Cross-Domain Recommendation via Clustering on Multi-Layer Graphs

Aleksandr Farseev, Ivan Samborskii, Andrey Filchenkov, Tat-Seng Chua

By Aleksandr Farseev
http://farseev.com

Aug 8th, 2017
Collaborative Venue Category Recommendation – recommendation of venue categories (i.e. restaurant, cinema) to user using information about his/her profile (i.e. past visits) and/or information about users from the same domain.

Venue categories:
- Clothing Store
- Hotel
- Ice Cream Shop

Total 764 different categories
Idea 1: Utilization of Individual And Group Knowledge for Better Recommendation
We perform venue category recommendation based on both individual and group knowledge => naturally models the impact of society on an individual's behavior during the selection of a new place to go:

$$
rec(u) = \text{sort} \left( \gamma \cdot vec_u + \theta \frac{\sum_{v \in C_u} vec_v}{|C_u|} \right)
$$
What do we need user communities for?

+ Users from the same community (extracted from multi-source data) may have similar location preferences

+ Search within user community significantly reduces search space during the recommendation process
Example of User Communities (1)

Community 1: Gingers

Community K: Darker Hair
One way to find user communities is to model users' relationships in the form of a graph so that dense subgraphs are considered to be user communities.
One of the commonly formulations is MinCut problem.

For a given number $k$ of subsets, the MinCut involves choosing a partition $C_1, ..., C_k$ such that it minimizes the expression:

$$
cut(C_1, ..., C_k) = \sum_{i=1}^{k} W(C_i, \bar{C}_i)
$$

*W is the sum of weights of edges attached to vertices in $C_i$*
How to solve MinCut problem?

Approximation of MinCut as standard trace minimization problem:

$$\min_{U \in \mathbb{R}^{n \times k}} \text{tr}(U^T L U), \text{s.t. } U^T U = I$$

which can be solved by Spectral Clustering:

1. Calculates Laplacian matrix $L \in \mathbb{R}^{n \times n}$
2. Builds matrix of the first $k$ eigenvectors $U \in \mathbb{R}^{n \times k}$ correspond to the smallest eigenvalues of $L$
3. Clusters data in a new space $U$ using i.e. $k$-means algorithm
Idea 2: Utilization of Multi-Source Data
Most of user actively use $\approx 3$ social networks

Multi-source data describe user from multiple views
Cross-Domain
Venue Category Recommendation

Cross Domain - Venue category recommendation – recommendation of venue categories (i.e. restaurant, cinema) using information about his/her profile (i.e. past visits) and/or information about users from other sources (i.e. images, texts, location types).

Multi-Source Data:
- Clothing Store
- Hotel
- Ice Cream Shop
Community Detection must performed in a Cross-Source Manner...

Problems:

• Data source integration
• Community detection
How to represent multi-source data?

Multi-layer graph – graph $G$, where $G = \{G_i\}$, $G_i = (V, E_i)$
Extending definition of spectral clustering

\[
\min_{U \in \mathbb{R}^{n \times k}} \sum_{i=1}^{M} \text{tr}(U^T L_i U), \text{ s.t. } U^T U = I
\]

\[
\min_{U \in \mathbb{R}^{n \times k}} \text{tr}(U^T L_{\text{sum}} U), \text{ where } L_{\text{sum}} = \sum_{i=1}^{M} L_i
\]

Such approximation could suffer from poor generalization ability.
Regularized Clustering on Multi-layer Graph -1

Use Grassman Manifolds to keep final latent representation “close” to all layers of multi-layer graph*. Where projected distance between two spaces $Y_1$ and $Y_2$:

$$d_{proj}^2(Y_1, Y_2) = \frac{1}{2} \|Y_1 Y_1^T - Y_2 Y_2^T\|_F^2,$$

where $\|A\|_F$ is the Frobenius norm

$$d_{proj}^2(S, \{S_i\}_{i=1}^M) = kM - \sum_{i=1}^M \tr(SS^T - S_i S_i^T)$$

Regularized Clustering on Multi-layer Graph -2

Extends the objective function to introduce the subspace analysis regularization

\[
\min_{U \in \mathbb{R}^{n \times k}} \sum_{i=1}^{M} \text{tr} \left( U^T L_i U \right) + \alpha \left( kM - \sum_{i=1}^{M} \text{tr} \left( UU^T U_i U_i^T \right) \right), \text{s.t. } U^T U = I
\]

\[
\min_{U \in \mathbb{R}^{n \times k}} \text{tr} \left( U^T L_{\text{mod}} U \right)
\]

\[L_{\text{mod}} = \sum_{i=1}^{M} \left( L_i - \alpha U_i U_i^T \right)\]
Idea 4: Making use of Inter-Layer (Inter-Source) Relations
Incorporating inter-layer relationship (1)

By using distance on Grassman Manifolds, we present the new objective function for the \(i\)th layer:

\[
\begin{aligned}
\min_{\mathcal{O}_i \in \mathbb{R}^{n \times k}} & \quad \text{tr}(\mathcal{O}_i^T L_i \mathcal{O}_i) + \beta_i \left( kM - \sum_{j=1, j \neq i}^{M} w_{i,j} \text{tr}(\mathcal{O}_i \mathcal{O}_i^T U_j U_j^T) \right) \\
\end{aligned}
\]

\[
\begin{aligned}
\min_{\mathcal{O}_i \in \mathbb{R}^{n \times k}} & \quad \text{tr}(\mathcal{O}_i^T \hat{L}_i \mathcal{O}_i) \\
\end{aligned}
\]

\[
\hat{L}_i = L_i - \beta_i \sum_{j=1, j \neq i}^{M} w_{i,j} \text{tr}(U_j U_j^T)
\]
But how can we determine $w_{i,j}$ when computing $i$-th layer?

$$\min_{\mathcal{U}_i \in \mathbb{R}^{n \times k}} \text{tr}(\mathcal{U}_i^T \mathcal{L}_i \mathcal{U}_i)$$

$$\hat{\mathcal{L}}_i = \mathcal{L}_i - \beta_i \sum_{j=1, j\neq i}^M w_{i,j} \text{tr}(\mathcal{U}_j \mathcal{U}_j^T)$$

Inter-layer relationship graph $R(V, E)$ – weighted graph which represents the similarity between layers.

$$\forall (i, j) \in E, w_{i,j} = \frac{\sum_{k=2}^K \left(1 - \frac{||M_{i,k} - M_{j,k}||}{\sqrt{N(N-1)}}\right)}{K-1}$$

where $M_{i,k}$ is clustering co-occurrence matrix of layer $i$, $m_{a,b} = 1$, if users $a$ and $b$ assigned to the same cluster, and 0 otherwise.
Final objective function

Let’s combine equations from previous slides to define the final objective function:

\[
\begin{align*}
\min_{U \in \mathbb{R}^{n \times k}} \sum_{i=1}^{M} \text{tr} \left( U^T \hat{L}_i U \right) + \alpha \left( kM - \sum_{i=1}^{M} \text{tr} \left( UU^T \hat{O}_i \hat{O}_i^T \right) \right) = \\
= \min_{U \in \mathbb{R}^{n \times k}} \text{tr} \left( U^T \sum_{i=1}^{M} (\hat{L}_i - \alpha \hat{O}_i \hat{O}_i^T) U \right)
\end{align*}
\]
Problems

- Community detection
- Data source integration
Recall: Community-Based Cross-Domain Recommendation

We perform venue category recommendation based on both individual and group knowledge, where group knowledge is obtained from multiple sources:

\[
rec(u) = \text{sort} \left( \gamma \cdot vec_u + \theta \frac{\sum_{v \in C_u} vec_v}{|C_u|} \right)
\]
NUS-MSS Dataset

Dataset* is presented as a set of features, extracted from user-generated data in three social networks:
- text based from Twitter (LDA, LIWC, text features)
- image based from Instagram (concepts)
- location based from Foursquare (LDA, categories, Mobility Features)

Foursquare categories is split into two parts: 3 months data (train) and 2 months (test).

Data Sources

Text Features: LIWC; Latent Topics; Writing behavior

Location Features: Venue Category Distribution; Areas of Interest (AOI)

Image Features: Image Concept Distribution (Image Net)

Data Sources:
- LIWC
- LDA
- Mobility
- Location Type Preferences
- Google Net
- Image Concepts
## Evaluation Baselines

### Recommender Systems

**Popular (POP)** — recommendation based on user’s past experience

**Popular All (POP All)** — recommendation based on experience of all users

**Multi-Source Re-Ranking (MSRR)** — linearly combines recommendation results from all data modalities

**Nearest Neighbor Collaborative Filtering (CF)** — recommendation based on top k most similar Foursquare users

**Early Fusion (EF)** — fuses multi-source data into a single feature vector

**SVD++** — makes use of the “implicit feedback” information

**FM** — brings together the advantages of different factorization-based models via regularization.

### Community Detection Approaches

**$C^3R - \hat{L}_i$** — $C^3R$ recommendation without inter-layer regularization

**$C^3R - \hat{L}_i - \hat{L}_{\text{Mod}}$** — $C^3R$ recommendation without inter-layer regularization and sub-space regularization

**$C^3R - \text{Comm}$** — $C^3R$ recommendation without user community extraction

**$C^3R (\text{DBScan})$** — $C^3R$ recommendation, where user communities are detected by Density-Based clustering (DBScan)

**$C^3R (\text{x-means})$** — $C^3R$ recommendation, where user communities are detected by x-means clustering

**$C^3R (\text{Hierarchical})$** — $C^3R$ recommendation, where user communities are detected by Hierarchical Clustering

**$C^3R$** — Our Approach
Evaluation against other recommender systems
Evaluation against other community detection approaches

+ Incorporation of group knowledge is important
+ Multi-modal clustering performs better than single-source clustering
+ Incorporation of Inter-Source relationship is crucial.
Evaluation against source combinations

+ In different geo regions, different data sources are of different importance
  + Location data is more powerful than other data modalities

\[
W_R = \begin{pmatrix}
    tw & 4sq & inst & tmp & mob \\
    tw & 1 & 0.632 & 0.621 & 0.643 & 0.561 \\
    4sq & 0.632 & 1 & 0.614 & 0.631 & 0.570 \\
    inst & 0.621 & 0.614 & 1 & 0.621 & 0.551 \\
    tmp & 0.643 & 0.631 & 0.621 & 1 & 0.560 \\
    mob & 0.561 & 0.570 & 0.551 & 0.560 & 1
\end{pmatrix}
\]
Examples of detected user communities

<table>
<thead>
<tr>
<th>Name</th>
<th>Bag of Words for different modalities</th>
</tr>
</thead>
</table>
| Gadgets 832 users | **Text**: device, launcher, android  
                           **Visual**: mouse, digital clock, hard disc  
                           **Location**: electronics store, tech startup, technology building |
| Arts 538 users   | **Text**: painting, landscape, reflection  
                           **Visual**: obelisk, paintbrush, pencil box  
                           **Location**: arts & crafts store, arts & entertainment, museum |
| Food 446 users   | **Text**: dining, coffee, cooking  
                           **Visual**: pineapple, microwave, frying pan  
                           **Location**: italian restaurant, pizzeria, macanese restaurant |
Future Work

Community Detection is more useful when it is Source-Dependent

=> Introduce Supervision Into Clustering

How?

• Graph Construction Level – reweight edges according to prior knowledge about existing user communities

• Model Level – introduce community-related constraints into clustering
Summary

+ Multi-View Data is crucial for User Community Detection

+ For the task of venue category recommendation, both Group And Individual Knowledge are Important

+ Venue Category Recommendation is not a conventional recommendation task: users visit many venue types from the past. (items from the train set often occur in test set)
Our released large multi-source multi-modal datasets


The Released Datasets

NUS-MSS  NUS-SENSE

Our Tutorial on Multi-View Learning @ WST WSSS’17

http://tutorial.farseev.com
Thank You

Questions?

By Aleksandr Farseev
http://farseev.com
Normalized Discounted Cumulative Gain (NDCG) measure, which is defined as:

\[
NDCG@p = \frac{DCG@p}{IDCG@p}, \quad DCG@p = \sum_{i=1}^{p} \frac{2^{rel_i}}{\log_2(i + 1)}, \quad rel_i = \frac{Cat_i}{N_{Cat}},
\]

Average Precision (AP), which is defined as:

\[
AP@p = \frac{1}{\sum_{i=1}^{p} r_i} \left\{ \frac{\sum_{j=1}^{i} r_j}{i} \right\}, \quad r_i = \begin{cases} 1, & \text{i is in top p visited cat.} \\ 0, & \text{otherwise.} \end{cases}
\]