Proceedings of Machine Learning Research 95:678-693, 2018 ACML 2018

A Data Driven Approach to Predicting Rating Scores for New Restaurants

Xiaochen Wang wxc575843@sjtu.edu.cn
Yanyan Shen shenyy@sjtu.edu.cn
Yanmin Zhu yzhu@sjtu.edu.cn
Department of Computer Science and Engineering Shanghai Jiao Tong University

Editors: Jun Zhu and Ichiro Takeuchi

Abstract

This paper focuses on predicting rating scores of new restaurants listed in online restaurant review platforms. Most existing works rely on customer reviews to make an prediction. However, in practice, the customer reviews for new restaurants are always missing. In this paper, we mine useful features from the information of restaurants as well as highly available urban data to tackle this problem. We propose a deep-learning based approach called MR-Net to model both endogenous and exogenous factors in a unified manner and capture deep feature interaction for rating score prediction. Extensive experiments on real world data from Dianping show that our approach achieves better performance than various baseline methods. To the best of our knowledge, it is the first work that predicts rating scores for new restaurants without the knowledge of customer reviews.

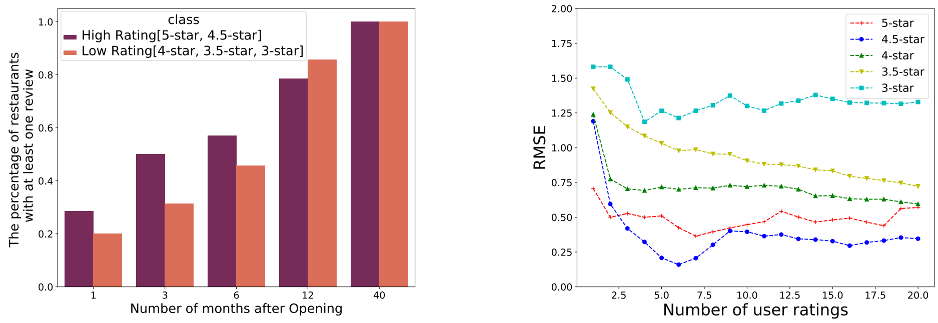
Keywords: multi-source data, deep learning, rating prediction

1. Introduction

The emergence of online restaurant review platforms and their smartphone applications, such as Dianping and Yelp, is the driving force of the proliferation of user-generated reviews on millions of restaurants. For instance, Dianping, China's largest restaurant review site, allows users to browse over 10 million restaurants and publishes more than 60 million reviews by around 200 million monthly active users in China[1]. In addition to the textual comments from customers' reviews, the rating scores of restaurants are undoubtedly one of the most important yet succinct signals to help users find and explore high-quality restaurants. Furthermore, the rating scores are typically the basis in many e-commerce applications to provide better recommendations for customers.

Various research efforts have been devoted to recognizing actual rating scores for restaurants from massive reviews Gan and Yu (2015). The most straightforward way is to aggregate or average rating scores provided by each customer. Due to the fact that the user-contributed rating scores can be unreliable in the presence of human water armies, work Mudambi et al. (2014) leverages machine learning methods towards deriving more fair and unbiased rating scores for restaurants. However, an important limitation of the existing approaches to rating score prediction is that they implicitly assume the existence of massive restaurant reviews from customers, which in practice, may not be available or easily

1. http://www.dianping.com/aboutus



(a) The percentage of restaurants with at least one review

(b) Error between the average of first n user ratings and final ratings.

Figure 1: Statistics of new restaurants.

acquired for new or recently opened restaurants. As shown in Figure 1(a), only less than 30% restaurants have at least one review in one month after opening. And less than 60% restaurants have been reviewed in half year. Even when new restaurants get a few reviews, this reviews cannot predict final ratings well. In Figure 1(b), we calculate the average of first n user ratings to predict final ratings. As the number of user ratings increase, the predicting error of most rating groups become lower. However, the error between the average of first 20 user ratings and final ratings are still high, which will lead misclassify of different rating group. All these observations inspire us to develop an effective approach to *predicting rating scores for this considerably large number of new or recently open restaurants without the knowledge of customer reviews.*

Generally, our problem is related to the cold-start challenge in recommender systems, i.e., predicting user preferences for new items. Many researchers Lin et al. (2013); Sedhain et al. (2014) to this challenge exploit auxiliary information such as location-based social network to measure the similarity among different items and adopt item-based recommendation methods to predict rating score of a new item for each user. But indeed, the two problems differ in one important aspect. That is, we aim to figure out a relatively objective rating score for a new restaurant while recommender systems focus on predicting subjective user-specific preferences. Note that our predicted scores can be further utilized by investors or practitioners to select appropriate restaurant sites in the very beginning.

To solve our problem, a deeper understanding of what kind of factors are contributed to the rating score for a new restaurant is critical. Intuitively, the rating score of a restaurant can be determined by both endogenous and exogenous factors. The endogenous features capture the *intrinsic restaurant properties* that affect the rating score. For instance, restaurant type is an important indicator. According to our analysis on the real world data, over 95% of Starbucks shops in Shanghai are rated between 3.5-4 stars, while the rating scores of hot pot restaurants are more diverse, varying from 3 to 5 stars. As for the exogenous features, they typically reflect the spatial environment surrounding a restaurant. We observe that most top-ranked restaurants are located in city centers or the places with convenient traffic, and those close to motorways are rated with lower scores. This inspires us to exploit urban data such as POIs and road networks to identify exogenous factors for rating score prediction, which have been widely used for urban computing Zheng et al. (2014).

In this paper, we introduce an end-to-end solution to predicting rating scores for new restaurants. We exploit restaurant data from the real-world review platform Dianping as well as urban data (i.e., POIs, road network, satellite light data) to identify important indicators that affect rating scores. As specified above, we consider indicators including average price, type and the number of reviews as endogenous restaurant features, while the indicators such as distances to light centers, the entropy of POI category or road type from urban data are utilized as exogenous features. We conduct an empirical analysis to evaluate the relationship between the selected features and rating score. Furthermore, we propose a deep multi-way rating network (MR-Net) which employs multiple convolution-based residual networks to model the spatial dependencies of different features in nearby regions. We then fuse the outputs from individual residual networks to learn a latent representation that is used for predicting the final rating score. We evaluate our approach using data from the largest China's online review platform Dianping. The results demonstrate 1) the effectiveness of the selected features for predicting rating scores of restaurants; and 2) the advantages of our approach in achieving better prediction performance, compared with 8 baselines.

Our contributions are summarized as below: (1) to the best of our knowledge, it is the first work that studies the problem of predicting rating scores for new restaurants without the knowledge of customer reviews; (2) we propose to exploit both restaurant and urban data to identify important features that affect rating scores of restaurants; (3) we introduce an MR-Net neural method to model endogenous and exogenous features in a unified manner; MR-Net can also learned the spatial relationships from different areas; (4) we conduct extensive experiments to demonstrate the effectiveness of the selected features and the superior performance of the proposed approach in rating score prediction.

The remainder of this paper is organized as follows. Section 2 gives the definitions and the proposed framework. Section 3 presents extracted features and empirical analysis. Section 4 provides MR-Net model for rating prediction. We show experimental results in Section 5, review related works in Section 6 and conclude this paper in Section 7.

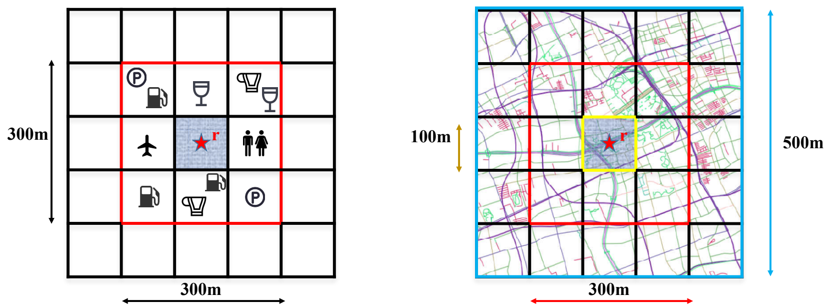
2. Overview

2.1. Definition and Problem

In this paper, we try to leverage both the restaurant data from online restaurant review platforms and the multi-source urban data to collaboratively predict rating scores for new restaurants.

Definition 1 (Restaurant Data D) *We consider a set D of restaurant records. Each record $r \in D$ represents a real restaurant recording the information such as its location coordinates $l_r = (lon, lat)$, restaurant type t_r, average price p_r, the number of customer reviews n_r and the average rating score s_r.*

Definition 2 (POI Data P) *The POI data P consists of a set of POIs (point of interest). Each POI record represents a specific location, with its name, category, spatial coordinates and several auxiliary attributes.*



(a) Covering grid for r (b) Neighbor grids with different θ

Figure 2: Example of the covering grid and neighbor grids for restaurant r. Left part shows POIs in 300m($\theta = 3$) around restaurant r. Right part shows Road Networks in 100m($\theta = 1$, yellow), 300m($\theta = 3$, red) and 500m($\theta = 5$, blue) around r.

Definition 3 (Road Network Data R) *The road network data R consists of a set of linked road segments. Each road segment is represented by several points in the segment, the segment length, and the road type (e.g., motorway, pedestrian).*

Definition 4 (Satellite Light Data L) *The satellite light data L contains a set of points with light intensity observed by remote sensing satellites. Formally, we denote by $\delta(p)$ the intensity of point p on earth's surface.*

As described in Section 1, the rating score of any new restaurant is potentially decided by both endogenous (e.g., restaurant type) and exogenous factors (e.g., flourishment of the restaurant location). While the urban data such as POIs and road networks involve important exogenous variables, the rating score of a particular restaurant is mainly affected by its surrounding environment rather than diverse and noisy situations over a wider area. For example, the rating score of a Michelin-starred restaurant in the business center of a city is independent of the existence of polluting factories in the remote areas of the city. To capture such spatial locality effects, we focus on the urban data within the *covering grid* (abbrev. *grid*) and *grid neighbors* of a restaurant for its rating score prediction, which are formally defined as follows.

Definition 5 (Covering Grid g_r and Grid Neighbors $NB_\theta(g_r)$) *Given a restaurant $r \in D$, the* covering grid *of r is defined as a squared rectangle $g_r = (c, \lambda)$, where the center point c of the rectangle is exactly the location l_r of r, and λ is half of the side length. In addition to the covering grid, we also consider the* neighbor grids *$NB_\theta(g_r)$ of g_r with respect to a window size $\theta \in \mathbf{N}^+$. For any grid $g' = (c', \lambda) \in NB_\theta(g_r)$, the distance between c, c' along each spatial coordinate is a multiple of λ, and the multiple is within the range of $[1, \theta]$.*

Without otherwise specified, we set λ to 100 meters by default in this paper. Figure 2 provides an example of the covering grid of a restaurant, and its neighbor grids w.r.t. different window sizes, i.e., $\theta = 1, 3, 5$. In Figure 2(b), yellow rectangle covering area with

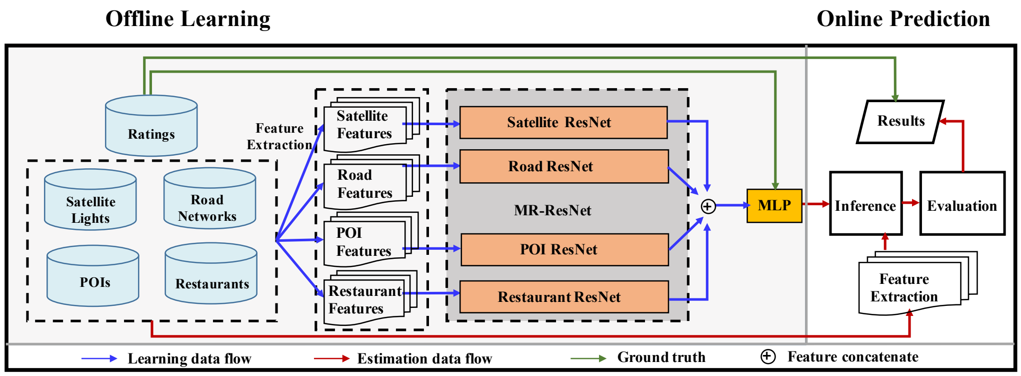


Figure 3: Proposed Framework

1 θ window size, red rectangle covering restaurant r itself and its neighbors within window size $\theta = 3$ and blue rectangle covering restaurant r and its neighbors in window size $\theta = 5$.
Problem Statement. Given restaurant data D and multi-source urban data P, R, L, we aim to predict the rating score s_r of any new restaurant $r \notin D$, where only the type and location of r are known.

2.2. Proposed Framework

Figure 3 provides the framework of our rating score prediction solution. It consists of two major components: offline learning and online prediction.

• **Offline Learning.** In the offline learning part, for each restaurant r, we first extract features from both restaurant and urban data with respect to the covering and neighbor grids of r. Specifically, we consider four groups of features: *restaurant features, POI features, road features, light features* extracted from four datasets. For each group of features, the corresponding feature values in each grid is represented by a vector. This allows us to obtain a feature map, regarding the covering and neighbor girds of r. Since different groups of features typically encode different information of a restaurant, we propose to employ individual convolution-based residual neural network to learn feature interactions for each group separately. The outputs from sub-networks are then fused to learn a unified representation that predicts rating score well. We use the actual rating scores from restaurant data as the ground-truth ratings. Our model is trained iteratively with the objective of minimizing the differences between predicted and actual rating scores.

• **Online Prediction.** The online prediction module tries to predict the rating score of a restaurant. Similar to offline learning, we first compute the feature maps for the restaurant. We then feed the maps into the trained MR-Net model and produce the predicted rating score as output.

3. Data Description and Feature Extraction

We start with describing the restaurant data from the real world restaurant review platform, Dianping. We then introduce the features extracted from the restaurant data and multi-source urban data for rating score prediction.

3.1. The Restaurant Dataset from Dianping

Dianping, often referred to as the "Yelp of China", was founded in 2003 and has now become a premier online review website that contains millions of restaurants in China. We crawled restaurant data from www.dianping.com and mainly focus on restaurants in Shanghai for empirical analysis in this study. Following Definition 1, we associate each restaurant with auxiliary information including the location, type, average price, the number of available reviews, and the overall rating score. We remove restaurants with incomplete information and obtain 24,331 restaurants in total. Without loss of generality, we assume that the overall rating score for each restaurant in the website matches its quality, but we allow using more trustful scores derived from textual reviews as indicated in Lei et al. (2016).

3.2. Feature Extraction from Restaurant Data and Urban Data

We first present the features used for rating score prediction. We then analyze the correlations between rating scores and features we extracted from the restaurant and the urban data. We focus on predicting the rating score of a restaurant r in location l_r with type t_r. To do this, we introduce several features extracted from the restaurant and the urban data, which are important indicators for the rating scores of restaurants. A vector of feature values is computed on a grid basis, i.e., covering grid of r or any of its neighbor grids. Let $G(r)$ denote $\{g_r\} \cup NB_\theta(g_r)$.

3.2.1. Restaurant features X_d^g.

Intuitively, the information on the nearby restaurants of r, especially those of the same type as t_r, reflects the potential popularity of r. We thus extract restaurant features for each grid $g \in G(r)$, including type-dependent and type-independent features. The type-dependent features include the *average rating score, average number of reviews, average price and the total number of restaurants with type* t_r in a grid. We also categorize rating score into 11 levels and measure *the restaurant count per rating level*, and the *entropy* based on the rating score distribution. Likewise, we compute the corresponding type-independent features by considering all the restaurants in the grid. We also consider citywide restaurant features, i.e., average rating score, average number of reviews, average price and the number of restaurants in the whole city area.

3.2.2. POI features X_p^g.

Generally, the number of POIs and the categorical information imply the functionality and the degree of prosperity of a grid. This is because high-variant POIs are often gathered in business centers, scenic spots and traffic hubs. For each grid $g \in G(r)$, we extract the *number of POIs in each category, total POI categories* and *category entropy* in g. In our POI dataset, Shanghai contains totally 724,304 POIs in 17 first-class categories and 122 second-class categories.

3.2.3. Road network features X_r^g.

The nearby road network structure reflects the traffic condition around the restaurants. In fact, in order to get popularity or highly rated, many high-quality restaurants are located

Table 1: Top-5 restaurant features with the highest Spearman, Kendall rank and Pearson correlation.

Feature	**Spearman**	**Tau**	**Pearson**
Average rating score with the same t_r in grid	0.89	0.87	0.55
Average rating score in grid	0.59	0.49	0.35
# Restaurants with 3 rating score in grid	-0.51	-0.47	-0.38
# Restaurants with 4.5 rating score in grid	0.42	0.40	0.44
Average rating score with the same t_r in city	0.41	0.32	0.40

in places with convenient traffic. In contrast, restaurants in remote areas are ranked with relatively lower scores. To capture the surrounding traffic condition, we measure the *number of segments in each type, total type number* and *type entropy* for each grid $g \in G(r)$ for restaurant r.

3.2.4. Satellite light features X_s^g.

Satellite light is a good indicator for areas with high prosperity degree or high-density commercial districts Jean et al. (2016). We use k-means method to cluster satellite lights and find city centers. We select the number of centers according to BIC Pelleg et al. (2000). As a result, we get 35 city centers in Shanghai from the satellite light data. For restaurant r, we calculate its distance to each city center and use the *minimum, mean and maximum distances* as the satellite features for r.

3.3. Empirical Analysis on Endogenous Features and Urban Features

3.3.1. Endogenous Interactions From Nearby Restaurants

To verify the effectiveness of the features in predicting rating scores of restaurants, we also compute the Spearman, Kendall rank and Pearson correlation coefficient for each feature against the rating scores in our restaurant dataset. Since rating scores of restaurants are rank order value, so we sort them by Spearman coefficient. Table 1 summarizes the restaurant features that report the top-5 values with the highest coefficient.

Considering a new restaurant without user's comments or feedbacks, a natural way to rating it is to leverage the rating scores of surrounding restaurants for reference, especially the nearby restaurants in the same type. The reason for such a prediction is that customers are likely to have a general assessment of a certain type of restaurants. For instance, most reviews for chain stores, such as Starbucks and McDonalds, are similar. Despite different store locations, they could get almost the same evaluations because of the analogous services and foods. On the other hand, it is observed that the number of highly rated restaurants, which get 4.5 stars or more in our marked grid, reflect the general standard of restaurants in the area. In reality, newly built restaurants are more likely to get a better rating if they are located in the region with many top-ranked restaurants and vice versa. To sum up, when assessing a new restaurant without user reviews, rating scores of surrounding and similar restaurants should be taken into account.

Table 2: Top 5 most attractive POI categories (left) and road types (right) for each rating group by Jensens inter-category coefficients.

Restaurant vs POI			
High Rating Score	5 ★	Bridge	2.839509
	4.5 ★	Coach station	2.209670
		Bus station	1.541488
		Port	1.538555
		Domestic service	1.485535
Low Rating Score	4 ★	Toll station	2.263459
	3.5 ★	Bridge	0.350564
	3 ★	Service area	0.220020
		Bus stop	0.218296
		Coach station	0.193498

Restaurant vs Road Network			
High Rating Score	5 ★	road	1.387220
	4.5 ★	motorway	0.907059
		motorway link	0.871768
		bridleway	0.846151
		tertiary	0.647711
Low Rating Score	4 ★	road	0.321133
	3.5 ★	motorway	0.122297
	3 ★	motorway link	0.105107
		unclassified	0.093082
		tertiary	0.090900

3.3.2. Spatial Interactions between Urban Features and Restaurants

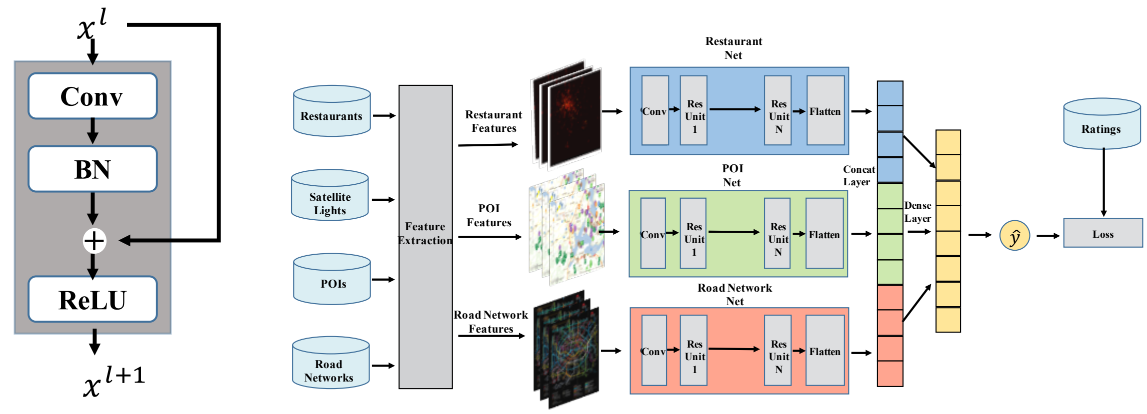
We use Jensen's inter-category coefficients Jensen (2006) to measure interactions between urban features and different rating restaurants. Jensen's inter-category coefficients was proposed by Jensen to assess retails' quality. The metric measures the attractiveness of two kinds of places, for instance, restaurants are near shopping centers and retail stores are near residential communities. Jensen coefficient greater than one represents two kinds of places attract with each other, otherwise, this two kinds of place repel each other.

We divided all restaurants into 2 groups according to their rating scores. We regard 5 star and 4.5 star restaurants as high rating restaurants, and the other restaurants are low rating restaurants. We selected POI and road network within 1km around each restaurant. We then calculate the attraction and repellent relationship of each type restaurant with 122 types of second-class POI and 33 types of roads respectively.

Table 2 left shows the result of the top 5 attractors for each rating group of restaurant corresponding to 122 second-class POI. We observe that restaurants with high rating scores are placed near bridges. This mainly because Shanghai is a coastal city. The city is run across by Huangpu River, and downtown areas are distributed along the Huangpu River or its tributaries. Meanwhile, high rating restaurants are gathered in downtown areas, so restaurants with high rating scores attract bridge. Moreover coach stations and bus stations both represent places with high popularity. As a result, high rating restaurants attract these places.

The Jensen coefficient of the toll station in low rating groups is large than high rating groups, which is due to that toll stations are mainly distributed in suburbs, and the average restaurant quality is lower than that in downtown. And restaurants with high rating scores are more attractive to all POIs than those have low rating scores. Due to the POIs are distributed denser in downtown than other places and restaurants with high rating scores are mainly gathered in downtown. Therefore, the Jensen's inter-category coefficients of high rating score restaurants are large than the lower ones.

We further measure the relationship between different types of roads with restaurants. The Jensen's inter-category is shown Table 2 right. Similar to POI, restaurants with high rating scores have greater attractions than lower ones for each type of roads.



(a) Residual Unit (b) Multi-source Rating Net

Figure 4: Residual Unit and Model Structure

4. Predicting Rating Scores Using Neural Networks

4.1. Spatial Dependencies from Endogenous and Urban Features

In the previous section, since the top 4 features in Table 1 is neighbor related features, the rating of a new restaurant is affected a lot by its neighbor. In order to capture the relationship between restaurant rating scores and endogenous features, we use the convolutional neural network on our grid-based input. Similar to endogenous features, urban data especially Road Network and POI have obviously spatial structures. So we use the same structure CNN to capture urban features. Inspired by He at el. He et al. (2016) we add the residual block to improve the performance. Residual unit in our model is shown as follow.

Residual Unit The residual unit is composed of two block types. The first type is convolutional block in which the input size is typically different from the output size. The second type is the identity block which has the same size between input and output. The structure of a residual unit in our model is shown in Figure 4(a). For identity block, since the input size and output size are identical, we add the input to the output directly. Different from identity block, convolutional block needs an extra convolutional layer to transform input to output. Formally, the residual unit can be defined as follow:

$$\begin{aligned} X_c^{(l+1)} &= Convolution(X^l) \\ X_b^{(l+1)} &= BatchNormalization(X_c^{(l+1)}) \\ X_d^{(l+1)} &= F(X^l) + X_b^{(l+1)} \\ X^{(l+1)} &= ReLU(X_d^{(l+1)}) \end{aligned} \tag{1}$$

where l is the layer number and $l \in [1, L]$, X^l is the input of layer l, $X^{(l+1)}$ is the output of layer l. We first perform batch normalization over the output of convolutional operation. We then add the input of layer l to the result of BN. If the unit is an identity block, we use $F(x) = x$. Otherwise, we use $F(x) = Conv(x)$, where the output shape of $Conv(x)$ is same as $X_b^{(l+1)}$. At last, we apply $ReLU$ activation function to get the unit output.

4.2. Multi-way Rating Net for Restaurant Rating Prediction

Due to the different distribution of restaurant features, road network features, and POI features, the combinations of different sources of info at the early stage of CNN not work well. To address this problem we use three different CNNs to capture different relationships between rating scores and our features. Figure 4(b) shows the structure of our Multi-way Rating Net(MR-Net), which has three sub-Nets: Restaurant Net (Net_d), POI Net (Net_p), and Road Network Net (Net_r), to absorb the corresponding input features. Since *Satellite Light Features* include distances from a restaurant to light centers, we combine light features with *Restaurant Features*. Now that we have three groups of features, we organize each group using an image-like feature map by regarding the covering and neighbor grids for a particular restaurant. The three Nets have the same structure, each of which consists of a convolutional layer followed by a sequence of residual units. Net_p and Net_r are used to find the effects of urban source data on rating score. Net_d is used to find the relations between rating scores and the intrinsic restaurant properties. As shown in Figure 4(b), for each restaurant r_i to be rated, we first transform three set of features described in Section 680 into a multi-channel image-like matrix over girds $G(r)$. We then feed these features into different Net accordingly.

4.2.1. Merge Layer

After obtaining the outputs from three Nets, we concatenate these latent features into a unified representation and feed it to a fully connected layer. The output of this merge layer is the predicted rating score for the restaurant. Formally, this layer can be defined as follows.

$$\begin{aligned} X_t = X_d^{N+1} \oplus X_p^{N+1} \oplus X_r^{N+1} \\ \hat{y} = F_t(X_t; \alpha_t) + b_t \end{aligned} \tag{2}$$

where $\oplus$ means the concatenate operation among vectors and F_t is a fully connected layer.

4.2.2. Model Training and Optimization

Because rating scores are numerical values, the restaurant rating problem can be treated as a regression problem. For the regression task, we use the total squared error as the objective function:

$$L_{loss} = \sum_{\mathbf{x}_r \in \mathbf{X}} (\hat{s_r}(\mathbf{x}_r) - s_r(\mathbf{x}_r))^2 \tag{3}$$

where $\mathbf{X}$ is the set of training data, $\mathbf{x}_r$ contains the set of input features for restaurant r, s_r is the ground-truth rating score of r, and $\hat{s}_r$ is the predicted rating score.

To optimize our objective function, we adopt Adam Kingma and Ba (2014) optimizer to train our model. The Adam optimizer allows tuning the learning rate dynamically during the training process. It also leads to faster convergence.

The overfitting problem causes smaller training error but larger test error. To prevent overfitting, we add L2 regularization term on our objective function, which has been widely used in machine learning models. As a result, the actual objective function we optimize is:

$$L_{loss} = \sum_{\mathbf{x}_r \in \mathbf{X}} (\hat{s_r}(\mathbf{x}_r) - s_r(\mathbf{x}_r))^2 + \lambda \|W\|^2 \tag{4}$$

where W denotes all the parameters in our model and λ is a hyperparameter to control the regularization strength.

5. Experiments

In this section, we present an experimental evaluation of the method proposed in this paper. The goals of our study are:

- To compare the performance of our approach with several baseline solutions: we consider several regression models including KNN, SVR, RandomForest, MLP and Single Net model.
- To evaluate the effectiveness of different features: we evaluate the prediction performance of regression models using different kinds of input features.
- To evaluate the effects of different values of parameters involved in our approach.

5.1. Datasets

Our experiments used Dianping dataset and multi-source urban data including POI, road network, and satellite light. We consider all these datasets in Shanghai, 2017. The details of each dataset are the following.
Dianping Dataset Dianping dataset has 56,053 stores where nearly half of them are restaurants. These restaurant distributed in 283 categories. Each store contains the information like rating score, average price, geographic coordinates, comment number and category message.
POI Dataset We employ a POI database from Baidu Maps to extract *POI Features* for Shanghai. There are 724,304 number of POIs in 17 categories.
Road Network The road network data is from Open Street Map. There are 197,561 road segments distributed in 33 types.
Satellite Light The satellite light data from light pollution map. There are 28,955 light points in this dataset.

Table 3 provides the statistics of Dingping review dataset and multi-source urban data in Shanghai. We split these datasets(80% and 20%) into training data and testing data. When we train the MR-Net Model, we use 20% of training data as validation data in order to tune the hyperparameters.

Table 3: Summary statistics of the data.

DataSet	Properites	Statistics
Dianping DataSet	Number of restaurants	24,331
	Number of Rating Classes	11
BaiduMap POI	Number of POIs	724,304
	Number of Classes	17
Satellite Light	Resolution	50m
Baidu Road Networks	Number of Roads	197,561
	Number of Road Types	33
	Total length	95,000 km

5.2. Evaluation Metric

We predict the rating score of a new restaurant using our trained model, and the ground truth is obtained from its real rating score in Dianping dataset. The performance of different model is evaluated in terms of root-mean-square-error (RMSE), which is defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - \hat{s}_i)^2} \tag{5}$$

where N is the number of restaurants, s_i and $\hat{s}_i$ are ground truth and prediction respectively.

5.3. Experiment Setup

We combine satellite light features with restaurant features. As a result, we have 3 set of features, restaurant features, POI features and road network features. We train a regression model that predicts the rating scores of a restaurant. We compare our proposed method MR-Net with the following baselines. To get a robust result we use 5-fold cross-validation for model comparison and tuning.

gird$_{mean}$: We use the mean rating scores in the grid to be the predicted result.

K Nearest Neighbor(KNN): This method selects K nearest neighbor which is most similar in all of three set of features, and compute the average rating scores as result. We select k to be 5 in our experiment.

Support Vector Regression(SVR): This is a support vector machine based regression model.

Random Forest(RF): Random forest is tree based ensemble method, which fits a number of classifying decision tree to improve the predictive accuracy.

Multi-layer Perceptron (MLP): This is a basic neural network with two hidden layers with 100 neurons for each.

Factorization Machines(FM): FM Rendle (2010) captures interactions between variables using factorized parameters.

Wide&Deep: Wide&Deep Cheng et al. (2016) captures feature interactions by a linear part and captures unseen feature combinations by neural network.

Single Net: A single CNN model without multi-way.

5.4. Comparison Results and Effects of Different Features

We use experiments to compare different features' effect on rating scores. First, we only input *Restaurant Features*, which contain most rating information. Then we use the combination of the *Restaurant Features* with *POI Features* and *Road Network Features* to predict the rating score. The results are shown in Table 5.4.

Comparing these methods, MR-Net performs best in every combination of feature sets. MR-Net achieves 0.4%, 2.3%, and 7% performance improvement than second best model random forest on $X_d^g + X_p^g$, $X_d^g + X_r^g$ and $X_d^g + X_r^g + X_p^g$ respectively. Although random forest performs well on single *Restaurant Features*, the performance when giving urban source data is lower than our model. This mainly because in urban source data, some types of POIs or roads may not exist around a specific restaurant r. As a result, there are many 0 in input urban source data. Generally tree-based models have low performance when much 0 in data. That means simple random forest can't model *Urban features* with *Restaurant Features* well, it can't capture the correlation between *Restaurant Features* and *Urban Features*. The $grid_{mean}$ method performs worst because it only uses the geographic coordinates. No other exogenous information has been used, in each gird there maybe many restaurants whose rating scores have large standard deviation. Comparing the results on different combinations of features, we can see that our MR-Net can make full use of multi-source urban features. The more feature adds to the model, the better result we get.

Comparing MR-Net with MLP, we can see that the former method can achieve great improvement, which indicates that MR-Net is able to capture the relations among different geographically related features using convolutional operations. Practically, sharply increasing feature amount could result in the worse performance of MLP due to the underfitting problem. Although Wide&Deep can model interactions and combinations between different features, it is hard to learn spatial relationship from a long 1D feature vector. In addition, by means of convolutional neural network, the number of parameters could be reduced without loss of performance. Regarding the different distributions of multi-source data, it is also necessary to compare MR-Net with a Single Net model. Single Net get a worse result as the type of input features increase. MR-Net achieves 22.3% performance improvement than single Net when using the combination of *Restaurant Features*, *POI Features* and *Road Network Features*. The reason for this could be that single Net cannot capture features from different sources well. Simply put all these features into Net may lead to unreliable results.

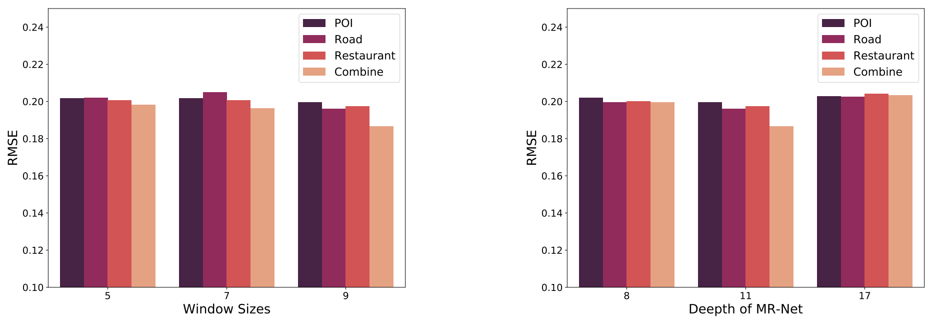
Compared 7 standard regression models with our method, we can see only MR-Net achieves lower $RMSE$ when using the combination of all the features. This means our method can learn complex relations between endogenous and exogenous factors which can hardly be modeled in classical regression methods.

Table 4: RMSE of Method on Different Features

Model	X_d^g	$X_d^g+X_p^g$	$X_d^g+X_r^g$	$X_d^g+X_p^g+X_r^g$
$grid_{mean}$	0.6369			
KNN	0.5270	0.5238	0.5216	0.5259
SVR	0.5559	0.5528	0.5508	0.5427
RF	0.1977	0.2005	0.2008	0.2007
MLP	0.3492	0.4055	0.5080	0.5816
FM	0.5299	0.5289	0.5390	0.5283
Wide&Deep	0.2803	0.3057	0.3562	0.3023
Net	0.1995	0.2173	0.2249	0.2402
MR-Net	**0.1974**	**0.1996**	**0.1961**	**0.1866**

5.5. Different Window Sizes in MR-Net

We study the performance of MR-Net model under different window size θ. Each gird is a 100m×100m square. We try three window sizes whose θ is 5, 7, 9. That means we test our method on 500m, 700m, and 900m squares. We use both single *Restaurant Features*, *POI Features*, and *Road Features* and their combination to test our model on different input sizes. The best result of each window size is shown in Figure $5(a)$. Observed from these results, we get worse results when the window size is very small. As the window becomes



(a) RMSE of Different Window Sizes. (b) RMSE of Different Depths.
Figure 5: Performance of different width and deepth.

larger, the RMSE is lower and get the best result at $\theta = 9$ which means the grid size is 900m×900m. Work Karamshuk et al. (2013) also point out that a point of interest can influence about 1km around it.

5.6. Effect of Different Depths of MR-Net

We evaluate the effect of different numbers of residual units in our method. We set default window size $\theta = 9$. The result is shown in Figure 5(b), we can find the performance of MR-Net11 is the best one. When the network becomes deeper the performance gets worse, this mainly because deeper networks have more parameters to learn, so we need more data. These deeper networks may be under-fitting due to the limited restaurant data.

6. Related Work

Rating Prediction. Work Tang et al. (2015) focus on performing semantic analysis to predict the rating score for each customer review. This paper exploits the effects of emotions in reviews to improve prediction accuracy. Beyond textual features, user information is also investigated in these works. McAuley et al. McAuley and Leskovec (2013) leverage both user profiles and textual features to predict rating scores, which can distinguish the same textual reviews from different customers. Our work is also related to the recommendation problem that leverage urban data such as POIs for predicting user preferences. Ye et al. Ye et al. (2011) consider the geographical influence for POI recommendation. Given user profiles, geographical features and POI dataset, the proposed method returns a list of recommended POI. We regard this as the rating scores for a list of POIs provided by the user. Matrix Factorization(MF) based methods are widely used in POI recommendation, including SVD++, and social MF Ma (2013). All of these works model the relationship between users and shops, however, there is little user information for a new restaurant. Moreover, since the customer review for new restaurants is missing, no textual information can be used to predict rating scores.

Residual Networks using Urban Source Data. Convolutional neural network is becoming more and more popular. Based on CNN Net allow us to train very deep network without gradient vanishing problem. Up to now, we can train more than 1000 layers deep network based on the residual neural network. Moreover, because urban shares the similar structure with an image, in Zhang et al. (2017), Zheng et al. applied residual network to traffic flow prediction. Both spatial and temporal factors have been considered in residual neural network. Wang et al. Wang et al. (2017) applied residual network on car demand prediction, and considered weather data and traffic data of a city. Albert et al. Albert et al. (2017) used residual network extract urban features from satellite images and solved the land use classification problem by using these features.

7. Conclusion

In this paper, we have tackled the problem of predicting rating scores for new restaurants, which is very important for customers to explore new restaurants and for investors to find appropriate business sites in the very beginning. We explore both online restaurant review data and urban data to identify endogenous and exogenous features that affect rating scores of restaurants. We conduct an empirical analysis to evaluate the correlation among features and propose an MR-Net model to absorb heterogeneous features in a unified way and capture feature interactions automatically. We evaluate the performance of the proposed approach using the real world online restaurant review data from Dianping. The results show that (1) our approach outperforms 8 baseline methods by achieving higher prediction accuracy, and (2) the utilization of urban features boosts the prediction performance.

Acknowledgments

This research is supported in part by 973 Program (No. 2014CB340303), and NSFC (No. 61772341, 61472254, 61572324 and 61170238). This work is also supported by the Program for the Program for Changjiang Young Scholars in University of China, the Program for China Top Young Talents, and the Program for Shanghai Top Young Talents.

References

Adrian Albert, Jasleen Kaur, and Marta C Gonzalez. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In *SIGKDD*, pages 1357–1366, 2017.

Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. Wide & deep learning for recommender systems. In *DLRS*, pages 7–10, 2016.

Qiwei Gan and Yang Yu. Restaurant rating: Industrial standard and word-of-mouth–a text mining and multi-dimensional sentiment analysis. In *HICSS 2015*, pages 1332–1340, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

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

Dmytro Karamshuk, Anastasios Noulas, Salvatore Scellato, Vincenzo Nicosia, and Cecilia Mascolo. Geo-spotting: mining online location-based services for optimal retail store placement. In *SIGKDD*, pages 793–801, 2013.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Xiaojiang Lei, Xueming Qian, and Guoshuai Zhao. Rating prediction based on social sentiment from textual reviews. *IEEE Transactions on Multimedia*, 18(9):1910–1921, 2016.

Jovian Lin, Kazunari Sugiyama, Min-Yen Kan, and Tat-Seng Chua. Addressing cold-start in app recommendation: latent user models constructed from twitter followers. In *SIGIR*, pages 283–292, 2013.

Hao Ma. An experimental study on implicit social recommendation. In *SIGIR*, pages 73–82, 2013.

Julian McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *RecSys*, pages 165–172, 2013.

Susan M Mudambi, David Schuff, and Zhewei Zhang. Why aren't the stars aligned? an analysis of online review content and star ratings. In *HICSS 2014*, pages 3139–3147, 2014.

Dan Pelleg, Andrew W Moore, et al. X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML*, volume 1, pages 727–734, 2000.

Steffen Rendle. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 995–1000, 2010.

Suvash Sedhain, Scott Sanner, Darius Braziunas, Lexing Xie, and Jordan Christensen. Social collaborative filtering for cold-start recommendations. In *RecSys*, pages 345–348, 2014.

Duyu Tang, Bing Qin, Ting Liu, and Yuekui Yang. User modeling with neural network for review rating prediction. In *IJCAI*, pages 1340–1346, 2015.

Dong Wang, Wei Cao, Jian Li, and Jieping Ye. Deepsd: Supply-demand prediction for online car-hailing services using deep neural networks. In *ICDE 2017*, pages 243–254, 2017.

Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, pages 325–334, 2011.

Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, pages 1655–1661, 2017.

Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *TIST*, 5(3):38, 2014.