Fuzzy GMDH Network Algorithm Model for Value Chain Analysis of Sports Industry Development

Hao Huaixia
School of Physical Education
Hainan University
Haikou, 570100, Hainan, China
{fztlsx545099@163.com}

ABSTRACT: *This article studies the analysis of the value chain of sports industry development using the fuzzy GMDH network algorithm model. The sports industry is an important component of modern service industry, and its development is of great significance for promoting economic growth, improving people's health level, and promoting social progress. Value chain analysis is an important means to understand the development status of industries and optimize resource allocation. The fuzzy GMDH network algorithm model is an algorithm model based on data mining and machine learning technology, which can extract valuable information by processing and analyzing a large amount of data.*

Keywords: Value Chain, Algorithm, Sports Industry Development, Discussion

Received: 21 May 2023, Revised 9 July 2023, Accepted 25 July 2023

DOI: 10.6025/jet/2023/14/4/105-112

1. Introduction

The sports industry value chain generally contains three parts: the value chain, the value chain enterprise relationship and the value production process [1]. The value chain is the performance of the number of enterprises on the value chain, which reflects the matching degree and the situation of the value chain enterprise and the expected value. In the value chain of enterprises, the relationship between each other is cooperation and competition coexist [2]. The relationship between composition of value chain and value chain relationship of enterprises is integrated into an industrial development network. Value production process is the embodiment of enterprise value, including the distribution of production value of resource utilization, which shows the relationship between enterprise input and output of value. Value chain enterprise relationship transforms organic industry network into organic value net [3]. The optimization of industrial value chain is generally carried out in the different sub catalogues of three dimensions, such as value chain formation, value chain enterprise relationship and value production process [4]. Each industry's understanding and application of big data need to be different and massive data need more visualization means to effectively use. Data mining technology mainly involves the main branches of mathematical statistics, artificial intelligence model, and database technology and visualization research. The research is extensive, and it is a new discipline with multi discipline, integration and interaction. At present, it is necessary to find the potential useful information from a large amount of data with noise, and find the direction of value chain optimization by scientific methods, which will inevitably choose to use data mining technology.

2. State of the Art

Most of the data mining technologies put forward higher requirements to the users, and the mathematical model will cause the quality instability because of human factors [5]. Group method of data handling (GMDH) uses cross, mutation and selection operations, and automatically establishes and confirms the model structure in the evolution principle [6]. GMDH algorithm model is a self-organizing data mining method, which can achieve the best balance between memory function and generalization function. It is the most widely used and universal modeling algorithm [7]. The GMDH algorithm has four main features. Firstly, the self-organization is strong, and there is no need to worry about the multiple linear and correlation of the data variables. The input variables are automatically screened with the best correlation, and the model structure of production is relatively simple (Parsaie A, et al. 2015) [8]. The second is the advantage of global optimization. Under the same conditions, the model generated by the system is the optimal model, and it can solve the small sample problem well [9]. The third is the objectivity of model selection. With the support of computer aided technology, GMDH network only needs input data, operation sequence and calculation standard. Other structural and variable parameters can be directly controlled by computer, and the model selection is more objective and accurate [10]. The fourth is the inductive feature of dynamic and static integration.

The value of industry can be viewed from two aspects: intrinsic value and market value. The intrinsic value is the net present value of cash flow realized by the industry in the foreseeable future, which reflects the profitability of the industry. In the development of socialist market economy, the free competition market is advocated, so the market value is the unified embodiment of the intrinsic value of the industry. The value of the industry ultimately depends on the future profitability of the portfolio of assets, which is the size of the future cash flow of the industry. Therefore, research on the value chain of sports industry was carried out in this paper. Mainly starting from the assessment of the value of the sports industry, the establishment of GMDH model was carried out by grouping method of data processing in order to help the sports industry to optimize and improve the value chain, promote the quality of sports resources integration, promote the rapid development of the sports industry and lasting benefits in the development of sports industry.

3. Methodology

3.1. GMDH Algorithm

Several publications related to the users' surveys are found in simulation literature. A dated survey carried out by Kleine 1, 2, has examined users' views of eleven discrete simulation languages [4]. The results of this survey showed that it was difficult to interpret the results mainly because a limited number of respondents were proficient in more than one language. In addition, the expertise of some respondents was difficult to specify.

GMDH Algorithm	
Parameter GMDH algorithm	Nonparametric GMDH algorithm
Combinatorial COMBI	Objective Computer Clusterization OCC
Multilayered Iteratiomal MIA	
Harmonial	Analogues Complexing AC
Objective System Analysis OSA	

Table 1. Classification of GMDH Algorithms

GMDH algorithm is a heuristic self-organizing mathematical model proposed based on multilayer neural network theory in the late 60s of last century. It is a combination of data processing methods. It is based on polynomial and uses continuous screening combinations to distinguish nonlinear systems, which has good identifiability for nonlinear systems. The neural network with GMDH organizational structure is a feed-forward neural network, which plays a predictive function mainly in the application and is also known as GMDH neural network. The main principle is that in a system if there are m variables X_i (1, 2, ..., m) and one output variable y, the K-G polynomial expression results of each fitting trajectory can be obtained according to formula (1). Among them, a_0, a_i. a_{ij}. a_{ijk},...(i, j, k = 1, 2, ..., m)are the coefficients. This function can describe past data trajectories and can approximate any linear function. It can be seen from the formula that the contribution of each item in the polynomial to the fitting degree is not the same. The system takes advantage of the characteristics of GMDH neural network self-organization, constantly removes the items with less contribution to the fitting degree, and reduces the calculation scale so that the model becomes more concise and practical and the effect is better.

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m} a_{ijk} x_i x_j x_k + \cdots \quad (1)$$

As a feedforward neural network, the main difference between a GMDH neural network and other types of neural networks is that the data training process is formed dynamically, and it will change constantly. When the neuron enters the next level to form a new neuron, the self-organization feature of the network automatically removes the neurons with poor contribution. This makes the number of neurons in each layer not fixed. Figure 1 is the GMDH network diagram. $\hat{y}$ represents the evaluation value of the network. X_{il} is the input variable *i* of the first sample, $i = 1, 2, \ldots, m$. The m is the number of input variables in the sample. $\hat{y}_{jkl}$ is the evaluation value of the *k* neuron in layer *j* of the first sample, $k = 1, 2, \ldots, m$. r_{jk}^2 is the evaluation index of the *k* neuron in layer *j*. The threshold in layer *j* is R_j.

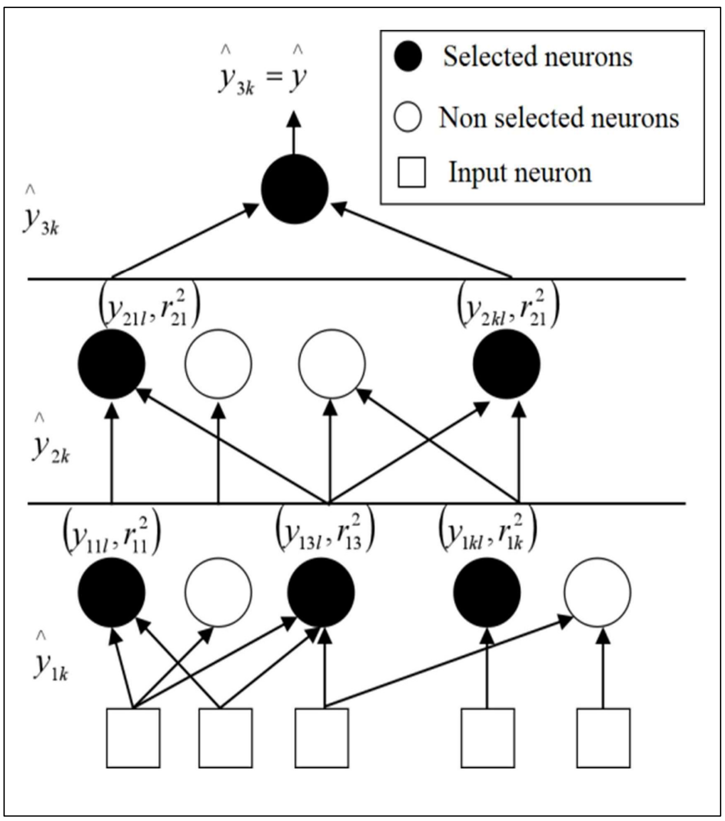

Figure 1. An iterative design model

The basic neuron of the GMDH neural network is a dual input and single output structure. A polynomial of two variables and two degrees represents the relationship between the input and the source.. Among them, the coefficients *a, b, c, d, e, f* are mainly determined by the least square method. The unit output here is the same as the content of the lower units. The production of the upper two different units is the input content of the lower unit, and the output variable of the processing unit is the polynomial with two degrees of the input variable. When the number of layers in the network is more than one layer, the polynomial order will be increased by 2. Finally, the network is a multinomial expression of 2k orders. GMDH neural network has shortcomings in practical application, mainly because some structures are unstable. This is because the intermediate variable is increasing at the iteration level, and the goodness of fit with the output variable is higher. This makes the correlation between the variables in the middle too high. The recognition rate of the least two-component system in a nonlinear system is not very good. This is because its retrieval principle is the use of gradient information, which will have the possibility of going to a local minimum to bring limitations and misleading to searching. To solve these shortcomings of the GMDH neural network, a fuzzy inference model was proposed to optimize the primary processing unit with the fuzzy inference model in this paper.

3.2. Optimize GMDH Algorithm

Reasoning is the process of drawing possibly imprecise conclusions from a data set that is inaccurate. The reasoning in human thinking is often possible and approximate. That is to say, there is a certain degree of fuzziness in human's natural thinking. Fuzziness is a transitional stage, which shows the intersection and integration of things in performance, attributes and other parameters. This is the natural state of the existence of things and also is an important research content in the study of artificial intelligence. Fuzzy reasoning generally uses two-valued logic or multi-valued logic to complete. Among them, the most crucial theory is composition rules that convert the conditional statement "if x is A, y is B" to a fuzzy relation, which Zade proposed in the 70*s* in the last century. On this basis, the fuzzy theory with different fuzzy logic methods with fuzzy truth values is introduced. Fuzzy inference is introduced to replace the basic processing unit in the GMDH algorithm to improve the effectiveness of the algorithm.

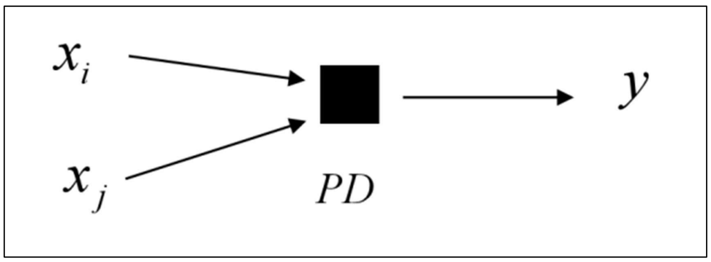

Figure 2. GMDH network basic processing unit

Figure 2 shows a neuron of fuzzy inference GMDH model. The fuzzy model here mainly contains k fuzzy rules. The expression of rule k ($1 \leq k \leq K$) is R_K: If x_1 is $A^k{}_1$ and x_2 is $A^k{}_2$ Then y is y^k Among them, x_1, x_2 stand for input, and y is the input of the model. are the membership functions of the input variables x_1, x_2. The Gauss function is chosen here, which is shown in formulas (2) and (3). Among them $a_{1k}, a_{2k}, b_{1k}, b_{2k}$,are the parameters of the model.

$$u_{1k}(x_1) = \exp\left\{-\frac{(x_1 - a_{1k})^2}{b_{1k}}\right\} \tag{2}$$

$$u_{2k}(x_2) = \exp\left\{-\frac{(x_2 - a_{2k})^2}{b_{2k}}\right\} \tag{3}$$

Figure 3 shows a normal distribution graph of the Gauss function shaped like an inverted clock. Parameter a is the peak value of the Gauss curve, b is the corresponding abscissa, and c is the RMS width of Gauss, which controls the width of the function.

The topological structure of the fuzzy inference model is composed of four layers. The first layer is the input layer, which is the direct connection between the neurons and the input variables. The second layer is the fuzzification layer. All nodes in this layer are divided into k groups, and each node represents the first half of a fuzzy rule. Each node is used to calculate the membership value of the input variable. The fuzzy inference layer in the third layer is mainly applied to the product principle to calculate the excitation intensity of each fuzzy rule. The formula is $u_k(x) = \prod_{i=1}^{2} u_{ik}(x_i)$. The defuzzification calculation is carried out on the fourth layer to output the calculated results. The calculation formula of input value is $y = \frac{\sum_{k=1}^{k} u_k(x)\omega_k}{\sum_{k=1}^{k} u_k(x)}$. The mixed projection method is used to estimate the parameters of the $a_{1k}, a_{2k}, b_{1k}, b_{2k}$ model here.

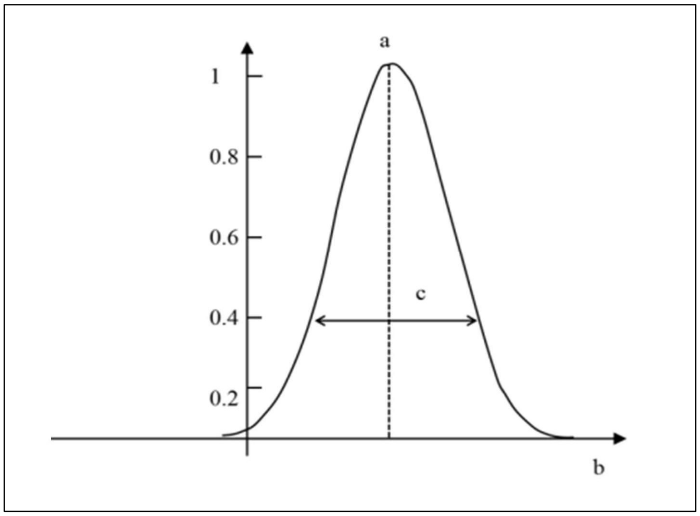

Figure 3. The normal distribution curve of the Gauss function

In the modelling of self-organizing data mining, the training samples are usually divided into different categories. This is to make the generated model have a certain balance in generalization ability and complexity. Prior knowledge must be able to control the complexity of the model effectively. In the modelling of GMDH algorithm, based on this principle, it is necessary to use effective external information so that the best model can be selected from the given data sample. The external information here is the information data that has not been used in the parameter estimation of the model. Figure 4 is the least fit variance map. The abscissa represents the complexity, and the ordinate is the least fitting square error. The least-square error in the selection data set is the minimum square error in the training set. With the increasing complexity of the model, the number of variables contained in the model and the maximum index of the polynomial corresponding to the model will

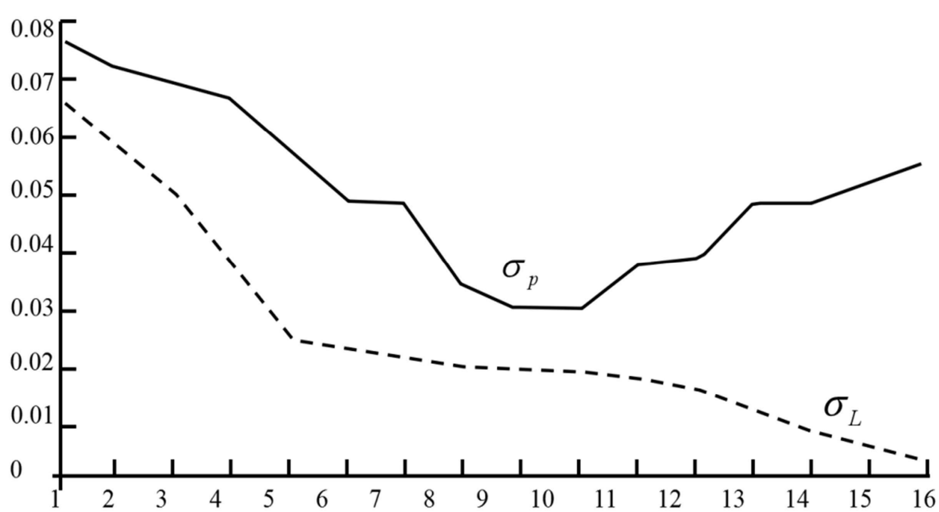

Figure 4. Minimum fitting variance

increase. The minimum squared error calculated in the training data set becomes smaller. It is impossible to determine whether the model has over-fitting or not only by this information, so it is necessary to use the external criterion to calculate the least square error when selecting the excluded training samples. The lowest value can be obtained, which is the corresponding minimum value. In self-organizing data mining, each step will select many models. When it is not necessary to improve the external criterion at a certain stage, the model is the optimal complexity model can be considered, and the best balance of fitting generalization of training samples can be realized so as to end the modeling process of GMDH algorithm.

4. Result Analysis and Discussion

In this paper, based on the basic data of the sports industry in M Province, the GMDH algorithm based on an optimized value chain was simulated and tested. Construction of the fuzzy GMDH network model was carried out based on Knowledge miner5.0 software. Firstly, sample data collection was carried out. For the value evaluation of the sports industry, the input of the model represents the indicators of the value-driving factors of the sports industry, and the output is the value of the sports industry in the training samples. September 1, 2016 was selected as the basis point, and sports industry-related enterprise data in *M* province were selected as a sample. The indicators were extracted to get the sample data for this paper. Then, the sample data were pre-processed to map the data to the (0, 1) range. The raw data were normalized to reduce the dynamic changes in the data processing process to increase the effectiveness of the evaluation results. Although the GMDH network has good data recognition and self-organization ability, data can also be trained without being processed; if the input data preprocessing can reduce the process and range of data recognition, the results are more practical and accurate. The formula used in data preprocessing is $x_i = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}}$. Among them, the x_i represents the data of the *i* row in any column. $x_{i,\ min}$ is the minimum value of all the data in the column in which x_i is located and $x_{i,\ max}$ is the maximum value of all the data in the column in which x_i is located.

$$
\begin{aligned}
A: & (y_1; x_{11,} x_{12,} \cdots x_{1n,}) \\
& (y_2; x_{21,} x_{22,} \cdots x_{2n,}) \\
& \ldots\ldots \\
& (y_s; x_{s1,} x_{s2,} \cdots x_{sn,})
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
B: & (y_{s+1}; x_{s+1,1,} x_{s+2,2,} \cdots x_{s+1,n,}) \\
& \ldots\ldots \\
& (y_m; x_{m1,} x_{m2,} \cdots x_{mn,})
\end{aligned}
\tag{5}
$$

Based on the construction of the GMDH network, the model is applied to fuzzy reasoning. The selected sample data are trained as input and output to establish a fuzzy GMDH value evaluation model. The processed sample data are divided into training samples and test samples. The sample data sequence is set as $\{y_i, x_{ij}\}$ ($i = 1, 2,.., m$), ($j = 1, 2, .., n$). The number of network input signals is *n,* the output variable is *y*, and the input variable is *x*. The training sample data set *A* and the test sample data set B are obtained, which are shown in formulas (4) and (5). The sequence generates the sample set *W*, which contains data samples with m inputs and outputs, $N_W = m$. 80% of them are training samples, and the rest are test sample data. In sample *A*, any X_i in the *N* inputs of the sports industry data is selected. When the input variables are set, the corresponding *y* can be obtained as the output variable. In the basic processing unit fuzzy reasoning model, $n(n-1)/2$ neurons can be generated. This can get the first layer of the initial network. Screening criteria can be used in neuronal screening. Assuming that that the R_j is the threshold of layer *j*. When $r_{jk}^2 < R_j$, the neuron is retained as the next input, the neuron that does not satisfy the condition is deleted. Finally, the value evaluation model of the sports industry established in the form of self-organization based on fuzzy GMDH network can be obtained, which is connected with the output layer directly connected with the neurons retained by the output layer.

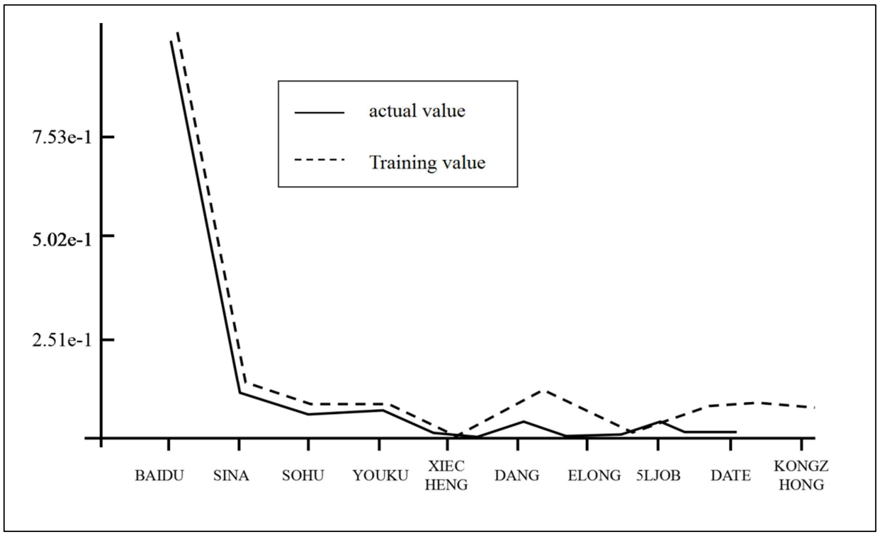

Figure 5. GMDH Comparison of data fitting process

To verify the practicability of the model, the basic GMDH network model and fuzzy GMDH network model were selected in this paper for comparative analysis, which provided the scientific basis for the study of sports industry development to provide industrial value. The basic GMDH network model is a linear output model. The result is that y = -6.987e - 2X5 + 9.528e - 1X11 + 3.934e - 2. The residual square sum of the prediction is 0.321 and the average absolute error percentage is 23.01%. The approximate error variance is 0.0187. The data fitting process is shown in Figure 5. Among them, the solid line represents the actual value, and the dotted line is the training value. It can be seen from the data that the fitting effect is not very good.

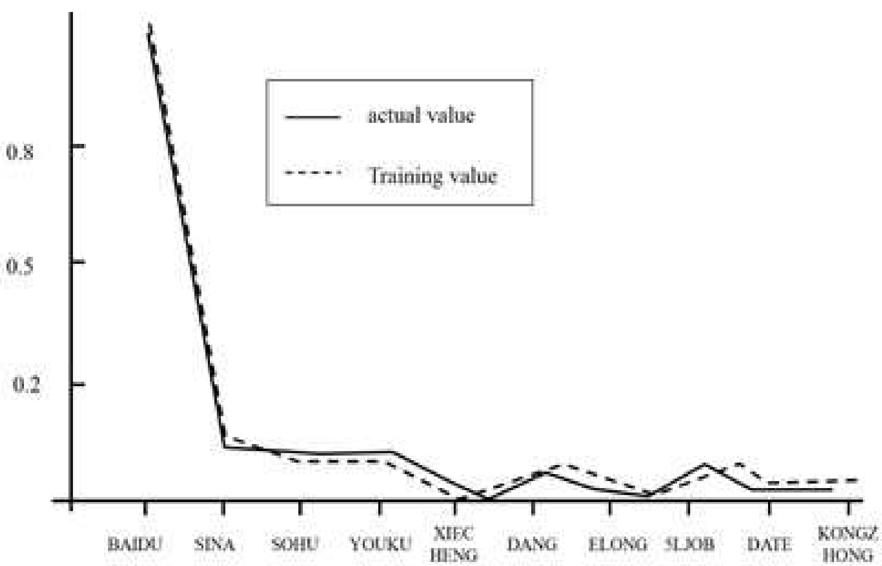

Figure 6. Comparison chart of data fitting process in fuzzy GMDH network model

In the simulation test, the structure of the fuzzy GMDH network model is $y = -9.815e - 1X12 + 5.7879e - 1$. The predicted value is 0.1785132, and the predicted value after reduction is 445031.1900003. The actual value is 45930.7300001, and the error is -1.95324%. The data fitting process is shown in Figure 6. The solid line represents the actual value, and the dotted line is the training value. It can be seen from the data that the fitting effect is better.

5. Conclusion

The advantages of the group method of data handling (GMDH) are that it can actively realize global optimization, adapt to solve small sample data, have strong self-organization, and automatically select and build concise model structures. The high growth of the sports industry determines the uncertainty of the future. The history of the sports industry is not long, and the data are limited, so it is difficult to evaluate and develop the value of the sports industry. Therefore, in this paper, the GMDH algorithm based on an optimized value chain was put forward to study the value of the sports industry, so as to provide a scientific basis for the prediction of the development of the sports industry. Based on the basic structure and process of the GMDH model, the advantages and disadvantages of the algorithm were analyzed. Fuzzy inference was introduced to replace the basic processing unit in the GMDH algorithm to improve the effectiveness of the algorithm. The fuzzy GMDH network was used to simulate the sports industry in *M* Province, and the degree of fit between the basic GMDH algorithm and the fuzzy GMDH network was compared. The results of this paper are as follows: the fuzzy GMDH algorithm based on the value chain can provide a more accurate prediction and evaluation basis for the development of the sports industry; the prediction accuracy is good, and the gap between fitting degree and prediction value is small. This proves that the study is successful. Of course, the results also show that the algorithm still has improved space. Therefore, further enhancing the model structure of the GMDH algorithm is the focus of future research.

References

[1] Yi, H. E., Chen, J. (2015). Research on Line Loss Data Pretreatment in Distribution Network Based on GMDH Algorithm. *Dianli Xitong Baohu Yu Kongzhi/Power System Protection & Control*, 43(9), 42-46.

[2] Kasaeian, A., Ghalamchi, M., Ahmadi, M. H., et al. (2017). GMDH Algorithm for Modeling the Outlet Temperatures of a *Solar Chimney Based on the Ambient Temperature. Mechanics & Industry*, 18(2), 216.

[3] Yin, W., Hu, W., Hui, F., et al. (2015). Inverse Determination of Material Parameters Based on Decoupled GMDH Algorithm. *China Mechanical Engineering*, 26(9), 1215-1221.

[4] Chang, F., Hwang, Y. (2015). A Self-Organization Algorithm for Real-Time Flood Forecast. *Hydrological Processes*, 13(2), 123-138.

[5] Antanasijevi, D., Antanasijevi, J., Pocajt, V., et al. (2016). A GMDH-Type Neural Network with Multi-Filter Feature Selection for the Prediction of Transition Temperatures of Bent-Core Liquid Crystals. *RSC Advances,* 6(102), 99676-99684.

[6] Osanaiye, O., Cai, H., Choo, K. K. R., et al. (2016). Ensemble-Based Multi-Filter Feature Selection Method for DDoS Detection in Cloud Computing. *Eurasip Journal on Wireless Communications & Networking*, (1), 130.

[7] Inbarani, H. H., Bagyamathi, M., Azar, A. T. (2015). A Novel Hybrid Feature Selection Method Based on Rough Set and Improved Harmony Search. *Neural Computing & Applications*, 26(8), 1859-1880.

[8] Parsaie, A., Haghiabi, A. H. (2015). Predicting the Longitudinal Dispersion Coefficient by Radial Basis Function Neural Network. *Modeling Earth Systems & Environment*, 1(4), 1-8.

[9] Najafzadeh, M. (2015). Neuro-Fuzzy GMDH Systems Based Evolutionary Algorithms to Predict Scour Pile Groups in Clear Water Conditions. *Ocean Engineering*, 99, 85-94.

[10] Bodyanskiy, Y., Tyshchenko, O., Kopaliani, D. (2015). A Hybrid Cascade Neural Network with an Optimized Pool in Each Cascade. *Soft Computing*, 19(12), 3445-3454.