Restaurant Survival Analysis with Heterogeneous Information

Jianxun Lian[*]
University of Science and
Technology of China
Hefei, China
lianjx@mail.ustc.edu.cn

Fuzheng Zhang
Microsoft Research
Beijing, China
fuzzhang@microsoft.com

Xing Xie
Microsoft Research
Beijing, China
xingx@microsoft.com

Guangzhong Sun
University of Science and
Technology of China
Hefei, China
gzsun@ustc.edu.cn

ABSTRACT

For shopkeepers, one of their biggest common concerns is whether their business will thrive or fail in the future. With the development of new ways to collect business data, it is possible to leverage multiple domains' knowledge to build an intelligent model for business assessment. In this paper, we discuss what the potential indicators are for the long-term survival of a physical store. To this end, we study factors from four pillars: geography, user mobility, user rating, and review text. We start by exploring the impact of geographic features, which describe the location environment of the retailer store. The location and nearby places play an important role in the popularity of the shop, and usually less competitiveness and more heterogeneity is better. Then we study user mobility. It can be viewed as supplementary to the geographical placement, showing how the location can attract users from anywhere. Another important factor is how the shop can serve and satisfy users. We find that restaurant survival prediction is a hard task that can not be solved simply using consumers' ratings or sentiment metrics. Compared with conclusive and well-formatted ratings, the various review words provide more insight of the shop and deserve in-depth mining. We adopt several language models to fully explore the textual message. Comprehensive experiments demonstrate that review text indeed have the strongest predictive power. We further compare different cities' models and find the conclusions are highly consistent. Although we focus on the class of restaurant in this paper, the method can be easily extended to other shop categories.

[*]This work was done when Jianxun Lian was an intern at Microsoft Research, Beijing, P.R. China.

Keywords

restaurant survival analysis; location-based services; data mining

1. INTRODUCTION

How business will thrive in the future is an important concern for all shopkeepers. Knowing about long-term trends, shopkeepers can take corresponding actions in advance. For instance, if they know the store will come to a crisis in a matter of months, they could take steps to avoid the misfortune such as to make changes on the style of the store, or even consider choosing a new placement to minimize economic losses. Usually store owners make long-term decisions based on empirical judgement. Due to limited data sources and lack of analytic tools, it is traditionally a challenge to make data-driven decisions.

Bankruptcy prediction is a common topic in management and finance literature. However, existing studies [31, 22, 25, 11, 20, 21, 13] are usually limited to the analysis of financial factors, such as *liquidity, solvency, and profitability*. In addition, the data-set used is usually small due to the challenge of obtaining financial data. With the development of information techniques, especially the growth of online location-based services, a large amount of business related data can be collected through the Internet. For example, people may post check-ins at some point of interest (POI) they are visiting; after consuming in a shop, they can write reviews on Yelp to show how they like the shop. Thus there is a potential to exploit heterogeneous information to build automatic business intelligence tools for enhancing the decision process. In this paper, we take advantage of both geographic analysis as well as user behavior analysis to study whether a physical store will close down in the future. Specifically, we explore various factors under the guidance of the following considerations:(H1) the geographical placement of the store play an important role in the store's operation; (H2) people's offline mobility patterns to the store as well as its nearby places influence the business; (H3) user's rating scores (e.g., on Yelp) are explicit evaluations of the store from the customers' point of view; (H4) besides well-formatted rating scores, review words contain more rich information which a simple numeric score does not cover.

Recent works have studied the economic impact of geographic and user mobility factors on the retailer store[19][12]. As formulated in these works, geographic signs contain the types and density of nearby places, and user mobility includes transitions between venues or the incoming flow of mobile users from distant areas. Inspired by them, we first analyze these two types of features. In fact user mobility features are quite correlated with geographical features, because the former reflects spatial character in terms of human popularity. We further bring in users' review as an important data source, including rating scores and review text. After consumption at a store, users can rate scores for the store on multiple aspects, such as environment, cost and taste on platforms like Yelp and Dianping. These numeric values summarize customers' overall opinion, but they are not as detailed as the words in the review. Through experiments we find that words provide more predictive power than the simple numeric values.

To make the work more specific and accurate, we focus on the shop category of restaurants. The main data source we use is from the Chinese website *www.dianping.com*, on which the overwhelming majority of venues are food service ones. Additionally, China has a variety of restaurant categories, for example, Cantonese restaurants, Szechuan restaurants, and Shaanxi noodle restaurants. Thus as a specific type of store, restaurants are interesting in diversity and worth in-dep mining. What is more, the analysis method is generalized and can be easily applied to other types of stores.

The contributions of this paper are summarized as follows:

- Different from traditional bankruptcy studies which usually focus on financial variables, we propose an approach to conduct restaurant survival analysis from exogenous factors which can be obtained from big data over the Internet. Note that our purpose is not to provide better accuracy than existing models which make full use of financial factors. The primary advantages are that our approach covers various angles from heterogeneous data and as well is scalable to large number of restaurant samples.

- We provide an in-depth analysis on geography, mobility, and user opinions. We demonstrate what are the relatively stronger predictors. For example, *neighbor entropy* turns out to be the best predictor from the perspective of geographic; users' textual message are far more important than their numeric ratings; restaurants which offer attractive group purchases but serve poor food have a higher probability to close; and restaurants holding core competitiveness (*time-honored brand, well-deserved reputation; crowded consumers, state-run*, etc.) tend to survive, which is in sync with common sense.

- We conduct comprehensive experiments on three different cities, and find that the conclusions are quite consistent. Meanwhile, integrating all the predictors can lead to the best accurate model, which demonstrates the necessary of including feature variety.

The rest of this paper is organized as follows. In Section 2 we describe the essential information about our dataset. In Section 3 we define and analyze the geographical features. In Section 4, we give an analysis of user mobility. In Section 5, we study online rating scores and exploit various methods to mine review text. After this we provide experiment details on combining different models and on different cities. In Section 7 we summarize the related works. Finally we give the conclusion in Section 8.

2. DATA AND PROBLEM STATEMENT

In this section we first provide some essential information about the dataset used, including the data collection process and the basic statistics of the collected data. Then we raise the restaurant survival prediction problem and show the performance of a naive solution.

2.1 Data Collection

The main data source we use in this paper is *Dianping.com*. Dianping, known as "Yelp for China", is the largest consumer review site in China. It offers multi-level knowledge through its diverse functions such as reviews, check-ins, and POI meta data (including geographical message and shop attributes). We use the *LifeSpec* data crawling platform [32] to retrieve all data related to the shop (from the shop's open time to our crawling time). Specifically, for each shop we crawl :

(1) the meta information, including name, location (city, latitude, longitude, and detailed address), category, and price;

(2) all the reviews written by consumers. A review is comprised of review words and 5 scores, including overall rating, taste, environment, service, and price.

(3) all the check-ins posted by users.

All the data we have crawled is publicly available on the website. The data crawling process finished in April 2014.

In the literature of churn analysis, a user is usually defined as a *churner* if he/she does not have any data during the last several periods of the dataset. However, some shops may not be popular online often resulting in receiving no reviews or check-ins for a long period, say several months. Therefore, it is not proper to define shop failure based on the review or check-in numbers across a period. Fortunately, we find that Dianping has an API to query the status of s shop. In general, all statuses can be grouped into four categories: (1) normal shop, which means the shop is still operating; (2) closed shop, meaning the shop has already closed down; (3) suspended shop, meaning the business is suspended for a certain time. The reason for suspension is various and unaware, and the shop may or may not reopen; (4) others, including a few special cases such as unqualified shops and applicative shops. We crawl shops' status at March 2016 and use the shop status as the label.

2.2 Basic Statistics

Our entire Dianping dataset captures the period ranging from April 2003 (when dianping.com was established) to June 2014 (when we finished crawling content data), as well as the shops' snapshot status at March 2016. Considering that spatial context may change over such a long time, for restaurant analysis we focus on a certain year and assume that geographical placement will not change a lot within one year. We decide to use the 2012 data because we have the most abundant data for this year. The basic statistics for
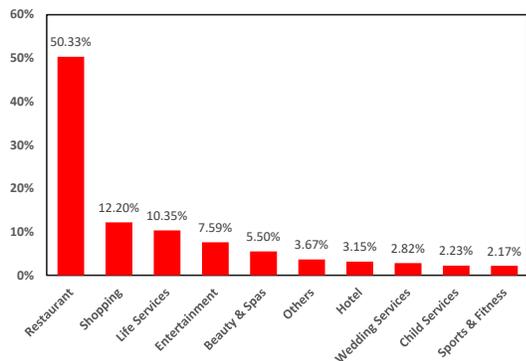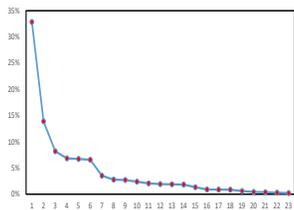
Figure 1: Shop categories and their percentage.

Table 1: Basic statistics of Dianping dataset for year 2012

#check-ins	#reviews	#cities	#shops
9,270,299	4,576,587	349	409,602

Table 2: Basic statistics of dataset for Shanghai, Beijing, and Guangzhou

City	#check-ins	#reviews	#shops
Shanghai	4,027,503	1,980,914	76,190
Beijing	1,710,396	826,772	50,917
Guangzhou	261,273	156,844	17,747

(a) The distribution of 23 subcategories for restaurants in Beijing. The highest point is snack bar, which accounts for 32.8%.
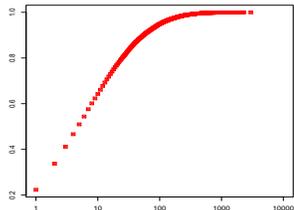
(b) The cumulative distribution function of the review count for each shop.

Figure 2: The restaurant subcategories distribution and the cumulative distribution function of the review count per shop.

2012 are listed in Table 1[1]. Among the 349 cities in China, Shanghai, Beijing, Guangzhou are the most popular cities in our dataset, and their statistics are shown in Table 2. As we can see, the three cities account for 64.8% of the total review count, 64.7% of the total check-ins, and 35.4% of the shop in amount.

Rather than building a unified model for all shops, in this paper we limit the study to restaurants. Restaurants is not only the largest shop category in quantity in our dataset, but also has the biggest number of subcategories. Figure 1 shows that half of the shops in our dataset belong to the restaurant category. There are 23 subcategories for restaurants and the distribution is shown in Figure 2a, from which we can observe that the biggest subcategory is snack bar.

---

[1] We only count shops which have at least one review from 2012.

Table 3: Statistics of restaurants which received no less than 10 reviews in 2012. Early closed restaurants(closed before the end of 2012) are removed. We use this set for training and testing.

	Shanghai	Beijing	Guangzhou
#restaurants	12,990	8,615	2,325
closed ratio	37.4%	28.4%	30.8%

In Figure 2b we plot the Cumulative Distribution Function (CDF) of the review number a shop has. As can be observed, 37.7% of the shops have no less than 10 reviews. Since we are mining users' online opinions, in the prediction task we focus on the shops which have no less that 10 reviews in order to ensure enough data to build the model. Next we plot the shop status distribution. As shown in Figure 3, the ratio of closed status in restaurant group is as high as 28.6%, which is significantly higher than that of non-restaurants. It indicates that restaurants differ from other types of shops not only in quantities or diversity but also in stability, warranting its own in-depth study. Finally, we remove the restaurants which closed before the end of 2012 from our learning set, since long term prediction is not relevant to them. The final statistics of the learning set are listed in Table 3. We randomly split the dataset into training set (70%), validation set (15%) and test set (15%) in the experiments.

2.3 Problem Statement

The restaurant survival prediction problem can be stated as follows: given the heterogeneous data (geographical information, user mobility data, online scores and review text) in 2012, we want to predict whether the restaurant will close down before March 2016. We use the restaurants which belong to *normal shops* or *closed shops* categories for study.

The first thing that comes into mind is that consumers' satisfaction may influence the future of a restaurant. This leads us to ask, can the task of restaurant survival prediction be solved simply using review scores and consumers' sentiment data? To verify, we use SnowNLP [2] to conduct a sentiment analysis on the review text. For each review of the shop, we can get a sentiment score $s(r) \in [0, 1]$, with 0 indicating negative and 1 indicating positive. We calculate the average/minimum/maximum sentiment score for each restaurant, and use these three scores as features to build a logistic regression model. The AUC is 0.52, which is just slightly better than a random guess. Similarly, we design several features based on consumers' rating scores. The AUC is 0.6136, which is not satisfactory. Now we ask: (1) Can we build a more accurate model for restaurant survival prediction? (2) What factors highly correlate with the future of restaurants?

3. GEOGRAPHICAL MODEL

We expect that a shop's business is to some extent dependent on its location. Motivated by [19, 12], we design spatial metrics and study their predictive power. When studying the performance of these metrics in our scenario, we use Beijing's data for illustration. Later we will compare the performance on different cities in Section 6.

Formally, we denote the set of all the shops in a city as **S**.

---

[2] https://github.com/isnowfy/snownlp
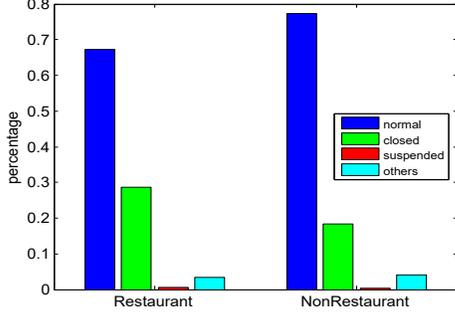
Figure 3: The status in March 2016 of restaurants and non-restaurants which are alive in 2012.

For each restaurant r, its neighbor set denoted by $N(r) = \{s \in \mathbf{S} : distance(r, s) \leq d\}$ are defined as all shops that lie within a d meter radius of it, and in the experiment we empirically set d to 500 meters. The category of a shop is denoted by $\gamma(r)$, and the entire category set by $\mathbf{\Gamma}$.

3.1 Predictors

Density: Although the prediction objects are restaurants, in predictor design, we consider all types of shops when studying context. Density is calculated as the number of shops in the restaurant's neighborhood. It is an indicator of the popularity around the restaurant:

$$f_r^D = |N(r)| \tag{1}$$

Neighbor Entropy: We refer to $N_\gamma(r)$ as the number of shops of cateogry γ near the restaurant r. Neighbor entropy metrics is defined as:

$$f_r^{NE} = -\sum_{\gamma \in \Gamma} \frac{N_\gamma(r)}{N(r)} \times log \frac{N_\gamma(r)}{N(r)} \tag{2}$$

A high entropy value means more diversity in terms of facilities within the area around the shop. A small entropy value indicates that the functionality of the area is biased towards a few specific categories, e.g. working area or residential area.

Competitiveness: Restaurants may have different cuisine styles and people may have different dinning preferences [33]. We assume that most competition comes from the nearby restaurants with the same category. For measuring competitiveness we count the proportion of the neighbors of the same category $C(r)$:

$$f_r^{Com} = \frac{N_{\gamma(r)}(r)}{N(r)} \tag{3}$$

Quality by Jensen: These metrics encode the spatial interactions between different place categories. It is first defined by Jensen et al.[18], and Dmytro et al.[19] use the inter-category coefficients to weight the desirability of the store for location choosing. Formally, we have:

$$f_r^{QJ} = \sum_{\beta \in \mathbf{\Gamma}} log(\kappa_{\beta \to \gamma(r)}) \times (N_\beta(r) - \overline{N_\beta(r)}) \tag{4}$$

where $\overline{N_\beta(r)}$ means how many shops of category β are observed on average around the shop of type $\gamma(r)$. And the

inter-type attractiveness coefficient $\kappa_{\beta \to \gamma(r)}$ is defined as:

$$\kappa_{\beta \to \gamma(r)} = \frac{N - N_\beta}{N_\beta \times N_{\gamma(r)}} \sum_{p:\gamma(p)=\beta} \frac{N_{\gamma(r)}(p)}{N(p) - N_{\gamma(p)}(p)} \tag{5}$$

Category Demand: Inspired by the Qualify by Jensen metrics, we propose a simplified category attractiveness measure, which is named with *category demand*:

$$f_r^{CD} = \frac{1}{N_{\gamma(r)}(r)} \frac{N_\beta(r)}{N(r)} \times \overline{N_{\gamma(r)}(\beta)} \tag{6}$$

where $\overline{N_{\gamma(r)}(\beta)}$ denotes how many shops of category $\gamma(r)$ are observed on average around the shops of type β. Basically, category demand is the ratio between the expected number of shops of category $\gamma(r)$ in its location and the real number of shops of category $\gamma(r)$. When f_r^{CD} is larger than 1.0, the shop meets the requirements of the position and is supposed to have a good business.

3.2 Results

We use logistic regression to perform binary classification based on the above predictors. We evaluate the performance in terms of area under the (ROC) curve (AUC)[6] because it is not influenced by the unbalanced instances problem. Figure 4a presents the AUC results. For individual features, neighbor entropy has better performance than the other three, which indicates that the heterogeneity of the nearby area play a more important role among the geographic attributes. Density, competitiveness, and quality of Jensen show similar levels of predictive power. Combining all the features could lead to a significantly(t-test p-value:0.026) better AUC than the best individual feature.

4. MOBILITY ANALYSIS

[19] find that deeper insight into human mobility patterns helps improve local business analytics. People's mobility can directly reflect a place's popularity. While various kinds of data source can be employed to mine mobility patterns, e.g., taxi trajectories and bus records as used in [10], we use check-ins to represent human mobility as [19] does. A check-in can be represented as a triple, $< u, s, t >$, containing user's id, shop's id, and event timestamp. To clean up the data, firstly we remove outliers by filtering out users who post check-ins too frequently(e.g., more than 100 times per day), and deleting successive check-ins in a very short period(e.g. in 1 minutes) or at the same place. The density distribution is shown in Figure 5b. As can be observed, check-in's distribution greatly coincides with shop's distribution (Figure 5a). There are four main concentrated regions which are considered the most flourishing areas in Beijing: Zhongguancun, China National Trade Center, Xidan Commercial Street, and Wangfujing Street.
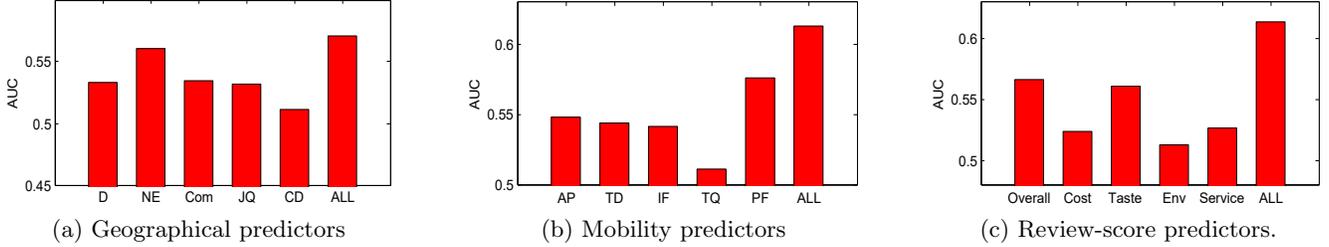
4.1 Predictors

Area Popularity: We use two values for mobility popularity: the number of total check-ins to the shop, and the check-ins near the shop:

$$f_r^{AP_2} = |\{< u, s, t > \in CI : distance(s, r) \leq d\}| \tag{7}$$

where CI denotes the entire check-in set.

Transition Density: We define a user transition as happening when a user posts two consecutive check-ins (c_i, c_j) within 24 hours, and denote the entire set of transition set

(a) Geographical predictors

(b) Mobility predictors

(c) Review-score predictors.

Figure 4: AUC performance comparison for individual predictors of different groups. In all the three charts we observe the best performance when combining all individual predictors. The classifier used is logistic regression.

(a) Shops distributions

(b) Check-ins distribution
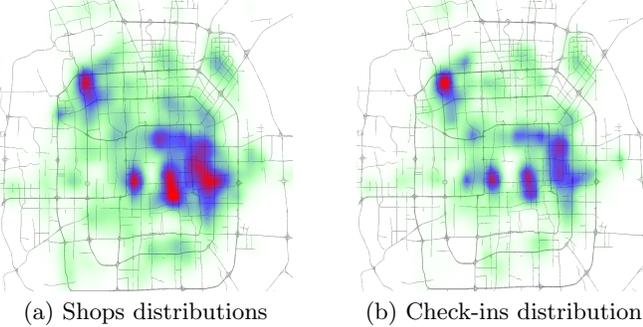
Figure 5: Heat maps of shops' and check-ins' density distribution in Beijing.

as Ts. Then transition density is defined as the number of transitions whose start and end location are both near shop r:

$$f_r^{TD} = |\{(c_i, c_j) \in Ts \ : distance(s_{c_i}, r) \le d$$
$$\&\& \ distance(s_{c_j}, r) \le d\}| \qquad (8)$$

Incoming Flow: The number of transitions whose start place is outside shop r's neighborhood but the end place is inside r's neighborhood:

$$f_r^{IF} = |\{(c_i, c_j) \in Ts \ : distance(s_{c_i}, r) > d$$
$$\&\& \ distance(s_{c_j}, r) \le d\}| \qquad (9)$$

This metrics indicate how well the area could attract customers from remote regions.

Transition Quality: This measures the potential number of customers that might be attracted from shop r's neighbors:

$$f_r^{TQ} = \sum_{s \in S : distance(s, r) \le d} \sigma_{\gamma(s) \to \gamma(r)} \times CI_s \qquad (10)$$

$$\sigma_{\gamma(s) \to \gamma(r)} = E\Big[\frac{|\{(c_i, c_j) \in Ts : s_{c_i} = s \ \&\& \ \gamma(s_{c_j}) = \gamma(r)\}|}{CI_s}\Big] \qquad (11)$$

where CI_s is the number of check-ins at shop s. $\sigma_{\gamma(s) \to \gamma(r)}$ is the expected probability of transitions from category $\gamma(s)$ to category $\gamma(r)$.

Peer Popularity: These metrics assesses shop r's relative popularity in comparison with shops of the same category:

$$f_r^{PP} = \frac{CI_r}{\overline{CI_r}} \qquad (12)$$

where $\overline{CI_r}$ means how many check-ins the shops of category $\gamma(r)$ have on average. We use f_r^{PP} instead of the restau-

rant's absolute check-in number to eliminate popularity bias caused by store nature. People are more likely to post check-ins at some types of shops like Starbucks, while they do not like to check in at Shaxian Refection, which is a famous low-cost restaurant in China. So f_r^{PP} reflects mobility popularity better through normalization .

4.2 Performance

Figure 4b presents the AUC performance of mobility predictors with logistic regression. Among the individual predictors, Peer Popularity is the strongest one and transition quality is the weakest one. This is reasonable because (1) the shop's own popularity can better reflect its business status than the popularity of the area around it; (2)people tend to choose nearby restaurants for dinner. Again, by combining all mobility features the AUC is significant (p-value < 0.01) better than the best individual one (peer popularity).

5. PREDICTING WITH ONLINE REVIEWS

Online reviews directly reflect customers' satisfaction with the restaurants, thus the data is a big fortune worth mining. In Section 2.3, we provide the initial results from rating scores only. In this section we go deeper with review data.
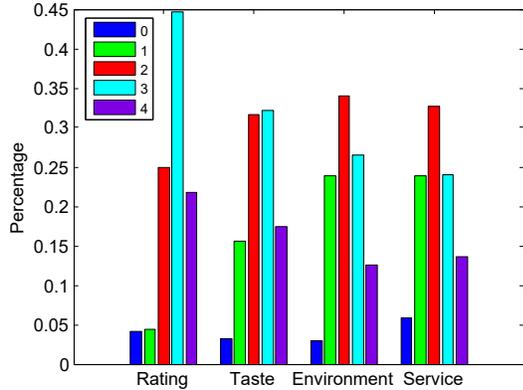
5.1 Rating Values

When writing a review for a restaurant, the consumer is asked to provide five scores on different aspects including: (1)overall rating (2)consumption level (3)taste (4)environment and (5) service quality. The rating scores are scaled from 0 to 4 except for consumption level. The distribution of scores is shown in Figure 6. The majority of users prefer to give a medium score, like 3 or 2.

For each type of score we compute the average, maximum and minimum values as features. The prediction results are shown in Figure 4c. The best individual score is overall rating, which can be regarded as a score summarizing all the other four ratings. Using of all rating scores yields a significantly better (p-value<0.01) performance than using only the best individual one. It means that explicitly using all scores as predictors can provide more comprehensive information than compressing them into one score.

5.2 Review Text

Besides numerical scores, users may write comments to express their opinions as well. Here we list two real review descriptions posted by customers as an example:

1. The portions were too small. We spent more than 500 yuan and ordered the beef combo, and he only offered

Figure 6: The distribution of different review ratings.

(a) BOW with various vocabulary size.

(b) Comparison of different textual models.

Figure 7: Performance analysis for review text. Logistic regression is used as the classification method.

us a small piece of beef steak and two drinks. I will never come here again!!!

2. The food here is terrible. We almost ate nothing and left in a moment. This restaurant is far worse than the one next to the museum.

The first reviewer points out that the price is too high while the food is too little. The second reviewer is complaining about the taste of the food. Both provide helpful knowledge about the potential problems with the restaurant. Inspired by this as well by the observations in Section 5.1, we examine whether we can mine more knowledge besides conclusive scores by exploiting textual information.

Bag-Of-Words. First we employ the Bag-Of-Words(BOW) model and use words as predictors. We collect all the reviews and then segment the sentences into words. To remove infrequent and helpless words, we use χ^2 statistic (CHI) [26] to select the top k_{voc} most useful words as the text representation:

$$f_r^{bow} = (wc_1, wc_2, ..., wc_{k_{voc}}) \quad (13)$$

Where wc_i is the frequency of the i-th word. Figure 7a illustrates the AUC of the BOW model in the Beijing dataset with different k_{voc} settings. We observe that $k_{voc} = 1000$ performs better than other settings. Ideally we should observe a non-decreasing trend when increasing k_{voc}. Performance in our case drops from $k_{voc} = 8000$ due to the limited number of training instances. In Beijing dataset we have 8615 instances, so when k_{voc} increase, the curse of dimensionality occurs and the amount of our data is not sufficient to train an optimal model. Thus in the next step we focus on language models which can reduce the dimension of textual features.

Word Embedding. Word embeddings are dense and low-dimensional representation of words[29][28]. Each word is represented as a D_{we}-dimensional continuous-valued vector, where D_{we} is a relatively small number(i.e., 100 in our experiment). Similar words have similar vectors. We train a model on our corpus using Word2Vec toolkit[3] and get a vector for each word w:

$$\pi^w = (\pi_1^w, \pi_2^w, ..., \pi_{D_{we}}^w) \quad (14)$$

[3]https://code.google.com/archive/p/word2vec/
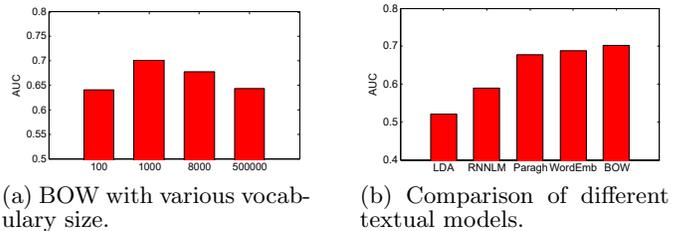
Similarly, we refer to a restaurant's representation as a D_{we}-dimensionlal vector, which is a TF-IDF weighted average of all words that have ever appeared in the restaurant's reviews:

$$\pi_i(r) = \sum_{p \in RV(r)} wc_p \times log \frac{1}{dc_p} \times \pi_i^p \quad for \ 0 \le i < D_{we} \quad (15)$$

$$f_r^{WE} = (\pi_1(r), \pi_2(r), ..., \pi_{D_{we}}(r)) \quad (16)$$

Where $RV(r)$ indicates the review set of restaurant r, and dc_p means the number of restaurants containing word p in its review set. Finally we refer to f_r^{WE} as the representation of the restaurant.

Paragraph Vector. Based on the word embedding model, Quoc et al. [23] propose *Paragraph Vector*, which learns continuous distributed vector representations for pieces of texts. We use this model in Gensim [4] to generate an embedding for each review. Then the representation of a restaurant is the average of its reviews' embeddings weighted by the review's text length.

Neural Language Model. [3] proposes a method to learn micro-post representation based on the Elman network[15][27]. Basically, it is a recurrent neural network-based language model aimed at predicting the next word's probability distribution given previous words. The architecture is shown in Figure 8. The input $\mathbf{w}(t) \in \mathbb{R}^{k_{voc}}$ is the one-hot representation of a word at time t. The hidden layer $\mathbf{h}(t) \in \mathbb{R}^{D_{rnn}}$, also known as the context layer, is computed based on $\mathbf{w}(t)$ and $\mathbf{h}(t-1)$, which is the context layer at time t-1. D_{rnn} is the size of hidden layer dimension, which is set to 100 in the experiment. The output $\mathbf{y}(t) \in \mathbb{R}^{k_{voc}}$ is the probability distribution of a word at t+1.

We use the neural language model to generate the restaurant's representation. We build a recurrent neural network using CNTK[1], and the training runs 3.5 days with GPU NVIDIA GK107. Then for each review p_i, we feed its textual content into the neural network word by word. We use $\mathbf{h}(T_{p_i})$ to be the representation of review p_i, where T_{p_i} is the word count of p_i and $\mathbf{h}(T_{p_i})$ is the state of hidden layer at the last word of p_i. For a restaurant, its representation is the average of all its reviews' representation.

$$f_r^{NLM} = \frac{1}{|RV(r)|} \times$$

$$(\sum_{p \in RV(r)} h(T_p)_1, \sum_{p \in RV(r)} h(T_p)_2, ..., \sum_{p \in RV(r)} h(T_p)_{D_{rnn}}) \quad (17)$$
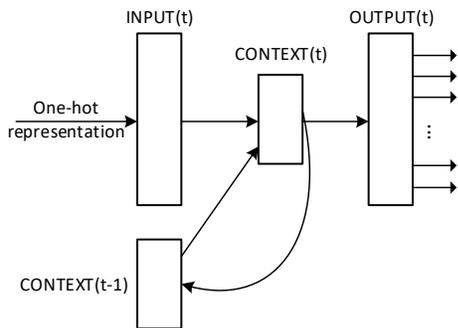
[4]https://radimrehurek.com/gensim/

Figure 8: The recurrent neural network for language modelling

Topic Model. Another way to represent a restaurant is to generate its topic distribution. We concatenate all the reviews belong to the same restaurant to form a document. We exploit Latent Dirichlet Allocation (LDA)[5] to model the topics. In LDA, each document is represented as a probability distribution over topics, and each topic is represented as a probability distribution over words. Thus we refer to the topic distribution vector as the restaurant's representation.

5.3 Results

We set $k_{voc} = 1000$ in BOW predictors for its superiority. Figure 7b shows the performance of each model. The RNN language model does not work as well as word embedding and bag-of-words. One possible reason is that textual context is not as important as the words themselves in our case. Another possible reason is that a simple recurrent neural network might not be able to keep long-term dependencies. Unlike tweets which are usually short, reviews might contain much longer textual content in length. Due to the vanishing gradients, the RNN model can not model the connection between final outputs and earlier input words. In the future we will enhance RNN with Long Short Term Memory units[16].

Unexpectedly, LDA doesn't work in our task. The possible explanations are: (1)the *topic* in our case can be regarded as restaurants' characteristics, such as food-style. However, for restaurant survival prediction, *topic* is not a good signal compared with *opinion*. (2)it might not be proper to concatenate all the reviews of the same restaurant together to form a document, since different reviews may concentrate on different aspects.

Paragraph2Vector performs slightly worse than *word embedding*, which to some extent verifies our guess that textual context is not as important as words themselves in our task. Since the two models share very similar algorithms, we only use one of them for further experiments. In the next section we will use *word embedding* and *BOW* as textual predictors.

Since review text plays such an important role in prediction, we use χ^2 score to select top 10 words related to alive and closed restaurants respectively, and we list them in Table 4 (words are translated from Chinese). Row *alive* lists the key words which indicate a higher probability for a restaurant to survive. These words describe the strength of the restau-

Table 4: Top informative words for the two types of restaurants (translated from Chinese). Words are selected based on χ^2 score.

type	top words
alive	time-honored brand; from childhood; well-deserved reputation; crowded; a dozen years; well-know; early morning; not tire; state-run; must-try
closed	group purchase; original price; four people set meal; sluggish; Meituan; double meal; lack of customers; booth; catfish; leaflet; LaShou Group

rant, including already having a long history (*time-honored brand, from childhood, a dozen years*), having strong reputations (*well-deserved reputation, well-known*), being popular (*crowded*), serving delicacies (*not tire, must-try*), and foundation (*state-run*). Key words for closed restaurants are more interesting. *Meituan*[5] and *LaShou Group*[6] are two famous Chinese group buying websites. It seems that restaurants which offer attractive group purchases but actually serve disappointing food have a higher probability of closing in the next few years. On the other side, the story behind words like *original price, double meal, and leaflet* is that consumers are complaining about the food or service: consumers feel that the reality of the food is a long way from the image on the leaflet. Lastly, words like *sluggish* and *lack of customers* directly describe the gloomy status of the business, which obviously make it hard for the restaurant to survive.

6. COMBINING MODELS

In previous sections we have studied various features' individual predictive power. Now we want to figure out how performance can be improved by combining features from different groups. In order to test the generality of models, we conduct experiments separately on three cities, i.e. Beijing, Shanghai and Guangzhou, which are the most popular cities in our dataset. In each experiment, we train a model based on parameters tuned from a validation set, and then report the performance in the test set. We examine the performance of logistic regression(LR), gradient boosted decision tree (GBDT)[9][7], and supported vector machine (SVM).

Results are shown in Table 5. Rows from G to E present the detailed performance of different individual models. For all three cities, textual models(BOW and WE) significantly outperform geographical, mobility and rating models. Geographical metrics and people mobility patterns are implicit factors reflecting the spatial demand within an area for the restaurant. However, most of the time, before a merchant opens a new retail store, he/she will carefully choose an optimal location to place the store, e.g., McDonald's restaurants are often placed near train stations; a new Muslim restaurant may open to meet people's dietary requirements if there are no existing Muslim food shops around. On the other side, people's online reviews are explicit feedbacks about the restaurant. The rating scores may not be directly connected to the future survival of the restaurant. Take environment score for instance. *Shaxian Refection* is a low-cost restaurant

---

[5]http://www.meituan.com

[6]http://www.lashou.com

Table 5: AUC performance of model combination for Beijing, Shanghai, and Guangzhou. The best result for each city is highlighted in bold. Significance test (denoted by *) indicates the best model significantly outperforms the others with *p-value<0.05*. G=geographical predictors, M = mobility predictors, B=bag-of-word predictors, E=word embedding predictors. GM=geographical+mobility predictors, ALL=using all predictors, -GM=use all predictors except geographical and mobility predictors, and the similar goes for -R, -B, -E.

	Beijing			Shanghai			Guangzhou		
	GBDT	LR	SVM	GBDT	LR	SVM	GBDT	LR	SVM
G	57.03	57.10	56.89	56.40	56.30	56.13	59.76	60.80	61.21
M	61.09	61.32	61.29	58.45	57.75	57.25	59.99	60.67	61.59
R	63.10	61.36	59.44	64.03	61.36	61.32	64.64	64.40	61.35
B	70.82	70.14	70.06	71.85	69.46	69.07	73.53	72.05	72.02
E	67.35	68.83	68.19	71.33	71.08	70.48	72.14	73.13	73.15
GM	61.25	61.78	61.47	59.43	58.89	58.47	60.61	60.58	61.40
ALL	**72.10***	71.72	71.48	**73.46***	72.05	72.02	**75.56***	74.21	74.45
-GM	71.54	71.12	71.16	73.04	71.70	70.70	74.77	73.55	73.45
-R	71.88	71.49	71.10	73.05	71.59	71.54	75.03	74.02	74.00
-B	69.98	70.26	69.94	72.08	71.69	71.26	72.83	73.99	73.71
-E	71.77	71.21	71.12	72.64	70.82	70.90	74.33	72.84	73.16

with a cost per person below 3 dollars, and the environment inside is usually dirty and noisy. The rating scores will be low compared with an expensive restaurant. Besides, as revealed in [4], low-cost restaurants tend to receive less number of reviews. However, many people still go to *Shaxian Refection* because they want to save money. This can explain why some restaurants have poor rating scores or few reviews in quantity but still survive for a long time. Rather, textual information reflects the restaurant's advantages and disadvantages best. If the restaurant continuously has bad service which is deemed unacceptable by customers, it will close down in the future.

Geographical and mobility predictors seem to be highly correlated, so adding geographical predictors to mobility predictors does not lead to much improvement. Rows of *-GM, -R, -B, -E* show how performance decreases with cutting part of the components. The best model is GBDT using all predictors, which is consistent among the three cities. It indicates that although textual models have the strongest predictive power, geographical metrics, human mobility and rating scores can still provide supplementary knowledge which improves the accuracy of the model.

7. RELATED WORK

Bankruptcy prediction is a common topic in management and finance literature [31, 22, 25, 11, 20, 21, 13]. Among these works, various machine learning algorithms have been tried to produce accurate and automatic models. Olsen et al. published the first study on predicting bankruptcy in the restaurant industry [29]. Gu et al.[14, 20] adapted M-DA and Logit regression for hotels and restaurants analysis. Kim and Upneja [20] studied restaurant financial distress prediction and recommended the use of the AdaBoosted decision tree model because of its best prediction performance in their dataset. Li and Sun [23] investigated the imbalanced data problem in hotel failure prediction challenge, and proposed a new up-sampling approach to help produce more accurate performance. [10] constructed a hotel bankruptcy prediction model using a probabilistic neural network, which not only provided high accuracy but also was able to determine the sensitivity of the explanatory variables. However,

all of these prior works studied finance-related variables (e.g. liquidity, efficiency, leverage, profitability) or very few non-financial variables (e.g. level of quality, zone of destination, the category of firm). Usually these data are hard to acquire, which results in that most of the prior works conduct experiments on a very small dataset. In this paper, we fully utilize the advantage of online big data and perform restaurant survival analysis through extracting exogenous factors (e.g., geographical nature, consumer opinions).

Meanwhile, the growth of online location-based service brings a broad range of new technologies to study the world around us [19, 12, 34]. Dmytro et al. studied the problem of optimal retail store placement in the context of location-based social networks[19]. They collected data from Foursquare and mined two general signals, i.e. geographic and mobility, and demonstrate that the success of a business depends on both of the two factors. Similar technique was used in [12], where the target was to model the impact of Olympic Games on local retailers. Our target is quite similar to this work. While they were studying the future of retail stores from the perspective of impacts from big events, we are projecting the restaurant's future by its own properties. Both above works didn't mine users' opinion knowledge. As demonstrated in this paper, users' opinion is the best predictor when it comes to studying the restaurant business. Yuan et al.[34] proposed a location to profile(L2P) framework in order to infer users' demographics based on user check-ins. When constructing the tensor model, they clustered check-ins according to words for location knowledge flattening. It's kind of similar with clustering words into topics. In our case we use LDA and find that it didn't work well. In the future we can consider modifying Yuan's framework to generate restaurant's latent topics through a tensor factorization method.

There are also some literature that integrate users' opinion knowledge into the prediction model. Hadi et al.[3] proposed utilizing RNNs to learn short text representation for detecting churny contents. Even though the RNN model doesn't work well in our scenario, one of their conclusions is consistent with ours: the combination of the bag-of-words model and the language model could yield better performance than using only one of them. Yanjie et al.[10] exploited LDA mod-

el to extract topic from user check-ins. However, handling reviews through topic model is proven not effective in our scenario.

Restaurant survival prediction is also related to customer churn prediction[2]. Churn means the customer leave a product or service. Existing research works have explored various user features through their historical behavior[8][2], and with the fast growth of online social network, several works have studied social influence on churn analysis[35][30]. However, shop survival analysis is obviously different from traditional churn analysis. To some extent shop's failure could be regarded as all or the vast majority of its customers' churn.

There are some research works that deserve a mention because are related to restaurant analysis. [17] showed that atmospherics and service functioned as stimuli that enhanced positive emotions, which mediated the relationship between atmospherics/services and future behavioral outcomes. [24] conducted experiments to show negative reviews could influence customer's dinning decision. [33] provided a comprehensive study on restaurants and embodied dinning preference, implicit feedback and explicit feedback for restaurant recommendation. [4] studied how restaurant attributes, local demographics and local weather conditions could influence the reviews of restaurants.

8. CONCLUSION

This paper discusses the problem of restaurant survival prediction by modeling four perspectives: geographical metrics, user mobility, rating scores, and review text. We provide detailed analysis on each perspective separately and demonstrate its predictive power. We find that if used properly, review text can reflect a restaurant's operating status best. Comprehensive experiments show that integrating different predictors can lead to the best model, and it is consistent among different cities.

In the future study, we are going to : (1) investigate more appropriate language models to extract better knowledge from review text; (2) design a unified model to incorporate heterogeneous learning algorithms so that the performance will not limited by a single learning algorithm such as GB-DT.

9. REFERENCES

[1] A. Agarwal, E. Akchurin, C. Basoglu, G. Chen, S. Cyphers, J. Droppo, A. Eversole, B. Guenter, M. Hillebrand, R. Hoens, et al. An introduction to computational networks and the computational network toolkit. Technical report.

[2] J.-H. Ahn, S.-P. Han, and Y.-S. Lee. Customer churn analysis: Churn determinants and mediation effects of partial defection in the korean mobile telecommunications service industry. *Telecommunications policy*, 30(10):552–568, 2006.

[3] H. Amiri and H. Daumé III. Short text representation for detecting churn in microblogs. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[4] S. Bakhshi, P. Kanuparthy, and E. Gilbert. Demographics, weather and online reviews: A study of restaurant recommendations. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 443–454, New York, NY, USA, 2014. ACM.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[6] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 25–32. ACM, 2004.

[7] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[8] G. Dror, D. Pelleg, O. Rokhlenko, and I. Szpektor. Churn prediction in new users of yahoo! answers. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 829–834. ACM, 2012.

[9] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[10] Y. Fu, Y. Ge, Y. Zheng, Z. Yao, Y. Liu, H. Xiong, and J. Yuan. Sparse real estate ranking with online user reviews and offline moving behaviors. In *2014 IEEE International Conference on Data Mining (ICDM)*, pages 120–129. IEEE, 2014.

[11] M. A. F. Gámez, A. C. Gil, and A. J. C. Ruiz. Applying a probabilistic neural network to hotel bankruptcy prediction. *Encontros Científicos-Tourism & Management Studies*, 12(1):40–52, 2016.

[12] P. Georgiev, A. Noulas, and C. Mascolo. Where businesses thrive: Predicting the impact of the olympic games on local retailers through location-based services data. *arXiv preprint arXiv:1403.7654*, 2014.

[13] Z. Gu. Analyzing bankruptcy in the restaurant industry: A multiple discriminant model. *International Journal of Hospitality Management*, 21(1):25–42, 2002.

[14] Z. Gu and L. Gao. A multivariate model for predicting business failures of hospitality firms. *Tourism and Hospitality Research*, 2(1):37–49, 2000.

[15] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation*, volume 1. Basic Books, 1991.

[16] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[17] S. S. Jang and Y. Namkung. Perceived quality, emotions, and behavioral intentions: Application of an extended mehrabian–russell model to restaurants. *Journal of Business Research*, 62(4):451–460, 2009.

[18] P. Jensen. Network-based predictions of retail store commercial categories and optimal locations. *Physical Review E*, 74(3):035101, 2006.

[19] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo. Geo-spotting: Mining online location-based services for optimal retail store placement. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 793–801, New York, NY, USA, 2013. ACM.

[20] H. Kim and Z. Gu. A logistic regression analysis for predicting bankruptcy in the hospitality industry. *The Journal of Hospitality Financial Management*, 14(1):17–34, 2006.

[21] H. Kim and Z. Gu. Predicting restaurant bankruptcy: A logit model in comparison with a discriminant model. *Journal of Hospitality & Tourism Research*, 30(4):474–493, 2006.

[22] S. Y. Kim and A. Upneja. Predicting restaurant financial distress using decision tree and adaboosted decision tree models. *Economic Modelling*, 36:354–362, 2014.

[23] Q. V. Le and T. Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 1188–1196, 2014.

[24] C. C. Lee. Understanding negative reviews' influence to user reaction in restaurants recommending applications: An experimental study.

[25] H. Li and J. Sun. Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples–evidence from the chinese hotel industry. *Tourism Management*, 33(3):622–634, 2012.

[26] T. Liu, S. Liu, Z. Chen, and W. Ma. An evaluation on feature selection for text clustering. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, pages 488–495, 2003.

[27] T. Mikolov. Recurrent neural network based language model.

[28] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[29] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[30] R. J. Oentaryo, E.-P. Lim, D. Lo, F. Zhu, and P. K. Prasetyo. Collective churn prediction in social network. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 210–214. IEEE Computer Society, 2012.

[31] M. Olsen, C. Bellas, and L. V. Kish. Improving the prediction of restaurant failure through ratio analysis. *International Journal of Hospitality Management*, 2(4):187–193, 1983.

[32] N. J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu, and X. Xie. We know how you live: Exploring the spectrum of urban lifestyles. In *Proceedings of the First ACM Conference on Online Social Networks*, COSN '13, pages 3–14, New York, NY, USA, 2013. ACM.

[33] F. Zhang, N. J. Yuan, K. Zheng, D. Lian, X. Xie, and Y. Rui. Exploiting dining preference for restaurant recommendation. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 725–735, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee.

[34] Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie. You are where you go: Inferring demographic attributes from location check-ins. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, pages 295–304, New York, NY, USA, 2015. ACM.

[35] Y. Zhu, E. Zhong, S. J. Pan, X. Wang, M. Zhou, and Q. Yang. Predicting user activity level in social networks. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, CIKM '13, pages 159–168, New York, NY, USA, 2013. ACM.