Research on Network Sales Development Plan Based on Big Data Neural Network

Lingyan Hu(✉) and Yumeng Dong

Beijing University of Technology, Beijing 100124, China
duoerxs@163.com

Abstract. In order to help Sunshine develop the sales plan based on the data, we first preprocess the data to prepare for the subsequent qualitative analysis of the model. Then it analyzes the relationship between the review and the reference values of other variables, and then further uses the multi-layer structure in the neural network, uses Star_ratings as the main variable, and calculates the relationship between some parameters and can be used as measures. Then select a hierarchical cluster to process the relevant data. With the help of Euclidean distance and Euclidean matrix, the most suitable variable is found, which can be used to track the similarity between different variables. Finally, a special model that uses time as an independent variable is created, and rating is used as a direct indicator of product reputation.

Keywords: MLP · Hierarchical · Clustering · Interaction Model

1 Introduction

Network sales are to sell products through the Internet, in essence is to use the Internet as a tool for sales. Like many new disciplines, "network marketing" also does not have a recognized and perfect definition [1]. In a broad sense, all marketing activities that take network marketing as the main means to achieve certain marketing goals can be called network marketing, that is to say, network marketing runs through the whole process of enterprises to carry out online business [2]. This article revolves around Sunshine's plans to launch and sell three new products in the online market: microwave ovens, baby pacifiers and hair dryers. To help Sunshine develop a sales plan based on the data, we first preprocess the data. The process involves removing invalid data from three products and converting complex text audits.

2 The Mathematical Principles Behind Neural Network-Multilayer Perceptron (MLP)

2.1 Prophase Pretreatment

The essence is a kind of feedforward neural network, which can map the input layer data to the output layer (only one-way propagation, no feedback) [6][7].

Z. Zhan et al. (Eds.): ICBDIE 2022, AHCS 5, pp. 1331–1342, 2023.
https://doi.org/10.2991/978-94-6463-034-3_137

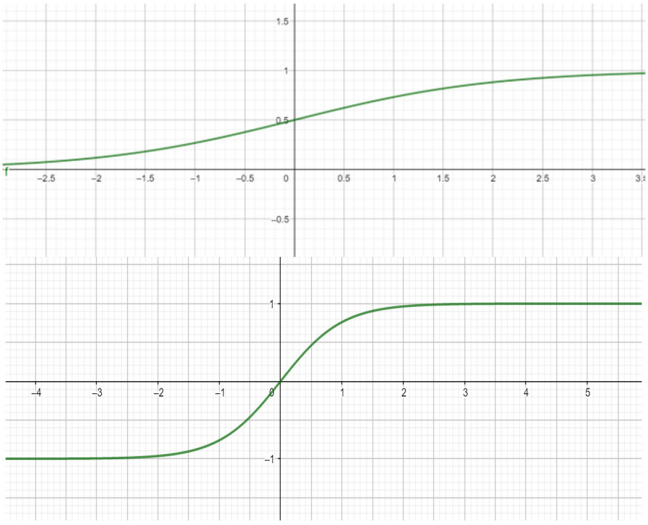


Fig. 1. (1) and (2) corresponding picture (original)

Compared with a single-layer perceptron, which can only process the input layer data with a linear function, it has multiple layers. The advantage of the perceptron is that it can use non-linear functions. This function is reflected in artificial neurons in the hidden layer. Artificial neurons will use non-linear excitation functions. From a statistical perspective, this process can also be seen as a combination of logistic regression models.

2.2 Non-linear Excitation Function Used in the Hidden Layer

(1) Sigmoid function can transform the value of an element to the interval (0, 1)

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

(2) Tanh function, which transforms the value of an element to the interval $(-1, 1)$ (Fig. 1)

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{2}$$

2.3 Steps and Results

(a) First, we use star_ratings and quantized reviews as dependent variables, vine, verified_purchase, helpful notes, and total notes as independent variables. We use the multilayer perceptron function in the SPSS neural network for analysis, where vine and verified_purchase are categorical variables, belong to factor; helpful notes and total notes are data variables, belong to covariates [3].

Covariate dimensions are different and need to be standardized.

(2) After setting the variables, the initial modeling begins.

The paper first sample 70% of the training samples to build the neural network model, 30% of the support samples to evaluate the function of our model, and finally observe the percentage of incorrect predictions from the results [4]. If the incorrect rate exceeds the maximum, we need to prevent overtraining in the next steps.

(3) Finally, perform a second modeling, observe the output of the SPSS system, and then analyze.

3 Analysis of Three Typical Articles

3.1 Hairdryer

For this neural model, we choose the dependent variable as star_ratings, whose data is at the output layer, and the output layer activation function is softmax; the independent variables are vine [5], verified purchase, helpful_votes, total_votes, whose data is input to the input layer, and the activation function of the hidden layer is tanh (x). In the network diagram, the blue line indicates that the synaptic weight is less than 0, and the gray line indicates that the synaptic weight is greater than 0 (Table 1 and Figs. 2 and 3).

The second column of the Table 2 the proportion of importance of the four independent variables, and the third column shows the proportion of the importance of the four independent variables that need to be normalized.

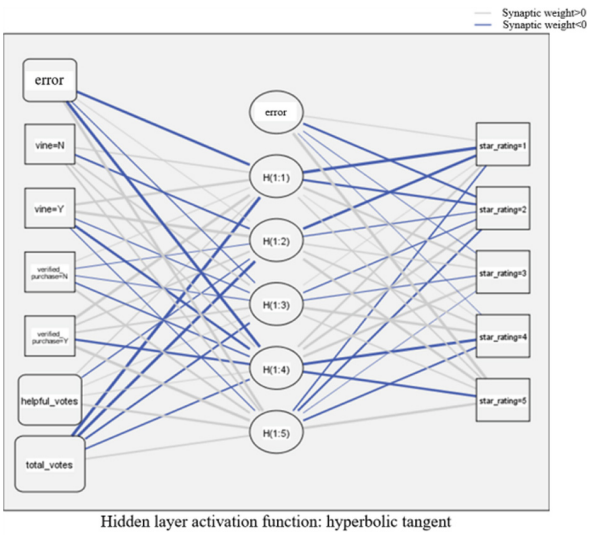


Fig. 2. Network topology (original)

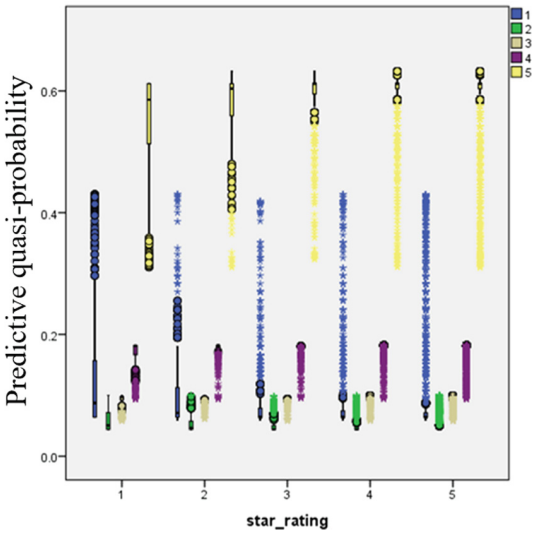


Fig. 3. Predictive quasi-probability (original)

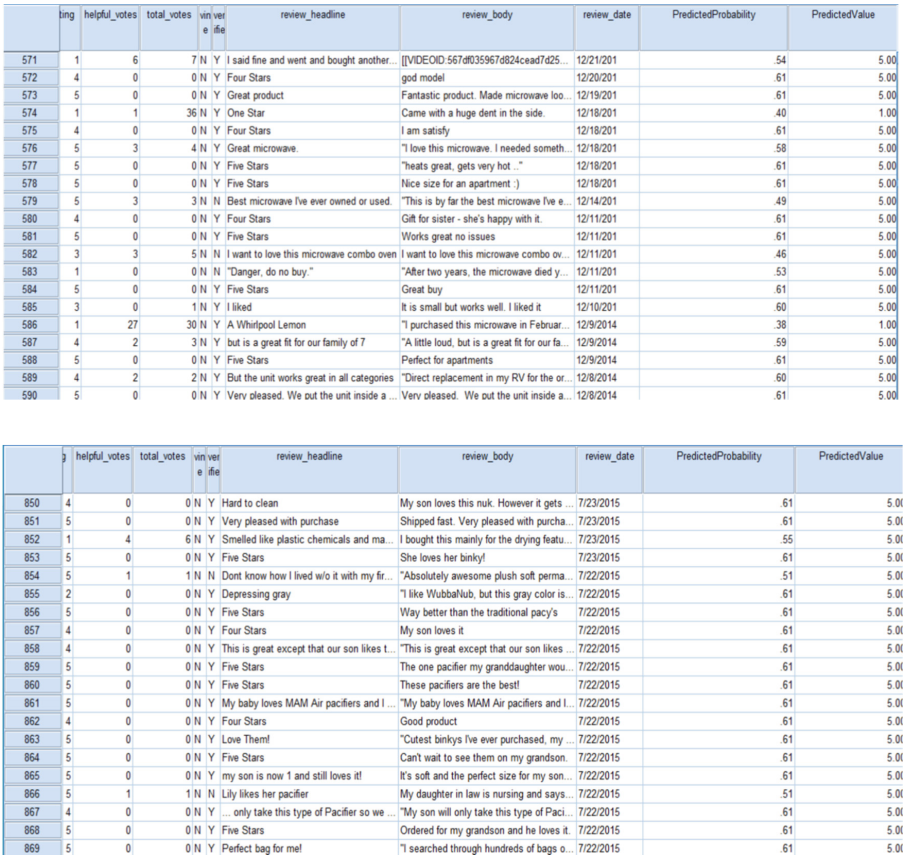
Table 1. Classification

Sample	measurement	Forecast					
		1	2	3	4	5	Correct percentage
Training	1	37	0	0	0	690	5.1%
	2	7	0	0	0	444	0.0%
	3	11	0	0	0	671	0.0%
	4	22	0	0	0	1435	0.0%
	5	60	0	0	0	4623	98.7%
	Overall percentage	1.7%	0.0%	0.0%	0.0%	98.3%	58.3%
Inspection	1	20	0	0	0	285	6.6%
	2	3	0	0	0	185	0.0%
	3	4	0	0	0	313	0.0%
	4	9	0	0	0	630	0.0%
	5	42	0	0	0	1979	97.9%
	Overall percentage	2.2%	0.0%	0.0%	0.0%	97.8%	57.6%

These two pictures show that the accmmended that companies use only 5 Star review predictiuracy rate of the five-star rating predicted by this neural network is very high, almost close to 100%, but the accuracy of other ratings is almost 0, and the overall accuracy rate is about 60%. Therefore, it is recoon results.

Table 2. Self-variable importance

	Importance	Normalization importance
Vine	.151	39.1%
verified_purchase	.155	40.1%
helpful_votes	.308	79.7%
total_votes	.386	100.0%



	ting	helpful_votes	total_votes	vine	verified	review_headline	review_body	review_date	PredictedProbability	PredictedValue
571	1	6	7	N	Y	I said fine and went and bought another...	[[VIDEOID:567df035967d824cead7d25...	12/21/201	.54	5.00
572	4	0	0	N	Y	Four Stars	god model	12/20/201	.61	5.00
573	5	0	0	N	Y	Great product	Fantastic product. Made microwave loo...	12/19/201	.61	5.00
574	1	1	36	N	Y	One Star	Came with a huge dent in the side.	12/18/201	.40	1.00
575	4	0	0	N	Y	Four Stars	I am satisfy	12/18/201	.61	5.00
576	5	3	4	N	Y	Great microwave.	"I love this microwave. I needed someth...	12/18/201	.58	5.00
577	5	0	0	N	Y	Five Stars	"heats great, gets very hot .."	12/18/201	.61	5.00
578	5	0	0	N	Y	Five Stars	Nice size for an apartment :)	12/18/201	.61	5.00
579	5	3	3	N	N	Best microwave I've ever owned or used.	"This is by far the best microwave I've e...	12/14/201	.49	5.00
580	4	0	0	N	Y	Four Stars	Gift for sister - she's happy with it.	12/11/201	.61	5.00
581	5	0	0	N	Y	Five Stars	Works great no issues	12/11/201	.61	5.00
582	3	3	5	N	N	I want to love this microwave combo oven	I want to love this microwave combo ov...	12/11/201	.46	5.00
583	1	0	0	N	N	"Danger, do no buy."	"After two years, the microwave died y...	12/11/201	.53	5.00
584	5	0	0	N	Y	Five Stars	Great buy	12/11/201	.61	5.00
585	3	0	1	N	Y	I liked	It is small but works well. I liked it	12/10/201	.60	5.00
586	1	27	30	N	Y	A Whirlpool Lemon	"I purchased this microwave in Februar...	12/9/2014	.38	1.00
587	4	2	3	N	Y	but is a great fit for our family of 7	"A little loud, but is a great fit for our fa...	12/9/2014	.59	5.00
588	5	0	0	N	Y	Five Stars	Perfect for apartments	12/9/2014	.61	5.00
589	4	2	2	N	Y	But the unit works great in all categories	"Direct replacement in my RV for the or...	12/8/2014	.60	5.00
590	5	0	0	N	Y	Very pleased. We put the unit inside a ...	Very pleased. We put the unit inside a...	12/8/2014	.61	5.00

	g	helpful_votes	total_votes	vine	verified	review_headline	review_body	review_date	PredictedProbability	PredictedValue
850	4	0	0	N	Y	Hard to clean	My son loves this nuk. However it gets ...	7/23/2015	.61	5.0(
851	5	0	0	N	Y	Very pleased with purchase	Shipped fast. Very pleased with purcha...	7/23/2015	.61	5.0(
852	1	4	6	N	Y	Smelled like plastic chemicals and ma...	I bought this mainly for the drying featu...	7/23/2015	.55	5.0(
853	5	0	0	N	Y	Five Stars	She loves her binky!	7/23/2015	.61	5.0(
854	5	1	1	N	N	Dont know how I lived w/o it with my fir...	"Absolutely awesome plush soft perma...	7/22/2015	.51	5.0(
855	2	0	0	N	Y	Depressing gray	"I like WubbaNub, but this gray color is...	7/22/2015	.61	5.0(
856	5	0	0	N	Y	Five Stars	Way better than the traditional pacy's	7/22/2015	.61	5.0(
857	4	0	0	N	Y	Four Stars	My son loves it	7/22/2015	.61	5.0(
858	4	0	0	N	Y	This is great except that our son likes t...	"This is great except that our son likes ...	7/22/2015	.61	5.0(
859	5	0	0	N	Y	Five Stars	The one pacifier my granddaughter wou...	7/22/2015	.61	5.0(
860	5	0	0	N	Y	Five Stars	These pacifiers are the best!	7/22/2015	.61	5.0(
861	5	0	0	N	Y	My baby loves MAM Air pacifiers and I ...	"My baby loves MAM Air pacifiers and I...	7/22/2015	.61	5.0(
862	4	0	0	N	Y	Four Stars	Good product	7/22/2015	.61	5.0(
863	5	0	0	N	Y	Love Them!	"Cutest binkys I've ever purchased, my ...	7/22/2015	.61	5.0(
864	5	0	0	N	Y	Five Stars	Can't wait to see them on my grandson.	7/22/2015	.61	5.0(
865	5	0	0	N	Y	my son is now 1 and still loves it!	It's soft and the perfect size for my son...	7/22/2015	.61	5.0(
866	5	1	1	N	N	Lily likes her pacifier	My daughter in law is nursing and says...	7/22/2015	.51	5.0(
867	4	0	0	N	Y	... only take this type of Pacifier so we ...	"My son will only take this type of Paci...	7/22/2015	.61	5.0(
868	5	0	0	N	Y	Five Stars	Ordered for my grandson and he loves it.	7/22/2015	.61	5.0(
869	5	0	0	N	Y	Perfect bag for me!	"I searched through hundreds of bags o...	7/22/2015	.61	5.0(

Fig. 4. Result analysis (original)

3.2 Microwave Oven and Pacifier

Judging from the prediction results, the prediction accuracy of this neural network is basically 60%. The prediction rates for 1-star and 5-star ratings are high, and the prediction has little significance for other star ratings (Fig. 4).

Table 3. 5 × 5 Euclidean matrix

	A	B	C	D	E
A	0	B-A	C-A	D-A	E-A
B	A-B	0	C-B	D-B	E-B
C	A-C	B-C	0	D-C	E-C
D	A-D	B-D	C-D	0	E-D
E	A-E	B-E	C-E	D-E	0

4 Model of Hierarchical Clustering

4.1 Selection of Data and Methods

Due to the large sample size, we first performed a simple random sampling of 10%. Since the sample size after sampling is larger than 1000, which is a large sample, and the type of data to be tested involves categorical variables, we use the Two Step Cluster method for clustering.

Its advantage is:

(1) To automatically determine the optimal number of classifications
(2) To quickly process large data sets.

4.2 Data Initialization

Set rated star_rating samples to A, helpful votes samples to B, total votes samples to C, vine to D, and verified_purchase to E, respectively.

Calculate the Euclidean distance between the two types of samples to construct a 5 × 5 Euclidean matrix. The diagonal is the distance from itself, so all values are 0.

Euclidean distance:

(See Table 3)

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3}$$

4.3 Clustering with SPSS

The system SPSS merges the two categories with the closest nature (the closest distance) into one category, and then finds the closest two categories from the n-1 category to merge.

Table 4. Cluster distribution

		Number of cases.	As a percentage of the combination.	As a percentage of the total
Cluster	1	5847	51.0%	51.0%
	2	142	1.2%	1.2%
	3	2086	18.2%	18.2%
	4	1791	15.6%	15.6%
	5	1604	14.0%	14.0%
	Combination	11470	100.0%	100.0%
Total		11470		100.0%

For variable clustering, the SPSS system generally defaults the correlation to the correlation (The squared Euclidean distance mentioned above). And so on, until all variables / observations are grouped into one category, the user then decides into several categories based on the specific problem and clustering results. The paper will use the final clustering results to determine which variable or variables will have a great correlation with the rating of the reviews, and the distance mentioned above will become a criterion for the company.

4.4 Analysis Based on the Results Below

The 3 and 4 columns are the percentages of the five clusters in the combination and the total number, respectively (Table 4).

In this example, the final classification is 5 categories, and its practical significance indicates five stars, of which the size of the first category accounts for 51.0% of the effective sample, the second category is 1.2%, the third category is 18.2%, and the fourth category It is 15.6%, and the 5th category is 14.0%. It can be seen that, because the amount of data provided by each star is not even, the case size varies greatly between categories (Table 5).

a. The amount of change is based on the number of previous clusters in the table.
b. Change ratio changes with respect to dual polyproducts.
c. The distance measurement ratio is based on the current cluster number rather than the previous cluster number.

Table 5. Automatic clustering results

Number of clusters.	Schwartz Bayesian criterion (BIC).	BIC change a.	BIC change ratio b.	Distance measurement ratio c
1	55181.948			
2	37683.584	−17498.365	1.000	1.355
3	24794.020	−12889.563	.737	2.153
4	18858.643	−5935.377	.339	1.131
5	13619.881	−5238.762	.299	1.900
6	10907.282	−2712.600	.155	1.383
7	8972.343	−1934.939	.111	1.201
8	7376.762	−1595.581	.091	1.498
9	6342.917	−1033.845	.059	1.177
10	5478.524	−864.393	.049	1.098
11	4699.842	−778.682	.045	1.333
12	4138.862	−560.981	.032	1.460
13	3783.927	−354.934	.020	1.346
14	3544.225	−239.702	.014	1.406
15	3400.732	−143.493	.008	1.437

In the second-order clustering, there are five variables we selected, namely total_votes, helpful_votes, verified_purchase, star_rating, and vine. The SPSS software automatically and comprehensively determines the final number of clusters based on four indicators: “BIC value”, “BIC change”, “BIC change rate”, and “distance measurement ratio” (Table 6).

Taking the result of star_rating as an example, in the 1–5 rating, 1 and 3 can be classified into the third category; 2 and 3 can be classified into the third category; 3, 4 are classified into the third category with themselves, and 5 and 1 can be classified into For the first category. The above classification is selected based on the maximum percentage.

The clustering quality of this example gives a good evaluation. Among the five variables, the importance of vine is the smallest, and the other four have reached the importance of 1.

Table 6. Star_rating results

		1		2		3		4		5	
		Frequency.	Percentage	Frequency.	Percentage	Frequency.	Percentage	Frequency.	Percentage	Frequency.	Percentage
Cluster	1	0	0.0%	0	0.0%	0	0.0%	0	0.0%	5847	87.2%
	2	28	2.7%	7	1.1%	12	1.2%	20	1.0%	75	1.1%
	3	728	70.5%	508	79.5%	850	85.1%	0	0.0%	0	0.0%
	4	0	0.0%	0	0.0%	0	0.0%	1791	85.4%	0	0.0%
	5	276	26.7%	124	19.4%	137	13.7%	285	13.6%	782	11.7%
	Combination	1032	100.0%	639	100.0%	999	100.0%	2096	100.0%	6704	100.0%

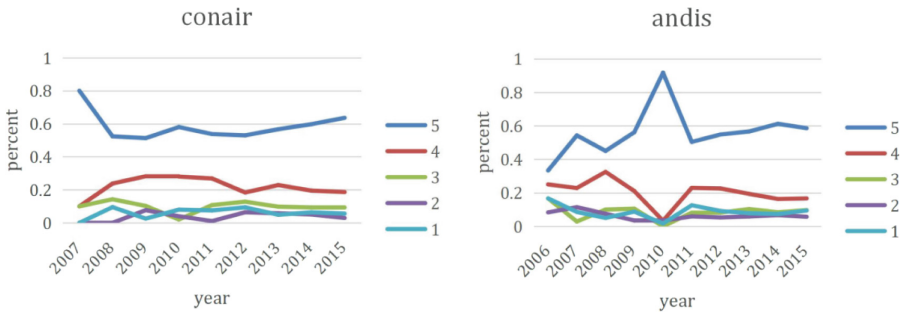


Fig. 5. Conair and andis results (original)

Fig. 6. Philips and wubbanub results (original)

5 Model Analysis of Time Measurement

(1) Hair Dryer

It can be seen from this graph that from the 10 years from 2006 to 2015, a total of 1993 andis products were sold on Amazon's website, and the conair was 3,095 (Fig. 5).

The two are compared horizontally. The five-star ratings of the two are at five different levels. All of the ratings are dominant, but after comparison, it can be seen that except for 2010, the proportion of 5-star ratings of conair is higher than that of andis, and the overall data is more stable.

(2) Pacifier

From the Fig. 6, it can be seen that from 2010 to 2015, Philips products sold a total of 1811 on Amazon.com, and from 2008 to 2015, Wubbanub products sold a total of 3051 on Amazon.

Wubbanub's data is similar to Philips's. Star ratings 1–3 have a lower variable percent in each year, only about 5%. The 4-star rating variable percent increased and reached its peak in 2008, however, it began to decline and stabilized at about 15% from 2009. Five-star ratings dominated the market, which was lower than the variable percent of the previous year in 2009, and rose slightly to 80% afterwards. In general, Wubbanub's

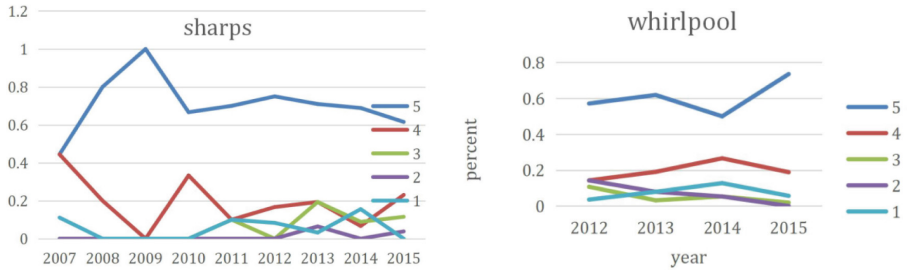


Fig. 7. Sharp and Whirlpool results (original)

5-star rating variable percent can be maintained at about 80%. It can be considered that users have a high evaluation of this product and have a good experience.

(3) Microwave oven

From the Fig. 7, it can be seen that from 2012 to 2015, whirlpool products sold a total of 238 products on the Amazon website. From 2007 to 2015, sharps products sold a total of 145 products on the Amazon website.

The amount of data in the microwave oven is much smaller than that of the hair dryer and pacifier, so the reference value is not great. It is recommended that the company select products with a large time span and sufficient sample size for analysis.

6 Conclusion

This paper analyzes and models and solves the actual network distribution problem of Sunshine Company, uses the multi-layer neural network method to take Star_ratings as the main variable, and calculates some parameters and the relationship that can be used as a measure. Then the hierarchical clustering algorithm is selected to deal with the relevant data. With the help of Euclidean distance and Euclidean matrix, the most suitable variable is found, which can be used to track the similarity between different variables. Finally, a special model that uses time as an independent variable is created, and rating is used as a direct indicator of product reputation.

References

1. Deng Lei. Research on the Interaction Mechanism of Individual Decision and Viewpoint in the Internet [D]. Beijing Jiaotong University, 2013.
2. Tomas M., Kai C., Greg C., Jeffrey D. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781v3[cs.CL]
3. Jacob D., Ming-Wei C., Kenton L., Kristina T. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2[cs.CL]
4. Ashish V., Noam S., Niki P., Jacob U., Llion J., Aidan N. G., Lukasz K., Illia P. (2017). Attention Is All You Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

5. Zhou Feiyan, Jin Linpeng, Dong Jun. A review of convolution neural networks [J]. Journal of computer Science, 2017. 40 (06): 1229-1251.
6. Chen Xianchang. Research on Deep Learning algorithm and Application based on Convolutional Neural Network [D]. Zhejiang Industrial and Commercial University, 2014.
7. Zhangling, Zhang cymbal, Yin Haifeng. Cross-coverage design algorithm for multi-layer feedforward network [J]. Journal of Software, 1999, 4 (07): 737742.