Joint Modeling of User Check-in Behaviors for Point-of-Interest Recommendation

Hongzhi Yin[1], Xiaofang Zhou[1]
Yingxia Shao[2], Hao Wang[3], Shazia Sadiq[1]

[1]The University of Queensland, [2]Peking University
[3]Institute of Software, Chinese Academy of Sciences

Outlines

- Introduction
- JIM Model
- Experiment
- Conclusion
- Q&A

Introduction

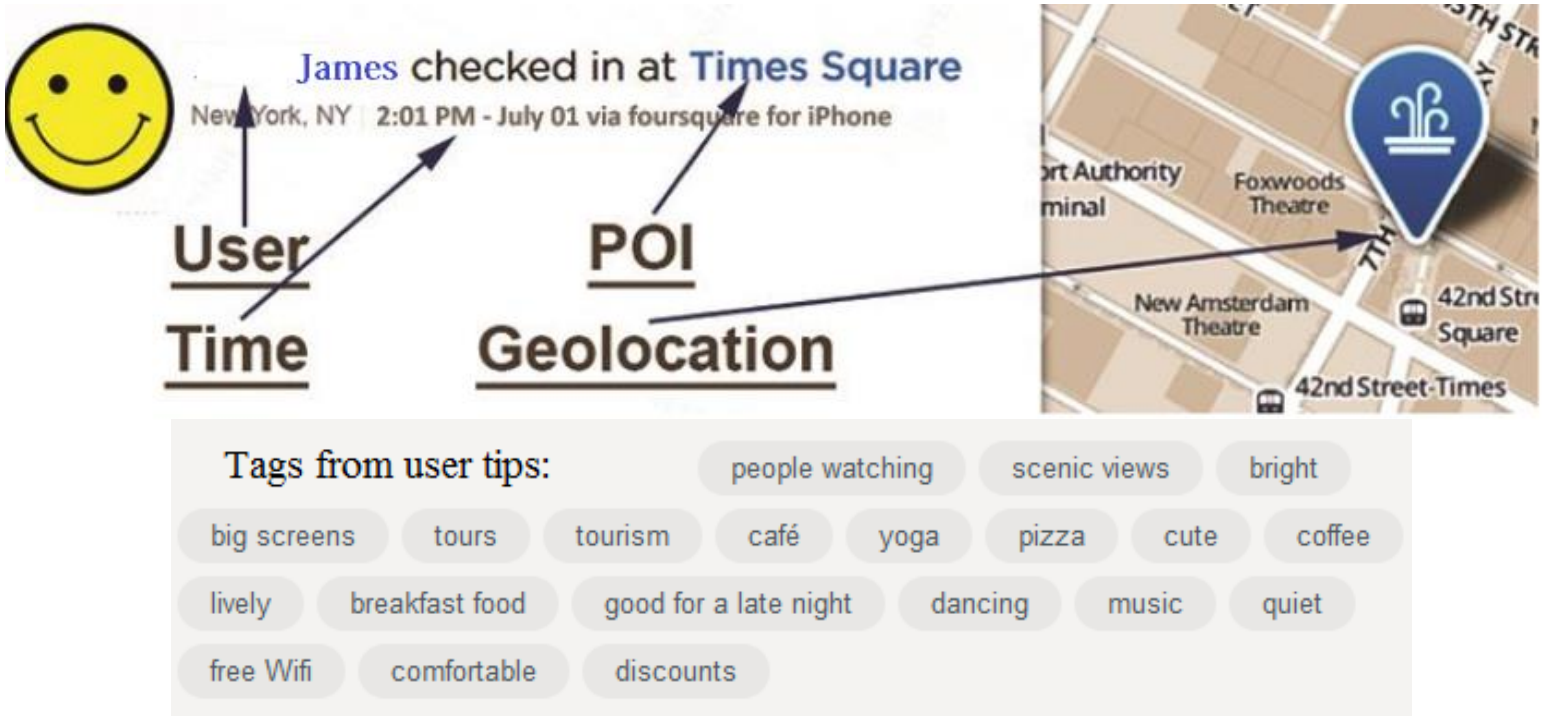
Motivation

- Location-based social network (LBSN) is popular
  - 65% of Americans access social networks on mobile devices
  - 48% of smartphone owners use LBSNs
  - Location-based marketing is expected to be a $39.87 billion business worldwide by 2016.

Refer to: http://www.marketsandmarkets.com/PressReleases/location-based-service.asp

Motivation

- Users can post their physical locations or geo-tag information via "check-in", and share their visiting experiences with their friends.



POI Contents

A check-in record consists of a five-tuple (user, POI-ID, check-in time, geo-location, check-in contents)

Motivation

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

- **Recommendation is important**
  - Too many choices!!
  - "best match" potentially billions of locations to billions of users globally.

PROBLEM 1. ***(POI Recommendation)*** *Given a check-in activity dataset D and a querying user u_q with his/her current location l_q and time t_q (that is, the query is $q = (u_q, t_q, l_q)$), our goal is to recommend a list of POIs that u_q would be interested in. Given a distance threshold d, the problem becomes an* ***out-of-town recommendation*** *if the distance between the target user's current location and his/her home location (that is, $|l_q - l_u|$) is greater than d. Otherwise, the problem is a* ***home-town recommendation****.*

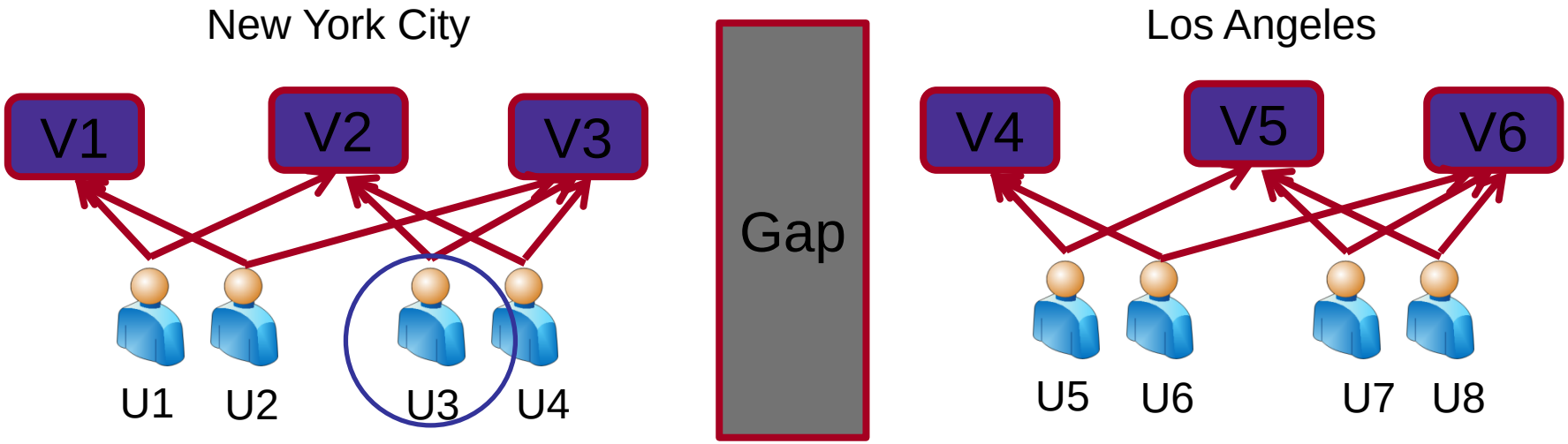
Challenges

- **Sparsity problem**
  - Billions of POIs around the world.
  - A user visits ~100 POIs.
  - The densities of user-POI matrix are 0.55% and 0.02% for Foursquare and Twitter datasets used in our experiments
- **Travel locality**[1]
  - 45% of users travel 10 miles or less, 75% travel 50 miles or less, and more than 97% users travel 100 miles or less
  - According to our investigations, the check-in records generated by users in their non-home cities are very scarce and only take up 0.47% of the check-in records left in their home cities, which aggravates the data sparsity problem with POI recommendation for out-of-town users.

[1] Levandoski *et al.* Lars: A Location-Aware Recommender System. In ICDE, 2012

Example of Out-of-Town Recommendation

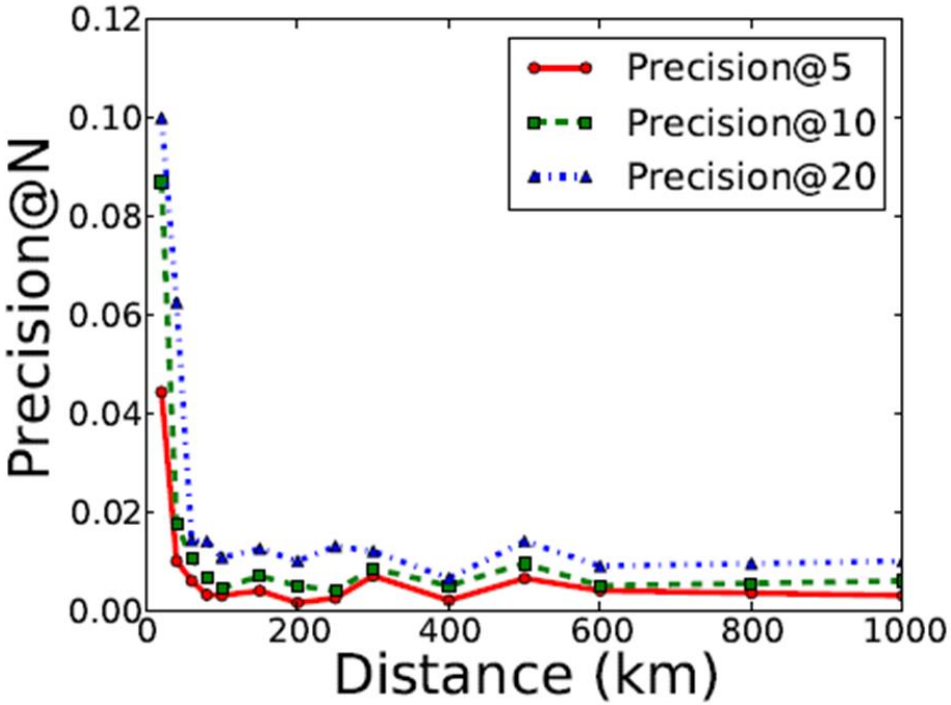
Existing Methods



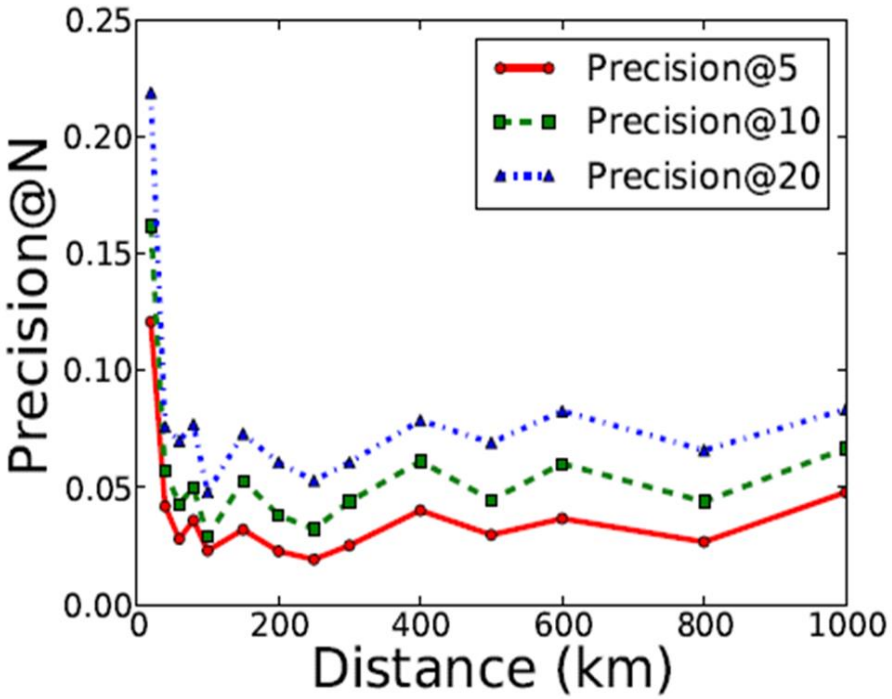
Travel Locality: When U3 travels to Los Angeles that is new to her since she has no check-in history there, how can we recommend POIs to her? In other words, how to link the users in one side to the items in the other side?

Both Graph-based methods and Collaborative Filtering methods would fail in this scenario.

Performance of Mainstream RS



Performance of Traditional user-based CF



Performance of Matrix Factorization

On the Foursquare Dataset

Our Solution - JIM

How Decision Process be Influenced

- **Geographical Influence:** People tend to explore POIs near their current positions or the ones that they have visited before. So, POIs visited by users often form spatial clusters, i.e., people tend to check in around several centers (e.g., “home” and “office”)
- **Temporal Effect**: Users’ activity contents exhibit strong temporal cyclic patterns. For example, a user is more likely to go to a restaurant rather than a night-life spot at lunch time.
- **Content Effect**: A user only prefer a small number of categories of POIs. The check-in activities of users exhibit a strong semantic regularity, e.g., the entropy of categories of POIs an individual user checks-in is very small.

How Decision Process be Influenced

- **Word-of-Mouth Effect:** The region-level popularity of POIs also affects user visiting behaviors. In fact, the probability of a user visiting a POI is largely affected by the local word-of-mouth about the POI, especially when users travel in unfamiliar regions.
  - Local attractions. For example, a professor never goes gambling when he lives in Beijing. But when he travels in Las Vegas, he is most likely to visit casinos.
- Need a model that jointly encodes the personal interests, geographical influence, temporal effect and word-of-mouth effect in a unified way to alleviate the issue of data sparsity for effective POI recommendation.

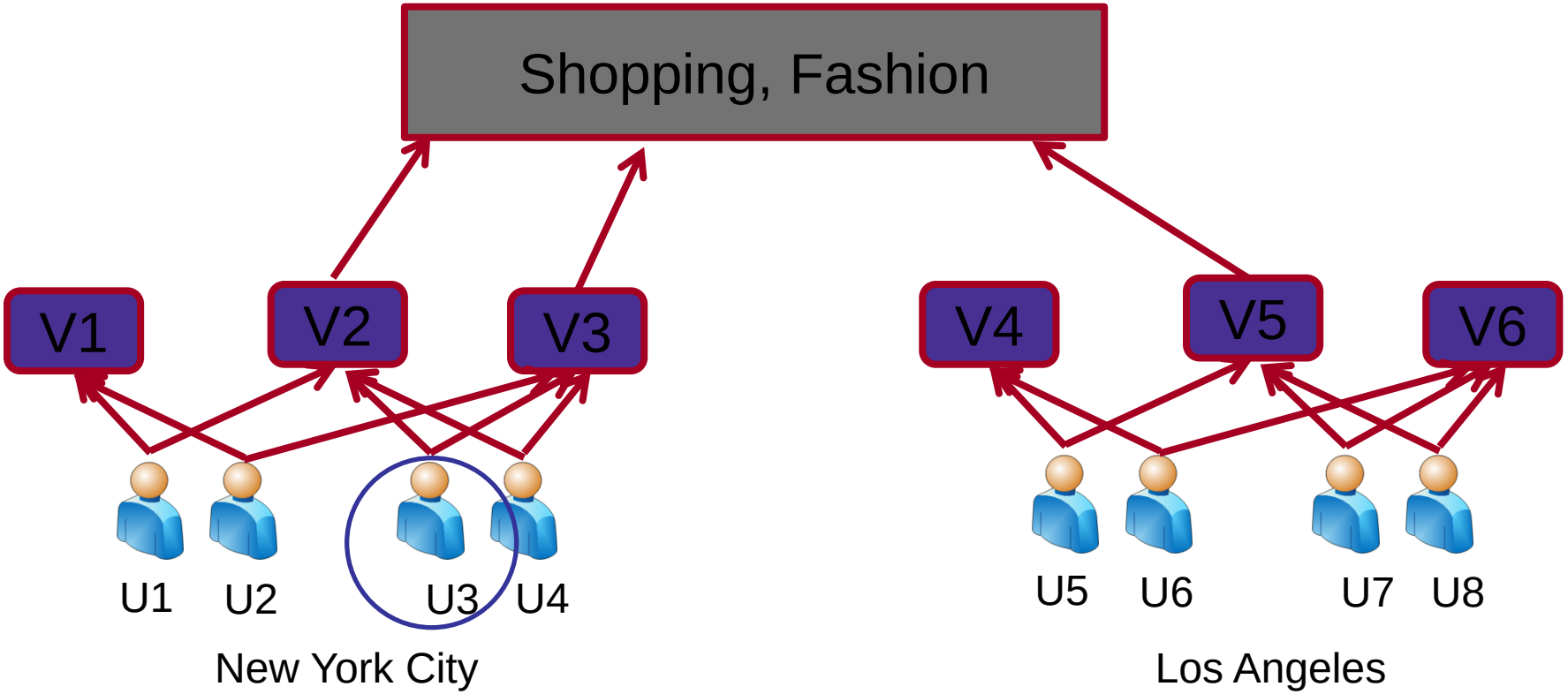
JIM Model

- As a POI has both semantic and geographical attributes, JIM consists of two components:
  - Time-aware User Interest Component (TIC) to exploit content effect and temporal effect to model personal interests
  - Popularity-aware User Mobility Component (PMC) to exploit geographical influence and word-of-mouth effect to account for users' spatial mobility

ID-Name: Darling Harbour
Location: Longitude: 151.200, Latitude: -33.877
Content: plaza, scenic views, harbour, bar, cafés, entertainment sunsets, harbors, museums, leisure, marina

JIM-TIC

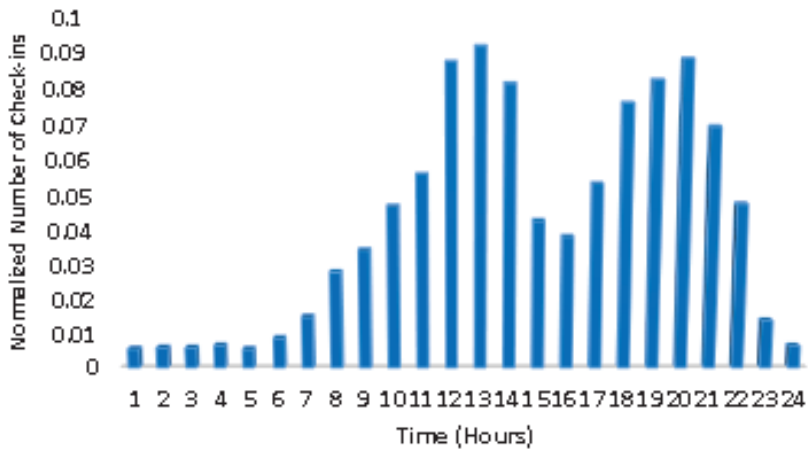
- TIC adopts latent topics to characterize users' interests, i.e., we infer individual user's interest distribution over a set of topics according to the contents (e.g., tags and categories) of her visited POIs.



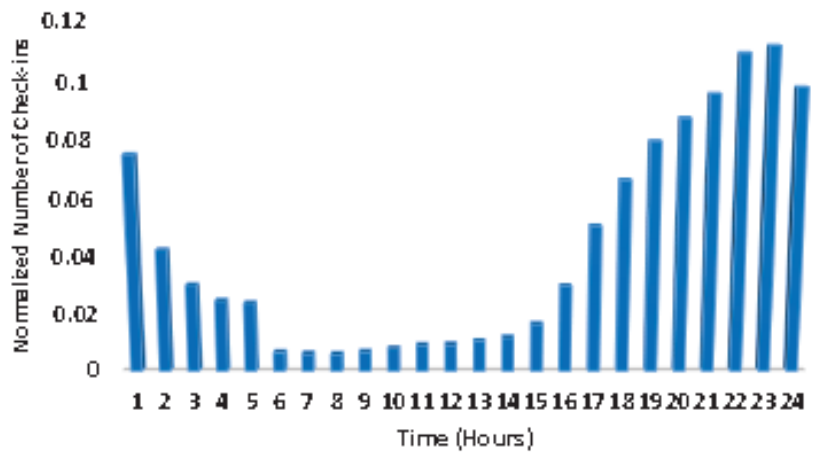
The contents play the role of medium which can transfer users' interests inferred at their home regions to out-of-town regions

JIM-TIC

- As users' interests and activities are influence by time, so the learnt topics should be not only semantic-coherent, but also time-aware.
- Each topic in our TIC is not only associated with a distribution over words, but also with a distribution over time. This design is also helpful to improve the quality of topics, since the check-in time provides important clues about the contents of POIs.



Topic: Restaurant & Eating

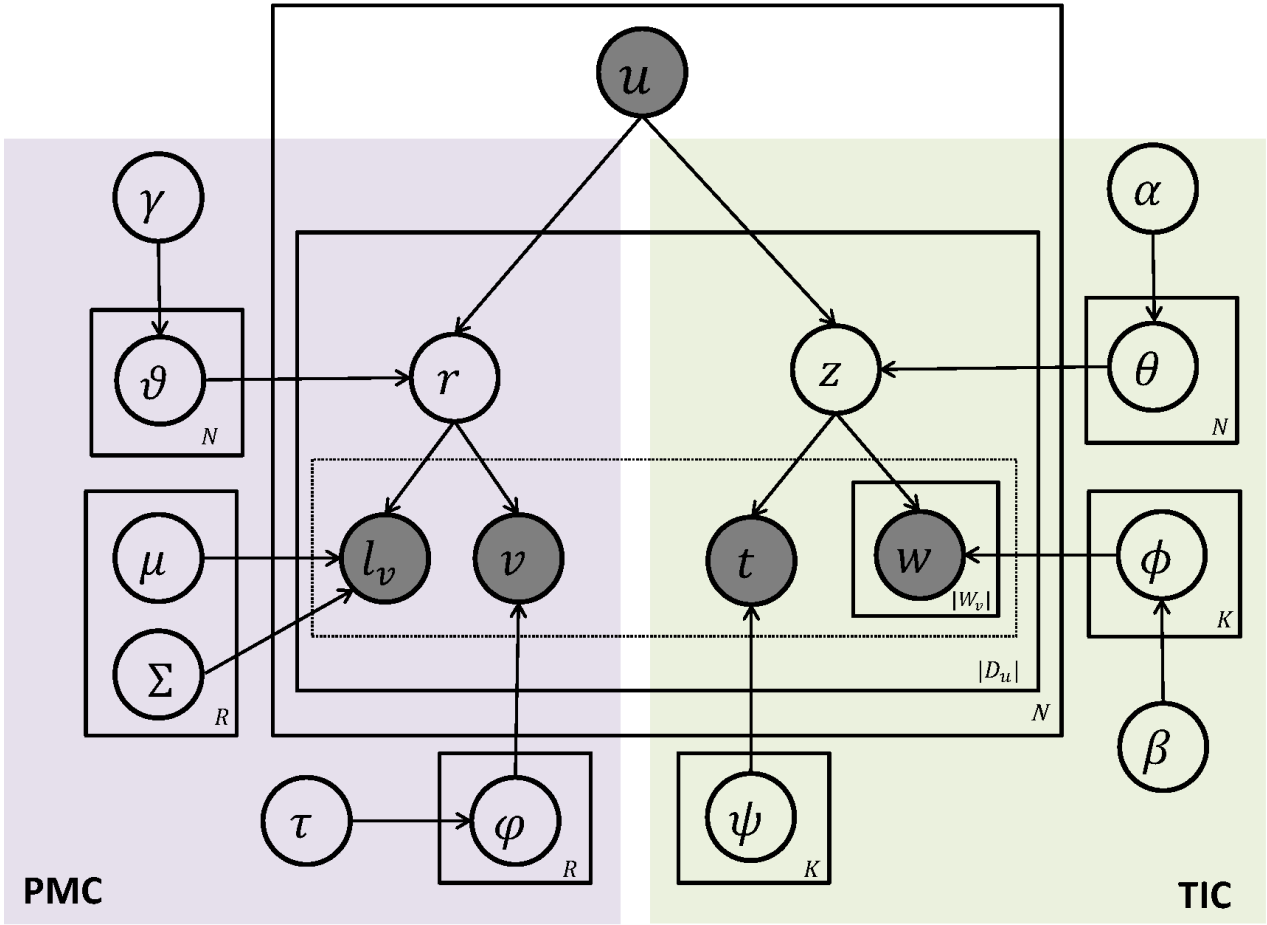


Topic: Night Life Spot & Entertainment

JIM-PMC

- The spatial clustering phenomenon indicates that users are most likely to visit a small number of POIs, and these POIs are usually limited to some geographical regions.
- The geographical space is divided into R regions. We assume a Gaussian distribution for each region r, and the location for POI v is characterized by: $l_v \sim \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$
- We estimate the probability that the user visits each region according to her historical visited POIs or her current location.
- Then, we infer the probability of the POI generated from each region, according to both the public's preferences at that region and the distance between the POI and the centre of that region.

Generative Process of JIM Model



for *each topic* z **do**
  Draw $\boldsymbol{\phi_z} \sim Dirichlet(\cdot|\beta)$;
end
for *each region* r **do**
  Draw $\boldsymbol{\varphi_r} \sim Dirichlet(\cdot|\tau)$;
end
for *each user* u **do**
  Draw $\boldsymbol{\theta_u} \sim Dirichlet(\cdot|\alpha)$;
  Draw $\boldsymbol{\vartheta_u} \sim Dirichlet(\cdot|\gamma)$;
end
for *each* D_u *in* D **do**
  for *each check-in* $(u,\ v,\ l_v,\ W_v,\ t) \in D_u$ **do**
    Draw a topic index $z \sim Multi(\boldsymbol{\theta_u})$;
    Draw a time $t \sim Beta(\psi_{z,1}, \psi_{z,2})$;
    for *each token* $w \in W_v$ **do**
      Draw $w \sim Multi(\boldsymbol{\phi_z})$;
    end
    Draw a region index $r \sim Multi(\boldsymbol{\vartheta_u})$;
    Draw a POI index $v \sim Multi(\boldsymbol{\varphi_r})$;
    Draw a location $l_v \sim \mathcal{N}(\boldsymbol{\mu_r}, \boldsymbol{\Sigma_r})$;
  end
end

Recommendation Using JIM

- After we estimate the model parameters using the collapsed Gibbs sampling, given a query user u_q with the query time t_q and location l_q, $i.e., q = (u_q, t_q, l_q)$, we compute the probability of each unvisited POI v w.r.t. the query q, and then select the top-k ones with highest visiting probabilities as recommendations.

$$P(v, l_v, W_v | u_q, t_q, l_q, \hat{\Psi}) \propto \sum_r \left[P(r) P(l_q | r, \hat{\Psi}) P(l_v | r, \hat{\Psi}) P(v | r, \hat{\Psi}) \sum_z P(z | u_q, \hat{\Psi}) P(t_q | z, \hat{\Psi}) \left(\prod_{w \in W_v} P(w | z, \hat{\Psi}) \right)^{\frac{1}{|W_v|}} \right]$$

$$= \underbrace{\boxed{\sum_r \left[P(r) P(l_v | \hat{\mu}_r, \hat{\Sigma}_r) P(l_q | \hat{\mu}_r, \hat{\Sigma}_r) \hat{\varphi}_{r,v} \right]}}_{\text{PMC}} \underbrace{\boxed{\sum_z \left[\hat{\theta}_{u_q,z} \hat{\psi}_{z,t_q} \left(\prod_{w \in W_v} \hat{\phi}_{z,w} \right)^{\frac{1}{|W_v|}} \right]}}_{\text{TIC}}$$

Experiment

Data sets

- Foursquare
  - Publicly available
  - Contain 483,813 check-in records of 4163 users who live in the California, USA
- Twitter
  - Publicly available
  - Contain 1,434,668 check-ins of 114,058 users who live across whole USA
- Distributions

Foursquare

Twitter

Comparative Approaches

- State-of-Art POI Recommendation Methods
  - LCA-LDA[1]
    - Consider both personal interests and local preferences of the target city
  - TCAF [2]
    - Time-Aware collaborative filtering method
  - UPS-CF [3]
    - A mixture of user-based collaborative filtering and social-based collaborative filtering
  - SVDFeature [4]
    - A feature-based matrix factorization model

[1] Yin *et al.* Lcars: A location-content-aware recommender system. In KDD, 2013

[2] Yuan *et al*. Time-aware point-of-interest recommendation. In SIGIR, 2013

[3] Ference *et al.* Location recommendation for out-of-town users in location-based social networks, In CIKM, 2013

[4] Chen *et al.* SVDFeature: a toolkit for feature-based collaborative filtering, In JMLR, 2012

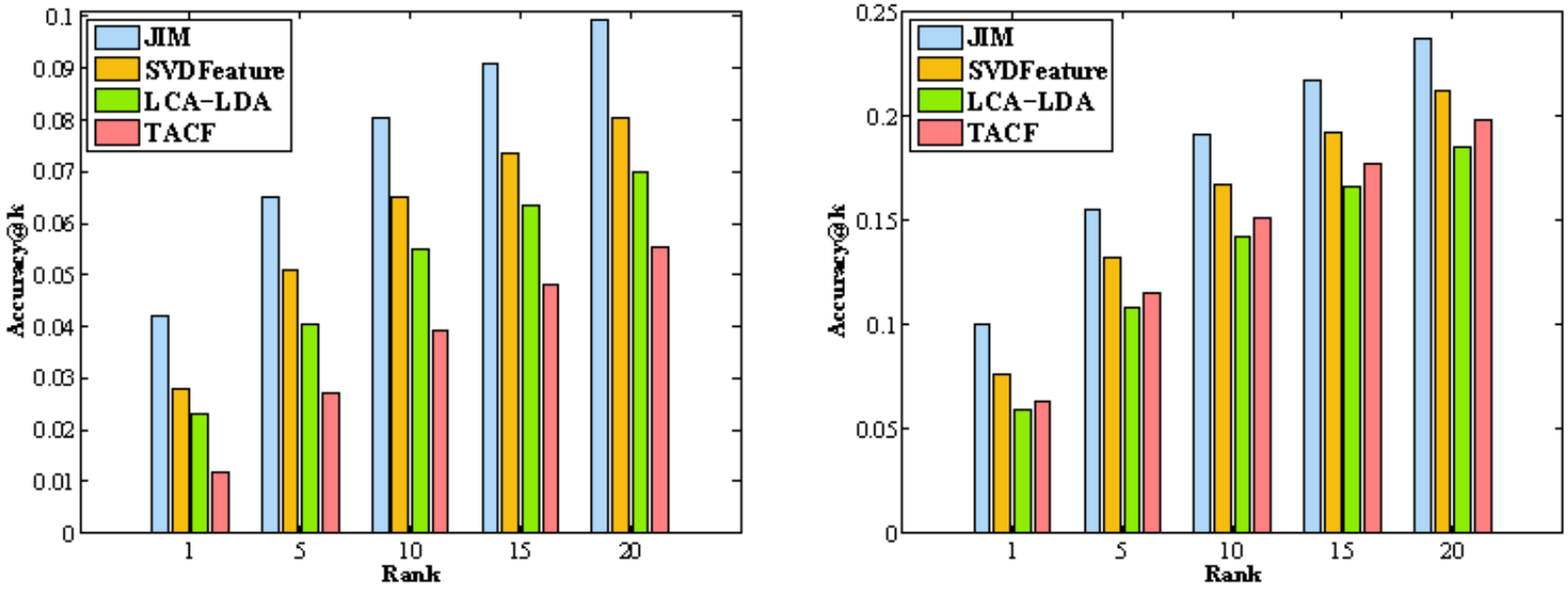
Comparative Approaches

- Variant versions of JIM
  - JIM-S1: Without consideration of temporal effect
  - JIM-S2: Without consideration of content effect
  - JIM-S3: Without consideration of word-of-mouth effect
  - JIM-S4: Without consideration of geographical influence

Methods \ Features	Spatial	Temporal	Social or Crowd	Textual
SVDFeature	♦	♦	♦	♦
LCA-LDA	♦		♦	♦
TACF	♦	♦		
UPS-CF	♦		♦	
JIM	♦	♦	♦	♦
JIM-S1	♦		♦	♦
JIM-S2	♦	♦	♦	
JIM-S3	♦	♦		♦
JIM-S4		♦	♦	♦

Features of Different Methods

Recommendation Effectiveness



(a) Out-of-town Recommendation (b) Home-town Recommendation

Figure 3: Top-k Performance on Twitter Dataset

1) The recommendation accuracies of all methods are significantly higher in the home-town scenario than in out-of-town scenario;
2) The same methods have different performances in these two recommendation settings, indicating that the two recommendation scenarios are intrinsically different, and should separately evaluated.

Impact of Different Factors

Methods	Out-of-Town Scenario			Home-Town Scenario		
	Ac@1	Ac@10	Ac@20	Ac@1	Ac@10	Ac@20
JIM-S1	0.052	0.108	0.134	0.101	0.204	0.253
JIM-S2	0.045	0.096	0.119	0.117	0.226	0.280
JIM-S3	0.049	0.101	0.125	0.112	0.219	0.271
JIM-S4	0.056	0.113	0.140	0.106	0.212	0.262
JIM	**0.062**	**0.121**	**0.149**	**0.124**	**0.241**	**0.298**

Table 4: Recommendation Accuracy on Foursquare Dataset.

Methods	Out-of-Town Scenario			Home-Town Scenario		
	Ac@1	Ac@10	Ac@20	Ac@1	Ac@10	Ac@20
JIM-S1	0.033	0.072	0.089	0.084	0.162	0.199
JIM-S2	0.028	0.064	0.079	0.095	0.179	0.220
JIM-S3	0.031	0.067	0.083	0.091	0.173	0.213
JIM-S4	0.037	0.075	0.093	0.088	0.168	0.206
JIM	**0.041**	**0.080**	**0.099**	**0.100**	**0.191**	**0.235**

Table 5: Recommendation Accuracy on Twitter Dataset.

According to the importance of the four factors, they can be ranked:

1) Content Effect > Word-of-Mouth > Temporal Influence > Geographical Influence for the out-of-town scenario;

2) Temporal Influence > Geographical Influence > Word-of-Mouth > Content for the home-town scenario.

Obviously, the content information plays a dominant role in overcoming the issue of data sparsity of out-of-town recommendation scenario, while the temporal effect is most important to improve home-town recommendation.

Impact of Different Factors

- The importance of the same factor w.r.t. the two recommendation scenarios is different.
- This is because the two recommendation scenarios have different characteristics:
  - Most of users have enough check-in records in their home towns while few check-in activities are left in out-of-town regions;
  - The limitation of travel distance in the out-of-town scenario does not matter as much as that in home town;
  - Users' daily routines may change when they travel out of town.

Conclusions

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

- To alleviate data sparsity and travel locality, we design a joint probabilistic generative model JIM to model users' check-in behaviours
  - Time-aware User Interest Component (TIC) : temporal effect and content effect
  - Popularity-aware User Mobility Component (PMC): geographical influence and word-of-mouth effect
- Study the importance of each factor in improving both home-town and out-of-town recommendation on two real datasets, and find
  - Content information and the wisdom of crowd play a dominant role in out-of-town recommendation scenario
  - Temporal influence and geographical influence are most important to improve home-town recommendation
- Most recommendation methods display different performance in these two recommendation scenarios, showing the two recommendation scenarios are intrinsically different, and should separately evaluated.

Thank you!