Research of Product Data Measurement Mode Based on Neural Network

Zhanwei Shen
School of Automation, Wuhan technology University
Wuhan, China
893060664@qq.com

Bei Feng*
School of Marxism, Wuhan University of Technology
Wuhan, China
46950255@qq.com

Yunlu Li
School of Automation, Wuhan technology University
Wuhan, China
907439438@qq.com

***Abstract*—With the evolution of traditional brick and mortar retail stores to online shopping, consumers are posting reviews directly on product pages in real time. We ultilize the RBF network to predict the sales number. We construct a GM (2,1) model to predict the sales of hair dryers, microwave ovens, and pacifiers in the next five years. The state transfer probability between positive, neural and negative reviews is built. The GINI correlation coefficient method is applied to determine whether the reviewer s attitude has a strong relationship with the star rating. The innovation of this article is that we build a comprehensive model with neural network, Markov and GINI correlation coefficient. Besides, Pearson correlation coefficient is used to test the GINI correlation coefficient which make the result more complete.**

Keywords- Grey prediction; Markov chain; RBF neural network; Bayesian network; Correlation coefficient.

I. INTRODUCTION

In this paper, our focus is the data mining problems provided by Sunshine Company. We ultilize the RBF network to predict the sales number. We set the four indicators (star rating, rated review, review validity and review reliability) of the products as the input and the sales number as the output of network. It proves that this network can be used as a data metric to determine product sales and reveal potentially successful products.

We construct a GM (2,1) model to predict the sales of hair dryers, microwave ovens, and pacifiers in the next five years. The total sales of all three are increasing rapidly. It indicates that these three types of products will have greater demand in the future market.

The state transfer probability between positive, neural and negative reviews is built. Markov chain model is made based on the annual review data, and then Bayesian network is used to solve the causal relationship between the reviews. The 1,2 stars of each product will lead to negative reviews while high star ratings do not necessarily lead to positive reviews.

The GINI correlation coefficient method is applied to determine whether the reviewer s attitude has a strong relationship with the star rating. It is clear that people's specific quality description of goods is strongly related to the star rating. Pearson correlation is consistent with the GINI coefficient method, which shows that our results are true and valid.

The innovation of this article is that we build a comprehensive model with neural network, Markov and GINI correlation coefficient. Algorithms such as Bayesian Network have obtained the intrinsic relationship among sales of brands of products, star ratings and reviews which process data with high accuracy and fast speed. Besides, Pearson correlation coefficient is used to test the GINI correlation coefficient which make the result more complete.

II. PREVIOUS RESEARCH

In article [1-4], sentiment analysis or language mining is used to analyze product reviews in detail and provide insights for future sentiment analysis. Article [5, 6] studies how people respond to user-generated product reviews (UGPRs) on various websites. The paper [7-9] studies when and how manufacturing companies adapt their marketing strategies to these comments. A normative model [10-13] was developed to solve several important strategic issues related to consumer review. Comprehensive analysis of various survey data analysis methods [14,15] can effectively improve the quality of information obtained from customer satisfaction surveys. Paper [16] proposes an online product rating model for customer satisfaction, which includes pre-purchase expectations of customers and actual product performance as determining factors. The influence of product type and comment nature on comment search behavior was investigated [17].

III. ASSUMPTIONS AND JUSTIFICATIONS

In order to simplify the problem and facilitate us to simulate real life conditions, we made the following basic assumptions:

978-1-7281-8145-5/20/$31.00 ©2020 IEEE
DOI 10.1109/ICEDME50972.2020.00138

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 19:25:46 UTC from IEEE Xplore. Restrictions apply.

- The number of product reviews represents the sales volume of the product.
- The collected stars and reviews are authentic and reliable.
- It is assumed that the state transition probability is entirely based on the conditional probability.

IV. Regularized Radial Basis Neural Network Model

The structure of a regularized RBF network is shown in the figure 1.

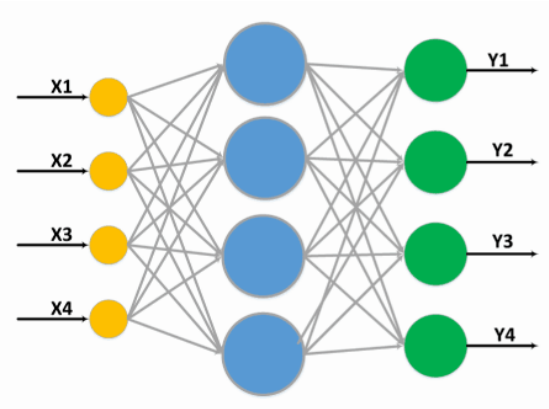

Figure 1: RBF network

A. FrequencyRegularized Radial Basis Neural Network

It has N input nodes, P hidden nodes, and I output nodes; the number of hidden nodes in the network is equal to the number of input samples; the activation function of the hidden nodes is usually a Gaussian radial basis function [20].The Gaussian radial basis function is set as:

$$\varphi(r) = e^{\frac{-r^2}{2*\delta^2}} \tag{1}$$

Distance from the center to the horizontal axis is the radius r. When the distance is equal to 0, the radial basis function is equal to 1, and the further the distance is, the faster the attenuation is. The parameter delta of the gaussian radial basis in the support vector machine is called the arrival rate or the speed at which the function drops to zero. Red line refers to delta=1, blue: delta=5, green: delta=0.5, the smaller the reach rate, the narrower it is.

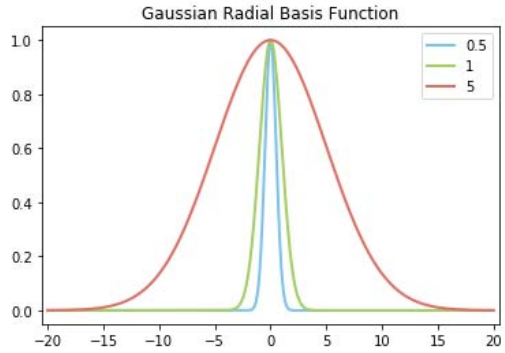

Figure 2: Gaussian radial basis function

Let any node in the input layer be i, any node in the hidden node j, and any node in the output layer k. We set the four indicators of the products as the input and the sales number as the output. Thus, the mathematical description of each layer is as follows: The input vector:

$$X = (X_1, X_2, \cdots, X_N)^T \; X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}) \; i = 1,2,\cdots,N \tag{2}$$

Where $X_{ij} \;\; j = 1,2,3,4$ refer to the four indicators of the product: star rating, rated review, review validity and review reliability. Activation function of any hidden node:

$$\varphi_j(x) \quad j = 1,2,3,\cdots,p \tag{3}$$

Gaussian functions are generally used. Output weight matrix is set as W

$$w_{jk} \quad j = 1,2,3,\cdots,0, \;\; k = 1,2,3,\cdots,l \tag{4}$$

Where is synaptic weights of the j-th node of the hidden layer and the k-th node of the output layer. Output vector:

$$Y = (y_1, y_2, \cdots, y_l) \tag{5}$$

The output layer neurons use a linear activation function. Given any unknown non-linear function f, a set of weights can always be found to make the regularization network's approximation to f better than all its possible choices.

B. Specific Parameter and Error Analysis

To identify data measures based on ratings and reviews that are most informative for Sunshine Company to track and determine combinations of text-based measure(s) and ratings-based measures, we select a total of N brands to form a new dataset for three products. $\left[\frac{9}{10} * N\right]$ is used as the training set to determine the relationship between the four input indicators and the output sales. The remaining $\left[\frac{1}{10} * N\right]$ is used as a test set to verify the relationship we get. The comparison of the real number and the predicted number is given with the residual

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 19:25:46 UTC from IEEE Xplore. Restrictions apply.

error.

To reflect the difference between each true value and the predicted value, the residual error is calculated to show that most predicted values are close to the real value except for some extremely large number. We give the pacifier in figures.

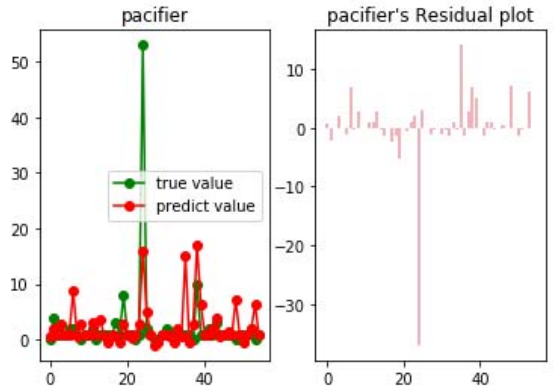

Figure 3: Comparison figure for the pacifier

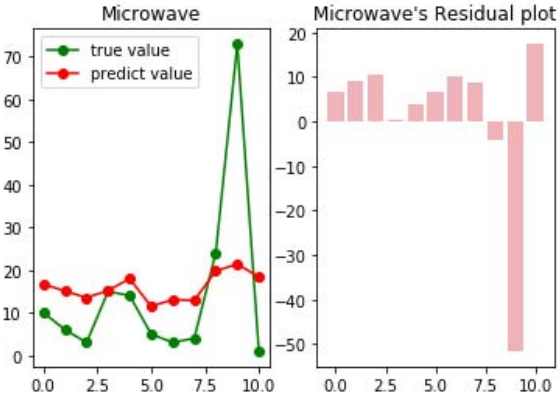

Figure 4: Comparison figure for the microwave

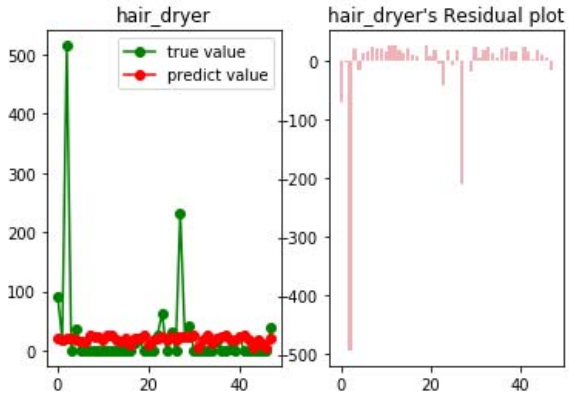

Figure 5: Comparison figure for the hair dryer

Besides, we also give the accruracy rate of the comparison between the real number and the predicted number in table 1. It is clear that the network works well with the high accuracy.

Table 1: The accuracy of the prediction

Hairdryer	Microwave	Pacifier
0.764	0.896	0.953

V. MODEL2:GRAY PREDICTION MODEL

A. *Gray Model Vs Time Serie*

The gray model is a differential equation model that uses discrete random numbers to generate significantly reduced randomness and generates regular numbers. The formula is set as:

$$\alpha^1 x^0(K) + a_1 x^0(K) + a_2 z^1(K) = b \tag{6}$$

$\alpha^1 x^0$ is the $1-IAGO$ array, x^0 is $1-AGO$ array.

A time series is a sequence of consecutive observations of the same phenomenon at different times, t is the observed time and Y is the observed value. Then $Y_i(i=1,2,\cdots,n)$ is the observed value at time t_i .The components of the time series can be divided into 4 types, namely trend(T), seasonality(S), periodicity(C), randomness(I). Its manifestation is:

$$Y_t = T_t * S_t * c_t * i_t \tag{7}$$

B. *Result Comparison and Prediction*

The figure 6 shows that the data is not stable even after the three order difference operator, so we cannot use time series to make predictions.

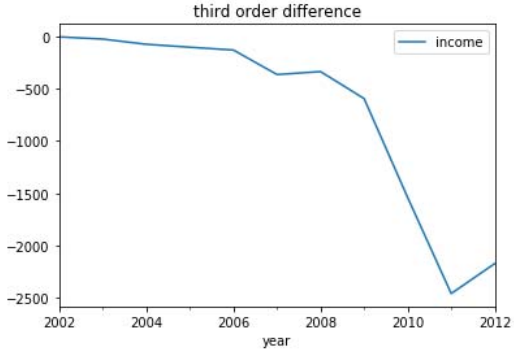

Figure 6: Three order difference of time series

Try the GM(2,1) to do the predictions, the results are as follows (table 2):

Table 2: Rating on the comment validity and comment credibility

	2016	2017	2018
hair dryer	4153.604237	6067.962247	8864.630265
microwave	569.4811761	974.7796404	1668.528105
pacifier	6224.474295	9513.032293	14539.02436

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 19:25:46 UTC from IEEE Xplore. Restrictions apply.

What matters is that three products' reputation is increasing in the online marketplace over the following years.

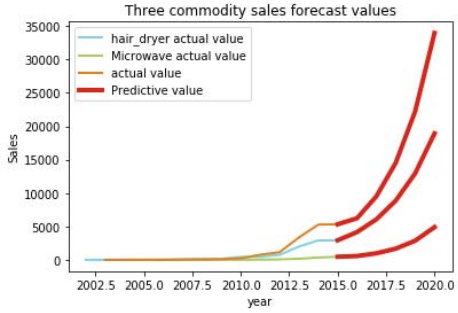

Figure 7: Three products' predicted reputation by GM(2,1)

Specifically, top three sales brands for each product are also predicted. These three brands will undergo a fourishing trend in the following years.

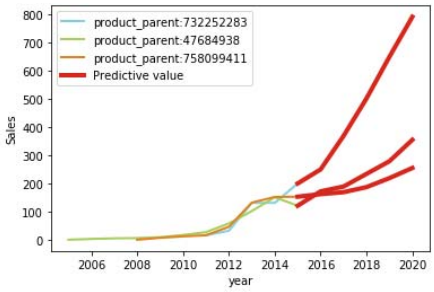

Figure 8: Three hair dryer commodity of sales forecast values

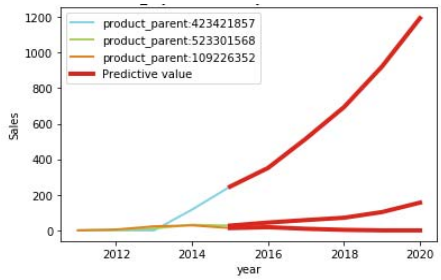

Figure 9: Three microwave commodity of sales forecast values

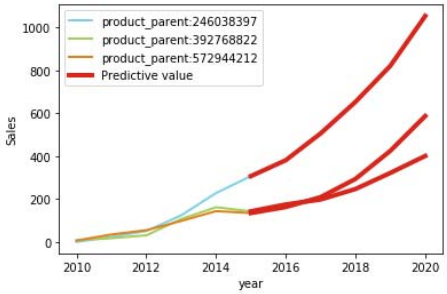

Figure 10: Three pacifier commodity of sales forecast values

Three hair dryer commodity, microwave commodity, pacifier commodity are increasing over the following 5 years, as well.

VI. Model3:Markov Chain and Bayesian Network Model

A. Markov Chain and Bayesian Network

Markov Chain describes the random transfer process with discrete time and state. The state of the next period depends only on the state and transition probability of the current period [22]. It is known that the present and future have nothing to do with the past (no aftereffect).

The state is set as:

$$X_n = 1,2,\cdots,k \quad (n = 1,2,\cdots) \tag{8}$$

The state probability is set as:

$$a_i^{\,n} = P(X_n = i) \quad i = 1,2,\cdots,k \quad n = 1,2,\cdots \tag{9}$$

The state transfer probability is:

$$P_{ij} = P(X_{n+1} = j \,|\, X_n = i) \tag{10}$$

Basic equation can be listed as:

$$a_i(n+1) = \sum_{j=1}^{k} a_j(n) P_{ji} \quad i = 1,2,\cdots,k \tag{11}$$

Specifically, we count the according number of rated reviews and star ratings. The data of hair dryer from year 2002 to year 2015 is separated into three catagories: rated review is 0 and star ratings are 1 or 2; rated review is 0.5 and star ratings are 3; rated review is 1 and star ratings are 4 or 5. The transfer probability matrix can be made as:

Table 3: Transfer probability matrix

	0_1,2	0.5_3	1_4,5
0_1,2	P_{11}	P_{12}	P_{13}
0.5_3	P_{21}	P_{22}	P_{23}
1_4,5	P_{31}	P_{32}	P_{33}

The random variables involved in a research system are independently plotted in a directed graph according to whether or not the conditions form a Bayesian network. It is mainly used to describe the conditional dependencies between random variables. Random variables are represented by circles, and conditional dependencies are represented by arrows. For any random variable, the joint probability can be obtained by multiplying the respective local conditional probability distributions:

$$P(X_1,\cdots,X_k) = P(X_k \,|\, X_1,\cdots,X_{k-1}) \cdots P(X_2 \,|\, X_1) P(X_1) \tag{12}$$

B. Results and Analysis

We choose the top sale brand in three products to give the pre probability.

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 19:25:46 UTC from IEEE Xplore. Restrictions apply.

Table 4:Pre probability for products in five star ratings and three reviews

	1 star	2 star	3 star	4 star	5 star	0	0.5	1
studio salon collection pearl hair dryer	0.09	0.04	0.10	0.18	0.59	0.16	0.09	0.75
danby 0.7 cu.ft. countertop microwave	0.12	0.05	0.12	0.29	0.43	0.20	0.12	0.68
philips avent bpa free soothie pacifier	0.05	0.04	0.06	0.12	0.74	0.05	0.16	0.79

Bayesian undirected graph for three products are listed below.

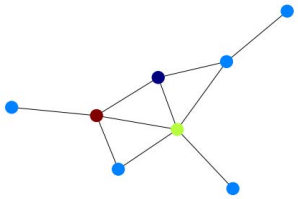

Figure 11: Bayesian undirected graph for hair dryer

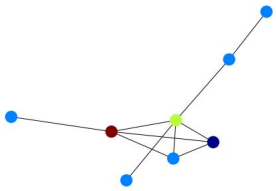

Figure 12: Bayesian undirected graph for pacifier

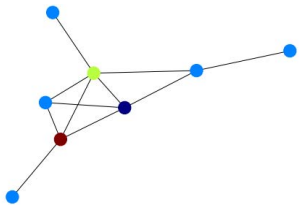

Figure 13: Bayesian undirected graph for microwave

The cause relationship is also achieved as follows:

Table 5: Cause relationship

	1 star	2 star	3 star	4 star	5 star	0	0.5	1
studio salon collection pearl hair dryer	0	1 star	2 star	0.5	4 star	3 star	1	5 star
danby 0.7 cu.ft. countertop microwave	2 star	0.5	3 star	1	5 star	4 star	0	1 star
philips avent bpa free soothie pacifier	1	0.5	3 star	2 star	1 star	0	4 star	5 star

We choose star ratings and reviews which are not only location-close but also logical alike. For the hair dryer, the negative review leads to 1 star rating. For the microwave, the positive review leads to 5 star ratings. For the pacifier, constant reducing star ratings result in the extremely negative review.

VII. Model 4: Correlation Coefficient Model

A. Pearson correlation coefficient Vs Gini Factor

Pearson correlation coefficient is to examine the degree of correlation between two variables. The larger the absolute value of the correlation coefficient, the stronger the correlation [21]. Assuming there are two variables X, K, then the Pearson correlation coefficient between the two variables can be calculated by the following formula:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2)E^2(X)}\sqrt{E(Y^2)E^2(Y)}} \tag{13}$$

At the same time,we can apply the gini factor. We assume that there are *K* classes, the probability that the sample belongs to the *K* class is p_k. Thus, the gini factor can be defined as:

$$Gini(p) = \sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k=1}^{K} {p_k}^2 \tag{14}$$

For the 2-class classification, if the probability that sample belongs to 1 class is *p,* then the gini factor is:

$$Gini(p) = 2p(1-p) \tag{15}$$

For the specific sample set D, the gini factor is:

$$Gini(p) = 1 - \sum_{k=1}^{K} \left(\frac{|C_k|}{|D|} \right)^2 \tag{16}$$

Where C_k is the subset which belongs to *k* class in *D, K* is the number of sets.

B. Result Analysis

As we extract the rated indicators from text-based reviews in Model 1, we further discuss its association with rating levels by the Pearson correlation coefficient and the gini factor (table 6).

Table 6: Correlation coefficient of Pearson and gini

	person	Gini
hair_dryer	0.238	0.668
Microwave	0.4141	0.820
pacifier	0.2259	0.615

Microwave ovens are the most closely related, indicating that microwave ovens have the most functions and belong to long-term supplies, so people will give more specific quality descriptions when rating stars. However, pacifiers are short-term supplies and need to be replaced after a period of time, so even if people rate stars, there is less description of their

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 19:25:46 UTC from IEEE Xplore. Restrictions apply.

specific quality. The pearson coefficient is consistent with the GINI coefficient method, which shows that the results are true and valid.

VIII. Strengths and Weaknessese

A. Strength

Use the IF-IDF method to convert the text information of the review into digital information for easy processing.

B. Weakness

1) For brands with few reviews, there is not much data analysis.

2) When using the IF-IDF method to convert the text information of a review into a number, human calibration is required.

References

[1] Xing Fang and Justin Zhan. Sentiment analysis using product review data. *Journal of Big Data.* 2(1):5, 2015.

[2] Kelley A OReilly, Amy MacMillan, Alhassan G Mumuni, and Karen M Lancendorfer. Factors affecting consumers online product review use. *Qualitative Market Research: An International Journal,* 2018.

[3] Meng-Xiang Li, Chuan-Hoo Tan, Kwok-Kee Wei, and Kan-Liang Wang. Sequentiality of product review information provision: An information foraging perspective. *Mis Quarterly*. 41(3):867-892, 2017.

[4] Jared Watson, Anastasiya Pocheptsova Ghosh, and Michael Trusov. Swayed by the numbers:the consequences of displaying product review attributes. *Journal of Marketing,* 82(6):109-131,2018.

[5] Danny Weathers, Scott D Swain, and Varun Grover. Can online product reviews be more helpful? examining characteristics of information content by product type. *Decision Support Systems,* 79:12-23, 2015.

[6] Kihan Kim, Yunjae Cheong, and Hyuksoo Kim. User-generated product reviews on the internet: the drivers and outcomes of the perceived usefulness of product reviews. *International Journal of Advertising.* 36(2):227-245, 2017.

[7] Jingjing Liu, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL),* pages 334-342, 2007.

[8] Yubo Chen and Jinhong Xie. Third-party product review and firm marketing strategy. *Marketing Science,* 24(2):218-240, 2005.

[9] Mengxiang Li, Liqiang Huang, Chuan-Hoo Tan, and Kwok-Kee Wei. Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce,* 17(4):101-136, 2013.

[10] Yubo Chen and Jinhong Xie. Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management science, 54(3):477-491, 2008.

[11] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168-177, 2004.

[12] Monic Sun. How does the variance of product ratings matter? Management Science,58(4):696-707,2012

[13] Yuren Wang, Xin Lu, and Yuejin Tan. Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. Electronic Commerce Research and Applications, 29:1-11, 2018.

[14] Silvia Angilella, Salvatore Corrente, Salvatore Greco, and Roman Siowiriski. Musa-int: Multicriteria customer satisfaction analysis with interacting criteria. Omega, 42(1):189-200,2014.

[15] Ron S Kenett and Silvia Salini. Modern analysis of customer satisfaction surveys: comparison of models and integrated analysis. Applied Stochastic Models in Business and Industry, 27(5):465-475, 2011.

[16] Tobias H Engler, Patrick Winter, and Michael Schulz. Understanding online product ratings: A customer satisfaction model. Journal of Retailing and Consumer Services, 27:113-120,2015.

[17] Jing Luan, Zhong Yao, FuTao Zhao, and Hao Liu. Search product and experience product online reviews: an eye-tracking study on consumers9 review search behavior. Computers in Human Behavior, 65:420-430, 2016.

[18] Hans Christian, Mikhael Pramodana Agus, and Derwin Suhartono. Single document automatic text summarization using term frequency-inverse document frequency (tf-idf). ComTech: Computer, Mathematics and Engineering Applications, 7(4):285-294, 2016.

[19] Song-yun XIE, Da-qun DONG, and Ben-gang WANG. A new approach to target recognition based on gray correlation analysis [j]. Acta Simulata Systematica Sinica, 2, 2002.

[20] Sheng Chen, ES Chng, and K Alkadhimi. Regularized orthogonal least squares algorithm for constructing radial basis function networks. *International Journal of Control.* 64(5):829-837, 1996.

[21] Per Ahlgren, Bo Jarneving, and Ronald Rousseau. Requirements for a cocitation similarity measure, with special reference to pearson[9] s correlation coefficient. *Journal of the American Society for Information Science and Technology.* 54(6):550-560, 2003.

Authorized licensed use limited to: UEL - Universidade Estadual de Londrina. Downloaded on November 06,2025 at 19:25:46 UTC from IEEE Xplore. Restrictions apply.