Regional Development of E-Commerce Based on Big Data Evaluation Model

LinLi1, FangQin1, *, CanWang1, JianyanSun1, WeiJiaZeng1, LinLinYu1
1, *Dalian University of Science and Technology

Abstract: In order to improve the authenticity and reliability of the evaluation data of regional e-commerce for reflecting the regional credit level in the business platform, this paper proposes a regional credibility evaluation model based on big data, which integrates the behavior characteristics of regional economic development and the judgment of the credibility of comments. Based on the analysis of regional economic behavior, the features include regional activity, regional reliability and regional contribution, while the credibility of comments is calculated based on the quality of regional text. The experimental results show that this method can quickly screen out the high reliability region, and can effectively alleviate the cold start problem. In the regional credibility evaluation model, two groups of features are proposed: regional behavior feature and regional text quality feature. The credibility of regional behavior, the credibility of regional social relations and the quality of regional comments are comprehensively evaluated. These two sets of features solve the problems of incomplete features and data cold start in the evaluation model of regional credibility.

1 Introduction
Digitalization is changing the way companies conduct business. Digital technology enables enterprises to benefit from many aspects, including cheaper resources, larger market scope and more efficient communication, which can break the old system for sales and distribution. Also it changes the consumer purchase behavior of many products and services [1]. Digitization and new technology require us to reconsider the key assumptions of the E-commerce Regional Evaluation Theory. Existing theories may need to be adjusted, or new theories may be applied to better explain the new phenomena in the international digital economy [2]. However, so far, few e-commerce regional evaluation studies can explore the impact of emerging digital technologies. Therefore, we urgently need to develop new theories and revise existing theories, and to determine how enterprises can benefit from digitalization when they conduct business around the world. Taking international enterprises engaged in cross-border e-commerce as the main research objects, this study reexamines the applicability, explanatory power and possible evolution direction of traditional theories such as principal-agent theory and control theory in the context of cross-border e-commerce transactions, which is an important supplement to the existing research on e-commerce regional evaluation in the era of digital economy, and has theoretical value to some extent [3].

When it comes to the text quality evaluation mechanism, the paper proposes two methods including the text quality evaluation method based on the theme model and the text quality evaluation method based on deep learning are proposed in this paper. The experimental results show that the two methods used are more accurate than other methods in the study of grammar analysis. At the same time, the evaluation method based on deep learning is better than that based on the theme model. Through this mechanism, the problem of insufficient text processing in previous research can be effectively solved.
The paper constructs a regional credit evaluation model of e-commerce based on blockchain by using big data technology, which provides new ideas, new perspectives and new methods for the research of e-commerce credit evaluation.

2 Statement of Regional Economic Development

According to the data evaluation model, the region is divided into high credibility region and common region, the definitions are as follows: Definition 1 (data evaluation model). For a given BRS region $U = \{u_1, u_2, ..., u_n\}$, in which each region $u_i \in U$ is represented as a feature vector $u_i = \{u_{i1}, u_{i2}, ..., u_{in}\}$, the data evaluation model is to learn a function $f(\cdot)$, to map each region $u_i \in U$ to a category label $c_i \in c = \{\text{high-credible}, \text{ordinary}\}$, so for any $u_i$, the following formula could be conducted:

$$f(u_i) = \begin{cases} 1, & \text{if } u_i \text{ is high-credible} \\ 0, & \text{if } u_i \text{ is ordinary} \end{cases}$$ (1)

Based on Definition 1, in order to solve the problem of data evaluation model, the experiment can be conducted through two steps, including: (1) extracting the required features, and (2) finding a suitable classification algorithm [4]. The evaluation model is shown in Figure 1.

![Data evaluation model](image)

Figure 1 Data evaluation model

The regional development of e-commerce is based on the regional review, which establishes a perfect commodity review and regional communication mechanism. However, the rampant of low-quality comments has caused damage to the interests of the region. Therefore, this paper proposes
a data evaluation model method based on the fusion of behavior analysis and comment credibility judgment [5]. The evaluation method is divided into three steps: regional feature extraction, text quality evaluation and region classification. Figure 3.1 shows the overall process for using evaluation data to judge model. Firstly, features are extracted from the personal data and comments of the region. Based on the features, a classifier is used to determine the credibility of each region. In order to solve the problem of data imbalance in the test process, three techniques are used to solve the problem, namely: random over sampling, random under sampling and SMOTE oversampling (SMOTE Tomek) [6].

3. Feature Representation of Region Behavior and Region Text
All the features are listed used in Figure 2. In the experiment, three features are used, which are region activity feature, region reliability feature and text quality feature.

![Character graph](image)

3.1 Characteristics of Regional Activity
Regional activity is mainly reflected in the degree of regional participation in the development of e-commerce. Regions with high credibility are usually more active and involved. Therefore, in terms of regional activity, five characteristics are introduced to evaluate the regional activity, which are: the regional level of the area, the average number of monthly comments, the number of regional bookmarks, the number of regional check-in and the contribution of the region [7].

(1) The first is the regional level. At present, many online business review services provide the function of establishing social relations for regions. Regions are a network connecting different areas, and what connects those areas is the same hobby. For both direct and indirect regions, this has shorten the distance between them. The areas with higher activity gain more reputation in the region, while other areas also prefer to listen to their opinions and suggestions. The level of the area is selected as one of the characteristics of regional activity [8]. In online business reviews, region level is usually used to express the richness of their experience. The "TALK" module is often used in YELP.

(2) The second is average number of comments per month in the region. Usually, people with more shopping experience will post more reviews. Therefore, the average number of comments per month in the region shows the number of regional purchasing behavior to a certain extent. If the number of comments published in a region is not enough or the publishing time is too concentrated, the region may be a malicious spam comment publisher to a certain extent. On contrast, if the number of monthly reviews released by a region is relatively even, then the region could be seen as having more reasonable purchasing habits and behaviors [9]. In reality, the comments published by low credible regions are more concentrated in one month, while the release of other months tends to zero. In the experiment, the number of regional reviews published each month is selected as one of the indicators of the data evaluation model. The calculation formula is as follows:
Within the formula, $M_i$ represents the number of comments published in the $i$th month, while $N_{non+1}$ represents the number of all months, and $U_{month}$ is the divisor of the total number of comments published in all months and the month of publication.

(3) The third is number of regional bookmarks. In the online business review service, when you see the goods you like or are interested in, you can put them into the bookmark folder through the collection function. The more active the area is, the more goods or stores it browses. Therefore, the number of bookmarks in the area can also be used as one of the evaluation indicators of the area activity, and affect the data evaluation model to a certain extent [10].

(4) The forth is number of regional check-in. With the continuous development of location-based services and wireless communication services, people can share their location information on a variety of applications through smart phones or tablets, and use text-based assistance. This is the current situation of social networks [11]. Whether it's Weibo or Twitter, the region can be punched in anywhere to show that people have been there. Therefore, the development of many e-commerce regions also provides such work synchronously. Regions can be checked in when people visit a special place.

(5) The fifth is regional contribution value. E-commerce regional development uses regional level as the overall embodiment of regional activity, which is reflected in many websites, such as public comments or YELP. In general, the website will set different treatment according to the level of each region, and for regions with loyalty and high consumption power, the discount will be based on a greater degree [12]. On the one hand, it promotes the enthusiasm of regional consumption, while increasing the regional stickiness of the website. Therefore, each website designs a new statistical method at the regional level to keep the region active. In the data evaluation model, the comprehensive evaluation of each region of the website is used as a feature that affects the regional activity. But the difference is that websites have different calculation methods. In e-commerce regional evaluation, any regional behavior may cause the increase of regional level. For example, the region that people can use mobile payment to consume in the store, evaluates a store or a commodity, or uses the check-in function, and so on. These operations will make the regional level increase accordingly. However, there is no direct description of regional level on YELP website. According to the statistical method of e-commerce regional evaluation, the regional level corresponding to YELP is fitted. Finally, the regional grade UC is selected as one of the indicators of the evaluation data evaluation model. Regional contribution value is the cumulative embodiment of regional activity behavior [13].

3.2 Regional Reliability Characteristics
Regional reliability characteristics essentially describe the degree to which a region can be trusted by other regions. Therefore, based on the social reality of e-commerce regional development, if region A trusts the text content published by region B and the content published by region B has high quality, region A will "focus" on region B [14]. At the same time, region A can also get any public information of region B anytime and anywhere. According to this principle, the reliability of a region is directly proportional to the number of regional concerns. If a region pays attention to a considerable number of regions with high contribution value, then the region should also have a high degree of reliability. Therefore, the data evaluation model is combined with the number of regional concerns and the contribution value of regional concerns. According to this principle, the regional reliability characteristic is proposed to express the regional reliability Uscore, the calculation formula is as follows:

$$U_{score} = \sum_{i=1}^{N_{total}} (u_i \times N_i)$$
Within the formula, \( U_i \) is the weight value of each region level, \( N_i \) is the region level of the region, and \( N_i \) is the number of concerns of the region in the \( i \)th level. The essence of regional reliability is the sum of the specific number of regions of interest corresponding to each level of the region multiplied by the corresponding weight number of the region level.

### 3.3 Text Quality Characteristics

The text quality of regional reviews is a very important factor in the data evaluation model system. A region that often publishes high-quality comments will attract a large number of attention, and people would trust the comments it publishes. In general, a high-quality review should be substantial and cover all aspects of the store or food being evaluated. Therefore, if a comment can cover the vast majority of aspects needed for high-quality evaluation, then the comment could be high-quality [15].

The number of words per comment is used as one of the characteristics to measure the quality of the text, the calculation formula for \( U_{\text{length}} \) is as follows:

\[
U_{\text{length}} = \frac{\sum L_i}{N}
\]

(4)

Within the formula: \( L_i \) is the text length of the comment, and \( N \) is the total number of all comments in the region. \( U_{\text{length}} \) represents the average length of all text.

High quality reviews usually cover attributes of the commodity in many ways. The fine-grained emotional analysis of regional reviews is of great value to understand the intention of regional reviews and to explore regional emotions. At present, it has been widely used in intelligent recommendation system, intelligent search and other fields. Through analysis, the emotional tendency of comment text under each attribute is judged. However, the long text length comment can not represent a high-quality evaluation. Many comments often repeat a sentence several times to meet the requirements for specific word number. The number of words in these comments is usually very long, but the quality is not high. There are also cases where some comments focus particularly on one aspect of the product, so the aspects involved are not comprehensive. Therefore, two methods of evaluating text quality are studied, one is based on LDA, the other is text quality evaluation based on deep learning.

1. **A text quality evaluation model based on LDA topic model.**

   For the text quality evaluation, we should consider not only the length of the comments, but also whether the text covers as many topics as possible. LDA is used to evaluate the text quality of the comments. LDA is a kind of topic model, which regards different aspects of goods or stores as different topic distribution in LDA. That is, a topic represents one aspect of goods, such as quality of goods, taste of food, service of stores and so on. Firstly, the LDA model is trained with large-scale labeled data sets, and then the variance of the topic distribution of the comment is used to measure the coverage of different aspects of the product, \( \lambda_i \) could be seen as variance of the topic distribution, the calculation formula of which is as follows:

\[
VAR(\lambda_i) = \frac{\sum_{j=1}^{K} (\lambda_{ij} - \bar{\lambda}_i)^2}{K}
\]

(5)

Within the formula, \( \lambda_{ij} \) is the j-th component of \( \lambda_i \), \( \bar{\lambda}_i \) is the mean value, and \( K \) is the number of topics in LDA topic model. \( VAR(\lambda_i) \) is the variance of all topic features and is expressed as Utopic. The Utopic of the region is inversely proportional to the coverage of the product theme. In other word, the smaller the region's Utopic value is, the better the coverage become and so it does for the quality of the regional reviews. For each region, there is a unique value to indicate the text quality of the region.

2. **Text quality evaluation model based on deep learning**

   LDA topic model is used to judge the topic distribution of comments. At present, the development of deep learning is undergoing great popularity, the use of deep learning method to evaluate whether
the distribution of comments topic has better effect. As a popular natural language processing method, BERT has achieved good results in many natural language processing tasks. The content of the text evaluated by BERT uses the relationship between the context of the text to a certain extent, so it has high accuracy.

According to the task requirements, a multitask classification model based on BERT is designed. The structure of the model is shown in Figure 3.

![Figure 3 BERT model](image)

The model is divided into two modules, the first part is the share layers, the second part is the task-specific layers. In the data sharing layer, the context representation of text information is calculated based on the pre training model of BERT. The best used relies on the Chinese pre training model provided by Google and the mass data provided by the American group comments as pre training, which can better represent the context information and extract the deep semantic information. In the second part, the specific task used multiple Attention+Softmax mechanism to predict the classification of each word.

LDA topic model statistics is to calculate the distribution of each valid word in the string through the model. Calculation formula of which is as follows:

$$ e_i = \frac{n_i}{N} $$

(6)

Within the formula, $e_i$ is the topic distribution of topic i in the comment, ni is the number of words belonging to topic i in the high comment, and N is the number of effective words in the comment.
Then, the variance of the topic distribution of each comment is used to evaluate the coverage of each comment on each topic. \( \lambda_i \) indicates the topic distribution of each comment. The calculation formula of \( \lambda_i \) is as follows:

\[
VAR(\lambda_i) = \frac{\sum_{j=1}^{K} (\lambda_i(j) - \bar{\lambda}_i) \lambda_i(k)}{K}
\]  

Among which, \( \lambda_i(j) \) is the j-th component of \( \lambda_i \), \( \bar{\lambda}_i \) is the mean value of \( \lambda_i \), and K is the classification number of the BERT model.

### 3.4 Normalization

The normalization is to map value of each number to \([0,1]\) through min or max. Therefore, for a certain region, the normalized calculation method of its characteristics is as follows: (3.8)

\[
x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}
\]  

Within the formula, max and min represent the maximum and minimum values of the region, indicating the specific value to be normalized. Projection represents the result of normalization.

In LDA topic model or BERT, the basic processing unit is a word item, which is judged according to a single word item. In the comments, there are many words that affect the experimental results, which are called stop words. In the segmentation model, these words are often separated to form an independent term. In order to reduce the influence of stop words on the experimental results, a list of stop words is listed and the stop words are removed in the experiment. The words listed in the stop list have no meaning and have no influence on the understanding of the sentence.

### 3.5 Reliability Evaluation Model

Three groups of different features were extracted namely the features of regional activity, regional reliability and text quality. The purpose of the experiment is to evaluate the reliability of the region. All features are integrated into a vector with nine components, each of which corresponds to one feature. The samples in training set and test set can be expressed as vectors and used as input of classification model. In order to reduce the problem of the large difference of data magnitude between the same features, the normalization is carried out.

In the experiment, three different data evaluation model classifiers are tested to evaluate the final performance, including support vector machine (SVM), random forests and logical regression. All the classifiers are implemented in Weka.

The number of low reliability regions is larger than that of high reliability regions, which results in the imbalance of two kinds of experimental data, and affects the final performance of the data evaluation model. In order to solve the text caused by data imbalance, the sampling mechanism is used in the process of classifier learning. In the experiment, three sampling mechanisms are used, namely random oversampling, random under sampling and SMOTE-Tomek. Three sampling mechanisms are used to process the training sets in the experiment, and the training sets are used to train the three classifier models.

### 4 Experiment and Analysis

#### 4.1 Experimental Environment

This experiment is based on Python. The specific configuration information is shown in Table 1.
Table 1 Experimental configuration information

| Surroundings | Configuration Information |
|--------------|---------------------------|
| Hardware     |                           |
| CPU          | i7-10700                  |
| RAM          | 32GB                      |
| Hard disk    | 1TB                       |
| Network      | 300Mbps                   |
| Operating system | Windows10              |
| Software     |                           |
| Programming language | Python 3.6.7          |
| IDE          | Pycharm                   |
| Database     | Mysql                     |

4.2 Performance Evaluation Test

The performance level, performance evaluation index, calculation formula of each index are showed as follows:

\[
\eta = \frac{(TP + TN)}{Total} \quad (9)
\]

\[
\mu = \frac{TP}{(TP + FP)} \quad (10)
\]

\[
\sigma = \frac{TP}{(TP + FN)} \quad (11)
\]

\[
F_1 = \frac{2\eta \mu}{\eta + \mu} \quad (12)
\]

Among them, TP is the number of positive samples correctly classified, FN is the number of positive samples wrongly classified, TN is the number of counter samples correctly classified, and FP is the number of counter samples wrongly classified. F1 is an evaluation index that balances precision and recall.

4.3 Effectiveness Test of Text Quality Evaluation Mechanism

In LDA model, the test parameters include super parameter \( \alpha, \beta \) and topic number K respectively. In order to verify the sensitivity of topic number, several LDA models are tested and used to calculate the text quality features introduced before. Then, different LDA models are used to calculate the text features of comments, and their performance is tested by accuracy.

In the whole process of the experiment, we use the random forest classifier model. In total there are three variables, two of them are controlled, the remaining one is changed to evaluate the effect of the model. First, fix the value of \( \beta \) at 0.01, and change the value of \( \alpha \) in the range of K, with the step size of 3. At the same time, for each specific value of each \( \alpha \). Figure 4 and figure 5 show the experimental results of DRD and YELPYRD when applying and evaluating data, respectively.
The value of $\alpha$ is fixed to 5.6 and the value of $\beta$ is changed to a range between 0.01, 0.1, 0.25, 0.5, 0.75 and 1. For each fixed value, test the impact of $K$ on the experimental method when $K$ equals 5, 7, 10, 15 respectively.

Test data analysis shows that the number of LDA topics has a great impact on the experimental results. In the data evaluation of DRD, when $K = 7$, the model produced the best effect, and the accuracy of the model was improved by more than 10% compared with the condition when $K = 15$. On YELPYRD, the best results can be obtained when $K = 10$, and the relative accuracy is increased by 7% compared with the condition when $K = 15$. Therefore, the number of topics will have a greater impact on the overall experimental results.

Therefore, attention should be paid to the number of topics when choosing an LDA model. Choosing an appropriate number of topics has a higher effect on the experimental results. The effect of the change of super parameters and parameters on the experimental results is less than 2%. According to the
performance of other studies, the change of super parameter has little effect on the experimental results. At the same time, the change of super parameter has some influence on the operation speed and convergence of the model in the experiment, and the super parameter may have some influence on the experimental efficiency.

5 Conclusion
In this paper, a data evaluation model based on regional behavior analysis and comment credibility judgment is proposed. The method consists of region activity, region reliability and region text quality. The extraction and processing methods of each group of features are introduced in detail, as well as the methods for the text quality evaluation, which include text quality evaluation method based on topic model and text quality evaluation method based on deep learning. This paper introduces the process of the whole model and the concrete steps of the data evaluation model. Finally, considering the uneven distribution of data sets, this paper introduces the algorithm of processing data sets by using resampling technology.

Acknowledgement
This work was supported by 2019 Scientific Research Funding Project of Educational Department of Liaoning Province (Project No.: W2019003)

References
[1] Han, J.H.&Kim, H.M. The role of information technology use for increasing consumerinformedness in cross-border electronic commerce: An empirical study[J]. Electronic Commerce Research and Applications, 2019, 34: 82-106.
[2] Liu P, Diao H, Zou L, Deng A. Uncertain Multi-attribute Group Decision Making Based on Linguistic-Valued Intuitionistic Fuzzy Preference Relations[J]. Information Science, 2020, 508: 293-308.
[3] Liu X, Tan X, Zhang Y, Zou L. Credibility Factors Reasoning Based on Linguistic Truth-Valued Intuitionistic Fuzzy Hesitancy Degree[J]. Journal of Multi-Valued Logic and Soft Computing, 2018, 30: 285-301.
[4] Zhang Y, Huang D, Lin H, Zou L. Knowledge Reasoning Approach with Linguistic-valued Intuitionistic Fuzzy Credibility[J]. International Journal of Machine Learning and Cybernetics, 2020, 11: 169-184.
[5] Qian T, Wei L, Qi J. Constructing Three-Way Concept Lattices Based on Apposition and Subposition of Formal Contexts[J]. Knowledge Based Systems, 2017, 116: 39-48.
[6] Zou C, Deng H, Wan J, Wang Z, Deng P. Mining and Updating Association Rules Based on Fuzzy Concept Lattice[J]. Future Generations Computer Systems, 2018, 82: 698-706.
[7] Zhi H, Li J. Granule Description Based on Formal Concept Analysis[J]. Knowledge Based Systems, 2016, 104: 62-73.
[8] Zhi H, Li J. Granule Description based on Positive and Negative Attributes[J]. Granular Computing, 2019, 4: 337-350.
[9] Zhi H, Li J. Granule Description Based Knowledge Discovery from Incomplete Formal Contexts via Necessary Attribute Analysis[J]. Information Sciences, 2019, 485: 347-361.
[10] Liu P, Cui H, Cao Y, Hou X, Zou L. A Method of Multimedia Teaching Evaluation Based on Fuzzy Linguistic Concept Lattice[J]. Multimedia Tools and Applications, 2019, 78: 30975-31001.
[11] Chen J, Mi J, Xie B, Lin Y. A Fast Attribute Reduction Method for Large Formal Decision Contexts[J]. International Journal of Approximate Reasoning, 2019, 106: 1-17.
[12] Wang C, Shao M, He Q, Qian Y, Qi Y. Feature Subset Selection Based on Fuzzy Neighborhood Rough Set[J]. Knowledge-Based Systems, 2016, 111: 173-179.
[13] Wang C, Huang Y, Shao M, Fan X. Fuzzy Rough Set-based Attribute Reduction using Distance Measures[J]. Knowledge-Based Systems, 2019, 164: 205-212.
[14] Nguyen N, Sartra W. On Reduction of Attributes in Inconsistent Decision Tables Based on Information Entropies and Stripped Quotient Sets[J]. Expert Systems with Applications, 2019, 137: 308-323.

[15] Nguyen N, Sartra W. A New Approach for Reduction of Attributes Based on Stripped Quotient Sets[J]. Pattern Recognition, 2020, 97: 1-13.