Machine Learning Algorithm Study on Refrigerator Parameter Analysis

Wenjin Xu[1,2], Yuan Feng[1,*], Di Zhou[3]

[1]Ocean University of China, Qingdao, Shandong, 266101, China

[2]Qingdao University of Science&Technology, Qingdao, Shandong, 266069, China

[3]The 91049 Army of the PLA

* Corresponding Author

Keywords: Machine Learning; K-means; Support Vector Machine; Information Analysis

Abstract: In this paper, we systematically study the two kinds of data mining algorithm and software simulation process in refrigerator data analysis. This paper mainly introduces the related content of machine learning and data mining algorithm, including the concept, background of machine learning, model and the relevant methods. Then we implement two kinds of data mining algorithms: K - means algorithm and support vector machine (SVM) algorithm in refrigerator parameter analysis. Finally, we discuss the result of real-time data mining algorithm, using simulation software to realize the analysis of the data instance.

1. Introduction

In this paper, K-means algorithm and support vector machine (SVM) algorithm are discussed. K-means algorithm is a typical clustering algorithm based on distance, and the distance is the similarity evaluation index. Support vector machine (SVM) algorithm is put forward according to the statistical theory of a new machine learning method; it is founded on the VC dimension theory and based on structural risk minimization principle[1]. Compared with other machine learning algorithms, K-means algorithm and SVM algorithm are more suitable for application of big data processing in industry, and they are more advanced for operating[2].

The purpose of this article is to systematically study the two kinds of data mining algorithm and software simulation process in refrigerator data analysis. This paper mainly introduces the related content of machine learning and data mining algorithm, including the concept, background of machine learning, model and the relevant methods. Then we implement two kinds of data mining algorithms: K - means algorithm and support vector machine (SVM) algorithm in refrigerator parameter analysis. Finally, we discuss the result of real-time data mining algorithm, using simulation software to realize the analysis of the data instance.

This paper is divided into the following four parts: The first part mainly deals with the related theory of machine learning (including: the concept of machine learning, the development process, the main technology and algorithm) and the related knowledge of data mining technology (including: the definition of data mining, evolution process, and technical issues); The second part mainly studies the basic content, mathematical analysis and implementation steps of K-means algorithm; the third part mainly introduces the related content of SVM, such as the formation and development of SVM, the basic principle and training process of SVM, and the algorithm of SVM; the fourth part mainly use the K-means algorithm to do the data mining of refrigerator. The method is combined with the mathematical analysis of SVM algorithm and large data examples, and simulated by MATLAB and LibSVM to achieve the effect of estimation and analysis.

2. Machine learning and data in refrigerator data analysis

Data mining is a hot issue in the field of artificial intelligence and database. Data mining is a non-trivial process that reveals hidden, previously unknown and potentially valuable information from a large number of data in a database. Data mining is a decision support process. It is mainly based on artificial intelligence, machine learning, pattern recognition, statistics, database, visualization technology, etc[3]. It analyzes enterprise data automatically, makes inductive reasoning, excavates potential patterns, helps decision makers adjust market strategies and reduces risks. Make the right decision.

The knowledge discovery process consists of iterative sequences of the following steps:

(1) data cleaning (eliminating noise and deleting inconsistent data).

(2) data integration (multiple data sources can be grouped together).

(3) data selection (extracting data related to task analysis from the database).

(4) Data transformation (through aggregation or aggregation operations, data transformation and unification into a form suitable for mining).

(5) data mining (basic steps, using intelligent methods to extract data patterns).

(6) pattern assessment (based on a measure of interest, identifying a truly interesting pattern representing knowledge).

(7) knowledge representation (using visualization and knowledge representation technology to provide knowledge to users).

Step 1-4 is the different form of data processing to prepare data for mining. Data mining steps may interact with users or repositories[4]. Interesting patterns are provided to users or stored in knowledge base as new knowledge.

In the development of traditional refrigerators, detection is an important part. It is not only a test of whether the refrigerator meets the national standards can be mass-produced threshold, but also a summary of the research and development process.

Traditional refrigerator testing is generally required by national standards, in the specified temperature, the specified time, a long time (generally not less than 48 hours) start-up testing, time-consuming and energy-consuming, does not conform to the current trend of low energy consumption.

And according to the traditional inspection process, the final test results can not be obtained without the prescribed time.

But the design of refrigerator can not meet the final requirements at one time. It is often necessary to test the prototype of refrigerator in the design rules to check whether it meets the design requirements. If the voltage is unstable, the sensor data collected by the test is not up to standard, or the supervisor fails, such as the wrong operation of the collected data consistently losing some data and so on, it often occurs that the test is not completed at one time, thus further prolonging the development cycle and even delaying the production of new products.

The design scheme is to use artificial intelligence machine learning algorithm to dig and analyze the test data of refrigerator deeply, which has reached the early time of refrigerator test[5]. By collecting the data, we can predict whether the prototype refrigerator meets the research and development standard or not. Quasi, can be completed in advance of the refrigerator prototype research and development testing, and the use of in-depth analysis of its data to preliminary find the reasons for improvement. This can greatly reduce the workload of inspection, reduce energy consumption, improve the efficiency of research and development, reduce the time of research and development.

This method is based on artificial intelligence machine learning technology, through in-depth analysis of relevant parameters in refrigerator inspection.

There are more than 20 types of data collected by refrigerator inspection, including voltage, current, instantaneous power, temperature in different areas, including the temperature of refrigerator interior and refrigerator refrigeration system. Data collection is usually every 20 to 30 seconds. Traditionally, it takes at least 48 hours to check once. Manual monitoring of so many types of data, such a large amount of data.

3. Method of refrigerator data Analysis

The paper can tell whether the refrigerator design meets the design requirements.

In this paper, the first decision path is to use after real-time data collection, machine learning dimension reduction algorithm to find the most important data for refrigerator design, according to the design requirements of the most relevant data.

Then the algorithm uses artificial intelligence on-line learning target tracking algorithm, real-time analysis, to determine whether the selected data meet the design standards.

In the on-line learning and tracking analysis of the data, this method introduces the historical data of refrigerator test as a comparison, which can greatly improve the efficiency and reliability of the algorithm.

Finally, based on the national standards and R&D objectives, this method determines whether the prototype of refrigerator design meets the national standards and the established goals of R&D based on artificial intelligence classification algorithm[6].

In this paper, the second decision path is to cluster the collected data based on artificial intelligence clustering algorithm, so that singular points can be found, and then anomaly detection of singular points, if it does not meet the national standards and research and development goals, it can be judged in the detection process is not qualified, need to redesign.

The third determination path of this method is based on the same qualified refrigerator historical data and collected real-time data regression analysis to see whether the two are consistent, or test whether the refrigerator meets national standards or design requirements, do not meet the inspection process immediately determine unqualified, redesign.

This algorithm can greatly improve the efficiency of the refrigerator design prototype detection, because the refrigerator is a mature product, its technical performance will not fluctuate greatly[7]. Unqualified products are often identified at the initial stage of testing, such as refrigerator energy consumption level does not conform to national standards, it used to take 48 hours of testing, and then to collect energy consumption data for calculation and comparison, often takes 3-4 working days or even longer time, this algorithm can analyze data in real-time, comparison[8]. Often in the beginning of detection 2-4 hours can be judged from the algorithm whether it is qualified, unqualified can be immediately redesigned, saving time. Start to determine eligibility, on the end of real-time inspection guidance has been carried out, which can also save data processing time after inspection.

4. The result of analysis of refrigerator data

Data storage is in the form of Excel tables, data capacity is large, the number of rows more, through the MATLAB 2012 B version of the system with the function import. The data source is the mining and analysis of refrigerator related parameters. The following table shows that the table is D, E, F, I column in the table below.

Table 1 Refrigerator Related Parameters Test Table

	A	B	C	D	E	F	G	H	I	J
1	日期	时间	相对时间	柜口	中心	胆底	环温1	环温2	功率	耗电量
2	2013-11-12	09:16:10	1640.0	-18.1	-19.7	-18.8	25.0	24.9	47.0	0.2652
3	2013-11-12	09:16:40	1640.5	-18.1	-19.7	-18.7	25.0	24.9	46.0	0.2655
4	2013-11-12	09:17:10	1641.0	-18.0	-19.5	-18.6	25.0	24.9	45.0	0.2659
5	2013-11-12	09:17:40	1641.5	-17.9	-19.5	-18.6	25.0	24.9	45.0	0.2663
6	2013-11-12	09:18:10	1642.0	-17.9	-19.5	-18.6	25.0	24.9	45.0	0.2667
7	2013-11-12	09:18:40	1642.5	-17.9	-19.5	-18.6	25.0	24.9	44.0	0.2670
8	2013-11-12	09:19:09	1643.0	-17.8	-19.5	-18.6	24.9	24.9	44.0	0.2674
9	2013-11-12	09:19:39	1643.5	-17.8	-19.5	-18.6	24.9	24.9	43.0	0.2678
10	2013-11-12	09:20:09	1644.0	-17.8	-19.5	-18.7	24.9	24.9	43.0	0.2681
11	2013-11-12	09:20:39	1644.5	-17.8	-19.6	-18.7	25.0	24.9	43.0	0.2685
12	2013-11-12	09:21:09	1645.0	-17.8	-19.7	-18.8	25.0	24.9	42.0	0.2688
13	2013-11-12	09:21:39	1645.5	-17.9	-19.7	-18.8	25.0	24.9	42.0	0.2692
14	2013-11-12	09:22:09	1646.0	-17.9	-19.8	-19.0	25.0	24.9	42.0	0.2695
15	2013-11-12	09:22:40	1646.5	-17.9	-19.9	-19.0	25.0	24.9	42.0	0.2699
16	2013-11-12	09:23:10	1647.0	-18.0	-20.1	-19.2	25.0	24.9	41.0	0.2702
17	2013-11-12	09:23:40	1647.5	-18.1	-20.1	-19.2	25.0	25.0	1.0	0.2705
18	2013-11-12	09:24:10	1648.0	-18.1	-20.3	-19.4	25.0	25.0	1.0	0.2705
19	2013-11-12	09:24:40	1648.5	-18.2	-20.4	-19.5	25.0	25.0	1.0	0.2705
20	2013-11-12	09:25:10	1649.0	-18.3	-20.5	-19.6	25.0	25.0	1.0	0.2705

Table 2 Refrigerator Related Parameters Test Table two

	A	B	C	D	E	F	G	H	I	J
1	日期	时间	相对时间	柜口	中心	胆底	环温1	环温2	功率	耗电量
2	2013-11-11	07:21:09	85.0	-18.2	-19.5	-19.0	9.8	9.9	1.0	0.0129
3	2013-11-11	07:21:39	85.5	-18.1	-19.4	-18.8	9.8	9.9	42.0	0.0135
4	2013-11-11	07:22:09	86.0	-18.1	-19.3	-18.8	9.8	9.9	43.0	0.0139
5	2013-11-11	07:22:40	86.5	-18.0	-19.2	-18.7	9.8	9.9	43.0	0.0142
6	2013-11-11	07:23:10	87.0	-17.9	-19.2	-18.6	9.8	9.9	42.0	0.0146
7	2013-11-11	07:23:40	87.5	-17.9	-19.1	-18.6	9.8	9.9	41.0	0.0149
8	2013-11-11	07:24:10	88.0	-17.9	-19.1	-18.5	9.9	9.9	41.0	0.0153
9	2013-11-11	07:24:40	88.5	-17.8	-19.1	-18.6	9.9	9.9	40.0	0.0156
10	2013-11-11	07:25:10	89.0	-17.8	-19.1	-18.6	9.9	9.9	39.0	0.0159
11	2013-11-11	07:25:40	89.5	-17.8	-19.1	-18.7	9.9	9.9	39.0	0.0163
12	2013-11-11	07:26:10	90.0	-17.8	-19.2	-18.8	9.8	9.9	38.0	0.0166
13	2013-11-11	07:26:39	90.5	-17.8	-19.3	-19.0	9.8	9.9	38.0	0.0169
14	2013-11-11	07:27:10	91.0	-17.8	-19.4	-19.1	9.8	9.9	1.0	0.0170
15	2013-11-11	07:27:39	91.5	-17.9	-19.7	-19.3	9.9	9.9	1.0	0.0170
16	2013-11-11	07:28:09	92.0	-18.0	-19.8	-19.5	9.9	9.9	1.0	0.0170
17	2013-11-11	07:28:39	92.5	-18.1	-20.0	-19.7	9.9	9.9	1.0	0.0170
18	2013-11-11	07:29:09	93.0	-18.2	-20.1	-19.8	9.9	9.9	1.0	0.0170

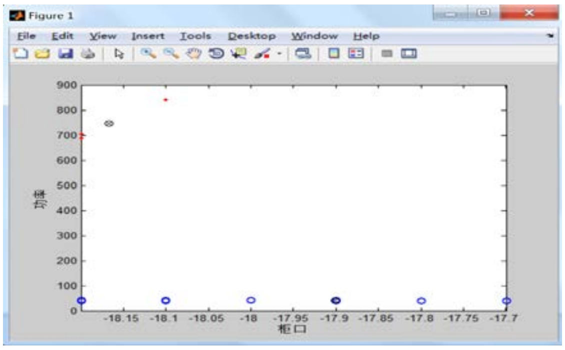


Figure 1. Tank Mouth-Power Relations and Center Locations (10℃)

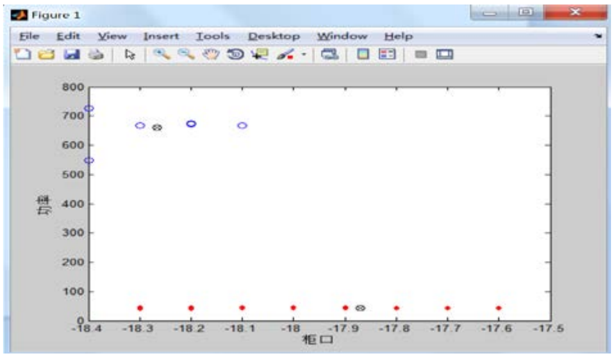


Figure 2. Tank Mouth-Power Relations and Center Locations (25℃)

4.1 Two-dimensional clustering results of single plot analysis

As you can see from the differences in the points shown in the figure, this test aggregates the data into two categories: the red dot represents the normal point, the blue circle represents the outlier, and the outlier (or singularity) below 5% indicates that the test results are good. The black circle represents the cluster center (that is the point of average power). As you can see from the graph, the circle formed by the clustering radius is the optimum radius automatically selected by the machine (that is, the point with 95% power is inside the circle). The smaller part radius of the normal point indicates that the data is more concentrated and the power variation range is smaller, while the larger part radius of the abnormal point indicates that the data is more dispersed, that is, the power variation range is larger.

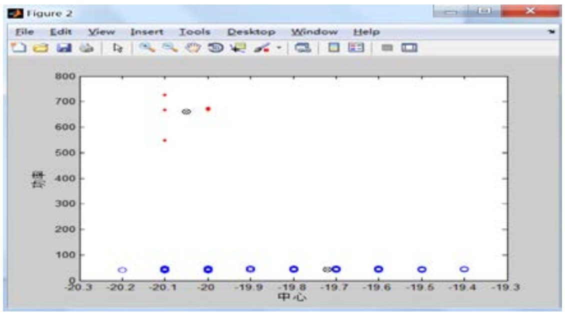


Figure 3. Center-Power Relations and Center Locations (25℃)

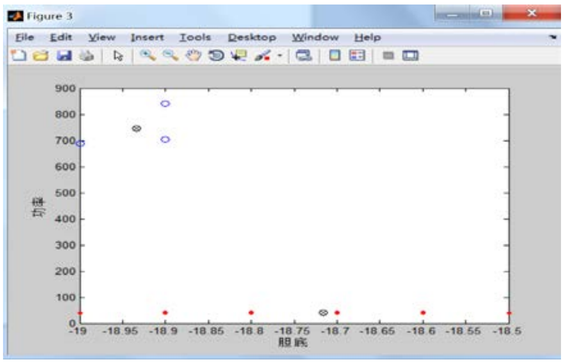


Figure 4. The Gallbladder-Power Relations and Center Locations (10℃)

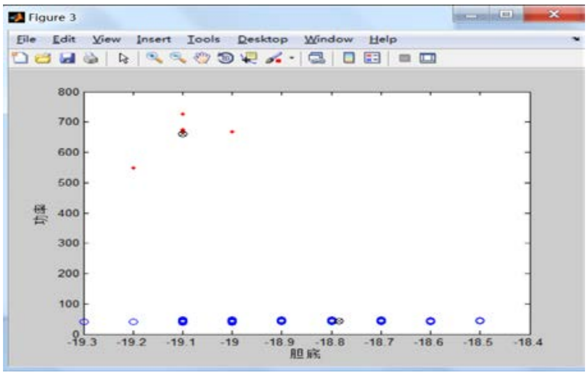
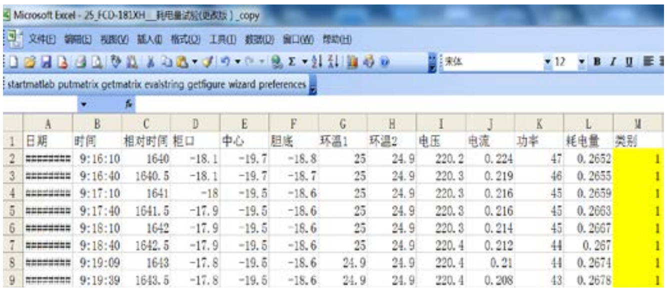


Figure 5. The Gallbladder-Power Relations and Center Locations (25℃)

4.2 Two dimensional clustering analysis of two loop temperatures

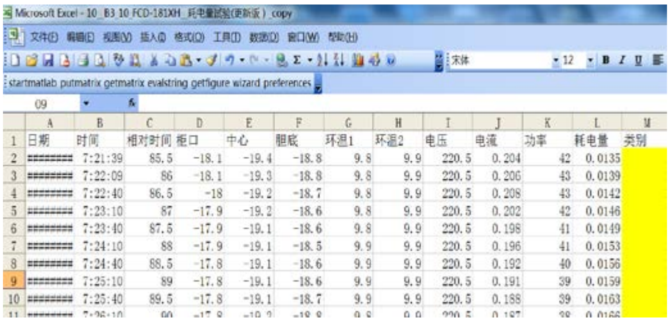
Comparing with the temperature of 10 C and 25 C, the distribution of abnormal points is relatively dispersed at 25 C and concentrated at 10 C. There is no obvious difference in the distribution of normal points.



Microsoft Excel - 25_FCD-181XH__耗电量试验(更改版)_copy

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	日期	时间	相对时间	柜口	中心	胆底	环温1	环温2	电压	电流	功率	耗电量	类别
2	########	9:16:10	1640	-18.1	-19.7	-18.8	25	24.9	220.2	0.224	47	0.2652	1
3	########	9:16:40	1640.5	-18.1	-19.7	-18.7	25	24.9	220.3	0.219	46	0.2655	1
4	########	9:17:10	1641	-18	-19.5	-18.6	25	24.9	220.3	0.216	45	0.2659	1
5	########	9:17:40	1641.5	-17.9	-19.5	-18.6	25	24.9	220.3	0.216	45	0.2663	1
6	########	9:18:10	1642	-17.9	-19.5	-18.6	25	24.9	220.3	0.214	45	0.2667	1
7	########	9:18:40	1642.5	-17.9	-19.5	-18.6	25	24.9	220.4	0.212	44	0.267	1
8	########	9:19:09	1643	-17.8	-19.5	-18.6	24.9	24.9	220.4	0.21	44	0.2674	1
9	########	9:19:39	1643.5	-17.8	-19.5	-18.6	24.9	24.9	220.4	0.208	43	0.2678	1

Figure 6 .The Four-Dimensional Data Output(25℃)



Microsoft Excel - 10_B3 10 FCD-181XH_耗电量试验(更新版)_copy

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	日期	时间	相对时间	柜口	中心	胆底	环温1	环温2	电压	电流	功率	耗电量	类别
2	########	7:21:39	85.5	-18.1	-19.4	-18.8	9.8	9.9	220.5	0.204	42	0.0135	
3	########	7:22:09	86	-18.1	-19.3	-18.8	9.8	9.9	220.5	0.206	43	0.0139	
4	########	7:22:40	86.5	-18	-19.2	-18.7	9.8	9.9	220.5	0.208	43	0.0142	
5	########	7:23:10	87	-17.9	-19.2	-18.6	9.8	9.9	220.5	0.202	42	0.0146	
6	########	7:23:40	87.5	-17.9	-19.1	-18.6	9.8	9.9	220.5	0.198	41	0.0149	
7	########	7:24:10	88	-17.9	-19.1	-18.5	9.9	9.9	220.5	0.196	41	0.0153	
8	########	7:24:40	88.5	-17.8	-19.1	-18.6	9.9	9.9	220.5	0.192	40	0.0156	
9	########	7:25:10	89	-17.8	-19.1	-18.6	9.9	9.9	220.5	0.191	39	0.0159	
10	########	7:25:40	89.5	-17.8	-19.1	-18.7	9.9	9.9	220.5	0.188	39	0.0163	

Figure 7. The Four-Dimensional Data Output (10℃)

4.3 Four dimensional clustering description

Because the MATLAB software can not realize the output of the four-dimensional image, the test results are output in the form of Excel tables. In the new EXCEL tables, there are many columns named category, and the test results are read into the tables in the form of 1,2 to produce classification.

The data is to use conventional keyword, the number of objective things, properties, location and the relationship between the abstract representation, in order to fit in this field using artificial or natural preservation, transmission and processing. Summarized by the data, interpretation, comparison of means to carry out excavation, making it a valuable part of settling down, and combined with the existing event, you can form event logic.[9] The information is time-sensitive, have a certain meaning, logic, after processing, decision-making valuable data stream. This study will establish the logical naming of the event, name resolution, authentication, security, routing, caching mechanism and other mechanisms.

The logic of events is the timing relationship between the data leading to the incident or causal. Huge amounts of data in abstract event logic is a very difficult scientific problems, different time, place, will produce different data .These data may be different sensors receive a variety of network protocols, transmission, and ultimately be stored in different database. From these space-time discrete data abstracted event logic to fuzzy logic theory and data mining technology. [10]In the event logic above, the effective data. From data to information in the middle of the fusion event logic to effectively improve the efficiency of the use of data, reduce data redundancy and improve the time validity.

SVM Support vector machine algorithm single graph analysis results

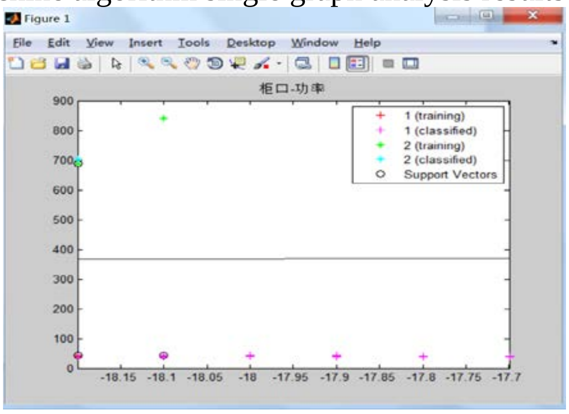


Figure 8.Tank-Power Relations Analysis Diagram(10°C)

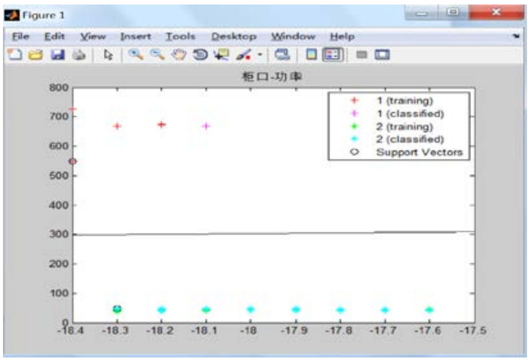


Figure 9. Tank-Power Relations Analysis Diagram(25°C)

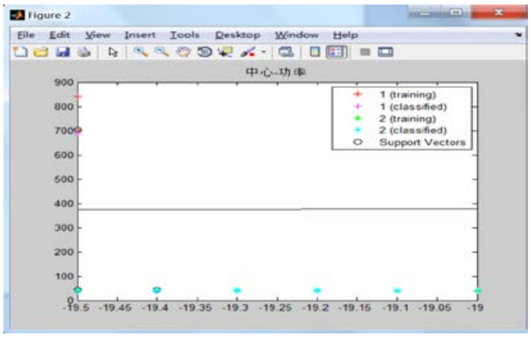


Figure 10. Center-Power Relations Analysis Diagram(10°C)

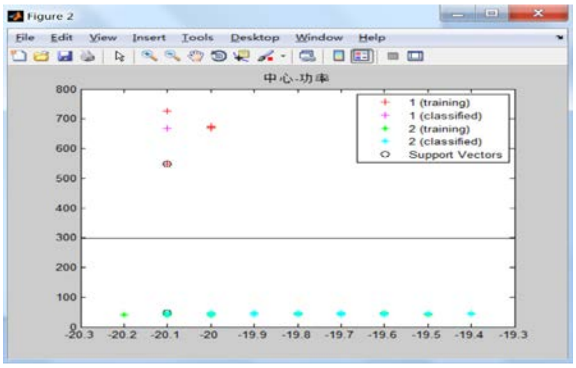


Figure 11. Center-Power Relations Analysis Diagram(25℃)

Traning represents the training sequence, classified represents the points classified, and circles represent the support vectors found according to the training sequence. The black lines on this map represent a hyperplane. One of them represents the handling of outliers (that is, other working conditions), and two represents the handling of normal points (that is, general working conditions). The test can be divided into two categories. The farther the distance from the hyperplane is, the better the test result is. The classification value of the normal point matches the training sequence better, but the classification value of the outlier matches the training sequence poorly. The percentage of the points in the normal value fraction is greater than 95%, indicating better results. For outliers, we need to analyze (continue data analysis or physical explanation).

4.4 Support vector machine algorithm two loop temperature comparative analysis results

When the ring temperature is 10, its hyperplane position is about 380 w. When the ring temperature is 25, its hyperplane position is about 300 w. At 10, the training sequence is relatively concentrated. At 25, the training sequence is relatively dispersed. But at this temperature, the training sequence and the classified points are relatively good. Because the distance between normal point and hyperplane is longer at 10 C, the classification effect is more obvious. At 25 C, because the distance between normal point and hyperplane is closer, the classification effect is relatively poor.

Acknowledgments

This work is supported by Shandong Critical ProjectcNo.2007GG10004018, and No.2018GGX105005, Shandong Key Auditing Project No. 1516SDSJ0102, and Qingdao National Labor for Marine Science and Technology Open Research Project QNLM2016ORP0405.

References

[1] Tom Mitchell.Machine Learning [M].2003-01

[2] LANG LEY Elements of machine learning [M].MorgnKaufmann Publisher Inc, 1998:1.

[3] Micheline Kamber, Data Mining:Concepts and Techniques,University of Illinois at Urbana-Champaign, 2006

[4] F. Lampariello, G. Liuzzi. A filling function method for unconstrained global optimization [J]. Computational Optimization and Applications, 2015, 61(3).

[5] Chih-Ming Chen,Hahn-Ming Lee. An Efficient Gradient Forecasting Search Method Utilizing the Discrete Difference Equation Prediction Model [J]. Applied Intelligence, 2002, 16(1).

[6] ROBERT TIBSHIRANI, Regression Shrinkage and Selection via the Lasso, 1996.

[7] Gary King Langche Zeng, Logistic Regression in Rare Events Data, 2001

[8] Hosmer,David W.;Lemeshow,Stanley. Applied Logistic Regression, 2000

[9] "How to Interpret Odds Ratio in Logistic Regression?". Institute for Digital Research and Education.

[10] Frank, E. X. Logistic regression and boosting for labeled bags of instances [J].Lecture Notes in Computer Science, 2004, (56): 272-281.