POIReviewQA: A Semantically Enriched POI Retrieval and Question Answering Dataset

GIR’18, Nov. 2018

Gengchen Mai\textsuperscript{1}  Krzysztof Janowicz\textsuperscript{1}
Cheng He\textsuperscript{2}  Sumang Liu\textsuperscript{2}  Ni Lao\textsuperscript{2}

\textsuperscript{1} STKO Lab, University of California Santa Barbara
\textsuperscript{2} SayMosaic Inc., Palo Alto, CA, USA
Location-Based Services & POI Searching

Location-based services (LBS) and the underlying Point of Interest (POI) datasets play a increasingly important role in our daily interaction with mobile devices.
Location-Based Services & POI Searching

Platforms such as Yelp, Foursquare, Google Map allow users to search nearby POIs based on their names, place types, or tags, which requires manual data annotation.

Place Name Search

Place Type Search

Tag Search
Problems in Current POI Searching

- **Predefined Tags**: The tags (take reservations, delivery, take out) are also **predefined** and **manually added** which may not satisfy users’ needs and require a lot of manually efforts.
Problems in Current POI Searching

- **Predefined Schema**: The place type schema is **predefined** and unbalanced.
**Introduction**

- Do we utilize all info in Yelp for search?
- Yelp contains billions of Points of Interest (POIs) and their attributes:
  - name
  - location
  - type
  - open times
  - rating
  - reviews
  - ...
- Can we use other **crowdsourcing information** e.g. reviews to **enrich POI** and facilitate POI searching process?
ANSWERING QUESTIONS BY POI REVIEWS

Saint Archer Brewery

$ - Breweries, Beer, Wine & Spirits

Write a Review

12/3/2017
Awesome brewery. There are so many breweries in San Diego but this one definitely stands out. It's fairly inexpensive and it's dog friendly. Both indoor and outdoor seating areas, you know since it's sunny in San Diego year round.

As far as the beer goes, what's your speed? From stouts to IPAs, they have you covered.

Is this place dog friendly?

They have an outdoor patio and I've seen customers with dogs sitting outside.

11 months ago • 1 person found this helpful

View 1 more answer
Answering Questions By POI Reviews

Fufu's Mid East Grill And Hookah
- Claimed
- 324 reviews
- Mediterranean, Hookah Bars, Middle Eastern

Ask the Community

Hello, do you serve **HALAL** meat? If so, do you have the certification that shows proof? Thank you!

Sarah B. of Fufu's Mid East Grill And Hookah

Business Owner

Yes everything we serve is 100% **halal**. And we do have the certificate of proof at the restaurant.

Thanks

2 months ago • 3 people found this helpful

View question details

The food is **halal** too, so for the Muslim community this is THE place to go to!

Only negative is that there is no WiFi, so I just about checked checked in on Yelp here.

Mark B. and 1 other voted for this review
Answering Questions By POI Reviews

Farm & Craft • Claimed

541 reviews • Details

$$ • American (New), Breakfast & Brunch, Cocktail Bars

Ask the Community

Dog friendly patio?

Yes!! Saw lots of cute dogs sitting in the patio

View 5 more answers

Is your salmon poke bowl gluten free and dairy free?

Kevin B. of Farm & Craft

Hey Bailee, the dish is dairy free, the dish is not gluten free because the salmon marinade has a slight amount of gluten in it from the soy sauce. However, upon request we can use gluten free soy sauce for the salmon but it may take a bit longer to... Read more
POI Review IR & QA

- **A Challenging Task:** semantically enriching POIs from unstructured data, e.g. POI reviews, in order to support open-domain search and question answering (QA).

- **One Approach:** Given a question, e.g. find night clubs near me which are 18+, build a model to:
  - Generate the answer to the question given the current POI review
  - Rank POI reviews based on whether they can answer it and the distance from POIs to the user location.

- **Problem:** No existing benchmark dataset to train and evaluate such kind of models
  - **No GIR Corpus:** Only a few existing GIR benchmark datasets (e.g., GeoCLEF); lack in rich annotations.
  - **No QA Corpus:** Many QA datasets for reading comprehension such as SQuAD; no spatial/platial component; very different query sets.
Our Contribution

- We introduce an evaluation benchmark dataset **POIReviewQA** which consists of 20k questions for 1022 Yelp business types to support open-domain POI search and QA.
- We build two baseline models: a Lucene-based baseline model and a sentence embedding-based model which indicate that the dataset presents a challenging problem for future research by the GIR community.
We created POIReviewQA based on the Yelp Challenge 11 (YC11) dataset and the “Ask the Community” section of POI pages.

Yelp platform is dominated by popular business types such as restaurants.

We need to produce a balanced query set for all business types.
Query Set Generation

- Count the **frequencies of POI name suffixes** (single words) in YC11
- For every suffix (freq \( \geq 10 \)), we create a **quoted search query**, and collect QA page URLs from **Google search engine**:
  - site:yelp.com/questions/ "Restaurant Ask the Community"
  - 1,701 quoted search queries
  - Up to 100 search results for each query
- Collect questions and answers from the **community QA pages**
Query Set Generation

- Since the suffix often indicates the place type of a POI, the collected query set has **a wide coverage of different place types**.
- **20K questions** were collected from Yelp business question pages for **1022 Yelp business types**.
- Each question is associated with one or multiple POIs with several POI types.

**Question: Age to get in?**

**Context:**
- **Echoplex** [Music Venues, Bars, Dance Clubs]
- **Paper Tiger Bar** [Cocktail Bars, Lounges]
For each question, **10 review candidates** are selected:

- Apply **Elastic Search** to POI reviews based on the question with **constraint to the associated POI types**
- Do **stratified sampling** on the search result
Relevance and Answer Annotation

Any gluten-free food available?

Here shows some search options:

Return 1000 results

1. Humble Pie
   Italian, Pizza, Restaurants
   21.35636
   Very sad that gluten free crust any available topping. Crust needs something more to be pizza. Had to leave to get dinner.

2. The Cheesecake Factory
   American (New), American (Traditional), Food, Desserts, Restaurants
   21.218565
   We always look forward to coming up to the Cheesecake Factory. They offer quite the variety of food. Also most all their pastas are available in a gluten free option, and tastes great. Sad to say most of the time I’m so stuffed to make it to their cheesecake. Which I think is some sort of food code violation. But you can get a piece to go or better yet a whole cake frozen to go. Just make sure when you order gluten free to remind the kitchen to make any meat toppings gluten free also. (Like grilled chicken instead of breaded)

3. Famous 48
   American (New), Restaurants
   20.990713
   Great for gluten free menu! We had the deviled eggs, spinach con queso and BBQ baby back ribs. Great food, service and atmosphere. Gluten free pizza crust and buns available!

4. PPG Paints Arena
   Stadiums & Arenas, Event Planning & Services, Venues & Special Event Spaces, Music Venues, Nightlife, Arts & Entertainment
   20.95517
   5 stars for layout, cleanliness, friendly staff, and overall quality for a venue. 1 star for gluten free options. There are close to zero gluten/allergen free items in this entire arena, save the fruit options at some concession stands. The nachos are gluten free but processed with other wheat products and also fried in soybean oil, the salad has granola mixed in with it, and the sushi is gluten free, but no safe soy sauce is available. They sell Redbridge lager, which is great, but I found myself having beer for dinner because I couldn’t eat anything. They will also

5. Make It Yourself Pizza
   Gluten-Free, Restaurants, Pizza
   20.874434
   Awesome fresh baked pizza. No more reheated slices. They’ll put any amount of toppings on it for one price! Gluten free crust is available too!!

6. Cafe Monte
   restaurants, Food, Breakfast & Brunch, French, Bakeries
   20.846643

7. Cafe Rio
   Mexican, American (Traditional), Salad, Restaurants
   20.7883

8. Giovanni's Pizza & Pasta
   Restaurants, Gluten-Free, Pasta
   20.779844

---

POIReviewQA

Gengchen Mai\textsuperscript{1}, Krzysztof Janowicz\textsuperscript{1}, Cheng He\textsuperscript{2}, Sumang Liu\textsuperscript{2}, Ni Lao\textsuperscript{2}
We developed a crowd-facing Web server and deployed it on Amazon Mechanical Turk to let raters annotate the answer for each question given POI reviews.

Each rater is given a question and 10 review candidates associated with it which are divided into sentences.

For each sentence:
- Whether it answer the current question
- What the corresponding answer is

Each question is annotated by 3 different raters to avoid subject biases.
Relevance and Answer Annotation

Survey Instructions (Click to collapse)

Help Label Yelp Business Reviews
This task aims at choosing which Yelp review sentences answer the given business question.

- For this task, we are giving you a question which was posted on Yelp regarding a specific business (Source Business), i.e., Age to get in?
- Source Businesses and Types are provided to let you have a better understanding of the context of this question, i.e., Echoplex [Music Venues & Bars, Dance Clubs]; 10 Yelp reviews from different businesses are provided. Each review is divided into sentences;
- Please select one or multiple sentences which answer the question asked, i.e., Wide variety of people of all ages 21+;
- Please also summarize the answer(s) you chosen in your own words, i.e., 21 or older;
- Note: if none of the sentences is correct, please ONLY select None of them and leave the summarization text box EMPTY;
- Note: these 10 reviews are not related to each other. Each review comes from a different business, which may provide different answers. All them are considered correct, i.e. Review A: 21 or older or Review B: 20 or older

An Example Question and Review Pair:

Question: Age to get in?
Context:
- Echoplex [Music Venues, Bars, Dance Clubs]
- Paper Tiger Bar [Cocktail Bars, Lounges]

Review 1. [Source Business]
1. Please select the sentence(s) from this Yelp business review which answer(s) the current question:
   - Excellent place to country dance and make new friends.
   - Drinks are reasonably priced, especially domestic beer before 10pm on certain nights.
   - Wide variety of people of all ages 21+.
   - This is my favorite country bar in the east valley.
   - None of them
2. If you find (a) sentence(s) which answers the question, please summarize the answer and enter it below:
   - 21 or older
Relevance and Answer Annotation

Survey Instructions (Click to expand)

Question: Am I able to bring in my own bottle?
Context:
- Waterman Grille [American (New), Seafood]

Review 1. [Source Business]
1. Which of the following answer the above question:
   - [ ] Absolute wonderful place!
   - [ ] My family was in town for a wake/urinal and due traffic did not have time for dinner before the viewing.
   - [ ] We called ahead to make sure they would be able to accommodate our party later in the evening.
   - [ ] We also mentioned that we had Campari lovers and the owner went out and got a bottle for us!
   - [ ] Our server was delightful and the food was outstanding!
   - [ ] I ordered the steak and fries and it was delicious.
   - [ ] Everyone at the table ordered something different and each person loved their meal!
   - [ ] Next time we are in town we will DEFINITELY be back!
   - [ ] None of them
2. Please summerize your selection with a few words:
   Enter your answer here...

Review 2. [Source Business]
1. Which of the following answer the above question:
   - [ ] Great wine and great customer service!
   - [ ] We were in Scottsdale for my friend’s bachelorette party, and we did the “bottling party” at Su Vino.
   - [ ] We had a big table that sat all 8 of us, with cheese, fruit and crackers, and our own personal wine maker bringing us tastes and explaining the different wines.
   - [ ] Then we were able to bottle our own wine, and put our own label on it (we had a picture of all of us girls).
   - [ ] It was a blast!
   - [ ] I would definitely recommend a visit here!
Relevance and Answer Annotation

- **4100** out of 20K questions have been annotated until now.

### Table 1. The Statistic of POIReviewQA

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Annotated question</td>
<td>4,100</td>
</tr>
<tr>
<td>% of questions WITHOUT related reviews</td>
<td>11.4%</td>
</tr>
<tr>
<td>Avg. # of related <em>reviews</em> per question</td>
<td>4.61</td>
</tr>
<tr>
<td>Avg. # of 1 rater agreeing on relevant <em>sentence</em> per question</td>
<td>2.19</td>
</tr>
<tr>
<td>Avg. # of 2 raters agreeing on relevant <em>sentence</em> per question</td>
<td>1.08</td>
</tr>
<tr>
<td>Avg. # of 3 raters agreeing on relevant <em>sentence</em> per question</td>
<td>0.83</td>
</tr>
</tbody>
</table>
For these 4100 questions, **11.4%** of them are **without** any related reviews.

**Table 1. The Statistic of POIReviewQA**

<table>
<thead>
<tr>
<th># of Annotated question</th>
<th>4,100</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of questions WITHOUT related reviews</td>
<td>11.4%</td>
</tr>
<tr>
<td>Avg. # of related <em>reviews</em> per question</td>
<td>4.61</td>
</tr>
<tr>
<td>Avg. # of 1 rater agreeing on relevant <em>sentence</em> per question</td>
<td>2.19</td>
</tr>
<tr>
<td>Avg. # of 2 raters agreeing on relevant <em>sentence</em> per question</td>
<td>1.08</td>
</tr>
<tr>
<td>Avg. # of 3 raters agreeing on relevant <em>sentence</em> per question</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Relevance and Answer Annotation

For these 4100 questions, **4.61** out of 10 reviews are related per question.

Table 1. The Statistic of POIReviewQA

<table>
<thead>
<tr>
<th># of Annotated question</th>
<th>4,100</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of questions WITHOUT related reviews</td>
<td>11.4%</td>
</tr>
<tr>
<td>Avg. # of related reviews per question</td>
<td>4.61</td>
</tr>
<tr>
<td>Avg. # of 1 rater agreeing on relevant sentence per question</td>
<td>2.19</td>
</tr>
<tr>
<td>Avg. # of 2 raters agreeing on relevant sentence per question</td>
<td>1.08</td>
</tr>
<tr>
<td>Avg. # of 3 raters agreeing on relevant sentence per question</td>
<td>0.83</td>
</tr>
</tbody>
</table>
**Relevance and Answer Annotation**

- **Interrater agreement** (given 3 raters per review sentence):
  - Divide the sentences into 4 sets based on the number of raters that agreed on each sentence: $R_0$, $R_1$, $R_2$, $R_3$
  - **Randomly sample 20 sentences** from each of the last three sets ($R_1$, $R_2$, $R_3$)
  - **Manually inspect** the relevance of these sentences to the corresponding questions and calculate the accuracy for each set

<table>
<thead>
<tr>
<th></th>
<th>Acc. of $R_1$</th>
<th>Acc. of $R_2$</th>
<th>Acc. of $R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45%</td>
<td>90%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Relevance and Answer Annotation

- The average number of sentences from each set \((R_1, R_2, R_3)\) per question

<table>
<thead>
<tr>
<th>Table 1. The Statistic of POIREviewQA</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Annotated question</td>
</tr>
<tr>
<td>% of questions WITHOUT related reviews</td>
</tr>
<tr>
<td>Avg. # of related reviews per question</td>
</tr>
<tr>
<td>Avg. # of 1 rater agreeing on relevant sentence per question</td>
</tr>
<tr>
<td>Avg. # of 2 raters agreeing on relevant sentence per question</td>
</tr>
<tr>
<td>Avg. # of 3 raters agreeing on relevant sentence per question</td>
</tr>
</tbody>
</table>
Evaluation Metrics

- Mean average precision (MAP)

\[ \text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP}(q)}{Q} \]

- Area Under receiver operating characteristic Curve (AUC)

Figure 5. AUC (Area under the ROC Curve).
**Baseline Model**

- **Lucene based model**: For each question, we rerank all the sentences from the 10 reviews selected by Elasticsearch by **TF-IDF** model.
**Baseline Model**

- **Sentence embedding model** (Sanjeev Arora et al. ICLR 2017)

**Algorithm 1 Sentence Embedding**

**Input:** Word embeddings \( \{v_w : w \in \mathcal{V}\} \), a set of sentences \( \mathcal{S} \), parameter \( a \) and estimated probabilities \( \{p(w) : w \in \mathcal{V}\} \) of the words.

**Output:** Sentence embeddings \( \{v_s : s \in \mathcal{S}\} \)

1. for all sentence \( s \) in \( \mathcal{S} \) do
2. \[ v_s \leftarrow \frac{1}{|s|} \sum_{w \in s} \frac{a}{a + p(w)} v_w \]
3. end for
4. Form a matrix \( X \) whose columns are \( \{v_s : s \in \mathcal{S}\} \), and let \( u \) be its first singular vector
5. for all sentence \( s \) in \( \mathcal{S} \) do
6. \[ v_s \leftarrow v_s - uu^T v_s \]
7. end for
Evaluation

- We evaluate two baseline models against POIReviewQA by using AUC, MAP

<table>
<thead>
<tr>
<th></th>
<th>Lucene Based</th>
<th>Sentence Embedding</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>77.0%</td>
<td>79.2%</td>
</tr>
<tr>
<td>MAP</td>
<td>48.8%</td>
<td>41.0%</td>
</tr>
</tbody>
</table>
## Evaluation

- **Q**: Question
- **T**: POI Type
- **C**: Review Sentence
- **A**: Answer
- **H**: Human judgements
- **L**: Lucene
- **E**: Sentence Embedding

<table>
<thead>
<tr>
<th>Reason</th>
<th>Example</th>
<th>Ranking (H/L/E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paraphrase</td>
<td><strong>Q</strong>: About <strong>how long</strong> should I expect my visit to be? &lt;br&gt;<strong>T</strong>: Venues &amp; Event Spaces; Kids Activities &lt;br&gt;<strong>C</strong>: We were there for about 2 hours, including the show. &lt;br&gt;<strong>A</strong>: took 2 hrs</td>
<td>1/107/88 out of 158</td>
</tr>
<tr>
<td>Hyponym</td>
<td><strong>Q</strong>: Any good vegan choices? &lt;br&gt;<strong>T</strong>: Restaurants→Cajun/Creole &lt;br&gt;<strong>Sent</strong>: After scanning the menu for a bit however, I was able to find the tofu wings. &lt;br&gt;<strong>A</strong>: Tofu wings could be a choice</td>
<td>2/49/18 out of 83</td>
</tr>
<tr>
<td>Synonymy</td>
<td><strong>Q</strong>: Any recommendations on how to score a table? ... &lt;br&gt;<strong>T</strong>: Restaurants→French &lt;br&gt;<strong>C</strong>: I made a reservation a day in advance thinking it will be busy. &lt;br&gt;<strong>A</strong>: A day in advance</td>
<td>1/63/14 out of 98</td>
</tr>
<tr>
<td>Deduction</td>
<td><strong>Q</strong>: Are there classes for seniors? &lt;br&gt;<strong>T</strong>: Art Galleries; Art Schools &lt;br&gt;<strong>C</strong>: Great studio for all, including kids! &lt;br&gt;<strong>A</strong>: There are classes for seniors</td>
<td>1/60/15 out of 72</td>
</tr>
<tr>
<td>Common Sense</td>
<td><strong>Q</strong>: Do they <strong>buy comic books</strong>? &lt;br&gt;<strong>T</strong>: Shopping→Comic Books &lt;br&gt;<strong>C</strong>: The concerns: The store currently has no consignment or new issues.  &lt;br&gt;<strong>A</strong>: No</td>
<td>1/45/53 out of 62</td>
</tr>
</tbody>
</table>
Conclusion

- We introduce a evaluation benchmark dataset POIReviewQA\(^1\) which consists of 20k questions for 1022 Yelp business types to support open-domain POI search and QA.
- We build two baseline models: a \textbf{Lucene-based baseline model} and a \textbf{sentence embedding-based model} which indicate that the dataset presents a challenging problem for future research by the GIR community.
- The POIReviewQA Task follows the idea of \textbf{semantic signatures} for social sensing by which we can study POI types using patterns extracted from human behavior, e.g., what people write about places of a particular type.

\(^1\)http://stko.geog.ucsb.edu/poireviewqa/