Restaurant Failure Prediction Based on Multi-View Online Data

Xiaoxiao Sun, Ping Liang, DongjinYu*
School of Computer Science and Technology
Hangzhou Dianzi University
Hangzhou, China
sunxiaoxiao@hdu.edu.cn, liangpingprivate@gmail.com, yudj@hdu.edu.cn

***Abstract*—Predicting future development trends of restaurants (especially failure judgement) helps entrepreneurs to identify potential downward trends in their business and supports potential investors' investment decisions. Review apps, such as Yelp, are generating massive restaurant-related online data every day, which provides a solid data source for the prediction through big data technology rather than applying commercial data with limited access and poor time efficiency. In this paper, we propose a novel multi-view restaurant failure prediction model named Semantic Business Cluster Effect Model (SBCM) based on online review data. Specifically, our model consists of three views: semantic view (we capture semantic features of reviews via a neural network and different reviews are assigned with different importance according to their reviewers' habits), business attribute view (we select the most influential business attributes from datasets), and business cluster effect view (we identify business clusters based on density and differentiate restaurants into different clusters). All attributes are then input into a LightGBM model to conduct the prediction. Experiments on public Yelp datasets of Toronto and Las Vegas from 2016 to 2017 demonstrate that SBCM averagely outperforms SVM and XGBoost by 14% and 3% respectively in terms of AUC. Furthermore, we find that credit card support, lunch support and noise level are the three most significant business attributes that influence the restaurant popularity online.**

Keywords: restaurant failure prediction; big data analysis; semantics extraction; cluster effect; LightGBM; Yelp

I. INTRODUCTION

Business failure prediction is a scientific field with long history, whose accurate results help entrepreneurs to identify potential downward trends in their business performance and give them timely warning to change business strategies in advance. Meanwhile, according to National Restaurant Association (NRA) [1], restaurant industry sales are projected to total $863 billion in 2019 and equal 4 percent of the U.S. gross domestic product. Meanwhile, restaurant workforce is about 10% of the overall U.S. workforce. Restaurants have played an essential role in the economy of a thriving society. Therefore, restaurant failure prediction is worthy of deep studying.

With the development of mobile Internet technology, restaurants are changing their traditional business patterns and starting to pay more attention to online advertisement. Apps such as Yelp provide platforms for restaurants to advertise their foods online and for customers to share their dining experiences. These reviews provide important references accordingly for other customers to select restaurants. Studies show that online data (i.e. reviews and check-ins) are related to restaurant performance and using them to predict restaurant failure is feasible [2,3,4,5]. However, few researchers have made in-depth studies on the relationship between the abundant semantics of reviews (The taste, environment, service, price, etc.) and business performance. Even some did, they ignored the fact that different customers have different preferences on giving ratings and reviews. For example, some customers like giving high ratings to almost all restaurants and some customers prefer to give low scores. In addition, some people tend to use personalized words to express their point of view (e.g., "good" is used to express satisfaction by some strict customers and express borderline by some lenient customers). Yelp also defines various attributes, such as credit card, Wi-Fi, parking etc., which can be used to predict the future performance of restaurants [3]. Moreover, the success and failure of restaurants are usually affected by their surrounding business districts, which is not considered in most studies.

To address the above problems, in this paper, we propose a novel prediction model named Semantic Business Cluster Effect Model (SBCM) based on the review semantics, business attributes and cluster effect to predict business failure of restaurants. In SBCM, we first design a neural network to capture semantic features from the newest and most popular reviews of each restaurant. Secondly, we design a review importance weight metric to match reviews with reviewing habits of different customers. Thirdly, we identify the importance of different attributes provided by Yelp and select the most important ones as the input of the prediction. Fourthly, we identify business clusters by Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [6] and differentiate restaurants into different clusters. Finally, semantic features, business features and cluster effect features are integrated and input into LightGBM[7] to obtain the final prediction value. Experiments on public Yelp datasets demonstrate that SBCM averagely outperforms SVM and XGBoost by 14% and 3% respectively in terms of AUC.

The rest of the paper is structured as follows. After discussing related work in Section 2, we introduce some basic definitions and raise our problem in Section 3. Section 4 presents

DOI reference number: 10.18293/SEKE2020-077

our prediction model in detail. Experiments and visualization results are given in Section 5. Finally, Section 6 concludes the paper and introduces the future work.

II. RELATED WORKS

Since the late 1960s, business failure prediction has been widely investigated through statistical techniques and discriminant analysis [8]. Logit analysis [9], generalized extreme value [10], machine learning techniques [11], neural networks [12] etc., have also been applied in the prediction in recent years. The common ground of these business failure prediction methods is that they mostly rely on commercial data, such as stock prices, working capital and debt, which are more suitable for large enterprises rather than medium-scale businesses, especially not applicable for restaurants that we focus on in this paper. In addition, commercial data are usually statistical-based, which lack timeliness and sustainability.

In recent years, online review apps become popular with the proposal of concepts such as big data and smart city, which provide consistent and substantial data of small and medium-scale businesses, giving researchers a new idea to predict restaurant failure. For example, Zhang et al. found that the rating stars, sentiments, and photos of reviews are closely associated with restaurant survival [2]. Snow et al. [3] studied the influence of different business attributes and reviews on restaurant failure. Wang et al. [5] incorporated check-in data captured from location-based services to predict restaurant failure and obtained better results than using business characteristic variables only. The aforementioned works, however, neglect the fact that business performance of a restaurant is not only affected by its own factors, but also affected by its surrounding neighbors. Hu et al. [13] proved that applying the neighbors' business performance to predict the rating of a business is feasible, and they observed positive correlations between business individual's ratings and its neighbors' ratings.

In conclusion, restaurant failure is a complex problem affected by various factors. But until now, these factors have not been adequately considered simultaneously in previous researches. This paper is dedicated to proposing a prediction model that synthesize all influence factors, including review semantic features incorporated with the corresponding review habit of customers, influential business attributes and business cluster effect.

III. THE PREMIMINARY

In this section, we define several basic concepts, which are designed based on Yelp dataset, and are also applicable in other datasets.

Since Yelp data only shows whether one restaurant is still open and does not indicate the specific closure date, existing researches approximates the business status of restaurants by the time of reviews or check-in records [2,3,5]. Therefore, we use the similar method and define *Restaurant Failure* as follows:

Definition 1. Restaurant Failure. *The date of the first review submitted is regarded as the opening date of a restaurant, and the date of the last review submitted is regarded as the closure date of the restaurant.*

Generally, business entities such as restaurants choose to cluster together. Studies have shown that clusters significantly promote business booming [14,15], and the incentive effect is called cluster effect [16]. One good example of this phenomenon is that customers prefer to choose a venue with many restaurants rather than a place with one standalone McDonald's.

Definition 2. Restaurant Cluster. *The restaurants are mapped to the map by latitude and longitude. A certain number of restaurants gathering geographically forms a restaurant cluster.*

The problem to be solved in this paper is to predict whether a restaurant will fail in some time, which is defined as follows:

Problem Definition. *Build a prediction model* $\mathcal{F}$*, which contains the following structures:*

Input: (a)heterogeneous data (review, check-in and business attributes) of a target restaurant; and (b)the cluster information data (the latitude and longitude) of all restaurants in the same city.

Output: whether the target restaurant will fail in the year of $t+1$.

IV. THE FRAMEWORK

The model that we propose in this paper mainly consists of three views: semantic view (we capture semantic features of reviews via a neural network and different reviews are assigned with different importance according to their reviewers' habits), business attribute view (we select the most influential business attributes from datasets), and business cluster effect view (we identify business clusters based on density and differentiate restaurants in different clusters). After capturing features of these views, we input the compositive feature vector into lightGBM to get the prediction result as shown in Fig.1.

A. Extractation of Influence Factors

1) Extracting Semantic Features: To capture the abundant semantics about different aspects of a restaurant, review texts are firstly converted to machine learnable sequences. Since one-hot encoding leads to too long vector, the output of word embedding tools such as word2vec[17] or GloVle[18] is still too long (e.g. 100 dimensions is a common length of word vector, but a review with only 10 words is converted to 10*100 dimensions). Convolutional neural network (CNN), which is famous for its ability of high-dimension information extraction, is applied to reduce the size of vectors. The intermediate vectors outputted from CNN represent the highly condensed semantic features and contain the same semantics with origin sentence. So, we design a deep learning sentiment classification model to obtain the review representation vector from CNN layers as shown in Fig.2. When customers submit reviews, they are required to attach rating stars at the same time, which contain the same emotion with reviews. Lots of studies use the rating stars as sentiment label to train sentiment classification model [19,20]. In this paper, we also employ the rating star as our sentiment label.

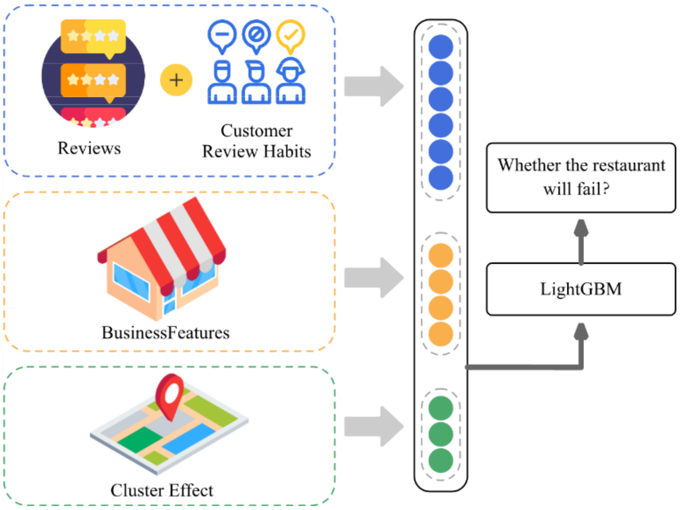


Figure 1. The Framework of SBCM.

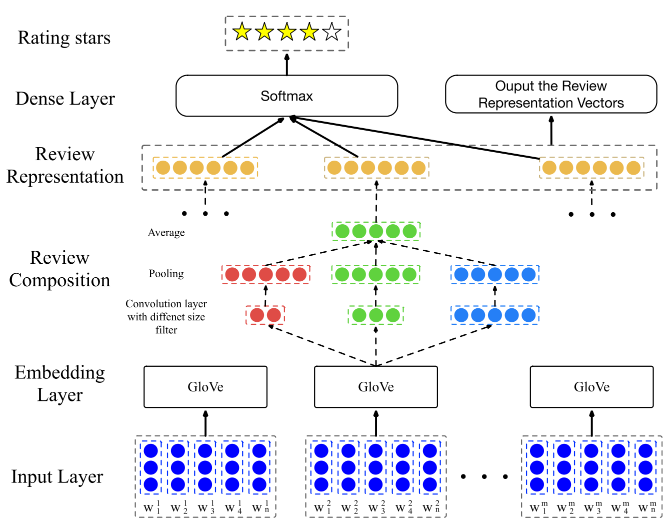


Figure 2. The method to reduce dimensions of review representation vectors.

For each restaurant α, we choose the most popular and recent m reviews to input into our model to get the review representation vectors. The input matrix of input layer is created by concatenating the word vectors of a review. The neural network accepts fixed length vector as input, but the length of reviews is usually unfixed. To solve this mismatching problem, we cut out n words from a review as a review vector (If the length of some reviews is less than n, the reviews are padding with zero). After assembling the review vectors, we get a $m * n$ review matrix, which is defined as follows:

$$M_r = (w_1^m; w_2^m; \dots; w_n^m) \tag{1}$$

where w_n^m stands for the n-th word in the m-th review, $M_r \in \mathbb{R}^{m*n}$ is the review matrix.

In our method, GolVe is employed to generate embedding word vectors. All the word vectors in the GloVe are stacked in a word embedding matrix $M_w \in \mathbb{R}^{d\times|V|}$, where d is the dimension of word vector and $|V|$ is the vocabulary size. We employ the pre-trained M_w from GloVe's official website[1] to ensure the efficiency of the word vector. In the embedding layer, every word in M_r is converted into a vector of floating number by finding every word vector in the M_w, which is defined as follows:

$$M_e = F_e(M_r, M_w) \tag{2}$$

where F_e denotes the operation of embedding. The matrix M_e is a set of m reviews, in which every word is converted.

Then, CNN is applied to compute representation vectors of reviews and to reduce the length of vectors. Several convolutional filters of different widths are used in the convolution layer to capture different semantic of various granularities. For example, a convolutional filter with a width of 2 captures the semantics of phrases in a sentence and a width of 5 captures the semantics of short sentences in a sentence as Fig.2 shows. The process of the CNN convolution is defined as:

$$M_c = f(W_e * M_e + b_e)$$

where $*$ denotes the operation of convolution and f is an activation function. W_{rl} and b_{rl} are learnable parameters. We input M_c into a pooling layer to reduce the size. Then an average pooling layer is employed to capture the whole semantics of the review. So far, we get the review representation feature vector matrix $M_o \in \mathbb{R}^{m*k}$, in which k is the output length of the review representation vector.

2) Integrating Review Importance Weight: Reviews are given by different reviewers with respective reviewing habits. Yelp has a simple and intuitive weight metric, i.e. review vote, which is not able to reflect this difference. Inspired by this, we design a review importance weight metric, considering both review vote and reviewer attributes. In Yelp, there are three tree type of votes, i.e., useful, funny and cool, all of which are positive vote. We employ the number of all votes as the weight of a review, which is defined as follows:

$$W_r = \theta_u^r + \theta_f^r + \theta_c^r \tag{3}$$

where θ_u, θ_f, and θ_c denote the number of useful, funny and cool votes. We think that the more votes a review have received the more important a review is. We then adopt the average received votes as the weight of a reviewer, which is defined as follows:

$$W_u = \frac{\zeta_u}{\delta_u} \tag{4}$$

where δ_u denotes the number of reviews that a reviewer has written. ζ_u denotes the number of all kinds of votes that a reviewer have received. η_r denotes the rating star of a review. η_u denotes the average rating star submitted by the reviewer. $\eta = |\eta_r - \eta_u|$ stands for how different between the sentiment of review and the reviewer's ordinary habits. Intuitively, the bigger η is, the more influential a review is. Considering the above factors, the review weight metric is defined as follows:

$$I = ln\big((W_r + W_u) * \eta + 1\big) \tag{5}$$

where I is in [0,1). By calculating every I of m reviews, we can get matrix $M_I \in \mathbb{R}^{m*1}$. To concatenate M_o with M_I, we get a matrix containing semantics and corresponding weights, which is denoted as $M_s \in \mathbb{R}^{m*(k+1)}$.

[1] http://nlp.stanford.edu/data/glove.6B.zip

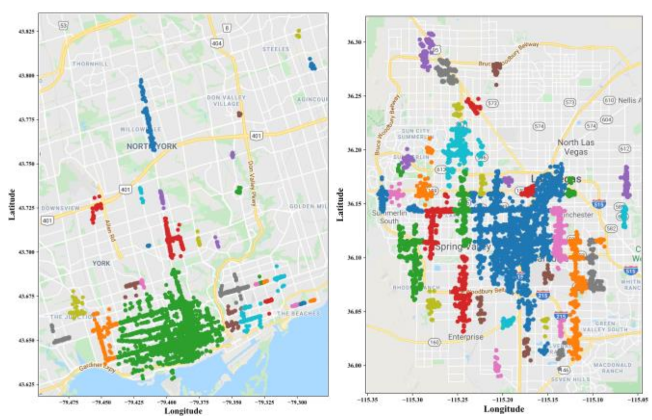
3) Screening Business Attributes:

As we mentioned above, commercial data are commercial secrets with limited access. Thanks to Yelp, we obtain business attributes of the restaurants instead, such as credit card, Wi-Fi, parking etc. We explore the importance of each attribute on restaurant failure by inputting all business attributes as a vector into our prediction component and output the weights.

By removing the zero-impact and low-impact attributessuch as music type and atmosphere, we select the most influential business attributes. Then we concatenate these attributes a_i into a business attribute vector $\Gamma_b = a_1 \oplus a_2 \oplus \ldots \oplus a_i$, where $\oplus$ represents vector connection. Finally, by flattening M_s into a vector and combining with Γ_b, we obtain a vector Γ_{sb}, which contains semantics and business attributes.

4) Capturing the Influence of Cluster Effect :

A restaurant cluster is a geographical location where enough resources and competences amass and reach a critical threshold, which is close to the density cluster. We employ DBSCAN to cluster the restaurants. DBSCAN is an algorithm to discover arbitrary-shaped clusters and to distinguish noise points simultaneously. In detail, DBSCAN accepts a radius value ε and a minimal value $MinPts$, which means that there are at least $MinPts$ points within the area of ε radius. Fig.3 shows the restaurant clusters of Toronto and Las Vegas in 2017 calculated by DBSCAN.



(a) (b)

Figure 3. The restaurant cluster of (a) Toronto and (b) Las Vegas in 2017. In the figure, different colors denotes different restaurant clusters.

Researchers find that increasing the productivity of business clusters increases the competitive advantage of their individual [16]. Therefore, we employ the total number of review and check-in of restaurants in a cluster to reflect its competitive advantage, which is defined as follows:

$$E = \sum (review_\alpha + checkin_\alpha), \alpha \text{ in } C_\alpha \qquad (6)$$

where α denotes a restaurant and C_α denotes the restaurant cluster where α is located in.

By concatenating Γ_{sb} with E, we get a restaurant feature vector Γ_{sbc}, which contains semantic features, business features and cluster effect.

[2]https://www.yelp.com/dataset

B. Prediction Component

LightGBM [7] is a fast, distributed, high-performance gradient boosting framework based on Gradient Boosting, which provides a good way to solve classification and regression problems by combining many tree models into a more accurate one. Compared to other Gradient Boosting Decision Tree (GBDT) algorithms using level-wise tree growth strategy, LightGBM produces more complex trees by following leaf-wise split approach, which is the main factor in achieving higher accuracy. In addition, it supports parallel and GPU learning and has compatibility in handling large-scale data. Therefore, LightGBM is adopted in this paper to predict whether restaurants will fail in the future, which is essentially a classification problem as succeed or fail.

We split the restaurant dataset into a training set and a testing set. After each restaurant in these two sets going through the process and forming a feature vector Γ_{sbc}, as shown in Fig.4, we get a training feature vector set ϕ_{train} and a testing feature vector set ϕ_{test}. Then we input ϕ_{train} into LightGBM to train a model $\mathcal{M}$. Finally, we input ϕ_{test} into model $\mathcal{M}$ and get the final prediction value set ψ, in which every prediction value $\hat{y}$ denotes whether the corresponding restaurant will fail.

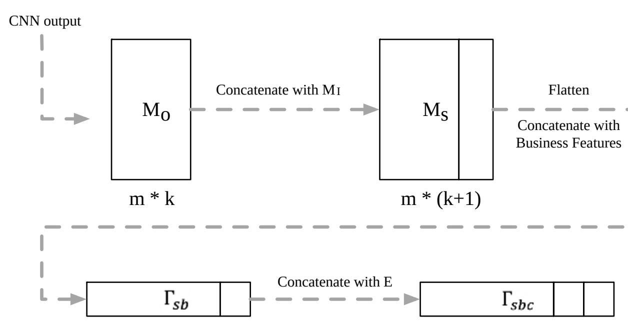


Figure 4. The process to generate a combined feature vector Γ_{sbc}.

V. EXPERIMENTS

A. Experimental Setup

1) Datasets: In this paper, we select the open dataset from Yelp[2] for experiments, which includes more than 190,000 restaurants, more than 5 million review data and more than 1 minllion user data in various cities from October 2004 to November 2018. After analyzing the datasets, we find that Las Vegas and Toronto have the highest number of reviews from different countries, which indicates the popularity of yelp in these cities. Therefore, we select the latest data from 2016 to 2017 of Las Vegas and Toronto as our experiment dataset (the data of 2018 is incomplete). The detailed statistics of the datasets are shown in Table Ⅰ. In our experiment, restaurants that receive less than 10 reviews are filtered to ensure enough data for semantic extraction. We split 80% of the dataset as training data and 20% of the dataset as testing data.

TABLE I. STATISTICS OF DATASETS

	Las Vegas		Toronto	
	2016	***2017***	***2016***	***2017***
#Restaurant	22515	24004	14286	14760
#Closed Restaurant	883	873	744	668

2) Parameters Settings: In our experiment, the number of the most popular and recent reviews(i.e., m)is set as 10. The length of every review(i.e., n) is set as 20. We implement the method to get semantic feature vector by Keras, which is a fast experimentation neural networks API running on top of TensorFlow. Our experiments were run on a cluster with four NVIDIA 1080Ti GPUs.

3) Evaluation Metrics: As Table Ⅰ shows, the numbers of the closed restaurants and surivial restauants are imbalance. Due to this, we employ ROC curve (receiver operating characteristic curve) and AUC (receiver operating characteristic's area under curve) as evaluation metrics. An ROC curve is defined by FPR (false positive rate) and TPR (true positive rate) as x and y axes, respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). The area under the curve is defined as AUC.

4) Comparison Methods: We compare our model with two other methods. As for the SBCM that we proposed, we also conduct multiple experiments without semantics and without cluster effect.

- SVM [21]: A supervised learning model that uses classification algorithms for two-group classification problems.
- XGBoost [22]: A gradient boosting tree model, which has gained widely popularity and attention recently after many winning teams of competitions using it.
- SBCM with no semantics: A variant model of SBCM, which does not contain semantic features.
- SBCM with no cluster effect: A variant model of SBCM, which does not contain cluster effect features.

B. Experimental Results

1) Performance Comparison:

The comparisons between SBCM and other methods is shown in Fig.5. As we can see, SBCM obtains the best performance on the datasets of Toronto and Las Vegas both in 2016 and 2017. On average, SBCM outperforms SVM and XGBoost by 14% and 3%, respectively in terms of AUC. Besides, the model performance on four datasets in terms of AUC also shows that our proposed model is most stable, and the range of SBCM in terms of AUC is 0.05. In general, if the AUC score of a model is above 0.7, it is regarded as a "fair model", and the average AUC of our proposed model is 0.78, which is above the standard.

In order to verify the validity of SBCM, we remove part of the structure in our model and conduct the same experiment. The results, shown in Fig.6, indicate the importance of semantics and cluster effect. Specifically, SBCM outperforms SBCM without semantics and SBCM without cluster effect by 8% and 2%, respectively in terms of AUC, which demonstrates that semantics is more important than cluster effect in our model. In addition, we notice that results in 2017 is worse than 2016 both in Toronto and Las Vegas. The possible reason is that restaurants experienced bad periods of closures in 2016 [3], which reduces the training datasets for 2017 prediction and leads to the low performance.

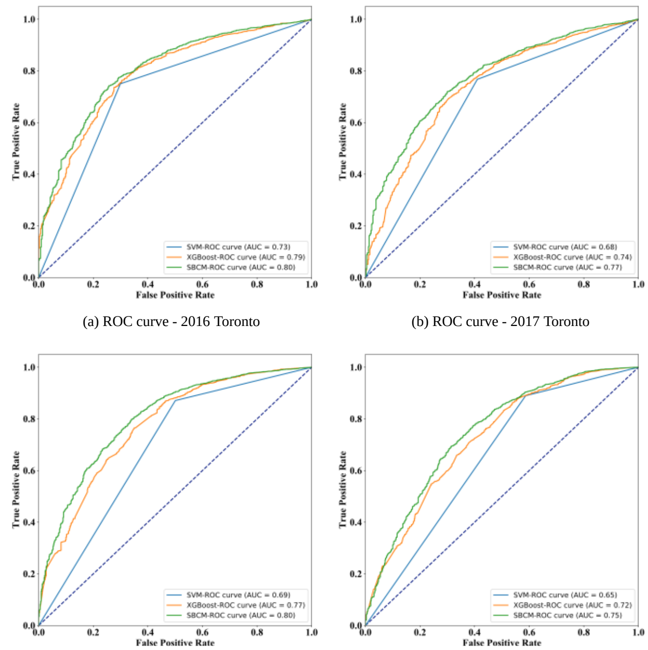


Figure 5. Performance comparisons of different methods.

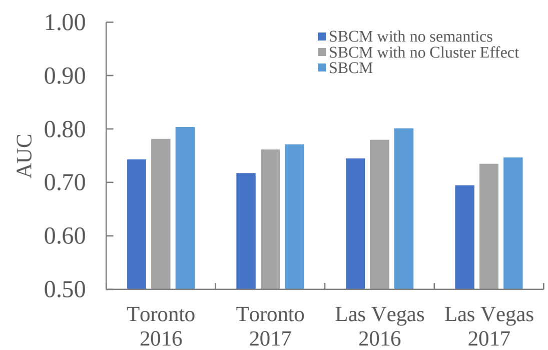


Figure 6. Performance comparisons of SBCM with different structures.

2) Importance ranking of business attributes:

We also explore the importance of business attributes, as shown in Fig.7. We notice that the three most important attributes that affect the future performance of restaurants are credit card support, lunch support and noise level. The importance of credit card support is comprehensible as it is safer and more convenient than cash. We also infer from the results that whether a restaurant provides lunch is closely relevant to customer's choice. Moreover, customers attach importance to the dinning environment such as noise influence.

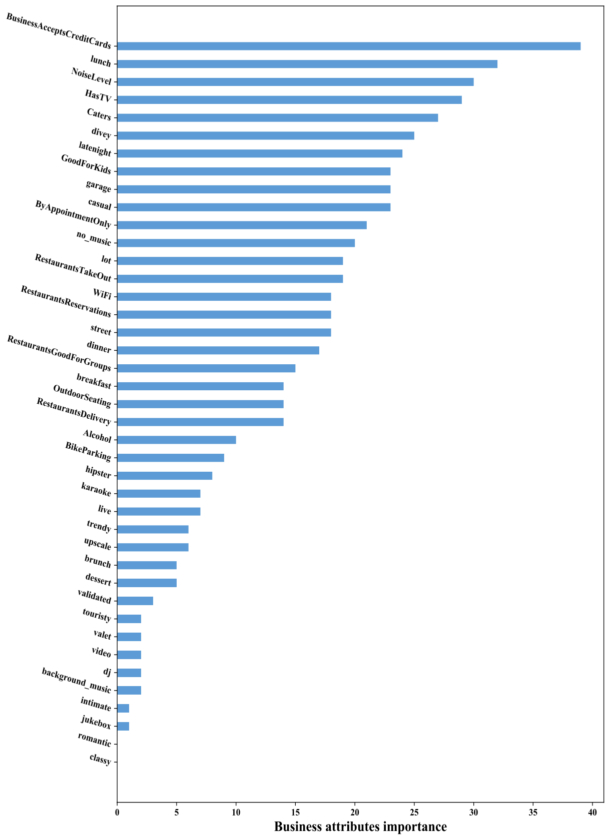


Figure 7. The importance ranking of different business attributes.

VI. CONCLUSION

In this paper, we propose a novel prediction model named SBCM based on review semantics, business attributes and cluster effect to predict business failure of restaurants. Specifically, our model consists of the following steps: 1) we capture semantic features of reviews via a neural network and different reviews are assigned with different importance according to their reviewers' habits; 2) we select the most influential business attributes from datasets; 3) we identify business clusters based on density and differentiate restaurants in different clusters; 4) The above semantic features, business features and cluster effect features are combined and input into LightGBM [7] to get the final prediction value. Experiments on public Yelp datasets of Toronto and Las Vegas from 2016 to 2017 demonstrate that SBCM averagely outperforms SVM and XGBoost by 14% and 3% respectively in terms of AUC.

In the future, we will further study the following issues: (a) explore the semantic influence of specific words that represent restaurant failure; (b) introduce more heterogeneous information to complete the model; and (c) improve the business cluster method to better simulate the actual restaurant clusters.

REFERENCES

[1] "National restaurant association," https://www.restaurant.org/Downloads/PDFs/Research/SOI/restaurant_industry_fact_sheet_2019.pdf, 2019.

[2] Zhang M, Luo L. Can user generated content predict restaurant survival: deep learning of yelp photos and reviews[J]. Available at SSRN 3108288, 2018.

[3] Snow D. Predicting Restaurant Facility Closures[J]. Available at SSRN 3420490, 2018.

[4] Hegde S, Satyappanavar S, Setty S. Restaurant setup business analysis using yelp dataset[C]//2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2017: 2342-2348.

[5] Wang L, Gopal R, Shankar R, et al. On the brink: Predicting business failure with mobile location-based checkins[J]. Decision Support Systems, 2015, 76: 3-13.

[6] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Kdd. 1996, 96(34): 226-231.

[7] Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[C]//Advances in neural information processing systems. 2017: 3146-3154.

[8] Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy[J]. The journal of finance, 1968, 23(4): 589-609.

[9] Jones S, Hensher D A. Predicting firm financial distress: A mixed logit model[J]. The accounting review, 2004, 79(4): 1011-1038.

[10] Andreeva G, Calabrese R, Osmetti S A. A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models[J]. European Journal of Operational Research, 2016, 249(2): 506-516.

[11] Pal R, Kupka K, Aneja A P, et al. Business health characterization: A hybrid regression and support vector machine analysis[J]. Expert Systems with Applications, 2016, 49: 48-59.

[12] Tkáč M, Verner R. Artificial neural networks in business: Two decades of research[J]. Applied Soft Computing, 2016, 38: 788-804.

[13] Hu L, Sun A, Liu Y. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction[C]//Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. 2014: 345-354.

[14] Canina L, Enz C A, Harrison J S. Agglomeration efects and strategic orientations: Evidence from the US lodging industry[J]. Academy of management journal, 2005, 48(4): 565-581.

[15] McClanahan B, Gokhale S S. Centrality and cluster analysis of yelp mutual customer business graph[C]//2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC). IEEE, 2016, 1: 592-601.

[16] Porter M E. The competitive advantage of nations. New York: Free Press[J]. PorterThe Competitive Advantage of Nations1990, 1990.

[17] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

[18] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

[19] Tang D, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015: 1014-1023.

[20] Wen S, Li J. Recurrent Convolutional Neural Network with Attention for Twitter and Yelp Sentiment Classification: ARC Model for Sentiment Classification[C]//Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence. 2018: 1-7.

[21] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.

[22] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-79