International Journal of Business Intelligence Research
Volume 11 • Issue 2 • July-December 2020

Recommendation and Sentiment Analysis Based on Consumer Review and Rating

Pin Ni, University of Liverpool, Liverpool, UK
https://orcid.org/0000-0003-4516-1249

Yuming Li, University of Liverpool, Liverpool, UK
https://orcid.org/0000-0003-2219-9033

Victor Chang, Teesside University, Middlesbrough, UK
https://orcid.org/0000-0002-8012-5852

ABSTRACT

Accurate analysis and recommendation on products based on online reviews and rating data play an important role in precisely targeting suitable consumer segmentations and therefore can promote merchandise sales. This study uses a recommendation and sentiment classification model for analyzing the data of beer product based on online beer reviews and rating dataset of beer products and uses them to improve the recommendation performance of the recommendation model for different customer needs. Among them, the beer recommendation is based on rating data; 10 classification models are compared in text sentiment analysis, including the conventional machine learning models and deep learning models. Combining the two analyses can increase the credibility of the recommended beer and help increase beer sales. The experiment proves that this method can filter the products with more negative reviews in the recommendation algorithm and improve user acceptance.

KEYWORDS

CNN, Consumer Reviews, LSTM, Rating, Recommendation Algorithm, RNN, Sentiment Analysis, Spark-ALS

1. INTRODUCTION

Online review and rating, as the two most important customer reference factors in online shopping platforms, have a greater influence on consumers' willingness to buy. At the same time, e-commerce platforms or online advertising agencies also need to use these data as a basis to make accurate recommendations or advertising for customers with different preferences. However, rating data cannot reflect the specific characteristics of the product. And in many cases, reviews that lack rating data often make it difficult to judge the user's specific tendency of the product (especially in the case of ambiguous, overly simplistic or worthless reviews). Therefore, how to comprehensively use these two types of data to support the construction of a more intelligent recommendation system is a subject worth exploring.

DOI: 10.4018/IJBIR.2020070102

This paper mainly analyses the data based on users' reviews on beer. However, the current researches focus on the intrinsic quality of the beer, and there are few researches on beer review and rating mining. Analyzing online reviews can not only help manufacturers develop products that are more in line with consumer preferences, but also promote sales. Therefore, the study is based on beer rating data and recommends beer products through the Spark-ALS collaborative filtering algorithm and compared 10 classification models including conventional machine learning and deep learning for consumer review analysis. Finally, a recommendation model based on Spark-ALS and LSTM was built to provide more accurate and credible recommendations.

Our main contributions are as follows:

1. We combine customers' review text data and rating data to support the construction of recommendation model and improve its effectiveness. Experiments have shown that our method achieves effective performance on beer product recommendation task;
2. Comparing 10 classifiers including mainstream conventional machine learning methods and deep learning methods for sentiment analysis task of recommendation model 3. We conducted a relatively comprehensive literature review of previous research in customer review mining, product rating analysis, and provided some technical and application analysis and suggestions for related business intelligence fields.

The rest of the article is structured as follows: Section 2 reviews the related works and is followed by the methodology in Section 3. The experiment description is presented in Section 4. Next, Sections 5 reports the result of the experiment. 5. Section 6 illustrates the limitations of the study, discusses, and analyzes the value of related tasks from both technical and application perspectives. Finally, the last section is the conclusion.

2. LITERATURE REVIEW

The reputation of the product has become an important factor for consumers to influence purchase intention. To a certain extent, it can be regarded as a filter in the current Internet environment with massive consumption information to help people make better decisions. Online reviews are one of the most important mediums that reflect the reputation of the product. And as an emerging field of Web information mining, online reviews' sentiment analysis involves a wide range of research topics, e.g., identifying the attributes of the products being reviewed, determining the attitudes of customers, and mining online reviews of products.

2.1. Mining for Customers' Review

Customers' review has become an essential reference for consumer consumption in today's product consumption (Duan, Gu, & Whinston, 2008). Therefore, the role of customers' reviews in business has attracted the attention of scholars in many fields. Customers' reviews are the most valued information for companies and manufacturers to understand customers' feedback on their products so that they could use this information to improve the quality of their products (Chong, Ch'ng, Liu, & Li, 2017). Customer reviews also can provide retailers with a better way to understand the specific preferences of each customer. Furthermore, consumers' review reflects the process of their purchase decisions, and by exploring the key factors that led them to purchase, it can help improve product sales (Wei, Chen, Yang, & Yang, 2010).

Park et al. (Park, Lee, & Han, 2007) used a likelihood model to explain how the degree of online product reviews and product participation affect consumer behavior, and the result was that the quality of reviews has a positive influence on consumers' purchase intentions, and low-engagement consumers are affected by the number, not the quality of the reviews. Ifrach et al. (Ifrach, Maglaras,

Scarsini, & Zseleva, 2019) proposed that in a certain price range, users usually use Bayesian models to infer product quality inversely based on product ratings, thereby conducting research on product pricing issues. Singh et al. (Singh et al., 2017) established a machine learning model that uses the characteristics of the text to predict the contributions of consumer reviews to potential users. The results of the study encourage buyers to write more effective reviews, thereby assisting other consumers in making purchasing decisions, and also help merchants improve their product websites. Salehan et al. (Salehan & Kim, 2016) conducted big data statistics and sentiment analysis on the evaluation of online users. The experimental results show that reviews with a higher degree of positive emotions will be read by more users, and the reviews with a neutral emotion in the text are also considered more helpful. Proserpio et al. (Proserpio & Zervas, 2017) analyzed user reviews in the hotel industry and concluded that when hotels respond to the customers, the hotel will receive fewer negative reviews. Lee et al. (Lee, Yang, Chen, Wang, & Sun, 2016) proposed an approach of mining perceptual mapping to automatically construct perceptual maps and radar charts from online consumer reviews. This approach can help related merchants to positioning new products' market and formulate corresponding marketing strategies. Additionally, customers' reviews may include not only text reviews but also product ratings and other sales information. Chen et al. (L. Chen, Li, Liu, Zhang, & Woodbridge, 2017) proposed a product recommendation algorithm based on Apache Spark. This machine-learning algorithm can recommend the most suitable product to users according to the customers' ratings. Filieri et al. (Filieri, Hofacker, & Alguezaui, 2018) used a detailed likelihood model to study consumer perceptions and found that long length reviews might not necessarily be helpful, while highly relevant reviews and products ranking scores were considered two important pieces of information.

2.2. About Sentiment Analysis in Online Reviews

According to different types of texts, sentiment analysis can be divided into subjective text analysis and objective text analysis (Witten, 2004), including text sentiment polarity analysis and text sentiment polarity intensity analysis. The polarity of sentiment is divided into positive and negative poles, and some scholars have joined the neutrality pole. Although this classification method is simple, it can meet the needs of most practical applications, such as judging whether consumers are positive or negative reviews of goods and whether they support or oppose some opinions. However, the multi-classification of sentiment is a difficult task in classification.

Cao (Cao, Duan, & Gan, 2011) analyzed the semantic features in the review text and found that the semantics in the review could more effectively influence consumers' decisions. Moreover, the reviews with more extreme language expressions were more influential. Jonathan et al. (Jonathan, Sihotang, & Martin, 2019) study the restaurant reviews in a restaurant scoring application named Zomato and used them for sentiment analysis. This article uses the term frequency-inverse frequency (TF-IDF) to create word feature. The accuracy of the positive, negative, and neutral emotions obtained in the experiments is 92%, 93%, and 96%, respectively. Ruder et al. (Ruder, Ghaffari, & Breslin, 2016) proposed using a hierarchical bidirectional LSTM model to model the content of customer reviews, and then used aspect-based sentiment analysis to process. The experimental results show that this model has obtained results that compete with the most advanced results and surpasses the most advanced technology on multilingual and multi-domain datasets. Zhang et al. (Zhang, Zhou, Duan, & Chen, 2018) proposed a bidirectional GRU-based sentiment analysis model for multi-label sentiment analysis tasks, and experiments proved the effectiveness of the model in computing efficiency. Chen et al. (H. Chen et al., 2018) used the LSTM network to perform fine-grained sentiment analysis on customer reviews on online shopping platforms. The experimental results achieved an accuracy of 90.74% and an F1 score of 65.47%, proving the feasibility and effectiveness of the LSTM network. In addition, the performance of LSTM networks in fine-grained sentiment analysis is significantly better than conventional machine learning methods. Jebbara et al. (Jebbara & Cimiano, 2016) divided the emotion analysis task into two sub-tasks: aspect extraction and specific sentiment extraction.

Compared with the conventional single-task sentiment analysis, this method is more flexible and more practical. It has been well verified in the ESWC-2016 semantic sentiment analysis challenge.

3. METHODOLOGY

The user review data in the dataset consists of two parts: the user's rating of the beer and the other part is the user's text review on the beer. The data mining of the data is also divided into two parts: one is to build a recommendation mechanism based on the rating data; the other is sentiment polarity analysis based on the textual review.

3.1. Analysis of Rating Data

The Spark-ALS based collaborative filtering recommendation algorithm is used to recommend beers to users. The ALS recommendation algorithm is a matrix-based decomposition method that considers both User and Item aspects (Xie, Zhou, & Li, 2016). In general, users only buy a minimal number of products in the item and score it. Such a rating matrix containing users and products is quite sparse. Using the matrix decomposition function in the Spark MLlib machine learning library (Meng et al., 2016), find a k-dimensional (low-order) matrix similar to the "user-item" matrix, which is a matrix of $m \times n$ is obtained by multiplying two matrices of $m \times k$ and $k \times n$ $(k << m, n)$:

$$A_{m\times n} = U_{m\times k} \times V_{k\times n} \tag{1}$$

These two matrices one for representing the user $m \times k$ -dimensional matrix, and a $k \times n$ -dimensional matrix that characterizes the item. These two matrices are called factor matrices, which are multiplied to get a rating for each product for each user.

The task of machine learning is to find $U_{m\times k}$ and $V_{k\times n}$. It can be seen that $u_i^T v_j$ is the preference of user *i* for commodity *j*, and the Frobenius norm is used to quantify the error generated by reconstructing *U* and *V*. Since many places in the matrix are blank, i.e., the user does not score the product, for this case we do not need to calculate the unknown element, only the observed (user, commodity) set *R* (Winlaw, Hynes, Caterini, & De Sterck, 2015).

U and *V* are coupled to each other in the objective function, so an alternating square algorithm is used. That is, the initial value $U^{(0)}$ of *U* is assumed first, so that the problem is transformed into a least-squares problem. $V^{(0)}$ can be calculated according to $U^{(0)}$, and $U^{(1)}$ is calculated according to $V^{(0)}$, this iteration continues until iterates a certain number of times, or converges:

$$A - UV^T \ {}_F^2 \rightarrow min \sum_{(i,j)\in R} \left(a_{ij} - u_i^T v_j\right)^2 \tag{2}$$

3.2. Sentiment Analysis for Review Data

The sentiment polarity analysis of product reviews can meet the basic needs of the platform or user in practical applications (Mostafa, 2018). According to the user's overall rating of the product, the sentiment polarity of the text data set is annotated, and the polarity of the sentiment is mainly divided into three types of positive, negative, and neutral. Pre-processing the already annotated text data set mainly removes duplicate data and fills in null values during the pre-processing process.

After the processed dataset is obtained, each piece of text is segmented, and the characters and stop words are removed. Use the Word2Vec, a word embedding model to vectorize words in the text.

Compared with the conventional natural language statistical modeling methods, it solves the problems of the dimensional explosion, word similarity and model performance to some extent (Liu, 2019).

The combination of the deep learning algorithm and text sentiment analysis can provide a better solution for sentiment polarity classification (Da Silva, Hruschka, & Hruschka Jr, 2014). This paper will compare the conventional machine learning classification model and the neural network classification model through experiments to obtain a better solution in the sentiment analysis of text reviews.

3.2.1. Convolutional Neural Network

Convolutional neural network (CNN) is a feed-forward neural network, which has an excellent performance in large-scale image processing and has been widely used in image classification, text classification, and other fields. The CNN is constructed by imitating the biological visual perception mechanism and can perform supervised learning and unsupervised learning. The sharing of convolution kernel parameters in the hidden layers and the sparseness of the connection between layers enable the convolutional neural network to learn grid-like topology features (e.g., pixels and audio) with a small amount of computation. And it has a stable effect and no additional feature engineering requirements for the data. The overall structure of Convolutional Neural Network mainly includes three layers:

- **Convolutional Layer:** This layer consists of filters and activation functions. Generally, the hyper-parameters to be set include the number, size, and step size of filters, and whether the padding is "valid" or "same", and what activation function is selected;
- **Pooling Layer:** The parameters of this layer have been set, such as MaxPooling or Average pooling. In addition, it is needed to specify the hyper-parameters, including whether it is Max or average, the window size, and the step size;
- **Fully Connected Layer:** This layer is a row of neurons. This layer is called "fully connected" because each unit is connected to each unit in the previous layer.

3.2.2. Recurrent Neural Network

Recurrent Neural Network (RNN) has the characteristics of parameter sharing, memory, and Turing completeness, and it has great advantages in learning the nonlinear characteristics of the sequence. As a significant model of deep learning, it has been widely used in the field of natural language processing (e.g., language modeling, machine translation, speech recognition, etc.) and time series related prediction tasks. The RNN constructed after introducing the convolutional neural network can be used to solve computer vision tasks with sequence input. RNN is a type of neural network used to process time-series data. Time-series data refers to data collected at different points in time; this type of data reflects the changing state of a thing or phenomenon over time or degree. The reason why RNN is called recurrent neural network is that the current output of a sequence is also related to the previous output. The specific manifestation is that the network memorizes the previous information and applies it to the current output calculation, that is, the nodes between the hidden layers are not connected but the layers are fully connected, and the input of the hidden layer includes not only the output of the input layer. It also includes the output of the hidden layer from the previous moment. It mainly consists of an input layer, a hidden layer, and an output layer.

Suppose $t-1$, t represents time series, X is the input sample, S_t is the memory of the sample at time t, W is the input weight, U is the weight of the input sample at the current moment, and V is the output sample weight. The final output value can be obtained by:

$$h_t = Ux_t + WS_{t-1} \tag{3}$$

$$s_t = f\left(h_t\right) \tag{4}$$

$$o_t = g\left(VS_t\right) \tag{5}$$

Among them, f and g are activation functions. And f can be $tanh$, $relu$, $sigmoid$ and other activation functions, g can be $softmax$ or other. And W, U, V are always equal, also defined as weight sharing.

3.2.3. Long Short-Term Memory

Long Short-Term Memory (LSTM) networks, a variant RNN network, is designed to solve the problem of long dependence (e.g., Forget the key information of long-distance context. It can also be regarded as "Gradient disappearance problem" of RNN). It is suitable for processing and predicting important events with relatively long intervals and delays in time series (e.g., a special reference in context). The LSTM network can delete or add information to the cell state through a structure called a gate, and the gate can selectively decide which information to pass. LSTM controls the state of cells by three gates, which are called forget gate, input gate, and output gate.

The first step in LSTM is to decide what information needs to be discarded from the cell state. This part of the operation is handled by a sigmoid unit called the forget gate. It uses h_{t-1} and x_t information to output a vector between 0-1. The 0-1 value in this vector indicates which information in the cell state C_{t-1} is retained or discarded.

The next step is to decide what new information to add to the state of the cell. First, use h_{t-1} and x_t to decide which information to update through an operation called an input gate. Then use h_{t-1} and x_t to obtain new candidate cell information $\widetilde{C_t}$ through a $tanh$ layer, and this information may be updated into the cell information.

After updating the cell state, it needs to determine which state characteristics of the output cell according to the inputs h_{t-1} and x_t. Then the input is passed through a sigmoid layer called an output gate to obtain a judgment condition. Then the cell state is passed through the tanh layer to obtain a vector between -1 and 1, which is multiplied with the judgment condition obtained by the output gate to obtain the final output of the LSTM unit.

4. EXPERIMENTS

4.1. Dataset

The dataset from Craft Beer dataset (Kaggle.com)[1], which mainly describes users' reviews on beer from 1999 to 2012. The Beer dataset file (CSV) has 37,500 rows and 19 columns. The user's review data includes textual reviews of beer and ratings of four features (aroma, appearance, taste and palate) and overall for beer. The information about the reviewers, such as age and gender, and other data columns irrelevant to the experiment, were excluded in this work. In the sentiment analysis experiment, the text dataset is classified according to the overall rating in the data, which is divided into three categories: neutral, positive, and negative (Table 1).

4.2. Experiment of Rating Data

1. Pyspark library is used to get a matrix <user, item, rating>;
2. Count the number of unique users and beer and calculate the sparsity of the matrix;
3. Training ALS model (Hidasi & Tikk, 2012) in pyspark.ml.recommendation library;

Table 1. Data description

Column Name	Quantity	Type
beer/beerId	37500	int64
beer/brewerId	37500	int64
beer/name	37500	object
beer/style	37500	object
review/appearance	37500	float64
review/aroma	37500	float64
review/overall	37500	float64
review/palate	37500	float64
review/taste	37500	float64
review/text	37490	object
user/profileName	37495	object

4. Training and get the score of each beer from each user, sort the score of each beer from the user, and recommend a beer to the user;
5. By comparing the ALS recommendation model, the SVD recommendation model (Ba, Li, & Bai, 2013) and KNN-based recommendation model (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994), select the optimal recommendation algorithm.

4.3. Sentiment Analysis Experiment for Text Review

4.3.1. Contrast Experiments

There are conventional machine learning classification models (Pedregosa et al., 2011), including the Decision Tree model, Random Forest model, Extra Trees model, Naive Bayesian model, and Logistic Regression model and Stochastic Gradient Descent classification model:

1. Decision Tree is a tree structure in which each internal node represents a judgment on an attribute, each branch represents the output of a judgment result, and each leaf node represents a classification result;
2. Random Forest refers to a classifier that uses multiple trees to train and predict samples. In which randomness is mainly reflected in two aspects: random sampling (row random) and random selection features (column random), which can prevent the occurrence of overfitting; multiple decision trees can prevent the occurrence of low generalization of the model;
3. Extra Trees algorithm has more randomness. When selecting the optimal split value for continuous variable features, the effect of all split values will not be calculated to select the split feature. Instead, for each feature, within its feature value range, a split value is randomly generated and then calculated to see which feature is selected for splitting;
4. Naive Bayes method is a classification method based on Bayes' theorem and independent assumptions of feature conditions. It is simplified based on the Bayesian algorithm, that is, it is assumed that the attributes are independent of each other when given target value;
5. Logistic Regression is a machine learning method for solving binary classification (0 or 1) problems. It is used to estimate the probability of certain events.

Deep Learning models: CNN (Kim, 2014), RNN (Mikolov, Karafiát, Burget, Černocký, & Khudanpur, 2010), LSTM (Sundermeyer, Schlüter, & Ney, 2012) classification model.

The detail of the above 3 Deep Learning models has been introduced in 3.2.1, 3.2.2 and 3.2.3.

4.3.2. Experiment Process

- Text Pre-processing uses the Tokenizer in the NLTK library to segment the text and get the word list for each textual review;
- Use the stop word dictionary in NLTK library to remove stop words, numbers and special symbols also removed from the list;
- Use the Word2vec model to learn the word vector of the data, using the default parameter settings. The pre-processed word sequence is entered into the learning word vector in the model;
- Use the generated word vector to input into each classification model for learning classification, and all models classify the data into three categories.

5. RESULT AND ANALYSIS

In the preliminary analysis of the data, the correlation matrix is shown in Figure 1.

Figure 1. Correlation Matrix

	review/appearance	review/aroma	review/palate	review/taste	review/overall
review/appearance	1	0.538076	0.555833	0.531676	0.498733
review/aroma	0.538076	1	0.608922	0.711809	0.616117
review/palate	0.555833	0.608922	1	0.732092	0.69722
review/taste	0.531676	0.711809	0.732092	1	0.78522
review/overall	0.498733	0.616117	0.69722	0.78522	1

It indicated that the most relevant to the overall rating of beer is the taste rating, followed by the palate rating. The least relevant is the appearance rating of beer.

Through the ALS recommendation model, beer can be recommended for each user. For example, the results of recommending beer for users 11 are as follows (Figure 2).

The beer recommendation system is a comprehensive recommendation based on the overall rating and four features' rating (aroma, appearance, palate and taste) to recommend beers for each

Figure 2. Recommendation results

```
User 11's Recommendations:
+-------+-------+----------+--------------------+
|beer_id|user_id|prediction|           beer/name|
+-------+-------+----------+--------------------+
|     39|     11| 4.2370405|Founders CBS Impe...|
|     14|     11|  4.250751|Barrel Aged B.O.R...|
+-------+-------+----------+--------------------+
```

user. Compared to the recommendation algorithm recommended only by one kind of rating, Users may likely to accept the beer recommended by the former system. However, due to the imbalance of the user rating data in the database, the accuracy rate of the ALS recommendation algorithm is 62%.

Compared with other SVD recommendation algorithms and KNN-based recommendation algorithms, the RMSE results are 0.65 and 0.82, respectively. The KNN-based recommendation algorithm is based on finding neighboring users and making recommendations based on items purchased by neighboring users. This method has a higher RMSE value of 0.82. Both the SVD recommendation model and the ALS model use the decomposition matrix to calculate the user's predicted score for the item. While the SVD first fills the entire sparse matrix and then calculates the dimensionality of the matrix by calculating the percentage of the square of the odd value. The RMSE values of the SVD model and the ALS model are similar in this dataset. However, the SVD model is not only computationally complex but also consumes a considerable amount of storage space. Thus, by comparison, the ALS recommendation model is best suited for this beer dataset.

In the experiment of sentiment analysis, the classification metrics (accuracy, recall, F1-Score) obtained by the conventional machine learning classification model shown in Table 2.

It can be seen from Table 2 that the accuracy of Logistic Regression, SGD Classifier, MLP Classifier, Extra Tree, Decision Tree and Random Forest model are all around 0.67, and the values of Precision and Recall are similar (around 0.69 and 0.67, respectively). The difference is the F1-Score value, which is the harmonic average of accuracy and recall. In the above four models, the highest F1-Score value is the Random Forest model, the Recall value is 0.695, and the recall values of the remaining five models is 0.69.

Overall, the best performing conventional machine learning model is the Random Forest model.

In comparative experiments, CNN, RNN and LSTM models were used to classify text sentiments. In order to facilitate experimental comparison, the main parameters of the three deep learning models are the same. the main parameters are shown in Table 3.

Table 2. Conventional model results

	Accuracy	Recall	Precision	F1-Score
Decision Tree	0.666	0.686	0.667	0.674
Random Forest	0.675	0.695	0.675	0.684
Extra Tree	0.676	0.697	0.676	0.679
Naive Bayes	0.614	0.631	0.647	0.634
Logistic Regression	0.677	0.697	0.677	0.679
SGD Classifier	0.677	0.697	0.676	0.679
MLP Classifier	0.676	0.697	0.676	0.680

Table 3. Parameter list

batch size	32
vocab dim	100
kernel size	3
dropout	0.5
num classes	3
epochs	100

The changes in the accuracy and loss of the LSTM model are shown in Figures 3 and 4.

The LSTM model has approached stationary in about 60 steps, which indicates that the model converges around 60 steps. The accuracy of the model reaches 0.75, which has been greatly improved compared to the conventional machine learning model. The second model of this experiment is the CNN model. As can be seen from Figures 5 and 6, the CNN model converges faster than the LSTM model and has converged in about 20 steps. However, the accuracy of the model is far less than that of the LSTM model, and its accuracy is 0.67. The accuracy is the same as that of the conventional machine learning model, but the complexity of the model is larger than the conventional machine learning model.

Figure 3. The Epoch accuracy of LSTM training process

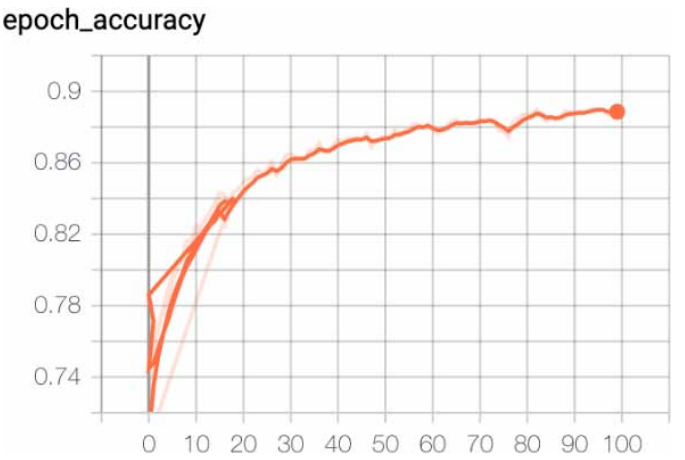
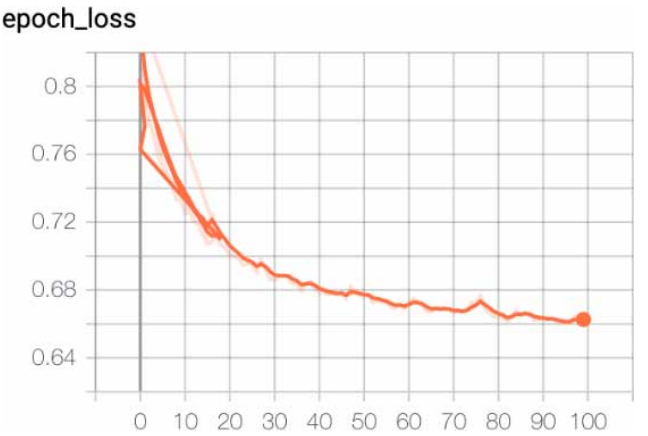


Figure 4. The Epoch loss of LSTM training process



The last model in this experiment is the RNN model. The loss of the RNN model did not converge well within the 100 steps (Figure 7). However, it can be seen in Figure 8 that the model has converged after 40 steps, and the accuracy is the same as the CNN model of 0.67.

Figure 5. The Epoch loss of CNN training process

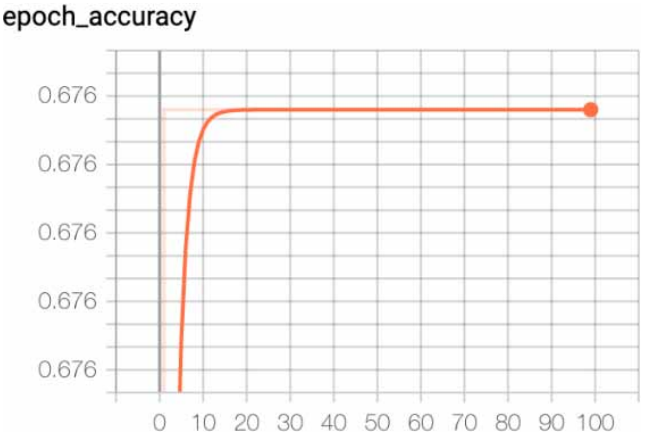


Figure 6. The Epoch loss of CNN training process

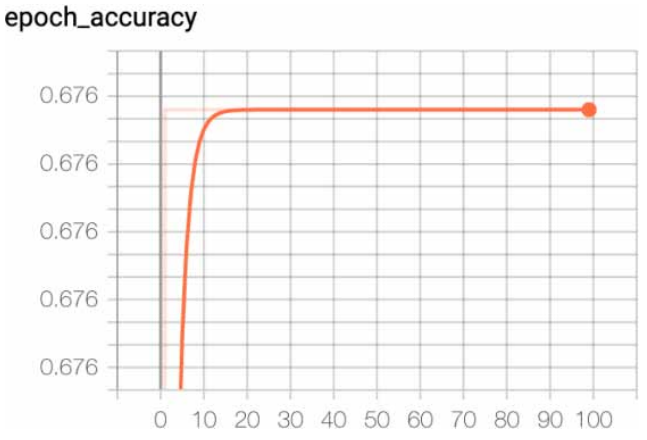


Figure 7. The Epoch loss of RNN training process

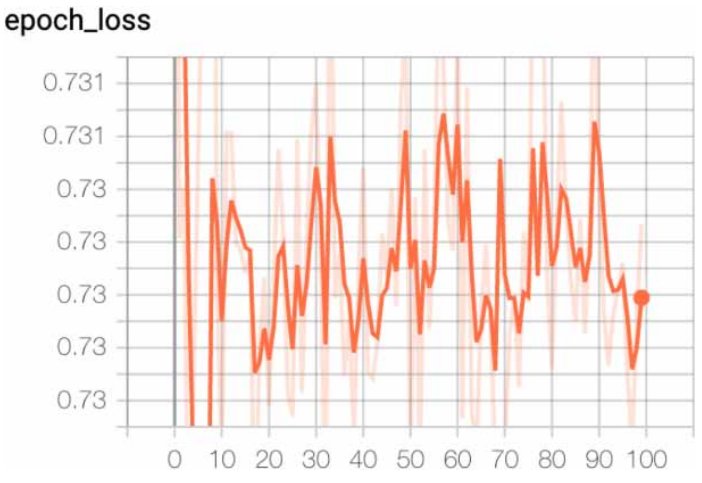
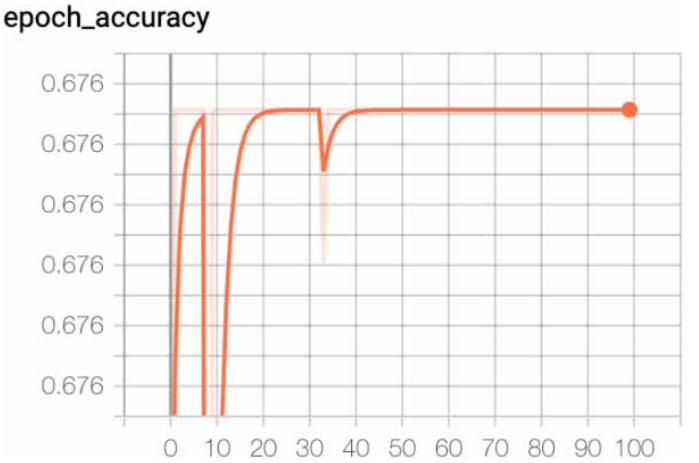


Figure 8. The Epoch accuracy of RNN training process



Comparing all the classification models in the experiment, the LSTM model performed best and achieved an accuracy of up to 0.75.

The recommendation system combined with the LSTM sentiment analysis system could increase the credibility of the recommended beer. According to the recommended beer by the recommendation algorithm, the LSTM model conducts a sentiment analysis of the reviews on these recommended beers. Although the recommendation algorithm is based on the user's recommendation for beer ratings, if the recommended beer has a lot of negative sentiments due to some other factors, these beers should also be removed from the recommended beer list. The beer recommended by the results of the integrated LSTM sentiment analysis system and recommendation algorithm can increase the user's acceptance. For example, according to the beer recommendation results of user No.62, after the LSTM sentiment analysis, the negative sentiments of the first three beer were less than 20%, and the positive reviews were higher than 50%. But the negative reviews on the fourth beer were as high as 51%, and the positive reviews were only 14%. Therefore, the beer of the fourth kind should be removed from the recommended beer list. Through the combination of the two analysis methods, the credibility of the recommendation system can be increased, and the probability that the user accepts the recommended product becomes larger. This is beneficial for users who buy beer online and can also help e-commerce to increase beer sales (Table 4).

5. LIMITATIONS AND DISCUSSIONS

Rating data is a measure that can directly quantify the quality of products in customer evaluation. As a typical quantitative data, products can be directly sorted through statistical forms. This is one of the

Table 4. Example results

Beer Type	Recommend	Negative	Neutral Positive
Nine Man Ale Golden Ale	14%	36%	50%
Benchwarmer Porter	3%	28%	69%
Strike Out Stout	9%	40%	51%
Amstel Light	51%	35%	14%

most intuitive and effective data types for recommendation systems, but it also has many shortcomings: 1. Rating data is affected by sales, and there is a statistical preference for products with fewer samples but higher average rating data. This is often unfair for products with a larger rating sample available. 2. Rating data represents the preferences of most customers, but it does not have a high reference value for customers with special preferences. Therefore, the single rating type of data cannot comprehensively reflect the real situation of the product. At the same time, there is a lack of more dimensional information for modeling. Therefore, customer review data, as a more comprehensive content carrier, contains richer details of the product. By combining review data and other product quantitative indicators (e.g. rating data, etc.) and other types of data into the recommendation agent, it can provide more accurate recommendations for customer groups with special preferences.

As a carrier of user questions, suggestions, and attitudes, the review data is extremely valuable for product evaluation, improvement, and optimization. Merchants can use text analysis to analyze user's concerns, main discussion topics, user's sentiment tendencies, and the main subject of reviews, etc. Therefore, this distillation processing of massive unstructured data can also transform complex and abstract semantics into quantifiable semi-structured or structured data. After such refined processing, a large amount of information implied in the textual review data can be effectively used. In particular, the potential multi-dimensional information in the reviews can be used to improve the recommendation effect of the recommendation model through dimensional modeling. In addition, the information extracted from the review text can also be directly presented to the user in a variety of report forms, which can include heat maps for sentiment analysis, radar chart of users' evaluation, word cloud, etc.

In the future study, the text mining task of reviews will be enriched more and more fine-grained (e.g. fine-grained topic analysis, etc.). At the same time, more quantitative data can be introduced into the recommendation model (e.g. other purchase record data of customers with different special preferences, and the duration of online browsing of related products, etc.). These sales-related data are an important reference source for consumers to quickly judge from mass products to meet their own needs. As a typical method of using Collective Intelligence, the recommendation model can use a collaborative filtering mechanism to discover a small number of products that match the preferences of target users from a large number of users. Compared with Collective Intelligence in the traditional sense, this collaborative filtering mechanism retains the characteristics of individuals (i.e. individual preferences) to a certain extent, so it can be used as the core idea of personalized recommendation algorithms. By adding more types of data as model inputs, the recommendation model can make more precise recommendations for their specific needs when serving the customer group with more complex and diverse preferences. This will undoubtedly enable merchants to track the appropriate target customer group better and faster, promote the purchase rate and favorable rate of related products, and provide efficient and high-quality services for customers with different needs.

6. CONCLUSION

By including more dimensions of online sales data, the effect of the online product recommendation system can be improved, and more accurate and high-quality product recommendations can be provided to customers with different preferences. It is a potentially feasible solution by combining customer rating data (quantitative type data) and review data (unstructured data) for recommendation system construction and performance improvement. This paper analyzes the rating and reviews of the beer dataset, and through the analysis of the rating, a beer recommendation system was constructed, which recommended beer according to four features' rating. Sentiment analysis of text reviews can effectively let the brewery or sales platform understand the consumer's feelings about a specific beer. By comparing the ten machine learning classification models, the LSTM model performed best, and

the accuracy of data classification reached 0.75, which was about 0.1 higher than the accuracy of other models. By combining the LSTM sentiment analysis model with the recommendation algorithm, the credibility of the recommended beer can be increased, allowing the user to accept the recommended beer, thereby facilitating the purchase and sale of the beer.

ACKNOWLEDGMENT

This work is partly supported by VC Research (VCR 0000044).

REFERENCES

Ba, Q., Li, X., & Bai, Z. (2013). *Clustering collaborative filtering recommendation system based on SVD algorithm.* Paper presented at the 2013 IEEE 4th International Conference on Software Engineering and Service Science.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, *50*(2), 511–521. doi:10.1016/j.dss.2010.11.009

Chen, H., Li, S., Wu, P., Yi, N., Li, S., & Huang, X. (2018). Fine-grained Sentiment Analysis of Chinese Reviews Using LSTM Network. *Journal of Engineering Science & Technology Review*, *11*(1), 174–179. doi:10.25103/jestr.111.21

Chen, L., Li, R., Liu, Y., Zhang, R., & Woodbridge, D. M.-k. (2017). *Machine learning-based product recommendation using apache spark.* Paper presented at the 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). doi:10.1109/UIC-ATC.2017.8397470

Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: The roles of online promotional marketing and online reviews. *International Journal of Production Research*, *55*(17), 5142–5156. doi:10.1080/00207543.2015.1066519

Da Silva, N. F., Hruschka, E. R., & Hruschka, E. R. Jr. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, *66*, 170–179. doi:10.1016/j.dss.2014.07.003

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007–1016. doi:10.1016/j.dss.2008.04.001

Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, *80*, 122–131. doi:10.1016/j.chb.2017.10.039

Hidasi, B., & Tikk, D. (2012). *Fast ALS-based tensor factorization for context-aware recommendation from implicit feedback.* Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. doi:10.1007/978-3-642-33486-3_5

Ifrach, B., Maglaras, C., Scarsini, M., & Zseleva, A. (2019). Bayesian social learning from consumer reviews. *Operations Research*, *67*(5), 1209–1221. doi:10.1287/opre.2019.1861

Jebbara, S., & Cimiano, P. (2016). *Aspect-based sentiment analysis using a two-step neural network architecture.* Paper presented at the Semantic Web Evaluation Challenge. doi:10.1007/978-3-319-46565-4_12

Jonathan, B., Sihotang, J. I., & Martin, S. (2019). *Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier.* Paper presented at the Abstract Proceedings International Scholars Conference. doi:10.35974/isc.v7i1.1003

Kim, Y. (2014). *Convolutional neural networks for sentence classification.* arXiv preprint arXiv:1408.5882

Lee, A. J., Yang, F.-C., Chen, C.-H., Wang, C.-S., & Sun, C.-Y. (2016). Mining perceptual maps from consumer reviews. *Decision Support Systems*, *82*, 12–25. doi:10.1016/j.dss.2015.11.002

Liu, H. (2019). Agricultural Q&A System Based on LSTM-CNN and Word2vec. *Revista de la Facultad de Agronomía*, *36*(3).

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., & Owen, S. et al. (2016). Mllib: Machine learning in apache spark. *Journal of Machine Learning Research*, *17*(1), 1235–1241.

Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). *Recurrent neural network based language model.* Paper presented at the Eleventh annual conference of the international speech communication association.

Mostafa, M. M. (2018). Mining and mapping halal food consumers: A geo-located Twitter opinion polarity analysis. *Journal of Food Products Marketing*, *24*(7), 858–879. doi:10.1080/10454446.2017.1418695

Park, D.-H., Lee, J., & Han, I. (2007). The effect of online consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, *11*(4), 125–148. doi:10.2753/JEC1086-4415110405

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Dubourg, V. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Proserpio, D., & Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, *36*(5), 645–665. doi:10.1287/mksc.2017.1043

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. doi:10.1145/192844.192905

Ruder, S., Ghaffari, P., & Breslin, J. G. (2016). *A hierarchical model of reviews for aspect-based sentiment analysis.* arXiv preprint arXiv:1609.02745

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30–40. doi:10.1016/j.dss.2015.10.006

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, *70*, 346–355. doi:10.1016/j.jbusres.2016.08.008

Sundermeyer, M., Schlüter, R., & Ney, H. (2012). *LSTM neural networks for language modeling*. Paper presented at the Thirteenth annual conference of the international speech communication association.

Wei, C.-P., Chen, Y.-M., Yang, C.-S., & Yang, C. C. (2010). Understanding what concerns consumers: A semantic approach to product feature extraction from consumer reviews. *Information Systems and e-Business Management*, *8*(2), 149–167. doi:10.1007/s10257-009-0113-9

Winlaw, M., Hynes, M. B., Caterini, A., & De Sterck, H. (2015). *Algorithmic acceleration of parallel ALS for collaborative filtering: Speeding up distributed big data recommendation in spark*. Paper presented at the 2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS). doi:10.1109/ICPADS.2015.91

Xie, L., Zhou, W., & Li, Y. (2016). Application of improved recommendation system based on spark platform in big data analysis. *Cybernetics and Information Technologies*, *16*(6), 245–255. doi:10.1515/cait-2016-0092

Zhang, L., Zhou, Y., Duan, X., & Chen, R. (2018). *A hierarchical multi-input and output Bi-GRU model for sentiment analysis on customer reviews.* Paper presented at the IOP Conference Series: Materials Science and Engineering. doi:10.1088/1757-899X/322/6/062007

ENDNOTE

[1] https://www.kaggle.com/applied-computing/beers

Pin Ni received the MRes degree from the University of Liverpool, UK in 2020. He is the co-founder of Research Lab for Knowledge and Wisdom (KnoWis), Xi'an Jiaotong-Liverpool University, China. His current research interests include Natural Language Processing, Knowledge Graph, Semantic Web, Data Mining and Knowledge Discovery.

Yuming Li received the MRes degree from the University of Liverpool, UK in 2020. She is the co-founder and member of the Research Lab for Knowledge and Wisdom (KnoWis), Xi'an Jiaotong-Liverpool University, China. Her current research interests include Natural Language Processing, Information System, Knowledge Graph, Deep learning in the financial field, etc.

Victor Chang (PhD) is a Professor of Data Science and IS at Teesside University, UK. He was a Senior Associate Professor, Xi'an Jiaotong-Liverpool University between June 2016 and Aug 2019. He was a Senior Lecturer at Leeds Beckett University, UK between Sep 2012 and May 2016. Within 4 years, he completed Ph.D. (CS, Southampton) and PGCert (HE, Fellow, Greenwich) while working for several projects. Before becoming an academic, he achieved 97% on average in 27 IT certifications. He won an IEEE Outstanding Service Award in 2015, best papers in 2012, 2015 & 2018, 2016 European award: Best Project in Research, 2017 Outstanding Young Scientist and numerous awards since 2012. He is widely regarded as a leading expert on Big Data/Cloud/IoT/security. He is a visiting scholar/PhD examiner at several universities, an Editor-in-Chief of IJOCI & OJBD, former Editor of FGCS, Associate Editor of TII & Info Fusion, founding chair of international workshops and founding Conference Chair of IoTBDS, COMPLEXIS, FEMIB & IIoTBDSC. He was involved in projects worth more than £13 million in Europe and Asia. He published 3 books and edited 2 books. He gave 18 keynotes internationally as a top researcher.