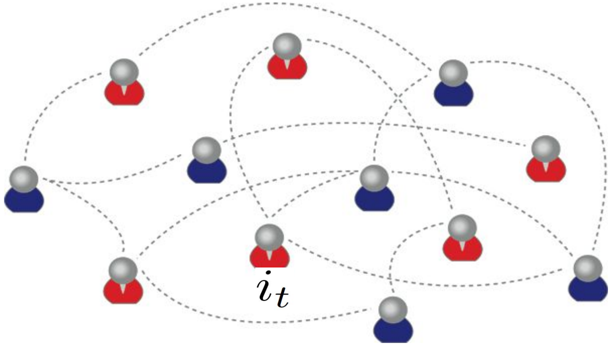
Horde of Bandits using GMRFs

Sharan Vaswani, Mark Schmidt, Laks Lakshmanan
University of British Columbia

Introduction

Input:



i_t

$\mathcal{C}_t$

Protocol:

For t = 1 to T

- Recommend an item j to target user
- Observe rating $r_{i,j}$
- Obtain new estimate for preferences of user i

Introduction

- New marketer without any user meta-data or available rating information.
- Each item j can be described by its content $\mathbf{x}_j$
- The generative model for ratings is linear, i.e.

$$\underbrace{r_{i,j}}_{\text{Rating by user i on item j}} = (\underbrace{\mathbf{w}_i^*}_{\text{"True" user preference}})^T \mathbf{x}_j + \underbrace{\eta_{i,j,t}}_{\text{Zero mean sub-gaussian noise}}$$

Framework: Contextual Bandits[1] - Sequential framework to trade-off exploration (learning about user preferences) and exploitation (making good recommendations).

Li, Lihong, et al. "A contextual-bandit approach to personalized news article recommendation", 2010.

Introduction

Aim: Minimize regret across a time horizon T.

$$R(T) = \sum_{t=1}^{T} \left[\max_{j \in \mathcal{C}_t} (\mathbf{w}_{i_t}^{*T} \mathbf{x}_j) - \mathbf{w}_{i_t}^{*T} \mathbf{x}_{j_t} \right]$$

Regret

Set of context vectors

“True” preference for the target user

Context vector for item j

Item chosen by the algorithm

Gang of Bandits (GOB)[1]

Motivation: Users interact with each other. Especially true for RS associated with social networks. Eg:

Basic Idea:

- Extend the contextual bandits to use the social network to make better recommendations by sharing feedback between users.
- Assume homophily[2]: Users connected in the network have similar preferences

1.Cesa-Bianchi et al. "A gang of bandits", 2013..

2. McPherson, et al. "Birds of a feather: Homophily in social networks", 2001

Estimating user preferences in GOB

$$\mathbf{w}_t = \underset{\mathbf{w}}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \sum_{k \in \mathcal{M}_{i,t}} (\mathbf{w}_i^T \mathbf{x}_k - r_{i,k})^2 \; + \lambda \mathbf{w}^T (L \otimes I_d) \mathbf{w} \right]$$

dn-dimensional mean estimate of concatenated preferences

Set of items rated by user i until round t

Observed rating

Laplacian of the social network

Limitations of previous work:

- Algorithm in [1] has $O(n^2 d^2)$ space and time complexity. Not scalable
- Clustering based approaches[3] lose personalization.

1.Cesa-Bianchi et al. "A gang of bandits", 2013..

2. Gentile et al. "Online clustering of bandits", 2014

Contributions

- Propose a scalable solution for estimating the mean by making a connection to Gaussian Markov Random Fields.
- Analyze the computational complexity and prove regret bounds for Epoch-Greedy and Thompson sampling.
- Propose a heuristic to learn the graph on the fly.

Scaling up GOB

Basic Idea: Mean estimation in GOB is equivalent to MAP estimation in a GMRF

Likelihood: $r_{i,j} \sim \mathcal{N}(\mathbf{w}_i^T \mathbf{x}_j, \sigma^2)$ **Prior:** $\mathbf{w} \sim \mathcal{N}(0, (\lambda L \otimes I_d)^{-1})$

Posterior: $\mathcal{N}(\hat{\mathbf{w}}_t, \Sigma_t^{-1})$ $\hat{\mathbf{w}}_t = \frac{1}{\sigma^2} \Sigma_t^{-1} \mathbf{b}_t$

$$\Sigma_t = \frac{1}{\sigma^2} X_t^T X_t + \lambda (L \otimes I_d) \quad \mathbf{b}_t = X_t^T \mathbf{r}_t$$

Covariance matrix

dn x dn dimensional block diagonal matrix

Number of CG iterations

- Solve by conjugate gradient in time $O(\kappa(nd^2 + d \cdot \text{nnz}(L)))$
- Space: $O(nd^2 + \text{nnz}(L))$

Contributions

- Make a connection to Gaussian Markov Random Fields and propose a scalable solution for estimating the mean
- Analyze the computational complexity and prove regret bounds for Epoch-Greedy and Thompson sampling.
- Propose a heuristic to learn the graph on the fly.

Algorithms - Upper Confidence Bound

UCB rule:

$$j_t = \mathrm{argmax}_{j \in \mathcal{C}_t} \left(\underbrace{\langle \mathbf{w}_t, \mathbf{x}_{i_t,j} \rangle}_{\text{Requires O(d) time}} + \alpha_t \underbrace{\sqrt{\mathbf{x}_{i_t,j}^T \Sigma_t^{-1} \mathbf{x}_{i_t,j}}}_{\text{Requires } O(\kappa(nd^2 + d \cdot \mathrm{nnz}(L))) \text{ time}} \right)$$

- Not scalable if the number of context vectors is large

Algorithms - Epoch-Greedy

- **Algorithm:** Divide T into Q epochs. In each epoch,
  - Do 1 round of random exploration (Pick an available item at random)
  - Do "some" steps of exploitation i.e. $j_t = \mathrm{argmax}_{j \in \mathcal{C}_t} \langle \mathbf{w}_t, \mathbf{x}_{i_t, j} \rangle$

- **Regret Bound:**

$$R(T) = \tilde{O}\left(n^{1/3} \left(\frac{\mathrm{Tr}(L^{-1})}{\lambda n} \right)^{\frac{1}{3}} T^{\frac{2}{3}} \right)$$

Connectivity of the graph

Sub-optimal dependence on T

Algorithms - Thompson Sampling

- **Algorithm:**
  - Obtain a sample from the posterior, i.e. $\tilde{\mathbf{w}}_t \sim \mathcal{N}(\bar{\mathbf{w}}_t, \Sigma_t^{-1})$
  - Pick greedily using this sample i.e. $j_t = \mathrm{argmax}_{j \in \mathcal{C}_t} \langle \tilde{\mathbf{w}}_t, \mathbf{x}_{i_t,j} \rangle$
- **Naive Sampling**:
  - Compute sparse Cholesky factor from prior covariance = $L \otimes I_d$
  - Make rank-1 updates to it to obtain Cholesky factor for covariance at round t

Problem: Cholesky factor gets dense, leading to $O(d^2n^2)$ cost for sampling

Algorithms - Thompson Sampling

- **Proposed Sampling[1]:**

$$\Sigma_t \tilde{\mathbf{w}}_t = (L \otimes I_d)\tilde{\mathbf{w}}_0 + X_t^T \tilde{\mathbf{r}}_t$$

MAP estimation !

Sample from prior covariance

Perturbed (with standard normal noise) ratings

- **Regret Bound:**

$$R(T) = \tilde{O}\left(\frac{dn\sqrt{T}}{\sqrt{\lambda}}\sqrt{\log\left(\frac{3\,\mathrm{Tr}(L^{-1})}{n} + \frac{\mathrm{Tr}(L^{-1})T}{\lambda d n^2 \sigma^2}\right)}\right)$$

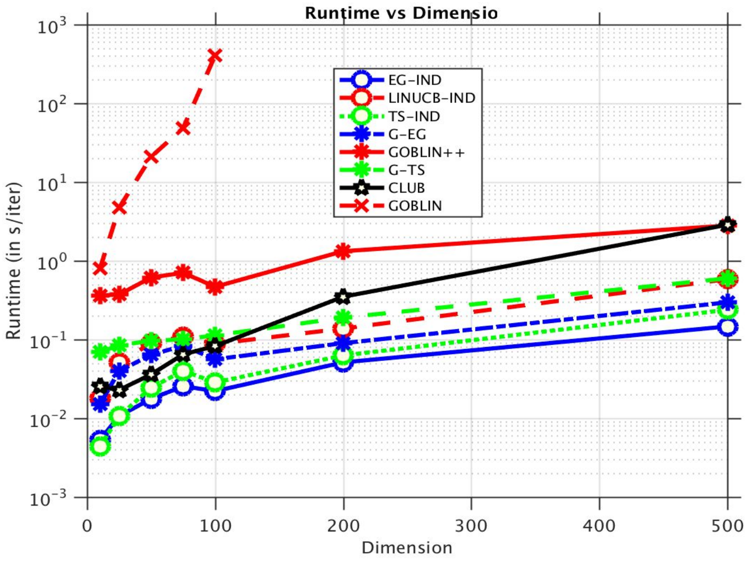
Near-optimal dependence on T

1. Papandreou et al. "Gaussian sampling by local perturbations

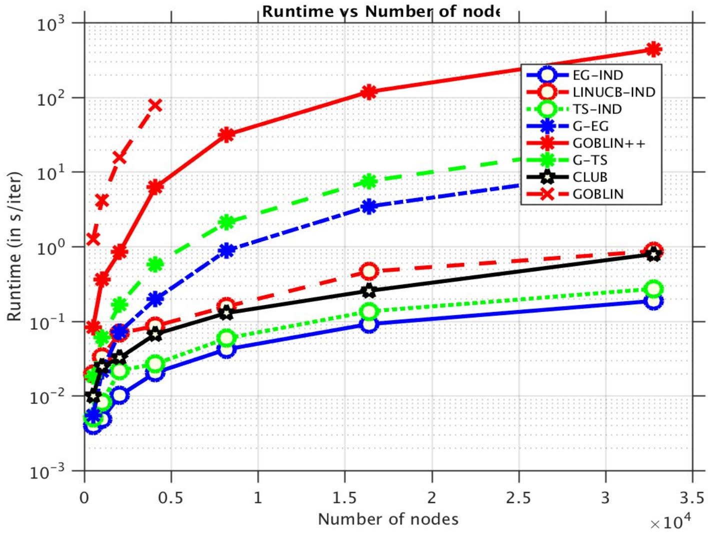
Experiments - Scalability

Graph Based: G-EG, GOBLIN (Cesa-Bianchi'13), GOBLIN++ (scalable GOBLIN), G-TS

Baselines: No sharing: EG-IND, LINUCB-IND, TS-IND; **Clustering**: CLUB

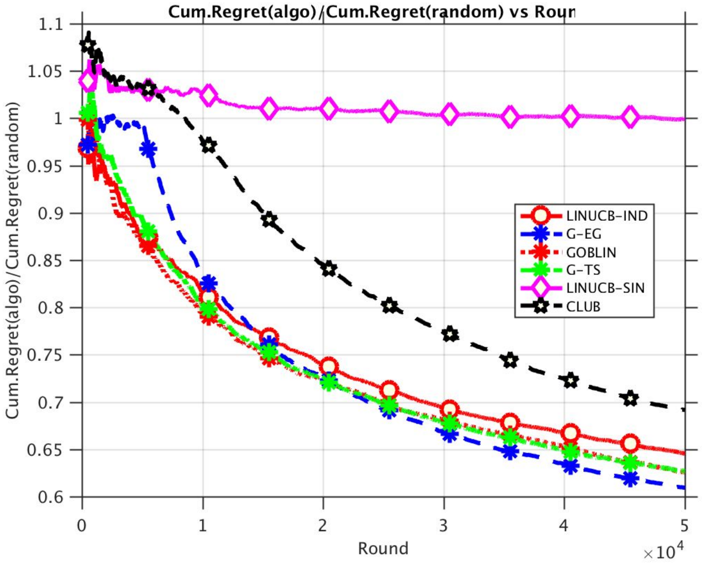


Scaling with Dimension

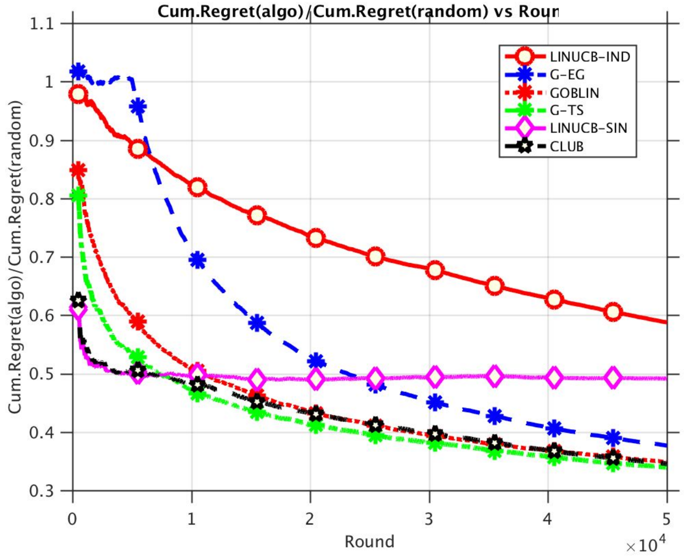


Scaling with Network Size

Experiments - Regret



Delicious



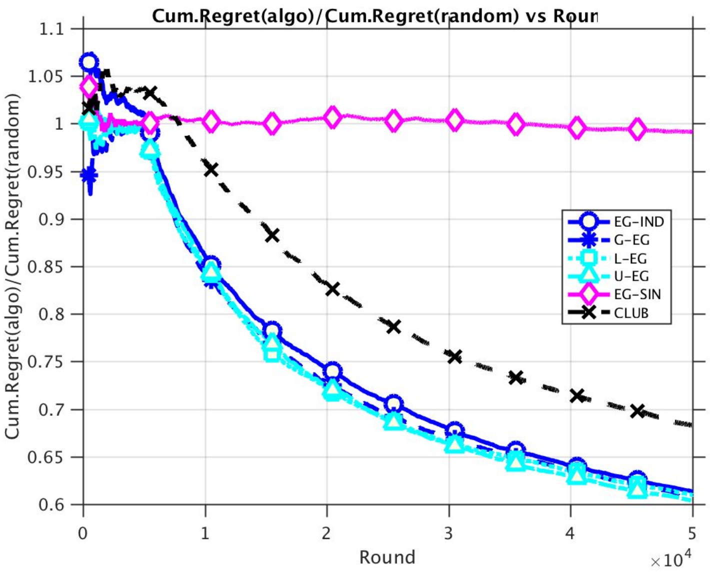
Last FM

Contributions

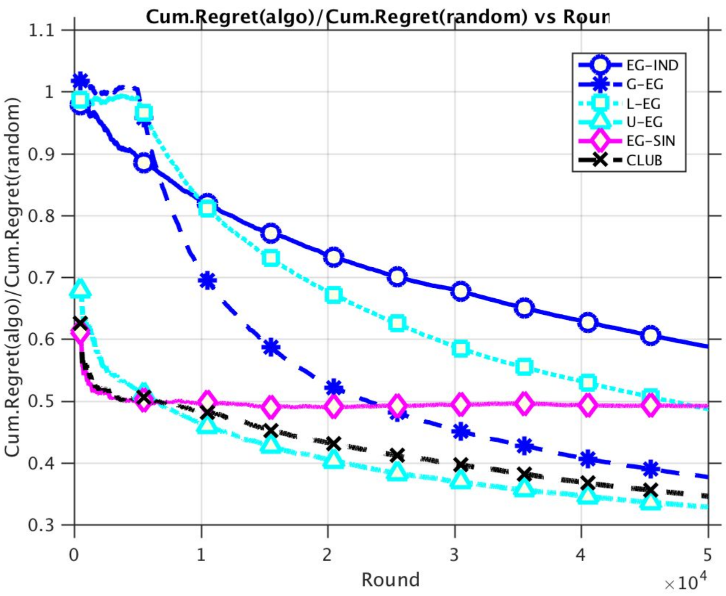
- Make a connection to Gaussian Markov Random Fields and propose a scalable solution for estimating the mean
- Analyze the computational complexity and prove regret bounds for Epoch-Greedy and Thompson sampling.
- Propose a heuristic to learn the graph on the fly.

Learning the Graph

L-EG: Learning starting from empty graph; **U-EG:** Updating starting from given graph



Delicious



Last FM

Future Work

- Tighten the regret bound for Thompson Sampling
- Prove regret bounds for “Learning the graph” variant

Contributions

- Make a connection to Gaussian Markov Random Fields and propose a scalable solution for estimating the mean
- Analyze the computational complexity and prove regret bounds for Epoch-Greedy and Thompson sampling.
- Propose a heuristic to learn the graph on the fly.

Backup slides

Learning the Graph

$$[\mathbf{w}_t, \underset{\text{Estimate of the inverse covariance matrix}}{V_t}] = \underset{\mathbf{w},V}{\operatorname{argmin}} \underbrace{||\mathbf{r}_t - \Phi_t \mathbf{w}||_2^2}_{\text{Data fitting term}} + \underbrace{\mathrm{Tr}\,(V(\lambda W^T W}_{\text{Smoothness of preferences wrt to learnt graph}} + \underbrace{V_{t-1}^{-1}}_{\text{Smooth change in graph structure}})) + \underbrace{\lambda_2||V||_1}_{\text{Encourage a sparse graph}} - (dn+1)\underbrace{\ln|V|}_{\text{Penalize complexity of graph}}$$

- Solve by alternating minimization:
  - w-subproblem: Same as MAP estimation
  - V-subproblem: Same as Graphical Lasso

$$V_t = \underset{V}{\operatorname{argmin}}\, \mathrm{Tr}\,((V\underbrace{[\lambda \overline{W}_t^T \overline{W}_t + V_{t-1}^{-1}}_{\text{Empirical Covariance}})) + \lambda_2||V||_1 - (dn+1)\ln|V|$$