Chinese Transaction Behavior Analysis with Recommender System Based on Transaction Behavior Categories

Jianyang Yu[1], YuanyuanQiao[1], Nanfei Shu[2], Kewu Sun[2], Shenshen Zhou[2], Jie Yang[1]
[1]*Beijing Key Laboratory of Network System Architecture and Convergence*
Beijing University of Posts and Telecommunications, Beijing, China
[2]*Technology Research Institute, Aisino Corporation, Beijing, China*
{oceancontinent, yyqiao}@bupt.edu.cn, {shunanfei, sunkewu, zhoushenshen@aisino.com}, janeyang@bupt.edu.cn

***Abstract*—With the rapid development of Chinese economy, it is significant to examine the economic activities in China. Each transaction behavior is recorded by the invoice. The invoice contains the transaction content, the classification of the transaction behavior (in accordance with the Tax Classification and Coding for Commodities and Services issued by the state) and transaction price, etc. Our work uses real invoice data and conducts a multi-dimensional analysis of Chinese transaction behavior based on recommender system of transaction behavior categories. Firstly, we propose a recommender system based on compositional CNN-RNN model with attention mechanism to recommend the corresponding categories of transaction behavior collected from tax invoices that maps the transaction behavior recorded in the invoice to transaction code in the Tax Classification and Coding for Commodities and Services issued by the state. Preliminary experiments show that the top-one accuracy of recommending transaction behavior achieves 75%. Then, we focus on the quantity distribution of invoice data. We can draw a conclusion that the major category with larger number of invoice records is more diversified and accompanied by an increase in the number of subdivided categories. After that, we studied the price distribution of various transaction behaviors to discover the difference in price distribution between different industries. Prices in the major categories of goods are more concentrated in the middle or lower prices. We can analyze the regional industrial structure through the price distribution of the industry which makes sense to study the economy of the region from the perspective of industry.**

***Index Terms*—Tax Invoice, Transaction Behavior, Deep Learning, Price Distribution, Recommender System**

I. Introduction

Today the Chinese economy is the second largest in the world and it experienced massive growth in that 35-year span, authorities have taken a new approach to the economy called the new normal. China has been the largest contributor to world growth since the global financial crisis of 2008. Therefore, it is crucial to study the economic activities in China [1].

Invoice refers to the text issued by the buyer to the purchaser during the purchase and sale of goods, provision or acceptance of services, and other business activities [2]. The invoice includes the name, quality and agreed price of the product or service provided to the purchaser. In other words, the invoice records the transaction behavior of the company. Value-added tax (VAT), known in some countries as a goods and services tax (GST), is a type of tax that is assessed incrementally, based on the increase in value of a product or service at each stage of production or distribution.

Analysis of Chinese VAT invoice data can help to understand the economic behaviors among enterprises and consumers, and then analyze the industry, industry status, ecological chain, value chain, and changes in the development process. In order to accelerate the construction of tax modernization, facilitate taxpayers' issuance of VAT invoices in a convenient and standardized manner, and assist tax authorities in strengthening the administration of tax administration, the State Administration of Taxation has prepared the Tax Classification and Coding for Commodities and Services (for Trial Implementation) [3]. Transaction behavior is classified into 6 major categories (including Goods, Labor Service, Sales Service, Intangible Assets, Real Estate and Non-Tax Items) and 4226 sub-categories according to transaction codes. Transaction codes can help to classify a wide range of goods and services (collectively referred to as transactional behavior) to help understand economic behavior, capital flows, and to track down tax evasion. Tax rate varies from category to category. To name only a few, the tax rate for the cereal is 11%, which is different from the category of mineral products with 17%. So it is necessary to accurately recommend the corresponding sub-category for the transaction behavior.

In order to analyze the economic activities of various industries in China in a fine-grained manner, this paper intends to analyze the distribution of invoices in various industries and the price distribution of different industries. Nevertheless, a transaction behavior may be mapped to multiple transaction codes without considering the actual semantics. For instance, apple may not mean a fruit, but it may be the name of Apples products. Besides, the majority of the transaction behavior contents are in short text and are not as long as sentences, which cant provide enough words cooccurrence or shared context for a good similarity measure and is formidable to extracting valid feature information. In Fig. 1 we find that with

Identify applicable funding agency here. If none, delete this.

being segmented by the word segmentation, content shows a heavy-tailed distribution in the number of words. And the average number of words for content is about 6. In order to use the transaction code to analyze the above issues, this paper proposes a recommender system based on a compositional CNN-RNN model with attention mechanism to recommend the corresponding categories of transaction behavior that maps the transaction behavior in the invoice to transacation code in the tax classification and coding document issued by the state. Overall, the contributions of our work are summarized as follow:

1. In order to solve the problem that there are a lot of unlabeled data in invoice data, we propose a propose a recommender system to recommend the transaction code for transaction behavior according to the Coding of Taxes for Commodities and Service. While considering the characteristics of the invoice data, we use the normalized vector of code frequency to represent label instead of using one-hot encoding and then recommend the labels (transaction code) for data (transaction behavior) with our system.

2. In our analysis, we found that the transaction behaviors are more diverse in major category which has the larger number of sub-categories codes. We have focused on analyzing the price distribution of different industries and extracting in the top five categories of invoices in each major category, we find that the price distribution in different industries is related to the industry characteristics and the price of goods are mainly distributed in the middle-to-lower price range. However, in the major categories of sales services and intangible assets, the price distribution of these industry is relatively uniform and is not concentrated in a certain price range.

3. Our work is based on real mass tax invoice data, during the whole 2017, which contains more than 8 million companies and 200 million tax invoices. This wealth of data can help us deeply understand the transaction behavior of a region in a fine-grained manner, and then study the economic development of the region [4].

This paper is structured as follows. Related work is discussed in Section II. In Section III, we provide details about our real dataset, and present some basic observations of data. In Section IV, the architecture of recommender system based on transaction behavior categories is presented as well as model configuration and parameters. Then, we present the experiment details and conduct a deep analysis on the category distribution of tax invoices and price distribution from the view of different industries in Section V. At last, we summarize our discoveries and discuss potential investigations in Section VI.

II. Related Work

In order to study the transaction behavior of China, with real tax invoice data, this paper proposes a recommender system to recommend the corresponding categories of transaction behavior that maps the transaction behavior in the invoice to transaction code, and then studies the economic activities in different industries. This Section gives a detailed description of transaction behavior analysis and text recommender system.

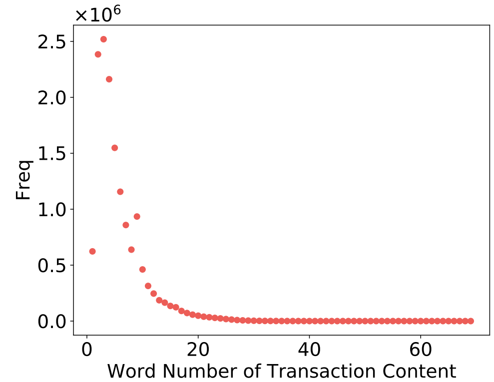


Fig. 1: The scatter figure shows the distribution of word number of transaction content that we have a distinct preprocess before.

A. Transaction Behavior Analysis

In China, most of transaction behaviors will be accompanied by the invoice. Whether it is a paper invoice or an electronic invoice, the tax controller will record the detailed transaction behavior, including transaction content, transaction code and so on. Transaction behavior can reflect the economic development of a country or region from the side. Research on transaction behavior can examines how market liberalization affects the profitability and productivity of Chinese firms [5]. Previous studies often used house prices [6], mobile communication data [7], etc. to study regional economic changes which have to face the problem of poor timeliness and coarse granularity, absolutely impossible to track changes in the industry. However, it is possible to characterize and detect those potential users of false invoices in a given year, depending on the information in their tax payment, their historical performance and characteristics [8]. Recently a study has been proposed that focuses on the analysis of intra-provincial trade growth in China through a database of invoices collected by the tax system [9]. But no studies have examined the economic activities of countries or regions from the perspective of the industry based on tax invoices.

This paper analyzes real value-added tax invoice data to study Chinese transaction behavior. In order to classify transaction behavior, the categories recommender system on transaction behavior is proposed and the transaction behavior is mapped to the transaction code, which gives the official category of transaction behaviors. It has provided new methods and ideas for the study of industry development.

B. Text Recommender System

Traditional text recommender methods concentrate on human-designed features. The most widely used methods is to represent text as a vector of terms, called Bag-of-Words. However, such methods have a serious problem regarding data sparsity. Then the distributed representation was first proposed by Hinton in 1986 [10]. The idea of its dependence is that the semantics of words are determined by contextual information,

TABLE I: Data Examples

Transaction Content	Transaction Code	Transaction Specification	Transaction Unit	Transaction Price
Steel Pipe	1020103000000000000	159*10	Ton	2790
Belt	1050202040000000000	450	Piece	46
Bearings	1090123010000000000	720	Set	48
Laptop	Null	13 inch	Set	16000

that is, words that appear in the same context have similar semantics. The word is mapped into a new space and expressed as a multidimensional continuous real number vector called "Word Embedding". Words use sequential vectors [11] for Embedding, which can be initialized using pre-trained word vector. Multi-Layer Deep Neural Network (DNN), CNN [12], [13], or Recurrent Neural Network (RNN) [14] is used for high-dimensional abstract feature extraction. There are also studies that concentrate on applying character representation [15] embedding and attention mechanisms [16] to text. But no studies have recommended the category of transaction behavior based on tax invoices. Hence, we propose a recommender system combining the feature from transaction unit using CNN and the feature from transaction content using word-level and char-level Bi-RNN with attention mechanism to recommend the corresponding codes for the transaction behavior.

III. Data Description

In this Section, we describe our dataset which is extracted from tax invoice, which contains the following five fields: transaction content, transaction code, transaction specification, transaction unit and transaction price. The sample data is shown in Table. I. The transaction code is a 19-digit string, which represents the category of transaction content. However, transaction code field in most of the data is empty (like the fourth row of laptops in Table. I) since this field is often not filled in during the invoicing. In order to classify every transaction behavior to an official transaction code, we will present a neural network model for classification in Section IV.

The transaction code set is a tree structure with a total number of 4226. It is divided into six major categories, i.e., the code of goods is starting with 1, the code of labor service starting with 2, the code of sales service is starting with 3, the code of intangible assets is starting with 4, the code of real property is starting with 5, and the code of no-tax item without sales action is starting with 6. The number of leaf nodes is 3,550. 90% of the total number of leaf codes belong to the category of goods. Simultaneously, the sales service category code accounted for 8%.

The data used for the experiment is collected during the whole 2017 from Zhejiang Province, China. Zhejiang is a large coastal province of China with a population of 56.57 million. The land area is 105,500 square kilometers. At the same time, the number of registered companies in Zhejiang Province is 8,282,121 and the number of invoice data record is 203,441,178. According to industrial structure, high-end equipment manufacturing industry is still the mainstay and the service industry is also growing rapidly. The growth rate (27.1%) of operating income of service industry enterprises ranks first in China. It is well-known that Alibaba Group locates in Hangzhou, Zhejiang Province, which is an e-commerce firm that provides consumer-to-consumer and business-to-business sales services via web portals and promote the development of all walks of life.

IV. Recommender System Based on Transaction Behavior Categories

In this Section, we present a neural network recommender system which extracts the word and character features [17] with attention machanism to recommend the transaction code. Then we describe the system architecture as shown in Fig. 2 and how to extract the features of the short text from the words and characters as well as configuration and parameters.

Transaction content can be regard as a short text. For a short text, we can split the text through jieba segment [18], [19], $S = \{w_1, w_2, , w_n\}$, where S is the set of words, and w_i represents the i-th word. The network takes the input as a sequence of N tokens, which are contained by a finite word dictionary.

Embedding layer: The emdedding layer transforms the short text into a matrix of embedding, denoted as $M \in R^{(n\times m)}$ as the input of the network, where n is the maximum number of words. And m is the dimension of word embedding. We obtain matrix M by concatenating the embedding of words together.

And $\bigoplus$ is the concatenation operation. Suppose the short text consists of n words, and $v_i^w \in R^m$ is an m-dimensional vector to represent the i-th word in the short text. We can obtain M by simply concatenating them:

$$M = v_1^w \oplus v_2^w \oplus ... \oplus v_n^w \tag{1}$$

In our experiment, we use the word vector pre-trained by word2vec pre-trained by Baidu Encyclopedia corpus [20].

Bi-RNN layer: We use a bi-directional RNN to get annotations of words by obtaining features from both directions for words, and therefore incorporate the contextual information. The bidirectional RNN consists of the forward RNN which reads the text from w_1 to w_T and a backward RNN which reads from w_T to w_1:

$$\overrightarrow{h_t} = \overrightarrow{RNN}(x_t), \overleftarrow{h_t} = \overleftarrow{RNN}(x_t), t \in [1, T] \tag{2}$$

We obtain an annotation for a given word w_t by concatenating the forward hidden state $\overrightarrow{h_t}$ and backward hidden state $\overleftarrow{h_t}$, i.e., $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$, which combines the information of the whole text centered around w_t. Similarly, the char-level simply replaces the words mentioned below with chars.

Attention layer: Not all words or chars contribute equally to the representation of the transaction content meaning. Hence, we use attention layer [16] to extract such words or chars that are significant to the semantic meaning of the text and weight the representation of those words or chars to compute a feature vector as follow:

$$u_t = relu(W_w h_t + b_w) \tag{3}$$

$$\alpha_t = \frac{exp(u_t^{\mathrm{T}} u_w)}{\sum_t exp(u_t^{\mathrm{T}} u_w)} \tag{4}$$

$$s = \sum_t \alpha_t h_t \tag{5}$$

We feed h_t through a simple multilayer perceptron(MLP) to get a hidden representation. Then, we use softmax function to normalize importance weight α_t computed by the similarity of u_t with context vector u_w. After that, we compute the vector s as a weighted sum annotations bases on the importance of word to text.

Convolution layer: The convolution layers are to extract higher level features from the word embedding matrix. To get different dimensions of features, we apply filters with different kernel sizes. Similar to many previous works, we fix the width of each filter as m and treat the height h of it as a hyper parameter. Given a filter $\omega \in R^{h \times m}$, a feature f_i is generated from a window of words $[v_i : v_{i+h-1}]$ by:

$$f_i = g(\omega[v_i : v_{i+h-1}] + b) \tag{6}$$

Here $b \in R$ is a bias. And g is a non-linear function. The filter is applied to all possible windows of words in M to produce a feature map $f \in R^n$ due to same padding mode.

Pooling layer: The intention of pooling layer is to further abstract the features generated from convolution layer by aggregating the values for each filter. In this work, we apply a max-over-time pooling operation over each feature map.

Full Connected layer: We use a non-linear full connected layer to combine various pooling features. In this layer, we can also apply dropout [21] as a mean of regularization by randomly setting to zero a proportion of elements of the feature vector to prevent over-fitting.

Output layer: Using the softmax activation function in the output layer to represent a categorical distribution over class labels, and obtaining the probabilities of each input belonging to a label. Due to the fact that the transaction content in the invoice data corresponds to multiple codes, we extract the top K class codes with the highest frequency for the same transaction content (fewer than K items are calculated by less than K), and normalizes them into probabilities.

$$p(c_i|content) = \frac{freq(c_i|content)}{\sum_{i=1}^{n} freq(c_i|content)} \tag{7}$$

Here, c_i stands by the class i, and $p(c_i|content)$ represents the probability that content belongs to class i. We fill the probability in the corresponding code and the rest codes with zero to generate the representing label vector.

In this work, we apply a max-over-time pooling operation over each feature map to further abstract the features generated from convolution layer as pooling layer. We use a non-linear full connected layer to combine various features. We also apply dropout as a mean of regularization by randomly setting to zero a proportion of elements of the feature vector to prevent over-fitting. Finally, we make use of the softmax function in the output layer to represent a categorical distribution over class labels, and obtaining the probabilities of each input belonging to corresponding codes.

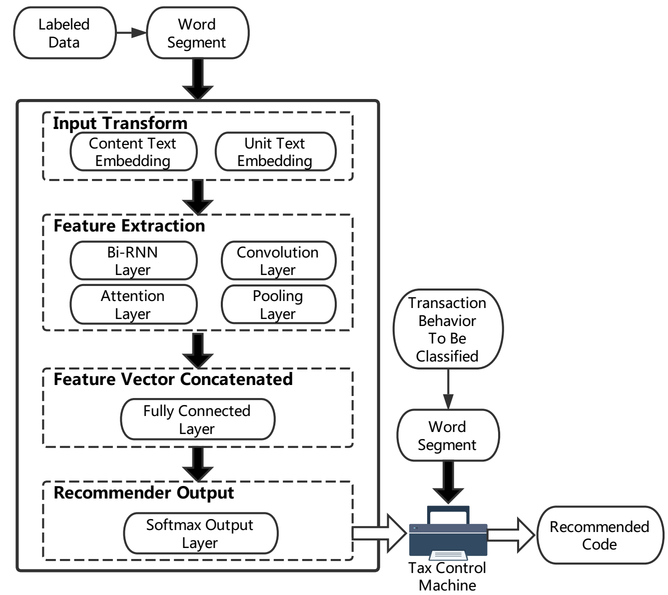


Fig. 2: System Architecture

V. Experiment and Analysis

A. Experiment

For the CPU we used Intel 6950x 3.0GHz with 64GB memory, and for GPUs we used Nvidia GTX 1080TI. In order to improve the speed of model training, make full use of the resources of GPUs, we use in-graph replication with synchronous training to parallelize the models training across multiple GPUs installed in the local machine. The controller device will be the CPU, meaning that all variables will live on this device and will be copied to the GPUs in each step. And the CPU has the task of summarizing the loss of each GPU and calculating gradients. We randomly select data sets according to the ratio of 7:2:1 and put them into the training set, test set, and verification set respectively. We use Adam as the optimization method, and the stopping strategy is that the loss does not decline in 1000 continuous iterations. From the Tab. II, we can see that our compositional recommender model performs better than others. And we show the details of the model configuration in Fig. 3, along with the hyper parameters of the model in the Tab. III.

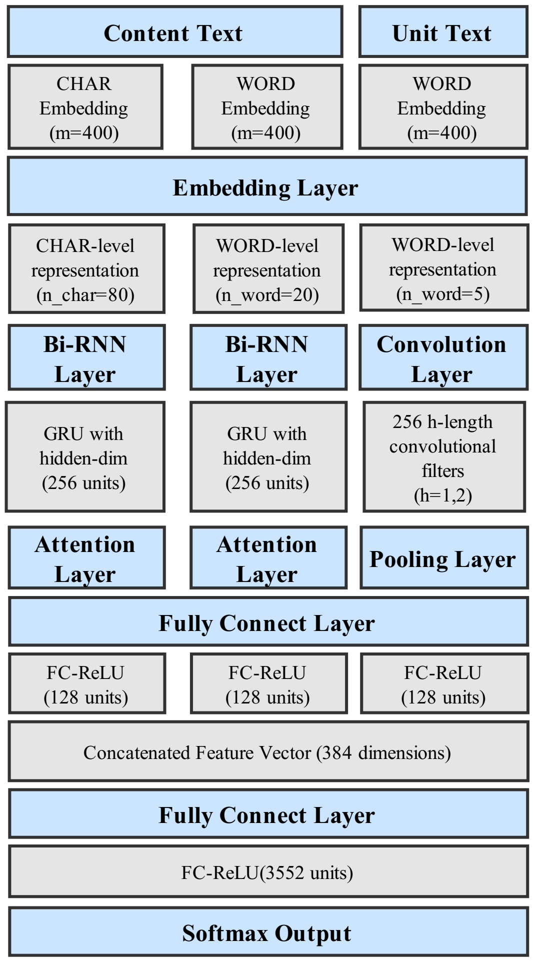


Fig. 3: Model Configuration

TABLE II: Experiment Results. We compared only using word-level RNN, word-level and char-level RNN without using feature of transaction unit, our compositional model without using attention and our compositional model on relevant indicators.(K=5)

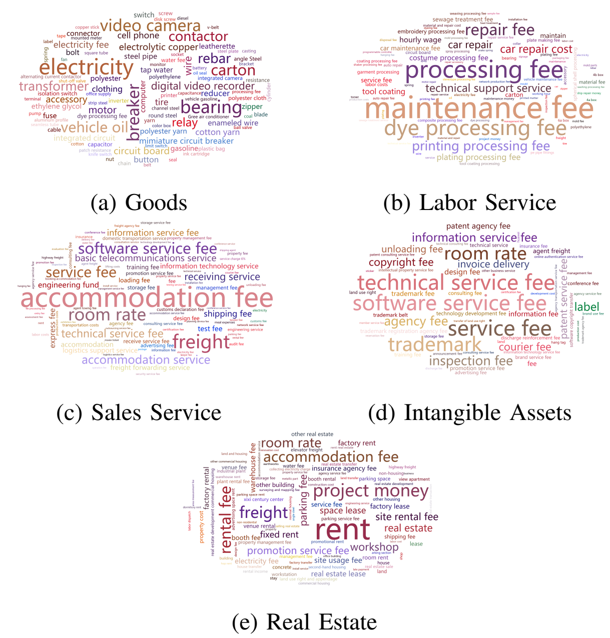
	Accuracy	Precision	F1-score
Word	0.58	0.57	0.57
Word+Char	0.69	0.67	0.67
No Attention	0.72	0.71	0.70
Compositional	0.75	0.73	0.73

B. Analysis

After using the recommender system proposed in the Section IV to recommend transaction code for unlabeled data, we label the transaction behavior with the top-one code in recommender system in the following analysis. Firstly, we focusing on the quantity distribution of invoice data. Meanwhile, we use visualization method to find out the characteristics of transaction contents in each major category. After that, we studied the price distribution of various transaction behaviors to discover the difference in price distribution between different industries.

TABLE III: Hyper Parameters

Parameter	Values
embedding dimension	400
sequence dimension	$content$: $[80, 20]$ $unit$: $[5]$
filter sizes	$[1, 2]$
filter kernel number	256
dropout rate	0.5
hidden layers dimension	256
learning rate	$\alpha = 0.01$



(a) Goods (b) Labor Service (c) Sales Service (d) Intangible Assets (e) Real Estate

Fig. 4: The word cloud for five major categoies except for Non-Tax Items

1) Word Cloud and Quantity Distribution of Invoice Data: In this part, we pay attention to the word cloud and quantity distribution of categories of invoice data. Fig. 4 demonstrates the word cloud for five major categories except No-Tax items. We can see that daily necessities and industrial products, such as electricity and oils are included in the word cloud of Goods category. In the labor service category, words with high frequency are about processing fees and maintenance fee. And the sales service includes life service and accommodation services, so it is reasonable to have more service fees and accommodation fees. In intangible assets, we can see more technical or software service fee, similar to patents and trademark related projects. It is quite clear that in the real estate category, since real estate is mainly related to buildings, it can be seen that there is a high probability that words with high frequency are related to buildings.

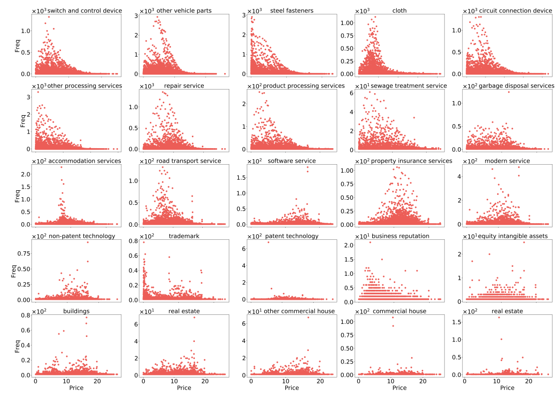
Table. IV shows the quantity distribution of the number of each major categories. The number of invoices of goods accounted for more than 90% of the total number of invoices followed by sales services and labor services. However, the number of invoices for non-tax item categories was only 318.

TABLE IV: The number of invoice record for major categories

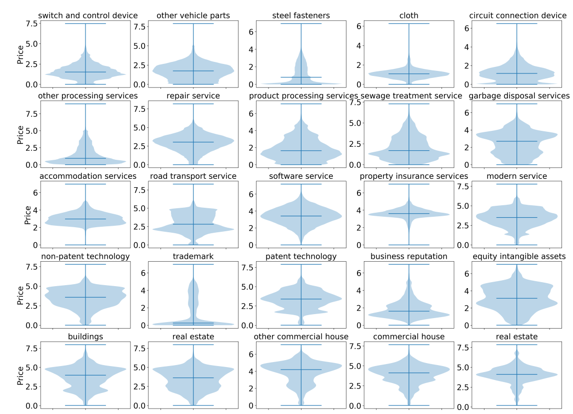
Category	Goods	Labor Service	Sales Service	Intangible Assets	Real Estate	Non-Tax Items
Numbers	1.53×10^8	3185832	11735649	132279	139727	318

It is quite a certain that there is a certain similarity between the quantity distribution of invoices and the distribution of code set issued by administration of state. Combining with the staticis in Section III, we can draw a conclusion that the major category with larger number of invoice records is more diversified and accompanied by an increase in the number of subdivided categories.

2) Price Distribution of Different Industry: We extract the top five sub-categories with the number of invoice records for each major category. Each small category actually corresponds to a certain industry, and then according to the unit price of the invoice data, the price scatter distribution for each small category is plotted in Fig. 5a and data distribution and probability density is shown in Fig. 5b. In longitudinal comparison, we can see that the points in the first row of the fig. 5 representing for goods are obviously more intensive than those of other categories. The results are consistent with the previous statistics: Firstly, the codes of Goods accounts for 85% of the tax classification code set, reflecting a wide variety of goods. At the same time, we can see that prices in the major categories of goods are more concentrated in the middle or lower prices.The second major category is mainly the labor service. The first is the other processing labor service, reflecting that the industry is mainly manufacturing-based. In the third major category (sales service), you can see the points in the Fig. 5 are much sparser, and the price of labor services tends to be cheaper. As we all know, Zhejiang is also a large province of labor input.The third major category is sales service. It can be seen that the overall price of software service fee category is relatively high, reflecting the rapid development of the technology industry led by the Internet which results in higher wages in this industry. In other sub-categories, we can see that the prices are more distributed in the middle range, which can also reflect the better development of the tertiary industry (service industry) in Zhejiang Province.The fourth category is intangible assets. Through the results of the word cloud, we can see the service fees mainly in terms of knowledge, trademarks, and property rights, such as technical service fees. Relatively, the prices of intangible assets are more widely covered than they are. Low prices embodied in the categories of goods and services occupy a dominant position.The distribution curves of the price of real estate is relatively uniform, and the frequency of high prices and low prices is also relatively similar. The fifth category of representative words is mainly real estate, which shows that real estate in this industry has both high prices and low prices. There are housing rental services, high-priced commercial housing (high prices in Zhejiang). It can reflect that the price structure of the industry is better and the industry can cover



(a) Scatter Plot (The horizontal and vertical coordinates are the price and it's frequency)



(b) Violin Plot (The vertical coordinate is the price)

Fig. 5: The price distribution for five major catrgories

the needs of various groups.

VI. Conclusion

In this work, we propose a recommender system based transaction behavior category to recommender the transaction code for transaction behavior. In this way, we map the diverse transaction behavior in the invoice to transaction code in the Tax Classification and Coding for Commodities and Services issued by the state administration of taxation, to help us deeply understand the transaction behavior of a region from the perspective of industry.

We focus on the characteristics in category distribution of invoice data. We find that the number of invoice records is highly diverse among the main categories of Goods.

More specifically, interesting results include: the prices of major categories of goods are mainly distributed in the middle-to-lower price range. The number of goods in the high price is small, and the industries in the labor service category have similar conclusions. However, in the major categories of sales services and intangible assets, there are subdivisions. The price distribution of the industry is relatively uniform and not concentrated in a certain price range, indicating the structure of that industry price is relatively healthy.

VII. Future Work

Our work only relies on analysis of invoice data. However, with the development of multi-source data fusion, we can obtain data from multi-source in the future. Then we can analyze the development of enterprises in different industries in the region more comprehensively. In this way we can provide real data driven strategies to analyze the development of the industry in the region, and further examine the methods we proposed according to feedbacks from industry.

VIII. Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (61701031, 61671078), Funds of Beijing Laboratory of Advanced Information Networks of BUPT, Funds of Beijing Key Laboratory of Network System Architecture and Convergence of BUPT, and 111 Project of China (B08004, B17007).

References

[1] Dilip K Das, "The role of china in asia's evolution to global economic prominence," *Asia & the Pacific Policy Studies*, vol. 1, no. 1, pp. 216–229, 2014.

[2] investopedia, "Bussiness activities," .

[3] chinatax, "Notice of the state administration of taxation on conducting the pilot program of tax classification and coding for commodities and services," .

[4] Zhanyu Ma, Jiyang Xie, Hailong Li, Qie Sun, Zhongwei Si, Jianhua Zhang, and Jun Guo, "The role of data analysis in the development of intelligent energy networks," *IEEE Network*, vol. 31, no. 5, pp. 88–95, 2017.

[5] Seung Ho Park, Shaomin Li, and K Tse David, "Market liberalization and firm performance during china's economic transition," *Journal of International Business Studies*, vol. 37, no. 1, pp. 127–147, 2006.

[6] John Muellbauer and Anthony Murphy, "Housing markets and the economy: the assessment," *Oxford review of economic policy*, pp. 1–33, 2008.

[7] Huina Mao, Xin Shuai, Yong-Yeol Ahn, and Johan Bollen, "Mobile communications reveal the regional economy in côte divoire," *Proc. of NetMob*, 2013.

[8] Pamela Castelln Gonzlez and Juan D. Velsquez, "Characterization and detection of taxpayers with false invoices using data mining techniques," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1427–1436, 2013.

[9] Weibo Xing and John Whalley, "The golden tax project, value-added tax statistics, and the analysis of internal trade in china," *China Economic Review*, vol. 30, pp. 448–458, 2014.

[10] Geoffrey E Hinton, James L McClelland, David E Rumelhart, et al., "Distributed representations," *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, no. 3, pp. 77–109, 1986.

[11] Zhanyu Ma, Jing-Hao Xue, Arne Leijon, Zheng-Hua Tan, Zhen Yang, and Jun Guo, "Decorrelation of neutral vector variables: Theory and applications," *IEEE transactions on neural networks and learning systems*, 2016.

[12] Jiachen Du, Lin Gui, Yulan He, and Ruifeng Xu, "A convolutional attentional neural network for sentiment classification," in *Security, Pattern Analysis, and Cybernetics (SPAC), 2017 International Conference on*. IEEE, 2017, pp. 445–450.

[13] Jin Wang, Zhongyuan Wang, Dawei Zhang, and Jun Yan, "Combining knowledge with deep convolutional neural networks for short text classification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, pp. 2915–2921.

[14] Wei Cao, Anping Song, and Jinglu Hu, "Stacked residual recurrent neural network with word weight for text classification.," *IAENG International Journal of Computer Science*, vol. 44, no. 3, 2017.

[15] Wang Ling, Tiago Luís, Luís Marujo, Ramón Fernandez Astudillo, Silvio Amir, Chris Dyer, Alan W Black, and Isabel Trancoso, "Finding function in form: Compositional character models for open vocabulary word representation," *arXiv preprint arXiv:1508.02096*, 2015.

[16] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480–1489.

[17] Xiang Zhang, Junbo Zhao, and Yann LeCun, "Character-level convolutional networks for text classification," in *Advances in neural information processing systems*, 2015, pp. 649–657.

[18] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, vol. 1, pp. 1555–1565.

[19] Omer Levy and Yoav Goldberg, "Neural word embedding as implicit matrix factorization," in *Advances in neural information processing systems*, 2014, pp. 2177–2185.

[20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[21] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.