Measuring the value of an online-review

As we know, online-review plays an important role in e-commerce for they have a major impact on the users' purchase decisions. However, instead of helping the customers better understanding the features of the products, the increasing number of reviews make more bias and misunderstanding. In order to solve this problem, most of the e-commerce industries have their ranking and filtering system for users to extract the reviews they want. But the extant ranking system can't represent the value of the reviews and could be easy to manipulated by attackers. Our study focus on finding a method to explore the real quality of a review by using text mining, which not only reflects the value of the reviews, but also be non-vulnerable.

General Terms
Algorithm, Experimentation, text mining

Keywords
Review quality, labels, sentiment, CNN

1. Instruction

In the world of abundant online-reviews, one of the main challenges is to identify the high-quality ones from others. Nevertheless, because of the diversity of natural language, it's difficult to compute the quality directly by the review contents. Instead, the common practice in e-commerce industries is to rank by the review update time or the review helpfulness. Unfortunately, due to the limitation of these ranking methods [4], it's hard to reflect the real value of a review[3]. Thanks to the development of text mining, we discover a method to reflect the quality of a review. In the remainder of this report, we will talk about why we use this method, what this method is, and how this method applies.

2. Motivation and objectives

Customers will make a better purchase decision depends highly on the real and objective reviews of a product. Sellers hope to have real and detailed feedbacks for the reviews are their word of mouth. The e-commerce platforms will gain increasing numbers of active users through the help of high quality reviews. Therefore, this is a triple win proposal. So our objective is to identify the real and objective reviews.

3. The value of a review and the measurement

The value of a review is an abstract concept. Typically, measurement around reviews is made of a mix of Qualitative and Quantitative data. It's impossible to simply say that 'A review is worth $X.XX' because all manner of things from tone to language to placement will affect the value of the review.

Generally speaking, a review has two kinds of value. One comes from the review itself, such as the update time or the level of the reviewer. The other one is given by individuals, such as how helpful the review is or the number of the comments. The measures of a single value above is so simple and straightforward. For example, measuring by update time could be easy because you can directly get the update time from the application. But these common practices have their limitations.[4] Therefore, we develop a new measurement which can reflect the two values at the same time and overcome the shortcomings.

Our approach is to analyze the amount of information in a review. In our consideration, more information, higher value is of a review. For example, “the steak is crispy and juicy” is better than “the food is good here”, because the first sentence has more information about how good the food is. And we found that more information is actually more detailed and more objective.

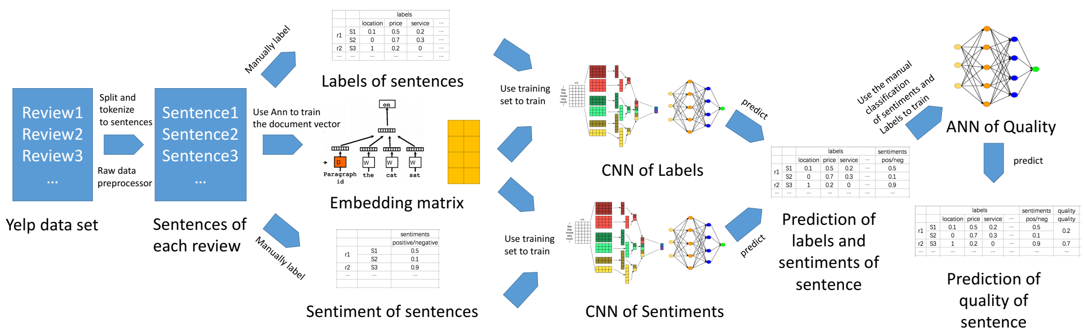
To represent the details of a sentence, we firstly create some labels, such as food, service, price etc. And then we manually score base on the relationship between a sentence and the labels we have. For example, we have 6 labels: amenities, environment, food, location, price and service. [0.5,0,0,0,0,0] means the sentence is about 50% probability related to amenities. What’s more, in the situation of sentiment, we use a constant value between 0~1 to represent how objective the review is, such as [0.8]. Finally, in order to represent the value of a review, we compute the matrix product of these two vectors . The higher the value is, the higher quality the review has.

4. The limitation of common practice in review ranking

Reviews are always ranked by update time and user's votes. But both ranking methods have their limitations. Firstly, ranking by update time doesn't reveal how people think of the review. In other words, it doesn't show how popular the review is and how people like the review. Secondly, ranking by user's vote, such as rank by review helpfulness or the number of comments, is vulnerable in that the ranking position can be manipulated. And what's more, the popular review will become more popular, because of its high ranking. To break these limitation, we introduce a new ranking method – rank by the value/quality of a review [3], which not only inherit the advantage of common ranking method, but also give up the shortcomings. Next, we will present the details of implementation of our ranking method.

5. Methodology

The main challenge of our study is to quantificat the information of a review. First of all, we need to quantificat the review content into an embedding matrix [5.2]. Then we need to make regression to change the matrix into a vector of labels and a vector of sentiment by using Convolution Neural Network [5.3,5.4]. At last, we apply a formula to compute the quality of the review.

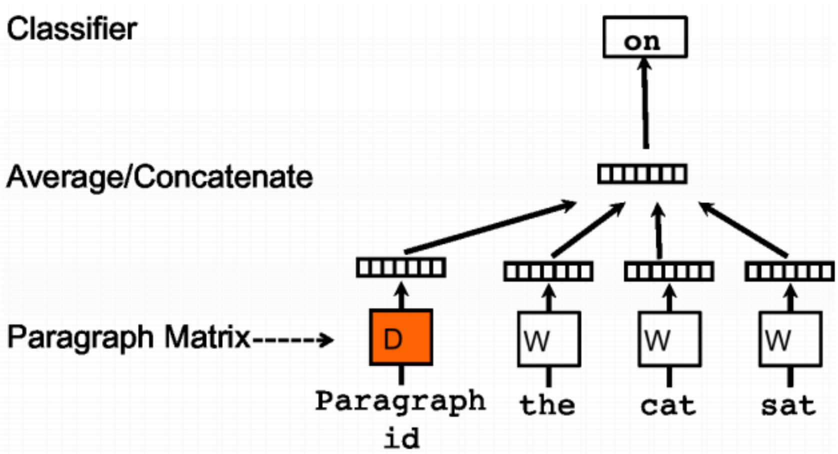


5.1 Data Set

We use a subset of the yelp official data set in this study, please refer to the attachment named word_sample.json. The Yelp dataset is a subset of their businesses, reviews, and user data for use in personal, educational, and academic purposes. It contains 4,700,000 reviews about business attributes like hours, parking, availability, and ambience in over 12 metropolitan areas.

5.2 Document Vector

In our case, we need to represent the review text input as a fixed-length vector. When it comes to texts, one of the most common fixed-length features is bag-of-words (Harris, 1954). Despite their popularity, bag-of-words features have two major weaknesses: they lose the ordering of the words and they also ignore semantics of the words. To overcome the weaknesses of bag-of-words, a concept called Paragraph Vector is proposed (Tomas, 2014), which is an unsupervised algorithm that learns fixed-length feature representations from variable-length pieces of texts. The algorithm represents each document by a dense vector which is trained to predict words in the document.



In our experiment, we split each review into sentences. Each sentence is a document/paragraph. Every paragraph is mapped to a unique vector, represented by a column in matrix D and every word is also mapped to a unique vector, represented by a column in matrix W. The paragraph vector and word vectors are averaged or concatenated to predict the next word in a context. After doing that, we transfer each sentences into the fixed-length vector.

5.3 Measuring details of a review

In order to represent that a sentence is relative to some particular areas, the common practice is to manually label the sentences with tags. For example, “The steak is juicy and crispy” can be labelled as “food”. “The sweet potato is cost $10” can be labelled as “price”. When tagging the sentence, we create a vector base on the list of total tags. In the beginning, all the initial numbers in the vector are 0. If the sentence hit one of the tags, the relative element of vector becomes 1. For instance, if we have total 6 tags: amenities, environment, food, location, price, and service. Given the sentence “The steak is juicy and crispy”, the output vector would be [0,0,1,0,0,0]. These output vector of training set is used to predict the next labels. Comparing the performance with other methodologies [6], we use Convolution Neural Network(“CNN”) as our model to train the data. In addition, instead of generating binary values to the vector, we generate a constant value between 0 and 1 to represent the relationship between the sentence and the particular areas. The value is closer to 1 if there are more details in that particular area. Besides, we still use CNN as the regression model.

		labels			
		location	price	service	···
r1	S1	0.1	0.5	0.2	···
	S2	0	0.7	0.3	···
r2	S3	1	0.2	0	···
···	···	···	···	···	···

5.4 Measuring sentiment of a review

Traditionally, the sentiment in a review can be positive, negative or neutral. And we only want to know how objective the review is. So we can mark neutral as 1 and positive/negative as 0. But it is still hard to do the agreement because of the bias of reviews. Therefore, we generate a constant value between 0 and 1 to represent the level of objectivity. More neutral the sentence is, closer to 1 the value is. We train another CNN to predict the sentiment of the next review based on the training data we labeled. Finally, we can predict the level of a review being objective.

		sentiments
		positive/negative
r1	S1	0.5
	S2	0.1
r2	S3	0.9
…	…	…

5.5 Calculating the quality of a review

Intuitively, the rank of the quality of the review can be: objective reviews with details > subjective reviews with details >= objective reviews without details > subjective reviews without details. For example, "the beef steak is crispy and juicy" > "the beef steak is delicious" >= "I am a meatatarian" > "I hate the food here". Therefore, we use a formula to represent the quality Q of a review: Q = (S*L) / n = $(\sum_{i=1}^{n}\sum_{j=1}^{m} S_j * L_{ij}) / n$ = ((S1*L11+S1*L12+…+S1*L1m)+(S2*L21+S2*L22+…+S2*L2m)+(Sn*Ln1+Sn*Ln2+…+Sn*Lnm))/n. S refers to the vector of the sentiment[5.4]. L points to the vector of labels[5.3]. n means there are n sentences in the review. m is the total number of the tags.

review id	review content	labels	sentiment	quality
7M2GCelba1uTJMJbVO7TKw	I really don't understand how anyone can eat the food from here.	0,0,0,0,0,0	0.3	0.06
7M2GCelba1uTJMJbVO7TKw	Granted, I am Chinese and enjoy authentic food, however I like good Chinese American food.	0,0,0,0,0,0	1	0.06
7M2GCelba1uTJMJbVO7TKw	For someone who usually doesn't leave food on her plate unless it is really bad, I could not finish my meal because it was tasteless and below average.	0,0,0.3,0,0,0	0.6	0.06
ByRzJ8rF2KJWLr-cUNU6EA	This place is horrible, we were so excited to try it since I got a gift card for my birthday.	0,0,0,0,0,0	0.2	0.0875
ByRzJ8rF2KJWLr-cUNU6EA	We went in an ordered are whole meal and they did not except are gift card, because their system was down.	0.3,0,0,0,0,0	0.9	0.0875
ByRzJ8rF2KJWLr-cUNU6EA	Unacceptable, this would have been so helpful if we would have known this prior!	0,0,0,0,0,0.2	0.4	0.0875
ByRzJ8rF2KJWLr-cUNU6EA	!	0,0,0,0,0,0	0.9	0.0875

review id	review content	labels	sentiment	quality
IlmdwOgDReucVoWEry61Lw	Location is everything and this hotel has it!	0,0,0,0.3,0,0	0.6	0.786
IlmdwOgDReucVoWEry61Lw	The reception is inviting and open 24 hours.	0.2,0,0,0,0,0.8	1	0.786
IlmdwOgDReucVoWEry61Lw	They are very helpful and have a lot of patience answering all my questions about where to go etc.	0,0,0,0,0,1	0.7	0.786
IlmdwOgDReucVoWEry61Lw	there is also a lounge open 24 hours with snack-type food.	1,0,0,0,0,0	1	0.786
IlmdwOgDReucVoWEry61Lw	Breakfast is continental-style so if you want heartier fare look elsewhere though you don't have to go far.	0,0,1,0,0,0	1	0.786
IlmdwOgDReucVoWEry61Lw	The bus and train stations are right across the street so it's easy access to the airport or anywhere else you may want to go.	0,0,0,1,0,0	0.9	0.786
IlmdwOgDReucVoWEry61Lw	Turn uphill to old town or cross the bridge to new town.	0,0,0,1,0,0	1	0.786
IlmdwOgDReucVoWEry61Lw	The room with a view i got was spacious and comfortable though it's a bit of a maze to find it-just follow the signs.	0.3,0.8,0,0,0,0	0.7	0.786
IlmdwOgDReucVoWEry61Lw	The windows are double paned so the room is quiet plus i was on the 5th floor which helps.	0.8,0.3,0,0,0,0	1	0.786
IlmdwOgDReucVoWEry61Lw	It's a bit pricey but still one of the best values i found!	0,0,0,0,0.3,0	0.7	0.786

low quality review vs high quality review

6. Evaluation Comparison

Label CNN Performance

```
{
  "mean_squared_error": 0.028211111177007368
}
```

Sentiment CNN Performance

```
{
  "mean_squared_error": 0.05753333207766245
}
```

Label sklearn.multioutput Performance

```
{
  "mean_squared_error": 0.03766666666666663
}
```

Sentiment sklearn.multioutput Performance

```
{
  "mean_squared_error": 0.06753333333333332
}
```

We use 350 training data and 150 validation data to train the model. The result shows the performance of CNN is better than the performance of multiclass regression in sklearn in our case.

7. Deploying

7.1 Install the dependencies

python 2.7

sklearn

matplotlib

keras

gensim

nodejs (admin)

npm (admin)

word_sample.json

data_sample2.csv

7.2 Training the samples

1. open jupyter
2. modify all retrain=0 to retrain=1 on the bottom of ReviewAnalyser.ipynb
3. run the cells

7.3 Start Restful API server

python RestfulAPI.py

7.4 Run the backend program

1. cd admin
2. npm install
3. npm start

7.5 Access the admin system

Access http://localhost:3001 in the browser

Reference

Theodoros Lappas, Gaurav Sabnis, Georgios Valkanas The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry

Quoc Le, Tomas Mikolov Distributed Representations of Sentences and Documents

Rob Thomas What is the value of an online review?

Amazon official website https://www.amazon.com/review/guidelines/top-reviewers.html/