Title:
Customer preference mining of electric vehicles for design decision-making using big sales data

Authors:
Xiaojie Gao, GGDgxj303682212@mail2.gdut.edu.cn, Guangdong University of Technology
Rui Li, 22rli1@stu.edu.cn, Shantou University
Deping Chen, cdp152418@126.com, Shantou University
Jian Zhang, jianzhang@stu.edu.cn, Shantou University
Qingjin Peng, qingjin.peng@umanitoba.ca, University of Manitoba

Keywords:
Electric vehicle, Customer preference, Design decision, Big sales data, Information entropy

DOI: 10.14733/cadconfP.2023.68-73

Introduction:
To increase the market sharing of electric vehicles, it is essential to improve efficiency of the vehicle design through the fast and accurate identification of customer preferences [4]. Product design needs accurate information of customer preferences for diverse market requirements. Existing studies on the customer preference are based on quantitative and qualitative analyses of customer preference data. The data are mainly collected through questionnaires, interviews, market surveys, and consumer evaluation for consumer preferences of vehicles, including the usage consumption, after-sales service, purchase cost, and vehicle quality [2-3]. These studies mainly consider consumer preferences without mapping the preferences to design specifications for the vehicle design improvement.

In the age of the Internet and e-commerce, there is a large amount of sales data available in webpages of vehicle companies, including specifications, parameters, prices and sales volume. It is difficult to obtain accurate and comprehensive customer preferences based on the context of traditional methods. The big sales data of vehicles provides new solutions for the customer preference research [1],[5]. 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. 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. This paper uses the big sales data of electric vehicles in the Chinese market in 2021 as the research object.

Main Idea:
Implementation of Preference information mining and design decision of electric vehicles
Based on our previous work [6], 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 Fig. 1.

- 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 analyze the module type of electric vehicle components by the amount of information of a single specification for the design decision of components that affect the electric vehicle specifications.

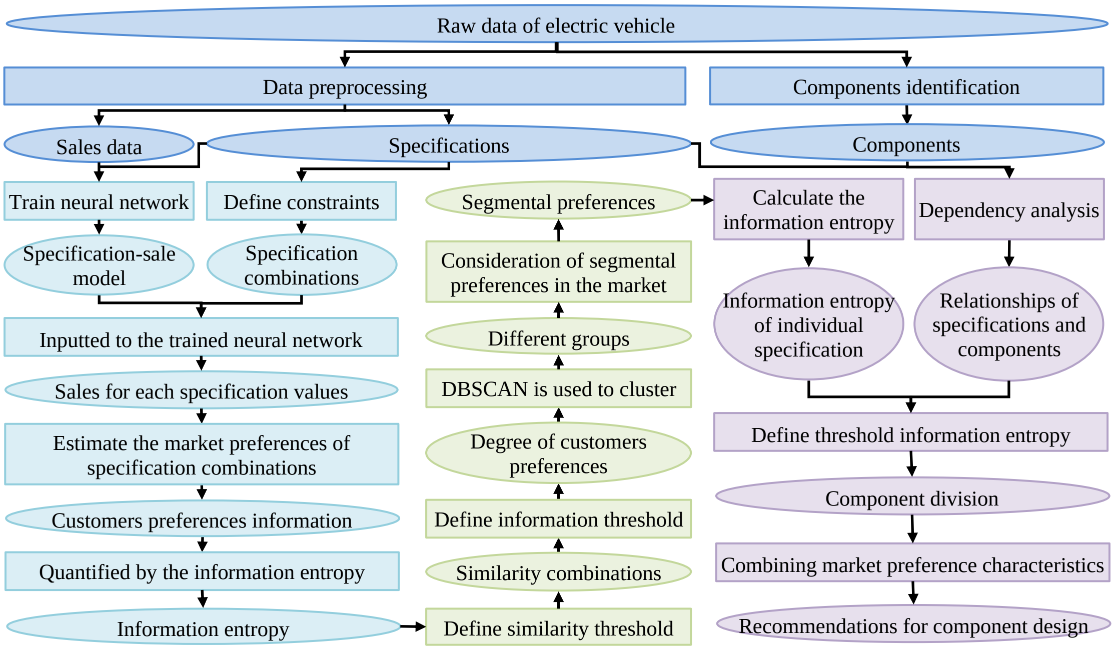

Fig. 1: Flow chart of preference information mining and product design decision of electric vehicles.

Dataset preparation of electric vehicles

Specifications of electric vehicles, such as the price, sales volume, are collected via the Internet. Data preprocessing is performed on incomplete and noise data through methods such as data cleaning and data transformation. The number of selected electric vehicles is reduced from 1482 to 430 to form a complete matrix dataset, where each row represents an electric vehicle, each column represents a specification or sales volume, and cells in the matrix are values of corresponding specifications for each electric vehicle.

After filtering and dimensionality reduction of the data, 18 electric vehicle specifications are finally retained, including specifications of the electric vehicle battery, motor, autonomous driving and the whole vehicle. In order to decide the probability of market preferences for different specifications, the vehicle specifications are discretized. Specifications with continuous values are divided to intervals, such as length and width into 5 intervals, height into 4 intervals, and selling price into 6 intervals. Specifications with non-continuous values are used as discrete values, such as the battery type, advanced driver assistance system and battery pack warranty. The final division results of all the retained electric

Proceedings of CAD'23, Mexico City, Mexico, July 10-12, 2023, 68-73

vehicle specifications are shown in Tab. 1. 27 key components of electric vehicles are finalized as shown in Tab. 2.

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}	Maximum number	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

Tab. 1: Electric vehicle specifications and their value ranges.

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

Tab. 2: Different types of components and their IDs of electric vehicles.

Market preferences based on the relationship model

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. The combination of 18 electric vehicle specifications is used as the input, and the corresponding sales volume is used as the output of the model. The training is until an expected result obtained. The method is repeated 10 times using the Bayesian model. Based on the result comparison, the best validation performance is 3.8767 at epoch 7.

Sales data of electric vehicles represent only the existing customer preference information. The new electric vehicle specifications are not included in the sales data. 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. 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.

Market segments through the cluster analysis of specification combinations
The electric vehicle specification combinations are clustered by Density-Based Spatial Clustering of Applications with Noise (DBSCAN). When the information entropy of the specification combinations is lower than 0.6, the specifications are clustered into one group. For non-contiguous specifications, such as the battery, motor, and parking brake, discrete values are identified to determine the combination of similar specifications. For continuous specifications, such as the acceleration, electric range and motor power, the absolute value of differences between discrete values is set as 1. Partial results of combinations of the clusters are shown in Tab. 3.

ID \ *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 1	1	5	3	4	3	3	4	4	3	2	3	4	1	3	4	5	3	3	3
	2	5	3	4	4	3	3	2	3	2	2	4	2	2	4	5	3	3	3
	3	5	3	4	4	4	4	2	3	2	2	5	1	2	4	5	3	3	3
Cluster 2	4	5	5	3	3	3	5	4	3	3	2	5	1	2	4	5	3	3	2
	5	5	5	3	4	4	5	4	3	3	2	4	2	1	4	5	3	3	2
Cluster 3	6	4	4	3	3	3	5	3	2	3	3	4	2	3	4	5	4	3	3
	7	4	5	3	2	3	4	3	2	3	2	3	2	2	4	5	4	3	3

Tab. 3: Clustering results.

Design of electric vehicle components based on the information quantity of specifications
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 paper, recommendations for the modular division of components are shown in Tab. 4.

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,	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

	these components are designed into a Reproduction module.	design, these components are designed into a Mutation module.

Tab. 4: Design recommendations and comparisons with current electric vehicle design.

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. These specifications have the same customer preferences in cluster 1, so 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.

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. The module design decisions are made based on the relationship between specifications and components. There are following advantages by using the information entropy to quantify customer preferences.

- 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.
- Based on existing electric vehicle specifications, the combination of specifications within the range of existing specifications is presented in its entirety. The specification combinations that conform to the physical laws can be filtered out by constraints.

References:

[1] Boone T, Ganeshan R, Jain A, et al. Forecasting sales in the supply chain: Consumer analytics in the big data era. International Journal of Forecasting, 2019, 35(1): 170-180. https://doi.org/10.1016/j.ijforecast.2018.09.003

[2] Liao F, Molin E, van Wee B. Consumer preferences for electric vehicles: a literature review. Transport Reviews, 2017, 37(3): 252-275. https://doi.org/10.1080/01441647.2016.1230794

[3] Ling Z, Cherry C R, Wen Y. Determining the Factors That Influence Electric Vehicle Adoption: A Stated Preference Survey Study in Beijing, China. Sustainability, 2021, 13(21): 11719. https://doi.org/10.3390/su132111719

[4] 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, 2021, 4(1): 372-404. https://doi.org/10.3390/smartcities4010022

[5] Zhang J, Xie G, Peng Q, et al. Product evolution analysis for design adaptation using big sales data. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2022. https://doi.org/10.1177/09544054221122844

[6] Zhang J, Lin P, Simeone A. Information mining of customers preferences for product specifications determination using big sales data, Procedia CIRP, 2022, 109:101-106 https://doi.org/10.1016/j.procir.2022.05.22

Proceedings of CAD'23, Mexico City, Mexico, July 10-12, 2023, 68-73