Motivation

- Discover open-ended contextual information from user reviews. Context doesn’t have to be predefined.
- Incorporate the open-ended contextual information together with ratings into a context sensitive recommender system.
- Make recommendations that better satisfy user goals.

Proposed Solution

- Our system is called Rich Context (RC) and its has two main components: RCMiner and RCRRecommender.

Assumptions

- Specific reviews describe experiences.
- Generic reviews give general reviews of products.
- Specific reviews contain more contextual information than generic ones.

Classifying Reviews

- We separate specific and generic reviews using a Logistic Regression classifier.
- Feature subset selection showed that the number of words and the number of verbs in past tense in a review are the most relevant features.

Topic modeling

- We use Latent Dirichlet Allocation (LDA) to find the topics that compose the specific reviews. Topics are composed solely of lemmatized nouns.

Example

**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- 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 wasn’t what I expected for being the Biltmore. However, everything else certainly lived up to that expectation.”

**Generic review**

"Nice hotel, all the amenities you need, great complex of pools. Just make sure your room is as far from the Vista Lounge as possible, otherwise you’ll be bombarded with crappy live music, fully audible from the lounge to all the surrounding rooms above it, for four hours a day. Horrible. The Internet access is not free, which is lame. The room service is good but overpriced by at least 40 percent. So it goes with resorts. Otherwise a very nice hotel.”

Experiments

- Yelp Data Challenge dataset of hotels (5034 reviews) and restaurants (159430 reviews).
- 150 topics, 5-fold cross validation.
- Compared against a matrix factorization baseline recommender.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Top-N Recall</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Context topics</td>
<td>All topics</td>
</tr>
<tr>
<td>Hotels</td>
<td>0.386</td>
<td>0.385</td>
</tr>
<tr>
<td>Restaurants</td>
<td>0.074</td>
<td>0.075</td>
</tr>
</tbody>
</table>

Future Work

- Improve RCMiner by designing new features and using different classifiers.
- Design a metric to measure the quality of contextual topics.
- Classify sentences within reviews; build the topic models on the specific sentences only.