

# **Combining the opinion profile modeling with complex context filtering for Contextual Suggestion**

**Peilin Yang and Hui Fang**

University of Delaware

# The Roadmap

Data Crawling



Profile Modeling



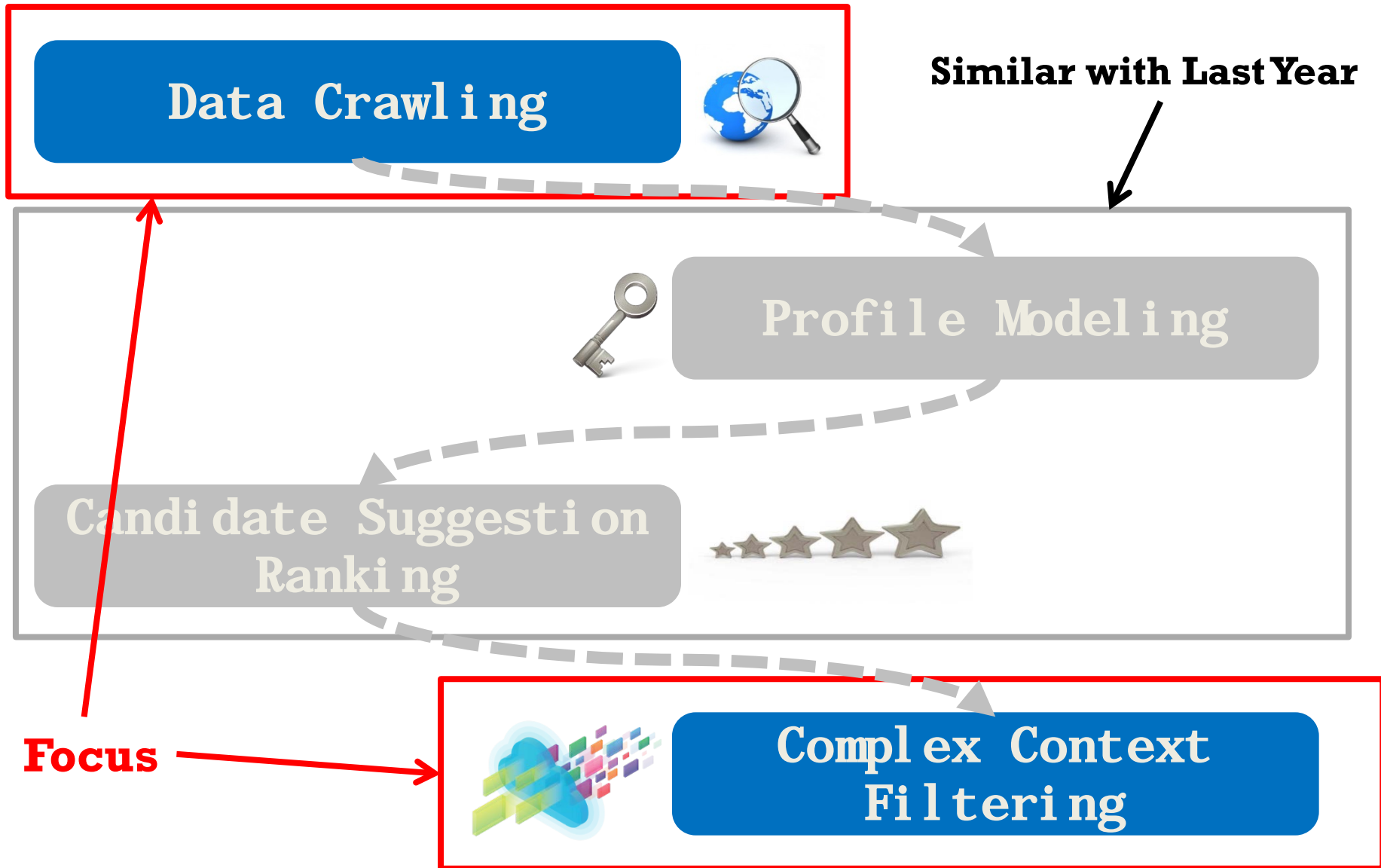
Candidate Suggestion  
Ranking



Complex Context  
Filtering



# The Roadmap



# The Roadmap

Data Crawling



Profile Modeling



Candidate Suggestion  
Ranking

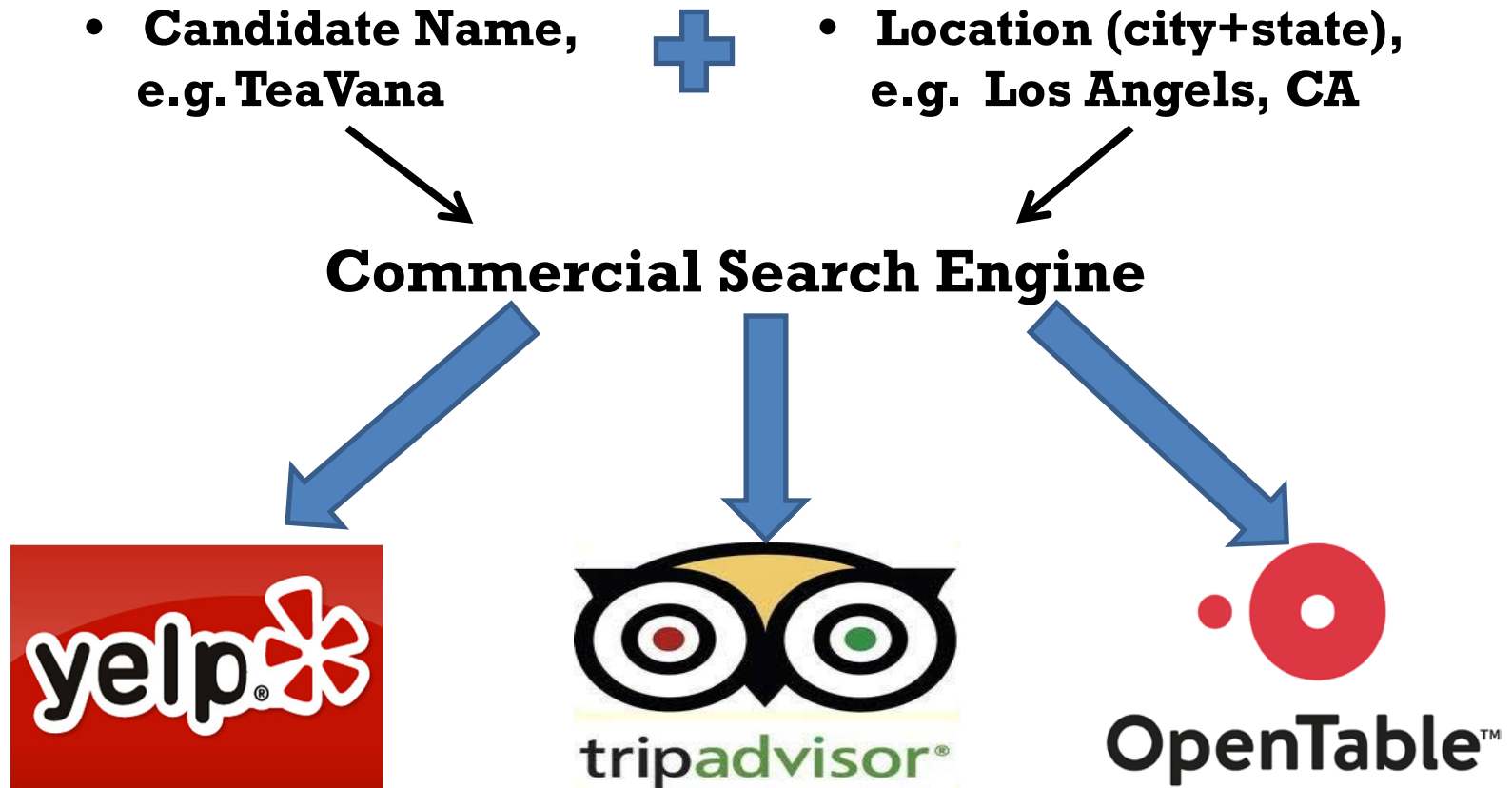


Complex Context  
Filtering




# Data Crawling

- **Best Effort (may have erroneous candidates)**

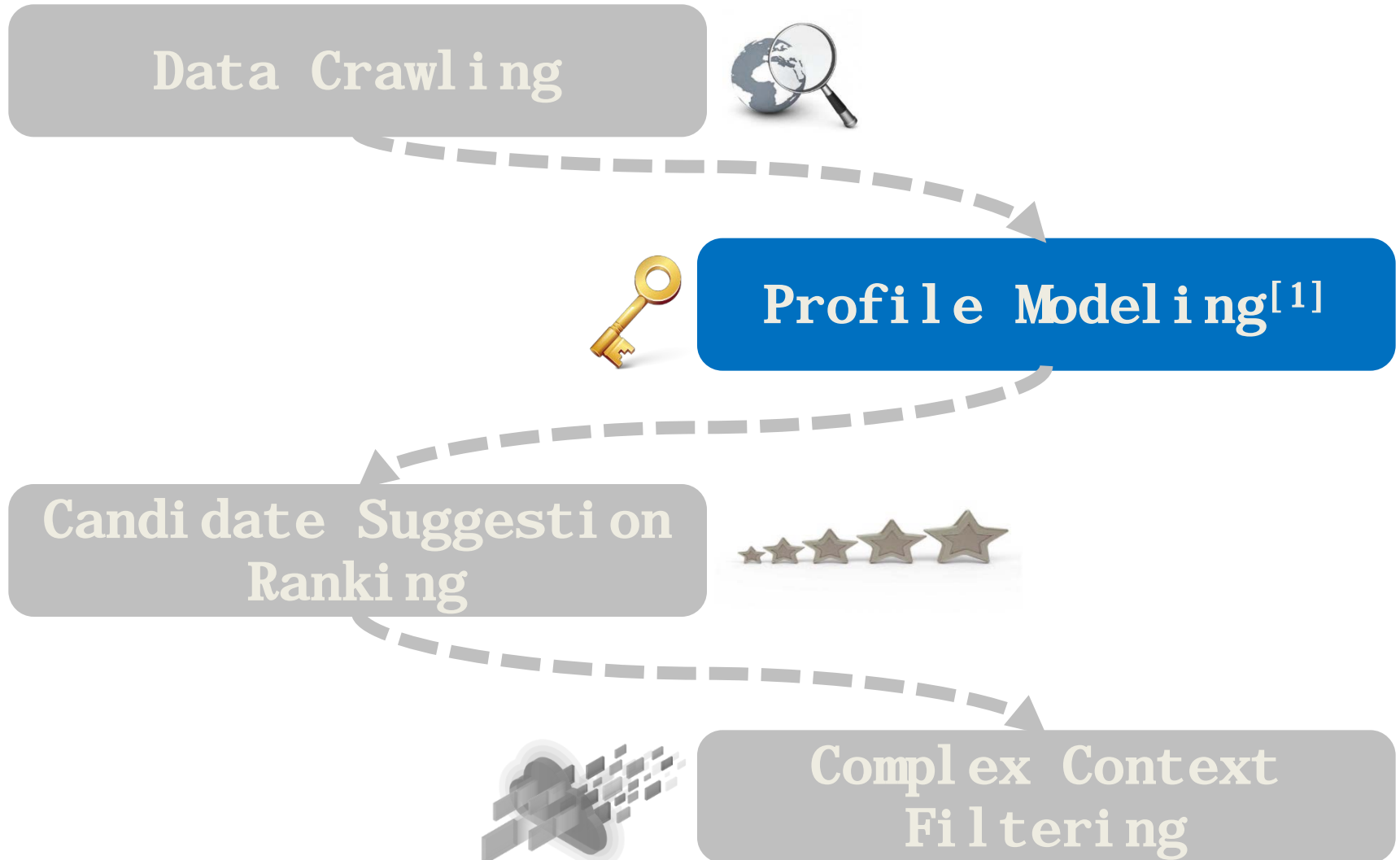


# Data Crawling

- **Total : 1,235,844**
- **Crawled: 161,907**

- 
- **Name**
  - **overall Rating**
  - **total\_review\_number**
  - **Categories**
  - **Business hours (if applicable)**
  - **Price Range (if applicable)**
  - **Reviews**
    - **rating**
    - **comment**

# The Roadmap



[1] P. Yang and H. Fang. Opinion-based user profile modeling for contextual suggestions. In *Proceedings of the 2013 Conference on the Theory of Information Retrieval, ICTIR'13*, pages 18:80–18:83, New York, NY, USA, 2013. ACM.

# **Inspiration**

**No single factor can capture the real reason...**

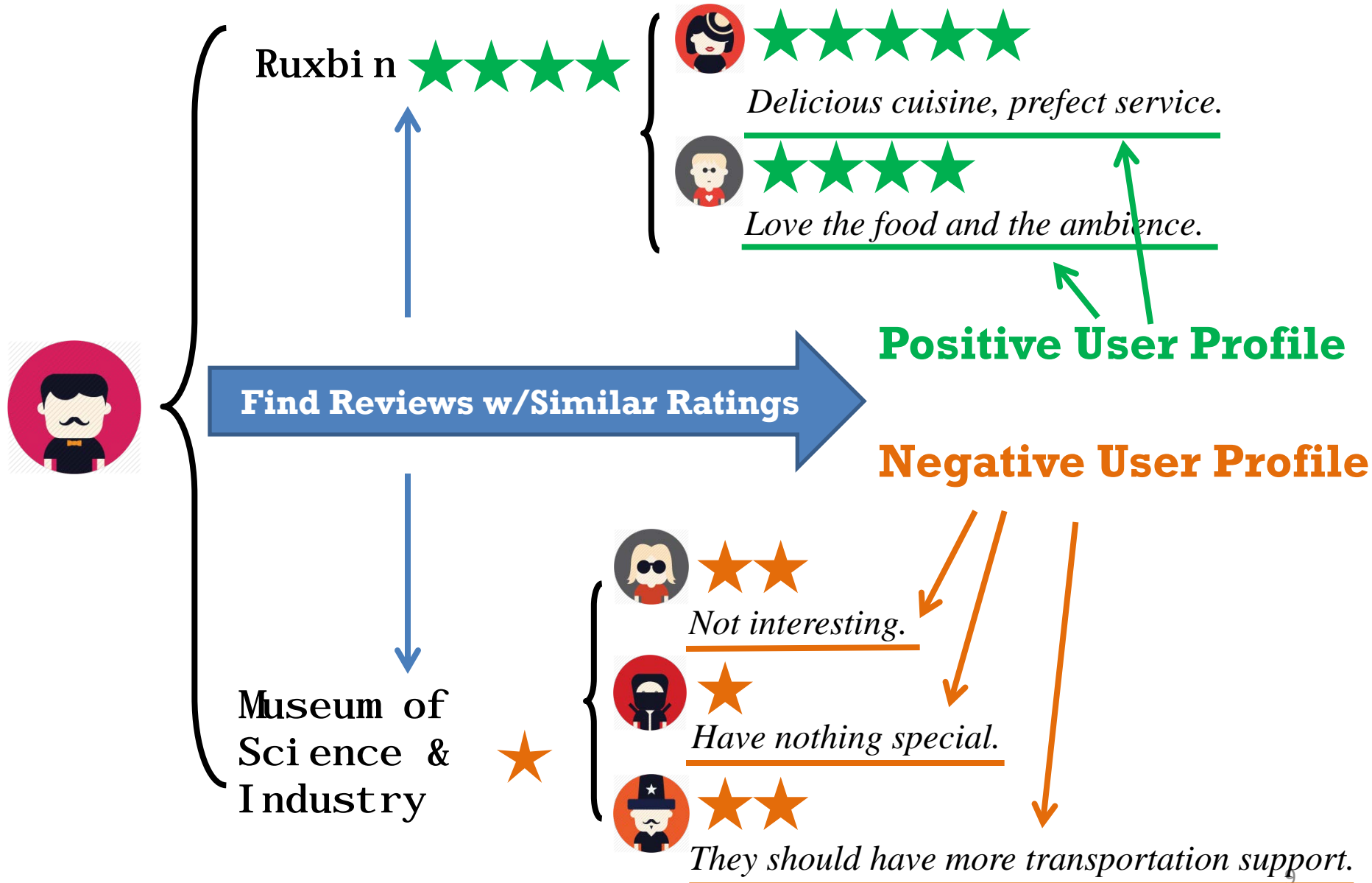
**Enrich!**

**Explore the opinions**

**Leverage the wisdom of crowds**



# Profile Modeling - User



# Profile Modeling – Candidate Suggestions



Absolutely amazing food and the most friendly and inviting atmosphere that I've ever experienced.



The food was very reasonably priced. No soggy dough and just the right amount of sweetness.

**Positive Candidate Profile**

**Negative Candidate Profile**



Rice ball for a dollar. I guess so? But nothing special.



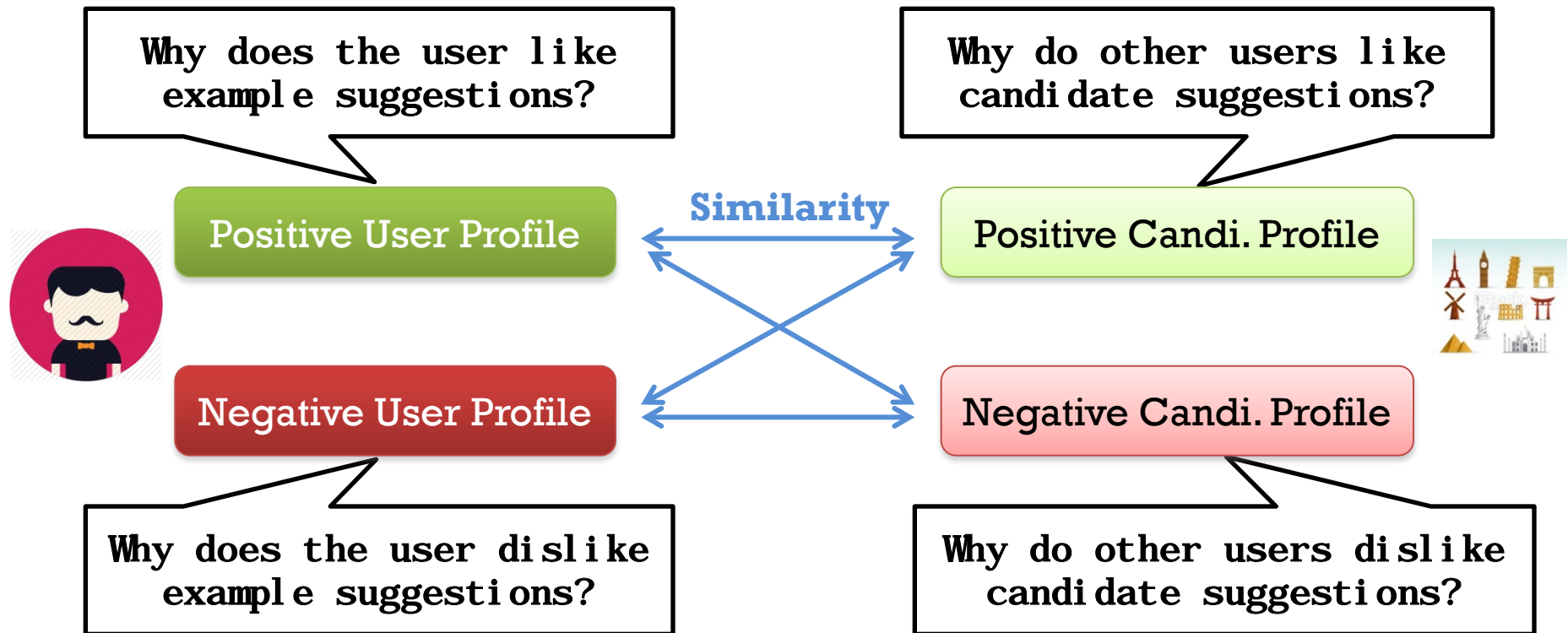
I've done some pretty disgusting things in my life... trust me, but I've always been considerate of other people and their boundaries.



# The Roadmap



# Candidate Suggestion Ranking



# Ranking Details

---

## Representations of Reviews when building user profile

- Full Reviews (**FR**): use full text in the review.
  - Nouns in Reviews (**NR**): nouns in the review.
- 

## Similarity Measurement: Linear Interpolation<sup>[1]</sup>

$$S(U, CS) = \alpha \times SIM(\mathcal{U}_{pos}, \mathcal{CS}_{pos}) - \beta \times SIM(\mathcal{U}_{pos}, \mathcal{CS}_{neg}) \\ - \gamma \times SIM(\mathcal{U}_{neg}, \mathcal{CS}_{pos}) + \eta \times SIM(\mathcal{U}_{neg}, \mathcal{CS}_{neg})$$

[1] Peilin Yang and Hui Fang. Opinions Matter: A General Approach to User Profile Modeling for Contextual Suggestion. *To Appear in Information Retrieval Journal*.

# The Roadmap

Data Crawling



Profile Modeling



Candidate Suggestion  
Ranking



Complex Context  
Filtering



# Complex Context Filter

- Basic Operations:
    - *Boost*: boost the score of the selected candidates
    - *Avoid*: remove the selected candidates
    - *Mix*: reorder the ranking list in round-robin way in terms of category
- \*\*\* Mix applies after Boost and Avoid. It actually reorders the ranking list. So we apply Trip Duration (where Mix occurs) filter last.



Context	Boost	Avoid	Mix
Trip Type			
- Business	Pricy hotels and restaurants	-	-
- Holiday	-	-	-
- Other	-	-	-

<b>Context</b>	<b>Boost</b>	<b>Avoid</b>	<b>Mix</b>
Trip Duration			
- Travelling Alone	-	-	-
- Travelling with a group of friends	-	-	-
- Travelling with family	amusement park	-	-
- Travelling with an other group	“good for groups” property (Yelp and OpenTable have such information)	-	-
Trip Season			
- Spring	-	-	-
- Summer	-	-	-
- Fall	-	-	-
- Winter	-	park, amusement park and zoo	-



<b>Context</b>	<b>Boost</b>	<b>Avoid</b>	<b>Mix</b>
Trip Duration			
- Night Out	bar, pub, theaters, music venues	venues that are closed at night, e.g. brunch restaurants.	-
- Day Trip	-	hotel, bar, pub, theaters and venues that are closed during daytime	-
- Weekend Trip	-	-	hotels, restaurants and landmarks for every 5 suggestions
- Longer	-	-	hotels and other types of venues for every 5 suggestions

# Results

<b>Runs</b>	<b>Review Representation</b>	<b>Context Filter</b>	<b>P@5</b>
FR_CF *	Full Review	Yes	0.5583
FR_NO_CF	Full Review	No	0.5972 
NR_CF *	Noun Review	Yes	0.5507
NR_NO_CF	Noun Review	No	0.6038 

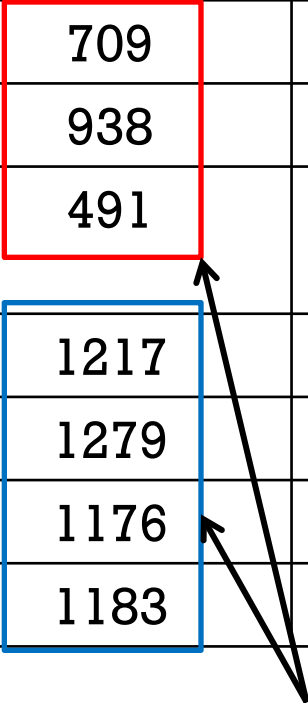
\* submitted runs

## Main Findings:

- All the runs generally perform good
- Context Filter does not work well as expected

# Analysis

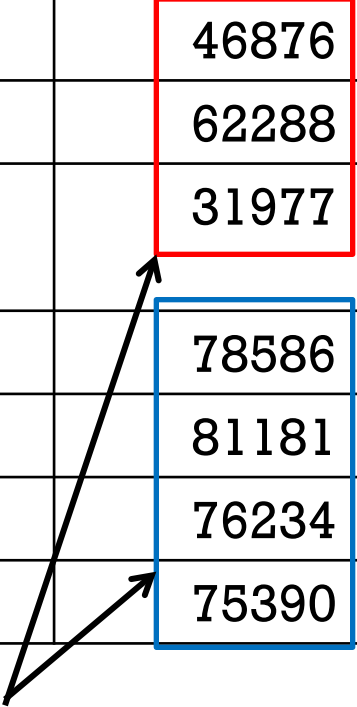
<b>Runs</b>	<b>reviews cnt. (mean)</b>	<b>reviews cnt. (std)</b>	<b>pos. terms cnt. (mean)</b>
All Candidates	709	1650	46876
Relevant Candidates	938	1943	62288
Non-Relevant Candidates	491	1275	31977
FR_CF (Top 5)	1217	1999	78586
FR_NO_CF (Top 5)	1279	2080	81181
NR_CF (Top 5)	1176	2017	76234
NR_NO_CF (Top 5)	1183	2019	75390



Number of reviews in top ranked results of our method is much larger than that of candidates (both relevant ones and non-relevant ones)

# Analysis

<b>Runs</b>	<b>reviews cnt. (mean)</b>	<b>reviews cnt. (std)</b>	<b>pos. terms cnt. (mean)</b>
All Candidates	709	1650	46876
Relevant Candidates	938	1943	62288
Non-Relevant Candidates	491	1275	31977
FR_CF (Top 5)	1217	1999	78586
FR_NO_CF (Top 5)	1279	2080	81181
NR_CF (Top 5)	1176	2017	76234
NR_NO_CF (Top 5)	1183	2019	75390



In terms of number of positive terms in the reviews, the difference is even larger.

# Analysis

<b>Runs</b>	<b>reviews cnt. (mean)</b>	<b>reviews cnt. (std)</b>	<b>pos. terms cnt. (mean)</b>
All Candidates	709	1650	46876
Relevant Candidates	938	1943	62288
Non-Relevant Candidates	491	1275	31977
FR_CF (Top 5)	1217	1999	78586
FR_NO_CF (Top 5)	1279	2080	81181
NR_CF (Top 5)	1176	2017	76234
NR_NO_CF (Top 5)	1183	2019	75390

**Our method favors venues with more reviews and positive review terms.**

**Thank you!**

**Questions?**