Rich-Context: An Unsupervised Context-Driven Recommender System Based On User Reviews

University College Cork

Francisco J. Peña

Supervisor: Dr. Derek Bridge
Problem
Context in Recommender Systems

Context Aware Recommender Systems

- Most of them predefine context
- Small number of features
- Small number of values

I’m travelling for:  ○ Work  ○ Leisure
Companion:  ○ Solo  ○ Couple  ○ Family

Open-ended context is very wide

- Context is richer, open-ended
- Birthday, anniversary, parking, accessibility, eat-in vs take away, pet friendly, ...
Goals

**Recommendation model**

- Treat context as open-ended
- Unsupervised (not predefined keywords)
- Good performance on sparse datasets

**Evaluation methodology**

- Datasets
  - Big (+10 000 records)
  - Sparse
  - Multiple from different domains
  - Publicly available
  - Real-world
- Third-party evaluation tool
Contributions

Recommendations without predefining context

New way of representing context.
Contributions

Recommendations without predefining context

- New way of representing context.

Better prediction performance

- New methodology for offline ranking evaluation.
- Better performance for brand-new users.
Contributions

Recommendations without predefining context
- New way of representing context.

Better prediction performance
- New methodology for offline ranking evaluation.
- Better performance for brand-new users.

Extract context from reviews in an unsupervised way
- Methodology for selecting the best topic modeling algorithm.
- New metrics to measure context-richness of topic models.
- Improved methodology to classify reviews.
Assumptions

Specific Review

• “During the summer, we like to take a mini staycation. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing.”

Generic Review

• “Nice hotel, all the amenities you need, great complex of pools.”
Assumptions

Specific Review

• “During the summer, we like to take a mini staycation. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing.”

Generic Review

• “Nice hotel, all the amenities you need, great complex of pools.”

Specific reviews contain more contextual information than generic ones.
Rich Context (RC)

Rich Context

- RCClassifier
- RCMiner
- RCRecommender
Reviews Classification

• 300 tagged reviews
• Random Forest Classifier
• Features
  • LogWords: log of number of words in the review + 1
  • Vsum: log of number of verbs in the review + 1
  • VBDSum: log of number of verbs in the past tense in the review + 1
  • ProRatio: ratio of log of number of personal pronouns + 1 to LogWords
Context Extraction

- Reviews
- Classifier
- Specific Reviews
- Topic Modeling Algorithm
- Topic Model
- Filter
- Contextual Topics
- Generic Reviews
Ensemble Topic Modeling

- More stable topic models.
- Context richer topics.

Source: Stability of Topic Modeling via Matrix Factorization. Belford et al

Advantages
Topic Model Validation

What topic modeling is more stable?

Which reviews data produces context richer topic models?
Topic Model Validation (Stability)

Term Difference

Topic Stability
Topic Model Validation (Context-Richness)

Part of speech types

Specific vs generic reviews

[Charts showing context richness scores for different datasets and part-of-speech categories]
Context Extraction

- Apply the topic model to both specific and generic reviews.
- Count the number of times topics appear in specific and generic reviews.
- The ones that appear more frequently in specific reviews are labeled as contextual topics.

Topic Distribution Among Reviews

Contextual Topics
Recommendations

User -> Text -> Context

Reviews & Ratings -> Context & Ratings

Context & Ratings -> Factorization Machines

Factorization Machines -> Recommendation
Evaluation - Dataset Description

Yelp Hotels
- 3,809 reviews
- 3,205 users
- 98 items
- 98.79% sparsity

Yelp Restaurants
- 147,864 reviews
- 35,021 users
- 2,550 items
- 99.83% sparsity

TripAdvisor Hotels
- 726,426 reviews
- 526,717 users
- 3,299 items
- 99.96% sparsity
Evaluation - Ranking Prediction (Recall@10)

Compared to best SOTA

- **Improvement (all users)**
  - Yelp Hotels (+2.32%)
  - Yelp Restaurants (+55.07%)
  - TripAdvisor Hotels (+30.15%)

- **Improvement (new users)**
  - Yelp Hotels (+3.67%)
  - Yelp Restaurants (+47.00%)
  - TripAdvisor Hotels (+20.50%)
Evaluation - Rating Prediction (RMSE)

Compared to best SOTA

 Improvement (all users)
- Yelp Hotels (+4.45%)
- Yelp Restaurants (+3.29%)
- TripAdvisor Hotels (+8.99%)

 Improvement (new users)
- Yelp Hotels (+5.44%)
- Yelp Restaurants (+4.54%)
- TripAdvisor Hotels (+9.96%)
Final Thoughts

Conclusions

- We present a context-driven recommender system that does not pre-defined contextual words.
- We improve recommendations by using the mined contextual information as side-information in factorization machines.
- The proposed model does not need expertise about contextual information.

Future Work

- Improve the topic model quality metrics in order to evaluate topic models without having to run the recommender (like a classifier).
- Use topic models to produce explanations of recommendations.
- Extrapolate the same model to other scenarios where documents are available and recommendations are needed, for instance using medical records, law, etc.
Thanks!

Acknowledgments
• Derek Bridge
• Diego Carraro
• Mesut Kaya
• Insight UCC
• SFI

francisco.pena@insight-centre.org
References


References


References


Acknowledgements

Dr. Derek Bridge
Diego Carraro
Mesut Kaya
Insight UCC
During the summer, we like to take a mini staycation. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing - it was great at the pool, Wrights and also at Frank and Alberts. The only reason I am not giving it a full 5 stars is the 'upgraded' room was just a nice basic room. Though it was certainly nice, it wasnt what I expected for being the Biltmore. However, everything else certainly lived up to that expectation.

Summer (0.04)
Weekend (0.02)
June (0.01)
...
Holiday (0.05)
Romantic (0.03)
Staycation (0.01)
...
Room (0.05)
Pool (0.04)
Sauna (0.01)
...
Free (0.03)
Cheap (0.02)
Expensive (0.01)
...

Topic Modeling
• Each document is a random mixture of corpus-wide topics
• Each topic is composed of words that co-occur along documents
Number Of Topics vs Performance

Yelp Restaurants

The number of topics matters!
Topic Model Validation (Stability)

Average Descriptor Set Difference

Yelp Hotels

Yelp Restaurants

TripAdvisor Hotels

Number of topics

Number of topics

Number of topics

ACSD

NMF
LDA
Ensemble

NMF
LDA
Ensemble

NMF
LDA
Ensemble
Topic Model Validation (Stability)

Average Term Stability

Yelp Hotels

Yelp Restaurants

TripAdvisor Hotels

ATS

Number of topics

5 10 20 40

NMF  LDA  Ensemble

Number of topics

5 10 20 40

NMF  LDA  Ensemble

Number of topics

5 10 20 40

NMF  LDA  Ensemble
# Topic Model Validation (Context-Richness)

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>family</td>
<td>sushi</td>
<td>wife</td>
<td>service</td>
</tr>
<tr>
<td>0.194</td>
<td>0.271</td>
<td>0.358</td>
<td>0.435</td>
</tr>
<tr>
<td>sunday</td>
<td>bar</td>
<td>date</td>
<td>atmosphere</td>
</tr>
<tr>
<td>0.086</td>
<td>0.055</td>
<td>0.026</td>
<td>0.023</td>
</tr>
<tr>
<td>town</td>
<td>town</td>
<td>birthday</td>
<td>customer</td>
</tr>
<tr>
<td>0.069</td>
<td>0.021</td>
<td>0.020</td>
<td>0.018</td>
</tr>
<tr>
<td>brunch</td>
<td>spot</td>
<td>anniversary</td>
<td>table</td>
</tr>
<tr>
<td>0.029</td>
<td>0.020</td>
<td>0.019</td>
<td>0.012</td>
</tr>
<tr>
<td>weekend</td>
<td>saturday</td>
<td>weekend</td>
<td>drink</td>
</tr>
<tr>
<td>0.021</td>
<td>0.014</td>
<td>0.015</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Score: 0.33  Score: 0.014  Score: 0.438  Score: 0.0

**Topic Model Score:** 0.1955

\[
ts(t) = \sum_w (p_{wt} \times v_{wt})
\]

The topic model score is the average of the topic scores.
## Evaluation - Generated Topic Models (Yelp Restaurant)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Ratio</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Word 3</th>
<th>Word 4</th>
<th>Word 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>2.03</td>
<td>night</td>
<td>dinner</td>
<td>friend</td>
<td>saturday</td>
<td>friday</td>
</tr>
<tr>
<td>Topic 2</td>
<td>1.61</td>
<td>lunch</td>
<td>today</td>
<td>day</td>
<td>friend</td>
<td>yesterday</td>
</tr>
<tr>
<td>Topic 3</td>
<td>1.34</td>
<td>time</td>
<td>couple</td>
<td>week</td>
<td>minute</td>
<td>hour</td>
</tr>
<tr>
<td>Topic 4</td>
<td>1.1</td>
<td>breakfast</td>
<td>morning</td>
<td>sunday</td>
<td>club</td>
<td>day</td>
</tr>
<tr>
<td>Topic 5</td>
<td>1.07</td>
<td>review</td>
<td>yelp</td>
<td>experience</td>
<td>star</td>
<td>read</td>
</tr>
<tr>
<td>Topic 6</td>
<td>0.99</td>
<td>scottsdale</td>
<td>location</td>
<td>town</td>
<td>experience</td>
<td>tempe</td>
</tr>
<tr>
<td>Topic 7</td>
<td>0.93</td>
<td>restaurant</td>
<td>phoenix</td>
<td>area</td>
<td>mexican</td>
<td>week</td>
</tr>
<tr>
<td>Topic 8</td>
<td>0.85</td>
<td>chicken</td>
<td>pizza</td>
<td>burger</td>
<td>sandwich</td>
<td>cheese</td>
</tr>
<tr>
<td>Topic 9</td>
<td>0.75</td>
<td>place</td>
<td>area</td>
<td>bar</td>
<td>love</td>
<td>home</td>
</tr>
<tr>
<td>Topic 10</td>
<td>0.72</td>
<td>food</td>
<td>service</td>
<td>mexican</td>
<td>atmosphere</td>
<td>price</td>
</tr>
</tbody>
</table>
## Evaluation - Ranking Prediction

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Recall@10</th>
<th>Yelp Hotels</th>
<th>Restaurants</th>
<th>TripAdvisor Hotels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rich-Context</td>
<td>0.381</td>
<td>0.086</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>Factorization Machines</td>
<td>0.372</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPMF</td>
<td>0.150</td>
<td>0.014</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>Biased MF</td>
<td>0.068</td>
<td>0.016</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>CAMF-C</td>
<td>0.085</td>
<td>0.016</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>CAMF-CI</td>
<td>0.235</td>
<td>0.016</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>CAMF-CU</td>
<td>0.239</td>
<td>0.017</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>CAMF-CUCI</td>
<td>0.082</td>
<td>0.015</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>DCR</td>
<td>0.125</td>
<td>Out of memory</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td>DCW</td>
<td>0.125</td>
<td>Out of memory</td>
<td>Out of memory</td>
<td></td>
</tr>
<tr>
<td><strong>Improvement (all users)</strong></td>
<td><strong>2.32%</strong></td>
<td><strong>55.07%</strong></td>
<td><strong>30.15%</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Improvement (new users)</strong></td>
<td><strong>3.67%</strong></td>
<td><strong>47.00%</strong></td>
<td><strong>20.50%</strong></td>
<td></td>
</tr>
</tbody>
</table>
Evaluation - Rating Prediction

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>RMSE</th>
<th>Yelp</th>
<th>TripAdvisor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hotels</td>
<td>Restaurants</td>
</tr>
<tr>
<td>Rich-Context</td>
<td>1.050</td>
<td>1.056</td>
<td>0.962</td>
</tr>
<tr>
<td>Factorization Machines</td>
<td>1.106</td>
<td>1.092</td>
<td>1.057</td>
</tr>
<tr>
<td>BPMF</td>
<td>1.495</td>
<td>1.405</td>
<td>1.435</td>
</tr>
<tr>
<td>Biased MF</td>
<td>1.109</td>
<td>1.183</td>
<td>1.104</td>
</tr>
<tr>
<td>CAMF-C</td>
<td>1.100</td>
<td>1.184</td>
<td>1.084</td>
</tr>
<tr>
<td>CAMF-CI</td>
<td>1.160</td>
<td>1.213</td>
<td>1.104</td>
</tr>
<tr>
<td>CAMF-CU</td>
<td>1.343</td>
<td>1.228</td>
<td>1.420</td>
</tr>
<tr>
<td>CAMF-CUCI</td>
<td>1.123</td>
<td>1.200</td>
<td>1.104</td>
</tr>
<tr>
<td>DCR</td>
<td>1.224</td>
<td>Out of memory</td>
<td>Out of memory</td>
</tr>
<tr>
<td>DCW</td>
<td>1.226</td>
<td>1.321</td>
<td>Out of memory</td>
</tr>
<tr>
<td>Improvement (all users)</td>
<td>4.45%</td>
<td>3.29%</td>
<td>8.99%</td>
</tr>
<tr>
<td>Improvement (new users)</td>
<td>5.43%</td>
<td>4.54%</td>
<td>9.96%</td>
</tr>
</tbody>
</table>
Topic Model Validation (Stability)

Average Descriptor Set Difference

Average Term Stability
Topic Model Validation (Context-Richness)

Part of speech types

Specific vs generic reviews
Evaluation - Ranking Prediction

Evaluation metric Recall@10

9 SOTA

Rich-Context vs best SOTA

6 CARS

3 non-CARS

Yelp Hotels (+2.32%)

Yelp Restaurants (+55.07%)

TripAdvisor Hotels (+30.15%)
Contributions

Recommendations without predefining context
- New way of representing context.

Better prediction performance
- New methodology for offline ranking evaluation.
- Better performance for brand-new users.

Extract context from reviews in an unsupervised way
- Methodology for selecting the best topic modeling algorithm.
- New metrics to measure context-richness of topic models.
- Improved methodology to classify reviews.