Design Engineering

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

A Research on Visualization Method of Middle-aged Women's Wear User Demand Data Based on t-SNE

Jun Yao, Jianhui Chen*

Fashion Institute, Donghua University, Shanghai 200051, China

*Corresponding Author: Jianhui Chen

Abstract:

The high-dimensional, fuzzy and sparse characteristics of clothing users often bring inconvenience to designers. With this regard, this paper will attempt to improve the rationality and accuracy of the analysis method of the clothing users' demand by applying computers technology of t-distribution stochastic neighbor embedding (t-SNE) in order to solve the problem of analysis and processing of high-dimensional data by reducing the dimension of data. Starting from the middle-aged women's wearing users in China, this paper uses t-SNE to map the complex and massive demand information relationship among users to 3D space, and combines the visualization method to show the closeness and relevance of the demand information. The correlation between the needs of middle-aged women's wear users in China and their age, disposable income, purchase frequency is also studied in this research together with the evaluation of the relevance so as to verify the effectiveness of the method. This paper proposes a visualization method to analyze the needs of clothing users, getting rid of the reliance on "experience ", from human subjective judgment to machine recognition, and provides a reference for the research of perfecting the needs of middle-aged women's wear users.

***Keywords**: t-SNE, Middle-aged women's wear, User demand, Data visualization, Relevance.*

I. INTRODUCTION

In recent years, the rapid development of artificial intelligence, machine learning, deep learning and other technologies in the field of computer [1] has enabled machines to assist designers' complex work, so as to change mode of judging and guessing based solely on subjectivity, experience, blindness [2] and also to offer help to obtain user needs. At present, the clothing field advocates the user-centered [3] designing concept, and takes the user demand as the starting point of product design [4]. However, one of the important problem at this stage is how to analyze and deal with massive, fuzzy and complex user demand data. This paper will apply current advanced computer technology t-SNE (t-distribution stochastic neighbor embedding, t-SNE) [5] to solve the problem of high dimensional data analysis and processing by reducing the dimension of data, in order to enhance the rationality and accuracy of clothing user demand analysis method and get more accurate analysis and evaluation of middle-aged women's clothing user demand.

II. RESEARCH OVERVIEW

2.1 Middle-aged Women's Clothing Demand

The World Health Organization divides 45-59 years old into middle-aged people [6], and middle-aged women play an important role in social life, such as daughters, mothers, wives and so on, due to their unique physical conditions, social status, economic strength, aesthetic preference and so on. At present, the research on middle-aged women's wear is mainly in the field of clothing fit, color pattern design and intelligent clothing design, but there is less research on the demand of users of this age group. Jiang Tingting[7]used a joint analysis method to obtain the preference degree of middle-aged and elderly users for each attribute and attribute level affecting the purchase of coats in spring and autumn, and measured the relative importance of each element of clothing.

According to the classification of clothing user needs, Zhao Ping and his partners [8] divided clothing needs into basic needs (including cold protection, heat prevention, protection of the body), advanced needs (that is, social, psychological, emotional, beautiful, etc.). Cui Jian et al. [9] believes that the user's clothing demand information dimension mainly includes clothing function, clothing performance and clothing customization. Based on the above theory, this paper summarizes the needs of middle-aged women's wear users in TABLE I.

Design Engineering

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

TABLEI. User Requirements for Middle-aged Women's Wear

Function of Clothing	Aesthetics of Clothing	Quality of Clothing	Brand of Clothing	Epidemic of Clothing
Warm	beauty	material	Brand positioning	trend
Heatstroke Prevention	pattern	craftwork	Brand style	Hot style
Protection of Body	color		Brand awareness	
	type			

2.2 The Acquisition of User Demand

According to the needs of users, many scholars and institutions have carried out a lot of basic researches [10,11] and application exploration [12,13]. The main analytical methods are quality functionexpansion [14], fuzzy comprehensive evaluation method [15], analytic hierarchy process, data mining clustering method and neural network method [16-18].

Document [19-21] provides a general framework for the study of user requirements through the analysis of requirements through QFD (quality function deployment) technology. However, its limitation is that it relies too much on the knowledge structure of domain experts and the calculation of matrix solving is too large when the users have more demand. Shen et al. [22] constructed a set of user resource requirements with different functional attributes by classifying user requirements, and used particle swarm optimization algorithm to mine the association rules user demand sets. Liu Bin and others [23] mine the knowledge of association rules of user's needs through particle swarm optimization algorithm by means of data mining and extension transformation. Then the extension transformation method is used to generate new extension transformation knowledge so that designers can better understand and predict the potential needs of users. Liu and his partners [24]used the multi-linear regression method to establish the association between availability requirements and design elements. Chen Xingyuet al. [25] proposed an ontology-based user requirement representation method, which refers to integrating a complete set of user requirements representation guidance method through natural semantic processing, intelligent machine learning and other artificial intelligence algorithms. Jin

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

Yingleiet al. [26] proposed a user demand factor characterization and extraction method to establish an extensible kano model by Extension element. In previous studies, there are some deficiencies in the acquisition of fuzzy requirements, for the reason that the psychological needs of users usually do not have linear characteristics and it is difficult to characterize the fuzzy needs of users by linear methods based ona normal distribution. To train a neural network model needs to input a large number of samples for self-organization, self-learning, self-adaptation, in order to achieve the appropriate generalization ability, while the general user demand survey is difficult to meet the artificial intelligence algorithm for large sample size requirements.

In the field of clothing user demand research, Cui Jian et al. [9] proposed the structure diagram of user demand information model in PLM garment enterprises. Cheng Guo [27] suggests a personalized clothing recommendation algorithm based on interactive genetic algorithm, which refers to gaining the personalized demand of clothing from Human-Computer Interaction and establishing the user's preference model for clothing by genetic algorithm. Jia Bingbing et al. [28] recommends two kinds of methods based on network to obtain the information of clothing user's demand: One is to acquire user demand through human-computer interaction mode and the other is data mining via user access records. Zhou Haimei[2]applied the human-computer interaction model to the clothing interactive design system, thinking that the user is both the consumer and the clothing designer who has the demand for the clothing perceptual design. The system retrieves and combines the design modules according to such requirements, and then presents the design results to the user.

The difficulty of clothing user demand research is that the user demand is a high-dimensional and nonlinear system composed of multi-dimensional variables, while a single statistic can only reflect the user's needs from a certain aspect and the accuracy rate is low and the result is fuzzy, so it is hard for the designer to judge directly according to it. And previous studies have not solved the complex nonlinear relationship of these requirements well, and cannot handle more complex high-dimensional data, which results in information loss. Therefore, based on t-SNE research data of middle-aged female users, this paper attempts to explore the feasibility of t-SNE technology application in garment user demand analysis.

III. t-SNE PRINCIPLE AND EXPERIMENTAL FLOW

3.1 t-SNE Principle

t-Distribution Stochastic Neighbor Embedding(t-SNE) is nonlinear manifold learning algorithm on the basis of Stochastic Neighbor Embedding(SNE) [29] in order to maintain the

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

consistency between the neighborhood distribution characteristics of high dimensional data and those of low dimensional data. The idea of t-SNE algorithmic is to express the similarity among data points in high dimensional space in the form of conditional probability. Application of t-SNE technology can display the three-dimensional space based on the internal structure of high-dimensional data, so as to reveal the inherent classification characteristics of data and to intuitively express the similarity among data by data visualization. t-SNE technology is easier to optimize than the original method, and it can reduce the problem of data set [30]. Its main algorithm principle is as follows:

The similarity degree used to calculate the data points x_i to x_j is the conditional probability $p_{j/i}$, σ_i is expressed as the Gaussian distribution variance of the data center point x_i. When the distance between points is relatively close, $p_{j/i}$ is relatively high, and when the distance between points is far away, $p_{j/i}$ is close to infinite low. The formula of conditional probability $p_{j/i}$ is:

$$p_{j/i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \quad (1)$$

If the mapping point of the low-dimensional space in the initialized sample is a:

$$\mathrm{a} = \{\mathrm{a}_1, \ \mathrm{a}_2, \ \mathrm{a}_3, \cdots, \ \mathrm{a}_n\} \quad (2)$$

High-dimensional samples use Gaussian probability distribution, while low-dimensional samples use t distribution with degrees of freedom 1. The joint probability density function between high-dimensional space data points and low-dimensional space data points is expressed by p_{ij} and g_{ij} respectively. The formulas are as follows:

$$p_{ij} = \frac{p_{j/i} + p_{i/j}}{2n} \quad (3)$$

$$g_{ij} = \frac{(1 + \|a_i - a_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|a_k - a_i\|^2)^{-1}} \quad (4)$$

SNE use gradient descent to minimize the sum of the Kullback-Leibler divergences of all

ISSN: 0011-9342
Issue: 11 | Pages: 37 - 51

sample data points and mismatch between g_{ij} and p_{ij}, the C is to define cost functions:

$$C = KL = \sum_i \sum_j p_{ij} \log_2 \frac{p_{ij}}{g_{ij}} \tag{5}$$

The gradient formula is:

$$\frac{\partial C}{\partial a_i} = 4\sum_j (p_{ij} - g_{ij})(a_i - a_j)(1 + \|a_i - a_j\|^2)^{-1} \tag{6}$$

The gradient update output is:

$$A^{(t)} = A^{(t-1)} + h\frac{\partial C}{\partial A} + w(t)(A^{(t-1)} - A^{(t-2)}) \tag{7}$$

Here t denotes the number of iterations, h the learning rate, w the momentum factor, and w (t) represents the momentum term of the t iteration. Then iterate repeatedly until the number of iterations is satisfied.

3.2 Experimental Flow

This paper introduces the t-SNE into the data research, and uses the t-SNE algorithm to reduce the dimension of the actual high-dimensional data and project it into the three-dimensional space to realize the data visualization and reveal the correlation between the data intuitively [30]. The main experimental tools include: basic programming language python 3.6.5, t-SNE algorithm by the use of scikit-learn, data visualization by using matplotlib. The program runs in computers of Intel i7-7700K 4.20GHz, memory 32 GB and Windows 10 64bits. The experimental flow is shown in diagram 1.

Design Engineering
ISSN: 0011-9342
Issue: 11 | Pages: 37 - 51

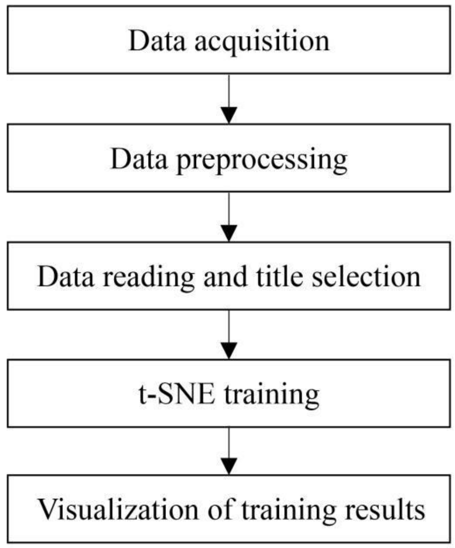

Fig 1:Experimental Flow

The experimental procedures are as follows:

1. Data acquisition. According to the above division of user needs, a questionnaire is set up based on this theory. The age, income and purchase frequency of users in the first three questionnaires were all single-topic topics, and the other six questions were all multi-topic topics related to clothing demand in order to increase the data dimension, with a total of 46 options (In order to better distinguish the age gap, this paper investigates the middle-aged Chinese women aged 40 to 59). A total of 513 valid questionnaires were obtained in the survey.

2. Data preprocessing. Considering that the data stored in the database contains some errors and invalid data, the data preprocessing is carried out before the analysis to ensure the correctness of the data before use and improve the quality of the data decision. Firstly, the data of 513 text result of the questionnaires is discretized, which means the single-topic result of basic information is processed by label encoding while the multi-topic result of clothing demand is by one-hot encoding.

3. Data reading and title selection. Data reading is realized by means of pandas dataframe. After reading data, training data and label are specified. In this experiment, the age, income and purchase frequency of the subjects were labeled respectively, and the training data were the

ISSN: 0011-9342
Issue: 11 | Pages: 37 - 51

results of multi-topic selection involving clothing demand.

4. t-SNE training.The initialization algorithm in t-SNE experiment is pca, iteration of 1000 times, and the confusion perplexity is set to 80. After performing the t-SNE, the model parameters are adjusted repeatedly, and when the parameters are set to 80, the ideal experimental results are obtained.

5. Visualization of training results. By using t-SNE algorithm, the high-dimensional data obtained from the survey is reduced and is projected into three-dimensional space, and the data is visualized to obtain the visual distribution map. The options are numbered 1,2,3...7 in the figure, and the different options are distinguished by different colors, which is beneficial to be reflected in the experimental results. There are 513 mapping points in the experiment, each representing a user under investigation.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Through experiments, the t-SNE is introduced into the study of clothing user's demand, and dimensionality reduction method the t-SNE is used to effectively distinguish the demand differences of different ages, income and purchase frequency. The dimensionality reduction of high-dimensional text and the corresponding low-dimensional mapping points are obtained, so that the low-dimensional manifold embedded in high-dimensional data sets is well characterized. The map points corresponding to text with low spatial similarity are relatively far while the mapping points corresponding to the text with higher similarity are close [30].

4.1 Experimental Results of t-SNE Dimension Reduction with Age Label

Fig 2 shows the distribution map obtained after t-SNE dimension reduction with age label, which indicates that t-SNE users with higher similarity of demand preference in raw data sets become more compact, while users with lower similarity are more distant. Therefore, it can be concluded that if t-SNE does not lose the characteristics of the data itself, the user needs are easier to identify. In Fig 2, the mapping points numbered 1 and 2 are for users aged 40 to 49 and users aged 50 to 59 years are numbered 3 and 4, which shows that the 1 and 2 mapping points are relatively close and are far away from the 3 and 4 mapping points, indicating that the user needs of 40 to 49 years old are similar but are different from those of users over 50 years old. Meanwhile, the mapping points of 1 and 2 are distributed in the upper left and middle of the graph to form two clusters, and there are partially overlapped mapping points, which shows a difference in the needs of users aged 40 to 49 and the similarity of the overlapped mapping points is more compact. The relatively long distance between map points 1 and 4 indicates that users aged 40 to 44 years have low similarity to users aged 55 to 59 years. According to the experimental results, we can find that the demand of middle-aged women's wear users is related to their age: the difference of user demand under 50 years old is not obvious; the difference of user demand between the same age range is different; the greater the difference of age is, the greater the difference of demand becomes, which shows positive correlation.

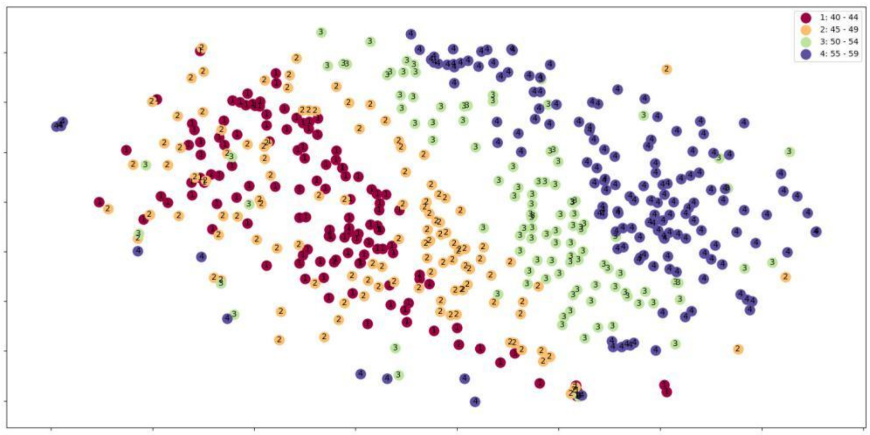

Fig2: Results of t-SNE Dimension Reduction with Age Label

4.2 Results of t-SNE Dimension Reduction Experiment with Income Label

Generally, in Fig 3, the number 5 and 6 which refers to the users with more than RMB 10,000 per month, are usually distributed above and close, and the distance to the users with relatively low income, numbered 1 and 2, is relatively long, indicating that the clothing demand among high income users is similar but it is different from that of low income users. In the Fig 3, number1 and 2 are split into two clusters, among which users with similar needs in the left side gather together to form one cluster, and the middle and lower sides 1 and 2 gather together to form another cluster, indicating that the user needs of the income range of 1,000 to 3,000 yuan are different. At the same time, the distribution distance of 2 and 3 is relatively close, which indicates that the user demand of this income range is similar while the relatively long distances between 2 and 4 indicate a large income gap and a large difference in user demand. In the same income interval in the graph, for example, number 3 is distributed in the left and right clusters and is far away from users 5 and 6, which shows that users in the same income range have

Design Engineering

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

different demand preferences and are different from income users above 10,000 yuan. According to the figures above, we can find the t-SNE experimental results demonstrates that the demand preference of middle-aged women's wear users is correlated with their monthly disposable income and there is also difference in demand in the same income range, and the greater the income gap is, the greater the difference in demand is, which shows positive correlation.

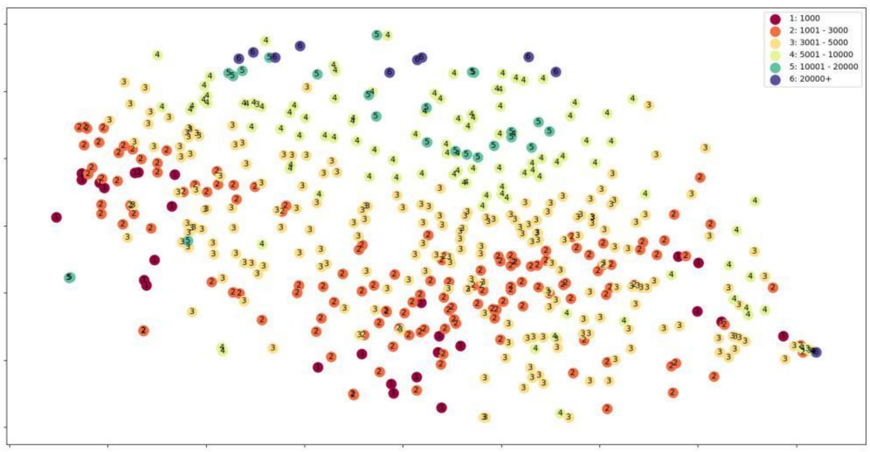

Fig 3: Results of t-SNE Dimension Reduction with Income Label

4.3 Experimental results of t-SNE dimension reduction based on purchase frequency label

In the Fig 4, clothing which is purchased once a week, every month, every three months, every six months, every year, two years and above and on a regular basis, is numbered 1,2,3...7, respectively. From it, we can see that the users who buy clothing on a regular basis on the 7th are obviously far away from other users, forming a separate cluster. The 2 and 3 mapping points become more compact, indicating similar user needs to buy clothing every month and every three months. Map point 5 is relatively far from other map points, indicating that users who buy it once a year differ from other users' needs. Point 4 and 5 are far away from the mapping point 1, indicating that there is a difference between users who buy once a week and those who buy more than half a year. Therefore, the experimental results show the correlation between the demand for middle-aged women's clothing users and their purchase frequency. Demand differs between high purchase frequency and low purchase frequency. The higher the purchase frequency, the greater the demand difference, and the difference in the purchase frequency of

Design Engineering

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

more than every three months is not obvious.

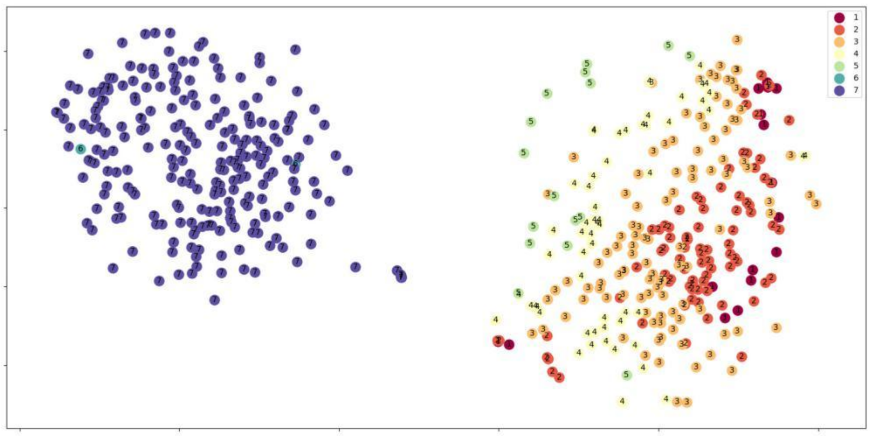

Fig 4: Experimental results of t-SNE dimension reduction based on purchase frequency label

4.4 Relevance Assessment

A new survey of 192 middle-aged female users was conducted with the same questions. Re-testing the experimental method with the income label, the results are shown in Fig 5. From it, the result is similar to that of the original experiment. Map points 1 and 2 are close and relatively far away from map points 4, 5 and 6, separated into two clusters and relatively close to map points 3 while map point 3 is relatively far away from map points 5 and 6. Therefore, this experiment verifies the conclusion that the greater the income gap, the greater the demand gap, and there is also difference in demand in the same income range, which proves the effectiveness of t-SNE technology.

ISSN: 0011-9342
Issue: 11 | Pages: 37 - 51

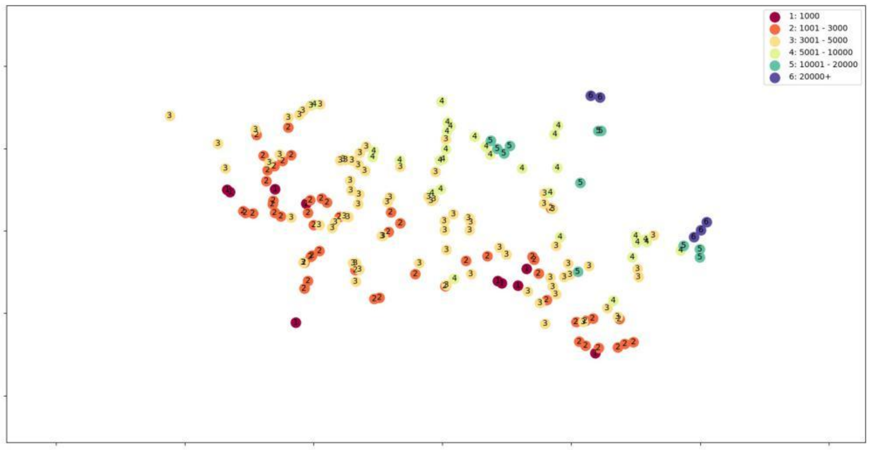

Fig 5: Results of revenue-labeled validation

Based on the t-SNE user requirements, on the basis of similarity measurement of the higher dimension Euclidean distance, the above information is reflected in the concept of distance in the process of requirement analysis, so the more easily identifiable results can be obtained. t-SNE, which is based on user needs, does not contain "empirical" evaluation assignments, basically getting rid of the dependence on "experience " [30], from human subjective judgement to machine recognition.

V. CONCLUSION

This paper aims at the problem of high dimensional, complex and massive demand of clothing users from the point of view of clothing user's demand, proposes a data visualization method based on t-SNE, and summarizes the correlation between the needs of middle-aged women's wear users and their age, disposable income and purchase frequency. t-SNE technology does not deliberately highlight a user parameter values on the demand results in the user requirements classification process and each variable is analyzed according to the characteristics of the user's overall needs, in order to maximum possible preservation of all information of the data while visualizing the data.

This paper has certain reference value for the study of the needs of middle-aged women's wear users in China, but due to the limitations of the research sample in this study, the specific

ISSN: 0011-9342
Issue: 11 | Pages: 37 - 51

analysis of user needs will be further summarized in the follow-up research to meet the needs of practical design.

ACKNOWLEDGMENT

This work was financially supported by the Fundamental Research Funds for the Design Category IV Peak discipline construction, fashion Painting Art Master Studio in Shanghai. Project number: DC17007. Project leader: Chen Jianhui.

REFERENCES

[1] Hinton, Geoffrey E., Ruslan R. Salakhutdinov (2006) Reducing the dimensionality of data with neural networks.Science 313. 5786:504-507

[2] Zhou Haimei (2015) Interactive Design System Based on Modularized Garment Style and Perceptual Demands. Soochow University, Su Zhou, March

[3] Saffer D (2010) Designing for interaction: creating innovative applications and devices. Berkeley, CA: New Riders. ISBN-10: 0321643399

[4] Aoussat A, Christofol H, Le C M (2000) The new product design: atransverse approach. Journal of Engineering Design 11(4): 399-417

[5] Hinton G, Roweis S (2003) Stochastic neighbor embed—ding. Advances in Neural Information Processing Systems15: 833—840

[6] World Health Organization (2012) "World Health Day 2012: ageing and health: toolkit for event organizers". World Health Organization. Available at https://apps.who.int/iris/handle/10665/70840. Accessed 2 November 2019

[7] Jiang Tingting, Xu Yaping (2018) Kansei cognitive and consumption preference of elderly women's coats. Wool Textile Journal 46(04): 27-30

[8]Zhao Ping, Lu Yihua, Jiang Yuqiu (2004) Introduction to clothing psychology. Beijing: China Textile Press. ISBN9787506430111

[9] Cui Jian, Cai Min, Ding Xianghai (2010) Information model of PLM customer requirement on costume enterprise. Journal of Textile Research 31(04): 132-138

ISSN: 0011-9342

Design Engineering

Issue: 11 | Pages: 37 - 51

[10] Kano K H, et al (1984) How to delight your customers. Journal ofProduct and Brand Management5(2):6-17

[11] Abraham H. Maslow (2012) Motivation and Personality. Beijing: China Renmin University Press. ISBN9787300158655

[12] Yuan Chang-feng, Pang Gui-bing (2012) Approach of customer requirement analysis based on requirement element and improved Ho Q in product configuration design. Journal of Software7(3): 691-698

[13] Luo Tian-hong, Ran Xian-sheng (2012) Customer requirements driving NPD method based on integrated CAX technology.Advanced Materials Research 479:1728-1732

[14] Sivasamy K, Arumugam C, Devadasan SR (2016) Advanced models of quality function deployment: a literature review. Quality and Quantity 50(3): 1399-1414

[15] Lou Jianren, Zhang Shuyou, Tan Jianrong (2004) Research on Expressing and Processing Client Demands for Mass Customization. China Mechanical Engineering (08):29-31

[16] Chen C H, Khoo L P, Yan W (2002) A strategy acquiring customer requirement patterns using laddering technique and ART2 neural network. Advanced Engineering Informatics 16(3): 229-240

[17] Shieh Meng-Dar, Yan Wei, Chen Chun-Hsien (2008) Soliciting customer requirements for product redesign based on picture sorts and ART2 neural network. Expert Systems with Applications 34(1):194-204

[18] Liu YX (2010) Study on application of apriori algorithm in data mining. In: Proc. of the 2010 2nd Int'l Conf. on Computer Modeling and Simulation. Taiyuan: IEEE Computer Society: 111-114

[19] Harding J A, Popplewell K, Fung R, et al (2001) An intelligent information framework relating customer requirements and product characteristics. Computers in Industry44(1):51-65

[20] Hong L, Wang W, Zhao HQ (2012) A service quality management approach based on QFD.In: Proc. of the Service Systems and Service Management (ICSSSM).West Sussex: John Wiley and Sons Ltd 11-14

[21] Chen CL, Shi WH, Yan HG (2012) Equipment quality monitoring analysis based on QFD.In: Proc. of the Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQ2MSE). Beijing: IEEE Computer Society: 1329-1332

[22] Shen YG, Liu J, Shen J (2010) The further development of wea base on positive and negative association rules.In: Proc. of the Intelligent Computation Technology and Automation (ICICTA 2010). Beijing: IEEE Computer Society: 811-814

ISSN: 0011-9342

Issue: 11 | Pages: 37 - 51

[23] Liu Bin, Zhu Ming, Wang Jinghua et al. (2011) Research on customer requirement capturing based on extension data mining. Journal of Hefei University of Technology 34(12):1823-1826

[24] Liu Yuanyuan, Zhou Jian, Chen Yizeng (2014) Using fuzzy non-linear regression to identify the degree of compensation among customer requirements in QFD. Neuro computing 142:115-124

[25] Chen Xingyu, Huang Junwen, Zhou Zhan, et al (2017) Ontology-based user requirements representation in the context of big data. Journal of Shenzhen University Science and Engineering 34(02):173-180

[26] Jin Yinglei, Lu Jian, et al (2017) A method of user demand characterization and extraction based on kano model. Modular Machine Tool and Automatic Manufacturing Technique (07): 22-26+31

[27] Chen Guo (2010) Research and application on style decision making model of personalized garment pattern system. DongHua University, ShangHai, March

[28] Jia Bingbing, Qiu Jianxin, Xu Jian et al. (2013) Application of QFD technique to analysis of garment customer requirements. Journal of Shanghai University of Engineering Science 27(03):256-260

[29] Hinton G E, Roweis S (2002) Stochastic neighbor embedding‖ Advances in Neural Information Processing Systems. USA: The MIT Press

[30] Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. Journal of Machine Learning Research 9 (2605): 2579-2605