

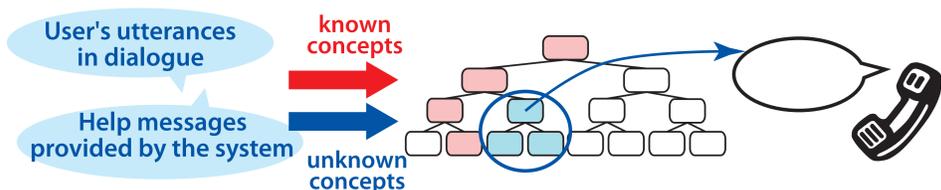


Introducing Utterance Verification in Spoken Dialogue System to Improve Dynamic Help Generation for Novice Users

Kazunori Komatani, Yuichiro Fukubayashi, Tetsuya Ogata, Hiroshi G. Okuno
Graduate School of Informatics, Kyoto University, Japan

Dynamic Help Generation

Help is generated by estimating the gap between user's mental model and the system.



1. Extract **known** (unknown) concepts from the utterances and help messages
2. Update "know degrees" of a domain concept tree.
3. Search an appropriate help message from the tree

Classification of In-grammar Utterances

Updating method of "known degree" has not been robust against ASR errors.

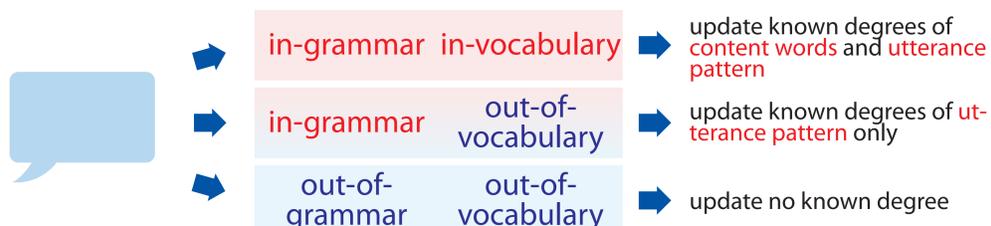
User's input: "Please tell me an approach to Yoshida Shrine." (out-of-vocabulary)
ASR output: "Please tell me the address to Yoshida Shrine."

- ➡ The user does not know the acceptable content word.
- ➡ The user knows the acceptable utterance pattern.

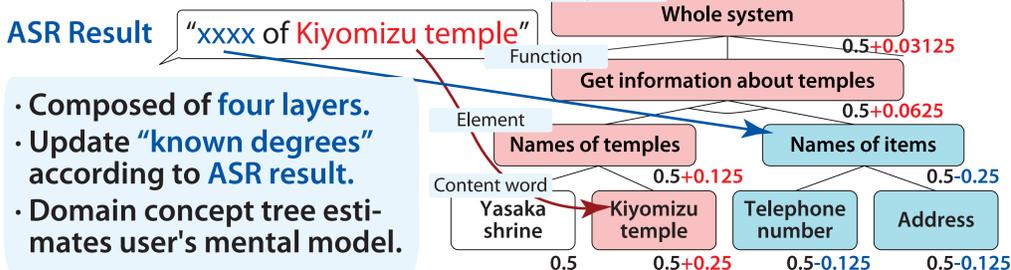
Simple keyword spotting can not distinguish in-grammar utterances, which contain out-of-vocabulary words.

Goal

Classify user's utterance into three groups and improve an estimation of user's mental model.



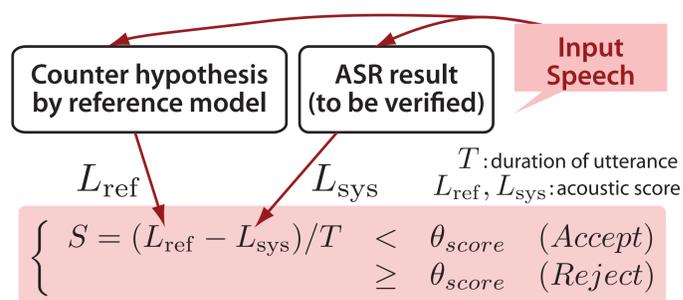
Domain concept tree



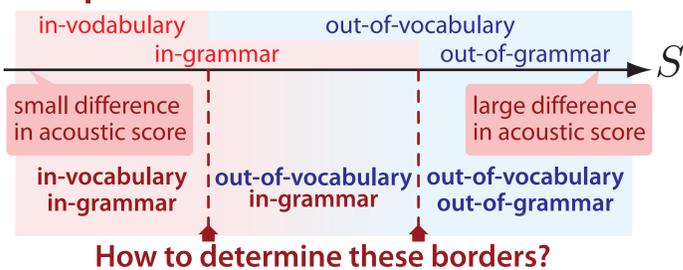
- Composed of four layers.
- Update "known degrees" according to ASR result.
- Domain concept tree estimates user's mental model.

Utterance Classification for Help Generation

Utterance verification with acoustic score



Assumption about acoustic score



Determination of thresholds with corpus

$SErr$ (slot error)
the ratio of correct utterances that are not accepted

FA (false acceptance)
the ratio of false acceptance

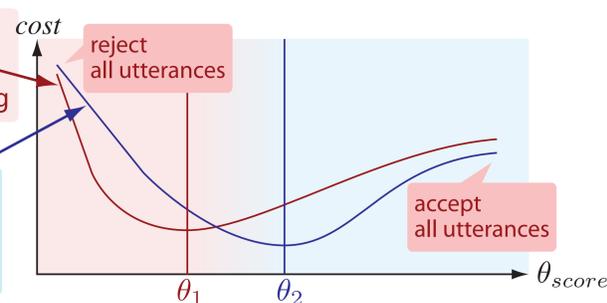
Acc
accuracy of language understanding

$$cost_1 = \frac{FA + SErr + (1 - Acc)}{3}$$

Considering accuracy of acceptance and language understanding

$$cost_2 = \frac{FA + SErr}{2}$$

Considering accuracy of acceptance only.



Classification of utterances

	grammar	vocabulary	Classification and updating
$S < \theta_1$	in-grammar	in-vocabulary	Accept LU results and update known degree of utterance pattern and content word.
$\theta_1 \leq S < \theta_2$	in-grammar	out-of-vocabulary	Reject LU results and update known degree of utterance pattern only.
$S \geq \theta_2$	out-of-grammar	out-of-vocabulary	Reject LU results and does not update known degrees.

Experimental Evaluation

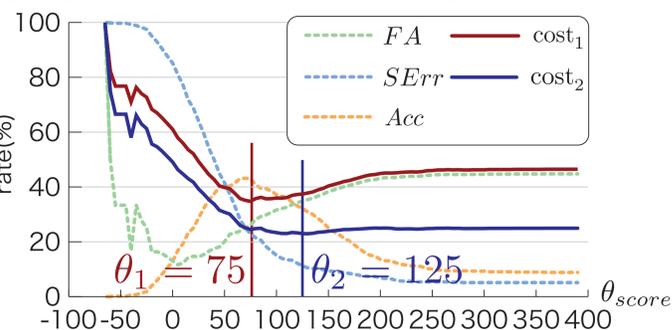
Dialogue data

- collected by using Kyoto sightseeing guide system
- 1512 utterances from 12 novice users
- grammar-based ASR engine (Julian)
- vocabulary size was 673 words
- average ASR accuracy was 42.9%

Reference model for verification

- statistical language model-based ASR engine (Julius)
- vocabulary size was 20000 words
- the model obtained from newspaper articles

Determination of the thresholds



Result

Comparison with the simple classification in which utterances with $S \geq \theta_2$ are simply rejected.

Correct answer of UV	LU result	S			Expected result
		$S < \theta_1$	$\theta_1 \leq S < \theta_2$	$S \geq \theta_2$	
Accept (in-grammar)	Correct (100%)	454	50	8	Most of the correct LU results were correctly accepted.
	Some errors (<100%)	84	29	34	
	No output	28	13	23	
Reject (out-of-grammar)	Some errors (insertion error)	158	104	185	Incorrect LU results were correctly rejected, and user's knowledge about utterance patterns was correctly reflected.
	No output (correct rejection)	166	86	98	

UV: utterance verification, LU: language understanding

Unexpected result

Some out-of-grammar utterances could not be rejected.

Correct LU results were incorrectly rejected, but user's knowledge about utterance patterns was correctly reflected.

- ➡ The utterance verification algorithm was very simple: only the difference in acoustic scores between two recognizers were used. Differences in acoustic scores were not large enough for short utterances: the difference are too small to classify the utterances.

Future work

- Extend our method with another features to improve the accuracy of classification.
- Integrate this method into the dynamic help generation.