LCARS A LOCATION-CONTENT-AWARE RECOMMENDER SYSTEM

Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, Ling Chen

Content

1. Introduction
2. Preliminary
3. Proposed LCARS method
4. Experiments
5. Conclusion

Introduction

- Event-based social network services (EBSNs):
  - Provide online platforms for users to establish social events which will be held in physical places.
  - Meetup (www.meetup.com) and DoubanEvent (www.douban.com/events/)
- Location-based social networking services (LBSNs):
  - Allow users to "check-in" at spatial venues and rate their visit via mobile devices
  - Foursquare (foursquare.com) and Gowalla (gowalla.com).
- The target of this paper is to mine more knowledge from the user activity history data in LBSNs and EBSNs:
  - If we want to visit venues in a city such as Beijing, where should we go?
  - If we want to attend local events such as dramas and exhibitions in a city, which events should we attend?

INTRODUCTION

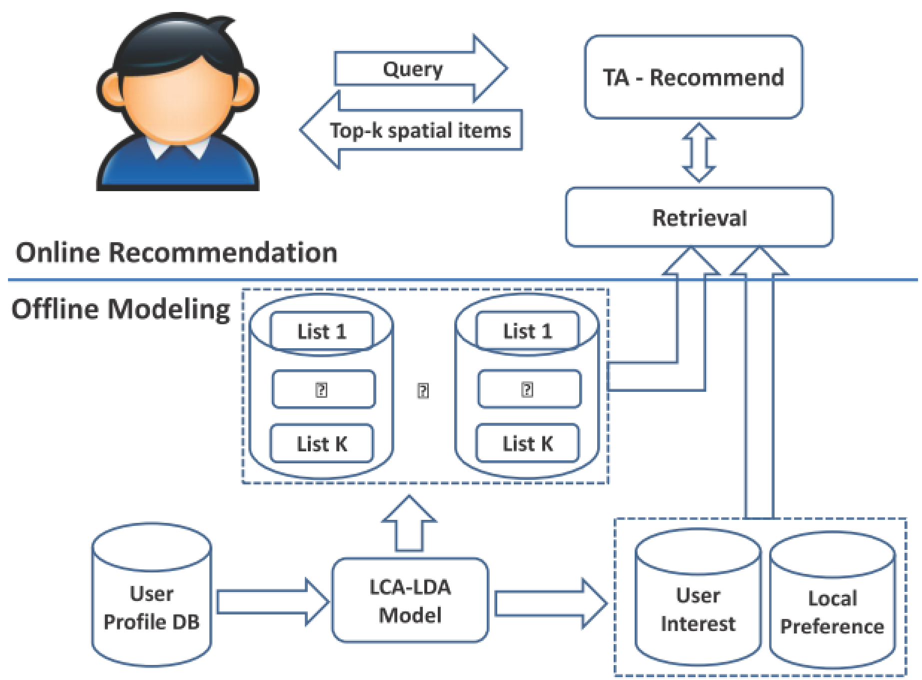
- Problem definition:
  - Given a querying user u with a querying city l_u, find k interesting spatial items (venues, events) within l_u that match the preference of u.
- Difficulties:
  - A user can only visit a limited number of physical venues and a limited number of social events.

    => Sparsity problem.
  - A user travels to a new place where he/she does not have any activity history.

    => Solely using a CF-based method is not feasible.
- A location-content-aware recommender system (LCARS) is proposed:
  - Both the location and content information of spatial items is exploited to alleviate the data sparsity problem.

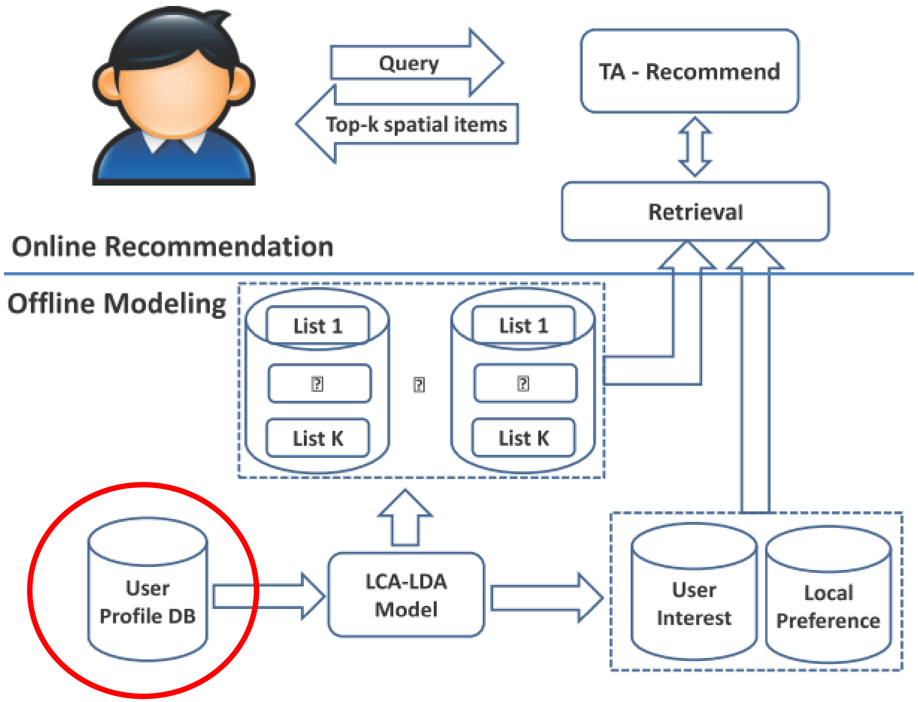
INTRODUCTION

- The Architecture Framework of LCARS:
  - *Offline model*: model user preferences to spatial items by simultaneously considering the *User Interest* and *Local Preference*.
  - *Online Recommendation*: computes a ranking score for each spatial item v within l_u by automatically combining user interest and local preference.



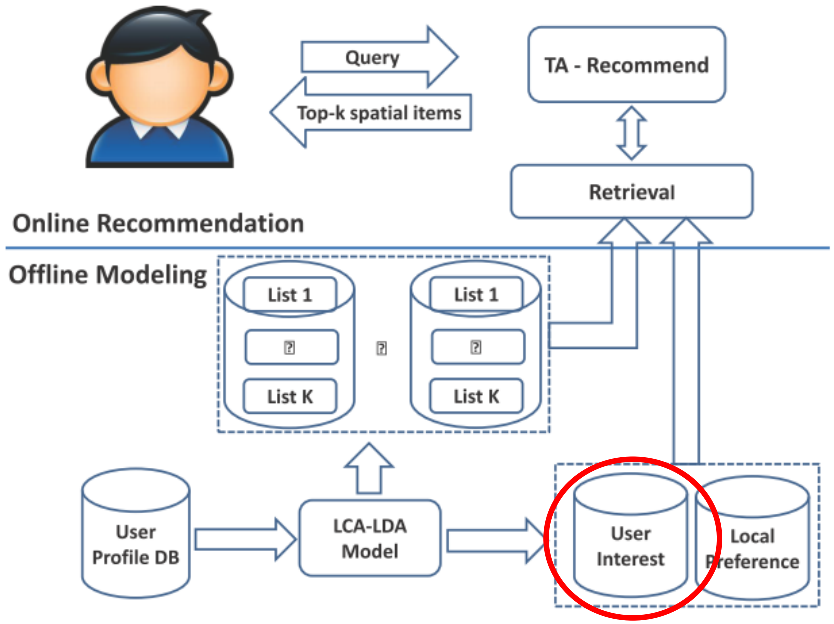
PRELIMINARY

- Spatial Item ($\boldsymbol{v}$): Either an event or venue.
- User Profile: A set of four-tuples (u, v, l_v, c_v) .
  - User u selects a spatial item v with the content word set c_v in location l_v.



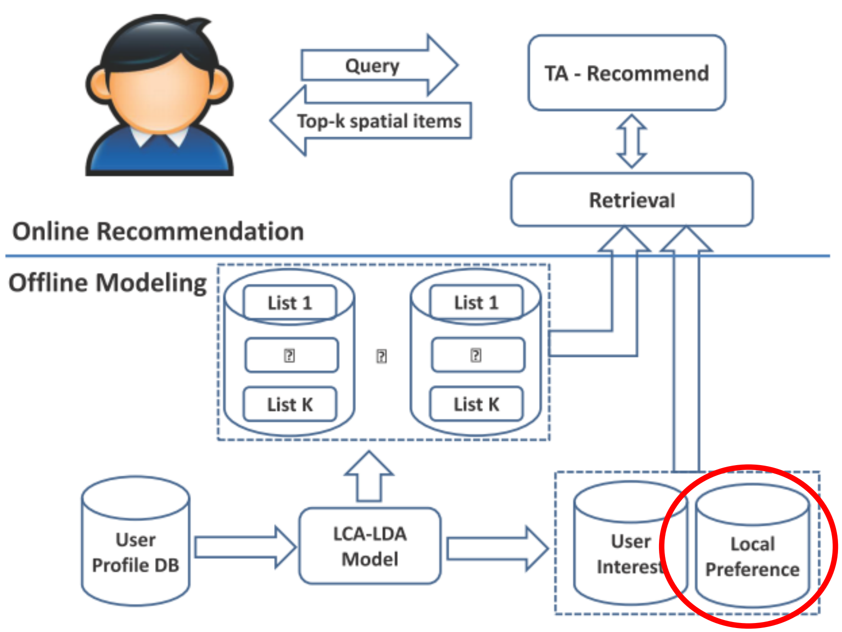
PRELIMINARY

- Topic (z) : represented by a topic model ϕ_z
  - A probability distribution over spatial item $\{ P(v \mid \phi_z) : v \in V \}$ or $\{ \phi_{zv} : v \in V \}$
  - A probability distribution over content words $\{ P(c \mid \phi_z) : v \in C \}$ or $\{ \phi'_{zc} : c \in C \}$
- User Interest (θ_u) : a probability distribution over topics (high probability value means high interest of user u in the topic).



Preliminary

- Topic (z) : represented by a topic model ϕ_z
  - A probability distribution over spatial item $\{ P(v \mid \phi_z) : v \in V \}$ or $\{ \phi_{zv} : v \in V \}$
  - A probability distribution over content words $\{ P(c \mid \phi_z) : v \in C \}$ or $\{ \phi'_{zc} : c \in C \}$
- Local Preference (θ'_l) : a probability distribution over topics (high probability value means that the topic more likely appears in city l).



Proposed LCARS method

- Offline modeling:
  - Given a querying user u with a querying city l_u , the likelihood that user u will prefer item v when traveling to city l_u :

$$P(v \mid \theta_u, \theta'_{lu}, \phi, \phi') = \sum_{c \in C_v} P(v, c \mid \theta_u, \theta'_{lu}, \phi, \phi')$$
$$= \lambda_u \sum_{c \in C_v} P(v, c \mid \theta_u, \phi, \phi') + (1 - \lambda_u) \sum_{c \in C_v} P(v, c \mid \theta'_{lu}, \phi, \phi')$$

  where λ_u is the mixing weight which controls the effect of personal interest and local preference.

  - Two terms of the probability function can be computed as:

$$P(v, c|\theta_u, \phi, \phi') = \sum_z P(v, c|z, \phi_z, \phi'_z) P(z|\theta_u)$$
$$= \sum_z P(v|z, \phi_z) P(c|z, \phi'_z) P(z|\theta_u)$$

$$P(v, c|\theta'_{l_u}, \phi, \phi') = \sum_z P(v, c|z, \phi_z, \phi'_z) P(z|\theta'_{l_u})$$
$$= \sum_z P(v|z, \phi_z) P(c|z, \phi'_z) P(z|\theta'_{l_u})$$

PROPOSED LCARS METHOD

- Offline modeling:
  - For computation:

$$P(v, c|\theta_u, \phi, \phi') = \sum_z P(v, c|z, \phi_z, \phi'_z) P(z|\theta_u)$$
$$= \sum_z P(v|z, \phi_z) P(c|z, \phi'_z) P(z|\theta_u)$$

$$P(v, c|\theta'_{l_u}, \phi, \phi') = \sum_z P(v, c|z, \phi_z, \phi'_z) P(z|\theta'_{l_u})$$
$$= \sum_z P(v|z, \phi_z) P(c|z, \phi'_z) P(z|\theta'_{l_u})$$

$$\hat{\theta}_{uz} = \frac{n_{uz} + \alpha_z}{\sum_{z'} (n_{uz'} + \alpha_{z'})}$$

$$\hat{\theta}'_{lz} = \frac{n_{lz} + \alpha'_z}{\sum_{z'} (n_{lz'} + \alpha'_{z'})}$$

$$\hat{\phi}_{zv} = \frac{n_{zv} + \beta_v}{\sum_{v'} (n_{zv'} + \beta_{v'})}$$

$$\hat{\phi}'_{zc} = \frac{n_{zc} + \beta'_c}{\sum_{c'} (n_{zc'} + \beta'_{c'})}$$

where:

- n_{uz} is the number of times that topic z has been chosen by user u
- n_{lz} is the number of times that topic z has been chosen in location l
- n_{zv} is number of times that spatial item v has been generated by topic z.
- n_{zc} is number of times that content word c has been generated by topic z.
- α, α', β, β', γ, and γ' are constant values.

Proposed LCARS method

- Online recommendation:
  - Ranking score *S:*

$$P(v|\theta_u,\theta'_{lu},\phi,\phi')=\sum_{c\in C_v}P(v,c|\theta_u,\theta'_{lu},\phi,\phi')$$

$$=\lambda_u\sum_{c\in C_v}P(v,c|\theta_u,\phi,\phi')+(1-\lambda_u)\sum_{c\in C_v}P(v,c|\theta'_{lu},\phi,\phi')$$

$$S(u,l_u,v)=\sum_z F(l_u,v,z)W(u,l_u,z)$$

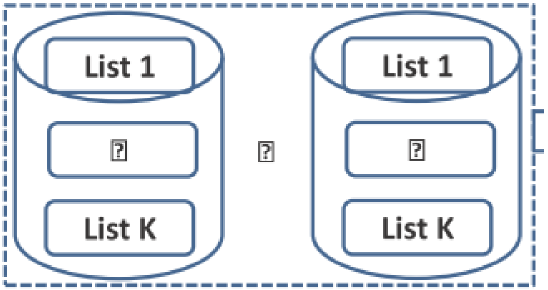
where

$$W(u,l_u,z)=\hat{\lambda}_u\hat{\theta}_{uz}+(1-\hat{\lambda}_u)\hat{\theta}'_{l_u z}$$ → Online scores

$$F(l_u,v,z)=\begin{cases}\hat{\phi}_{zv}\sum_{c_v\in\mathcal{C}_v}\hat{\phi}'_{zc_v} & v\in\mathcal{V}_{l_u}\\ 0 & v\notin\mathcal{V}_{l_u}\end{cases}$$ → Offline scores

PROPOSED LCARS METHOD

- Online recommendation:
  - Based on the ranking value, top-k spatial items for each query are selected.
  - Large number of spatial items becomes larger (e.g., millions) require millions of vector operations.
  - A scalable query processing technique for top-k recommendations is proposed:
    - Partition all spatial items into locations at a given level such as cities.
    - For each location, store K lists of spatial items.
    - Spatial in each list is sorted by the items' offline score.
    - At query time, retrieve k spatial items that have highest ranking from K lists.



Experimental Results

- Dataset:
  - DoubanEvent: 100,000 users, 300,000 events and 3,500,000 positive definite RSVPs.
  - Foursquare: 11326 users and 1385223 check-ins, no item content information.
- Comparative approaches:
  - User interest, social and geographical influences (USG).
  - Category-based k-Nearest Neighbors Algorithm (CKNN).
  - Item-based k-Nearest Neighbors Algorithm (IKNN).
  - LDA
  - Location-Aware LDA.
  - Content-Aware LDA (CA-LDA)

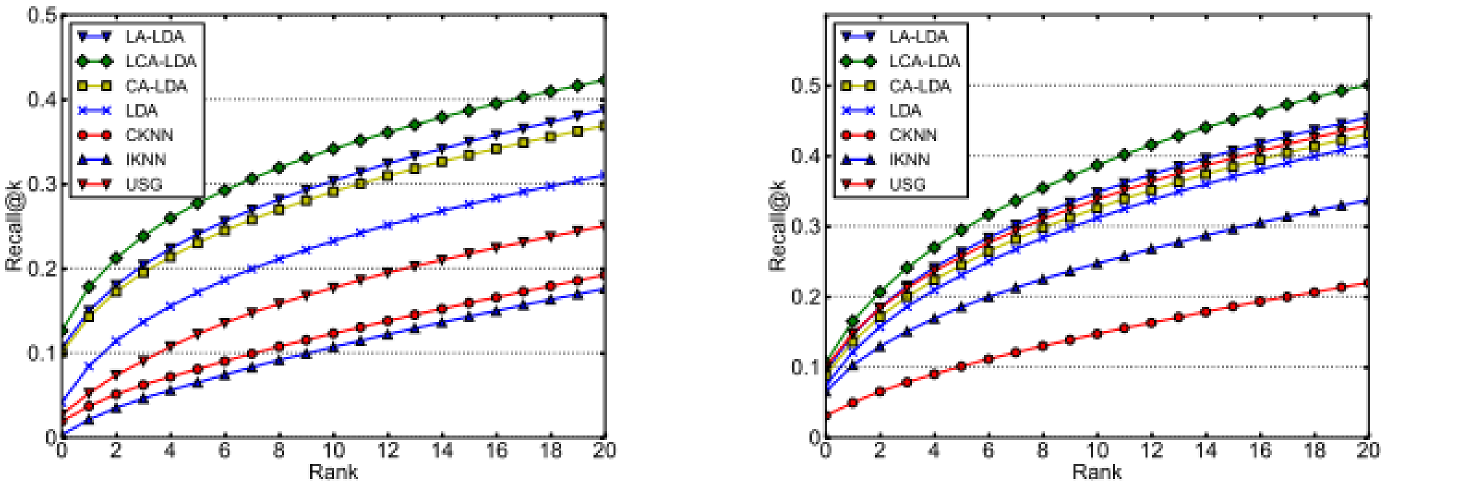
Experimental Results

- Evaluation method:
  - For each test case $(u,\ v,\ l_v)$, randomly select 1000 additional items located at l_v.
  - Compute the ranking score for the list of 1001 spatial items.
  - Let p the rank of test item v in the list.
  - if $p < k$, we have a hit.
  - The proposed measurement:

$$Recall@k = \frac{\#hit@k}{|S_{test}|}$$

Experimental Results

- Results:

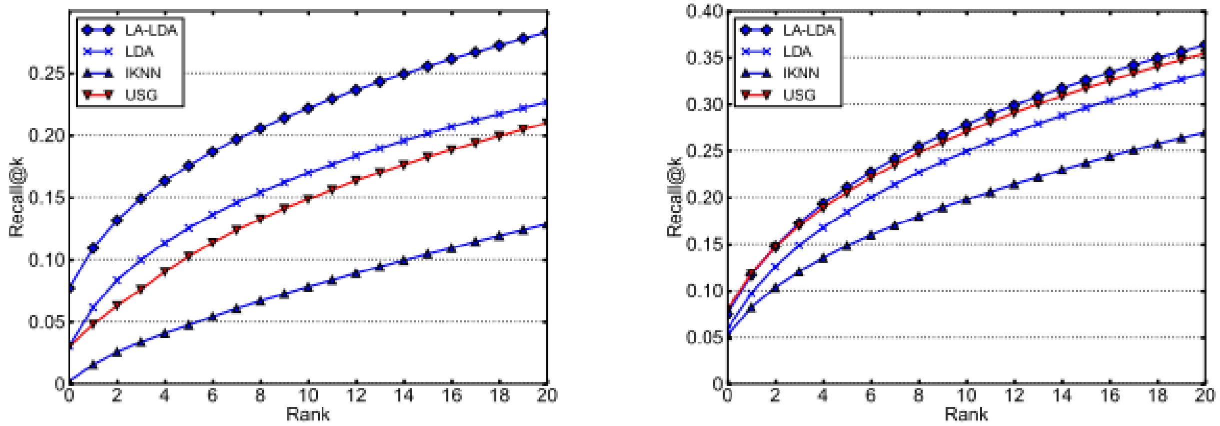


(a) Users Traveling in New Cities (b) Users Traveling in Home Cities

Figure 3: Top-k Performance on DoubanEvent

Experimental Results

- Results:

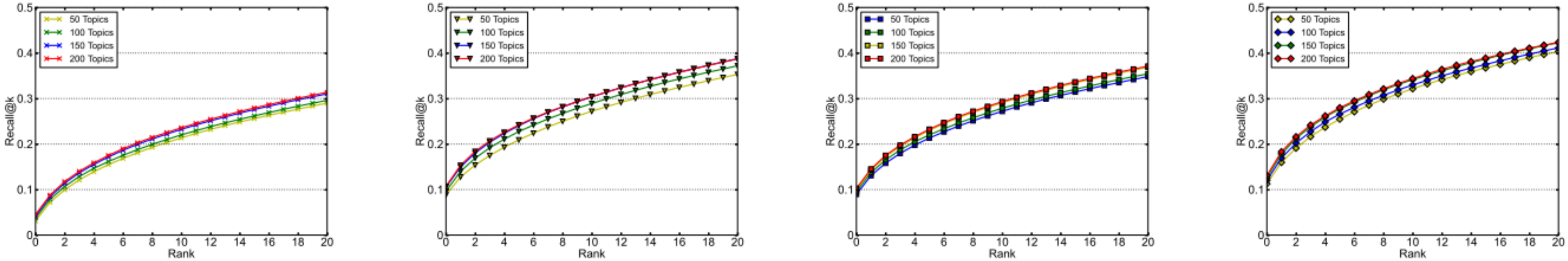


(a) Users Traveling in New Cities (b) Users Traveling in Home Cities

Figure 4: Top-k Performance on Foursquare

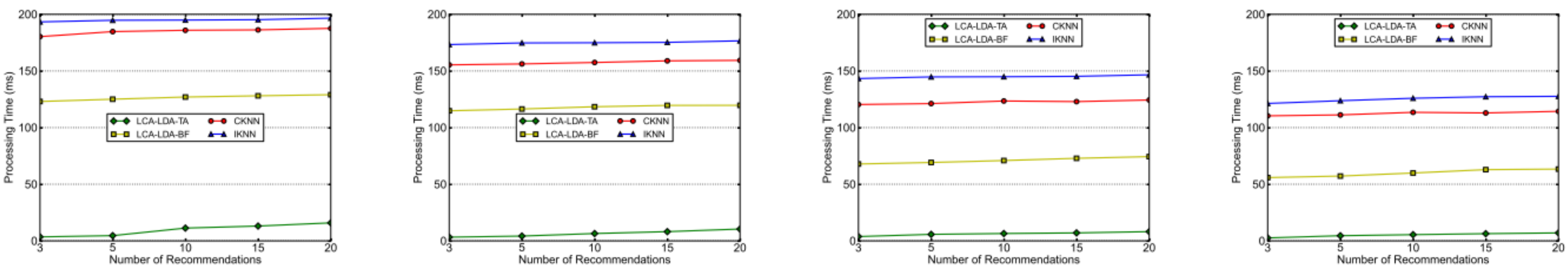
Experimental Results

- Results:



(a) LDA (b) LA-LDA (c) CA-LDA (d) LCA-LDA

Figure 5: Impact of the Number of Latent Topics



(a) Processing Time in Beijing (b) Processing Time in Shanghai (c) Processing Time in Guangzhou (d) Processing Time in Shenzhen

Figure 6: Efficiency w.r.t Recommendations

Conclusion

- This paper proposed a location-content-aware recommender system.
- Provides a user with spatial item recommendations within the querying city based on the individual interests and the local preferences.
- By taking advantage of both the content and location information of spatial items, the proposed system overcomes the data sparsity problem.
- The proposed scalable query processing technique improves the efficiency of the approach significantly.

THANK YOU