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

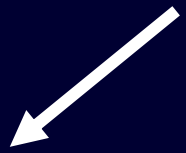
- Reduce the cost involved with hand-crafting a spoken language generation module
- Benefits:
  - Facilitate development of a spoken dialogue system
  - Achieve more natural system utterances using attested language examples from the web

NLG in Spoken Dialogue Systems

Semantic Representations

Assert-food_quality(Babbo, superb)

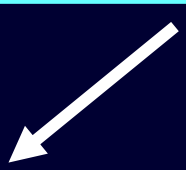
Assert-décor(Babbo, superb)



Mappings between semantic representations and realizations

Assert-food_quality(X, superb) ⇔ *X has superb food.*

Assert-décor(X, superb) ⇔ *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
  - $\mathcal{U}$ (Utterance), $\mathcal{R}$ (Semantic representation), and $\mathcal{S}$ (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 $\mathcal{U}$
2. For each $\mathcal{U}$,
   Derive semantic representation $\mathcal{R}$
   Derive syntactic structure $\mathcal{S}$
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 $\mathcal{R}$

Ratings

Food, Service, Value, Atmosphere, Overall

Tabular Data

Name, Food Type, Location

Distinguished Attributes

Meronymy relation with the review entity

Domain Ontology (relations)

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

Named-entities for 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 ➔ $\mathcal{R}$

Ratings

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

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!

Review Sentence $(\mathcal{U})$

The best Spanish food in New York.

DSyntS converter

DSyntS $(\mathcal{S})$

$$\begin{bmatrix}
\text{article} : \text{def} \\
\text{class} : \text{common_noun} \\
\text{lexeme} : \text{food} \\
\text{number} : \text{sg} \\
\text{ATTR} \begin{bmatrix} \text{class} : \text{adjective} \\ \text{lexeme} : \text{best} \end{bmatrix} \\
\text{ATTR} \begin{bmatrix} \text{article} : \text{no-art} \\ \text{class} : \text{common_noun} \\ \text{lexeme} : \text{FOODTYPE} \\ \text{number} : \text{sg} \end{bmatrix} \\
\text{ATTR} \begin{bmatrix} \text{class} : \text{preposition} \\ \text{lexeme} : \text{in} \\ \text{II} \begin{bmatrix} \text{article} : \text{no-art} \\ \text{class} : \text{proper_noun} \\ \text{lexeme} : \text{LOCATION} \\ \text{number} : \text{sg} \end{bmatrix} \end{bmatrix}
\end{bmatrix}$$

NE-tagged Review Sentence

The best {NE=Foodtype, string=Spanish} {NE=Food, string=food, **rating=5**} in {NE=Location, string=New York}.

Semantic Representation $(\mathcal{R})$

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

Filtering Example

Ratings	*Review Comment*
Food=5, Service=5, Value=5, Atmosphere=5, Overall=5	*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!*

Review Sentence $(\mathcal{U})$

- The best Spanish food in New York.

OK

Successfully created $\mathcal{R}$ and $\mathcal{S}$ → *Added to the dictionary*

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

NG

No distinguished attributes are found ➔ *cannot create $\mathcal{R}$* → *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

Relation \ Rating	1	2	3	4	5	Total
food	***5***	***8***	***6***	***18***	***57***	94
service	***15***	***3***	***6***	***17***	***56***	97
atmosphere	***0***	***3***	***3***	***8***	***31***	45
value	***0***	***0***	***1***	***8***	***12***	21
overall	***3***	***2***	***5***	***15***	***45***	70
Total	23	15	21	64	201	327

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.
[4] The food has gone downhill.
[5] Bad food.

[RESTAURANT has foodquality=2]

[1] Bad food.
[2] The food was very bland.
[3] Not enough food.
[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.
[2] The food is delicious.
[3] Keep up the great food.
[4] The food is delicious but simple.
[5] Great food.
[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.

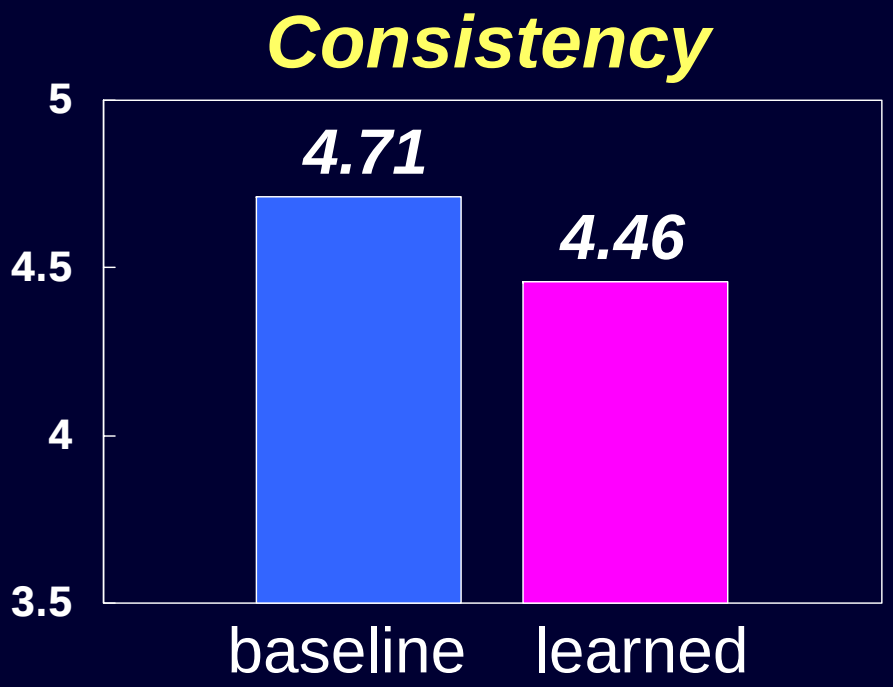
  With learned mappings: Babbo has the best overall quality among the selected restaurants because *atmosphere is exceptionally nice, food is excellent and the service is superb*.

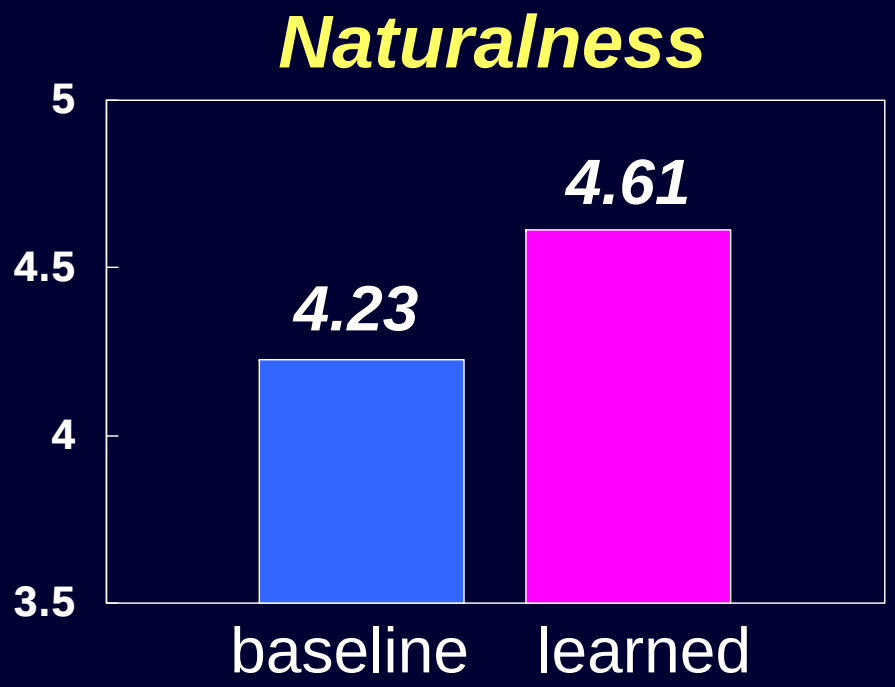
Successfully incorporated into
a conventional generation engine

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