Spectral clustering and Markov tree models for collaborative filtering

Dan-Cristian Tomozei
Thomson Paris Research Lab
dct@melix.net

Laurent Massoulié
Thomson Paris Research Lab
laurent.massoulie@thomson.net

Peter Marbach
University of Toronto
Dept. of Computer Science
marbach@cs.toronto.edu

***Abstract*—In this work we consider the collaborative filtering problem of predicting valuation of items by users, on the basis of observed valuations. While this is a generic problem with many applications, we are motivated by the prediction of user tastes for multimedia content. We take a model-based approach. Specifically, we assume that the valuations of each user are distributed according to a Markov random field with tree-based topology spanning the items. We propose a Maximum Likelihood approach for identifying the model and performing predictions. We validate our predictions on the Netflix database. We also present a way of deploying this technique in a distributed setting. We then propose enhancements of this method, replacing the unique Markov tree distribution by a mixture of such distributions. To identify mixture components, we develop a spectral clustering algorithm, which allows to partition users into similar profile groups. This algorithm is proven to identify hidden clusters under mild assumptions. We finally combine such clustering with the training of a specific Markov tree model per cluster. We find that the proposed approach has a reasonable computational cost, and yields performance comparable to that of the Netflix Cinematch recommender system.**

Keywords: Collaborative Filtering, Markov Random Fields, Spectral Clustering, Mixture Models

I. Introduction

We consider the problem of predicting valuation of items by users, on the basis of observed rankings. More precisely, we consider the situation where a population of users ranks items in a given set. Each user has a valuation for each item. However users reveal their valuation only for a small subset of items, namely the set items that the users have already ranked. The goal is then to design an algorithm that uses the known user valuations to accurately predict the user's valuations of all items, i.e. to answer the question "*What is the user's valuation of an unranked item given that we know the user's current rankings?*" Such a system (algorithm) is also referred to as a recommender system. This problem has many important applications, particular in the context of e-commerce sites that offer a large number of products. For example, Amazon.com uses a recommender system to suggest to customers books based on the book rankings that a customer has provided. Similarly, the online movie rental company Netflix uses a recommender system to suggest movies to customers. Recommender systems can potentially also be used in the context social networking applications that have become increasingly popular.

In this work, we assume that the valuations of each user are distributed according to a Markov random field with tree-based topology spanning the items. In section II we adapt a Maximum Likelihood approach for identifying the parameters of this model and performing value predictions. In section III, we propose enhancements of aforementioned method, replacing the unique Markov tree distribution by a mixture of such distributions. To identify mixture components, we develop a spectral clustering algorithm, which allows to partition users into similar profile groups. This algorithm is proven to identify hidden clusters under assumptions less restrictive than in prior work [3], [4]. We combine such clustering with the training of a specific Markov tree model per cluster. We find that the proposed approach has a reasonable computational cost.

II. Chow Liu Trees

Let $\mathcal{T} = (V, E)$ be a tree. Assume that in each vertex $v \in V$ of the graph we have a random variable X_v over a discrete space $\mathcal{S}_v$. Also assume each of the X_v have a probability mass function (PMF) T_v which is strictly positive in every point x_v of the corresponding space $\mathcal{S}_v$. We study, like in [1], distributions of the following form:

$$T : \prod_{v \in V} \mathcal{S}_v \to [0,1] \qquad T(\mathbf{x}) = \frac{\prod_{(u,v)\in E} T_{uv}(x_u, x_v)}{\prod_{v\in V} T_v(x_v)^{\deg(v)-1}} \quad (1)$$

The marginal distributions $T_u(x_u) = \mathbb{P}(X_u = x_u)$ and $T_{uv}(x_u, x_v) = \mathbb{P}(X_u = x_u, X_v = x_v)$ should satisfy

$$T_u(x_u) = \int T_{uv}(x_u, x_v)\, dx_v,\ \forall v \neq u \quad (2)$$

We say that the marginal distributions are *consistent*. This type of distribution is a particular case of a Markov Random Field (MRF). More general MRFs span over any kind of undirected graphs. We chose to study tree-based distributions, since trees have attractive properties which lead to low computational complexity.

We find a tree-based MRF of the form (1) which best models an n-th order multivariate distribution, based on a finite set of samples from it. Trying to find a model for the true distribution in the same form of an n-th order multivariate distribution is not feasible (its storage only would require an exponential amount of memory, namely $\prod_{v\in V} |S_v|$). We then employ this estimator for the Collaborative Filtering problem.

Algorithm 1 is proposed in [1]. It constructs the maximum weight spanning tree over the complete empirical mutual information graph. The resulting tree-estimator for the true law $\mathcal{P}(\cdot)$ is then shown to be the best one.

Algorithm 1 Chow-Liu algorithm for maximum likelihood tree structure estimation

CHOW-LIU($\{\mathbf{x_1}, \ldots, \mathbf{x_s}\}$)

1: Compute the sample marginal distributions P_v, P_{uv} for $u, v \in V$
2: Compute the sample mutual information $I_{uv} = \sum_{x_u, x_v} P_{uv}(x_u, x_v) \log \frac{P_{uv}(x_u,x_v)}{P_u(x_u)P_v(x_v)}, \forall u, v \in V$
3: Compute the maximum weight spanning tree E_T = MWST($\{I_{uv}\}$)
4: Set $T_{uv} := P_{uv}$, for $(u, v) \in E_T$

III. Spectral Clustering

We now describe our technique for separating users into disjoint clusters. Several methods relying on spectral properties of suitable matrices have been developed in the past, the closest to ours being those of Boppana [5], McSherry [4] and Ng et al. [3]. However, their approaches and results differ in several ways.

We assume that we are given a *conflict* matrix A. A is assumed to be symmetric, each entry A_{ij} is assumed to be non-negative, and interpreted as a measure of the disagreement over content ratings between two users i and j in our collaborative filtering context. Diagonal entries A_{ii} all equal 0. For instance, A_{ij} could be set to 1 if the two users issue distinct ratings on at least T different movies, for some threshold T. Alternatively, A_{ij} could count the fraction of distinct ratings made by i and j over the collection of movies they both rated, in which case A_{ij} could take any value between 0 and 1.

Our aim is to partition the index set into clusters, so that most conflicts are between indices from distinct clusters.

We show that the proposed algorithm successfully recovers some hidden cluster structure under specific statistical assumptions. Namely, we consider the following "planted partition" model, previously considered by McSherry in [4], a generalization of the model considered in [5] and [2].

The extended planted partition model is as follows. Indices are partitioned into K distinct clusters C_1, …, C_K. The conflict values A_{ij} are assumed to be random, with values in $[0, 1]$, and independent across all index pairs (i, j), $i < j$. Moreover, for all $i < j$, these variables are assumed to verify the following:

$$\mathbb{E}(A_{ij}) = \begin{cases} p_{kk} & \text{if both } i, j \in C_k \text{ for some cluster } C_k, \\ p_{k\ell} & \text{if } i \text{ and } j \text{ belong to distinct clusters } C_k \text{ and } C_\ell. \end{cases}$$

In this context, we prove the following

Theorem 1: Consider a conflict matrix A distributed according to the above planted partition model. Assume that the number of clusters K is fixed, the initial number of indices N is large, and the size of cluster C_k verifies $|C_k| \sim \alpha_k N$ for some fixed positive parameters α_k such that $\min_k \alpha_k > 0$.

Algorithm 2 Spectral clustering algorithm

SPECTRAL-CLUSTERING($\mathbf{z}$)

1: Associate to each index $n = 1, \ldots, N$ the corresponding $(K-1)$-dimensional vector $z_n := (z_1^{(n)}, \ldots, z_{K-1}^{(n)})$, consisting of the corresponding coordinates of the eigenvectors $z(1), \ldots, z(K-1)$ of A.
2: Pick M indices $n(1), \ldots, n(M)$ uniformly at random from $\{1, \ldots, N\}$, for some suitable M.
3: **repeat**
4: Identify the two indices $n(i)$, $n(j)$ that achieve the smallest Euclidean distance $||z_{n(i)} - z_{n(j)}||$ among all M indices.
5: Remove $n(j)$ and set M to $M - 1$.
6: **until** $M = K$
7: The remaining K indices, say $n(1), \ldots, n(K)$ now serve as cluster representatives. Assign any other index n to that representative that achieves the minimum in $\{||z_{n(i)} - z_n||, i = 1 \ldots, K\}$.

Assume that the conflict matrix P is such that $p_{k\ell} = b_{k\ell}\frac{\omega_N}{N}$, where $\lim_{N\to\infty} \omega_N = +\infty$ for some fixed $b_{k\ell}$ and

$$\forall k \neq k', \ \sup_{\ell} |b_{k\ell} - b_{k'\ell}| > 0. \qquad (3)$$

Let the initial number of candidate cluster representatives M be fixed. Then with high probability, the above algorithm partitions indices into the original clusters, except for at most $o(N)$ mis-classified indices.

References

[1] C.K. Chow, C.N. Liu, *Approximating Discrete Probability Distributions with Dependence Trees*, IEEE Trans. on Information Theory, Volume 14, Issue 3, pp. 462 – 467, May 1968.
[2] A. Condon, R.M. Karp, *Algorithms for Graph Partitioning on the Planted Partition Model*, Proc. 3rd Int. Workshop on Approx. Algorithms for Comb. Opt. Prob.: RANDOM-APPROX '99.
[3] A.Y. Ng, M.I. Jordan, Y. Weiss, *On Spectral Clustering: Analysis and an algorithm*, In Advances in Neural Information Processing Systems 14: Proc. 2001.
[4] F. McSherry. Spectral partitioning of random graphs. In FOCS, pp. 529–537, 2001.
[5] R.B. Boppana, *Eigenvalues and Graph Bisection: An Average-case Analysis*, Proc. FOCS 1987, pp. 280 – 285
[6] Z. Fűredi, J. Komlós, *The eigenvalues of random symmetric matrices* (1981), Combinatorica 1:3, pp. 233 – 241