CIRP Annals - Manufacturing Technology xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: http://ees.elsevier.com/cirp/default.asp

Product features characterization and customers' preferences prediction based on purchasing data

Jian Zhang [a,b], Alessandro Simeone [b], Peihua Gu (1)[a,b,*], Bo Hong [b]

[a] Intelligent Manufacturing Key Laboratory of Ministry of Education, Shantou University, Shantou, China
[b] Department of Mechatronic Engineering, Shantou University, Shantou, China

ARTICLE INFO

Keywords:
Design
Product development
Big data

ABSTRACT

Big data of online product purchases is an emerging source for obtaining customers' preferences of product features for new product development. This paper proposes a framework and associated method for product features characterization and customers' preference prediction based on online product purchase data. Specifications and components of products are firstly analyzed and the relationships between product specifications and components are then established for features characterization. The customers preferred specifications, features and their combinations are predicted for development of new products. The features characterization and customers' preferences prediction of toy cars were used as an example of illustrating the proposed method.

1. Introduction

Product requirements specifications play an important role in developing competitive products [1]. Various product design methods are applied on the basis of understanding customers' preferences of product features and characteristics [2]. Product sales data is a valuable source for obtaining the customers' needs and preferences [3]. In recent years, online shopping market has been growing very rapidly. Fig. 1 shows that the global market data trend of the online retail market and customer numbers are expected to reach 2.49 trillion US dollar and 1.79 billion customers respectively in 2018 [4]. Product feature preferences of customers are embedded in the big data of online purchases. It was found that existing research mainly focuses on mapping online purchasing behaviour based on customers reviews and feedback [5,6]. Since customers subconscious and sensorial perception on preference are difficult to fully describe [7], it is needed to develop methods for determining customers' preferences based on big data of online purchases.

To facilitate new product development, this paper introduces a framework and an associated method for product features characterization and customers' preferences prediction based on online purchasing big data.

2. The proposed method

In the proposed method, customer, functional and physical domains are considered as follows:

- Customers satisfaction index is defined in customer domain for the quantification of customer' satisfaction on a specific product. The better product sales lead to the higher customer satisfaction index in the market. Accordingly, the prediction of customers' preferences is evaluated by customers satisfaction index.
- In functional domain, product significant specifications are identified. Customers satisfaction index is influenced by significant specifications and their combinations. Intelligent modelling of relationships between customers' satisfaction index and product specifications is used to predict customers' preferred specifications and their optimal combinations.
- Product specifications are realized by physical components. The relationships between the specifications and the components of products are modelled for the specification clustering and features characterization. The specifications within the same

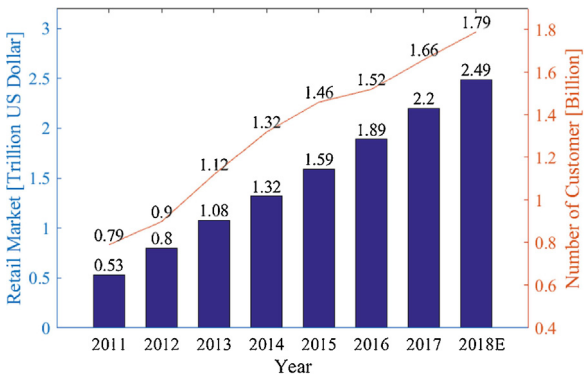

Fig. 1. Increasing trend of global online purchasing market [4].

* Corresponding author at: Intelligent Manufacturing Key Laboratory of Ministry of Education, Shantou University, Shantou, China.
E-mail address: peihuagu@stu.edu.cn (P. Gu).

https://doi.org/10.1016/j.cirp.2018.04.020
0007-8506/

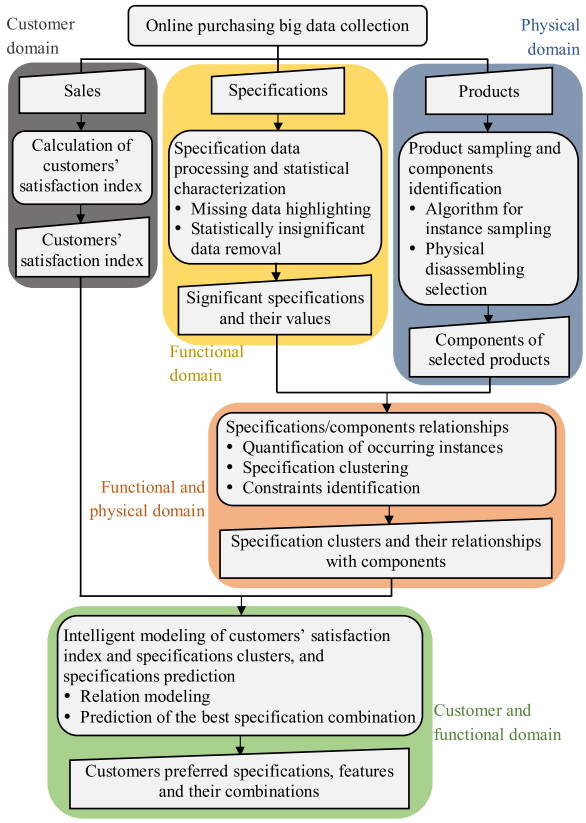

Fig. 2. Flow chart of the proposed method.

cluster depend on the same components. Physical constraints on the specifications clusters are identified based on components assessment.

A flow chart of the proposed method is shown in Fig. 2. Six operations provided in the flow chart are described below.

2.1. Online purchasing big data collection

In this step, market survey and big purchasing data set preparation are carried out. The market survey, aimed at the acquisition of online purchasing big data, needs to be carried out through accessing online shopping platforms and other sources i.e. market statistics and database for a given product category.

Variability of non-design factors such as product brand, advertisement etc., can be considered in the survey, by assigning a score to the brand proportional to its market share impact and to quantify the customer bias influence on sales.

The survey output is a catalogue of products available on market, including price range, number of sales and a comprehensive list of product specifications. The purchasing data set consists of a large matrix with rows representing the products and the columns representing the sales and specifications of products.

2.2. Customers' satisfaction index calculation

Customers' satisfaction index is defined and calculated in this step. Products sales are used for the evaluation of customers' satisfaction index. In this work, the total number of product sales N_T is calculated as follows:

$$N_T = \sum_{i=1}^{p} N_i \tag{1}$$

where p is the number of products considered in market survey, N_i is the purchasing number of specific product i, $i = 1, 2, \ldots, p$.

Customers' satisfaction index SI_i of product i is calculated by:

$$SI_i = \frac{N_i}{N_T} \times 100\% \tag{2}$$

Customers' satisfaction index is calculated for each product selected in the survey. According to the Eqs. (1) and (2), SI_i, can be used to evaluate the customers' preferences on product i in the market. The higher is the value of SI_i, the more customers preferred on product i. In general, a high value of customers' satisfaction index is required for new product development.

2.3. Specification data processing and statistical characterization

Specification data processing operation consists in reducing the big data complexity by removing the specifications which are not relevant for the purpose. Redundant or difficult to mathematize specifications need to be removed too.

To extract information from the processed big purchasing data set, and to enable its further reduction, statistical parameters i.e. mean and variance need to be computed for products (rows) and specifications (columns). This computation aims at highlighting the amount of missing/incomplete data and showing any statistical insignificance, due for example to an extremely low data variability. These operations allow a possible dataset reduction. Significant specifications and their values can be obtained for further analysis.

2.4. Product sampling and components identification

In this step, product samples and their components need to be identified for modelling relationships between components and specifications. Products sampling is used to select a subset of statistically representative product instances. Products can be sampled according to one precise specification, e.g. price range or sales. When the problem complexity increases and the selection needs to be carried out according to more specifications at the same time, the sampling techniques can be applied to the big data set [8]. The number of product instances to be selected depends on the specific application, the commercial availability and the economic feasibility.

A comprehensive study on design features is also carried out in this step for components identification. The objective of this exercise is to produce a list of components to be matched with specifications. For some products, the design features can be easily accessed or measured, for some others, instead, it is necessary to disassemble the product in order to retrieve the bill of components.

2.5. Specifications/components relationships

The objective of this step is to evaluate the components influence on the specifications. In this respect, a matrix of specifications/components relationships is established, in which the rows represent the components, the columns represent the specifications. The single cells contain the number of product instances for which that specifications/components relationship occurs.

Physical dependencies among specifications need to be investigated according to their relationships with one or more components reported in the matrix. Intelligent clustering algorithms can be used for this purpose, such as fuzzy k-means, hierarchical clustering, or self-organizing maps [9]. The choice of the algorithm depends on the specific purpose, the amount of available data and the computational complexity.

Based on the specifications/components relationships, physical constraints amongst specifications and their combinations within each cluster are identified through specific components assessment.

2.6. Intelligent modelling for the best combination of specifications

In this step, an intelligent model is set up to predict the best combinations of product specifications. For this purpose, a regression model is initially trained using specification data as

Please cite this article in press as: Zhang J, et al. Product features characterization and customers' preferences prediction based on purchasing data. CIRP Annals - Manufacturing Technology (2018), https://doi.org/10.1016/j.cirp.2018.04.020