Application of Deep Learning in Yelp Review Analysis

Renfeng Jiang[1], *Yimin Liu*[1], *Ke Xu*[2]
Department of [1]Energy Resources Engineering, [2]Electrical Engineering, Stanford University, USA

Motivation

- Millions of user reviews have been posted through Yelp. Automatic extraction of useful information from these reviews can be very beneficial for both users and businesses. Recently, Deep Learning models have been successfully applied in sentiment analysis in movie reviews [1]. And we would like to use this idea to help Yelp do better categorization for reviews and and also recommendation system based on specific item that customers are likely to search, ie., beef, nail.
- Our model learns the topics reflected from reviews through word2vec using deep learning.
- **Inputs:** Yelp Reviews $V \in R^{N*200}$, preprocessed by Stanford POS-tagger and generated by word2vec.
- **Outputs:** Business similarities$\alpha \in [-1\ 1]$ and clusters for categorization.

Dataset

- The Yelp Challenge dataset provides over one million user reviews and over half million tips. 5000 reviews of 500 different businesses were selected as our input data.
- Words in each piece of review are treated as features.

Reference

1. Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the conference on empirical methods in natural language processing (EMNLP). Vol. 1631. 2013.
2. Le, Quoc V., and Tomas Mikolov. "Distributed representations of sentences and documents." arXiv preprint arXiv:1405.4053 (2014).
3. word2vec: https://code.google.com/p/word2vec/

Future work

We plan to feed our model with all the reviews of over one million to:

- Categorize all the different types of businesses;
- Do more specific categorizations on restaurants;
- Get better search results with enough reviews feed.

Methodology

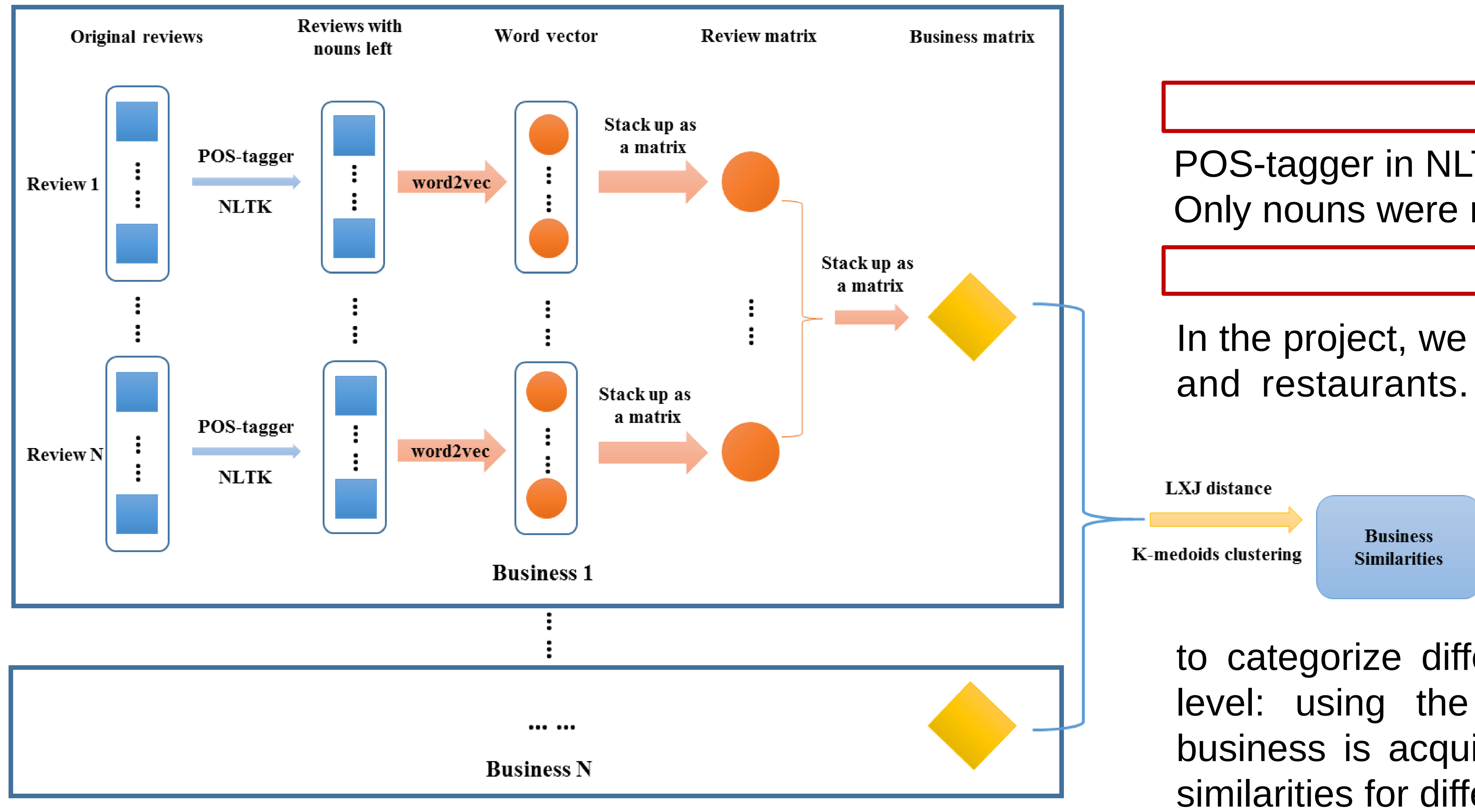


Fig. 1 Flow chart of the model

1. Data processing

POS-tagger in NLTK was applied to tag all the words in reviews.
Only nouns were retained as our features.

2. Model flow chart

In the project, we design a model (Fig. 1) with three levels: words, reviews, and restaurants. The basic level is the word level: inputting different words into word2vec pre-trained model generates different vectors corresponding different words. The second level is the review level: we simply stack up all the nouns in a review to form a review matrix. K-medoids clustering can be applied here to categorize different types of reviews. The third level is the business level: using the same techniques above, matrix representing each business is acquired. We use self-defined LXJ distance to calculate the similarities for different businesses to support the recommendation system. The LXJ distance, named by the first letter of the family name of our group members, is defined as follows to calculate the businesses similarities:

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \quad B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \quad A \otimes B = \underset{i}{mean}\left\{ \max_j \left(\frac{a_i^T b_j}{\|a_i\|_2 \|b_j\|_2} \right) \right\} \quad LXJ(A,B) = \frac{A \otimes B + B \otimes A}{2}$$

where A and B are two different business matrix, $\otimes$ are the operator defined by ourselves, which takes the max value among the inner product of different vectors and then take the average.

Results and Discussion

- After calculating the LXJ distance of 500 businesses, we used multidimensional scaling (MDS) to visualize these similarities and it turned out that we successfully categorized out different types of business, especially for restaurant, labeled in red in Fig. 2.
- To go one step further, we also categorized different types of restaurants, Asian, American, Mexican, etc (Fig. 3).
- Our model can further achieve food-wise search. For example, inputting beef can not only output Amercian beef but also Chinese beef. In addition, inputting beef China only outputs Chinese restaurants famous for beef.
- What is more, our model can also categorized reviews so as to help businesses get better insights on customer feedbacks, and to provide reviews highlights.

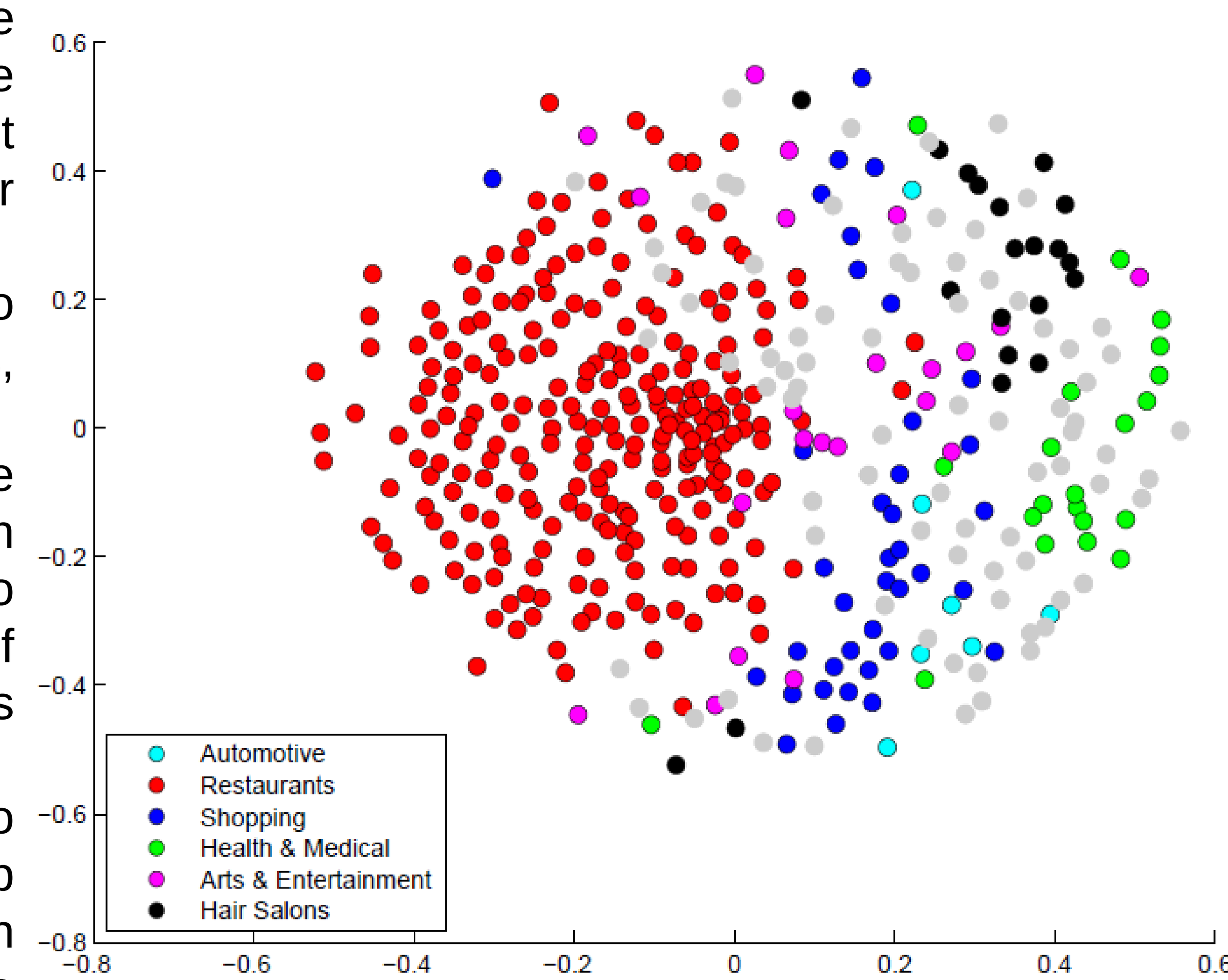
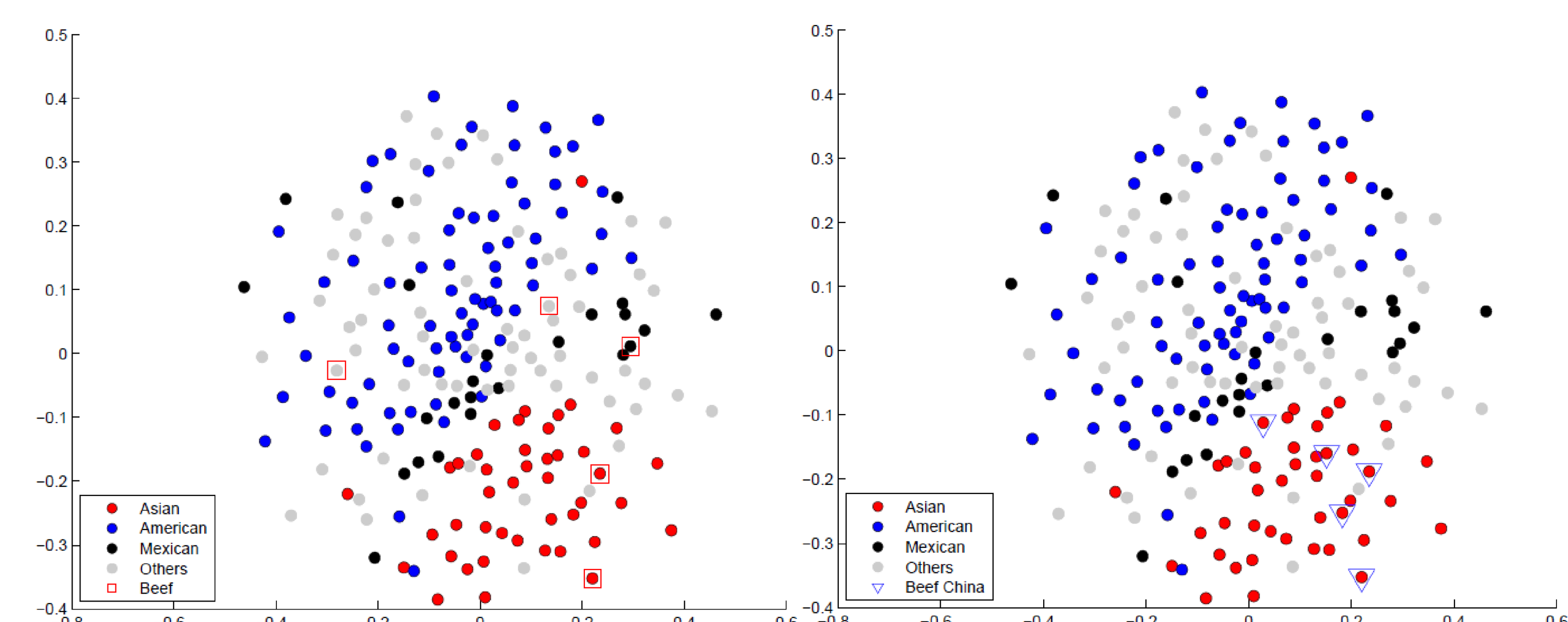


Fig. 2 Labeled businesses by similarities 1.



Input	
beef	beef China
Results	
Heng's Kitchen (Chinese Restaurants) Thé & Toast (Food Coffee & Tea Japanese Chinese Restaurants) Billy's BBQ (Soul Food Southern Barbeque Restaurants) ⋮	Heng's Kitchen (Chinese Restaurants) Thé & Toast (Food Coffee & Tea Japanese Chinese Restaurants) Pho Bistro(Vietnamese Chinese Restaurants) ⋮

Fig. 3 Labeled different types of restaurants and results by input beef & beef china