Customer Preference Mining of Electric Vehicles for Design Decision-making Using Big Sales Data

Xiaojie Gao[1], Rui Li[2], Deping Chen [3], Jian Zhang[4] and Qingjin Peng[5]

[1]Guangdong University of Technology, GGDgxj303682212@mail2.gdut.edu.cn
[2]Shantou University, 22rli1@stu.edu.cn
[3]Shantou University, cdp152418@126.com
[4]Shantou University, jianzhang@stu.edu.cn
[5]University of Manitoba, qingjin.peng@umanitoba.ca

Corresponding author: Jian Zhang, jianzhang@stu.edu.cn

Abstract. Rapid and accurate identification of customers preferences on specification combinations is critical to the development of competitive electric vehicles. Both conscious and subconscious preferences of customers are embedded in the big sales data of electric vehicles. Based on the big sales data of electric vehicles in the Chinese market in 2021, this paper provides a novel customer preference mining approach for supporting design decision-making of electric vehicles. Specifications and sales numbers of electric vehicles are collected and analyzed for customers preferences information mining. Based on the relationship between specifications and components, mutation and reproduction probabilities of electric vehicle components are calculated for meeting customers' diversified preferences on specification combinations. Recommendations of design decision-making for enhancing product adaptability are made through components clustering based on the mutation and reproduction probabilities of electric vehicle components. Similarities and differences between the recommended design and current design of electric vehicles are investigated to illustrate the rationality of the proposed method.

Keywords: Electric vehicle; customer preference; design decision, big sales data; information entropy
DOI: https://doi.org/10.14733/cadaps.2024.241-257

1 INTRODUCTION

In the context of improving air quality, mitigating climate change and saving energy, electric vehicles, a main strategic direction of new energy vehicles in China, are highly competitive [27]. To increase the market sharing of electric vehicles, the vehicle design efficiency must be improved by quickly and accurately identifying customer preferences. With the rapid development of the Internet and information technology, customers can quickly, comprehensively and accurately obtain relevant features and specifications about electric vehicles for comparison. Product design needs accurate information of customer preferences for enhancing product competitiveness [4],[29]. It is very important to obtain customer preference information quickly and accurately for supporting rationalized design decision-making of electric vehicles.

Data collection is essential for the customers preference analysis. Existing methods of customers data collections are mainly through customer interviews and questionnaires to obtain customers preference data on product features [13],[25]. Due to the lack of design expertise, it is difficult for customers to accurately express their subconscious requirements through semi-structured and unstructured expressions [1],[30]. With the development of Internet and e-commerce, customer's comments of vehicles have become an important source for collecting customers preference data [15],[22]. However, only customers who are far from their psychological expectations are more likely to provide their evaluation comments, such as customers who far exceed expectations or are overly disappointed. Both emotional positive evaluations and malicious negative evaluations from customers potentially lead undesired impacts on the accuracy of preference data collections. Due to development of the human-computer interaction technology, collecting the physiological data of customers in the process of using vehicles has also become an effective means to obtain customer data [19],[28]. However, collecting the physiological data of customers are costly and limited by the customer numbers and environment. Results of accurate measurements from a small number of customers are not generalizable and cannot be applied to the entire vehicle market.

To support rationalized design decision-making, preference information mining through the quantitative and qualitative analysis of collected customer data is also needed [14]. Due to influence of the customer demand expression and preference information transmission, it is necessary to analyze customers preference information through Metaphor Elicitation Technique, Analytic Hierarchy Process, online evaluation recognition and other methods to accurately identify potential customer preferences [5],[7]. By analyzing the behavior of customers through User Journey Map and User Experience Map, the requirement of customers in the using process of vehicles can be accurately obtained [16],[18]. However, the analysis process requires extensive research experience (e.g., the competence of the researcher, the choice of methodology) and professional ability (e.g., the level of knowledge of the vehicle industry), and the accuracy of the results are hardly to be replicated [17]. Analysis of quantitative data on the customer behavior, cognition, and attitudes through the Perceptual Engineering Scale method, Kawakita Jiro (KJ) method, and Kano model [34] can also be carried out for mining customers preference information. However, the accuracy of those methods depends on both the completeness and quality of the collected data.

In summary, the collected datasets of the existing methods for the customers preference analysis are not big enough, resulting in an insufficient reflection of customers preferences in the marketplace. In addition, customers preferences on specification combinations are not well considered in those existing methods. Moreover, customers preferences information can be distorted through inaccurate mapping of customers needs into design specifications, leading to improper design decision-making [32-33].

In the age of the Internet and e-commerce, there is a large amount of sales data on vehicle websites, including specifications, parameters, prices and sales volumes. According to the statistical analysis of China Association of Automobile Manufacturers (CAAM), the new energy vehicle market in China has changed from policy-driven to market-pulled in 2021, and the sales models in 2021 account for 75% of the total sales [9]. The big sales data of electric vehicles provide new solutions for the customer preference mining. Comparing with the traditional methods of collecting and analyzing customer requirements, the large amount of real-time information in sales data of electric vehicles provides timely and comprehensive information of customer preferences [23]. Designers can quickly capture the market change to improve the vehicle design for new preferences. Identifying the correlation between the customer preference and specification combination will help solving problems caused by inaccurate mapping of customer requirements and design specifications in the existing customer preference research process.

Adaptability is the capability of a design or product to be adapted for meeting changeable requirements [11-12]. A design with enhanced adaptability whose functional modules can be easily modified to satisfy changed requirements [6],[24],[31]. In addition to rapid and accurate information mining of customers' preferences, adaptability of electric vehicles also need to be enhanced for better meeting diversified and changeable preferences of customers [2],[21]. By using the big sales data of electric vehicles in the Chinese market in 2021, a novel preference information mining approach is proposed to support design decision-making of division and feature identification of electric vehicle modules.

2 MARKET PREFERENCE INFORMATION MINING AND DESIGN DECISION METHOD FOR ELECTRIC VEHICLES

Quantifying customer preference information through electric vehicle market big data to assist electric vehicle design decisions can improve the decision efficiency and customer satisfaction. Based on our previous work [32-33], this paper quantifies customer preferences at the level of electric vehicle specifications based on historical big sales data. Information is searched by a relationship model between electric vehicle specifications and sales data. Data analysis of customer preferences combines information axioms for product family segmentation recommendations. Design recommendations are made for electric vehicle components based on the relationship of electric vehicle specifications, components and size of information of a single specification. The method consists of 4 parts shown in Figure 1.

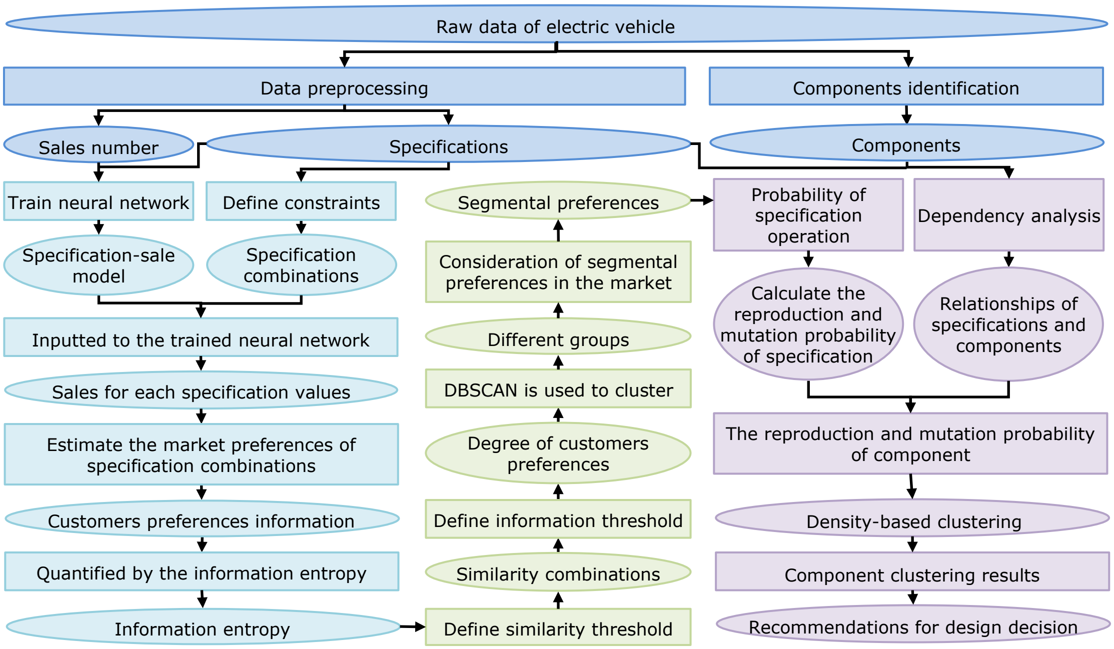

Figure 1: Flow chart of preference information mining and design decision of electric vehicles.

The method is mainly divided into the following 4 parts.

- Sales data of electric vehicles are collected to form the electric vehicle raw data set. The data are pre-processed to discretize values of electric vehicle specifications for datasets of electric vehicle sales, electric vehicle specifications and electric vehicle components.
- A relationship model is built between electric vehicle specifications and sales volume using neural networks to calculate and mine the amount of market preference information of the

combination of specifications that satisfy physical constraints in the full arrangement of existing specifications.

- Density clustering of electric vehicle specifications is identified according to the self-information quantity in dividing customer preferences and determining design requirements of electric vehicle specifications for corresponding market segments.
- The relationship between electric vehicle specifications and components is combined to determine the probability of component operations by the probability of specification operations. The module types of electric vehicle components are determined by the cluster analysis of the components.

3 COLLECTION AND PROCESSING OF ELECTRIC VEHICLE DATA

To mine electric vehicle customer preference information through big sales data, big sales data of electric vehicles are collected and sorted to create a data set of electric vehicle specifications and sales. This research collects the specification data of electric vehicles to obtain the corresponding sales data from websites of vehicle sales data. Supplement vehicle data are collected from the official website, user manuals and repair manuals, etc. The electric vehicle sales data and vehicle information are collected through the Internet including the electric vehicle price, sales volume, specifications, component, etc. Different styles and models of electric vehicles are organized according to the sales and specification characteristics to form a data set of electric vehicle sales and indicators. Finally, data of 1482 electric vehicle models are collected in the Chinese market from January 2017 to March 2022, including 236 specifications and 5,489,559 sales data. Those sales data of electric vehicles are mainly collected through publicly available electric vehicle data platforms, integrated with data from other platforms, such as the official websites of vehicle brands, user manuals, and vehicle repair manuals. The data are shown in Table 1, where each row represents an electric vehicle, each column represents a kind of specifications, and cells in the matrix are the corresponding specification values of each electric vehicle.

ID	*Price*	*Maximum power*	*Maximum torque*	*Maximum speed*	*Official 100km/h acceleration*	*Length*	*Width*	*…*
1	118800	40	140	120	-	2986	1676	…
2	251900	160	360	180	7.7	4930	1940	…
3	281900	160	360	180	8.3	4930	1940	…
4	241900	160	360	180	7.8	4788	1940	…
5	256900	160	360	180	7.8	4788	1940	…
6	261900	160	360	180	8.4	4788	1940	…
7	276900	160	360	180	8.4	4788	1940	…
…	…	…	…	…	…	…	…	…

Table 1: Electric vehicle raw data set.

3.1 The Electric Vehicle Specification Data Set

The data collected through the electric vehicle data platform contain noisy data that are biased or erroneous with respect to real values. The noisy data, such as missing values, outliers, and values inconsistent with common sense, affect the data analysis and information mining. To facilitate the data mining, the data is pre-processed for incomplete and noisy electric vehicle data by data cleaning and data transformation [26]. The specification data are removed when they affect the model construction. In this research, the acceptance threshold for missing data is set at 40%. Missing specifications that exceed the threshold are removed and missing values that below the threshold are supplemented using constants, mean or median values of similar electric vehicle specifications. In addition, as the focus of this research is on design-related characteristics, all specifications that are not relevant to design purposes are removed. To simplify the analysis, the influence of non-physical factors on electric vehicle customer preferences is not considered. After filtering and dimensionality reduction of the raw data, the number of selected electric vehicles is reduced from 1482 to 430, and only 18 critical specifications as shown in Table 2 are considered for further analysis.

This research reserves data of electric vehicle specifications related to battery, motor and frame. The specifications related to the battery are the NEDC electric range, battery type, battery capacity, fast charging time, slow charging time, and battery warranty. The specifications related to the motor are the maximum speed, 0-100km/h acceleration, maximum motor power, etc. The specifications related to the frame are the length, width, height, chassis height, maximum weight, and number of seats. This research discretizes the continuous electric vehicle specifications to decide the probability of the market preference. For example, values of the length and width are divided into 5 intervals, values of the height are divided into 4 intervals, and values of the price are divided into 6 intervals. For the specifications that do not have continuous values, their values are discrete values, such as the battery type and battery warranty. The final division of the electric vehicle specification values is shown in Table 2.

3.2 The Components of Electric Vehicles

Electric vehicles are powered by batteries and driven by electric motors to satisfy all requirements of road traffic and safety regulations. The composition of electric vehicles includes electric drive and control system, transmission and other mechanical systems, and working devices to accomplish the tasks. The electric drive and control system consists of the drive motor, battery and speed control device of the motor. The electric drive and control system, which has the biggest difference from internal combustion engine vehicles, is the core of electric vehicles. The other devices of electric vehicles are basically the same as those of internal combustion engine vehicles. Some of the electric vehicle components are shown in Figure 2.

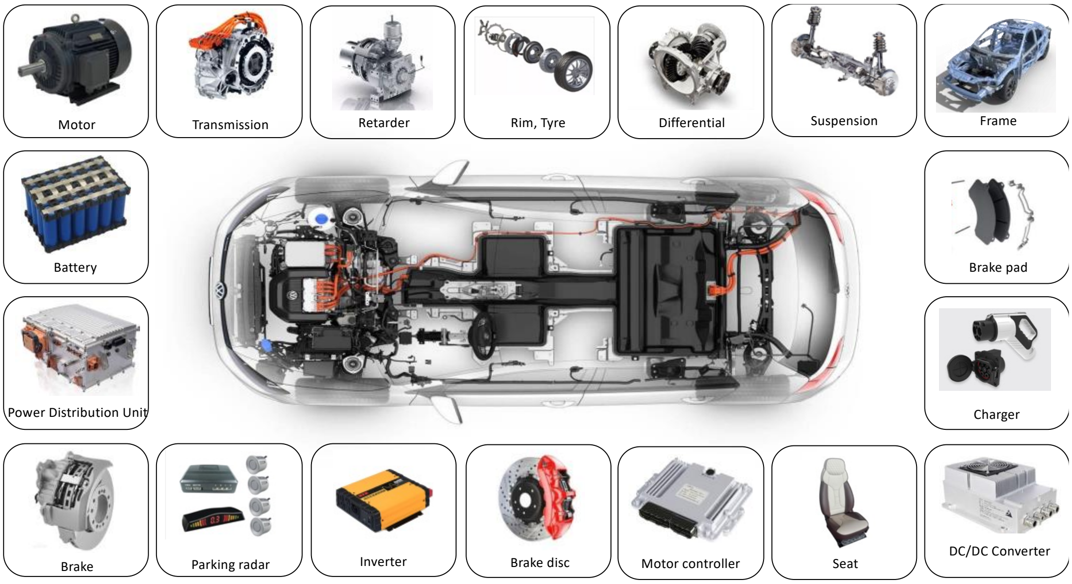

Figure 2: Different types of components in electric vehicles.

ID	Specifications	Value	Unit
S_1	Overall length	1:<3000; 2:≥3000-3500; 3:≥3500-4000; 4:≥4000-4500; 5:≥4500	mm
S_2	Overall width	1:<1600; 2:≥1600-1700; 3:≥1700-1800; 4:≥1800-1900; 5≥1900	mm
S_3	Overall height	1:<1450; 2:≥1450-1550; 3:≥1550-1650; 4:≥1650	mm
S_4	Maximum speed	1:<120; 2:≥120-140; 3:≥140-160; 4:≥160-180; 5:≥180	km/h
S_5	0-100km/h acceleration	1:<7; 2:≥7-10; 3:≥10-13; 4:≥13-16; 5:≥16-19; 6:≥19	s
S_6	NEDC electric range	1:<200; 2:≥200-300; 3:≥300-400; 4:≥400-500; 5:≥500-600; 6:≥600	km
S_7	Maximum motor power	1:<100; 2:≥100-200; 3:≥200-300; 4:≥300-400; 5:≥400	kw
S_8	Battery type	1: Lithium-ion battery; 2: Lithium iron phosphate Battery; 3: Ternary Lithium Battery	-
S_9	Battery capacity	1:<30; 2:≥30-50; 3:≥50-70; 4:≥70	kwh
S_{10}	Fast charging time	1:<0.6; 2:≥0.6-0.8; 3:≥0.8-1.0; 4:≥1.0-1.2; 5:≥1.2-1.4; 6:≥1.4	h
S_{11}	Slow charging time	1:<5; 2:≥5-7; 3:≥7-9; 4:≥9-11; 5:≥11-13; 6:≥13	h
S_{12}	Battery warranty	1: Unlimited years/ mileage for the first owner; 2: 8 years or 150,000 km; 3: 8 years or 120,000 km; 4: 6 years or 60,000 km; 5: 8 years or 160,000 km; 6: 8 years or 300,000 km; 7: 10 years or 200,000 km; 8: 5 years or 500,000 km	-
S_{13}	Chassis height	1:<130; 2:≥130-150; 3:≥150-170; 4:≥170	mm
S_{14}	Maximum weight	1:<200; 2:≥200-300; 3:≥300-400; 4:≥400-500; 5:≥500-600; 6:≥600	kg
S_{15}	Number of seats	2; 3; 4; 5; 6; 7	-
S_{16}	Advanced Driver Assistance System	1:0; 2:1; 3:2; 4:3	-
S_{17}	Parking brake	1: Hand brake; 2: Foot break; 3: Electric Parking Brake	-
S_{18}	Price	1:<100000; 2:≥100000-200000; 3:≥200000-300000; 4:≥300000-400000; 5:≥400000-500000; 6:≥500000	CNY

Table 2: Specifications and their value ranges.

In order to achieve customer preference analysis and design smart decisions for electric vehicles, this research collects electric vehicle components in Technical Conditions for Pure Electric Vehicles released and implemented in 2012 and Technical Conditions for Pure Electric Vehicles (Draft for Comments) released in June 2021. 32 key electric vehicle components related to specifications are selected. The final division of the electric vehicle components is shown in Table 3.

ID	Component	ID	Component	ID	Component
C_1	Motor	C_{10}	Tyre	C_{19}	Power distribution
C_2	Transmission	C_{11}	Seat	C_{20}	Traction control

C_3	Retarder	C_{12}	Brake	C_{21}	High voltage charging module
C_4	Rim	C_{13}	Charger	C_{22}	Brake auxiliary device
C_5	Differential	C_{14}	Parking radar	C_{23}	Body stability control
C_6	Front suspension	C_{15}	Inverter	C_{24}	Lane departure warning
C_7	Rear suspension	C_{16}	Motor controller	C_{25}	Lane keeping assist
C_8	Frame	C_{17}	DC/DC Converter	C_{26}	Road traffic identification
C_9	Battery	C_{18}	LCDA	C_{27}	Active brake / Active safety

Table 3: Selected components and their IDs in electric vehicles.

4 DATA MINING OF CUSTOMERS PREFERENCES FOR ELECTRIC VEHICLES

4.1 Relationship Modeling Between the Electric Vehicle Specification and Sales Volume

Sales volumes of electric vehicles depend on the degree of customer preferences in the market, and high sales are due to the design of electric vehicles that satisfy customer requirements. Therefore, this research assumes that the higher sales volume of electric vehicles in the market lead to the higher customer satisfaction with the combination of specification values. By modeling the relationship between specifications and sales through the existing electric vehicle data in the market, customer preferences mining and design decisions for electric vehicles can be made efficiently and accurately. Seventy percent of the electric vehicle data are randomly selected as the training data set, fifteen percent of the data are randomly selected as the validation data set, and the remaining data are used as the test data set. A neural network is built to train the relationship model by ANN (artificial neural network) fitting [10]. The combination of 18 electric vehicle specifications summarized in Table 2 is used as the input, and the 5,489,559 corresponding sales volume is output of the model. The training proceeds until an expected result obtained. The method is repeated 13 epochs using the Bayesian model. Based on the result comparison, the best validation performance is 3.6784 at epoch 7, the correlation of the training data is about 0.87664, the correlation of the test data is 0.43734, the correlation of the validation data set is 0.62248, and the correlation of all data is about 0.76981.

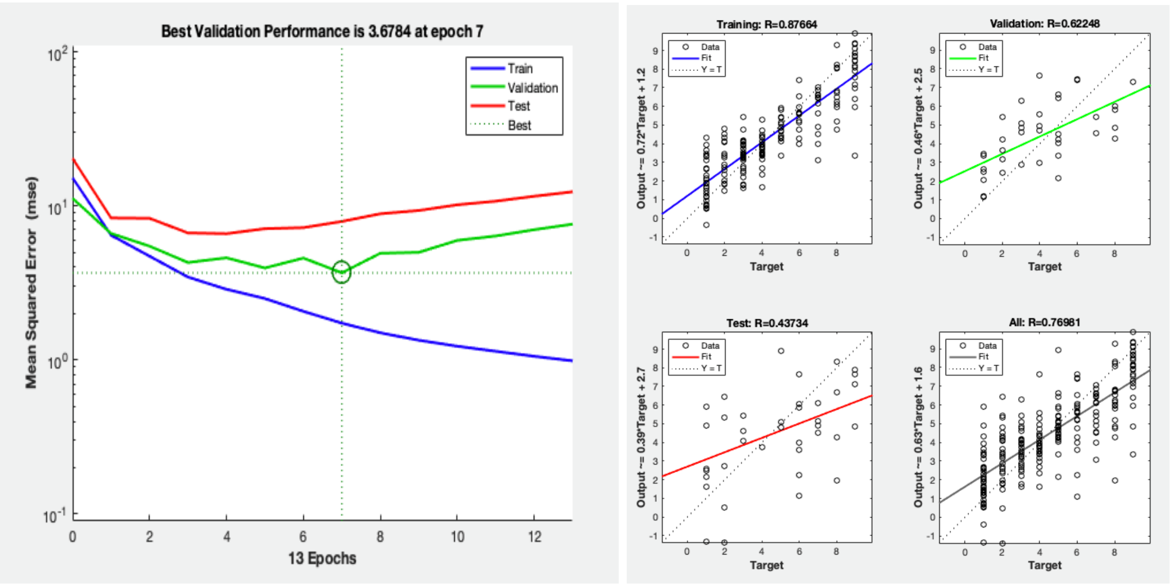

Figure 3: Model error variation.

4.2 Physical Constraints for Electric Vehicle Specifications

Sales data of electric vehicles only represent the existing customer preference, new electric vehicle specifications are not included in the sales data. The full arrangement of existing electric vehicle specifications can obtain combinations of electric vehicle specifications that are not available in the market. Since not all electric vehicle designs with new electric vehicle specification combinations can be realized, electric vehicle specification combinations that violate physical rules and economic constraints in design are excluded. These combinations that violate the objective rules will not appear in the future electric vehicle market, so they are not of practical reference. Considering combinations of electric vehicle specifications that may violate physical feasibility and economic constraints, this research includes the constraint range for the specification combinations to ensure the accuracy of their corresponding market sales. Some of the constraints are shown in Table 4. For an EV manufacturer, technical constraints should be predefined according to their manufacturing capability.

ID	*Additional constraints*
1	Price ≤ 200,000 yuan, NEDC electric range ≤ 300 km
2	The maximum motor power ≤ 200kw, Maximum speed ≤ 170 km/h
3	Number of seats ≥ 5, Overall length ≥ 3000mm

Table 4: Some of the corresponding constraints for the specification combinations.

4.2.1 Preference probabilities for specification combinations satisfying physical constraints

Based on the relationship model, the expected value to satisfy the physical constraints is predicted and the amount of information for the entire market is estimated. The electric vehicle specification combinations that satisfy the physical constraints are added into the relationship model of electric vehicle specifications and sales to obtain the expected sales of the new electric vehicle specification combinations. The preference probability of the corresponding specification combination can be calculated based on the estimated sales of electric vehicles. A higher value of the probability of preferences for an electric vehicle means that more customers prefer the electric vehicle with this combination of values. Results of the preference probabilities of the electric vehicle specification combination are shown in Figure 4.

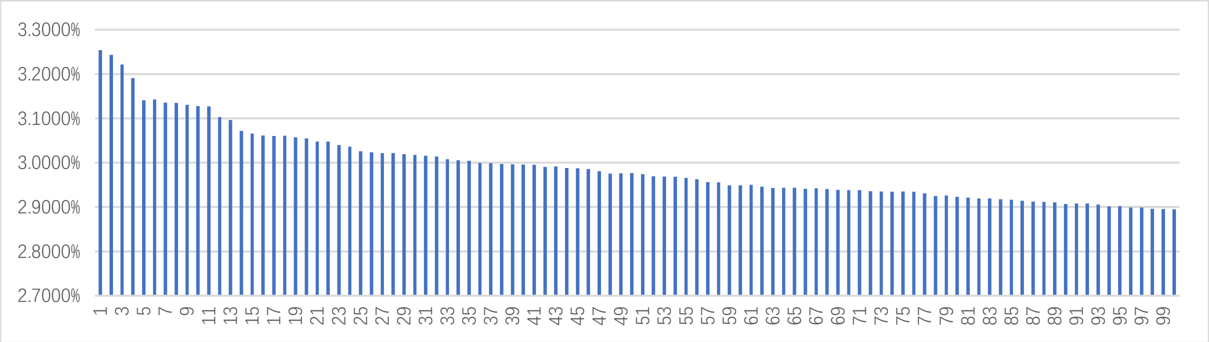

Figure 4: Preference probabilities of the electric vehicle specification combination.

4.2.2 Information entropy for electric vehicle specifications

Information entropy is usually used to quantify the information obtained. The larger information entropy, the lower uncertainties on customers preferences of product specifications. Based on the market preference probabilities of different electric vehicle specification combinations, the information entropy of each electric vehicle specification combination can be decided according to the information entropy algorithm, and the information entropy of the whole electric vehicle market preference can be identified [32]. The results are shown in Figure 5.

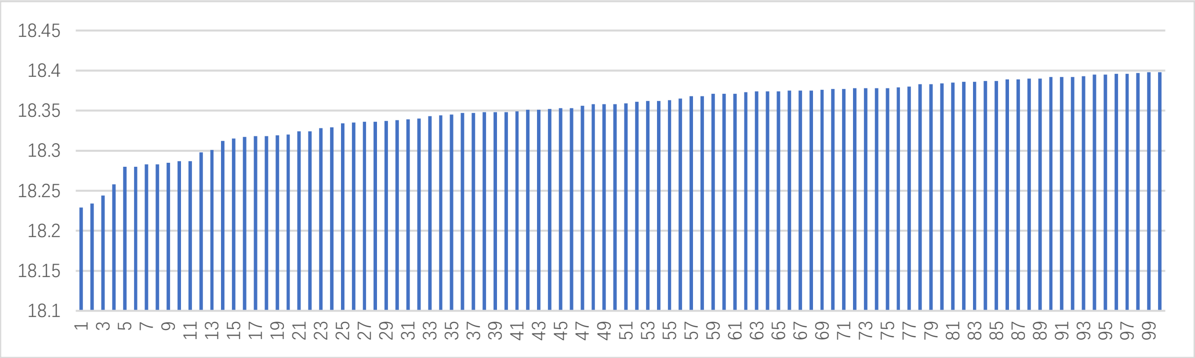

Figure 5: The information entropy of preference for electric vehicle specifications.

4.3 Clustering Analysis of Electric Vehicle Specification Combination Based on Information Entropy

In the electric vehicle design process, different electric vehicle design strategies are devised for different market segments to satisfy different customer preferences. Different models of electric vehicles are designed differently, so the customer groups are different. The information entropy of the electric vehicle specification combinations is also different.

Based on the content and extent of electric vehicle customer preferences, the electric vehicle specification combinations are clustered by the information entropy. Customers who purchase similar electric vehicles have similar preferences, so they are considered having similar customer preferences that are grouped into the same cluster [3]. Customers who purchase different types of electric vehicles have different preferences and they are grouped into the different clusters. Clustering specification combinations of electric vehicles based on information entropy can provide insight into the distribution of customer preferences for electric vehicles. According to the clustering results, the specification combination of electric vehicles with the similar information entropy and similar specifications are designed as the same electric vehicle to avoid homogeneous competition effectively.

Density-based clustering is an algorithm that determines the clustering structure based on the closeness of the sample distribution [8]. Similar electric vehicle designs are identified by density clustering of electric vehicle specification combinations. The threshold value of density-based clustering corresponds to the information entropy size of the electric vehicle specification combination. In this paper, the threshold value of density clustering corresponding to the information entropy of electric vehicle specification combinations is set to 0.6. Electric vehicles that satisfy the information entropy threshold and are similar in distance within the specification range are classified into the same category.

For specifications that use non-continuous values whose specification values are physically or mathematically unrelated to each other (e.g., battery type and brake type), it is not possible to quantify their design similarity based on the specification discrete values. Therefore, this research requires that the discrete values of electric vehicle specifications are identical to determine the similarity of the specification combinations.

Specification	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}

Cluster																		
Cluster 1	5	3	4	3	3	4	4	3	2	3	4	1	3	4	5	3	3	3
	5	3	4	4	3	3	2	3	2	2	4	2	2	4	5	3	3	3
	5	3	4	4	4	4	2	3	2	2	5	1	2	4	5	3	3	3
Cluster 2	5	5	3	3	3	5	4	3	3	2	5	1	2	4	5	3	3	2
	5	5	3	4	4	5	4	3	3	2	4	2	1	4	5	3	3	2
Cluster 3	4	4	3	3	3	5	3	2	3	3	4	2	3	4	5	4	3	3
	4	5	3	2	3	4	3	2	3	2	3	2	2	4	5	4	3	3

Table 5: Clustering results.

By clustering the preference information entropy, the electric vehicle specifications are divided into three clusters, and the customer preference market is divided into three segments. The specification combination with the smallest information entropy in each cluster is selected as the design requirement for the electric vehicle specification in the corresponding market segment. The corresponding specification combinations are shown in Table 5.

4.4 Probabilities of Specification Operations for Electric Vehicles

The characteristics of electric vehicles are mainly represented by the specifications of electric vehicles in the form of numerical values. The evolution of electric vehicles is the reproduction and mutation of electric vehicle specifications to form new electric vehicle specifications [3]. Therefore, by analyzing the electric vehicle specifications, the evolution mode and evolution law of electric vehicle specifications are obtained. The electric vehicle specification reproduction mainly means that the electric vehicle retains some features of the historical model in the design process. The reproduction of electric vehicle specification does not change the structural parameters of the specification, such as the chassis height, maximum weight, and number of seats. The electric vehicle specification mutation is the process by which a specification of electric vehicles is created from scratch. The specification mutation makes fundamental changes to the electric vehicle specification values, resulting in new electric vehicle specification values, such as the battery capacity, maximum motor power, and advanced driver assistance system.

Based on the reproduction and mutation of electric vehicle specifications, the reproduction and mutation probabilities of electric vehicle specifications can be decided based on the proportion of electric vehicles sales with reproduced and mutated specifications to the total number of electric vehicle sales [33]. In the data preparation process, the sale and specification of electric vehicles are prepared. Finally, the reproduction and mutation probabilities of electric vehicle specifications in each time are obtained. According to results for reproduction and mutation probabilities of electric vehicle specifications, a map for reproduction and mutation probabilities of electric vehicle specifications can be formed as shown in Figure 6.

The reproduction operations of electric vehicle specifications have the characteristic of electric vehicle specifications that remain unchanged. Mutation operations of electric vehicle specifications have the characteristic of electric vehicle specification changes. According to the operation probability calculation method, the reproduction and mutation probabilities of each electric vehicle specification during all the phases can be decided [32-33]. The results are shown in Table 6. The table shows that the reproduction probability of the electric vehicle specifications is high, such as the length, width, height, chassis height, maximum weight, and number of seats. The high reproduction probability of electric vehicle specifications indicates that these specifications change less during the design evolution and the customer requirement for that part of the electric vehicle specification changes less. The mutation probability of electric vehicle specifications is high, such as the NEDC electric range, advanced driver assistance system, battery type, battery capacity, and maximum motor power.

Figure 6: Thermodynamic diagram of the electric vehicle specification evolution: (a) Reproduction, (b) Mutation.

The higher mutation probability of electric vehicle specifications indicates that the possibility of changes in the specifications is higher and the customer requirements for this part of electric vehicle specifications are more variable.

4.5 Probability of Component Operations for Electric Vehicles

4.5.1 Relationship between specifications and components of electric vehicles

Electric vehicle specifications and components are interrelated, and functions expressed in the specifications are implemented through the component design [20]. There are two types of relationships between electric vehicle specifications and components: related and unrelated. Related relations mean that a change in one of the electric vehicle specifications or components will cause a change in the other. Unrelated relations mean that when one of the electric vehicle specifications or components change, the other does not change. Therefore, this research constructs a relationship model based on the dependency between electric vehicle specifications and components. The relationship between specifications and components is set to "1" if there is an influence relationship, or "0" if there is no influence relationship. The relationship between specifications and components of electric vehicles are shown in Table 7.

Specification	*Reproduction probability*	*Mutation probability*	*Specification*	*Reproduction probability*	*Mutation probability*
S_1	0.8942	0.1058	S_{10}	0.8744	0.1256
S_2	0.8901	0.1099	S_{11}	0.5704	0.4296
S_3	0.8892	0.1108	S_{12}	0.8635	0.1365
S_4	0.8626	0.1374	S_{13}	0.9773	0.0227
S_5	0.8824	0.1176	S_{14}	0.8984	0.1016
S_6	0.6431	0.3569	S_{15}	0.9006	0.0994
S_7	0.818	0.182	S_{16}	0.6373	0.3627
S_8	0.5627	0.4373	S_{17}	0.8777	0.1223
S_9	0.8405	0.1595	S_{18}	0.8325	0.1675

Table 6: Specification probability of electric vehicle.

	Length	*Width*	*Height*	…	*Driver assistance system*	*Parking brake*	*Price*
Motor	0	0	0	…	0	0	1
Transmission	0	0	0	…	0	0	0
Retarder	0	0	0	…	0	0	0
Rim	0	0	1	…	0	0	0
…	…	…	…	…	…	…	…
High voltage battery charging module	0	0	0	…	1	0	0
Lane keeping assist	0	0	0	…	1	0	1
Road traffic identification system	0	0	0	…	1	0	1
Active brake/active safety system	0	0	0	…	1	0	1

Table 7: The relationship matrix between specifications and components of electric vehicle.

4.5.2 Reproduction and mutation probabilities of electric vehicle components

Since there is a correlation between the specification and component, the probability of a component with the relation to the specification can be obtained based on the probability of electric vehicle specifications [32-33]. There is also reproduction probability and mutation probability of the electric vehicle component. If a reproduction probability of electric vehicle specifications is high, the corresponding electric vehicle component has less changes. The high reproduction probability of electric vehicle specifications means that the component corresponding to the specification is retained or changed less during the design process. Therefore, this type of electric vehicle component can be mass-produced to reduce production costs. The high mutation probability of electric vehicle specifications means that the possibility of changing the component corresponding to the specification is high, and the component requires adapting to the market changes during the design process.

Component	*Reproduction probability*	*Mutation probability*	*Component*	*Reproduction probability*	*Mutation probability*	*Component*	*Reproduction probability*	*Mutation probability*
C_1	0.8339	0.1661	C_{10}	0.8904	0.1096	C_{19}	0.7687	0.2313
C_2	0.8489	0.1511	C_{11}	0.8758	0.1242	C_{20}	0.7775	0.2225
C_3	0.8476	0.1525	C_{12}	0.8576	0.1424	C_{21}	0.7402	0.2598
C_4	0.9069	0.0931	C_{13}	0.7573	0.2427	C_{22}	0.8025	0.1975
C_5	0.8725	0.1275	C_{14}	0.7825	0.2175	C_{23}	0.7775	0.2225
C_6	0.8994	0.1007	C_{15}	0.8725	0.1275	C_{24}	0.7349	0.2651
C_7	0.8994	0.1007	C_{16}	0.8077	0.1923	C_{25}	0.7349	0.2651
C_8	0.8931	0.1069	C_{17}	0.8592	0.1408	C_{26}	0.7349	0.2651
C_9	0.7702	0.2298	C_{18}	0.7349	0.2651	C_{27}	0.7349	0.2651

Table 8: Component probability of electric vehicle.

4.6 Electric Vehicle Evolution Analysis for Design Adaptation

4.6.1 Component clustering based on the component probability of electric vehicles

Component cluster by agglomerative clustering is based on the reproduction and mutation probabilities of electric vehicle components. Each component is considered as a class. The probability differential value between electric vehicle components is calculated, and components with the smallest value of probability differential are clustered into a new class. This research sets the agglomerative clustering threshold to 0.03. Components are clustered according to the component reproduction probability results, to finally obtain six component classes. The results are shown in Table 9.

Cluster	*Components*	*Probability*
1	Rim, Differential, Front suspension, Rear suspension, Frame, Tyre, Seat, Inverter	0.8887
2	Retarder, Brake, DC/DC Converter	0.8533
3	Motor	0.8339
4	Parking radar, Motor controller, Brake auxiliary device	0.7768
5	Battery, Charger, Power distribution, Transmission, Traction control, Body stability control	0.7702
6	LCDA, High voltage charging module, Lane departure warning, Lane keeping assist, Road traffic identification, Active brake / Active safety	0.7358

Table 9: Agglomerative clustering results.

4.6.2 Adaptable design of the component modular for electric vehicles

Electric vehicle specifications are physically realized by components. A relationship matrix is formed based on the dependency between specifications and components. For different customer preferences represented by three market segments, recommendations of the module division are based on the relationship between specifications and components. In this research, recommendations for the modular division of components are shown in Table 10.

Comparisons	*Design recommendations*	
	Mutation module	*Reproduction module*
Similarities	C_2, C_{16} and C_{19} are recommended to be designed into a Mutation module. These components are related to S_4, S_5, S_6 and S_7. This is same as the current design. C_{14}, C_{18}, C_{20}, C_{22}, C_{23}, C_{24}, C_{25}, C_{26} and C_{27} are recommended to be designed into a Mutation module. These components are related to S_{17}. This is same as the current design.	Reproduction module is suggested for C_6, C_7, C_8 and C_{11}. These components are related to S_1, S_2, S_3, S_{14} and S_{15}. This is same as the current design. Reproduction module is suggested for C_3, C_{12} and C_5. These components are related to S_{17}. This is same as the current design. Reproduction module is suggested for C_4 and C_{10}. These components are related to S_{13}. This is same as the current design.
Differences	C_{13}, C_{17} and C_{21} are recommended to be designed into a Mutation module. These components are related to S_{10}, S_{11} and S_{12}. In the current design, these components are designed into a Reproduction module.	Reproduction modules are suggested for C_1 and C_9. These components are related to S4, S_5, S_6, S_8 and S_9. In the current design, these components are designed into a Mutation module.

Table 10: Design recommendations and comparisons with current electric vehicle design.

4.6.3 Similarities and differences between the recommended design and current design

Components in the mutation module:

- Same components include the transmission, motor controller, power distribution and components related to advanced driver assistance system. They are same as the current design for the product modularity and adaptability.
- Different components are the charger module related to charging time and battery warranty. In the current design, these components are designed as a reproduction module, charging time is changed by the battery capacity.

Components in the reproduction module:

- Same components are the front suspension, rear suspension, frame and seat for the vehicle length, width, height and weight. Rim and tyre are related to the chassis height. these components are designed into a reproduction module.
- Different components are motor and battery. Specifications affected by components can be achieved with other components. The components designed into reproduction modules can reduce production costs.

5 CONCLUSIONS

Based on the big sales data of electric vehicles, this paper quantifies customer preferences using the information entropy. A relationship model between the specification combinations and sales volume is built. Customer preferences are predicted for specification combinations to satisfy physical constraints. Mutation and reproduction probabilities of electric vehicle components are calculated based on the relationship between specifications and components. By considering the operational probabilities of components and their physical connections, design recommendations of module divisions and their feature identification are provided for enhancing electric vehicle adaptability. Comparisons between the current design and recommended design are also provided to illustrate the rationality of the proposed method. There are following advantages by using the newly developed method.

- Since customer preferences are related to the purchasing behavior, the research of the customer preferences based on big sales data provides a comprehensive and objective understanding of customer preferences for the electric vehicle design.
- Based on the big sales data, the relationship between specifications and customer preferences is searched by training a neural network. The mapping from customer preferences in customer domain into specification combinations in functional domain explores the electric vehicle design to satisfy the customer preference demand.
- The reproduction and mutation probabilities of electric vehicle components are obtained based on the relationship between electric vehicle specifications and components and the evolutionary probability of specifications.
- Adaptability of electric vehicle to diversified and changeable customers preferences on specification combinations can be significantly enhanced through identifying mutation and reproduction modules.

Despite the progress, a number of factors need to be carefully addressed as the accuracy of results is highly sensitive to following factors: (1) data collection and pre-processing for preference information mining and (2) determinations of both specification similarity threshold and component clustering threshold. In addition, design recommendations of product modularity and adaptability need to be carefully examined as other factors, such as manufacturing and operational constraints, should also be considered. Nevertheless, emerging big sales data provide valuable resources for supporting competitive product design and development. Our ongoing research activities include the game theory-based specification optimization method for enhancing product competitiveness using big sales data, and software tools development for the intelligent design and management of various modules using big sales data.

6 ACKNOWLEDGEMENTS

The authors wish to thank the National Key R&D Program of China (No: 2018YFB1701701), for providing financial support to this research.

Xiaojie Gao, http://orcid.org/0009-0002-1227-1839
Rui Li, http://orcid.org/0009-0003-7165-6568
Deping Chen, http://orcid.org/0009-0004-3962-6040
Jian Zhang, http://orcid.org/0000-0002-4658-3630
Qingjin Peng, http://orcid.org/0000-0002-9664-5326

REFERENCES

[1] Adeoye-Olatunde, O. A.; Olenik, N. L.: Research and scholarly methods: Semi-structured interviews, Journal of the American College of Clinical Pharmacy, 4(10), 2021, 1358-1367. https://doi.org/10.1002/jac5.1441

[2] Albers, A.; Bursac, N.; Scherer, H.; Birk, C.; Powelske, J.; Muschik, S.: Model-based systems engineering in modular design, Design Science, 5, 2019, E17. https://doi.org/10.1017/dsj.2019.15

[3] Balakrishnan, P. V.; Jacob, V. S.: Genetic algorithms for product design, Management Science, 42(8), 1996, 1105-1117. https://doi.org/10.1287/mnsc.42.8.1105

[4] Baxter, K.; Courage, C.; Caine, K.: Understanding your users: A practical guide to user research methods, Morgan Kaufmann, Burlington, US, 2015.

[5] Bloomfield, J.; Fisher, M. J.: Quantitative research design, Journal of the Australasian Rehabilitation Nurses Association, 22(2), 2019, 27-30. https://doi.org/10.33235/jarna.22.2.27-30

[6] Bonvoisin, J.; Halstenberg, F.; Buchert, T.; Stark, R.: A systematic literature review on modular product design, Journal of Engineering Design, 27(7), 2016, 488-514. https://doi.org/10.1080/09544828.2016.1166482

[7] Chatterjee, S.: Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents, Decision Support Systems, 119, 2019, 14-22. https://doi.org/10.1016/j.dss.2019.02.008

[8] Chen, Y.; Hu, X.; Fan, W.; Shen, L.; Zhang, Z.; Liu, X.; Du, J.; Li, H.; Chen, Y.; Li, H.: Fast density peak clustering for large scale data based on kNN, Knowledge-Based Systems, 187, 2020, 104824. https://doi.org/10.1016/j.knosys.2019.06.032

[9] China Association of Automobile Manufacturers (CAAM) [EB/OL]. http://www.caam.org.cn/tjsj.

[10] Dering, M. L.; Tucker, C. S.: A convolutional neural network model for predicting a product's function, given its form, Journal of Mechanical Design, 139(11), 2017, 111408. https://doi.org/10.1115/1.4037309

[11] Engel, A.; Browning, T. R.: Designing systems for adaptability by means of architecture options, Systems Engineering, 11(2), 2008, 125-146. https://doi.org/10.1002/sys.20090

[12] Gu, P.; Xue, D.; Nee, A. Y. C: Adaptable design: Concepts, methods, and applications, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 223(11), 2009, 1367-1387. https://doi.org/10.1243/09544054JEM1387

[13] Jiao, J.; Chen, C.: Customer requirement management in product development: A review of research issues, Concurrent Engineering, 14(3), 2006, 173-185. https://doi.org/10.1177/1063293X06068357

[14] Hayashi, Jr P.; Abib, G.; Hoppen, N.: Validity in qualitative research: A processual approach, The Qualitative Report, 24(1), 2019, 98-112. https://www.proquest.com/openview/c200412f0d74bd6a137110fb51805eeb/1?cbl=55152&loginDisplay=true&pq-origsite=gscholar.

[15] Hou, T.; Yannou, B.; Leroy, Y.; Poirson, E.: Mining customer product reviews for product development: A summarization process, Expert Systems with Applications, 132, 2019, 141-150. https://doi.org/10.1016/j.eswa.2019.04.069
[16] Huo, F.; Zhao, Y.; Chai, C.; Fang, F.: A user experience map design method based on emotional quantification of in-vehicle HMI, Humanities and Social Sciences Communications, 10(1), 2023, 1-10. https://doi.org/10.1057/s41599-023-01761-4
[17] Islam, M. A.; Aldaihani, F. M. F.: Justification for adopting qualitative research method, research approaches, sampling strategy, sample size, interview method, saturation, and data analysis, Journal of International Business and Management, 5(1), 2022, 1-11. https://rpajournals.com/jibm
[18] Kim, S.; Chang, J. J. E.; Park, H. H.; Song, S. U.; Cha, C. B.; Kim, J. W.; Kang, N.: Autonomous taxi service design and user experience, International Journal of Human–Computer Interaction, 36(5), 2020, 429-448. https://doi.org/10.1080/10447318.2019.1653556
[19] Klaib, A. F.; Alsrehin, N. O.; Melhem, W. Y.; Bashtawi, H. O.; Magableh, A. A.: Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies, Expert Systems with Applications, 166, 2021, 114037. https://doi.org/10.1016/j.eswa.2020.114037
[20] Koziolek, H.; Brosch, F.: Parameter dependencies for component reliability specifications, Electronic Notes in Theoretical Computer Science, 253(1), 2009, 23-38. https://doi.org/10.1016/j.entcs.2009.09.026
[21] Lihra, T.; Buehlmann, U.; Graf, R.: Customer preferences for customized household furniture, Journal of Forest Economics, 18(2), 2012, 94-112. https://doi.org/10.1016/j.jfe.2011.11.001
[22] Ma, S. C.; Fan, Y.; Guo, J. F.; Xu, J. H.; Zhu, J. N.: Analysing online behaviour to determine Chinese consumers' preferences for electric vehicles, Journal of Cleaner Production, 229, 2019, 244-255. https://doi.org/10.1016/j.jclepro.2019.04.374
[23] Nicoletti, L.; Mayer, S.; Brönner, M.; Schockenhoff, F.; Lienkamp, M.: Design parameters for the early development phase of battery electric vehicles, World Electric Vehicle Journal, 11(3), 2020, 47. https://doi.org/10.3390/wevj11030047
[24] Peng, Q.; Liu, Y.; Gu, P.: Improvement of product adaptability by efficient module interactions, International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 46353, 2014, V004T06A044. https://doi.org/10.1115/DETC2014-35183
[25] Petiot, J. F.; Grognet, S.: Product design: a vectors field-based approach for preference modelling, Journal of Engineering Design, 17(03), 2006, 217-233. https://doi.org/10.1080/09544820500275032
[26] Ridzuan, F.; Zainon, W. M. N. W.: A review on data cleansing methods for big data, Procedia Computer Science, 161, 2019, 731-738. https://doi.org/10.1016/j.procs.2019.11.177
[27] Sanguesa, J. A.; Torres-Sanz, V.; Garrido, P.; Martinez, F. J.; Marquez-Barja, J. M.: A review on electric vehicles: Technologies and challenges, Smart Cities, 4(1), 2021, 372-404. https://doi.org/10.3390/smartcities4010022
[28] Sawangjai, P.; Hompoonsup, S.; Leelaarporn, P.; Kongwudhikunakorn, S.; Wilaiprasitporn, T.: Consumer grade EEG measuring sensors as research tools: A review, IEEE Sensors Journal, 20(8), 2019, 3996-4024. https://ieeexplore.ieee.org/abstract/document/8945233
[29] Schräder, T.; Stolte, T.; Jatzkowski, I.; Graubohm, G.; Maurer, M.: An approach for a requirement analysis for an autonomous family vehicle, IEEE Open Journal of Circuits and Systems, 2019, 1597-1603. https://ieeexplore.ieee.org/abstract/document/8813882
[30] Timoshenko, A.; Hauser, J. R.: Identifying customer needs from user-generated content, Marketing Science, 38(1), 2019, 1-20. https://doi.org/10.1287/mksc.2018.1123
[31] Zhang, J.; Xue, D.; Gu, P.: Adaptable design of open architecture products with robust performance, Journal of Engineering Design, 26(1-3), 2015, 1-23. https://doi.org/10.1080/09544828.2015.1012055
[32] Zhang, J.; Xie, G.; Peng, Q.; Gu, P.: Product evolution analysis for design adaptation using big sales data, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of

Engineering Manufacture, 237(8), 2023, 1241-1253. https://doi.org/10.1177/09544054221122844
[33] Zhang, J.; Lin, P.; Simeone, A.: Information mining of customers preferences for product specifications determination using big sales data, Procedia CIRP, 109, 2022, 101-106. https://doi.org/10.1016/j.procir.2022.05.221
[34] Zhao, S.; Zhang, Q.; Peng, Z.; Fan, Y.: Integrating customer requirements into customized product configuration design based on Kano's model, Journal of Intelligent Manufacturing, 31(3), 2020, 597-613. https://link.springer.com/article/10.1007/s10845-019-01467-y