Don't Forget to Tip — It Means a Lot

Ava Rezvani
W266: Natural Language Processing
University of California, Berkeley — School of Information
ava.rezvani@ischool.berkeley.edu

ABSTRACT

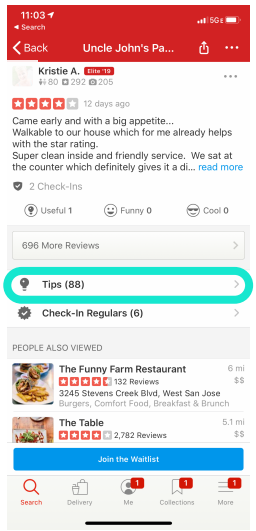
Yelp Tips data is a 500-character capped opportunity for users to provide key pieces of information, such as food items to order and happy hour tips, without needing to write a long review. While currently underutilized, I predict that Tips have the ability to supplement and enhance in the above-the-fold front page Yelp page for a restaurant. Using an LDA model, this paper attempts to unearth unique topics from Tips, attempting to do so with each individual Tip and also collectively for a restaurant. As document size increased, the ideal number of topics peaked 5, and with various increases in batch size and passes, the model's complexity increases while coherence score decreases. Topics that were outputted became more and more generic and not useful. Future work would be around manipulating and maximizing alpha and beta inputs for the model.

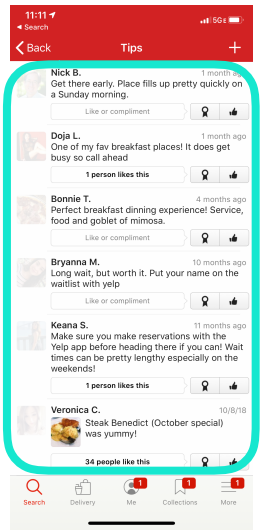
INTRODUCTION

Yelp is a crowd-sourced platform where users can leave reviews, ratings, and Tips for businesses. Founded in 2004 and used in primarily metropolitan areas, it has grown to have 192 millions reviews on its site [1].For a single business web page on yelp.com, the layout consists of many features, such as average star ratings, business contact information, user-uploaded photos, and user-curated reviews. Yet what are not shown on this page are Yelp Tips, which Yelp define as "a way to pass along some key information about a business -- such as the best time to go or your favorite dish -- without writing a full review about your experiences [2]."This feature can be found at the bottom of a businesses Yelp mobile-app page, after a user Tips on "Tips," where they are then given a list of 500 characters or less text written by other users in chronological order.

While many data scientists and researchers have focused on Yelp Review text analysis [12, I will be focusing on Yelp Tips text analysis. The Tips written by users provide useful and helpful information, that often overlap with other sections of the app such as the businesses specials, hours of operation, and "Explore the menu." See the sample of Yelp Tips below that address each of those sections, respectively:

"Pro Happy Hour Tip: burger, shot, beer for $15 - it's a steal."
"This place is closed until 27th Sept 2016"
"Pescatarians take note: while the day boat scallops are very good, the menu fails to mention that the dish includes chunks of bacon throughout and a big poached runny-yolk egg."





Would it make sense for a user to have to open the Yelp app on their phone, find the restaurant, scroll down to the bottom of the page, tap on "Tips" and peruse through the options to find that the restaurant is actually closed on the day that they wanted to visit? No — it does not seem like a great user experience.

My hypothesis is that, given Tip's low discoverability and usability, that Yelp currently believes that there is lower value in the Tips than there are in the reviews and has de-prioritized leveraging the rich data that users have left in there to supplement their current snapshots of businesses. Additionally, the differences between a review and a Tip is not as easily understood by most users — many users could be deterred from writing a review because of all of the fields and large body of text to fill (along with the possibility that the review gets deleted). Steering them towards Tips could be a better proposition to gathering feedback quickly from more people. The problem is that Tips have larger value despite how they are currently being utilized, and many users are unable to leave Tips nor read them. The information from them could play vital parts in enhancing the above-the-fold experience of a Yelp user, agnostic of the platform they are access it on (web or mobile).

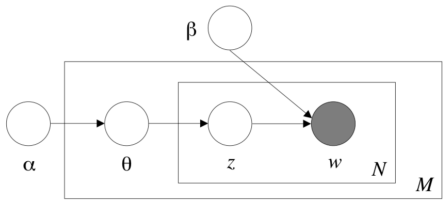
My primary approach to this will be with Latent Dirichlet allocation (LDA)[3] topic modeling, with the goal of deriving topics of Tips in order to enhance their usability and organizations as well as inform above-the-fold features of a Yelp business page.

BACKGROUND

The dataset for my analysis was originally acquired from http://www.yelp.com/dataset/challenge [3]and then replaced with data from Kaggle [4], as I issues with downloading it from the original link. The only differences between the two datasets was the Kaggle was from 2017, the former from 2018 — there was no additional clean-up of the data on the Kaggle platform.

While the dataset included many files (yelp_academic_dataset_business.json, yelp_academic_dataset_user.json, yelp_academic_dataset_Tip.json, yelp_academic_dataset_review.json, yelp_academic_dataset_checkin.json) it was important to identify and combine the business and tips datasets together.

When approaching this problem, it became clear that topic modeling was the best path forward, given that I am trying to understand the topics of the Tips in order to better leverage and organize them elsewhere in the Yelp experience. There are many models to explore that I considered with regards to semantic similarity, before deciding on Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA), both useful for unsupervised learning.



Graphical model of LDA. Outer rectangle is document, inner is repetition of topics and words within a document. [3]

For LSA, while one of the original topic modeling techniques, I knew that given my data set, there would be limitations due to the fact that there is a large weight on words that appear often in a document but not as often in the corpus.

Finally, there was Latent Dirichlet Allocation (LDA), which employs distributions over distributions, using the probability of the likelihood of seeing these distributions. While LDA is often used for documents or large bodies of text, I am interested in exploring its application to Yelp Tips, which can consist of up to 500 characters. Weng et al. in 2010 and Steinskog et. al in 2017 found it to be beneficial as a topic modeling technique applied to Twitter tweets, limited

to 140 characters, by pooling tweets into larger documents [6-7]. Interestingly, Naveed et al.'s work in 2011 strived to analyze the challenges of identifying topics in more sparse documents like microblogs (e.g. tweets) as well, applying LDA and introducing an "interestingness" as a static quality measure to combat sparsity [8]. My approach is that I am using different data and will not be pooling Tips responses to a particular user, given that the

METHODS

My first step in approaching this problem was cleaning up the data by converting the data from JSON to CSV, matching Tips to their respective Restaurants; removing punctuation, case and stop words; attempting stemming then deciding lemmatizing was better due to less loss of data; tokenizing the data; developing bigrams and trigrams; and conducting a word cloud exploration. Throughout the process I was exploring the data to understand its dimensions and potential for topic modeling.

I used the "gensim" library to develop and run experiments on my model [15]. To measure the quality of the learned topics, I will be calculating coherence scores to each model [9] to ensure the interpretability of the model topics, knowing that the there is more computation involved. Additionally, I will be interpreting the model perplexity [14].

RESULTS

Baseline

Coherence Score: *0.403906* — Model Complexity Score: *-8.499812*

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
best	good	try	great	time	like	love	sushi	breakfast	hour
ever	always	amazing	food	lunch	delicious	place	eat	taco	happy
place	pizza	menu	service	yummy	make	night	lot	everything	dont
awesome	burger	salad	good	wait	vega	clean	price	fun	get
favorite	fry	new	place	come	location	slow	big	never	drink
chicken	really	special	friendly	dinner	bettter	vegan	going	loved	order
one	wing	must	staff	first	sure	bomb	wine	like	open
roll	pretty	check	nice	closed	try	recommend	option	cool	even
ive	bbq	yum	excellent	day	tase	experience	healthy	bread	bar
town	small	beer	back	ever	sauce	chip	buffet	wonderful	pm

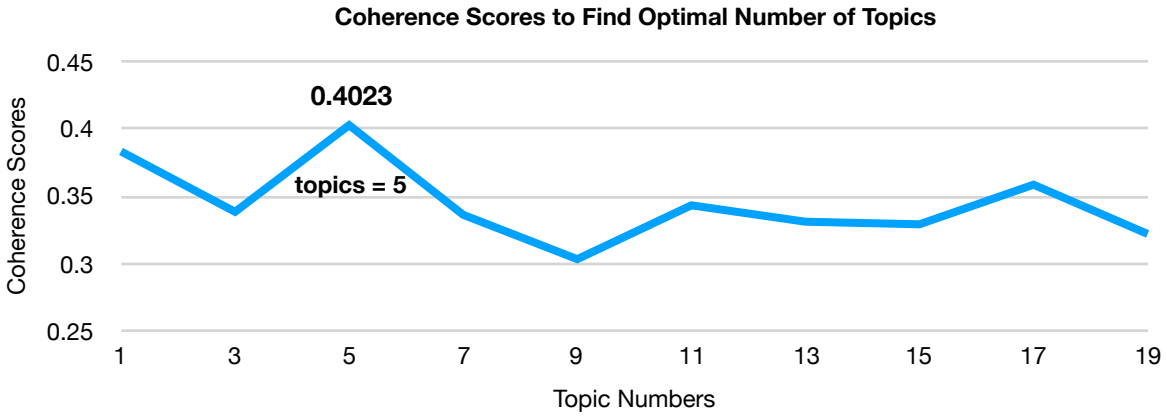
By reading through the topics and choices of words, it becomes apparent that certain topics. Topic 3 could be understood as new and popular dishes, 4 could be seen as a service topic, 10 could be regarding bars and happy hours.

Experiment 1 Large Document Formation

Coherence Score: *0.352937* — Model Complexity Score: *-7.832160*

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
taco	food	wing	food	pizza	hour	beer	great	burger	breakfast
food	good	tea	great	great	happy	great	place	fry	brunch
good	sushi	boba	service	best	great	selection	love	good	great
great	great	gelato	good	good	drink	wing	coffee	great	wait
best	place	milk	place	italian	place	tap	good	food	get
burrito	best	flavor	best	place	bar	good	best	get	good
mexican	service	slush	love	love	good	bar	food	place	food
salsa	roll	honey	amazing	sandwich	night	food	delicious	service	amazing
place	love	hot	awesome	slice	service	place	try	sandwich	waffle
margarita	lunch	shaved_ice	back	get	time	draft	sandwich	order	pancake

Once combined into larger documents per restaurant, I notice that the topics change from types of Tips feedback to types of the restaurants/establishments. Topic 1 is representative of a restaurant that serves Mexican food, 2 is Japanese food, 3 is drinks and desserts, 5 is Italian, 5 is a bar, 8 is a cafe, 9 is a burger shop, and 10 is a place that serves breakfast items.



Experiment 2 Large Document Formation - Topics = 5

Coherence Score: *0.315856* — Model Complexity Score: *-7.463464*

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
breakfast	pizza	beer	great	food
great	wing	great	food	great
place	best	good	service	good
good	slice	place	sushi	service
coffee	great	food	good	place
get	italian	bar	place	best
love	donut	wing	happy	love
sandwich	crust	drink	hour	chicken
service	pie	time	amazing	delicious
burger	cheese	night	best	amazing

Coherence scores and complexity scores are getting worse and the adjectives “great” and “good” are showing up in mulTiple places, along with nouns such as “food.” All of the topics are starting to look the same.

DISCUSSION

Once I had my cleaned and prepared data, I did a baseline attempt with an LDA model where I saved my corpus as my bigrams and applied my LDA model — the result was a Coherence Score of 0.403906 and Model Complexity Score of -8.499812.

After I ran my initial experiment and investigated the topics that were arising along with the scores, I began to question if LDA was truly the best model for my approach (see Methods). After reading several papers where it had been applied for smaller sizes of documents [6-8], I chose to investigate pooling the Tips per a restaurant establishment.

After combining the Tips for each restaurant into larger documents, I then re-ran the model to see how it performed. The coherence score was lower and complexity score higher, which made me want to look into the topic numbers. After viewing that 5 topics provided the highest coherence score (0.4023), I start optimizing the model around that, tweaking the chunk size for training as well as the number of passes to run the LDA model through. My hypothesis was that as I increased the chunk size and passes, I would be trading model run-time for a higher coherence and better complexity score. The opposite occurred — both numbers became worse (0.284794, -7.433212 are the respective scores where I halted the investigation).

As I noted in the Results section the words listed per topic in one of the final models is now a mix of adjectives and generic nouns that offer no substantial topics of use for the Yelp feature. With Experiment 1 at least, by concatenating the Tips into documents for each restaurant, the topics that did arise had more coherence but around types of restaurants and foods.

It seems like in our case the baseline model has the most useful results and metrics of success. I was wrong to assume that the LDA model could have been more useful had it been applied to a larger corpus.

CONCLUSION & FUTURE WORK

My conclusions include that Tips data alone may not be robust enough in its current state to provide information that reviews and current Yelp information already has. When Tips were combine for establishments, the topics were non-revelatory.

In order to improve our topic model coherence and complexity, a next step would be to investigate and identify the optimal alpha and beta values, which show document-topic density and word-topic density respectively. Additionally, as mentioned in the Discussion section, more work could have been done by going back to the Baseline model, after having disproved the assumption that the document sizes needed to be larger. I would explore the options of using Tips alone as a 500 character max document.

References

1. https://en.wikipedia.org/wiki/Yelp
2. https://www.yelp-support.com/article/What-are-Tips?l=en_US
3. D. M. Blei, A. Y. Ng, M. I. Jordan (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research 3 3 (2003) 993-1022.
4. https://www.yelp.com/dataset/challenge
5. https://www.kaggle.com/yelp-dataset/yelp-dataset
6. Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He. TwitterRank: finding topic-sensitive influential twitterers. Proceedings of the third ACM international conference of Web search and data mining. (2019) 261-270.
7. A Ottessen Steinskog, J. Foyn Therkelsen, B. Gambäck. Twitter Topic Modeling by Tweet Aggregation. Proceedings of the 21st Nordic Conference of Computational Linguistics, (2017) 77-86.
8. N. Naveed, T. Gottron, J. Kunegis, A. Che Alhadi. Searching microblogs: Coping with sparsity and document quality. Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM (2011).
9. M. Röder, Andreas B., Alexander H. Exploring the Space of Topic Coherence Measures. WSDM (2015).
10. S. Qaiser, R. Ali. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications 181 1 (2018).
11. J. Ramos. Using TF-IDF to Determine Word Relevance in Document Queries. Department of Computer Science, Rutgers University (2002).
12. R. Suchdev. Subtopics in Yelp Reviews. Master's Theses and Graduate Research, San Jose State University (2018)
13. D. M. Blei. Introduction to Probabilistic Topic Models
14. W. Buntine. Estimating Likelihoods for Topic Models. Lecture Notes in Computer Science (LNCS) Book Series 5828 51-64 (2009).
15. https://radimrehurek.com/gensim/