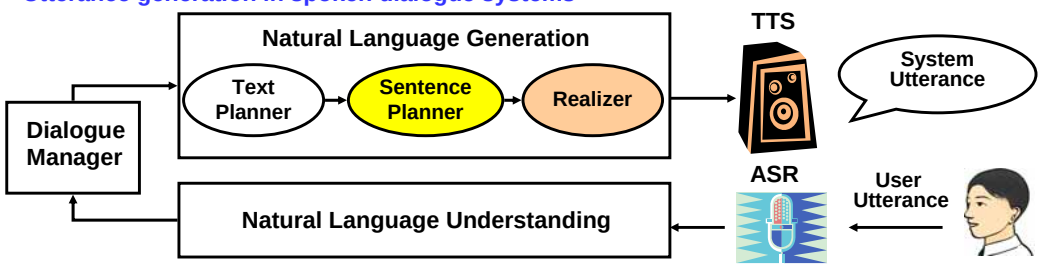
Augmenting variation of system utterances using corpora in spoken dialogue systems

Ryuichiro Higashinaka, Rashmi Prasad, and Marilyn Walker
NTT Communication Science Laboratories, NTT Corporation, Japan
Institute for Research in Cognitive Science, University of Pennsylvania, USA
Department of Computer Science, University of Sheffield, UK

Overview

- Current spoken dialogue systems speak very rigidly. (e.g., template-based, rule-based generation)
- We propose augmenting the variation of system utterances by incorporating useful sentences from corpora.
- Created a generator with more variation in the restaurant domain.

Utterance generation in spoken dialogue systems



Conventional Method

- Use trainable sentence planner to generate variation (SPaRKy).
- Variation is created by randomly combining basic syntactic structures.

Problem

- Proposition to basic syntactic structure mappings are prepared manually, which is very costly.
- The lack of mappings fundamentally limits the overall variation.

Text plan

Propositions:
p1 assert-best(Babbo)
p2 assert-food_quality(Babbo, superb)
p3 assert-decor(Babbo, excellent)
p4 assert-service(Babbo, excellent)
Relations:
justify(nuc:p1, sat:p2)
justify(nuc:p1, sat:p3)
justify(nuc:p1,sat:p4)

Random combination of associated basic syntactic structures

Example utterances of SPaRKy

Babbo has the best overall quality among the selected restaurants since it has excellent service with excellent decor and it has superb food quality.

Babbo has excellent decor and it has superb food quality with excellent service. It has the best overall quality among the selected restaurants.

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

Proposition to Basic Syntactic Structure mappings

assert-best(X)
assert-food_quality(X, superb)
assert-decor(X, excellent)
assert-service(X, excellent)
etc.

class:verb
lexeme: have
I [class:proper_noun, lexeme: X]
II [class:common_noun, lexeme: decor, ATTR [class:adjective, lexeme: excellent]]

Approach

- Automatically obtain proposition to basic syntactic structure mappings from corpora.
- Focus on restaurant recommendation utterances.

Step1 Create a corpus by collecting restaurant reviews on the web.

Step2 Create a proposition expression for each sentence in the corpus.

1. The best Spanish food in New York.
   Keyword spotting and named entity recognition
2. The best (food_type=Spanish) (rating_key=food) in (location=New York).
   Rating : "food=5"
3. {Food=5, location, food_type}
4. assert-food_quality(X, superb), assert-nbhd(X, Y), assert-food_type(X, Z)

Step4 Create mapping

class:common_noun
lexeme: food
article:def
ATTR best [class:adjective]
ATTR Food_Type [class:common_noun, article:no-art]
ATTR in [class:preposition, II Location [class:proper_noun]]

Step3 Convert sentences into basic syntactic structures. (using a rule-based converter)

Review Example

Ratings
Food=5, Service=5, atmosphere=5, Price/Value=5, Overall=5
User 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!

From the corpus of 18811 sentences, 514 mappings were obtained.

Experiment

- Comparison between SPaRKy and SPaRKy+ (restricted to use at least one of the obtained mappings).
- 5 subjects rated the output of SPaRKy and SPaRKy+ on the scale of 1-5.
- Generated 10 utterances for 15 text plans; obtained 750 ratings for each generator.

Results

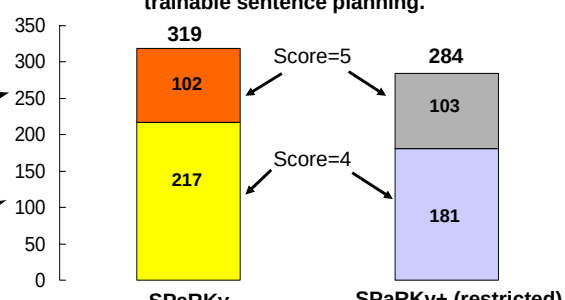
Example utterances of SPaRKy+

Great atmosphere, excellent food and friendly waiters. Babbo has the best overall quality among the selected restaurants.

Babbo has the best overall quality among the selected restaurants since the food, service and atmosphere are all excellent.

Babbo has the best overall quality among the selected restaurants. The food is excellent, the wait staff is professional and the decor is beautiful and very comfortable.

Number of highly-rated utterances
(NB. Only highly-rated utterances are important in trainable sentence planning.



- Comparable utterance quality using corpora
- By selecting to use/not to use corpora, system can speak with more variation