Research on the Openness of Regions Along the Belt and Road Based on Machine Learning

Taking Liaoning province as an example

Nan Wang[1,a*], Feng Li[2,b]

[1]School of Economics Management and Law, Shenyang Institute of Engineering, Shenyang, China.
[2]Guanghua School of Management, Peking University, Beijing, China.

[a*]wangnan@sie.edu.cn, [b]lifeng6559@stu.pku.edu.cn
*Corresponding Author: Nan Wang wangnan@sie.edu.cn

Abstract. The "Belt and Road" project has not only brought Liaoning province a major opportunity and challenge to revitalize its old industrial base but also a brand-new pilot free trade zone. In order to measure the degree of openness of cities in Liaoning province since the establishment of the pilot free trade zone seven years ago, this paper proposes a machine learning-based research method on the degree of openness. The index system of the degree of openness of regions along the "Belt and Road" is constructed from the three dimensions of trade, finance, and investment. The data of 12 cities in Liaoning province from 2016 to 2021 are clustered based on the K-means method, and the clustering results are used as a learning guide to train SVM for classification. The data of 2 other cities in the province were classified using the obtained model, and the openness classification of the 2 cities was obtained. The new combined model can significantly improve the quality of clustering and can be used to study the degree of openness of countries and regions along the "Belt and Road".

Keywords: openness; machine learning; K-means; SVM; the Belt and Road

1 Introduction

The Belt and Road Initiative is an important element of China's opening up to the outside world and is a major decision made in light of changes in the economic situation and development at home and abroad. On September 1, 2016, the much-anticipated third batch of China's Pilot Free Trade Zone finally surfaced. The Party Central Committee and the State Council decided to set up new pilot free trade zones in seven provinces, including Liaoning. After more than 7 years, it has created good economic benefits for the cities of Liaoning and injected new blood into the local enterprises.

Scholars at home and abroad have conducted some studies on the measurement of the degree of openness to the outside world, for example, Low et al. (1998) [1] and Cameron et al. (1999) [2] suggest that the openness of an economy can also be measured

B. K. Kandel et al. (eds.), *Proceedings of the 2023 3rd International Conference on Business Administration and Data Science (BADS 2023)*, Atlantis Highlights in Computer Sciences 19,
https://doi.org/10.2991/978-94-6463-326-9_17

in terms of trade and investment. Kaplan et al. (2006) [3] argue that a country's openness is analyzed in terms of imports and exports, tariff rates, and exchange rate indices. Domestically, Wang et al. (2023) [4], Zhao (2014) [5], Liu (2016) [6], and Zhang (2019) [7] measure the degree of opening to the outside world in Shouguang, Shandong, coastal areas of Liaoning, and Xinjiang using foreign trade and foreign investment data, respectively. Zhao (2020) [8], Cui (2023) [9], and Lian (2023) [10] measure the investment risk of cities on the Belt and Road Economic Belt using machine learning, respectively.

Researchers have made detailed studies and discussions on the degree of openness from different perspectives. However, most of the existing research objects stay in the study of a certain country or province. The methods are too simplistic, the metrics are homogenized, the dimensions included are not comprehensive, and it is difficult to be binding and convinced in the practical application. Therefore, this paper tries to propose a new method to measure the degree of openness of cities to the outside world.

2 Methodology

2.1 K-means

The k-mean clustering algorithm (K-means) is a cluster analysis algorithm for iterative solutions and is an unsupervised learning method. Its steps are listed as follows.

(1) Select the initialized k samples as initial clustering centers $a = a_1, a_2, \ldots, a_k$.

(2) For each sample x_i calculate its distance to the k clustering centers and classify it into the class corresponding to the clustering center with the smallest distance.

(3) For each category a_j, recalculate its clustering center. i.e., the center of mass of all samples belonging to that category using equation (1). Where a_j is the centroid of the jth class, c_i is the number of parameters. x is the coordinates of the contained point.

$$a_j = \frac{1}{|C_i|}\sum_{x \in C_i} x \tag{1}$$

(4) Repeat steps 2 and 3. The termination condition can be no (or a minimum number of) objects are reassigned to different clusters, no (or a minimum number of) cluster centers change again, or the sum of error squares is locally minimized.

In this paper, we choose the common elbow method to determine the value of K according to the sum of squared errors. Then use the Gap statistic method to evaluate.

2.2 SVM

Support Vector Machine (SVM) is a classification algorithm developed in the computer community in the 1990s. It has been proven to be effective in many problems and is considered to be one of the most widely adapted algorithms. SVM is a generalized linear classifier for binary classification of data by supervised learning. Its decision boundary is the maximum-margin hyperplane solved for the learning samples.

Non-linear classification can be performed by the kernel method, which is one of the common kernel learning methods. It needs the support of the kernel function. In this

paper, it is a nonlinear problem classification and the Gaussian function is chosen as kernel function. It is suitable for linear indivisible cases with more parameters. Its formula is equation (2) where x_1 and x_2 are the feature vectors of the sample points. σ is the bandwidth of the Gaussian kernel function. k represents the output value.

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|}{2\sigma^2}\right) \tag{2}$$

In this study, k-means is used to perform preliminary clustering analysis, and then the clustering results are used as a learning guide to train the SVM, and the acquired model is used to retest and reclassify the whole dataset, so as to make the classification results reach the global optimization.

3 Data

3.1 Data selection

Openness to the outside world refers to the degree of integration of a country or region into the international economy or its dependence on the international economy. In order to comprehensively and rationally assess the degree of openness of cities in Liaoning Province after their integration into the Belt and Road, it is measured from a number of perspectives. The selected indicators of the degree of openness of the regions along the Belt and Road are shown in Table 1 below.

Table 1. Openness indicator

Tier 1 indicators	Tier 2 indicators	Interpretation of indicators
Trad Openness	Total Imports	Total regional imports
	Total Exports	Total regional exports
	Import Dependency	Import Value/Regional GDP
	Export Dependency	Exports/Regional GDP
Financial Openness	Offshore Loans	Total foreign currency loans
	Foreign currency deposits	Total foreign currency deposits
Investment openness	Foreign Direct Investment	Total Foreign Investment

The data are mainly from the Liaoning Statistical Yearbook 2016-2021. 12 out of 14 cities are randomly selected for training, other 2 cities are for generalization.

3.2 Data preprocessing

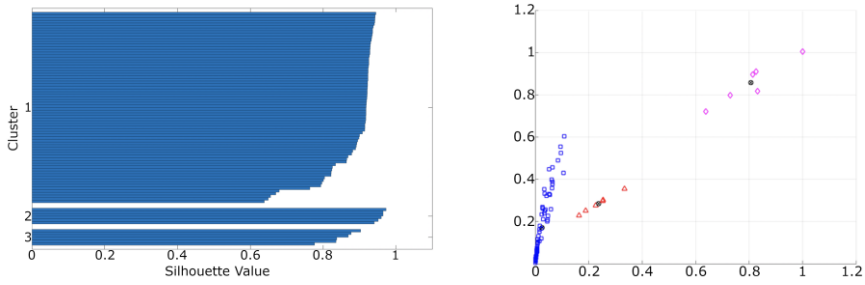
Both K-means and SVM require the preprocessing of indicator data. Given that there is a certain degree of variability in the nature of the various types of indicators and characteristics such as units. In order to ensure the scientific nature of the results of data analysis and facilitate data comparison and common processing, it is necessary to implement the dimensionless processing of various indicators to avoid dimensions with

large mean and variance that will have a decisive impact on the clustering and classification of data. In this paper, the z-score normalization method is chosen to normalize the data, using the following formula to map the data within the range of 0 to 1.

4 Results and discussions

4.1 Cluster analysis using the k-means method

The k-means method requires us to determine the optimal number of clusters. In this paper, we first choose the elbow method to determine the optimal number of clusters. By varying the number of k, we find that when k=3, and then continue to increase k, the decrease in SSE decreases and levels off. The number of clusters can be initially determined to be 3. Then use the Gap statistic method to verify whether the selection of k is reasonable. Verify the average profile coefficient for different values of k. It can be shown in Fig. 1(a) and (b). The closer to 1, the better the clustering effect. It can be seen that most of the points in the 3 clusters have large contour values (greater than 0.8) when k = 3, indicating that these points can be well distinguished from neighboring clusters. It is further determined that the cluster selected for clustering is 3.



(a) (b)

Fig. 1. K-means clustering results.

4.2 Supervised SVM classification learning using k-means results

According to the conclusion of k-means, k=3 was chosen. the degree of openness of the 12 cities in Liaoning was defined into 3 categories, excellent, good, and poor. The results obtained using the k-means method guided the SVM to be trained, and the whole data set was retested and reclassified using the acquired model, thus making the clustering results, globally optimal. The data points of the selected 12 cities were classified. For a more intuitive representation, the simulation results are mapped to a two-dimensional space, and the simulation data map is obtained as shown in Fig. 2(a).

4.3 Evaluation of the combinatorial classification models

Generally speaking, the higher the total amount of imports and exports, the more economic and trade exchanges the region has with the outside world. Researchers often

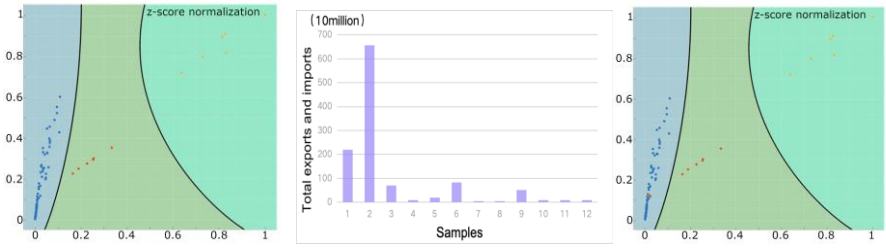
use the sum of the total amount of imports and exports to simply indicate the degree of the region's openness to the outside world. In order to verify the effect of the SVM combined classification model, the sum of the total amount of imports and exports of 12 cities in 2021 is selected for simple classification. The classification result can be seen as binary classification through Fig. 2(b). The combined classification model established in this paper is a multiclassification model, and in order to compare the classification effect, the multiclassification problem is simplified to n binary classification problems. The accuracy rate of the model is obtained by equation (3). Where TiPi is the positive prediction of the true classification i as classification i, and FjPi is the incorrect prediction of the true classification j as classification i.

$$accuarcy = \frac{\sum_{i=1}^{n} T_i P_i}{\sum_{i=1}^{n} T_i P_i + F_j P_i} \tag{3}$$

The accuracy rate is 91.67%, which indicates that the classification effect is good. But at the same time, it can be found that the use of simple variables can only realize coarse classification and the use of the combination of k-means and SVM in this paper can uncover the classification details and realize fine classification.

4.4 Empirical Extension of Combinatorial Classification Models

The data from the remaining 2 cities were brought into a classification model built on SVM to categorize the openness of these 2 cities. The results of the categorization are shown in Fig. 2(c). It can be intuitively found through the image that the red dots representing the two cities both fall in the blue region on the far left of the image. This region represents the category of poor degree of openness to the outside world.



(a) (b) (c)

Fig. 2. Training and generalization of the combinatorial classification model.

5 Conclusions

This study constructs the index system of the degree of openness of regions along the "Belt and Road" from three dimensions: trade, finance, and investment. The data of 12 cities in Liaoning Province from 2016 to 2021 are selected for clustering based on the K-means method. The elbow method is first chosen to determine the optimal number of clusters. By varying the number of k, we find that when k=3, continues to increase k, the SSE decreases and tends to level off. The Gap statistic method is then used to

verify the value of k is reasonable. Again, we find that the average profile coefficient converges to 1 when k is chosen as 3. The number of clusters can be determined as 3.

The clustering results are used as a learning guide to train SVM for classification, and the classification effect is further improved. Using the obtained model to classify the data of 2 other cities, the openness classification was obtained. The combined model uses machine learning methods to significantly improve the quality of clustering, and can be used to generalize the study of the degree of openness of countries and regions along the "Belt and Road". The degree of openness of countries and regions along the "Belt and Road" is an issue worthy of in-depth exploration, which is of great practical significance for the cooperation between China and the countries along the route.

Acknowledgment

This research was funded by Social Science Planning Fund Subjects of Liaoning Province, under grant number L16BJL009.

References

1. LOW P, OLARREAGA M, SUAREZ J. Does globalization cause a higher concentration of international trade and investment flows? [Z]. WTO staff working papers, 1998.
2. CAMERON G, PROUDMAN J, REDDING S, et al. Openness and its association with productivity growth in UK manufacturing industry [J]. Bank of England working papers, 1999. Bank of England working papers,1999.
3. KAPLAN M, ASLAN A. Quantifying international openness in Turkey: 1965-1995 [J]. Dogus University Journal, 2006, 7(1):965-996.
4. Wang Xiaofang, Yang Yongjian. Research on the Measurement and Influencing Factors of the Degree of Openness of Countries Along the "Belt and Road"[J]. Asia-Pacific Economy,2023(04):15-23. DOI: 10.16407/j.cnki.1000-6052.2023.04.013.
5. Z. Zhao, Y. Liu, Y. Fan, S. Liu, et al. Social risk in countries along the "Belt and Road" - an empirical analysis based on BP neural network[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Science Edition), 2023, 36(04): 105-114. DOI: 10.13766/j.bhsk.1008-2204.2021.0707.
6. Liu J. The relationship between the degree of opening to the outside world and economic growth in Xinjiang under the Silk Road Economic Belt[J]. Marketing World,2019(32):91-92.
7. Zhang Zhenjia. Measurement of the degree of opening to the outside world in the Liaoning coastal economic zone under the new normal state--Based on the comparative perspective with the Shenyang economic zone[J]. Science and Industry,2016,16(01):39-43.
8. Zhao Guangyan. An empirical study on the degree of opening up of county areas to the outside world--Taking Shouguang of Shandong as an example[J]. China Market,2014(08):29-31.
9. Cui Zhenguan, Jin Yide. Application of machine learning in early warning of China's overseas investment efficiency--evidence from countries along the "Belt and Road"[J]. Finance Theory and Practice,2023(05):13-25.
10. Lian Qiuyi. Analysis of Liaoning's all-around opening up to the outside world[J]. Liaoning Economy,2020(12):64-65. DOI: 10.14041/j.cnki.1003-4617.2020.12.027