LS-SVM Key Nodes Identification from Simple Indices and Complex Indices in Aviation Network

Haiqing Huang[1, a], Xusheng Gan[2, b, *], Jingjuan Sun[3, c], Shuangyan Chen[2, d]

[1]XiJing College, Xi'an, Shaanxi, 710123, China

[2]Air Traffic Control and Navigation College, Air Force Engineering University, Xi'an, Shaanxi, 710051, China

[a]ganxusheng123@163.com, [b] gxsh15934896556@qq.com, [c]sunjingj11@163.com, [d]chenshy21@126.com

Keywords: Aviation Network, Least Square, Support Vector Machine, Key Nodes Identification, Comprehensive importance

Abstract: Identifying key nodes in aviation network is significant in practice. A key nodes identification method based on Least Square Support Vector Machine (LS-SVM) is proposed for aviation network. Firstly, basic prinple of LS-SVM is introduced for evaluation modeling, then the comprehensive importance of nodes is calculated through complex indices, finally LS-SVM model is established the mapping relationship between the simple indices (as the input) and the comprehensive importance (as the output). The simulation shows that the proposed method is effective and feasible for identifying the key nodes in aviation network.

1. Introduction

The construction of the aviation network is an important reflection of the level of social and economic development, and it can also reflect the national strength from the side. In peacetime, the aviation network can promote economic development. In wartime, the aviation network can directly affect the progress of the war.

For example, Corley et al evaluated the importance of deleting nodes by studying the shortest path after deleting nodes [1]; Daniel et al compared the indices such as closeness centrality, betweenness centrality, and degree centrality, and introduced game theory to evaluate the importance of network node [2]; He et al studied the important ranking problem of complex network nodes according to node degree and efficiency [3]; Y. J. Tan et al proposed a node shrinkage method for evaluating node importance based on the definition of cohesion degree. The node with the largest degree of cohesion after the shrinkage is regarded as the most important node [4].

These studies are mainly applicable to unweighted networks, but basically do not consider the impact of airline traffic flow in aviation networks. Second, the method is too single, and generally only consider a certain property of the node. For example, F. H. Xie et al proposed a method for ranking the importance of complex network nodes based on weighted clustering coefficients according to the shortcomings of the key node identification algorithm of complex networks [5]; Chen et al measured its own shortcomings according to each centrality measure and used several different centrality measures for multi-attributes, and the analytic hierarchy process is used to aggregate the multi-attributes to obtain the impact evaluation value of each node [6]. On the basis of the local characteristics of nodes in the network, J. W. Wang et al proposed a method for measuring the importance of nodes based on the degree of neighboring nodes [7]. These studies are simple and efficient. The disadvantage is that the factors affecting the importance of airport nodes are too complicated, and it is often difficult to obtain accurate conclusions only considering individual properties. In the paper, we intend to use Closeness Centrality (*CC*), Betweenness Centrality (*BC*), Link Density (*LD*) and Network Efficiency (*NE*) to comprehensively evaluate the importance of nodes. However, in actual calculations, these indices usually involve high-complexity operations

  DOI: 10.25236/icited.2021.007

such as the shortest path, which causes the evaluation process to take too long and affects the identification of key nodes.

According to the above analysis, Least Square Support Vector Machine (LS-SVM) is introduced to establish the identification model to realize the accurate identification of the key nodes of the aviation network for the normal operation of aviation network in peacetime and defense and repair in wartime.

2. Least square support vector machine

SVM was originally used to solve the problem of pattern recognition, and then extended into regression estimation. In the paper, we can use the LS-SVM where the inequality constraint of the original problem is transformed into the equality constraint for solution. Comparing to the standard SVM [8][9][10], LS-SVM has a simple calculation and good prediction without considering the ε-insensitive loss function.

Suppose that the training set is $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ where $x_i \in R^d$, $y \in R$, $F = \{f \mid f : R^d \to R\}$. For the nonlinear regression problems, first the training sample x in the original input space R^d is mapped into a high dimensional feature space by a nonlinear function $\phi(\cdot)$, and then in the space the linear regression is executed to obtain the effect of nonlinear regression in the original space. Regression estimation function can be given as

$$f(x) = \langle w \cdot \phi(x) \rangle + b \tag{1}$$

in light of the principle of the structural risk minimization, the regression problem can be transformed into the following constraint optimization problem:

$$\min \ \frac{1}{2}\|w\|^2 + C\frac{1}{2}\sum_{i=1}^{n}\xi_i^2 \tag{2}$$

$$s.t. \ y_i = w^T\phi(x_i) + b + \xi_i \tag{3}$$

where C is the penalty parameter; ξ_i is the non-correlated random error. Lagrange method can be used to solve the constraint optimization problem. Then we can get the following equation:

$$\begin{bmatrix} 0 & 1_n^T \\ 1_n & \Omega + C^{-1}I_n \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{4}$$

where the Lagrange multiplier $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T$, $\Omega_{i,j} = \phi(x_i)^T\phi(x_j)$, $y = [y_1, y_2 \cdots, y_n]^T$, $1_n = [1, 1, \cdots, 1]$, $i, j = 1, 2, \cdots, n$, According to Mercer conditions, the kernel function $K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$, then α_i and b can be solved by Least Square method. The regression function can be given as

$$f(x) = \sum_{i=1}^{n}\alpha_i K(x, x_i) + b \tag{5}$$

It can be seen that, in the estimation of the regression function, we only needs to calculate $K(x_i, x_j)$ without calculating $\phi(\cdot)$. Furthermore, the regression function with different kernel function has a different estimation. the kernel function can be selected as:

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \tag{6}$$

where σ is the kernel width. Thus, LS-SVM only need to determine two parameters: C and σ. The optimal parameter combination (C, σ) can be determined by 10-fold cross validation method.

3. Evaluation process of node importance using LS-SVM

3.1 Simple indices

Index analysis is the focus of aviation network research. When the selected index is small, the accuracy of the evaluation will not meet the requirements. When the selected index is more comprehensive, the calculation time complexity will be very high. The simple index value is the training knowledge database. In this paper, node degree value, point strength, and K-shell value are selected as simple indices.

Table 1 Information about three simple indices

Index	Advantage	Disadvantage	Time complexity
Node degree	Reflect the ability of a node to influence neighboring nodes	Only consider node local information	O(N)
Point strength	Reflect airport route flow	Only consider node local information	O(N)
K-shell	Reflect the global information of the node	Poor accuracy	O(N)

Three indices in Table 1 are more representative and all have low time complexity. The following also normalizes simple indices:

For the degree value of the node v_i:

$$D_i = \frac{D(v_i) - \min D(v)}{\max D(v) - \min D(v)} \tag{7}$$

Do the same process for point strength S_i of the node v_i and the K-shell value Ks_i. In summary, n nodes can form a simple index value matrix

$$X = \begin{bmatrix} D_1 & S_1 & Ks_1 \\ D_2 & S_2 & Ks_1 \\ \vdots & \vdots & \vdots \\ D_n & S_n & Ks_n \end{bmatrix} \tag{8}$$

where X denotes the matrix of degree value, point strength and K-shell value for n nodes.

3.2 Complex indices

For social network analysis methods, Closeness Centrality (*CC*), Betweenness Centrality (*BC*), Link Density (*LD*) and Network Efficiency (*NE*) can be selected as evaluation indices.

Table 2 Information about four complex indices

Index	Advantage	Disadvantage	Time complexity
CC	Reflect node location information	Not suitable for large networks	$O(N^3)$
BC	Reflect node load capacity	Low accuracy/not suitable for large networks	$O(N^3)$
NE	Reflect the ability of network information transmission	Not suitable for large networks	$O(N^3)$
LD	Reflect the heterogeneity and load capacity of the network	Not suitable for large networks	O(N)

3.3 Establishment of LS-SVM evaluation model

Through the introduction on simple indices, complex indices and LS-SVM, the evaluation model can be established. The simple indices such as node degree value, point strength, and K-shell value are

taken as the inputs of LS-SVM, and the comprehensive importance calculated by the complex indices such as Closeness Centrality (*CC*), *B*etweenness Centrality (*BC*), Link Density (*LD*) and Network Efficiency (*NE*) is taken as the outputs of LS-SVM. To calculate the comprehensive importance, we need obtain the weight of each complex index by Analytic Hierarchy Process (AHP) [11]: W_{BC}=0.5789; W_{NE} =0.2055; W_{CC}=0.1592; W_{LD}= 0.0565, and then get the comprehensive importance $Y=[Y_1, Y_2,\cdots,Y_n]^T$ of the random nodes by $Y_i=0.5789BC_i+ 0.2055NE_i +0.1592CC_i+ 0.0565LD_i$. For the new nodes in the aviation network except for the training samples, only the simple indices X_t of the new nodes need to be calculated and input the LS-SVM model built to evaluate the comprehensive importance Y_t of the new nodes, completing the node sorting.

4. Simulation

In experiment, U.S. aviation network can be selected. The experiment data [12] include 332 airport nodes, 2126 edges (direct flight routes) and the weights of the edges. 20, 40, 60 and 80 nodes are randomly selected as training samples.

According to the proposed method, the parameter C and σ are determined by cross-validation, as shown in Fig. 1. In the process of node importance evaluation, the longest time-consuming is the establishment of knowledge database (using complex indices to evaluate the importance of nodes). If most of the nodes in the network are selected as training samples, the evaluation method can lose the advantage of low time complexity, becoming meaningless. Therefore, it is tested here whether needs only a small number of nodes LS-SVM model can accurately evaluate the importance of nodes. 20, 40, 60, and 80 nodes are randomly selected as training samples, and then compare the test results with the original importance values, as shown in Fig. 2.

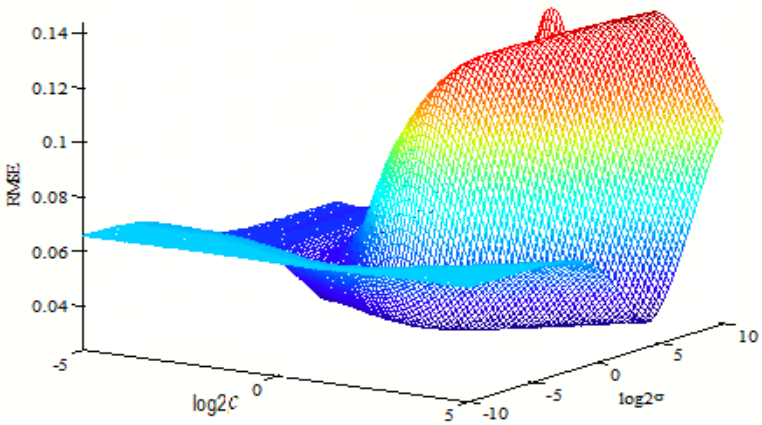
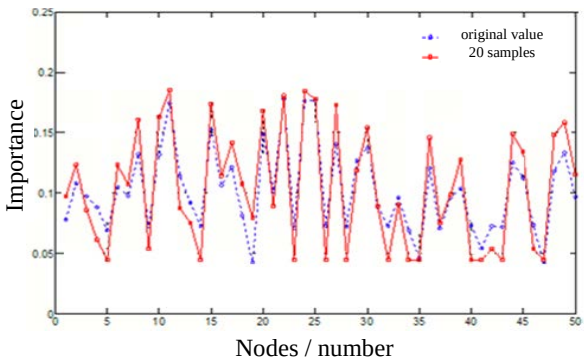
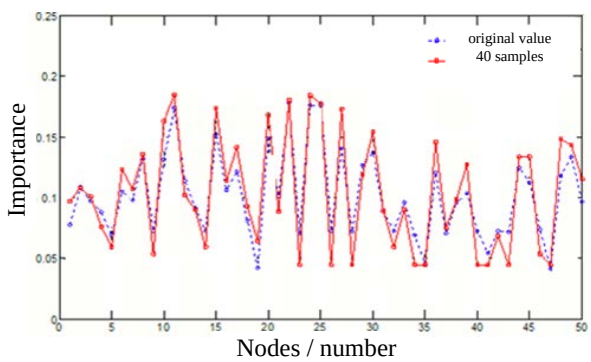


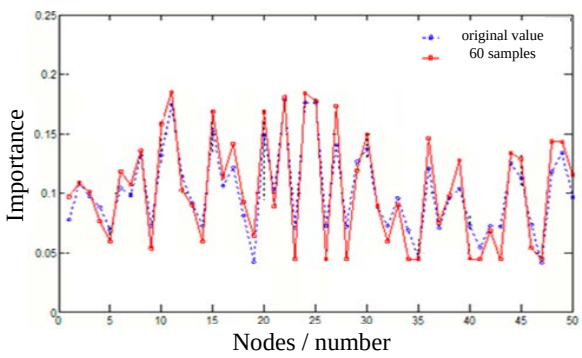
Fig. 1 Optimization process from parameter C and σ



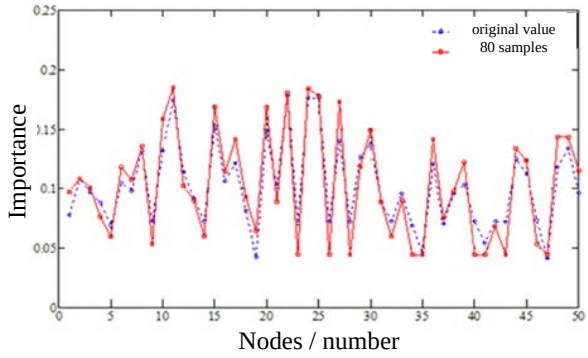
(a) 20 training samples



(b) 40 training samples



(c) 60 training samples



(d) 80 training samples

Fig. 2 Comparison of test results and actual values

It can be seen from the figure that, when 20 training samples are selected, the fitting effect between the test results and the original value is poor. When 40 training samples are selected, the fitting effect is significantly improved. Therefore, in American aviation network, only the complex index values of 40 nodes need to be calculated, which can greatly reduce the original calculation amount and improve the efficiency of identifying key nodes.

5. Conclusions

For key nodes identification in aviation network, the comprehensive importance of nodes is calculated based on Closeness Centrality (*CC*), *B*etweenness Centrality (*BC*), Link Density (*LD*) and Network Efficiency (*NE*), which overcomes the specific aviation network factors that affect the importance of nodes such as route flow and airport location that have not been considered in the field of complex network research. LS-SVM is applied to obtain the mapping relationship between simple indices and comprehensive importance. According to the evaluation result of each node, the nodes in aviation network can be sorting, identifying the key nodes. it shows that the proposed method is effective and feasible for key nodes identification.

References

[1] H. W. Corley, D. Y. Sha. Most vital links and nodes in weighted networks. Operation Research Letters, 1(4), (1982), 157-161

[2] D. Gomezb, G. A. Enrique. Centrality and power in social networks: a game theoretic approach. Mathematical Social Sciences, 46(1), (2003), 27-54

[3] N. He, D. Y. Li, W. Y. Gan, et al. Mining vital nodes in complex networks. Computer Science, 34(12), (2007), 1-5

[4] Y. J. Tan, J. Wu, H. Z. Deng. Evaluation method node importance based on node contraction in complex networks. System Engineering Theory and Practice, 26(11), (2006), 79-83

[5] F. H. Xie, D. W. Zhang, D. Huang, et al. Keywords extraction based on weighted complex network. Journal of System Science and Mathematical Science, 30(11), (2010), 1592-1596

[6] D. B. Chen, L. Y. Lv, M. S. Shang, et al. Identifying influential nodes in complex networks. Physica A, 391, (2012), 1777-1787.

[7] J. W. Wang, L. L. Rong, T. Z. Guo. A new measure method of network node importance based on local characteristics. Journal of Dalian University of Technology, 50(5), (2010), 822-826.

[8] V. Vapnik. Statistical Learning Theory. New York: Wiley, (1998)

[9] J. A. K. Suykens, V. T. Gestel, J. De Brabanter, et al. Least squares support vector machines. World Scientific, (2002)

[10] L. C. Jiao, L. F. Bo, L. Wang. Fast sparse approximation for least squares support vector machine. IEEE Transactions on Neural Networks, 18 (3), (2007), 685-697

[11] G. H. Yu, Y. J. Zhang, H. J. Huang, et al. A new method for consistency test of judgment matrix in AHP. Mathematics in Practice and Theory, 47(22), (2017), 189-198

[12] Pajek's list of network data sources: http://vlado.fmf.uni-lj.si/pub/networks/data