Why Restaurants Fail?

The Evidence from Yelp Restaurants Reviews

MACS 30250 Project Proposal

Li Liu

Research Question

What are the major factors that lead to restaurants' low rating scores in Yelp?

Importance:

- Discover the users experience and opinion by topic modeling
- Quantify the effect of users experience and business attributes on customers' satisfaction level
- Develop statistical models to predict the rating score

Theory/Literature

"Marketing is the business function that identifies current unfulfilled needs and wants, defines and measures their magnitude,....."

---Philip Kotler (marketing author; professor at Northwestern)

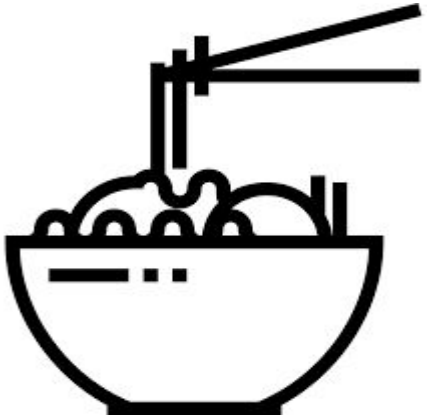
- Text as useful input for economic/business research (Gentzkow et al. 2017)
- Use sentence structure to infer and predict from customer reviews (Büschken and Allenby, 2016)
- Identify customer needs from user-generated content with ML (Timoshenko and Hauser, 2019)

- Gentzkow, Matthew and Kelly, Bryan T. and Taddy, Matt, Text As Data (February 15, 2017). Available at http://dx.doi.org/10.2139/ssrn.2934001
- Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. Marketing Science, 35(6), 953–975.
- Timoshenko, A., & Hauser, J.R. (2019). Identifying Customer Needs from User-Generated Content. Marketing Science, 38(1), 1-20.

Customers Experience (Examples)

Unreasonable prices

Bad services



Horrible food

Business Attributes (Examples)

Location
(City;
Neighborhood;
Competitors)

Category
(American;
Chinese; Thai;
Indian; French...)

Provide parking
lots or not

Data

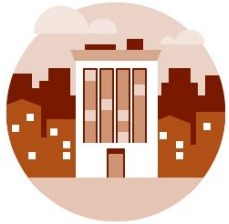
Source: Yelp Dataset Challenge (Round 13)

Size: 6 GB in JSON format

6,685,900 reviews

192,609 businesses

200,000 pictures

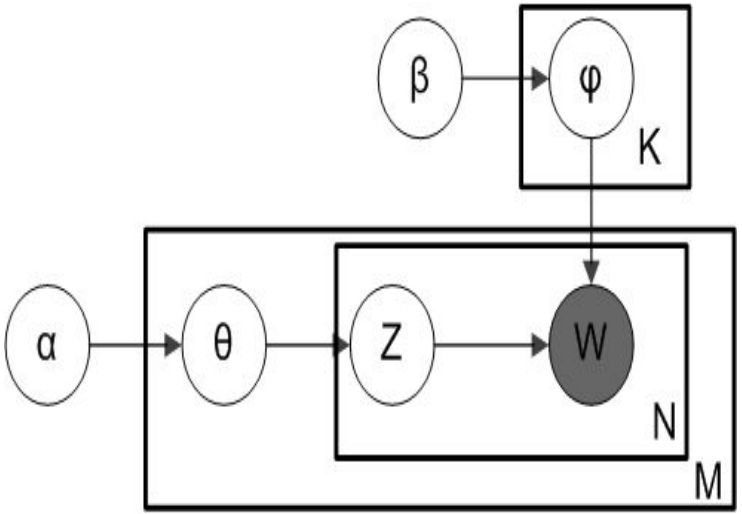


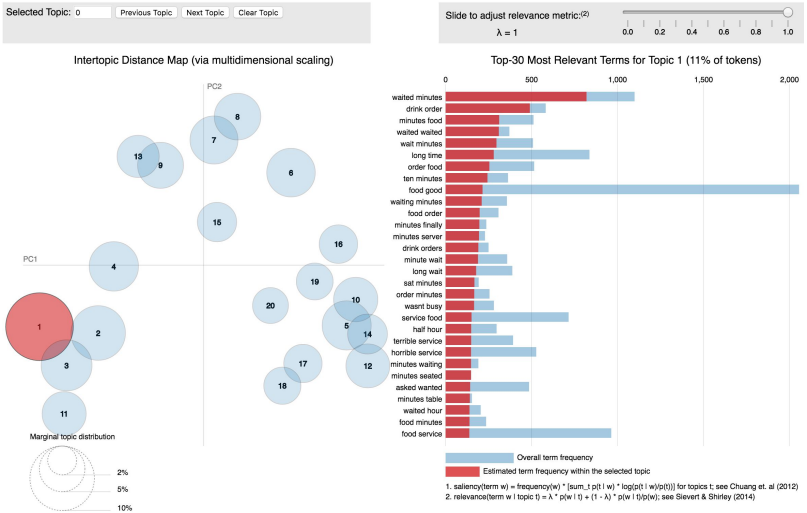
10 metropolitan areas

Topic Modeling for Reviews

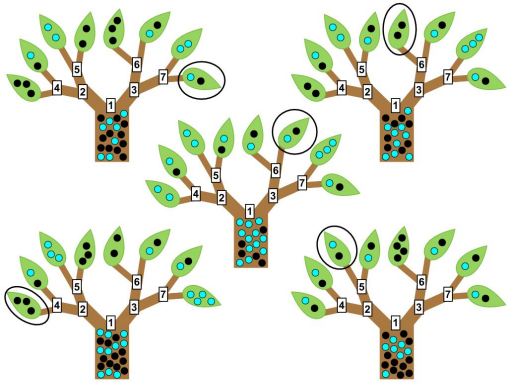
Purpose: find hidden semantic structures in documents

Algorithm: Latent Dirichlet allocation

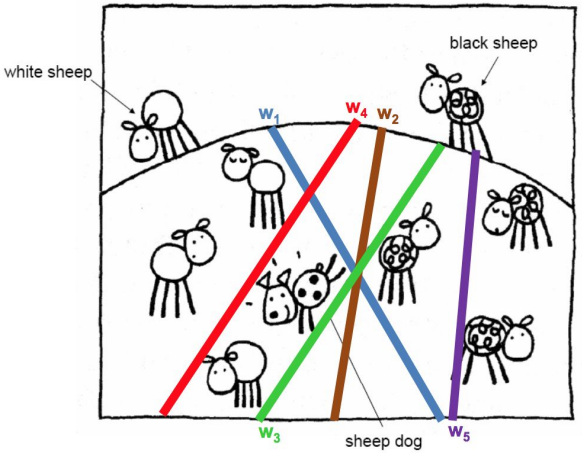




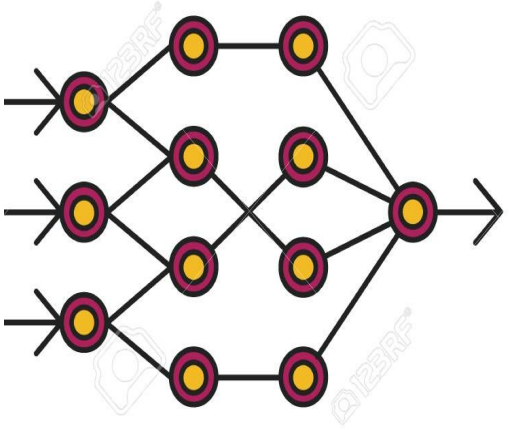
Predictive Model (Classification)



Tree Model
(Random Forest;
XgBoost;
Gradient Boosted Trees)



Support Vector
Machine
(with different
kernels)



Artificial
Neural
Network

Experiments

- Sentiment analysis as alternative metrics

  ( vs 3+ scores;  vs 3- scores)

- Grouping by cuisines type

  (Italian vs American-Chinese; Fine Dining vs KFC)

- Spatial analysis

  (Downtown vs Countryside; Las Vegas vs Madison)

Understanding why restaurants fail is crucial for avoiding future failing

- Customers' opinion deserve to be better heard and analyzed
- Predictive models provide marketing researchers with new metric of customers' satisfactory level

All images are from Google Image