# 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



#### Input: A bunch of reviews for a super-entity & its menu (restaurant: Rollin Smoke Barbeque)

#### Menu Smoke Barbeque Rollin Smoke Barbequ \$\$ Barbeque, Burgers, Soul Food 1. Brisket 🖈 🖈 🖈 🖍 1986 reviews 3185 S Highland Dr Vegas, NV 89109 Barbeque, Burgers, Soul Food 2. Pulled Piggy Sandwich 8/28/2015 3. Smoked Chicken 4. Po-Boy ove this little spot. Everything about this place is good. The brisket and the smol omemade barbecue sauce is smoky and delicious. Their sides are also really good as well. Let me tell you 5. Trio Sliders nething very important.. ALWAYS GET THE CORN NUGGETS. They are the most addicting thing ever. It's like ied creamed corn balls. Best balls you ever put in your mouth. Support this local business! They know how to



#### & 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.

### $(french s^{onion} s^{onion} - z] + (s|es|ies)?$

 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

delicious = $[0.22, 0.31, 0.7, 3.9, \dots, 8.2, 0.8, 1.2, 3.4]$ sauce = $[2.4, 0.9, 1.5, 4.8, \dots, 2.3, 0.66, 1.4, 2.4]$	$\frac{\text{Objective function:}}{O = -\sum_{(i,j)\in E} w_{ij} \log P(i,j).}$
2) Sub-Entity Ranking Output: A list $Q(b_i) = \sum_{j=1}^n f_{s_j} \cdot \cos(\vec{v}_{b_i}, \vec{v}_{s_j}) \mathbb{1}_{\{\cos(\vec{v}_{b_i}, \vec{v}_{s_j}) > 0\}}$	of recommended sub-eneities 1. House Sauce 2. Brisket 3. Smoked Chicken 4. The Pit Special 5 Meats 3 Sides 5. Brisket Three Ways 

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

Avg. sentiment-sub-entity proximal distance	2.102	2.592

NDCG	cit	ry1	cit	city2		city3		avg	
	@5	@10	@5	@10	@5	@10	@5	@10	
Ρ	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

different areas and incorporate other sentiment words or constructing a hierarchical entity graph.



Chih-Yu Chao and Yi-Fan Chu Dept. Computer Science, University of Taipei, 臺北福主大学 Taipei,Taiwan



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



Yelp

CIKM 2017 ©CY Chao, YF Chu, HW Yang, CJ Wang, MF Tsai