

LSTM을 이용한 생성모델 기반 대화 상태 추적기

Robust Generative Dialog State Tracker using LSTM

Youngsoo Jang*, Jiyeon Ham*, Byung-Jun Lee, Youngjae Chang, Kee-Eung Kim

URP Researcher Youngjae Chang, School of CS
(youngjae.chang@kaist.ac.kr)

Professor Kee-Eung Kim, School of CS
TA Byung-Jun Lee, School of CS

AILAB
Artificial Intelligence & Probabilistic Reasoning

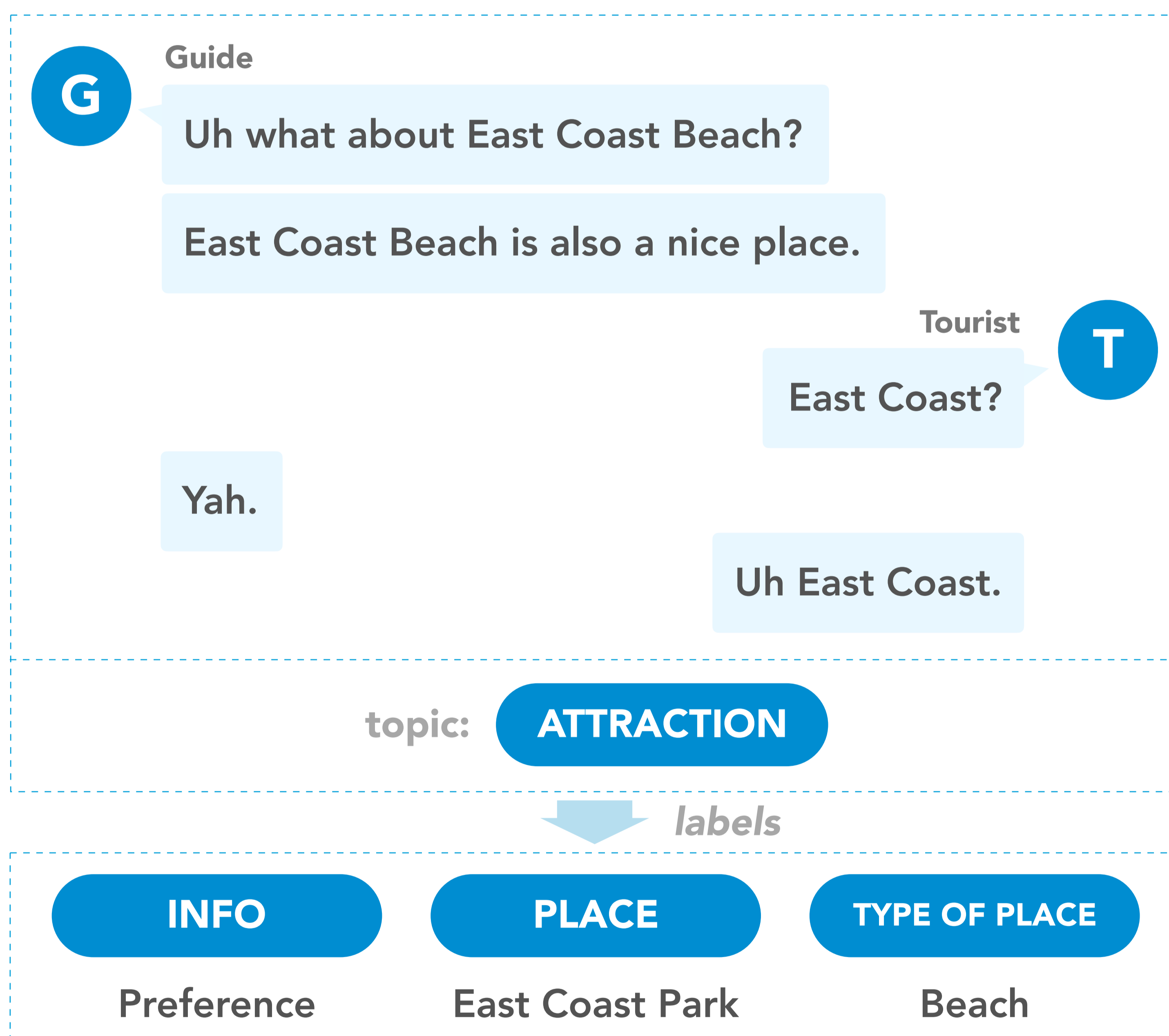
URP

Abstract

Dialog state tracker, a subcomponent of a spoken dialog system, records a state of dialog during a conversation. For example, in a travel planning system, the dialog state tracker will record parameters for the hotel a user is searching for, i.e. location, price range and desired star rating.

We've participated in the Dialog State Tracking Challenge 5 (DSTC5) developing a dialog state tracker with other fellow researchers. To extract dialog state from the cross-language human-human dialog with a limited training set, we've proposed a tracker that focuses on the meaningful words using an attention mechanism and a bi-directional long short term memory (LSTM). The nouns are vectorized using related texts crawled from the web, providing a good embedding for the words not presented in the training dialogs. Our proposed tracker achieved high accuracy without sophisticated pre- and post-processing.

Objectives



Challenges

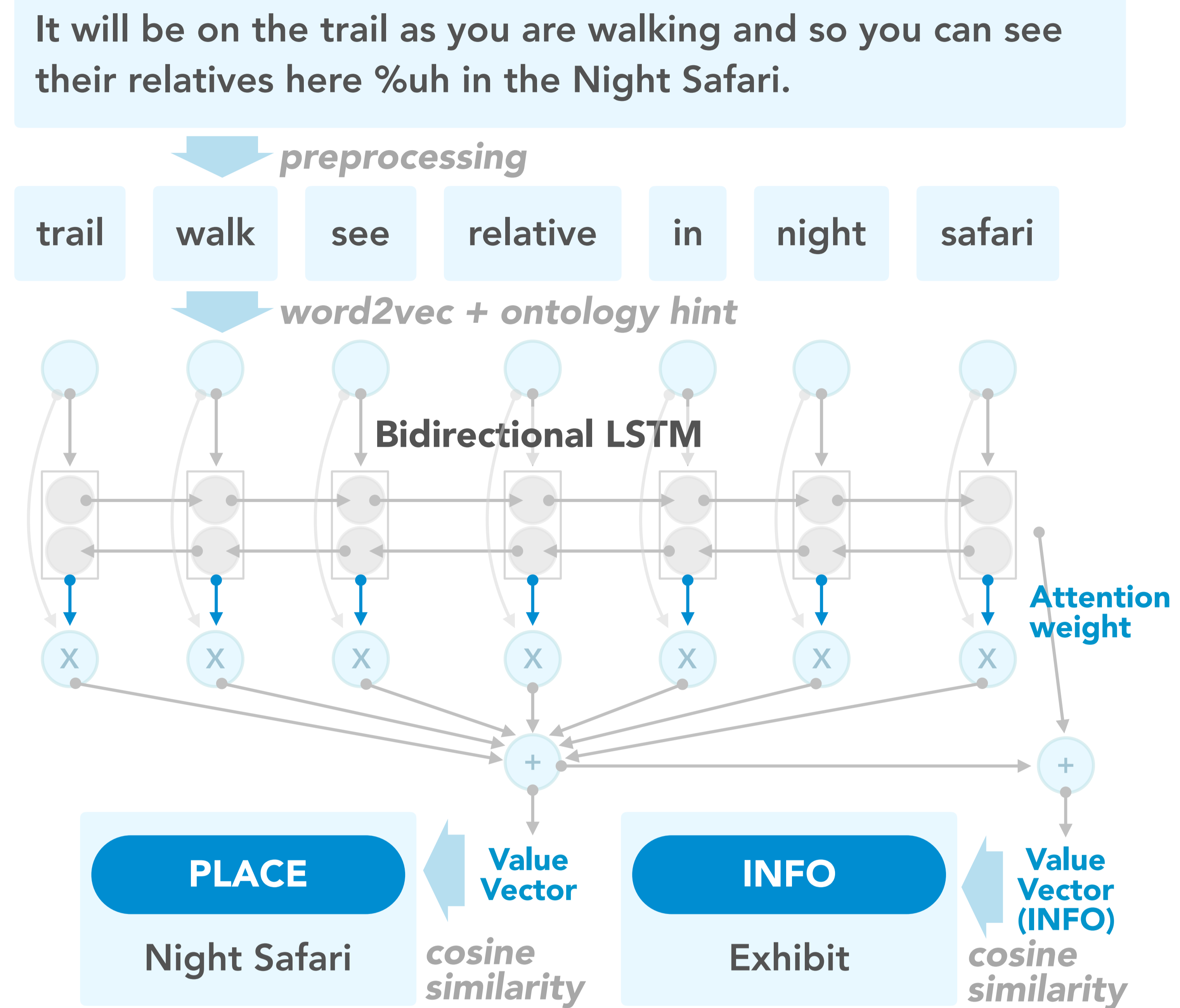
- Limited Dataset: 35 dialogs for training, 2 dialogs for validation
 - Only 56% of ontology keywords exists on training dialogs.
- Robustness: Input dialog may have errors
 - Validation & Examination is done on the Chinese dialog; hence the training set is in English. We've used the machine translated versions of Chinese dialogs.

Methodology

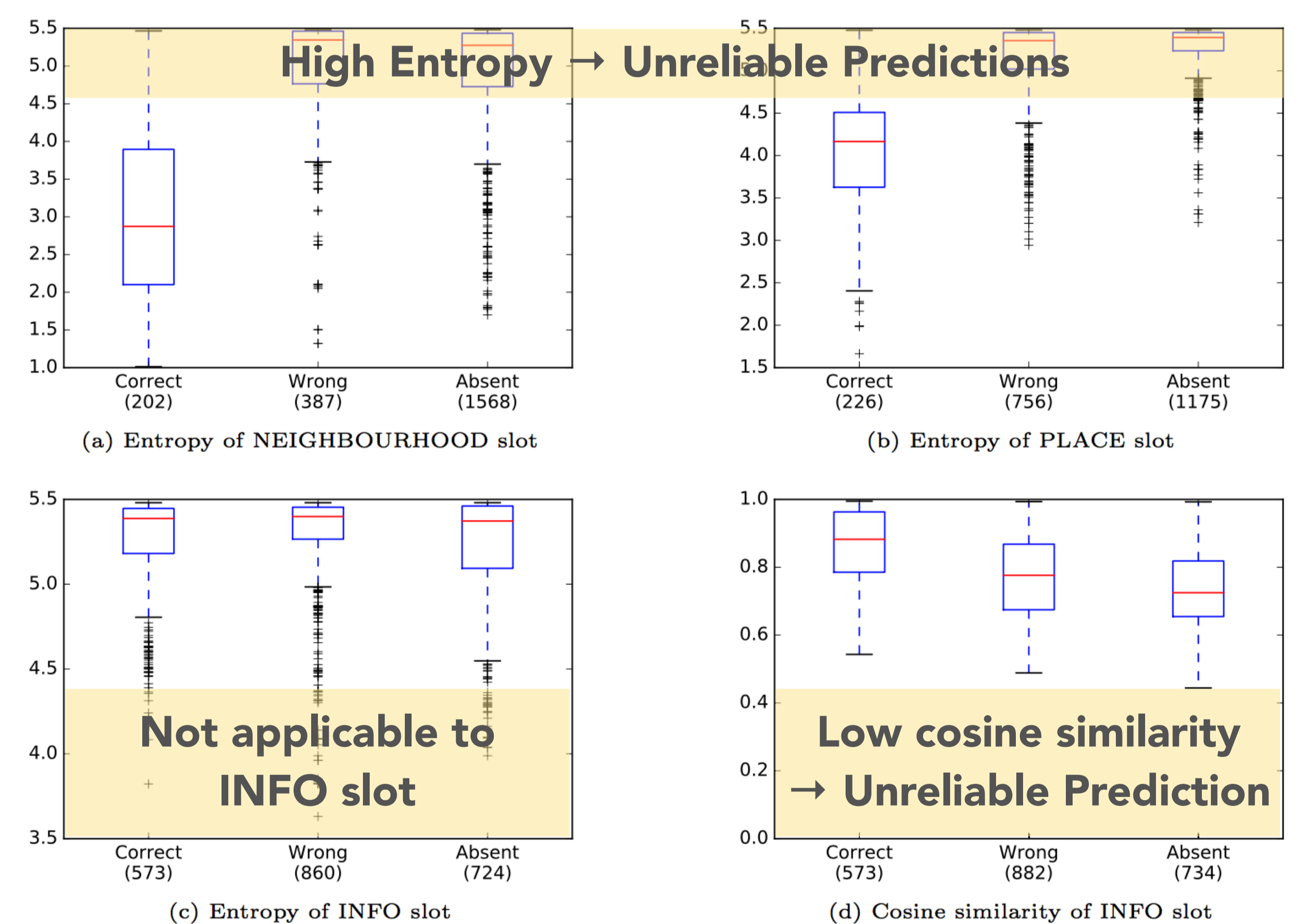
Preprocessing: POS-tagging, Stemming & Lemmatizing
Embed words to 100dim Vector (Word2Vec) using crawled reviews

Prediction: Attention weight for each word w/ Bidirectional LSTM
Exclude unreliable prediction w/ entropy & cosine similarity (INFO)

Architecture



Excluding Unreliable Predictions



Results

Team	Schedule 1				Schedule 2			
	Accuracy	Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure
Baseline 1	0.0250	0.1148	0.1102	0.1124	0.0321	0.1425	0.1500	0.1462
Baseline 2	0.0161	0.1743	0.1279	0.1475	0.0222	0.1979	0.1774	0.1871
Team 1	0.0417	0.3650	0.2795	0.3166	0.0612	0.3811	0.3548	0.3675
Team 2	0.0788	0.5195	0.3315	0.4047	0.0956	0.5643	0.3769	0.4519
Team 3	0.0351	0.3216	0.1515	0.2060	0.0505	0.3350	0.2045	0.2539
Team 4	0.0583	0.4008	0.2776	0.3280	0.0765	0.4127	0.3284	0.3658
Team 5	0.0330	0.3377	0.2318	0.2749	0.0520	0.3637	0.3044	0.3314
Ours	0.0491	0.4684	0.2193	0.2988	0.0643	0.4758	0.2623	0.3381
Team 7	0.0092	0.4287	0.0431	0.0783	0.0107	0.4000	0.0441	0.0794
Team 8	0.0192	0.3130	0.1048	0.1570	0.0214	0.3021	0.1046	0.1554
Team 9	0.0231	0.1139	0.1090	0.1114	0.0314	0.1412	0.1487	0.1449

The table summarizes the DSTC 5 final result.

Schedule 1 (utterance level evaluation): 3rd place on accuracy

Schedule 2 (segment level evaluation): 3rd place on accuracy