A Hierarchical Model of E-Commerce Sellers Based on Data Mining

Xiuyan Bai

Hangzhou Vocational and Technical College, Hangzhou 310018, China

Corresponding Author Email: adenbai@163.com

ABSTRACT

This paper attempts to accurately classify e-commerce sellers based on data mining. Firstly, the original data from an e-commerce platform were preprocessed, and the classification indices were identified from five categories (products, users, traffic, sales and basic attribute). Next, the principal component analysis (PCA) and the self-organizing feature map (SOM) were fused into a hierarchical model that divides e-commerce sellers into three categories: large sellers, medium sellers and small sellers. The effectiveness of our model was verified through experiments. Finally, several operating strategies were put forward for e-commerce sellers in each category. The research results provide a good reference for the development of the e-commerce industry.

1. INTRODUCTION

With the rise of online shopping, e-commerce has become a research hotspot. However, the relevant studies mainly focus on the interests of users, such as the security of online transactions. There are very few reports on sellers, not to mention their hierarchy and specific operations.

The hierarchy of e-commerce sellers is usually defined by gross merchandise value (GMV) [1]. Based on the GMV, e-commerce sellers can be divided into large, medium and small sellers [2, 3]. Considering the sheer number of sellers on e-commerce platforms, the simple and rough classification is unable to support refined operations. Therefore, it is necessary to develop a comprehensive hierarchical model for sellers across e-commerce platforms.

This paper mainly proposes a hierarchical model for e-commerce sellers based on data mining [4]. Firstly, the data mining tools selected for modelling were introduced, namely, the self-organizing feature map (SOM) and principal component analysis (PCA). Next, the data processing and index selection were described in details. After that, the SOM and the PCA were fused into our hierarchical model. Finally, the effectiveness of our model was verified through experiments [5-8].

2. PRELIMINARIES

2.1 SOM

The SOM was proposed by Finnish professor T. Kohonen, a doctor at the University of Helsinki. Prof. Kohonen held that the neurons in a neural network (NN) respond inconsistently to external stimuli, and the entire NN can be automatically divided into different regions based on the responses [9-11].

The SOM has a strong ability to learn the intrinsic attributes of every input, efficiently acquire the statistical features of the input, and iteratively update the parameters and structure of the NN. The competitive learning ability naturally leads to the excellence in self-organization. The SOM can continuously process and transmit the various dimensional information of the input dataset through network neurons, eliminating complex operations like derivation and differentiation.

The SOM outperforms other algorithms of its kind in computing efficiency and feature learning. On the one hand, the SOM updates the neuron weights iteratively based on the latest learning results, and assigns the highest weight to the optimal neuron in each round; in this way, SOM algorithm can quickly converge to the global optimal solution. On the other hands, the SOM can acquire the intrinsic statistical features of input samples and reflect the probability distribution of training samples. Therefore, the SOM is suitable for classification of e-commerce sellers based on a massive amount of data.

2.2 PCA

Empirical problems generally involve various inter-correlated influencing factors, a.k.a. indices. The correlation between some indices is rather strong. If all indices are combined into a dataset, there must be many overlaps between the information carried by different data elements. If there are too many indices, the problem analysis will become very complex and require a huge computing power. In severe cases, the algorithm may face overfitting or fail to find the optimal solution [12, 13].

The PCA provides a viable solution to the above defects. This dimensionality reduction strategy mainly transforms multiple indices into a few composite indices: the original data are reduced in dimensionality; then, the useful indices are selected and imported to the algorithm, improving the analysis accuracy. The PCA was selected to preprocess the original data in our research, because the data involve five primary indices (products, users, traffic, sales and basic attribute), each of which contains multiple closely coupled secondary indices [14, 15].

3. DATA PROCESSING AND INDEX SELECTION

The original data are the user behaviors (e.g. "page view", "add to cart" and "order placement") that reflect the traffic and sales of 140,000 e-commerce sellers on a Chinese e-commerce platform. The data were collected from multiple channels, including the app and website of the platform and instant messengers like WeChat and QQ.

The original data from different databases were allocated to Hadoop clusters, processed by MapReduce, and aggregated into the raw data. The raw data were examined in details and then preprocessed in five steps: missing value treatment, outlier processing, skewed distribution correction, normalization, and correlation analysis.

(1) Missing value treatment

The R-language was adopted to remove redundant entries and supplement the missing data [4, 16].

(2) Outlier processing

The calculation errors and outliers, i.e. the entries with business field of zero, were processed according to the business logic. First, the entries of the stores with no transaction were deleted; Second, the outliers containing the information of actual transactions were retained; Third, the other outliers that deviate greatly from the normal range were processed to ensure the accuracy of subsequent normalization [17].

(3) Skewed distribution correction

Most indices in the original data have left-skewed distributions, which reflect the Pareto principle: the few large sellers attract most traffic and sales on the e-commerce platform, i.e. the winner takes all. To solve the problem, the psych package of the R language was employed to obtain the descriptive value of each index whose distribution skews to the left [18]. Then, the bcPower function was called from the car package of R language to reduce the skewness of the skewed distributions.

(4) Normalization

For convenience, each number was normalized as a decimal between zero and one, and the dimensional expressions were made dimensionless. The normalization does not change the size of the dataset.

(5) Correlation analysis

If two indices in the same dataset are highly correlated (i.e. too close to each other), much computing power will be wasted to handle identical or highly similar information in the subsequent iterative computations. Here, the correlation coefficients between the indices in our dataset are calculated by the cor function of R language. Then, the correlations between indices were displayed with the ggcorrplot, a visualization tool [19, 20].

The indices of the preprocessed data were roughly divided into five categories: products, users, traffic, sales and basic attribute. The indices in the five categories are respectively explained in Tables 1-5, where SKU is stock keeping unit (a number or code used to identify products in most online stores), SPU is standard product unit (the smallest unit of product information used by most online stores), PV is page view (the number of pages viewed by a visitor), and UV is unique visitor (an aggregate of PVs generated by the same user during the same session).

Table 1. The indices in the category of products

Category	Symbol	Meaning
	SPN (SKU)	The number of sold products (SKU)
	SPN (SPU)	The number of sold products (SPU)
	SPR (SKU)	The ratio of sold products to displayed products (SKU)
	SPR (SPU)	The ratio of sold products to displayed products (SPU)
	VBCR	The conversion ratio of visitors to buyers
	ACPN (SKU)	The number of add-to-cart products (SKU)
Products	ACPN (SPU)	The number of add-to-cart products (SPU)
	ACPR	The ratio of add-to-cart products to displayed products
	VPN (SKU)	The number of visited products (SKU)
	VPN (SPU)	The number of visited products (SPU)
	DPN (SKU)	The number of displayed products (SKU)
	DPN (SPU)	The number of displayed products (SPU)
	OPN	The number of ordered products

As shown in Table 1, four indices in the category of products have both SKU and SPU dimensions: the number of sold products, the number of add-to-cart products, the number of visited products and the number of displayed products.

As shown in Table 2, the indices in the category of users all focus on users, and the users follow two kinds of items:

product and store.

As shown in Table 3, the SPV is the sum of the PVs on any page of the store, including but not limited to the front page, the product details page, the query page and the promotion page; similarly, the SUV is the sum of the UVs on any page of the store.

Table 2. The indices in the category of users

Category	Symbol	Meaning
	ACUN	The number of users that add products to cart
	FGUN	The number of users that follow one or more products
	FSUN	The number of users that follow one or more stores
Users	POUN	The number of users that place one or more orders
	OBN	The number of old buyers
	NBN	The number of new buyers
	30DRT	The 30-day repurchase rate
	90DRT	The 90-day repurchase rate

Table 3. The indices in the category of traffic

Category	Symbol	Meaning
	SPV	Store PV
	SUV	Store UV
	ATP	Average time on page (s)
	ADV	Average depth of visit
	BR	Bounce rate
Traffic	PE (SKU)	Product exposure (SKU)
	PE (SPU)	Product exposure (SPU)
	PPV	Product page views
	PVN	Number of product visitors
	PVPCR	The conversion ratio of product visitors to order placers
	SVPCR	The conversion ratio of store visitors to order placers

Table 4. The indices in the category of sales

Category	Symbol	Meaning
	OA	Order amount
	ON	The number of orders
Sales	PBA	Per buyer amount
	OPON	The number of orders paid online
	OPDN	The number of orders paid upon delivery
	OPTN	The number of orders paid by bank transfer

As shown in Table 4, the indices in the category of sales all focus on user purchases. There are three payment methods for e-commerce users: online payment, payment upon delivery and bank transfer.

Table 5. The index in the category of basic attribute

Category	Symbol	Meaning
Basic attribute	SOY	The number of years since the opening of the online store

As shown in Table 5, a store with large SOY tends to gain rich sales experience. The greater the SOY, the more likely it is for the seller to become a large seller. Therefore, the SOY is an important reference for the classification of e-commerce sellers.

4. PCA-SOM HIERARCHICAL MODEL

4.1 The PCA phase

The PCA is the most popular unsupervised algorithm for dimensionality reduction of features. Through the PCA, the correlated high-dimensional indices are linearly mapped to the low-dimensional space. The resulting low-dimensional indices are called principal components. The nonlinear mapping is comparable to transforming the original dataset to a new coordinate system containing n orthogonal coordinates. The first coordinate is the first principal component, the second coordinate is the second principal component, and the rest may be deduced by analogy. Based on the maximum variance theory, the first principal component has the greatest explanatory power of the original dataset, followed in descending order by the second to the *n*-th principal component [21, 22].

As the first part of our hierarchical model, the PCA is performed on the normalized dataset, using the FactoMineR and factoextra packages of R language. According to the Kaiser-Harris criterion, the principal components whose eigenvalues are greater than one should be selected to represent the feature space of the original dataset, because these components are useful and capable of explaining more than two original indices. To capture sufficient feature information, all the principal components whose eigenvalues are greater than 0.4 were selected to form the training set. The established training set contains 147,008 entries, each of which covers features in 14 dimensions.

4.2 The SOM phase

As shown in Figure 1, the SOM, the second part of our hierarchical model, is trained through the following steps:

First, the number of output layer neurons is configured and inputted into the model, and the weight of each output layer neuron is initialized as a small random number.

Second, the input data and weights are normalized. Although every index has been normalized in preprocessing, every entry and weight must be normalized again to form a consistent training set.

Third, some samples are collected from the dataset by a sampling module and taken as training samples. The sampling is highly necessary, as it is difficult to train the model with all the147,008 entries in the preprocessed dataset. Distance calculation is essential to model training. Thus, Euclidean distance and cosine distance are compared to find which is more suitable for the training of our model.

Fourth, weight update is implemented after each input for the neurons in the neighborhood radius of the winning neuron, and the updated weights are normalized again.

Fifth, the learning rate and neighborhood radius are updated as two functions, waiting to be called for model training. The two parameters both decreases with the growing number of iterations.

Sixth, the model training is terminated under one of the following conditions: the learning rate falls below the preset threshold; the number of iterations surpasses the preset maximum number of iterations.
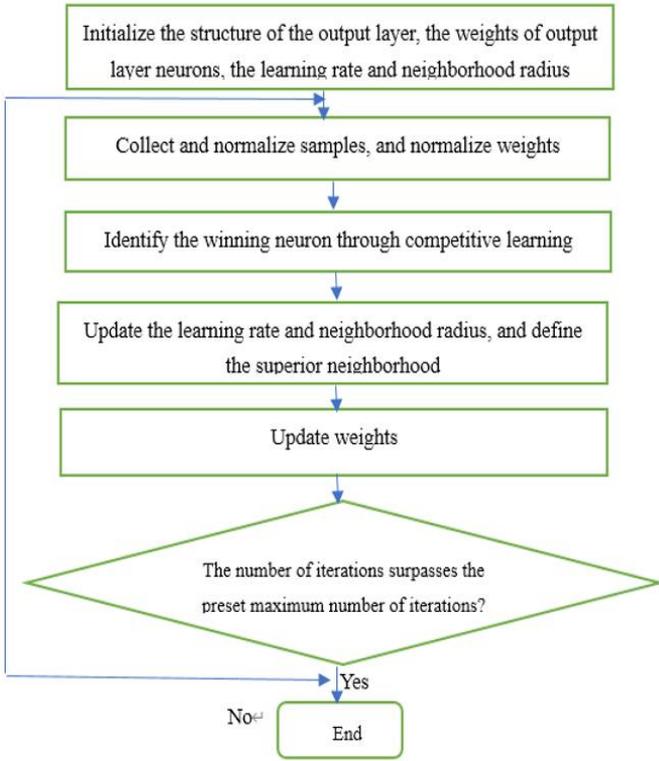
Figure 1. Workflow of the SOM

4.3 Workflow of PCA-SOM model

In general, the PCA-SOM model can be executed in nine steps:

Step 1. Dataset reading

Filter out the store IDs and save feature data of 14 dimensions. Read the data with the pandas Python library. To facilitate subsequent calculations, convert the obtained DataFrame into the datatype of the NumPy Array.

Step 2. Data initialization

First, set up the structure output layer neurons as a 2× 2tput l array, which contains four 14-dimensional neurons. Next, initialize the data of the 14×4 matrix in a random manner, plus another two parameters (i.e. size of sampling batch and number of iterations). Define these parameters as passable in the coding process, laying the basis for tuning and reference in subsequent experiments.

Step 3. Data normalization

Normalize the sampled dataset and output matrix of weight errors before each round of competitive learning:

$$x_i = \left. x_i \middle/ \sqrt{x_1^2 + x_2^2 + ... + x_n^2} \right. \quad (x_1, x_2, \cdots, x_n) \tag{1}$$

Step 4. Competitive learning

Compute the distances between each sample and the weights of the four output layer neurons. Since the data have been normalized, replace the distance computation with the dot product of weights, i.e. the multiplication between two matrices, and then select the winning neuron for each sample.

Step 5. Weight update

Based on the winning neuron and neighborhood radius, identify the weights to be updated, and update them at the preset learning rate:

$$W_j(n+1) = W_j(n) + \eta_{i(x)j}(n)(X(n) - W_j(n)) \tag{2}$$

Step 6. Update of neighborhood radius and learning rate

Update neighborhood radius and learning rate after weight update in each round, such that the model can adapt to the latest environment. The neighborhood radius can be updated by:

$$N = a - \frac{a*t}{iter} \tag{3}$$

where, a is the short-edge distance of the output layer; t is the current number of iterations; iter is the maximum number of iterations. Obviously, the neighborhood radius is negatively correlated with the number of iterations.

The learning rate can be updated by:

$$eta = \frac{e^{-n}}{t+2} \tag{4}$$

where, n is the current neighborhood radius; t is the current number of iterations. Obviously, the learning rate is also negatively correlated with the number of iterations.

Step 7. Checking termination condition

If the current number of iterations is greater than the maximum number of iterations, terminate the training and save the latest weights.

Step 8. Sample prediction

According to the latest weights, calculate the level (category) of each seller, and save the predicted dataset.

Step 9. Visualization

Visualize the predicted dataset with PyLab, such that each category is displayed in a unique color.

5. EXPERIMENTAL VERIFICATION

5.1 Parameter settings

Table 6. Correlations between 14 principal components and four categories in Parameter Set A

Label	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7
0	0.145462	0.114938	0.368108	0.391941	0.089378	0.263032	0.175507
1	0.252905	0.18781	0.218794	0.15147	0.623843	0.086672	0.120722
2	0.898651	0.140505	0.080103	0.080886	0.075871	0.063685	0.065003
3	0.24596	0.810467	0.105459	0.085603	0.114202	0.07422	0.077887
Label	**Dim8**	**Dim9**	**Dim10**	**Dim11**	**Dim12**	**Dim13**	**Dim14**
0	0.137724	0.080217	0.158189	0.086634	0.057333	0.063127	0.089902
1	0.132	0.083466	0.073325	0.067509	0.063904	0.074067	0.089163
2	0.046135	0.033628	0.040682	0.024839	0.022028	0.024663	0.030701
3	0.060589	0.056506	0.080123	0.03729	0.039968	0.04165	0.048224

Through repeated tunings, two sets of parameters were selected to train our model: Parameter set A: maximum number of iterations *iter*=5; size of sampling batch *batchsize*=20,000; Parameter set B: maximum number of iterations *iter*=10; size of sampling batch *batchsize*=20,000. After the training, the final dataset under each parameter set was saved to analyze the hierarchy of e-commerce sellers.

Since the data on each seller contain 14 principal components, the correlations between each principal component and each category were obtained through clustering. The results under each parameter set are recorded as a correlation matrix (Tables 6 and 7). The correlation matrix represents the distribution of eigenvalues of each category across the 14 principal components [23, 24].

Table 7. Correlations between 14 principal components and four categories in Parameter Set B

Label	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7
0	0.144169	0.13101	0.457498	0.319664	0.087179	0.246859	0.19463
1	0.215309	0.214191	0.146652	0.224041	0.594014	0.098243	0.124524
2	0.89625	0.141599	0.078692	0.081162	0.081033	0.064357	0.064039
3	0.246716	0.825063	0.093644	0.085554	0.111229	0.072239	0.072092

Label	Dim8	Dim9	Dim10	Dim11	Dim12	Dim13	Dim14
0	0.118006	0.089593	0.135805	0.08509	0.057211	0.057739	0.076073
1	0.147773	0.083255	0.100536	0.069718	0.063902	0.077903	0.108046
2	0.048089	0.033167	0.042368	0.025315	0.022536	0.025333	0.031392
3	0.058085	0.054796	0.077205	0.035749	0.039071	0.041042	0.046528

5.2 Results analysis

As shown in Table 6, the eigenvalues of category 0 mainly appeared in principal components 3 and 4, those of category 1 in principal component 5, those of category 2 in principal component 1, and those of category 3 in principal component 2. The comparison between Tables 6 and 7 shows that the eigenvalues were distributed similarly between the two parameter sets. The following results were drawn from the two tables:

(1) The top 5 principal components explain over 77% of the variance in the entire dataset.

(2) Principal component 1 (Dim1) explains 41% of the variance in the dataset; the greater the variance, the more important the component. This component has a significant positive correlation with the indices of absolute values in the categories of traffic, products and sales.

(3) Principal component 2 (Dim2) explains 16% of the variance in the dataset. This component has a significant positive correlation with the indices of ratios (conversion ratios) in the categories of products and traffic, and a significant negative correlation with indices about sold products, displayed products and visited products.

(4) Principal component 3 (Dim3) has a significant positive correlation with BR and ADV. It can be abstracted as the decoration quality and attractiveness of a store.

(5) Principal component 4 (Dim4) is positively correlated with BR and ADV, as well as several indices of ratios (conversion ratios).

(6) Principal component 5 (Dim5) has a relatively weak correlation with the indices. The only exceptions are the strong correlation with SOYN, and slight correlations with several indices of ratios (conversion ratios).

5.3 Discussion

The sellers in Category 0 are featured by high levels of BR, ADV, VBCR and SVPCR. The high BR and ADV indicate that the stores are well decorated and organized; Upon accessing a page of such a store, the user prefers to browse the related pages, rather than jump elsewhere in a short time. The high VBCR and SVPCR reflect the good traffic value of these stores, i.e. many visitors are converted to buyers per unit of traffic [25].

The sellers in Category 1 have a relatively long SOYN. Among the top 5 principal components, principal component 5 plays the greatest role among these sellers. Hence, many of these sellers have being operating for many years. Meanwhile, the retention and conversion ratios of these old sellers are not desirable, as they fail to update the operation mode according to the constant changes in user preference and platform policy [26].

The sellers in Category 2 stand out for the indices of absolute values in the categories of traffic, products and sales. These indices are yardsticks of the scale of stores. Thus, Category 2 sellers generally have many displayed products, high traffic and good sales. However, these sellers perform poorly in retention and conversion ratios. As shown in Tables 6-7, these sellers lag behind their counterparts of other categories in Dim1-4, especially in Dim1. These results show that the excellence of these sellers in sales comes from the traffic brought by the scale of products, instead of high traffic value.

The sellers in Category 3 have high ratios (conversion ratios) in the categories of products and traffic, but very low results in indices about sold products, displayed products and visited products. These sellers must have a small scale of products. The high ratios (conversion ratios) are attributable to the small base of products. In general, Category 3 sellers are small sellers on the e-commerce platform.

5.4 Countermeasures

The large sellers, mostly in Category 2, generally process well-known brands, numerous products and high traffic, and achieve high results on the indices in the category of sales. The common problem among these sellers lays in the poor traffic quality and low conversion ratios. To solve the problems, the e-commerce platform should encourage them to optimize the product structure (e.g. eliminating slow movers), and improve store decoration.

The medium sellers, mostly in Categories 0 and 1, have good traffic quality or a long SOYN, i.e. the sellers have been working hard for better sales. The e-commerce platform should guide these sellers to expand the scale of their stores, and provide supports to those with a long SOYN, aiming to optimize their store structure and quality.

The small sellers, mostly in Category 3, mostly operate a small store with very low SPN, DPN and VPN. Their high conversion ratios cannot cover up the serious operating problems. The e-commerce platform must acknowledge the problems of these sellers: small scale, i.e. low attractiveness and irrational product structure, and help them to solve each of the three problems. For example, the platform could encourage these sellers to refine store decoration, relax the promotion policy on them, and assign high weights to their products.

6. CONCLUSIONS

This paper puts forward a novel hierarchical model for e-commerce sellers based on the PCA and the SOM, uses the model to classify 140,000 sellers on an e-commerce platform, and designs operating strategies for the sellers in each category. The research results provide a good reference for the operation of e-commerce sellers.

The further research will focus on the following aspects: the hierarchical model will be improved with data involving more dimensions; the product range will be introduced to the classification problem; the SOM algorithm will be improved in the light of the actual demand of e-commerce sellers and the state of online shopping.

REFERENCES

[1] Zhang, J., Zhang, C., Yu, H. (2018). Research on e-commerce intelligent service based on data mining. In MATEC Web of Conferences, 173: 03012. https://doi.org/10.1051/matecconf/201817303012

[2] Najafi, I. (2019). Assessment and modeling of decision-making process for e-commerce trust based on machine learning algorithms. Fundamental Research in Electrical Engineering, 969-986. https://doi.org/10.1007/978-981-10-8672-4_74

[3] Ju, C., Wang, J., Zhou, G. (2019). The commodity recommendation method for online shopping based on data mining. Multimedia Tools and Applications, 78(21): 30097-30110. https://doi.org/10.1007/s11042-018-6980-7

[4] Shah, T.H., Naveed, N., Rauf, Z. (2018). A methodology for brand name hierarchical clustering based on social media data. Journal of Applied and Emerging Sciences, 8(1): 10-23. http://dx.doi.org/10.36785/jaes.v8i1.238

[5] Alsenan, S., Zemirli, N. (2016). PERSO-retailer: Modeling the retailer's business data: Toward recommender system of retailers' marketing plan for personalized CMS. 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, pp. 106-111. http://dx.doi.org/10.1109/CCAA.2016.7813699

[6] Mu, M., Liu, C. (2018). Research on the construction of o2o e-commerce and express industry collaborative development model in big data environment. In Proceedings of the 2018 International Conference on Internet and e-Business, 29-33. https://doi.org/10.1145/3230348.3230424

[7] Castro-Lopez, A., Alonso, J.M. (2019). Modeling human perceptions in e-commerce applications: A case study on business-to-consumers websites in the textile and fashion sector. Applying Fuzzy Logic for the Digital Economy and Society, 115-134. https://doi.org/10.1007/978-3-030-03368-2_6

[8] Zaim, H., Ramdani, M., Haddi, A. (2018). A model of e-commerce self-assessment system based on e-customer behavior. Smart Application and Data Analysis for Smart Cities (SADASC'18). http://dx.doi.org/10.2139/ssrn.3179244

[9] Sulova, S. (2018). Integration of structured and unstructured data in the analysis of e-commerce customers. International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & Mining Ecology Management, 18: 499-505. http://dx.doi.org/10.5593/sgem2018/2.1/S07.063

[10] Hsieh, P.H. (2019). A study of models for forecasting e-commerce sales during a price war in the medical product industry. International Conference on Human-Computer Interaction, 3-21. https://doi.org/10.1007/978-3-030-22335-9_1

[11] García, M.D.M.R., García-Nieto, J., Aldana-Montes, J.F. (2016). An ontology-based data integration approach for web analytics in e-commerce. Expert Systems with Applications, 63: 20-34. https://doi.org/10.1016/j.eswa.2016.06.034

[12] Lee, H.C., Rim, H.C., Lee, D.G. (2019). Learning to rank products based on online product reviews using a hierarchical deep neural network. Electronic Commerce Research and Applications, 36: 100874. https://doi.org/10.1016/j.elerap.2019.100874

[13] Goswami, A., Mohapatra, P., Zhai, C. (2019). Quantifying and visualizing the demand and supply gap from e-commerce search data using topic models. In Companion Proceedings of the 2019 World Wide Web Conference, 348-353. https://doi.org/10.1145/3308560.3316605

[14] Huang, H.J., Yang, J., Zheng, B. (2019). Demand effects of product similarity network in e-commerce platform. Electronic Commerce Research, 1-31. https://doi.org/10.1007/s10660-019-09352-9

[15] Yoo, B., Jang, M. (2019). A bibliographic survey of business models, service relationships, and technology in electronic commerce. Electronic Commerce Research and Applications, 33: 100818. https://doi.org/10.1016/j.elerap.2018.11.005

[16] Hamidi, H., Moradi, S. (2017). Analysis of consideration of security parameters by vendors on trust and customer satisfaction in e-commerce. Journal of Global Information Management (JGIM), 25(4): 32-45. https://doi.org/10.4018/JGIM.2017100103

[17] Wakil, K., Alyari, F., Ghasvari, M., Lesani, Z., Rajabion, L. (2019). A new model for assessing the role of customer behavior history, product classification, and prices on the success of the recommender systems in e-commerce. Kybernetes. https://doi.org/10.1108/K-03-2019-0199

[18] Zhao, J., Wang, L., Li, D.A., Li, Y., Yang, B., Zhu, B., Bai, R. (2018). Mining shopping data with passive tags via velocity analysis. EURASIP Journal on Wireless Communications and Networking, 2018(1): 1-13. https://doi.org/10.1186/s13638-018-1033-5

[19] Sohaib, O., Naderpour, M., Hussain, W., Martinez, L. (2019). Cloud computing model selection for e-commerce enterprises using a new 2-tuple fuzzy linguistic decision-making method. Computers & Industrial Engineering, 132: 47-58.

https://doi.org/10.1016/j.cie.2019.04.020

[20] Liu, S. (2016). The innovation study of e-business mode based on big database environment. In 2016 2nd International Conference on Education Technology, Management and Humanities Science, 416-419. https://doi.org/10.2991/etmhs-16.2016.92

[21] Behl, A., Dutta, P., Lessmann, S., Dwivedi, Y.K., Kar, S. (2019). A conceptual framework for the adoption of big data analytics by e-commerce startups: a case-based approach. Information Systems and e-Business Management, 17(2): 285-318. https://doi.org/10.1007/s10257-019-00452-5

[22] Leng, K., Jing, L., Lin, I.C., Chang, S.H., Lam, A. (2019). Research on mining collaborative behaviour patterns of dynamic supply chain network from the perspective of big data. Neural Computing and Applications, 31(1): 113-121. https://doi.org/10.1007/s00521-018-3666-z

[23] Jiang, H., Sabharwal, A., Henderson, A., Hu, D., Hong, L. (2019). Understanding the role of style in e-commerce shopping. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3112-3120. https://doi.org/10.1145/3292500.3330760

[24] Kaushik, V., Khare, A., Boardman, R., Cano, M.B. (2020). Why do online retailers succeed? The identification and prioritization of success factors for Indian fashion retailers. Electronic Commerce Research and Applications, 39: 100906. https://doi.org/10.1016/j.elerap.2019.100906

[25] Leung, W., Shi, S., Chow, W. (2019), Impacts of user interactions on trust development in C2C social commerce: The central role of reciprocity. Internet Research. https://doi.org/10.1108/INTR-09-2018-0413

[26] Sharma, H., Aggarwal, A. (2019), Finding determinants of e-commerce success: a PLS-SEM approach. Journal of Advances in Management Research, 16(4): 453-471. https://doi.org/10.1108/JAMR-08-2018-0074