

Methodology on Demand Extraction Method for Clothing Users Facing Online Chinese Comments

Jun Yao, Jianhui Chen*

Fashion Institute, Donghua University, Shanghai 200051, China

*Corresponding Author.

Abstract:

This paper applies a cutting - edge method to extract clothing user needs so as to transform the user's subjective needs into specific design requirements through obtained demand features. In view of the characteristics of massive, fuzzy and disorderly comments from online clothing customers, this method, taking the middle - aged women coat as an example, tries to use the relevant methods of natural language processing like the concept analysis, taxonomic relation and non-taxonomic relation to come up with the user demand characteristics of middle - aged women's coat, and realize the information extraction of clothing users for online comments. The effectiveness of the experimental results has been verified. This paper has also provided reference list to improve the clothing user demand extraction method.

Keywords: *Natural language processing, online reviews, clothing user demand, dependency syntax analysis.*

I. INTRODUCTION

In recent years, with the rapid development of Internet shopping, the network evaluation system formed by user online product reviews can directly reflects the rich potential demand of users[1]. However, the irregular use of language, the constantly changing meaning of words, massive information extraction and other problems among user comments have become more and more prominent in the field of demand acquisition of clothing users. In particular, users' comments on clothing products are relatively more subjective, more complex and fuzzier than on other products, which brings inconvenience to the extraction of users' needs.

In existing research, the research results in the field of user demand lay a theoretical foundation and reference framework for exploring the needs of garment users. Document[2-4] provides a general framework for studying the needs of clothing users through needs analysis

by QFD technology. Liu Bin et al.[5] combined data mining and extension transformation to mine the knowledge of association rules of consumer requirements through particle swarm optimization and used the extension transform method to generate new extension transform knowledge, so that designers can better understand and predict the potential needs of buyers. Tu Haili[6] constructed a requirement mining model for online comment data. Yin Pei et al.[1] proposed a method of product feature level emotion classification based on the feature view of domain ontology recognition review. Chen Xingyu et al.[7] used ontology to establish abstract semantic relationship between user requirements and product attributes, and then suggested an ontology-based representation of user requirements by which the initial data generation, requirement ontology generation and requirement representation library generation are integrated into a set of scientific guidance methods for user requirement representation. Jin Yinglei[8] studied the difficulty in recognition by computer in the extraction of user needs in the early stage of product design and in the disjoint with digital design and manufacturing technology, then analyzed the emotional dimension attribute of demand factors and finally proposed a theory of Extension element by establishing the user demand factor characterization and extraction method of extension kano model.

In clothing user demand field, Song Xin et al.[9] considered that the weight of garment product engineering characteristics reflects the importance of the project in realizing user demand and can be determined according to the weight of user demand and the relationship between user demand and engineering characteristics of garment product. Cui Jian et al.[10] proposed the clothing enterprise PLM user demand information model structure diagram via the combination of the user's own information, clothing function demand information, clothing performance demand information and clothing customization demand information. Cheng Guo et al.[11] proposed personalized clothing recommendation algorithm based on an interactive genetic algorithm, from the human-computer interaction to obtain the user's personalized demand for clothing, and through genetic algorithm to establish a user's preference model for clothing. Jia Bingbing et al.[12] studied two kinds of methods based on network to obtain consumer's information demand and found that one method is to obtain their information demand through human-computer interaction mode whose advantage is that it can obtain demand information more accurately but whose disadvantage is its poor initiative, while the other method is to obtain user's basic needs, style habits and personal preferences through data mining for clothing users.

There are the following three main problems in the previous studies:

- The existing research on the needs of clothing users is not deep enough, so the feasibility of user demand extraction technology in the field of clothing needs to be discussed, and

further optimization methods and new means should be discussed.

- For the features of large quantity, high dimension and disorder, the accuracy of traditional methods to acquire demand information is low and very limited, making it difficult to obtain a large amount of information and to ensure accuracy and completeness.
- The demand of clothing users reflects the fuzzy psychological needs of the users, and most of the previous studies rely on the subjective judgment of designers and experts, which means it is difficult to ensure the consistency of the users' real needs.

This study will combine the relevant methods in the fields of Natural Language Processing, machine learning, ontology and so on and presents an effective method to extract the needs of clothing users for online reviews of internet users based on the literature[7]. This method can improve the rationality and accuracy of clothing user demand acquisition and realize more efficient analysis of extraction of information, so as to transform the requirement feature into specific design requirements to guide the design work. The specific steps are shown in Fig 1.

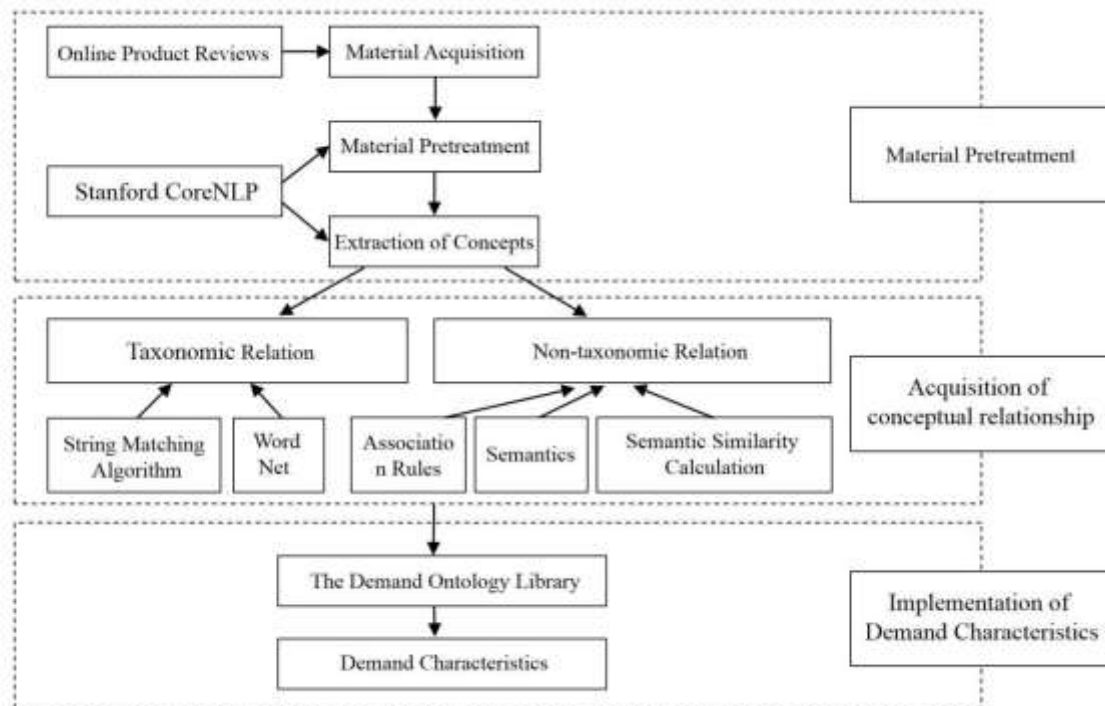


Fig 1: implementation steps of user demand characteristics

II. MATERIAL PRETREATMENT AND ACQUISITION OF CONCEPTS

2.1 Material Acquisition

The original data of this study came from the highest - selling coat products in the Tmall flagship store of a well - known middle - aged women's clothing brand in China, with a total of 5,099 effective online reviews, including favorable and negative comments. The objective of this study is mainly to analyze Chinese language network reviews, excluding the description of coats and contents of other languages by extracting and filtering the bytes in these original comments, and obtaining a corpus set such as TABLE I.

TABLE I. Example of online reviews of middle-aged women's coats

ORDER NUMBER	FAVORABLE COMMENTS	NEGATIVE COMMENTS
COAT 1	The clothes are very comfortable and white color looks good.	The clothes are too thin and pill just in 3 days from they are bought!
COAT 2	The design is reasonable, simple and generous, and you can wear it for a long time.	The color of the clothes is too dark. Putting them on immediately make people ten years older. The color is too dark.
COAT 3	Very good. It is very suitable to wear them in summer.	The fabric is particularly easy to hook.
COAT 4	They are fine workmanship, quite fit, and is genuine.	The sleeve is fat, we feel the sleeve is so big when wearing them.
COAT 5	The color is clear and refreshing, and this is cheaper than the store.	The style is outdated, the quality is ordinary, and the cost performance is relatively low.
COAT 6	I wear this coat for a period of time. It indeed does not pill. I like it very much.	The quality is so bad that it breaks under the armpit twice.
COAT 7	The pattern is very special, and I look better in it.	It is not worth the price, because it is ordinary.
COAT 8	Great style and delicate buttons.	The size is relatively bigger.
COAT 9	The color is very positive, make people look whiter, and the fabric is also not sultry.	The dress is dark red, making people look older.

COAT 10	It is very good. The color is fresh and bright. The kitting is soft and comfortable.	The clothes pill so much, and I am so disappointed.
---------	--	---

2.2 Material Pretreatment

As the clothing product evaluation is fuzzy and disorderly, this paper uses the Chinese participle, word annotation and dependent syntactic analysis using the high accurate language analysis tool Stanford CoreNLP[13], and generates the dependency tree and dependency arc to obtain more accurate user demand information.

Stanford CoreNLP is run in the python for natural language processing, parsing text. Take the following “Zhe jian hong se de wai tao hen piao liang” (this red coat is very beautiful) as an example, and the following parsing results is obtained:

- Chinese word segmentation .Stanford CoreNLP is able to divide the original corpus into separate syntax bytes and come up with [‘zhe’(this), ‘jian’(piece), ‘hongse’(red), ‘de’(of), ‘waitao’(coat), ‘hen’(very), ‘piaoliang’(beautiful), ‘.’].
- Part of speech tagging. The result is: [(‘zhe’(this), ‘DT’), (‘jian’(piece), ‘M’), (‘hongse’(red), ‘JJ’), (‘de’(of), ‘DEG’), (‘waitao’(coat), ‘NN’), (‘hen’(very), ‘AD’), (‘piaoliang’(beautiful), ‘VA’), (‘.’, ‘PU’)]. Among them, DT refers to qualifiers, M measures, JJ name modifiers, DEG possessive cases, NN nouns, adverbs, VA adjectives, PU indicators.
- Compliance with the syntactic analysis.This method is one of the important techniques of natural language processing, and is able to analyze and identify the syntactic structure in a sentence. It is also able to express the dependency relationship between the sentence words in words, and presents the semantic association of the sentence in the dependency structure to obtain the deep semantic information[14]. The operation results are as follows :

(ROOT

(IP

(NP

(IP

(ADVP (DT zhe) (M jian))
 (NP (JJ hongse) (DEG de)))
 (NP (NN waitao)))
 (VP
 (IP (AD hen)(VA piaoliang)))
 (PU .)))

[('ROOT', 0, 7), ('mark:clf', 1, 2), ('case', 3, 4), ('det', 5, 1), ('amod', 5, 3), ('nsubj', 7, 5), ('advmod', 7, 6), ('punct', 7, 8)]

The list of three-element sentences representing dependency is thus obtained. The det refers to determiner, a mark; clf sign; case the subordinate; the amod adjective; the noun subject; the advmod adverbial modifier; the punct punctuation.

- Visualization of analytical results. The result of semantic modification relation is obtained from step 3, and the dependency tree form of linear structure hierarchy is generated from Fig 2, and the dependency arc of dominant and dominant relation is represented in in Fig 3.

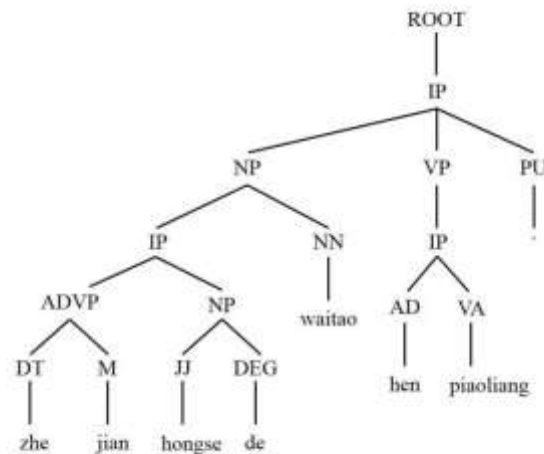


Fig 2: dependent grammar analysis tree

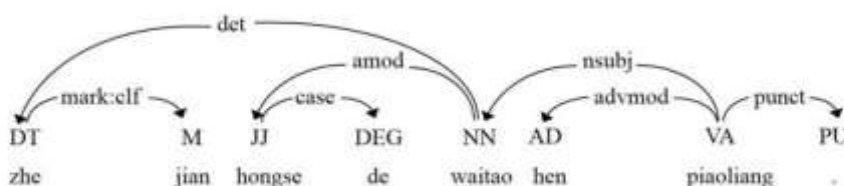


Fig 3: dependency arc

2.3 Extraction of Concepts

According to the results of Chinese word segmentation and part of speech tagging, this paper mainly extracts grammatical labels labeled as nouns and noun phrases according to the characteristics of clothing field, and removes interjections, auxiliary words, qualifiers, possessive cases, adjectives and definite articles, etc.[15] Furthermore, in the light of the relevant knowledge of clothing, synonyms and similar words are eliminated on the basis of high frequency. For example, “Yi fu”(clothes), “Yi shang”(clothes) and “Yi yi”(clothes) have the same meaning, but according to the frequency of words, the relevance of “Yi fu” is higher, but the utilization rate of “Yi shang” is lower, and “Yi yi” is more colloquial. Therefore, the words “Yi shang” and “Yi yi” are eliminated, and the word “Yi fu” are extracted. The concept of coat extracted from TABLE II is obtained.

TABLE II. The extracted concept of coat

CONCEPT	CONCEPT	CONCEPT	CONCEPT
Clothing	Body trim	Pilling	Possibility of fitting on
Style	Patterns	Dark color	Design
Fabric	Fashion	Color difference	Collars
Color	Buttons	Prices	Length
Size	Brands	Matching	Chest measurement
Craftwork	Quality products	Embroidery	Quality
Sleeves	Hooks	Manual embroidery	...

III. IMPLEMENTATION OF USER REQUIREMENTS FEATURES

According to the syntactic analysis results of word set and dependency extracted above, user demand features will be gained by means of obtaining conceptual taxonomic relation and non-taxonomic relation.

3.1 Taxonomic Relation Extraction

Taxonomic relation is also called definition concept, which embodies the upper and lower relation between concepts[16]. By means of string matching algorithm and electronic lexical WordNet, the taxonomic relation is obtained:

- String matching algorithm is a kind of method in which two words having the same head have a taxonomic relation. For instance, “quality” and “poor quality” contain the same head “quality”, so there is a taxonomic relation in concepts between “quality” and “poor quality”.
- The electronic vocabulary database WordNet[17], is to combine the meaning of the concept to form a set of words with the same meaning. For example, there is a taxonomic relation in the word “collar” and “coat” according to the WordNet, indicating “collar” and “coat” have a taxonomic relation. Besides, each polysemous word will be represented in the synonym set of its different meanings, and each word set of the same meaning contains its connection relation and represents a set of semantic conceptual with taxonomic relation[18].

The conceptual taxonomic relation pairs derived from the above two methods are shown in TABLE III.

TABLE III. Taxonomic relation extraction pairs

Quality—Quality of clothes—Quality of coats
Coat—Spring coat—Spring coat with long sleeves
Fabric—Cotton fabric—Cotton comfort
Design—Neckline design—Round neckline design
Design—Neckline design—Round neckline design

Color——Dark color——Black
...

3.2 Non-taxonomic Relation Extraction

According to dependency grammar, the predicate verb is the center of the whole sentence, and there is a directional dominant relationship between words and words, thus forming the dependency relationship[19]. In Fig 4, the semantic information of the sentence and the semantic relationship between words can be derived from this theory. Since the dependency grammar generally consists of the verb connecting two concepts[20], this paper will obtain the non-taxonomic relation between the concepts through the semantic annotation and dependency relationship based on the verb connection and dependence relationship. Specific methods are as follows:

- Based on association rules. According to the above analysis of dependency syntax, the verb-centered subordination and semantic dependency structure information in sentences are obtained[21]. Because dependent syntactic analysis holds that a verb is the central of a sentence, other words are subordinate to this verb. Therefore, this paper extracts the non-taxonomic relation which is centered on verbs. For example, in the sentence “Wai tao dong tian chuan qi lai hen bao nuan.”(Coat is very warm in winter), “chuan” (wear) as a verb is the central word, and the non-taxonomic relation of this sentence should be expressed as “wai tao”(coat), “chuan”(wear), “bao nuan”(warm). In another example: “Wai tao you kou dai.”(Coat has a pocket), is a sentence with “noun-verb-noun” as the structure, and its non-taxonomic relation should be: “wai tao”(coat), “you”(has), “kou dai ”(a pocket).

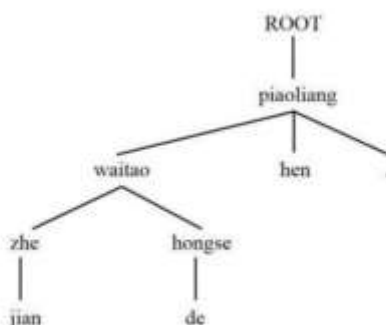


Fig 4: a syntactic analysis of verb-centered dependency

- Based on semantics. Because of the different semantic situations, the semantic segmentation of data set training can't correctly handle the non-taxonomic relation in different contexts. This study uses adaptive semantic analysis method to analyze the context of text context and obtain non-taxonomic relation that can't be processed by regular semantic analysis. For example, in the sentence "Jv hui chuan zhe jian wai tao hen shi shang."(This coat is very fashionable in parties), "jv hui"(party), "wai tao"(coat), "shi shang"(fashionable) will be obtained in normal process while through context analysis, the correct non-taxonomic relation should be: "wai tao"(coat), "hen"(very), "shi shangf"(fashionable).
- Semantic similarity calculation. For the non-taxonomic relation that the above two methods can't handle, this paper will use semantic similarity algorithm to cluster the text. According to the results of verb-centered dependency syntax analysis, this paper will calculate the semantic similarity by extracting the sentence trunk containing the verb framework according to the literature [22], so as to obtain the non-taxonomic relation. The methods are as follows: All verb frames are searched in the results of dependency analysis. The similarity of the two concepts can be drawn a Sim, and the sentence can be displayed as $S = \{F_1, F_2, \dots, F_h\}$, and the verb frame is $F_1 = \{E_{i1}, E_{i2}, E_{i3}\}$ $i \in [1, 2, \dots, h]$, and sentence frame is $E_{ij} = (r_{ij}, f_{ij}, h_{ij})$ $j \in [1, 2, 3]$, in which E_{ij} is the word in a sentence frame, r_{ij} is a dependency, f_{ij} is a part of speech, h_{ij} is a semantic description. The word weight in the sentence is: if the dependency r_{i1} and r_{i2} , part of speech f_{i1} and f_{i2} , semantic description h_{i1} and h_{i2} are all the same, the matching weight is 1; if the r_{i1} and r_{i2} , h_{i1} and h_{i2} are the same, but the f_{i1} and f_{i2} are different, the matching weight is 0.6, vice versa is 0^[22]. The formula is:

$$Sim(F_i, F_j) = \frac{1}{n} \sum_{1 \leq k \leq 3} W_k \quad (1)$$

$\sum_{1 \leq k \leq 3} W_k$ is the sum of the word weight matching, and the n represents the number of words

in the sentence. Suppose the semantic similarity threshold of sentence is 0.57. According to formula (1), when the result is greater than the similarity threshold, it is determined that there is a non-taxonomic relation between the concepts of the semantic role, and the verb in the sentence is the conjunction of the non-taxonomic relation.

From the above method, the forward concept and the subsequent concept are obtained according to the dependency structure, and the conceptual non-taxonomic relation pair are formed. See TABLE IV.

TABLE IV. Non-taxonomic relation extraction pairs

Coat—Have—Buttons	Loosen—Comfortable coat
Long coat—Wear—Warm	Scientific fabric—Practical coat
Bigger—Match—Ugle	Simple style—Easy-to-match coat
Length—Suitable—Spring	Bright color—Fashion coat
Buttom—Prevent—Warmth	Pattern—Funny coat
Color—Look—Younger	Elasticity—Suitable coat
Tighten—Look—Slim	Cotton—Comfortable coat
Style—Feel—Old	Pilling—Terrible coat
...	...

3.3 Implementation of Demand Characteristics

According to the above taxonomic relation and non-taxonomic relation, the demand ontology of middle-aged women's coat is obtained, as is exhibited in Fig 5. The taxonomic relation shows the inclusion relationship between the two concepts, while the non-taxonomic relation reflects the relationship other than the inclusion relationship.

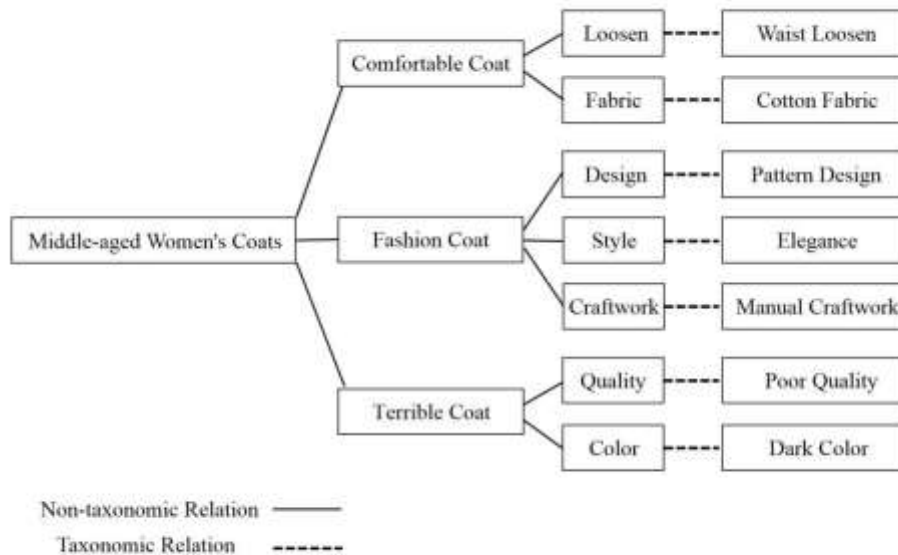


Fig 5: the demand ontology of middle-aged women's coats

Based on the relevant knowledge in the field of clothing and the design requirements of clothing designers, this research will integrate and classify similar user needs, and transform the taxonomic relation and non-taxonomic relation into demand characteristics one by one. The summary of the following middle-aged women coat user needs is demonstrated in TABLE V.

TABLE V. Middled-aged user demand generation

NUMBER	USER DEMAND
1	Coats in cotton is comfortable.
2	Coats with bright color is fashionable.
3	Red coats make wearers look older.
4	Simple design suits all.
5	Long coat keeps warmth.
6	Coats with hook have poor quality.
7	Coat in dark color make wearer look older.
8	Elegant style is stylish.
9	The coat with waistline make wearers thinner.
10	Coats which often pill have poor quality.
...	

Because the design of clothing products is mainly the design activities involving fabric, color, design, craftwork, style and function[23], designers need to classify and summarize the user demand characteristics based on these six sections, and thus obtain the design requirements of middle - aged women, see TABLE VI. Therefore, in the process of designing, it is suggested to focus on the specific user needs of the following six sections.

TABLE VI. Design Requirements for Middle-aged Women's Coats Based on User Needs

FABRIC	COLOR	DESIGN	CRAFTWORK	STYLE	FUNCTION
Cotton	Bright color	Loosen	Leveling needle	Elegance	Warm
Lightweight	No dark color	Not fat	Embroidery	Leisure	Sunlight-resistant

Elastic fabric	Comfortable color	Show body trim	No pilling	Simplicity	One dress with multiple purposes
Soft fabric	Red	Suit all figures	No hook	Lady style	Easy to wear
Comfortable fabric		Simple	Beautiful patterns		Easy to match
Wearing-resistant fabric		Slim waist line			
		Not exposing			
		Beautiful profile			
		Not easily outdated			

IV. ASSESSMENT OF USER NEEDS

The 30 randomly selected user needs are distributed to 50 middle-aged women, who are asked to score the results with a 5-level matrix scale, so as to assess the accuracy of user needs. The answer is divided into 5 ratings from 1 to 5 grade: very accurate, accurate, general, a little inaccurate, completely inaccurate. Then, Ratings are measured by numerical interval values from a very inaccurate 1 increasing to very accurate 5 points. After analyzing the reliability and validity of the findings, if a reliability α coefficient is 0.958, KMO validity was 0.921, it shows that the reliability and validity of the survey are high, which shows the stability and effectiveness of the questionnaire results. Then the average and variance of the survey results are calculated, with mean 4.39, Variance 0.58, which shows the results are stable and accurate, so the extraction results for the middle-aged woman coat user demand feature are reliable.

V. CONCLUSION

This study takes the middle-aged women's coat as an example, and suggests a method of clothing user demand extraction based on natural language processing is proposed in with regards to online user comments. This method can extract the characteristics of user demand from massive text, and transform the relevant knowledge of clothing field into specific design requirements, which provides a reference for the related research of clothing consumer needs. However, in the light of the flexibility of Chinese grammar, Chinese word segmentation, part of speech tagging and dependent syntactic analysis are rather superficial, so the accuracy in relevant researches needs to be improved. Besides, because of the particularity of Chinese

grammar, the meaning of sentences depends on different combination forms, not only on words, but also on the similarity of semantic similarity in the consent dimension which may be different in other dimensions, so the related research needs to be further discussed.

REFERENCES

- [1] Yi P, Wang H (2016) Sentiment classification for Chi-nese online reviews at product feature level through domainontology method. *Journal of Systems & Management* 25(1): 103-114
- [2] Harding JA, Popplewellk K, Fung RYK, et al. (2001) An intelligent information framework relating customer requirements and product characteristics. *Computers in Industry* 51-65
- [3] Hong L, Wang W, Zhao H (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
- [4] Chen C, Shi W, Yan H (2012) Equipment quality monitoring analysis based on QFD. In: *Proc. of the Qualit, Reliability, Risk, Maintenance, and Safety Engineering (ICQ2MSE)*; Beijing: IEEE Computer Society 1329-1332
- [5] Liu B, Zhu M, Wang J, et al. (2011) Research on customer requirement capturingbased on extension data mining. *Journal of Hefei University of Technology* 34(12): 1823-1826
- [6] Tu H, Tang X, Xie L (2015) Research on user needsmining model based on online based on online reviews. *Journal of the China Society for Scientific and Techni-cal Information* 34(10): 1088-1097
- [7] Chen X, Huang J, Zhou Z, et al. (2017) Ontology-based user requirements representation in the context of bigdata. *Journal of Shenzhen University Science and Engineering* 34(2): 173-180
- [8] Jin Y, Lu J, et al. (2017) A Method of User Demand Characterization and Extraction Based on kano Model. *Modular Machine Tool & Automatic Manufacturing Technique* 7: 22-26+31
- [9] Song X, Guo W, et al. (2010) Method of mapping user requirements to technical characteristics in QFD. *Journal of tianjin university* 43(2): 174-180
- [10] Cui J, Cai M, Ding X (2010) Information model of PLM customer requirement on costume enterprise. *Journal of Textile Research* 31(4):132-138
- [11] Chen G (2010) Research and Application on style Decision Making Model of Personalized Garment Pattern System. Dissertation. DongHua University, Shanghai.
- [12] Jia B, Qiu J, Xu J, et al. (2013) Application of QFD Technique to Analysis of Garment Customer Requirements. *Journal of ShangHai University of Engineering Science* 27(3): 256-260
- [13] Manning CD, Surdeanu M, Bauer J, et al. (2014) The Stanford CoreNLP Natural Language Processing Toolkit In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* 55-60
- [14] Ma J (2007) Research on Chinese Dependency Parsing Based on Statistical Methods. Dissertation. Harbin Institute of Technology, Harbin.
- [15] Wu M (2012) A Continuous Fuzzy Kano Model for Customer Requirements Analysis in Product Development. *Journal of Engineering Manufacture* 226(3): 535-546

- [16] Zong C (2013) Statistical Natural Language Processing, 2nd Edition. Publisher: Tsinghua University Press, China.
- [17] George A, Miller (1995) WordNet: A Lexical Database for English. Communications of the ACM 11: 39-41
- [18] Fellbaum C (1998) WordNet: An Electronic Lexical Database. Publisher: MA: MIT Press, Cambridge.
- [19] Du X, Li M, Wang S (2006) A survey on ontology learning research. Journal of Software 17(9): 1837-1847
- [20] Wang S, Zhao A, Ma W (2010) Approach to ex-tracting non-taxonomic relationships for Chinese ontology from web. Computer Engineering and Design 31(2): 451-454
- [21] Wen C, Shi Z, Xin Y (2009) Chinese non-taxonomic relation extraction based on extended association rule. Computer Engineering 35(24): 63-65
- [22] Gu L, Sun S (2012) Approach to chinese ontology non-taxonomic relation extraction based on semantic dependency. Computer Engineering and Design 34(4): 1676-1680
- [23] Liu Y (2010) Fashion design. Publisher: Higher Education Press, China.