Graph-based Recommendation System

Kaige Yang and Laura Toni
{kaige.yang.11, l.toni}@ucl.ac.uk
Department of Electrical and Electronic Engineering
University College London, London, UK

Abstract

We study recommendation systems in large-scale environments, modelled as contextual multi-armed bandit (MAB) problems. We propose a graph-based recommendation system that exploits the geometry of user space to efficiently learn optimal recommendations. We propose a graph-clustering to reduce the dimensionality of the recommendation problem, while preserving accuracy of MAB. We then study the effect of graph sparsity and clusters size on the MAB performance and provide exhaustive simulation results both in synthetic and in real-case datasets.

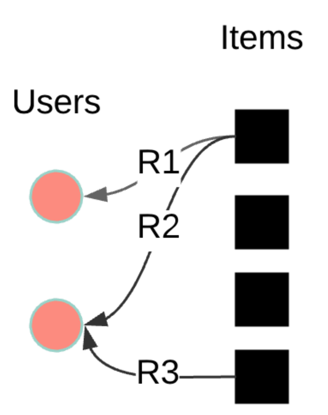
Background

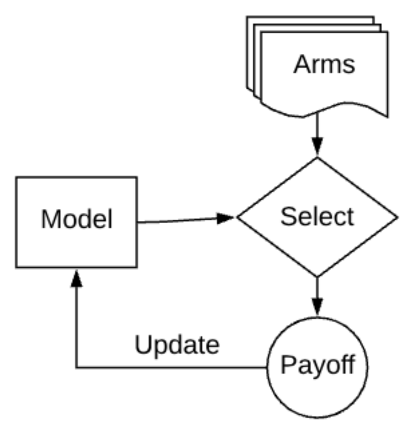
Content-based recommendation systems

- Three entities in the system: ***users, items and ratings.***
  - An item is recommended to a user. The user rates the item.
  - Ratings are a function of both item and user features.
- Item features and ratings are available. While user profiles are unknown.
- **Goal**: Given ratings of user-item pairs, learn user profiles and recommend items that bear high rating.

Multi-armed bandit framework (MAB)

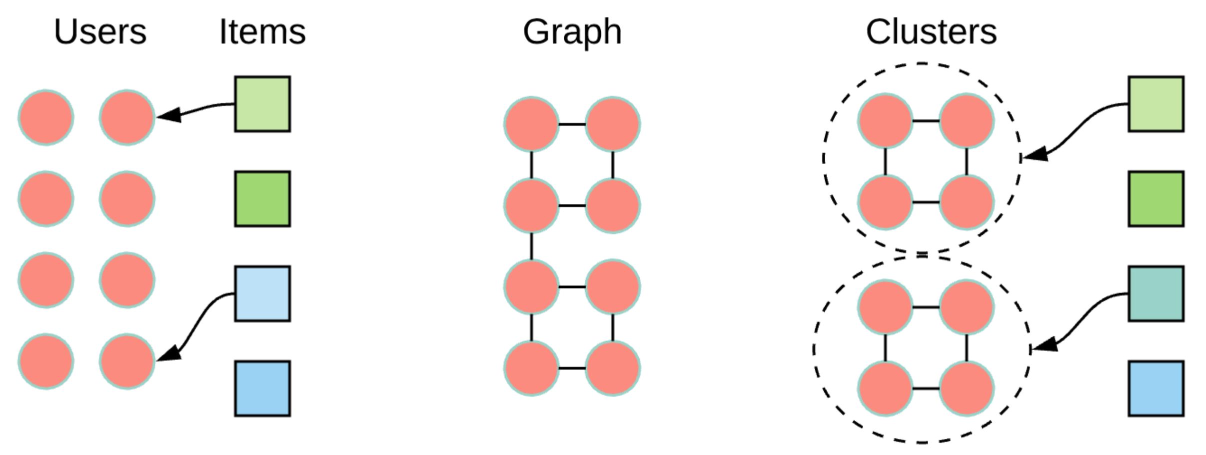
- This learning process can be formalised by multi-armed bandit framework.
- **Challenge**: The problem becomes intractable in scenarios with infinitely large strategy sets (e.g., large number of users)
- Solution: Clustering technique has been proposed to quantise the content space (i.e., user space).
- Limitation: In current algorithm, the number of cluster constantly increases over time, leading to meaningless clusters.





Graph-based recommendation system

- Motivation: Find clusters by graph-based clustering techniques to reduce the dimensionality and still preserve accuracy.



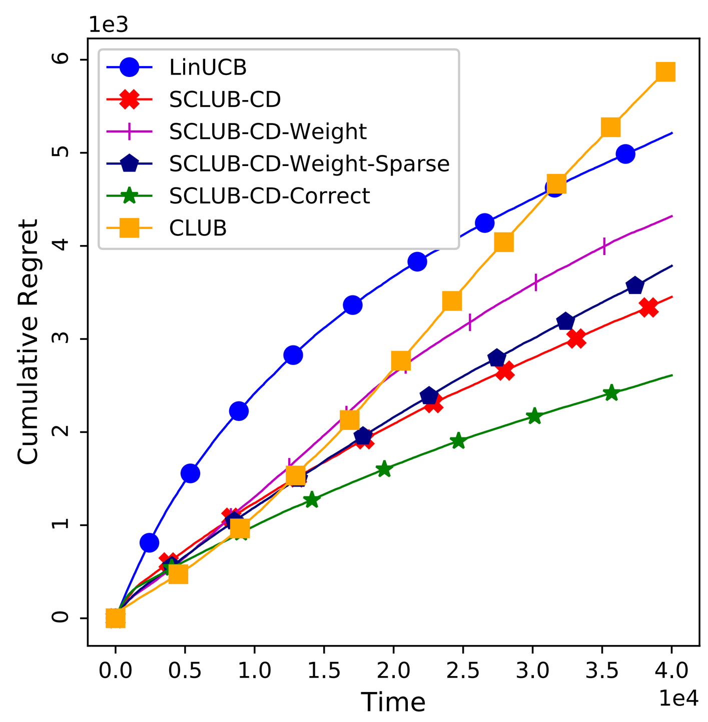
SCLUB-CD algorithm

- User profiles are estimated by Ordinary Least Square (OLS).
- At each step, the algorithm constructs an user graph via Gaussian *RBF-distance* between user profiles.
- To introduce sparsity and robustness to noise, select the strongest edges and delete the others.
- User clusters are derived via community detection applying the Louvain Method [2].
- Estimate of profile per cluster are average of user profiles within the cluster.
- Items are recommended to users based on cluster profile.

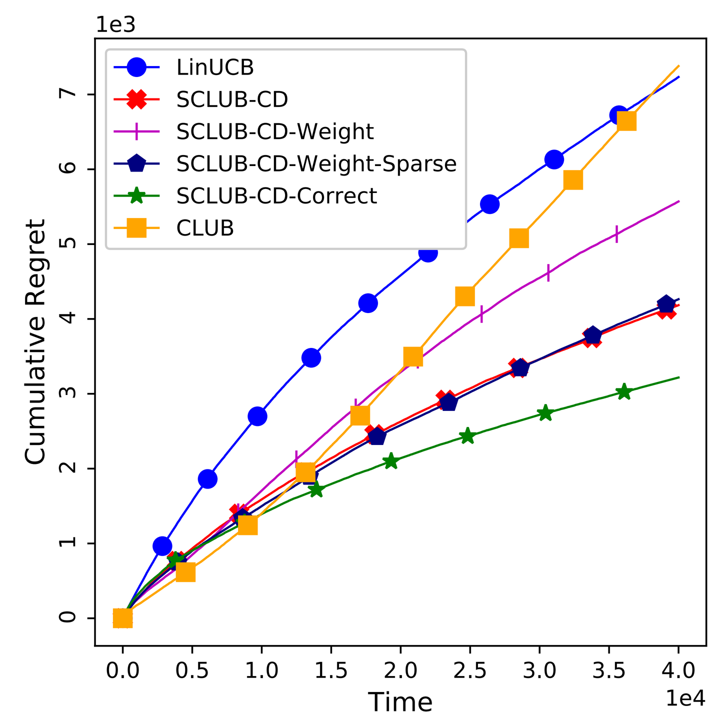
$$k_t = \arg\min_{k\in C_t} (\mathbf{u}^c_{j(i_t)}\mathbf{x}_k) + CB_{j(i_t)}\mathbf{x}_k$$

Results

- Synthetic data sets: σ_e *noise level,* σ_c *Intra – cluster noise level*

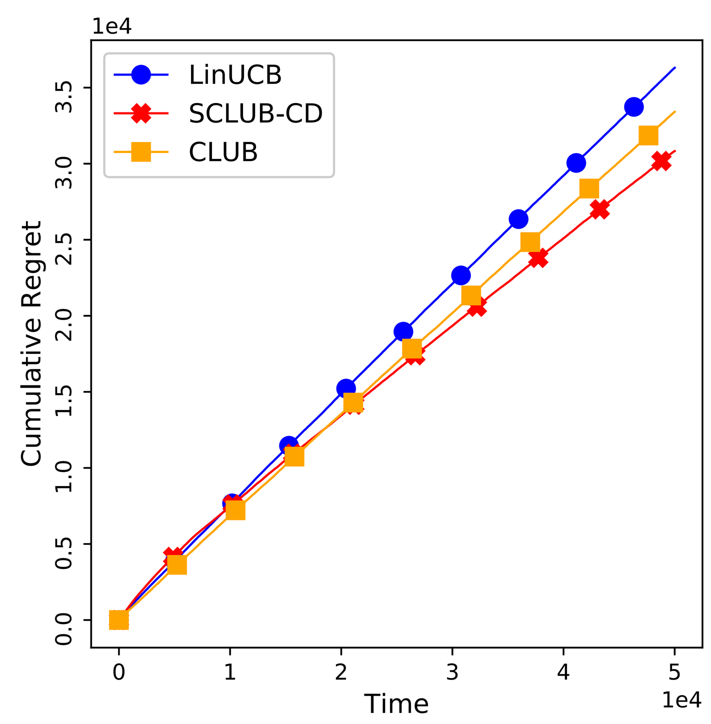


$\sigma_c = 0.25,\ \sigma_e = 0.25$

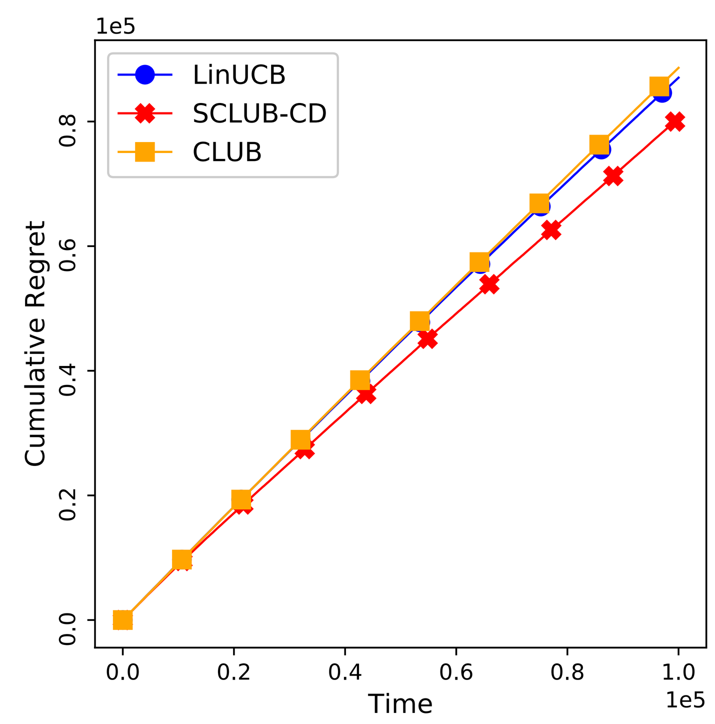


$\sigma_c = 0.5,\ \sigma_e = 0.25$

- Real world data sets: LastFM, Delicious



LastFM



Delicious

- SCLUB-CD outperforms its competitors consistently over all scenarios.
- Clustering the user space with proposed approach leads to a system improvement.
- The proposed approach is able to find the right tradeoff between dimensionality reduction and approximation in clustering users.
- The unweighted graph (SCLUB-CD) makes the system more robust to weights estimation errors.
- No so good in realistic datasets.

Conclusion

The graph-based bandit algorithm encodes users' similarity in profile by an undirected unweighted graph and group users into clusters.

- It adopts graph-based clustering to extract meaningful clusters.
- The unweighted graph (SCLUB-CD) makes the system more robust to weights estimation errors.
- All these components lead to an overall gain in terms of cumulative regret with respect to state-of-the-art algorithms.
- Future works to compensate for poor performance in datasets.

References

[1]Robbins, H. (1985). Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers* (pp. 169-177). Springer, New York, NY.
[2]Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.