Hyper-Local, Directions-Based Ranking of Places

Petros Venetis
Stanford University

Christian S. Jensen
Aarhus University

Hector Gonzalez
Google Inc.

Alon Halevy
Google Inc.
Search queries have local intent!

~25%  

~50%
Accurate locations for users

- IP-based positioning: 1-10s km
- Current smartphones: 10s m
Motivating example: looking for a cafe
Motivating example: looking for a cafe
Motivating example: looking for a cafe
New type of queries: hyper-local

- **Known** user location
- User determines
  - Interest (e.g., cafe)
  - Willingness to travel (e.g., 2 kms)
Problems

1) Which data sources?
   - Current solution: Reviews
     - Expensive
     - Lacking new businesses
     - Lacking time aspect
     - Sparse

2) How to scale?
Contributions

1) Directions query logs
   • Easy to access
   • Time aspect
   • Many, many more than reviews

2) Scalable ranking architecture
Outline

Data sources
Efficient ranking
Experiments
Directions query logs
Directions query logs

From

1900 Fifth Avenue, Seattle, WA 98101-1204
227 Yale Avenue North, Seattle, WA 98109-5

To

Suggested routes:
- Stewart St: 0.7 mi, 14 mins
- a St and John St: 0.7 mi, 14 mins

Walking directions to Espresso Vivace Alley 24

Or take Public Transit (Bus): 8 mins

Distance
Directions query logs

• Example database entries:
  • From, To: in (latitude, longitude) pairs

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Distance</th>
<th>Time Of Day</th>
<th>Day Of Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westin Seattle</td>
<td>Espresso Vivace</td>
<td>0.7 miles</td>
<td>12:20pm</td>
<td>Tuesday</td>
</tr>
<tr>
<td>House A</td>
<td>Espresso Vivace</td>
<td>2 miles</td>
<td>12:22pm</td>
<td>Tuesday</td>
</tr>
</tbody>
</table>

• Users *willingly* expose FROM and TO locations
How to use directions query logs?

- Destination popularity
- Distances traveled to reach a destination
- Co-located people's destinations
- Time-based destination popularity
How to use directions query logs?

- Destination popularity
- Distances traveled to reach a destination
- Co-located people's destinations
- Time-based destination popularity
Scoring model: example
Scoring model: example
Scoring model: example
Scoring model: example
Scoring model: example
Scoring model

- **User U, place P**

- **Popularity(P):**
  - Quality/Popularity of place P

- **Willingness(distance(U, P)):**
  - Willingness of user U to travel the distance to place P
  - Assumption:
    - Non-increasing function of the distance

- **Score(U, P):**
  - Popularity(P) x Willingness(distance(U, P))
Ranking process

Query

Top-k businesses

Ranking

Retrieval

Online

Offline

Business Directory

Combined Scorers

Region index
Ranking process (offline)

1) Divide area into regions
2) Within regions rank businesses **by** (offline) combined scores
Ranking process

Online

Offline

Business Directory

Combined Scorers

Region index

Query

Top-k businesses

Ranking

Retrieval
Ranking process

- Find interesting regions
- Calculate distances between user and regions
- Retrieve next most promising business from lists
- Stop when you have k items
Experiments

- How valuable are directions logs?
  - Comparison with reviews
- How efficient is our ranking architecture?
  - Comparison with baseline
Dataset

- Directions logs (Google)
  - July 2009
  - Subset of USA
  - ~19M unique destinations

- Business listing (Google)
  - ~150K businesses
    - Museums, hotels, restaurants, bars, clubs, landmarks
  - Data quality issues
Query distribution across locations

![Query distribution graph](image-url)
Many more queries than reviews

- Reviews from Google, Yelp, and other systems

- **Queries** for businesses = ~50M
- **Reviews** for businesses = ~550K

- 20% higher coverage
  - Businesses with **queries** = ~130K (150K)
  - Businesses with **reviews** = ~100K (150K)
Many more queries than reviews

- Reviews from Google, Yelp, and other systems

  - Queries for businesses = ~50M
  - Reviews for businesses = ~550K

- 20% higher coverage
  - Businesses with queries = ~130K (/150K)
  - Businesses with reviews = ~100K (/150K)
Many more queries than reviews

- Reviews from Google, Yelp, and other systems

- **Queries** for businesses = ~50M
- **Reviews** for businesses = ~550K

- 20% higher coverage
  - Businesses with **queries** = ~130K (/150K)
  - Businesses with **reviews** = ~100K (/150K)
Evaluation framework

- Humans evaluate businesses
  - Provide a query
  - Evaluate each result (0-4)
- 10 users, 45 queries, 675 businesses
Evaluation framework

• Compared
  • Destination queries
  • Number of Reviews
  • Average score of Reviews

• Metrics
  • Average evaluation score
  • nDCG metric (how consistent a method's ranking is to humans' evaluations)
    – 0 is bad
    – 1 is good
Evaluators agree with directions log

- Average Evaluation Score for top-5
  - Destination queries: 1.96
  - Number of reviews: 1.453
  - Score of reviews: 1.498
Ranking is questionable...

- nDCG for top-5
  - Destination queries: 0.787
  - Number of reviews: 0.827
  - Score of reviews: 0.845
Performance evaluation

![Graph showing performance evaluation](image-url)
More in paper...

- Time aspect
  - Example: brunch restaurants
- One range of sizes works best in all cases
  - Approximately squares of with ~2-3 kms edges
Future directions

- Explore different ranking functions for different scenarios
- Personalized ranking
Conclusions

- Direction Query Logs
  - Numerous
  - Cheap
  - Retrieve good businesses
- Scalable architecture
  - Fast comparing to database solutions
  - Incremental in nature
Thank you!

Questions?