

Targeted Poverty Reduction Model of the Needy Undergraduates Based on GA-SVM

Bairui Tao, Kaida Liu, Fengjuan Miao, Tongri Sun, Rui Miao

College of Communications and Electronics Engineering, Qiqihar University, Qiqihar, Heilongjiang 161006, China

Keywords: GA-SVM; Targeted Poverty Reduction Model; Needy Undergraduates; Precision-Targeted Subsidization.

Abstract: At present, the targeted funding plays an important role in solving the problem of needy undergraduates' schoolwork and promoting fair education. However, the current approach to identify needy undergraduates is full of uncertainty and the qualitative standard is lack of objective basis. This paper takes freshman of a major in S University as the research object, and employs the data of students' daily lives which objectively and truly reflects their life and behavioral characteristics to implement the simulation modeling for the study of targeted poverty reduction model of the needy undergraduates based on GA-SVM. This will be an effective supplement to the affirmation of the needy undergraduates in the work of a precision-targeted subsidization.

1. Introduction

In recent years, a series of major decisions of the Party Central Committee have laid a widespread stress on the relief work for impoverished students. For instance, the Fifth Plenary Session of the 17th CPC Central Committee, the 18th national Congress of the CPC, and the 3rd and 5th Plenary Sessions of the 18th CPC Central Committee have emphasized the importance of the education fairness^[1,2]. China vigorously advocates the education fairness and continuously improves the assistance for impoverished students, while establishing a sound relief system for impoverished students in colleges and universities to effectively help students with family economic difficulties^[3-5].

2. Overall Design for Grant Evaluation System

The paper employs GA-SVM model to simulate and forecast the data collected from the campus card records in canteen of freshman majored in internet of things from S University. The overall design diagram is shown in Fig.1.

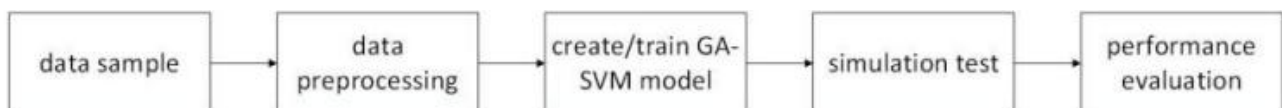


Fig. 1 System design block diagram

2.1 SVM for Regression Prediction

SVM(Support vector machine) is a neural network model developed by Vapnik in the 1990s to study small samples and small probability events to seek the optimal solution between the complexity of the model and the learning ability based on the limited sample information. It is assumed that the sample training set contains l vectors $(x_i \in \mathbb{R}^d (i=1,2,\dots,l))$ of d -feature space with, and $y_i \in \{-1,1\}$ is associated with each vector x_i .

The problem of finding the optimal hyperplane is transformed into solving the convex quadratic equation.

$$\begin{cases} \min: \frac{1}{2} \|w\|^2 \\ s.t.: y_i(w \cdot x_i + b) \geq 1, i = 1, 2, 3, \dots, l \end{cases} \quad (1)$$

w,b represent the coefficients of the hyperplane equation $f(x)=w \cdot x+b$. Lagrangian function is constructed based on the convex quadratic programming in the optimization theory, namely $\alpha_i(i=1,2,\dots,l)$ as Lagrangian multiplier in the equation.

$$L(w, \alpha, b) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i y_i \times (w \cdot x_i + b) + \sum_{i=1}^l \alpha_i \quad (2)$$

The corresponding dual problem after mapping to the high-dimensional feature space becomes:

$$\begin{cases} \max Q(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ s.t. \begin{cases} \sum_{i=1}^l \alpha_i y_i = 0, & i = 1, 2, \dots, l \\ 0 \leq \alpha_i \leq C \end{cases} \end{cases} \quad (3)$$

$K(x_i, x_j)$ represents the kernel function type used by SVM converting from low-dimensional space to high-dimensional space.

2.2 GA-SVM Algorithm Design

The paper adopts GA-SVM algorithm, an improved algorithm based on GA(Genetic Algorithm and SVM). The algorithm introduces GA into SVM prediction modeling so as to perform feature selection and parameter optimization simultaneously to improve the construction efficiency and predictive accuracy of prediction models of grants for impoverished students. Among them, GA is for feature selection and parameter optimization, and SVM for modeling and prediction. The flow chart of GA-SVM algorithm is shown in Figure 2.

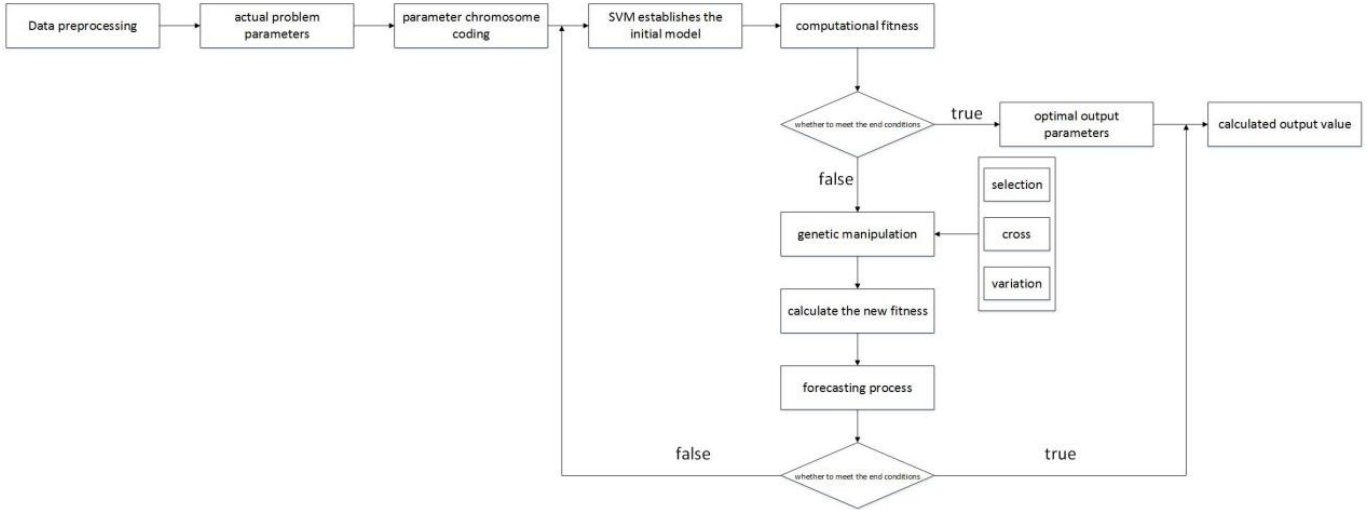


Fig.2 GA-SVM algorithm flow

3. Data Simulation Based on GA-SVM Algorithm

The experimental subjects are one-year campus card records in canteen of freshmen majored in the Internet of Things from S University. There are more than 20,000 data, including name, time(morning, afternoon and evening), location, and consumption situation. The corresponding data is shown in Table 1.

Table 1 Sample data

name	annual	month	day	consume	month	day	consume	month	day	consume	...
X ₁	Freshman	8	0	0	9	29	262.95	10	30	316.49	...
...
X ₆₅	Freshman	8	1	10	9	20	302.23	10	20	232.04	...

First, the kernel function K in equation (3) is set as RBF type, with which 33 input vectors are projected into a high-dimensional Hilbert space for analysis, thereby reducing the analysis complexity of the entire model. To improve the prediction accuracy of the model, the maxgen, the sizepop, the pCrossover, and the pMutation of GA are respectively set as 100, 20, 0.4 and 0.01. Since the high pMutation will aggravate the instability of the system while increasing the the diversity of the sample, it can be set too high. Then, the system will intelligently select the penalty term C and parameter g of the RBF kernel function which are substituted into the GA-SVM model for calculation.

The GA-SVM regression prediction model obtains a stable fitness value with the best fitness after 100 evolution calculations, as shown in Fig.3.

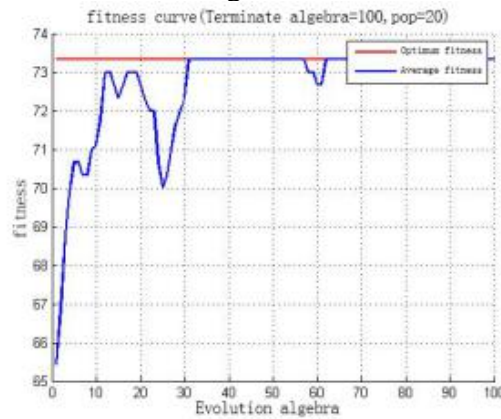


Fig.3 GA-SVM evolution curve

Fig.3 shows the optimal penalty item C and parameter suitable for the GA-SVM model are respectively set as 15.8383 and 0.7973 which are substituted into the GA-SVM model for calculation, and the training results of the training data set are shown in Fig.4. The test results of the test set with GA-SVM regression prediction model are shown in Fig.5.

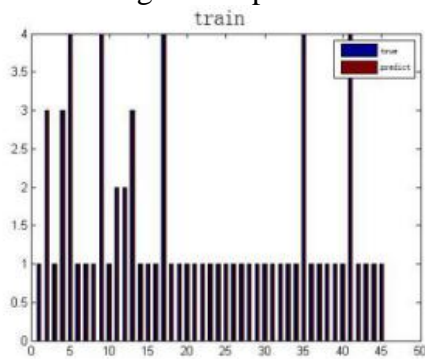


Fig.4 GA-SVM training results

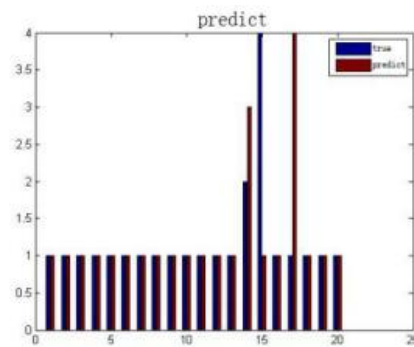


Fig.5 GA-SVM test results

Fig4-5 intuitively shows that GA-SVM can fit with the training data set very well with 100% accuracy (45/45), and its prediction results are consistent with the actual results with 85% accuracy (17/20).

In order to further verify whether the GA-SVM prediction model is superior to other machine learning algorithm models in the evaluation of poverty grants, the paper chooses the common BP neural network, SVM model and GA-SVM model to compare the prediction accuracy. Each model is tested five times and its accuracy and training time are shown in Table 2.

Table 2 The result comparison

	1		2		3		4		5	mean value		
	accuracy	training time	accuracy	training time	accuracy	training time	accuracy	training time	accuracy	training time	accuracy	training time
BP	60.00%	4.30s	60.00%	3.33s	65.00%	4.34s	80.00%	4.16s	50.00%	2.88s	50.00%	3.80s
SVM	75.00%	1.42s	65.00%	1.98s	70.00%	1.48s	55.00%	1.46s	65.00%	2.01s	65.00%	1.67s
GA-SVM	85.00%	13.70s	85.00%	14.60s	80.00%	14.86s	90.00%	12.27s	85.00%	13.02s	85.00%	13.69s

Table 2 shows that the BP neural network performs poorly in the evaluation of poverty stipends, and SVM is also superior to BP neural network and GA-SVM in prediction accuracy and time. However, GA-SVM has significantly improved prediction accuracy by optimizing the penalty term C and the parameter g compared to the original SVM model, keeping 85% accuracy.

4. Conclusion and Prospects

The SVM model proposed in the paper based on GA parameter optimization is applicable to the evaluation of poverty grants. With 85% prediction accuracy in average, it is superior to the traditional BP neural network and SVM model without parameter optimization in training accuracy and prediction accuracy, so as to provide some reference for the Precision-Targeted Subsidization.

Acknowledgments

We acknowledge the support of the National Natural Science Foundation of China (No. 61204127), Heilongjiang Education Science "Twelfth Five-Year" Plan for the Record (GBGH27), Science and Technology Project of Qiqihar (GYGG-201619, GYGG-201409), and New Century Excellent Talents in Heilongjiang Provincial University (1253-NECT025). This work was supported by the Supercomputing Center of Qiqihar University.

References

- [1] Sapiezynski J, Taratula O, Rodriguezrodriguez L, et al. Precision targeted therapy of ovarian cancer. [J]. Journal of Controlled Release, 2016, 243:250-268.
- [2] Baac H W, Ok J G, Maxwell A, et al. Carbon-Nanotube Optoacoustic Lens for Focused Ultrasound Generation and High-Precision Targeted Therapy[J]. Scientific Reports, 2012, 2(12):989.
- [3] Subbiah V, Gainor J F, Rahal R, et al. Precision Targeted Therapy with BLU-667 for RET-Driven Cancers[J]. Cancer Discovery, 2018:CD-18-0338.
- [4] Huang Chengwei.China's New Contributions to Global Poverty Reduction[J]. 2015(04):38-41.
- [5] Shi Li. Poverty Reduction and Effects of Pro-poor Policies in Rural China[J]. China & World Economy, 2014, 22(02):22-41.