Research of a Method for Discovering Sustainable Design Knowledge

Zhang Hui

School of Naval Architecture and Ocean Engineering

Zhejiang Ocean University

Zhoushan，China

zhongzhuan_zz@aliyun.com.cn

Abstract—to discover the associational knowledge between the structure and sustainable design features in sustainable design, a new method, namely, discovering knowledge based on patents, was proposed. It extracts the features of parts in the patent claims, analyzes the objective functions and the sustainable design features in the title or abstract, sets up training sets of parts features-objective functions and objective functions-sustainable design features. To obtain associational knowledge between sustainable design features-objective functions-parts features, a data mining algorithm is adopted to extract the association rules used for aiding product sustainable design. The method is verified by an example in this paper. The experimental result shows that the effectiveness of the method.

Key words- sustainable design; knowledge discovering; data mining; association rule; feature association

I. INTRODUCTION

Sustainable design is an innovative design to protect the environment and save energy. Patents contain many of the latest sustainable design technologies, products and processes, and have great usability. Accessing effectively to and reusing of knowledge of sustainable design can develop quickly sustainable and innovative products. Combining design knowledge in sustainable patents and the functional parts features, data mining technologies are used to acquire association rules between parts features and the objective functions and sustainable design features in patent documents. Then sustainable design strategies of patented products can be discovered.

The method for effectively discovering knowledge has been recognized as an important area at home and abroad. According to the different sources of knowledge, different ways are proposed. ZHEN [1] obtained knowledge maps, instances, rules knowledge based on ontology and TRIZ query from the program base, which to support product innovation design under the knowledge grid environment. LI [2] got the product creative design domain knowledge by extending HowNet and establishing of product semantics base to improve search technology. This method used semantic networks to extend meanings of key words to get the desired result. LIN [3] presented a method to obtain human knowledge about customer characteristics for product innovation through graphs and images and Visual technology, image scaling and oral analysis. But some researchers [4-6] used the lexical, syntactic and semantic analysis method in natural language processing techniques to extract the product feature information in patent documents. Among them, WANG [4] acquired functional knowledge and functional structure knowledge through natural language processing and artificial mining to aid functional and structural innovation. XUE [5] got design principles and conflict knowledge through the patent analysis method based on text mining. Namely, the regular method was used to extract summary and instructions, and collect patent to extract nouns, adverbs, adjectives, and nouns. Then TF-IDF filtering was used to generate structured data. Some keywords, such as recurred and related to the purpose, function, were mined. Last invented the principle was clustered. ZHANG [6] obtained functional knowledge by using data mining algorithms to extract features in the claims of the patent document components, analyze objective and artifice functions in the title or abstract.

In a word, the research about acquiring sustainable design knowledge is less. Discovering sustainable design knowledge from the patent literature is an effective way.

II. DISCOVERING SUSTAINABLE DESIGN KNOWLEDGE FROM PATENTS

Different functions hide the different sustainable design goals. Therefore, the objective functions and sustainable design features, the hidden objective functions and sustainable design knowledge in patents can be discovered by digging up the relationship between parts features and the objective functions.

A. Analysis of Sustainable Design Knowledge in Patents

Sustainable design is a design which needs not only innovation in the functional principle, but also friendly environment. Since the sustainable product patent consists of both innovation design and sustainable design, sustainable design knowledge typically includes a variety of knowledge:

1) Features knowledge of sustainable design: refers to reducing energy consumption, such as energy conservation, water saving, energy saving, fuel-efficient;

2) Objective functions knowledge: including the effects and characteristics of the products. Sometimes it lies in the summary or title of patents [7];

3) Artifice functions knowledge: mainly with the product knowledge of functional structure;

4) Principle effect knowledge: relates to all kinds of energy of sustainable design products. It is located in the abstract, claims and specifications [8];

5) Structural knowledge: including parts and shape knowledge, it is in the abstract, claims and inventing content and specific embodiment [9];

6) Material knowledge: nonpoisonous, harmless, decomposition of the material involved and associated with reusable design knowledge [10];

7) Process knowledge: mainly includes products knowledge such as surface treatment and process;

8) Environmental knowledge: access to products used in the production of environmental hazards.

According to the analysis of sustainable knowledge in patents, the knowledge in the initial patent is shown in Fig.1.
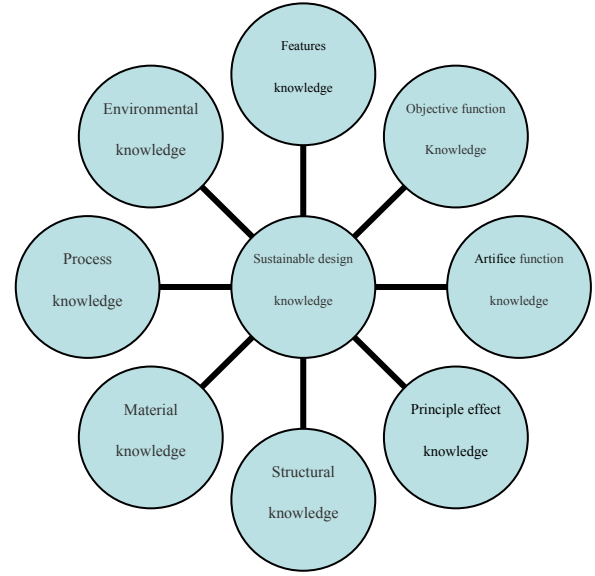

Figure 1. The types of sustainable design knowledge

Patent documents include a variety of different forms of knowledge representation. The knowledge about structure, material, technology, and environment is explicit knowledge that can be expressed and structured. This explicit knowledge can be used for mining sustainable design features, functions, and other knowledge and their relationships. Due to the structure, material, process features are clarity in the claim, an method of acquiring sustainable design knowledge based on patents is proposed.

B. Sustainable Design Knowledge Discovery Based on Association Rules

Such an approach would combine feature dictionary with PCAR algorithm [11] in the data mining. First it preprocesses the patent collection, and then obtains the frequent itemsets which are composed by objective functions-parts features, sustainable design features-objective functions, acquires association rules of objective functions-parts features, sustainable design features- objective functions. Finally, identifies associational knowledge and sustainable design strategies. The process is as shown in Fig.2.
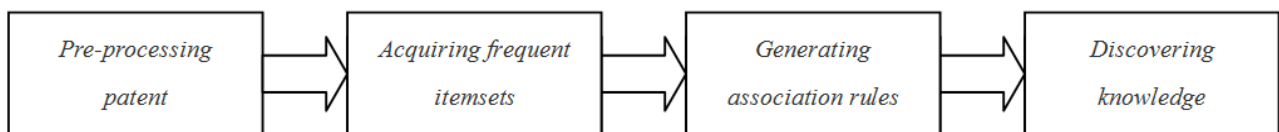

Figure 2. Flow chart of mining sustainable design knowledge

TABLE 1 SYMBOL TABLE OF SUSTAINABLE DESIGN FEATURES-FUNCTIONS-PARTS FEATURES

Patent Number	S	F	Parts Features		
			K	M	T
201320004969.7	S_1	$F_1F_2F_7F_8$	K_1K_2		
201210590660.0	S_1	$F_2F_{20}F_{21}F_{22}F_{24}F_{25}$	K_4	M_4	
201220651740.8	S_1	$F_2F_9F_{18}$	K_3	$M_1M_2M_3$	
201120221462.8	S_2	$F_{15}F_{16}F_{23}$		M_5M_6	
201310039117.6	S_2S_8	$F_{10}F_{15}$			T_1
201110278168.5	S_2S_9	F_3F_{25}		M_7	
200910219665.0	S_3	$F_{11}F_{26}$	K_6		
201110310667.8	S_3	$F_{11}F_{12}$	K_5		
200980154424.0	S_3	$F_{11}F_{26}$		M_8	
200610024412.4	S_4	$F_{12}F_{14}F_{19}F_{24}F_{26}$	K_7		
200410017967.7	S_4	$F_2F_{10}F_{12}F_{13}F_{14}F_{19}$	K_7		
200410067793.5	S_4	$F_2F_{10}F_{12}F_{13}F_{14}F_{19}$	K_8		
200610112601.7	S_5	$F_{17}F_{18}F_{19}$	K_{10}	M_{12}	T_3T_4
200920261179.0	S_5	$F_4F_{10}F_{15}$	K_9	M_9	T_1
200780008738.0	S_5	F_4F_{18}	K_7	$M_{10}M_{11}$	T_2
201220657923.0	S_6	F_4F_{25}	K_{16}		
201220738086.4	S_6	$F_{10}F_{15}F_{19}$	K_{11}		
201310047214.x	S_6	$F_5F_{10}F_{24}$	K_5K_7		
201210485262.2	$S_6 S_7$	$F_1F_6F_{19}F_{24}F_{25}$	$K_{12}K_{15}$		
201120526757.6	S_7	$F_{11}F_{25}$	$K_{13}K_{14}$		

Steps：

1) Pre-processing patent: sustainable products patents are divided into training set and test set. Objective functions and sustainable design features are analyzed in the training set. Some features, such as structure, material and process in protection parts are extracted. The data set of objective functions-parts features and sustainable design features- objective functions is established.

2) Acquiring frequent itemsets: PCAR algorithm is used to access all frequent itemsets in the dataset based on minimum support.

3) Generating association rules: under minimal support, association rules that meet the criteria are extracted.

4) Discovering knowledge: relationships knowledge among parts, sustainable design features and objective functions is discovered by obtained association rules in test set. Association rules are constantly revised based on the results in order to obtain satisfactory results.

III. EXAMPLES

A. Preprocessing Patents

Here mechanical products closely related to sustainable design are as examples to illustrate the specific method. According to the characteristics of sustainable products, the patent randomly divided into two parts when preprocessing. One part is the training set, which made up of 20 patents to extract association rules. Via human analysts, antibody/synonyms are inducted, specialized dictionaries are built, the symbol table of sustainable design features-functions-structural features is as shown in tab.1. Another part is the test set and is used to check the correctness of association rules. Sustainable design features are represented as S_1: noise reduction, S_2: nontoxic, S_3: low emission, S_4: vibration reduction, S_5: energy saving, S_6: removable, S_7: recyclable, S_8: solvent-free, the S_9: no fluorescence. Objective functions said for F_1: easily installation, F_2: easily manufacturing, F_3: easily degradation, F_4: easy maintenance, F_5: easy installation, F_6: easy recycling, F_7: improve acoustic performance, F_8: improve air Dynamics performance, F_9: improve stiffness, F_{10}: improve sealed sex, F_{11}: improve

efficiency, F_{12}: improve stability, F_{13}: improve by force uniform sex, F_{14}: improve control flexibility. F_{15}: longevity, F_{16}: enhanced compression capability, F_{17}: increased carrying capacity, F_{18}: lose weight, F_{19}: reduced size, F_{20}: reducing cavitations, F_{21}: reduction of cavitations, F_{22}: reduce flashing F_{23}: reduce the release of acid gases, F_{24}: simplify structure, F_{25}: lower costs, F_{26}: fuel economy. Shape features are said for K_1: gate mesh, K2: Hexagon, K_3: cellular structure, K_4: maze type, K_5: tapered, K_6: spiral structure, K_7: hollow body, K_8: Butterfly shaped, K_9: streamlined, K_{10}: leg-like, K_{11}: L shaped, K_{12}: semicircular shaped, K_{13}: oriented Cone, K_{14}: arc surfaces, K_{15}: tooth shaped, K_{16}: symmetric set. Feature representation M_1 materials: aluminum, M_2: rubber, M_3: elastic, M_4: hard materials, M_5: low smoke zero halogen flame retardant filler, M_6: polyurethane, M_7: Crude methyl esters of fatty acids, M_8: a thin layer of fabric, M_9: glass fiber, M_{10}: plastic, M_{11}: isobutylene, M_{12}: stainless steel wire. Process Features denoted by T_1: epoxy-coated, T_2: extrusion, T_3: hydrophobic, T_4: bending. After finishing as shown in tab.2, item set tables is set up.

B. Acquiring Frequent Itemsets

PCAR algorithm is particularly suitable for mining association rule at large volumes of items and small frequency of itemsets. While getting frequent itemsets based on the principles of top-down, It sets the minimum support MinSupport=2, and starts from frequent 8-set to mine frequent 7-set, until more frequent 2-set can't be found. Frequent itemsets got from the objective functions-parts feature sets are $L_4=\{F_{12}F_{14}F_{19}K_7\}$, $L_3=\{F_{10}F_{15}T_1\}$, $L_2=\{K_7F_{10}\}$. Similarly, remove the frequent itemsets $\{F_{24}F_{25}$, $F_{10}F_{15}\}$ that are not associated with sustainable design features. Frequent itemsets $L_7=\{S_4F_2F_{10}F_{12}F_{13}F_{14}F_{19}\}$, $L_5=\{S_4F_{12}F_{14}F_{19}K_7\}$, $L_3=\{S_3F_{11}F_{26}\}$, $L_2=\{S_1F_2$, S_2F_{15}, S_5F_4, S_5F_{18}, $S_6F_{19}\}$ are obtained from sustainable design features-the objective functions set.

C. Generating Association Rules

The association rule is "$c \Rightarrow (L-c)$"according to the minimum confidence threshold $conf_{\min}$, the number of records *Support_count (L)*containing frequent itemsets

L, and the number of records *Support_count（c）* containing set *c*.

Relationships between objective functions and parts features can be analyzed according to the product functions from parts features. Relation sustainable design features with the objective functions is analyzed based on the probability of the objective functions to sustainable design features. Its relevance can be evaluated by calculating the confidence.

TABLE 2 TABLE OF FEATURED ITEMSETS

Serial Number	Sustainable Design Features-Objective Functions Itemsets	Objective Functions-Parts Features Itemsets
1	S_1- $F_1F_2F_7F_8$	$F_1F_2F_7F_8$-K_1K_2
2	S_1- $F_2F_{20}F_{21}F_{22}F_{24}F_{25}$	$F_2F_{20}F_{21}F_{22}F_{24}F_{25}$-$K_4M_4$
3	S_1- $F_2F_9F_{18}$	$F_2F_9F_{18}$-$K_3M_1M_2M_3$
4	S_2- $F_{15}F_{16}F_{23}$	$F_{15}F_{16}F_{23}$-M_5M_6
5	S_2S_8- $F_{10}F_{15}$	$F_{10}F_{15}$-T_1
6	S_2S_9- F_3F_{25}	F_3F_{25}-M_7
7	S_3- $F_{11}F_{26}$	$F_{11}F_{26}$-K_6
8	S_3- $F_{11}F_{12}$	$F_{11}F_{12}$-K_5
9	S_3- $F_{11}F_{26}$	$F_{11}F_{26}$-M_8
10	S_4- $F_{12}F_{14}F_{19}F_{24}F_{26}$	$F_{12}F_{14}F_{19}F_{24}F_{26}$-$K_7$
11	S_4- $F_2F_{10}F_{12}F_{13}F_{14}F_{19}$	$F_2F_{10}F_{12}F_{13}F_{14}F_{19}$-$K_7$
12	S_4- $F_2F_{10}F_{12}F_{13}F_{14}F_{19}$	$F_2F_{10}F_{12}F_{13}F_{14}F_{19}$-$K_8$
13	S_5- $F_{17}F_{18}F_{19}$	$F_{17}F_{18}F_{19}$-$K_{10}M_{12}T_3T_4$
14	S_5- $F_4F_{10}F_{15}$	$F_4F_{10}F_{15}$-$K_9M_9T_1$
15	S_5- F_4F_{18}	F_4F_{18}-$K_7M_{10}M_{11}T_2$
16	S_6- F_4F_{25}	F_4F_{25}-K_{16}
17	S_6- $F_{10}F_{15}F_{19}$	$F_{10}F_{15}F_{19}$-K_{11}
18	S_6- $F_5F_{10}F_{24}$	$F_5F_{10}F_{24}$-K_5K_7
19	S_6S_7- $F_1F_6F_{19}F_{24}F_{25}$	$F_1F_6F_{19}F_{24}F_{25}$-$K_{12}K_{15}$
20	S_7- $F_{11}F_{25}$	$F_{11}F_{25}$-$K_{13}K_{14}$

D. Discovering Knowledge

The relationship knowledge about sustainable design features-objective functions-parts features can be discovered by testing the above rule into the test set. Rules would be revised according to test results, final rules obtained as shown in tab. 3. Sustainable design knowledge can be discovered by the extraction rules.

In order to improve the precision of the method, the relationships between the sustainable design features-parts features and the sustainable design features-objective functions added to the symbol table,

which was no found in patents. The accuracy can be improved after updating set table, fetching the new association rules and adding to table again for extraction of design knowledge.

TABLE 3 TABLE OF DISCOVERED RULES

	Sustainable Design Features-The Objective Functions (Minsup=0.5)	Objective Functions-Parts Features (Minsup=0.5)
1	rule 4: $F_2 \rightarrow S_1$ conf =3/5=0.5	rule 1: $T_1 \rightarrow F_{10}F_{15}$ Conf= 2/2=1
2	rule 3: $F_{11}F_{26} \rightarrow S_3$ conf =2/2=1	
3	rule 2: $F_{12}F_{14}F_{19} \rightarrow S_4$ conf =3/3=1	
4	rule1: $F_2F_{10}F_{12}F_{13}F_{14}F_{19} \rightarrow S_4$ conf=2/2=1	
5	rule 6: $F_4 \rightarrow S_5$ conf =2/3=0.67	
6	rule 7: $F_{18} \rightarrow S_5$ conf =2/3=0.67	

IV. CONCLUSIONS

A method is presented based on the data mining algorithms to find association rules among the sustainable design features-objective functions-parts features in the patents. Then product knowledge of sustainable design is discovered by obtained association rules. The example proves good results. With increasing in the number of patent documents in the database, test set can be automatically expanded, rules base can be perfect, and recall and precision of the sustainable design knowledge would be continuously improved. Gained associational knowledge among sustainable design features, objective functions, parts features can deepen understanding of sustainable design products and aiding sustainable products design.

REFERENCES

[1] ZHEN Lu, JIANG Zhuhua, SHU Hai. The knowledge grid aided innovation platform and its key techniques[J], Journal of Shanghai Jiaotong University, 2007, 41(6), pp. 876-880.
[2] LI Shumei, HE Weiping, ZHAO Feng. Method to search information from internet for product creative design domain[J], Application Research of Computers, 2007, 24(4) , pp. 117-119.
[3] LIN Li, YANG Mingqing, YANG Qin. Research on product innovation design based on human machine knowledge[J], Journal of Engineering Graphics, 2007, 2, pp.129-133.
[4] WANG Zhaoxia,QIU Qingying, FENG Peien. Patent knowledge mining for conceptual design[J]. Journal of Zhejiang University (EngineeringScience), 2008, 42(3) , pp. 522-527.
[5] XUE Ghi, Qiu Qingying, FENG Peien. Acquisition method for principle solution of mechanical patent[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(1), pp. 222-229.
[6] ZHANG Hui，QIU Qingying，FENG Peien，WANG Zhaoxia. Research of a method for acquiring design knowledge from product patents[J]. Journal of Harbin Engineering University, 2009, 30(7) , pp.785-791.
[7] ZHANG Hui. A tensioning device of underwater diamond wire saw machine: China，201220497623.0[P] . 2012-09-06.
[8]GUO Jianzhang, WANG Yufeng, LI Qingling. The greenhouse double collector solar chimney power generation device based on: China，201310109459.0[P], 2013-06-12.
[9] YU Jun. Inflatable Recyclable bolt: China，201210426191.9[P], 2013-02-20.
[10] WANG Chendong. Environmental protection non-toxic flame retardant wire: China，201120221462.8[P], 2011-06-28.
[11]FENG Peien, ZHANG Hui, QIU Qingying, et al. PCAR：an efficient approach for mining association rules[C]. In: Proceedings of the ICNC-FSKD 2008 International Conference on Fussy Systems and Knowledge Discovery, Jinan, China, Oct 2008, pp.605-609.