Learning to Generate Naturalistic Utterances Using Reviews in Spoken Dialogue Systems

Ryuichiro Higashinaka (NTT)
Rashmi Prasad (Univ. of Pennsylvania)
Marilyn Walker (Univ. of Sheffield)
Overview

• A new method for automatically acquiring a generation dictionary in spoken dialogue systems
• Uses user reviews on the web
• Uses ratings to map a sentence to its semantic representation with high precision
Objective

• Reduce the cost involved with hand-crafting a spoken language generation module

• Benefits:
  - Facilitate development of a spoken dialogue system
  - Achieve more natural system utterances using attested language examples from the web
NLG in Spoken Dialogue Systems

Semantic Representations

Assert-food_quality(Babbo, superb)  
Assert-décor(Babbo, superb)

Mappings between semantic representations and realizations

Assert-food_quality(X, superb) \iff X has superb food. 
Assert-décor(X, superb) \iff X has superb décor.

Utterances

Babbo has superb food.  
Babbo has superb décor.
Problem

• Mappings are created by hand
  ➔ *It is very costly*
• Quality may be left unchecked
  ➔ *Utterances can be unnatural*

We need an automatic method for acquiring mappings that generate naturalistic utterances
Related Work

• Create mappings from tagged corpora (Barzilay et al., 2002)
  ➔ Requires an hand-annotated corpus

• Find expressions from reviews
  - Adjectives for products (Hu and Liu, 2005)
  - Product features and adjectives with polarity (Popescu and Etzioni, 2005)
  ➔ Do not focus on creating mappings
Approach

• Automatically acquire mappings from user reviews on the web
  ➔ user reviews are widely available
• Uses ratings to derive accurate semantic representations for sentences
• A mapping is a triple consisting of
  - $U$ (Utterance), $R$ (Semantic representation), and $S$ (Syntactic structure)
  - Having syntactic structures is useful for full-NLG systems to perform syntactic transformation
Procedures

1. Collect user reviews on the web
   ➔ create a population of $U$
2. For each $U$,
   Derive semantic representation $R$
   Derive syntactic structure $S$
3. Filter inappropriate mappings
4. Add remaining mappings to dictionary
Collecting user reviews on the web

- Select review websites with individual ratings for review entities
- Collect review comments and ratings
- Collect tabular data

**Ratings**
Food, Service, Value, Atmosphere, Overall

**Tabular Data**
Name, Food Type, Location
Deriving semantic representation $\mathcal{R}$

**Domain Ontology**

(relations)

- RESTAURANT has foodquality
- RESTAURANT has servicequality
- RESTAURANT has valuequality
- RESTAURANT has atmospherequality
- RESTAURANT has overallquality
- RESTAURANT has FOODTYPE
- RESTAURANT has LOCATION

**Ratings**

Food, Service, Value, Atmosphere, Overall

**Tabular Data**

Name, Food Type, Location

**Distinguished Attributes**

Food = food, meal
Service = service, wait staff, …
Value = value, price, expensive, …
Atmosphere = atmosphere, décor, …
Overall = place, experience, …

**Utterance $\mathcal{U}$**

Named-entity Tagger

Distinguished attributes in $\mathcal{U}$

Corresponding relations in the domain ontology $\Rightarrow \mathcal{R}$
The best Spanish food in New York.

Review Comment
The best Spanish food in New York. I am from Spain and I had my 28th birthday there and we all had a great time. Salud!

Ratings
Food=5, Service=5, Value=5, Atmosphere=5, Overall=5

Review Sentence (U)
The best Spanish food in New York.

NE-tagged Review Sentence
The best {NE=Foodtype, string=Spanish} {NE=Food, string=food, rating=5} in {NE=Location, string=New York}.

Semantic Representation (R)
RESTAURANT has FOODTYPE
RESTAURANT has foodquality=5
RESTAURANT has LOCATION
(Rating-related relation has the same scalar value of that rating)
Filtering inappropriate mappings

- 6 Filters to guarantee the quality of mappings
- No Relations Filter, Other Relations Filter
  - Check whether a mapping has just the relations expressed in the ontology
- Contextual Filter
  - Checks whether $u$ can be uttered independently of the context
- Unknown Words Filter (typos, etc.), Parsing Filter, Duplicates Filter
<table>
<thead>
<tr>
<th>Ratings</th>
<th>Review Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food=5, Service=5, Value=5, Atmosphere=5, Overall=5</td>
<td>The best Spanish food in New York. I am from Spain and I had my 28th birthday there and we all had a great time. Salud!</td>
</tr>
</tbody>
</table>

**Review Sentence ($U$)**

- The best Spanish food in New York. **OK**

- I am from Spain and I had my 28th birthday there and we all had a great time. **NG**

- Salud!

**Filtered by No Relations Filter**
Experiment

• Obtaining mappings in the restaurant domain
• Reviews collected from we8there.com
  - 3,004 user reviews on 1,810 restaurants
  - 18,466 sentences
  - Obtained 451 mappings after filtering
• Objective and subjective evaluations
Objective Evaluation

- **Domain coverage**
  - Check how many of the relations in the domain ontology can be expressed by the obtained mappings

- **Linguistic variation**
  - Check the expressiveness of the mappings

- **Generativity**
  - Check whether the mappings can be incorporated in conventional generation engines
Domain Coverage

Distribution of mappings containing a single relation

<table>
<thead>
<tr>
<th>Relation</th>
<th>Rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>food</td>
<td></td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>18</td>
<td>57</td>
<td>94</td>
</tr>
<tr>
<td>service</td>
<td></td>
<td>15</td>
<td>3</td>
<td>6</td>
<td>17</td>
<td>56</td>
<td>97</td>
</tr>
<tr>
<td>atmosphere</td>
<td></td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>value</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>overall</td>
<td></td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>23</td>
<td>15</td>
<td>21</td>
<td>64</td>
<td>201</td>
<td>327</td>
</tr>
</tbody>
</table>

Domain Coverage is almost complete.
Linguistic Variation

• 137 syntactic patterns, 275 distinct lexemes, 2-15 lexemes per DSyntS (mean 4.63)

• Adjectival phrases
  - food=1 ➔ awful, bad, cold, burnt
  - service=1 ➔ silly and inattentive
  - atmosphere=5 ➔ comfortable, mellow

The mappings show good variation of language
Example mappings

**[RESTAURANT has foodquality=1]**

[1] Food was cold or burnt.
[2] The food was awful.
[3] The food was very ordinary.

**[RESTAURANT has foodquality=2]**

[1] Bad food.
[2] The food was very bland.
[4] The food is very bland.
[5] Food was not very good.
[6] Very bland flavored food. etc.

**[RESTAURANT has foodquality=5]**

[1] The food was excellent.
[4] The food is delicious but simple.
[6] The food was very good.
[7] The food was exquisite.
[8] Food is wonderful.
[9] The food was simply inspired.
[10] The food was absolutely great.
[11] Food was excellent.
[12] The food was great.
[13] The food was superb.
[14] All the food is terrific.
[15] The food was fantastic. etc.
Generativity

• Incorporating the learned mappings into SPaRKy generator (Walker et al., 2003)

**SPaRKy**: Babbo has the best overall quality among the selected restaurants with excellent decor, excellent service and superb food quality.

**With learned mappings**: Babbo has the best overall quality among the selected restaurants because *atmosphere is exceptionally nice, food is excellent and the service is superb.*

Successfully incorporated into a conventional generation engine
Subjective Evaluation

• 10 native English speakers evaluated baseline vs. learned mappings
  - Baseline: 27 hand-crafted mappings (taken from SPaRKy generator)
  - Learned: 451 learned mappings

• Evaluation criteria:
  - Consistency between semantic representations and realizations
  - Naturalness of realizations
  - 1-5 Likert scale
Results

- Consistency is significantly lower, but still high
- Naturalness is significantly higher

The method creates good quality mappings
Conclusion

• A new method for automatically acquiring a generation dictionary in spoken dialogue systems
  - Uses user reviews on the web
  - Uses ratings to obtain accurate semantic representations for sentences

• Experimental results showed the effectiveness of our approach