BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 05/05 pp143–152
Volume 3, Number 2, June 2020
DOI: 10.26599/BDMA.2019.9020023

Text-Based Price Recommendation System for Online Rental Houses

Lujia Shen*, Qianjun Liu, Gong Chen, and Shouling Ji

Abstract: Online short-term rental platforms, such as Airbnb, have been becoming popular, and a better pricing strategy is imperative for hosts of new listings. In this paper, we analyzed the relationship between the description of each listing and its price, and proposed a text-based price recommendation system called TAPE to recommend a reasonable price for newly added listings. We used deep learning techniques (e.g., feedforward network, long short-term memory, and mean shift) to design and implement TAPE. Using two chronologically extracted datasets of the same four cities, we revealed important factors (e.g., indoor equipment and high-density area) that positively or negatively affect each property's price, and evaluated our preliminary and enhanced models. Our models achieved a Root-Mean-Square Error (RMSE) of 33.73 in Boston, 20.50 in London, 34.68 in Los Angeles, and 26.31 in New York City, which are comparable to an existing model that uses more features.

Key words: price recommendation; natural language processing; sentence embedding; Long Short-Term Memory (LSTM); mean shift

1 Introduction

The sharing economy has changed users' consumption behaviors in the past decade. Notably, the emergence of short-term online rentals redefines the lodging business. Airbnb, which is available in over 34 000 cities with 1.5 million hosts and 50 million guests[1, 2], is one of the most well-known online platforms in this industry for people to discover and book unique accommodations around the world[3]. On this platform, hosts can flexibly provide lodging information (e.g., pictures, location, description, neighborhood, and room information) and price their listings, as shown in Fig. 1. Therefore, a better pricing strategy, along with suitable wordings, may result in more interest from guests, and even a better profit on Airbnb.

Several studies focus on providing better pricing strategies or constructing price recommendation systems from different aspects. Price is the most critical factor that affects guests' decisions to choose a specific listing or not[4–6]. While a lower price may decrease income, a higher price may lose customers. Also, location-based information (e.g., point of interest, and neighborhood) is used in Refs. [7–9]. For example, the distance between attractions and their neighboring listings conversely affects the listed prices[8]. However, such studies do not consider other key factors (e.g., room type and the number of rooms). Moreover, review scores play an essential role in most existing pricing strategy models[10].

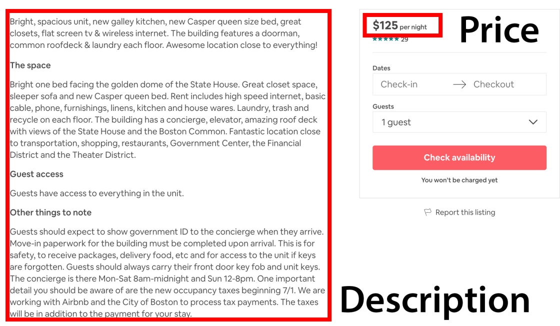


Fig. 1 Listing sample with description and price on Airbnb.

• Lujia Shen, Qianjun Liu, and Shouling Ji are with the Department of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China. E-mail: {shenlujia, liuqj0522, sji}@zju.edu.cn.

• Gong Chen is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA. E-mail: gchen@ece.gatech.edu.

∗ To whom correspondence should be addressed.

Manuscript received: 2019-12-09; accepted: 2019-12-10

However, the above-mentioned works fail to consider the complicated situation for listings without reviews.

We believe that the majority of hosts list lodging information for their rentals in a reasonable manner, that is, hosts make an effort to detail and price their properties to attract guests. According to our analysis, nearly all listings in the dataset provide detailed descriptions (e.g., house information, surrounding environment, and nearby attractions). Therefore, we aim to explore the relationship between the description of a listing and its price and then build a price recommendation system accordingly. Hosts of new listings need to come up with competitive and appropriate prices. In this paper, we used deep learning techniques to reveal the relationship and make price predictions. We vectorized each description with sentence embedding and then fed the vector into a feedforward network. Afterwards, we transferred the trained sentence embedding component to the enhanced model, which considers the outputs of both sentence embedding and location clustering components and also supplementary data (e.g., room type, bed type, and cancellation policy) as inputs for another feedforward network.

To evaluate our models, we employed two chronologically extracted Airbnb datasets from the same four cities (i.e., Boston, London, Los Angeles, and New York City), which total 207 000 distinct listings. During our study, we faced two challenges: (1) How can we thoroughly utilize the detailed description on each listing page; and (2) How can we only apply location information from the datasets instead of incorporating data from elsewhere. We tested four sentence embedding methods (i.e., Long Short-Term Memory (LSTM), BiLSTM, Self-Attention, and SIF) to convert the description of each listing into a vector. Also, we used mean shift to create location clusters as influence areas of attractions.

Our experiments reveal two points. First, a relationship exists between the description of each listing and its price. Factors that have a positive correlation to price, include indoor equipment (e.g., heater, microwave, stove, and TV), laundry, bar, and some influence areas of attractions (e.g., Freedom Trail, Boston Common, and Back Bay in Boston). Whereas, high-density area (e.g., Chinatown and Downtown Crossing in Boston) has a negative effect. Second, our LSTM-based preliminary model provides comparable predictions to another method[9]. The enhanced model obtains a better outcome, that is, the final Root-Mean-Square Error (RMSE) of 33.73 in Boston, 20.50 in London, 34.68 in Los Angeles, and 26.31 in New York City. As for those predicted prices having an offset of $ 200, we deduced two reasons: (1) For those higher than the real prices, hosts may be willing only to rent a room, but provide the descriptions related to their whole houses; and (2) For those lower than the real prices, it may be due to a lack of detailed descriptions for the listings.

The contributions of our paper are threefold.

(1) We are the first to employ text information, mainly only the description of every listing, to predict lodging prices. The outcomes are comparable to those of an existing work that uses more features.

(2) We explore the relationship between a listing's description and its price, and uncover some key factors that either positively or negatively affect a listing's price.

(3) We build TAPE, a Text bAsed Price rEcommendation system, according to the above-mentioned vital findings. It may further help hosts better advertise and price their properties on online lodging marketplaces.

2 Background

In this section, we briefly talk about four sentence embedding methods (i.e., LSTM, BiLSTM, self-attention, and SIF), and a location clustering method.

2.1 Long short-term memory

As a special architecture of the Recurrent Neural Network (RNN), LSTM[11] is well suited for classifying, processing, and making predictions based on time series data. LSTM is developed to alleviate the exploding and vanishing gradient problems that happen when training traditional RNNs. In this paper, we employed LSTM as the base model.

Bidirectional LSTM (BiLSTM) is designed to maintain contextual features from the past and future. Unlike LSTM, which has only one forward layer, the BiLSTM network has two parallel layers propagating in forward and backward directions, thus allowing both past and future information to be utilized and memorized in the cell[12].

2.2 Transformer

Transformer[13] is based solely on the attention mechanisms. Its encoder is composed of a stack of six identical layers. Each layer has two sublayers, including a multi-head self-attention mechanism and a simple feedforward network. Usually, the transformer has a significantly faster training speed, and achieves a better performance than RNN[14]. In this paper, we employed the encoder to transform a paragraph into new vectors, and fit these vectors into an LSTM network to obtain an embedding vector for the whole paragraph.

2.3 Smooth inverse frequency

Smooth Inverse Frequency (SIF)[15] is a simple sentence embedding method but performs pretty well in some complex-supervised learning problems. Given a sentence embedding vector $\boldsymbol{c}_s$, the probability of a word w being emitted in the sentence s is modeled by $Pr[w|\boldsymbol{c}_s]$. The probability of a sentence under prior sentence embedding can be calculated by multiplying all words in the sentence. Thus, sentence embedding can be estimated using maximum likelihood estimator.

2.4 Mean shift

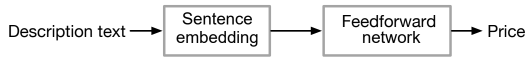
Mean shift[16] treats the points in the feature space as an empirical probability density function. The local maximum of the underlying distribution corresponds to the dense region in the feature space[17]. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the centroids of the distribution. Thus, data points associated with the same stationary point are considered as members of the same cluster.

3 Methodology

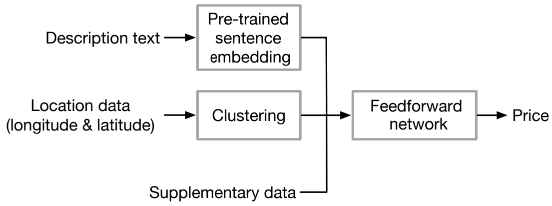
In this section, we discuss about the TAPE components. A feedforward network is used in the price prediction model because it outperforms in regression problems[18]. In the preliminary model, only the output of sentence embedding is used as an input of the feedforward network. Whereas in the enhanced model, two other components are prefixed to the feedforward network aside from the pre-trained sentence embedding model from preliminary. Here we elaborate on these three components (see Fig. 2).

3.1 Sentence embedding

Figure 2a depicts the preliminary model with only the sentence embedding component before a feedforward network. The sentence embedding model is used to convert description paragraphs to vectors, which serve as the input to a feedforward network. We tested the four different models (i.e., LSTM, BiLSTM, self-attention, and SIF), and selected the optimal one for further



(a) Preliminary model



(b) Enhanced model

Fig. 2 Model structures.

analysis. For the first three models, all words in description paragraphs are mapped to different word embedding vectors. Afterwards, word embeddings are trained with the entire model. Whereas, pre-trained word embedding is used for the SIF model. For the first three models, we obtain a trained preliminary model while getting the word embedding.

3.2 Location clustering

According to Ref. [8], listings are likely to cluster together around different attractions, and attractions affect their surroundings' listed prices. Therefore, clusters based on attractions are formed. Unlike Ref. [9] that used attraction information from TripAdvisor, we used the mean shift clustering[19] with the longitude and latitude of the listed properties to find the clusters. On the basis of the longitude and latitude of these listings, mean shift[19] is used to find the clusters of these listings. Similar to the method used in Ref. [20], these clusters represent a group of listings located within the vicinity of a landmark. Therefore, for each listing, location information is converted to a one-hot vector that consists of 0 s in all clusters except for a 1 in the cluster to which the listing belongs.

3.3 Other supplementary information

Table 1 lists all other useful features that were not extracted from the listing-related description. It is worthy

Table 1 Available information.

Available feature	Value
Room_type	Entire home/apt, etc.
Number of bathrooms	Numeric
Number of bedrooms	Numeric
Number of beds	Numeric
Bed_type	Real bed, futon, etc.
Host_is_superhost	Boolean
Host_has_profile_pic	Boolean
Host_identity_verified	Boolean
Cancellation_policy	Moderate, flexible, etc.

to mention that data, such as reviews, square feet of a listed house, number of guests allowed, and minimum nights, are not included. Reviews are not available for new listings, and other information is not accessible at the very start.

4 Evaluation

4.1 Datasets

We chronologically extracted datasets of the same four cities (i.e., Boston, London, Los Angeles (LA), and New York City (NYC)) from a website with Airbnb listings[21] within July 14 and September 14, 2019. Rationales behind choosing these cities are: (1) the data amounts of the four cities are different, thereby allowing us to evaluate TAPE's performance under different data amounts; and (2) two cities among the four were previously studied, so that we can compare TAPE's performance with that of the models proposed by other studies. The datasets are full of information (e.g., description, location, host, price, and house/room) with a total of 106 features in total. Figure 1 displays the information of facilities in a house (e.g., queen size bed, flat-screen TV, and wireless Internet), the nearby attractions and buildings (e.g., Boston Common and State House), and other information.

Table 2 displays the statistics of our dataset. The remaining listings exclude listings with empty descriptions and price outliers from the total listings. For example, over 95% of the listings in Boston are priced below 500 dollars; therefore, we neglected the remaining 5%. We split the dataset acquired on July 14, 2019 into 80% of training data and 20% of validation data. Furthermore, for the most recent one, we considered the listings without reviews as new listings and used them as validation data.

4.2 Preliminary model

In this section, we introduce the model setting of different sentence embedding methods first. Next, we analyze the effect of words/phrases on price. Lastly, we show two cases that cannot be predicted correctly by our model.

Table 2 Descriptive information of four different cities.

	Total listings	Remaining listings	Price range ($)
Boston	6241	5908	⩽499
London	83 850	77 899	⩽315
LA	44 620	41 418	⩽550
NYC	48 895	45 713	⩽358

4.2.1 Model selection

In the preliminary model, only the sentence embedding component is used before the feedforward network. Table 3 compares the performance of all four sentence embedding methods in the four cities by using RMSE as the evaluation metric to represent the difference between predicted values and true values. We chose the best model when the validation error reached the minimum.

For LSTM, we embedded every word of a sentence into a 512-dimensional vector, which returns a 512-dimensional hidden output as the final representative of the sentence. The 512-dimensional vector is then regressed along with a ReLU activation function to predict the listing's price. The setting of BiLSTM is similar to that of LSTM, and the only difference lies in the output dimension, which is twice the output of LSTM.

Similar to LSTM, we embedded every word of a sentence into a 512-dimensional vector in the self-attention model. Afterwards, we fed 512-dimensional embedded words to the encoder of a transformer. As mentioned previously, it consists of six identical layers, which include a self-attention block and a fully connected network. Finally, it outputs new embedded vectors. We took the average of these embedded words, and input them to a linear projection to predict the listing's price.

For SIF, the pre-trained word vector named GloVe[22] was used to calculate the expected sentence embedding. The SIF model generates sentence embedding of a 300-dimensional vector from a maximum likelihood estimator, given a known word embedding vectors. We then input this sentence embedding to a multilayer perceptron with two hidden layers to predict the price of a listing.

In Table 3, we observe that the LSTM model outperforms the other three models, and the LSTM model is also comparable to another model[9]. As an improved version of LSTM, the result of BiLSTM is contradictorily worse than that of LSTM in some cases. The self-attention mechanism has a comparable performance with SIF, whereas both models do not

Table 3 RMSE of the four embedding models for the four cities.

	LSTM	BiLSTM	Self-attention	SIF	Li et al.[9]
Boston	39.85	46.24	60.91	62.40	N/A
London	20.81	21.20	48.21	43.11	31.48
LA	35.21	69.06	68.79	71.31	33.89
NYC	26.72	25.27	57.34	55.47	N/A

perform better than LSTM in all cities. Besides, Li et al.[9] only focused on clustering houses with location information and review scores to predict prices. However, due to newly added listings, we could not use review scores for TAPE. Therefore, we selected LSTM for the sentence embedding component in our further analysis.

Figure 3 depicts our predicted prices versus ground truth prices for the four cities within scatterplots. Each green line is the diagonal that represent where the predicted price equals the true price, and each red line is the regression line of predicted prices under true prices. Most of the predicted prices are close to the true values. In Fig. 3a, the points are located alongside the green line, it implies that the listings' descriptions relate to their prices. Similarly, for Figs. 3b–3d, most points aggregate close to the green line. This evidence indicates that our model effectively finds the relationship.

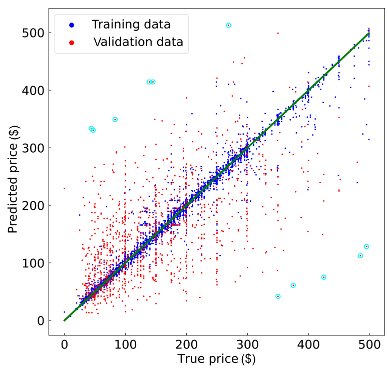
4.2.2 Relationship

To look into how words/phrases affect the predicted price, we conducted three steps after obtaining an original predicted price. (1) We masked a word in the sentences of a listing's description, and used these sentences to predict a new price. (2) We subtracted the original predicted price from the new price. If the difference is negative, then we considered the word as positive in the sentence; otherwise, the word is negative. However, we ignored the word that generates a difference of less than 0.01, due to its minor influence on price prediction. (3) We counted all positive words C_p and negative words C_n in our dataset, and labeled the word as positive if $C_p - C_n > \min\{C_p, C_n\}$, or negative if $C_n - C_p > \min\{C_p, C_n\}$. Otherwise, we labeled it as neutral. After finishing the above steps, we started over, and moved forward to the next word.

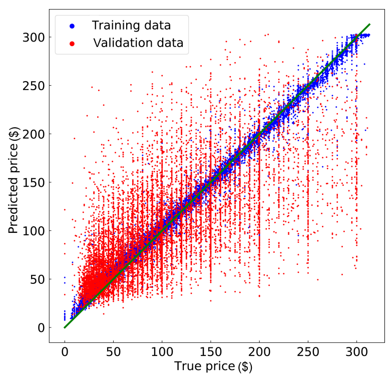
Table 4 lists a few keywords/phrases which may positively or negatively affect a listing's price in Boston.

Table 4 Positive and negative words or phrases.

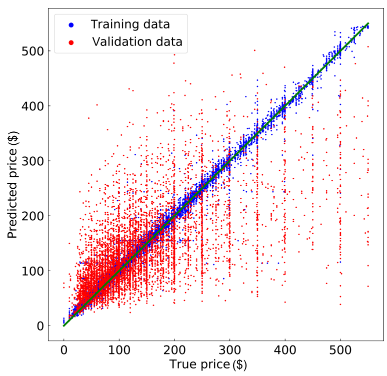
Word	Positive counts	Negative counts	Property
Freedom Trail	52	0	Positive
Boston Common	41	17	Positive
Back Bay	111	50	Positive
Chinatown	14	47	Negative
Downtown Crossing	0	31	Negative
Fenway	250	245	Neutral
Buses/shuttle	34	0	Positive
Heating	23	0	Positive
TV	274	109	Positive
Luxury/luxurious	246	106	Positive
Mini	0	42	Negative



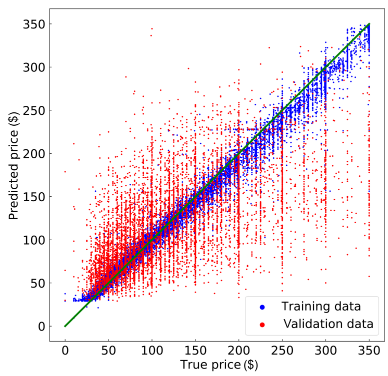
(a) Boston



(b) London



(c) Los Angeles



(d) New York City

Fig. 3 Predicted value vs. true value using LSTM.

Our model treats some notable scenic spots (e.g., Boston Common and Freedom Trail) as positive words/phrases. Such words/phrases lead to higher predicted prices, whereas, places such as Downtown Crossing and Chinatown have negative effects on the surrounding's listed prices. We deduced that people might have an impression that these places are crowded and untidy. Also, Fenway Park and some other regional spots are considered neutral. Aside from advertised spots in hosts' descriptions, TAPE also considers public transportation (e.g., buses and shuttles) and indoor equipment (e.g., heating and TV) as positive signs. Finally, TAPE can distinguish how adjectives influence a listing's price. For example, "luxury/luxurious" is positive, but "mini" is negative.

4.2.3 Case studies

For the Boston dataset, we obtained 37 outlier cases that have differences of over \$ 200 between predicted prices and true prices. We look into two examples to reveal the causes.

First, for a \$ 75 listing, TAPE predicts a price of \$ 300. By inspecting the following description, we found that \$ 300 would be reasonable for the entire house. However, the primary cause of such a low price may be due to the listing's distance to the downtown area. '*Free street parking. Private 3rd fl space, 3 bds, in well kept home. 2 bdrms, with full sz bds, rear bdrm also has twin bd. Futon sofa bd in Den. Good place to save, enjoy and explore Boston. Lots within a 10 minute walk: Grocery, laundry, cleaners, package store, pharmacy, coffee shops, existing and newly opened restaurants. Walk to bus stop or Trolly at Central Ave Milton which connects to Mattapan Station buses and Ashmont's redline Train. D'town Boston is approx 10 ml north. Please book the number of guest that will be staying. Private, compact apartment on the 3rd floor of home. Features: 2 bedrooms with 3 beds, each with smart TVs, 1 gigabit internet/WiFi...*'

Second, for a \$ 459 listing, TAPE predicts a price of \$ 150. In the following description, the host provides only the surrounding environment without mentioning any details about the house. Therefore, we believe that more details in the description will eliminate the issue. '*Enjoy the staying at one of the most beautiful building in Boston close to public transportation and the green line subway, restaurants, bars, and few steps from downtown Boston, enjoy the view of Charles river from your living room at 33rd floor, let the sun rays wake you up with the great view from your room, enjoy the amazing gym and the bbq area at the 35th floor and don't forget to get your free coffee from the great coffee machine at 2nd floor.*'

To sum up, detailed descriptions help hosts derive more accurate prices for their listed properties with TAPE. Other information, such as location and room type, may also aid TAPE to better price those listings. Therefore, we incorporated such information to create our enhanced model.

4.3 Enhanced model

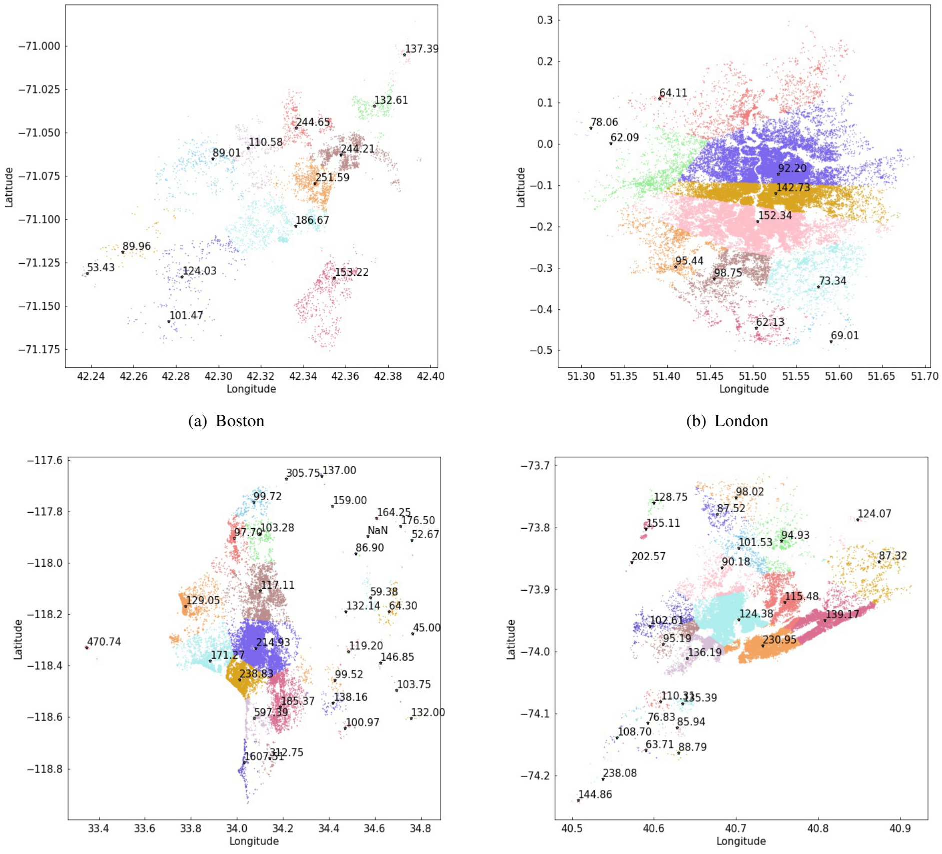
Figure 2b depicts the enhanced model with three components arranged in parallel before a feedforward network. The sentence embedding component is directly brought from our pre-trained sentence embedding component in the preliminary model. For the clustering component, we clustered location information (i.e., longitude and latitude) with mean shift in Fig. 4. Mean shift can automatically find clusters without presetting their number. Once found, we assign each listing a one-hot vector to represent the cluster to which it belongs, and then input it to the feedforward network. Also, other features, which are shown in Table 1, are fed into the network.

Table 5 (supp means supplementary data, as stated in Table 1) shows the RMSE results with different enhanced components. We observed from the dataset related to Boston that, any enhanced model built with new components greatly outperforms the preliminary model, and the result obtained by using the enhanced model with the location clustering component is more obvious. We deduced that, our enhanced model may greatly improve the performance when smaller datasets are used. In addition, the location clustering component has little side effects on the model performance for other datasets. Therefore, we decided to include all three components in the enhanced model for our further analysis.

Figure 5 shows all four cities' predicted prices versus true prices from the enhanced model. The dots closer to the green line indicate the perfectly predicted prices. Compared with the cyan dots in Fig. 3a, the outliers in Fig. 5a move closer to the green line. Such evidences reveal the effectiveness of using the two additional

Table 5 RMSE for two enhanced LSTM models.

	Boston	London	LA	NYC
With supp	34.11	20.65	34.52	26.27
With supp & clusters	33.73	20.50	34.68	26.31



(a) Boston (b) London

(c) Los Angeles (d) New York City

Fig. 4 Price clusters ($).

components in the enhanced model.

We then looked into the cases studied in Section 4.2.3. With the use of the enhanced model, the predicted price of the first case dropped from $ 300 to $ 230; whereas, that of the second case increased from $ 150 to $ 165. The existence of location information may prevent our model from overfitting. Therefore, predicted prices can be slightly corrected.

Finally, we used the September 14, 2019 dataset to evaluate new listings. To simplify the task, we only considered the listings with no reviews as newly added listings. We retrained the enhanced model with the new dataset's listings that contained reviews, as the Airbnb released datasets regularly to remove nonexistent listings and to add new listings. Table 6 shows the RMSE results of the validation data. From the performance perspective, the retrained enhanced model does not differ from the previous one because no differences were found in the description with or without reviews.

Table 6 RMSE for newly added listings.

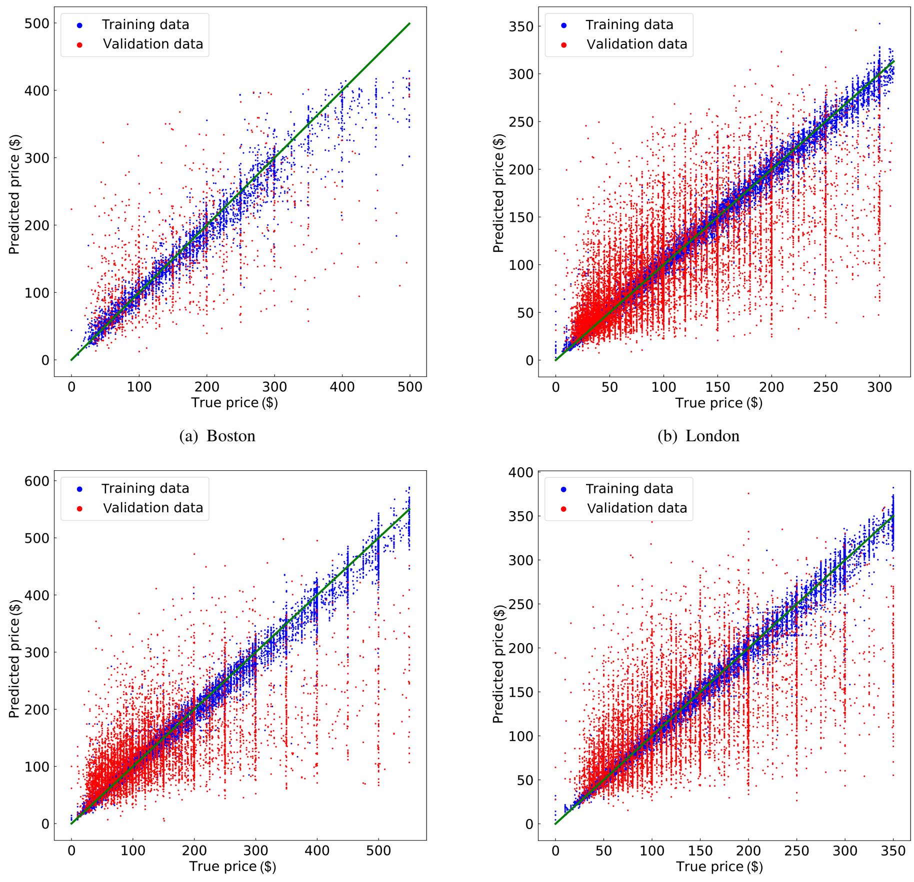
	Boston	London	LA	NYC
Enhanced model	65.24	44.37	76.34	56.94

5 Discussion

Two issues restrict the pricing studies for the lodging business: (1) In general, different cities may require different model adjustments; and (2) Specifically, different hosts may advertise their properties differently.

Modeling lodging prices must be based on cities. Different cities have their own city layout and attractions, population size, distribution, and economic situation. As a result, the number and average price of listings in a small town may differ from those in a big city. However, different cities may use the same names for their landmarks, such as SoHo and Chinatown. Therefore, we cannot have one universal model for all cities.

It is up to hosts themselves to edit the descriptions of their listed properties. Even though the majority of hosts reasonably publicize lodging information, a few may only leave very basic information or nothing. Usually, TAPE can precisely predict a listing's price with a description of more than 100 words. Therefore, according to our dataset extracted from Airbnb, more



(a) Boston (b) London

(c) Los Angeles (d) New York City

Fig. 5 Enhanced models.

than 90% of the hosts can use TAPE to price their listings.

6 Related Work

Predicting rental prices has been thoroughly studied by economists. Gallin[23] investigated the relationship between house prices and rents and concluded that rents and house prices tend to correct back to each other over three-year horizons. Also, Lee et al.[24] discovered that listing prices are significantly associated with sales.

Most research studied the price by using the house information, such as the number of bedrooms or bathrooms. Wang and Nicolau[10] used a total of 31 explanatory variables (e.g., host attributes, site and property attributes, amenities and services, rental rules, and online review ratings) to find the relationship between a price and its determinants. In Ref. [25], Choudhary et al. analyzed Airbnb listings in San Francisco to better understand how different attributes (e.g., bedrooms, location, and house type) can be used to accurately predict the price of a new listing, which is optimal in terms of the host's profitability yet affordable to their guests. These works studied price-related factors only and failed to provide any models for predicting rental prices.

In Ref. [7], Tang and Sangani labeled text information for nine handpicked classes, extracted image-related features, and finally used all features to predict a listing's neighborhood and its price. However, the price prediction model is merely a binary classifier that predicts whether listed prices are above or below the median price. Our goal is to provide reasonable prices instead of price categories. Also, researchers may use geographical information for better pricing strategies. In Ref. [8], Li et al. aggregated houses appropriately according to landmarks and facilities, and they used a linear regression model to predict prices. Overall, the above works are ineffective for newly added listings because review scores and comments are not available.

7 Conclusion

In this paper, we presented TAPE, a text-based price recommendation system for predicting a reasonable price for newly added listings. On the basis of our language model, we revealed the relationship between the description of a listing and its price. Experimental results on four cities show that, our LSTM-based preliminary model can adequately relate the description to its corresponding price, and is comparable to an existing model. Afterwards, we brought the pre-trained sentence embedding components, prefixed two other components to the enhanced model, and achieved an RMSE of 33.73 in Boston, 20.50 in London, 34.68 in Los Angeles, and 26.31 in New York City.

References

[1] K. Barron, E. Kung, and D. Proserpio, The sharing economy and housing affordability: Evidence from Airbnb, https://ssrn.com/abstract=3006832, 2018.

[2] S. Sheppard and A. Udell, Do Airbnb properties affect house prices? http://web.williams.edu/Economics/wp/SheppardUdellAir//bnbAffectHousePrices.pdf, 2016.

[3] G. Quattrone, D. Proserpio, D. Quercia, L. Capra, and M. Musolesi, Who benefits from the sharing economy of Airbnb? in *Proc. of the 25th International Conference on World Wide Web*, Montreal, Canada, 2016, pp. 1385–1394.

[4] Y. Chen and K. Xie, Consumer valuation of Airbnb listings: A hedonic pricing approach, *International Journal of Contemporary Hospitality Management*, vol. 29, no. 9, pp. 2405–2424, 2017.

[5] C. Gibbs, D. Guttentag, U. Gretzel, L. Yao, and J. Morton, Use of dynamic pricing strategies by Airbnb hosts, *International Journal of Contemporary Hospitality Management*, vol. 30, no. 1, pp. 2–20, 2018.

[6] G. Zervas, D. Proserpio, and J. W. Byers, The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry, *Journal of Marketing Research*, vol. 54, no. 5, pp. 687–705, 2017.

[7] E. Tang and K. Sangani, Neighborhood and price prediction for San Francisco Airbnb listings, http://cs229.stanford.edu/proj2015/236_report.pdf, 2018.

[8] Y. Li, Q. Pan, T. Yang, and L. Guo, Reasonable price recommendation on Airbnb using Multi-Scale clustering, in *Proc. of the 35th Chinese Control Conference (CCC)*, Chengdu, China, 2016, pp. 7038–7041.

[9] Y. Li, S. Wang, T. Yang, Q. Pan, and J. Tang, Price recommendation on vacation rental websites, in *Proc. of the 2017 SIAM International Conference on Data Mining*, Galleria Houston, TX, USA, 2017, pp. 399–407.

[10] D. Wang and J. L. Nicolau, Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb. com, *International Journal of Hospitality Management*, vol. 62, no. 1, pp. 120–131, 2017.

[11] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] U. Kamath, J. Liu, and J. Whitaker, Recurrent neural networks, in *Deep Learning for NLP and Speech Recognition*. New York, NY, USA: Springer, 2019, pp. 315–368.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in *Proc. of the Advances in Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 5998–6008.

[14] C. Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, C. Hawthorne, A. M. Dai, M. D. Hoffman, and D. Eck, An improved relative self-attention mechanism for transformer with application to music generation, arXiv preprint arXiv: 1809.04281, 2018.

[15] S. Arora, Y. Liang, and T. Ma, A simple but tough-to-beat baseline for sentence embeddings, presented at International Conference on Learning Representations, Palais des Congrès Neptune, Toulon, France, 2017.

[16] K. Fukunaga and L. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[17] K. G. Derpanis, Mean shift clustering, *Lecture Notes*, http://www.cse.yorku.ca/%7Ekosta/CompVis_Notes/mean_shift.pdf, 2005.

[18] G. K. Tso and K. K. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, *Energy*, vol. 32, no. 9, pp. 1761–1768, 2007.

[19] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[20] L. Cao, J. Luo, A. Gallagher, X. Jin, J. Han, and T. S. Huang, A worldwide tourism recommendation system based on geotagged web photos, in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, 2010, pp. 2274–2277.

[21] Airbird Inside Airbnb Data, http://insideairbnb.com/get-the-data.htm, 2019.

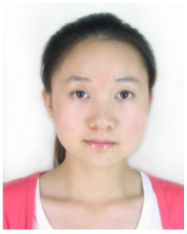
[22] J. Pennington, R. Socher, and C. Manning, Glove: Global vectors for word representation, in *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543.

[23] J. Gallin, The longrun relationship between house prices and rents, *Real Estate Economics*, vol. 36, no. 4, pp. 635–658, 2008.

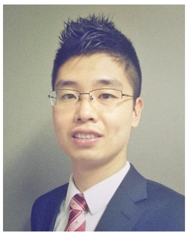
[24] D. Lee, W. Hyun, J. Ryu,, W. J. Lee, W. Rhee, and B. Suh, An analysis of social features associated with room sales of Airbnb, in *Proc. of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA, 2015, pp. 219–222.

[25] P. Choudhary, A. Jain, and R. Baijal, Unravelling Airbnb predicting price for new listing, arXiv preprint arXiv: 1805.12101, 2018.

Lujia Shen received the MS degree from Northeastern University, Boston, USA, in 2019, and now he is a research assistant in the College of Computer Science and Technology, Zhejiang University, Hangzhou, China. His research interests include data mining and analytics, machine learning robustness, and interpretability.



Qianjun Liu received the BS degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016, and now she is a PhD candidate in the College of Computer Science and Technology, Zhejiang University, Hangzhou, China. Her research interests include big data security, natural language processing, and code analysis.



Gong Chen is a postdoctoral fellow at the Georgia Institute of Technology, Atanta, USA. He received the PhD and MS degrees from Georgia Institute of Technology, the Diplôme d'Ingénieur degree in electronics and signal processing from Institut National Polytechnique de Toulouse-École nationale supérieure d'électrotechnique, d'électronique, d'informatique, d'hydraulique et des télécommunications, and the BEng degree in electronic science and technology from Beijing University of Posts and Telecommunications. His current research interests include Ad-related network security and privacy, big data security and privacy, IoT security, and adversarial machine learning.

Shouling Ji is a Zhejiang University 100-Young Professor in the College of Computer Science and Technology at Zhejiang University and a research faculty in the School of Electrical and Computer Engineering at Georgia Institute of Technology. He received the PhD degree from Georgia Institute of Technology in 2015 and the PhD degree in computer science from Georgia State University in 2013. His current research interests include big data security and privacy, big data driven security and privacy, and adversarial learning. He also has interests in graph theory and algorithms and wireless networks. He is a member of IEEE and ACM and was the membership chair of the IEEE Student Branch at Georgia State (2012–2013).