Semantically-Enhanced Pre-Filtering for Context-Aware Recommender Systems

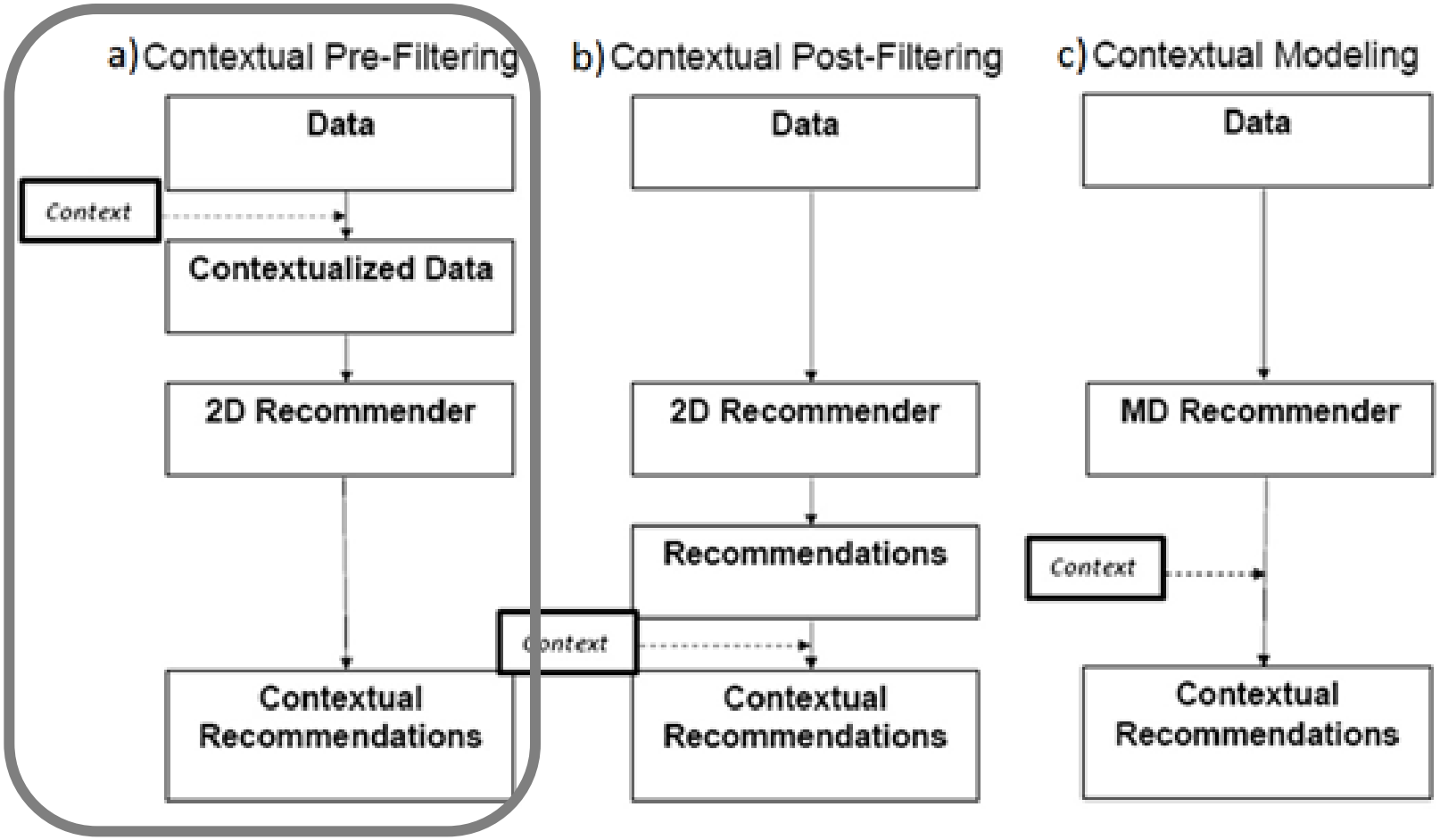
CaRR 2013

Victor Codina

Francesco Ricci

Luigi Ceccaroni

Main paradigms for exploiting contextual information in CARS



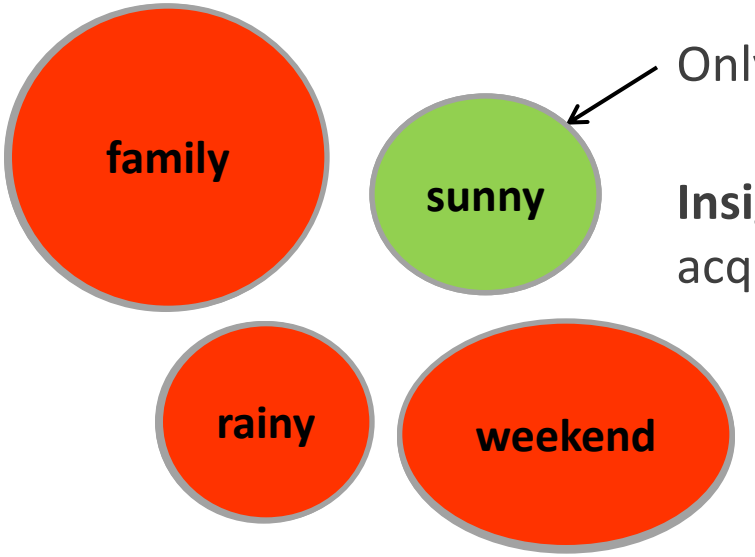
[Adomavicius & Tuzhilin, 2010]

Pre-filtering is very prone to suffer from sparsity-related problems

In traditional pre-filtering only the ratings acquired in **exactly** the same context of the target user are used

An example (places of interest recommender):

target context = sunny



Only this set of ratings would be used

Insight: Perhaps we can also use ratings acquired in **similar** contexts…

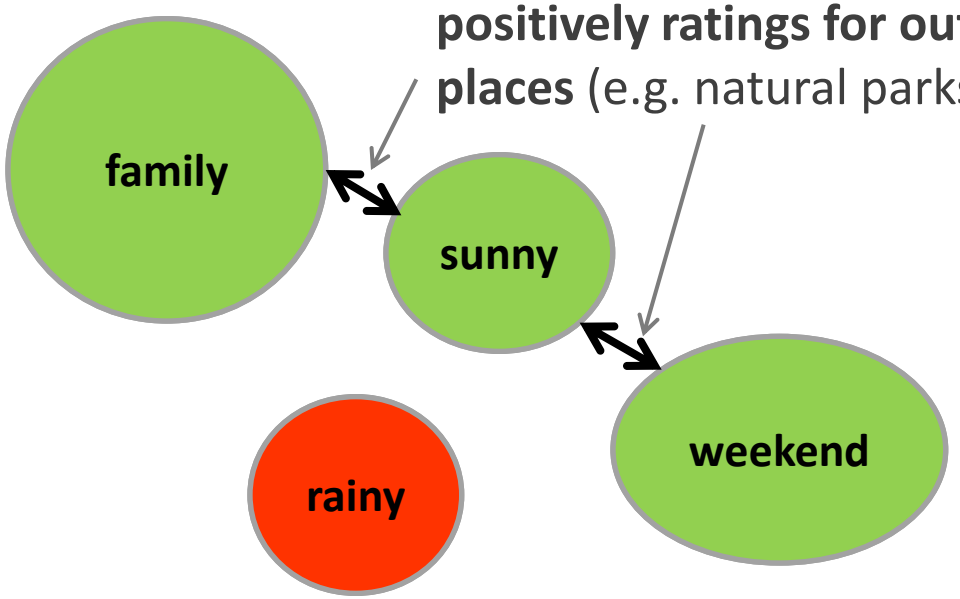
Our solution is based on using also ratings acquired in semantically similar contexts

We consider two contexts as **semantically similar** if they have a **similar effect on the users rating behavior**

Same example:

Target context = sunny

We know that sunny, family and weekend **influence positively ratings for outdoor places** (e.g. natural parks)



Outline

- Semantics acquisition
- Semantics exploitation
- Experiments

Outline

- **Semantics acquisition**
  - **Implicit semantics**
  - **Vector Space Model**
  - **Similarity calculation**
- **Semantics exploitation**
- **Experiments**

The implicit semantic similarity is based on the distributional hypothesis

In implicit semantics the **meaning** of a concept is **captured by its usage**

distributional hypothesis:

"concepts that share **similar usages** share **similar meaning**"

In Information Retrieval (IR) **usages are portions of textual data**:

- document
- paragraph
- sentence

The Vector Space Model (VSM) is normally used to find the implicit semantic similarity

Two concepts are similar if their usages overlap

An example of concept-usage matrix in IR (WordSpace)

usage = sentence (s1)

frequency-based weight

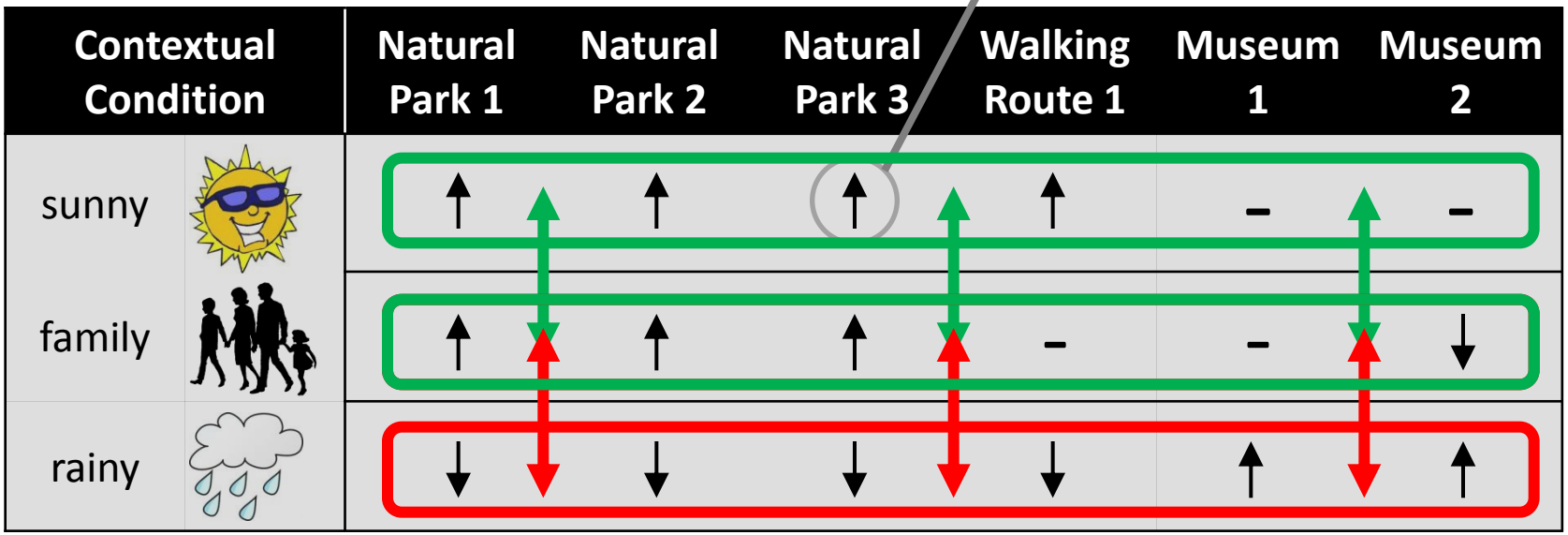
Concept	s1	s2	s3	s4	s5	s6	s7
glass 	1	1	0	1	0	2	0
wine 	2	1	0	0	1	2	0
spoon 	0	1	1	1	0	0	2

Two conditions are semantically similar if they influence the user's ratings similarly

We measure the influence of a condition from a **item-centered** perspective

usage = item (e.g. a museum)

Real value indicating the influence of a condition on the item's ratings (positive, neutral or negative)



Outline

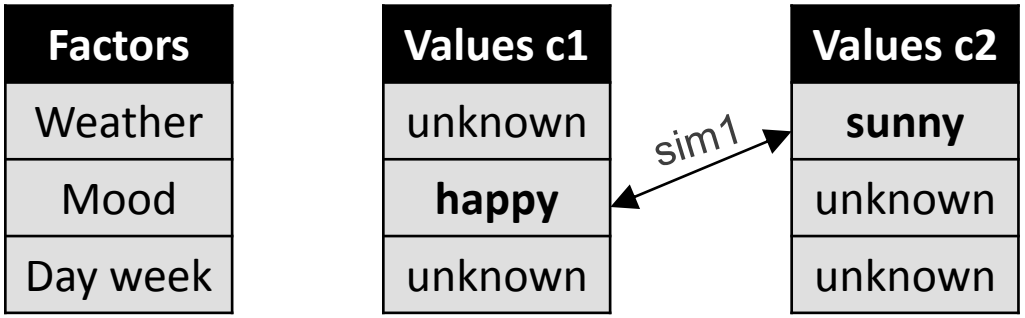
- Semantics acquisition
- Semantics exploitation
  - Comparing contexts
  - Pre-filtering algorithm
- Experiments

Similarities among contexts are calculated from the similarities of their conditions

A context can be defined by a single or multiple conditions

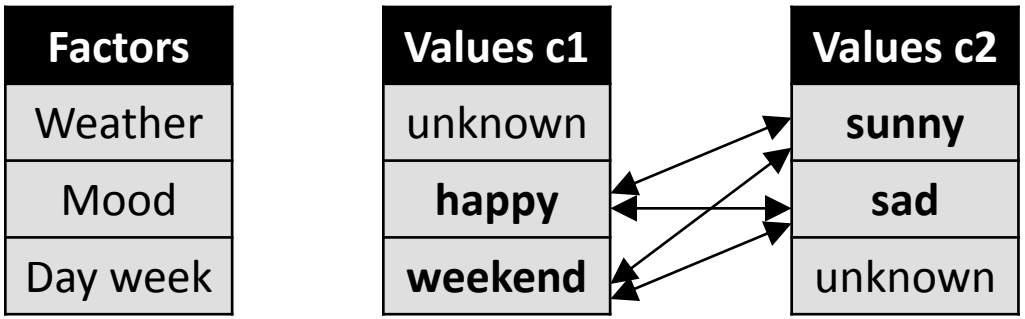
Two main scenarios when comparing two contexts:

1. Both contexts (c1, c2) have a **single** condition



Sim(c1,c2) = sim1

2. At least one of the contexts has **multiple** conditions



Sim(c1,c2) = **pairwise** combination

A similarity threshold is used to decide if two contexts are semantically similar

This threshold can be learnt at **different granularities**

We have experimented with two methods for learning the optimal threshold (β) from training data:

- Using a **global threshold** for all the possible target contexts
- Using a **different threshold** per target context

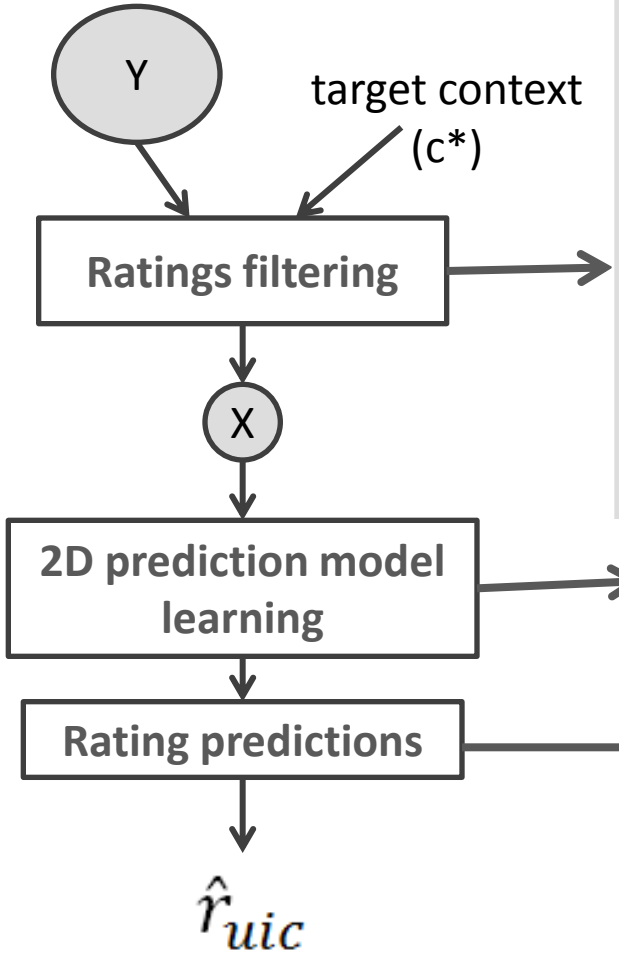
Sim(c1,c3)

	c1	c2	c3
c1	1	0.6	0.4
c2	0.7	1	0.2
c3	0.4	0.1	1

Similar contexts

Target context	Global (β=0.5)	Different threshold (β_1=0.7 β_2=0.5 β_3=0.3)
c1	c2	None
c2	c1	c1
c3	None	c1

Our pre-filtering approach uses also ratings whose context is similar to the target one



FOR EACH (user **u**, item **i**) in **Y** **DO**
  IF r_{uic*} exists **THEN**
    ADD r_{uic*} to **X**
  ELSE
    GET all r_{uic} where Sim(c,c*) ≥ β
    ADD rating average to **X**
  END IF
END FOR

Matrix Factorization (MF) model learnt using Stochastic Gradient Descent

$$\mu + b_u + b_i + q_i^T p_u$$

Overall rating average

User bias

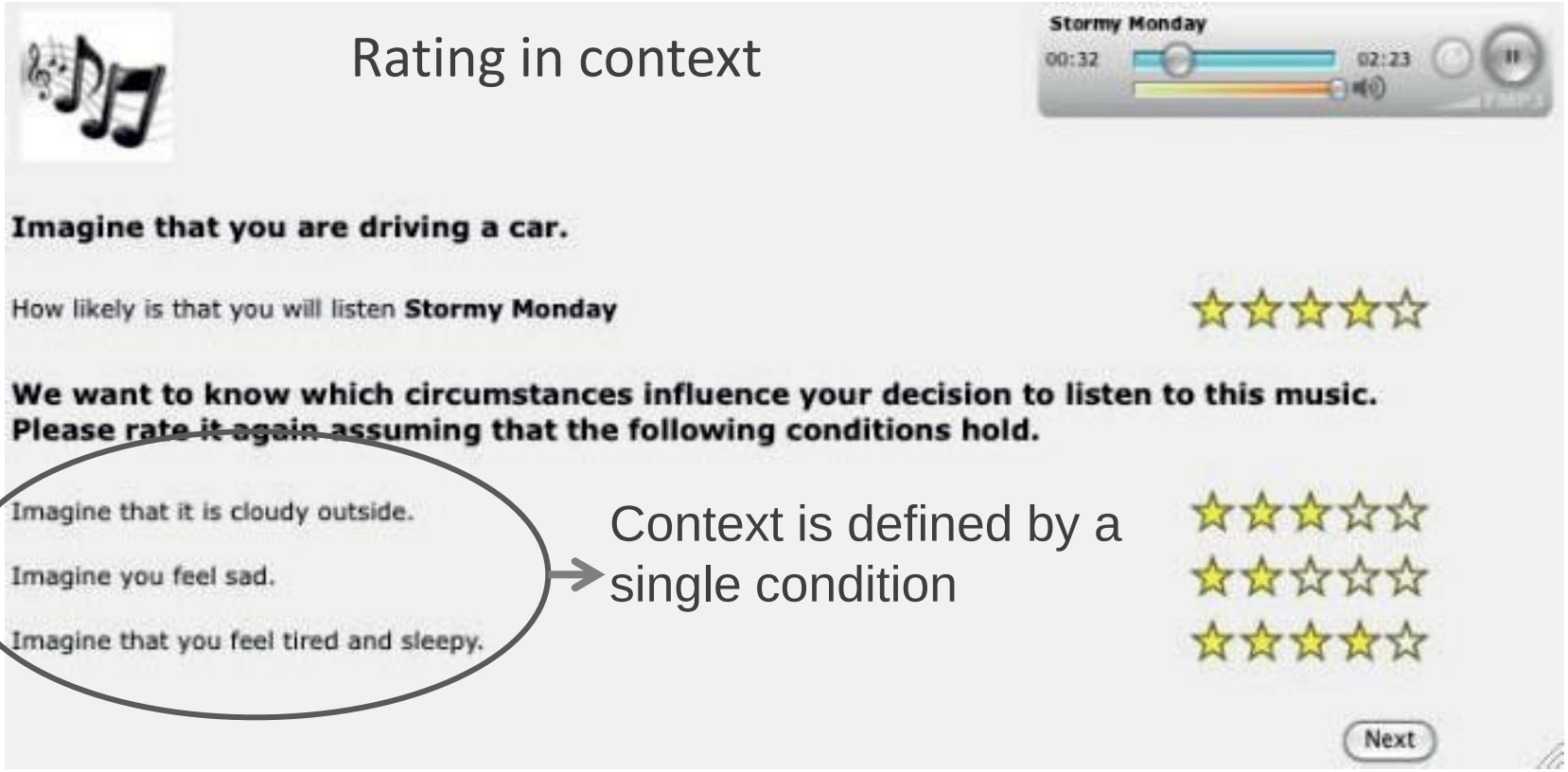
Item bias

Outline

- Semantics acquisition
- Semantics exploitation
- Experiments
  - Data sets
  - Considered prediction models
  - Experimental results

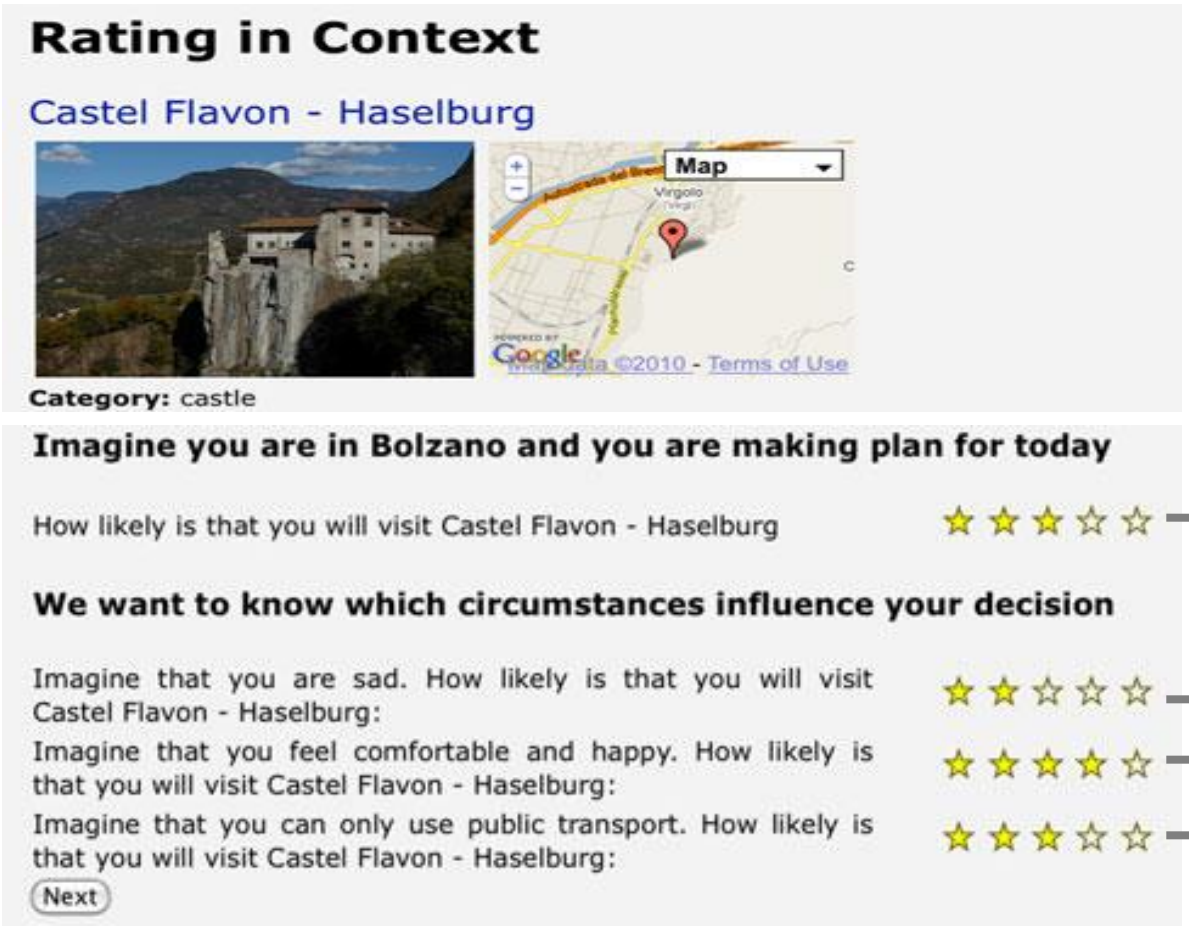
For the experimentation we used two real-world contextually-tagged data sets

One data set is about an In-Car music recommender



For the experimentation we used two real-world contextually-tagged data sets

The other data set is about a tourism recommender



Multiple ratings for the same item and user but in different conditions:

→ without context

→ if "sad"

→ if "happy"

→ by "public transport"

We compared the performance of our approach with two other approaches

The 2 proposed variants of the semantic pre-filtering:

Sem-Pref-gt – using a **global threshold**

Sem-Pref-dt – using a **different threshold** per context

A traditional approach based on exact pre-filtering:

Exact-Pref: **exact pre-filtering** using the same MF model

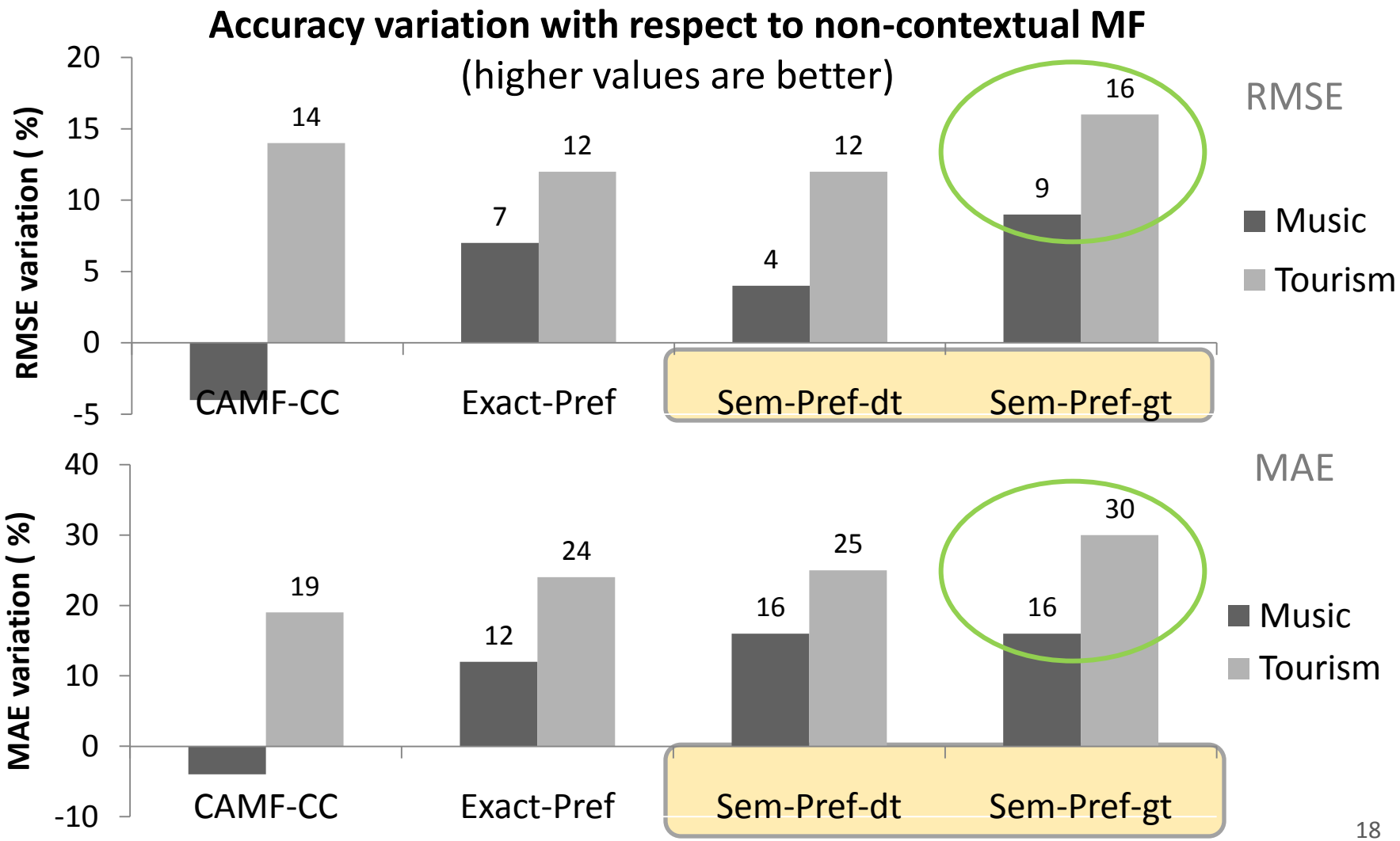
A state-of-the-art context-aware MF approach:

CAMF-CC: **multi-dimensional MF** model

$$\hat{r}_{uic} = \mu + b_u + b_i + b_{type(i),c} + q_i^T p_u$$

Bias of condition on the item's type

The semantic pre-filtering using a global threshold (Sem-Pref-gt) is the most accurate



Future work

Research in progress…

- New variants of the semantic pre-filtering approach
- New methods for measuring the implicit semantics of contextual conditions

Planned in the near future:

- Novel semantic multidimensional approaches
- Extend the evaluation by assessing the performance of the proposed approaches for ranking recommendation

Semantically-Enhanced Pre-Filtering for Context-Aware Recommender Systems

Any comments or questions?

Victor Codina

vcodina@lsi.upc.edu