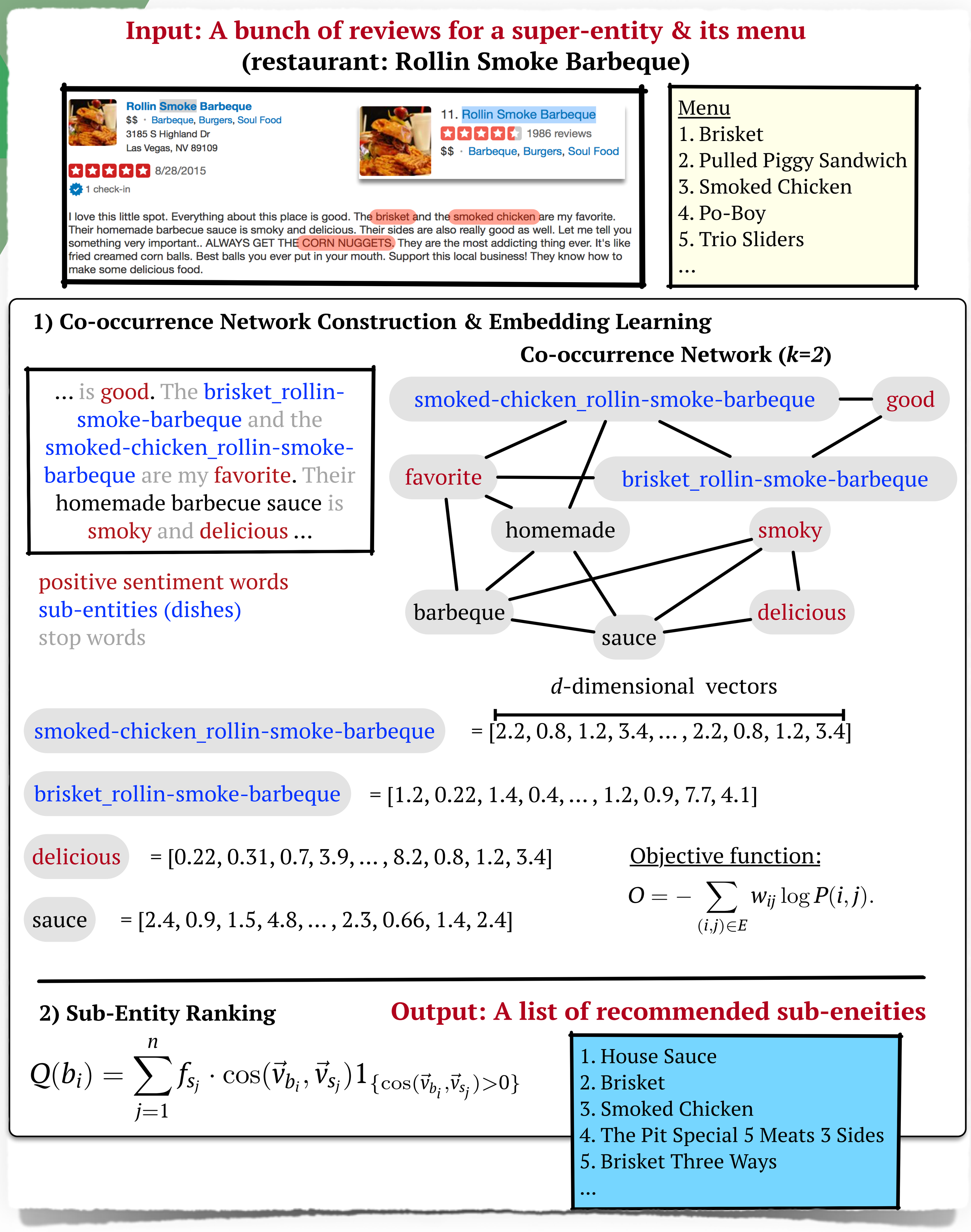
Text Embedding for Sub-Entity Ranking from User Reviews

This work attempts to conduct analysis for one certain type of user reviews; that is, the reviews on a super-entity (e.g., restaurant) involve descriptions for many sub-entities (e.g., dishes). To deal with such analysis, we propose a text embedding framework for ranking sub-entities from user reviews of a given super-entity. Experiments on two real-world datasets show that our method outperforms three baselines by a statistically signifiant amount. Intriguing cases from the experiments are discussed.



Methodology

Stage 1: Co-occurrence Network Construction & Embedding Learning

- Each word (sub-entity, sentiment words and other words) is a vertex.
- Each edge is associated with a positive weight w_{ij} , the frequency of word j occurring in the context of word i.
- Minimize the objective function to learn the representations of words and sub-entities.

Stage 2: Sub-Entity Ranking

- The sub-entities are ranked via a scoring function based on the learned word and sub-entity representations.

Experiments

- Regex is used to extract sub-entities.

  (french\s*|onion\s*)+sou[a-z]+(s|es|ies)?

- Ground Truth of TripAdvisor: the average rating stars of all user reviews for an attraction.

Datasets	TripAdvisor	Yelp
super-entity (city, restaurant)	25	256
Avg. # sub-entity per super-entity (attraction, dish)	20	104.8
# reviews	2,870,024	192,308
Avg. sentiment-sub-entity proximal distance	2.102	2.592

NDCG	city1		city2		city3		avg	
	@5	@10	@5	@10	@5	@10	@5	@10
P	0.761	0.934	0.602	0.651	0.624	0.743	0.657	0.734
B1	0.542	0.552	0.494	0.560	0.365	0.429	0.530	0.601
B2	0.542	0.578	0.494	0.547	0.365	0.429	0.534	0.610
B3	0.411	0.499	0.414	0.429	0.374	0.495	0.446	0.534

P: Proposed method B: Baseline

Top 5 dishes in Mount Everest India's Cuisine

Frequency (F)	Proposed method (P)
Naan (F:353, P:14)	Tandoori Chicken (F:48, P:1)
Chicken Tikka Masala (F:96, P:7)	Chicken Tikka (F:26, P:2)
Tandoori Chicken (F:48, P:1)	Gulab Jamun (F:15, P:3)
Mango Lassi (F:41, P:10)	Chicken Curry (F:27, P:4)
Chicken Makhani (F:28, P:5)	Chicken Makhani (F:28, P:5)

Conclusions

- A novel sub-entity ranking framework that incorporates the construction of co-occurrence networks and direct proximity embedding learning.
- In future work, the framework can be extended into different areas and incorporate other sentiment words or constructing a hierarchical entity graph.

TripAdvisor

Yelp

Chih-Yu Chao and Yi-Fan Chu
Dept. Computer Science,
University of Taipei,
Taipei,Taiwan

Chuan-Ju Wang and Hsiu-Wei Yang
Research Center for Information Technology Innovation,
Academia Sinica,
Taipei, Taiwan

Ming-Feng Tsai
Dept. Computer Science,
National Chengchi University,
Taipei, Taiwan

CIKM 2017