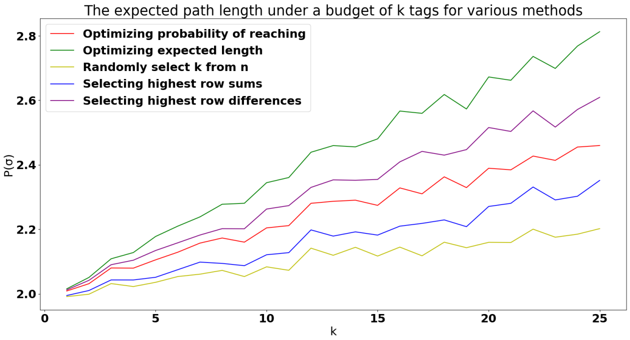

Figure: The expected path length under a budget of k tags for various methods

Experimental results

- Their model vs. our model: probability of reaching σ

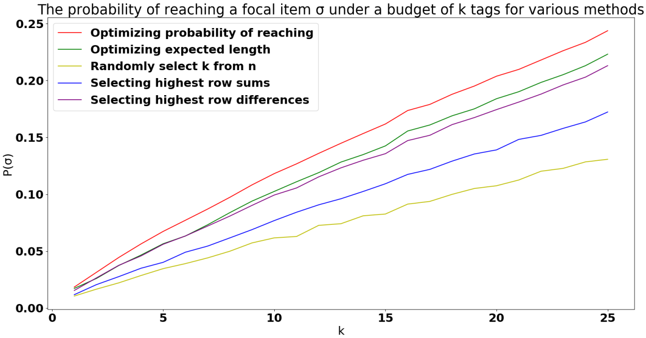

Figure: The probability of reaching a focal item σ under a budget of k tags for various methods

Conclusion

- Even without strong theoretical guarantees yet, *good* empirical evidence
- Model similarity suggests theoretical results around the corner, perhaps for restricted subset of instances
- Additional challenges for incremental computation on large, sparse datasets: LUP decomposition [Rei82]
- Far future: weight estimation techniques, live testing for causal results

References I

Ivan Cantador, Peter L Brusilovsky, and Tsvi Kuflik, *Second workshop on information heterogeneity and fusion in recommender systems (hetrec2011)*, ACM, 2011.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher, *An analysis of approximations for maximizing submodular set functions—i*, Mathematical Programming **14** (1978), no. 1, 265–294.

John Ker Reid, *A sparsity-exploiting variant of the bartels—golub decomposition for linear programming bases*, Mathematical Programming **24** (1982), no. 1, 55–69.

References II

Nir Rosenfeld and Amir Globerson, *Optimal tagging with markov chain optimization*, Advances in Neural Information Processing Systems, 2016, pp. 1307–1315.

Appendix: model details

Common setup:

- Markov chain over $n+1$ states, initial distribution π
- State $\sigma = n+1$ is eligible for k incoming transitions
- Decide $S \subseteq [n]$ of k states to "siphon" edge weights to σ
- Edge weights specified by $q_{ij}, \bar{q}_{ij}$ (connected vs. not connected):

$$\rho_{ij}(S) = \begin{cases} q_{ij} & i \in S \\ \bar{q}_{ij} & i \notin S \end{cases}$$

Their formulation:

- Constraint (leeching): $\forall j \neq \sigma, i, q_{ij} \leq \bar{q}_{ij} \qquad \forall i, \bar{q}_{i\sigma} = 0$
- Objective: for X_t the state of the chain at time t, find S maximizing

$$\Pr_S[X_t = \sigma \text{ for some } t \geq 0]$$

Appendix: model details

Common setup:

- Markov chain over $n+1$ states, initial distribution π
- State $\sigma = n+1$ is eligible for k incoming transitions
- Decide $S \subseteq [n]$ of k states to "siphon" edge weights to σ
- Edge weights specified by $q_{ij}, \bar{q}_{ij}$:

$$\rho_{ij}(S) = \begin{cases} q_{ij} & i \in S \\ \bar{q}_{ij} & i \notin S \end{cases}$$

Our formulation:

- Constraint (leeching): $\forall j \neq \sigma, i, q_{ij} \leq \bar{q}_{ij} \qquad \forall i, \bar{q}_{i\sigma} = 0$
- Constraint (local improvement): $\forall i, \sum_{j=1}^{n+1} q_{ij} \geq \sum_{j=1}^{n} \bar{q}_{ij}$
- Constraint (death rate): $\sum_{i}^{n} q_{ij} \leq 1$
- Objective: for X_t the state of the chain at time t, find S maximizing

$$\Pr_S[X_t = \sigma \text{ for some } t \geq 0]$$

Appendix: datasets

- Synthetic:
  - π selected from normalized $X \sim \mathcal{U}[0,1)^n$
  - For $i \in [n]$, each $\bar{q}_i$ also drawn from normalized $X \sim \mathcal{U}[0,1)^n$, then scaled according to $Y \sim \mathrm{B}(6,3)$
  - Each q_i is $\bar{q}_i$ scaled according to another $Y' \sim \mathrm{B}(6,3)$
  - $q_{i\sigma}$ drawn uniformly so that total of q_i is between total with $\bar{q}_i$ and 1.
- Future work: generate data using Last.fm, Delicious, Movielens datasets from HotRec 2011 workshop [CBK11]