A personalized Smart Tourism Recommender System based on social media data

D. Ferone, E. Fersini, E. Messina

3 Airo Young Workshop, 28th March 2019

Department of Informatics, Systems and Communication, University of Milano-Bicocca

SmartCAL project

SmartCAL project

Empowerment of the tourism offer of a region based on the preferences and needs of the modern tourist

SmartCAL project

Empowerment of the tourism offer of a region based on the preferences and needs of the modern tourist

Sites managers

- sentiment analysis about POIs;
- viral marketing.

SmartCAL project

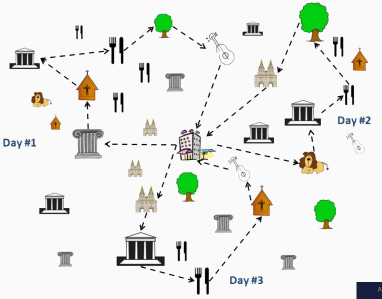
Empowerment of the tourism offer of a region based on the preferences and needs of the modern tourist

Sites managers

- sentiment analysis about POIs;
- viral marketing.

Tourists

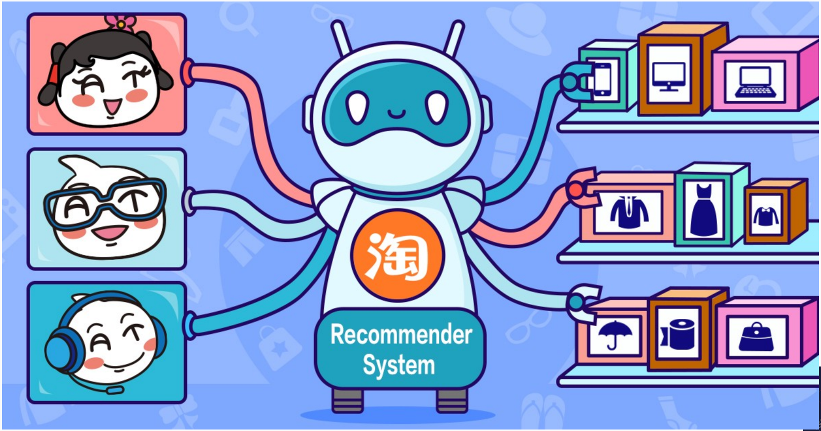
- visit plan on one or more days;
- integration with public transportation.



The idea

Recommender system

The goal of a recommender system (RS) is to find and suggest interesting and relevant elements, based on users' tastes and preferences



Is preference enough?

I like sea!

Is preference enough?

I like sea!

Is preference enough?

I like sea!

Is preference enough?

I like sea!

Is preference enough?

I like sea!

Sentiment Analysis on media data

- Perform sentiment analysis on reviews and calculate a *sentiment score*

Sentiment Analysis on media data

- Perform sentiment analysis on reviews and calculate a *sentiment score*
- Mediate the user preferences with sentiment score

Understand the preferences

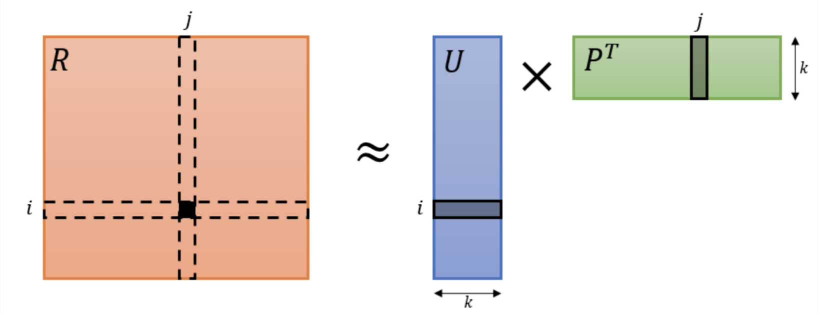
Ratings matrix

John	5	1	3	5
Tom	?	?	?	2
Alice	4	?	3	?

Ratings matrix

Tourist Category	Number of POIs	Tourist Profile	Description
Abbeys, Hermitages and Monasteries	36	*Religious*	A type of tourist mainly interested in visiting religious-related places, such as cathedrals and churches.
Cathedrals, Churches and Shrines	34		
Architecture and Landscape	29	*Historical*	A type of tourist mainly interested in visiting historical-related places, such as headlights, castles and bridges.
Excellences in Art	13		
Historical Headlights	2		
Itineraries	6		
Historical Bridges	6		
Archaeology	17	*Artistic*	A type of tourist mainly interested in visiting art-related places, such as museums and theatres.
Places of Art	10		
Museums	58		
Theatres	24		
Historical Places	31	*Food Lover*	A type of tourist mainly interested in visiting food-related places, such as bars.
Restaurants	35		
Public Gardens	12	*Nature and Relax Lover*	A type of tourist mainly interested in visiting nature-related and relax-related places, such as lakes, parks and gardens.
Lakes	8		
Parks	10		
Amusement Parks	11		
Cycle Paths	2		
Thermal Baths	3		
Tourist Trains	3		
Total: 20	**Total: 350**	**Total: 5**	

Latent Factor Model



R: Ratings matrix

U: User latent features

P^T: Tourist categories latent features

$$\min_{u,p} \sum_{i,j \in R} (R_{i,j} - u_i p_j^T)^2 + \gamma \left(\sum_{i=1}^{n} ||u_i||^2 + \sum_{j=1}^{m} ||p_j||^2 \right)$$

The user is more interest in the top-k POIs

Learning to Rank

The user is more interest in the top-k POIs

RankSVM is a LtR model:

input: latent features

output: ranking of POIs

Sentiment analysis

Time changes sentiment

Time changes sentiment

$$w(R) \equiv M \cdot e^{kD},$$

- M: max value;
- D: number of days;
- $k < 0$: decay constant.

Time changes sentiment

$$w(R) \equiv M \cdot e^{kD},$$

- M: max value;
- D: number of days;
- $k < 0$: decay constant.

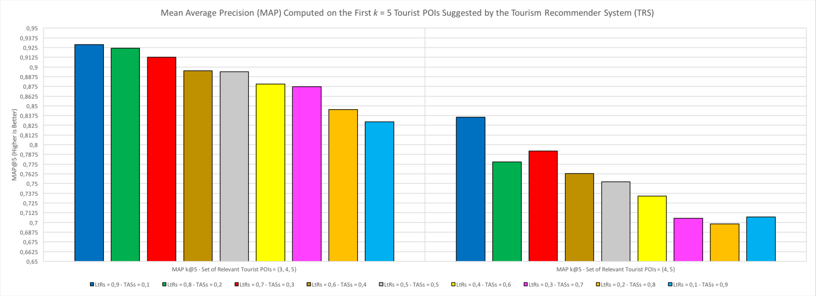
$$FS = \alpha \cdot \text{category score} + (1 - \alpha) \cdot \text{sentiment score}$$

Experiments

Dataset

- 350 POIs;
- TripAdvisor reviews;
- Ratings matrix with 5000 simulated users and 20 categories
- 25000 rating values
- 75 reals users

Results



Mean Average Precision (MAP) Computed on the First k = 5 Tourist POIs Suggested by the Tourism Recommender System (TRS)
MAP@5 (Higher is Better)
0,95
0,9375
0,925
0,9125
0,9
0,8875
0,875
0,8625
0,85
0,8375
0,825
0,8125
0,8
0,7875
0,775
0,7625
0,75
0,7375
0,725
0,7125
0,7
0,6875
0,675
0,6625
0,65
MAP k@5 - Set of Relevant Tourist POIs = {3, 4, 5}
MAP k@5 - Set of Relevant Tourist POIs = {4, 5}
LtRs = 0,9 - TASs = 0,1
LtRs = 0,8 - TASs = 0,2
LtRs = 0,7 - TASs = 0,3
LtRs = 0,6 - TASs = 0,4
LtRs = 0,5 - TASs = 0,5
LtRs = 0,4 - TASs = 0,6
LtRs = 0,3 - TASs = 0,7
LtRs = 0,2 - TASs = 0,8
LtRs = 0,1 - TASs = 0,9

AIROYoung

Conclusions and future work

- Recommender system is able to predict relevant POIs
- System in production
- Introduce the concept of Serendipity

01000101 01001110 01000100

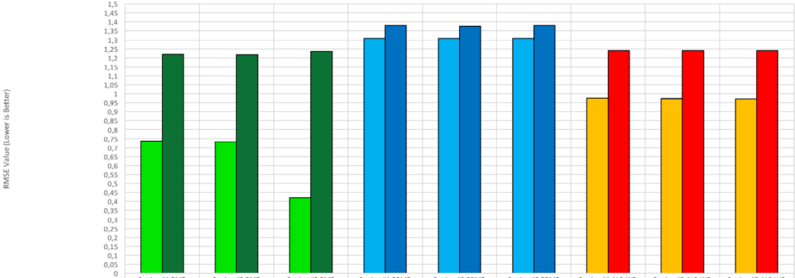
(**E** **N** **D**)

Thank you.

Latent Factor Model

- Probabilistic Matrix Factorization (Stochastic Gradient Descent);
- Bayesian PMF (Markov Chain Monte Carlo);
- Alternating Least Squares Weighted Regularization (Alternating Least Squares).

Latent Factor Model



	Setting #1 PMF	Setting #2 PMF	Setting #3 PMF	Setting #1 BPMF	Setting #2 BPMF	Setting #3 BPMF	Setting #1 ALS-WR	Setting #2 ALS-WR	Setting #3 ALS-WR
RMSE Value on the Training Data Set	0,735107	0,734075	0,420906	1,308920	1,308518	1,309065	0,975066	0,972040	0,970836
RMSE Value on the Testing Data Set	1,220643	1,217411	1,235429	1,379551	1,375454	1,379590	1,239613	1,239907	1,240516
Model Training and Testing Time	6,6311259	7,0639870	7,0321720	1,8864229	1,8363421	1,9844551	13,4015429	12,1312101	9,1136730