Building Useful Recommender Systems for Tourists

Francesco Ricci
Information and Database Systems Engineering
Free University of Bozen-Bolzano
fricci@unibz.it

unibz
Fakultät für Informatik
Facoltà di Scienze e Tecnologie informatiche
Faculty of Computer Science

Content

- Recommender systems – the *classical view*
- What makes **good** a recommendation
- How a RS can identify “good recommendations”?
- Modelling:
  - Groups of users with **similar behaviours may reveal the hidden utility of choices**
  - Expected utility is a **function** of item’s **features** and **context**
- *Inaccurate* recommendations for **single users that** deviates from the predicted choice may be good recommendations.

What we like may not be what we choose

Classical Recommendation Model

Three types of entities: **Users, Items and Contexts**

1. A **background knowledge**:
   - A set of **ratings – preferences**
     - *r: Users x Items x Contexts → {1, 2, 3, 4, 5}*
   - A set of “features” of the Users, Items and Contexts
2. A method for **<u>predicting</u>** the function *r* where it is unknown:
   - *r*(u, i, c) = Average ratings r(u’, i, c’): users u’ are similar to u and context c’ is similar to c*
3. A method for **<u>selecting</u>** the items to recommend (choice):
   - In context *c* recommend to *u* the item *i* with the largest predicted rating r*(u,i,c)*

This process should identify items that the user will happily choose

Predicting Choices

- More recent models - based on user action observations - predict choices (e.g. sequences of movie views)
- Ironically, they claim to be able to predict user preferences
- None is able to decouple preferences from choices.
- Some models can combine actions and preferences [Lavee et al., 2019]

G. Lavee, N. Koenigstein, O. Barkan. When actions speake louder than clicks: a combined model of purchase probability and long-term customer satisfaction, RecSys 2019.

Context Aware RS Algorithms

- Reduction-based Approach, 2005
- Exact and Generalized Pre Filtering, 2009
- Item Splitting, 2009
- Tensor Factorization, 2010
- User Splitting, 2011
- Context-aware Matrix Factorization, 2011
- Factorization Machines, 2011
- Differential Context Relaxation, 2012
- Differential Context Weighting, 2013
- UI splitting, 2014
- Similarity-Based Context Modelling, 2015
- Convolutional Matrix factorization, 2016
- Contextual bandit, 2018

Preference prediction

Knowing your goals

- *"what do I want?"* – addressed largely through internal dialogue
  - Depends on how a choice will make us feel
  - Not an easy task
- **Future:** what you expect an experience will make you feel is called **expected utility**
- **Present:** The way an item (movie, travel, etc.) makes you feel in the moment is called **experienced utility**
- **Past:** Once you had an experience (e.g. a movie), future choice will be based on what you remember about that: **remembered utility**.

Recommender Systems Limitations

- They analyse past experiences to predict the goodness of future experiences
- They can hardly predict our best choice because they do not know what options we are considering and how we feel now
- They build models collapsing all the recorded user’s experiences (ratings) in a single time point
- Sequential recommenders assumes that people repeat the same sequence of choices.

Good Travel Recommendations

- *When is cost effective*
- *When is liked by people that likes what we like*
- *When is good for the full family*
- *When we did not yet think about that*
- *When is not what we did last year*
- *When it has the features that we usually like*
- *When it has some impressive features*
- *When it is much better than other options*
- *When it is similar to what we did previous years*
- *When they are quite diverse*
- *When the weather will be great.*

Do you still believe that by simply mining a data set of users' ratings or choices we can generate good travel recommendations?

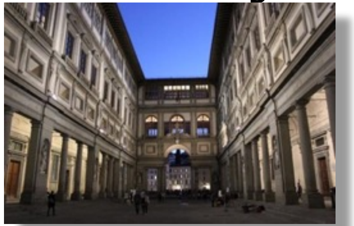
We need to **structure** the **knowledge** that can be derived from the data!

We need to better understand the **current** user's **goal**!

Recommendation Lists

List 1

- Gallerie degli Uffizi



- Piazzale Michelangelo

- Duomo

List 2

- Museo di San Marco

- Piazza S. Croce

- Santa Maria Novella

Recommendation Lists

List 1

- Gallerie degli Uffizi

- Piazzale Michelangelo

- Duomo

List 2

- Museo di San Marco

- Piazza S. Croce

- Santa Maria Novella

Points of Interest

- A number of features and contextual factors influence the pre-visit evaluation of a POI – intention to visit
  - **Traveller's knowledge of the place**
  - What she has already visited and when
  - Pictorial representation
  - Distinguished features
  - Travel party
  - Previous knowledge/usage of the app/recsys
  - **Popularity, fashionableness, trendiness, fame, prominence, prestige, reputation, visibility, rank.**

Expected vs. Experienced Utility

- *Should the system optimize expected or experienced utility?*
- *Should the system use behavioural data or ratings/judgements data?*

Expected

- \+ Match the **user values** at decision time
- – Match the **bias** of user's judgement
- \+ Based on an **unbiased sample** of observations
- – Based on observations without **meaning.**

Experienced

- \+ **Explicit** user assessment
- – **Incomplete** data
- \+ Depends on the consequence of choice (**context** is used)
- – Depends on the consequence of choices (outcome **wrong** but choice is **right**)

Good Travel Recommendations

- Criteria for establishing quality recommendations are highly subjective and contextual
- In practice is often impossible to predict what is a good recommendation for you now
- *Is it better to understand and match the user heuristics or use <u>solid</u> data mining prediction methods?*


ZAPATA
BREITLING

Grouping People

- We have recently addressed some of these problems with techniques that make use of groups
- **Group** and **model travellers** with observable similar behaviour and optimize the recommendations for them – not purely individual recommendations.

D. Massimo, F. Ricci:Harnessing a generalised user behaviour model for next-POI recommendation. RecSys 2018: 402-406

Behaviour and Recommendation

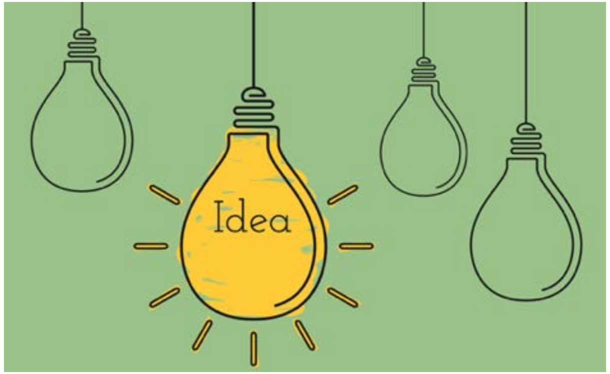
- **Behaviour learning** and **recommendation** should be **decoupled**
- The **learned behavioural model**, e.g., what points of interest a user is likely to visit may produce **uninteresting recommendations**
- Recommendation should also come from **expert knowledge** and the **optimization** of criteria the determine the behaviours (expected utility).

Behavioral Model Learning

- Learning user behaviour, but suggest to deviate from the usual behaviour

  - *The user is predicted to take a coffee at 8:00 at Walter Bar*
    - *The system suggests to get coffee at Rosy Bar – it is cheaper and better*

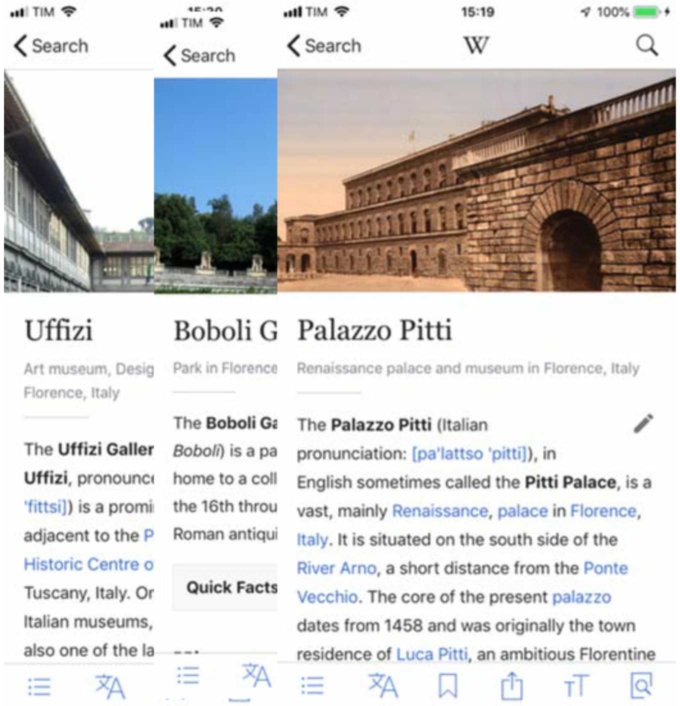
We must understand that the user likes good Italian and cheap coffe – not that he likes to go to Walter Bar at 8:00!



Grouping Travellers

Clustering Users' Visit Trajectories

- One visit to Florence:
  - Pitti Palace; Boboli Garden; Uffizi Museum
- Extract important keywords and combine them into a document visit
- Cluster visit documents
- Each cluster models a group of similar behaviours



5 Clusters in Florence

#Term	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
1	morning	hot	cloudy	warm	freezing
2	cold	afternoon	cold	cloudy	cloudy
3	square	century 16	church	century 14	afternoon
4	palace	palace	square	church	century 14
5	century 15	church	century 13	square	palace
6	century 13	square	palace	building	building
7	church	century 19	rain	palace	century 13
8	night	century 13	museo	ponte	church
9	dante	museo	brunelleschi	century 13	foggini
10	century 10	brunelleschi	tadda	century 19	century 19
#Traj.	368	339	341	297	153

1663 geo-localized temporally ordered trajectories of users' POI-visits, recorded via GPS sensors in the historic centre of Florence (Italy)

Inverse Reinforcement Learning

- **Assumption:** *the reward obtained by visiting a POI is determined by the POI's* ***features*** *and the visit* ***context***
- Inverse Reinforcement Learning **estimates** the hidden reward function (expected utility) that the users in a cluster apparently tried to maximise with the observed behaviour
- The reward is a **function** of the selected features and context
- The users choose visit actions with the **largest expected reward** (Q function).

Generating Recommendations

- Recommend to a user what is learned to be optimal for all the users in his cluster

	Q-BASE	SKNN
Reward@1	**0.073**	**-0.007**
Precision@1	**0.043**	**0.109**
Novelty@1	**0.061**	**0.0**
Reward@5	0.032	-0.010
Precision@5	0.045	0.068
Novelty@5	0.122	0.0

Have we correctly interpreted the user behaviour?

Alternative POI sample

- POIs were identified by action observation - *not corresponding to renowned ones*
- We repeated the test considering the subset of identified POIs present in TripAdvisor attractions – more popular

	Q-BASE	**SKNN**
Reward@1	0.369	0.097
Precision@1	**0.101**	**0.108**
Novelty@1	0.244	0.030
Reward@5	0.037	-0.061
Precision@5	0.056	0.062
Novelty@5	0.629	0.307

Now the precision is very similar

Why precision is a bad metric

- If we optimize for precision the system will learn to recommend the items that the user found autonomously – not «useful» recommendations
- When the precise recommendations are finished (already recommended) the system is unable to find novel recommendations
- Measured precision is typically very low (10% in our data set) – so the user is mostly exposed to imprecise recommendations.

S. M. McNee, J. Riedl, and J. A. Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06).

Advantages of the IRL method

- It is based on sequence mining but it can also *generalise* and suggest items never consumed before
- By grouping users it can fix the errors of models tuned individually on poorly represented users (few or erroneous data)
- Recommendations are not the predicted actions, they are optimal in some sense (which can be further tuned if we know the user values – expected utility).

Lesson Learned

- Distinction between **preferences** and **choices**
- Individual preference or behavior learning does not suffice – we need **choice modeling** (expected utility)
- Useful **recommendations** may be generated by deviating from the precited behavior (imprecise)
- Individual recommendation may be generated by assuming that **groups of similar user** are driven by a **hidden utility function.**

Thanks

- In particular to my students and collaborators who contributed to develop these ideas:
  - David Massimo
  - Linas Baltrunas
  - Laura Bledaite
  - Marius Kaminskas
  - Marko Gasparic
  - Marko Tkalcic
  - Matthias Braunhofer
  - Mehdi Elahi
  - Saikishore Kalloori
  - Tural Gurbanov
  - Thuy Ngoc Nguyen