frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fenrg.2022.958885

Demand-driven NEV supplier selection: An integrated method based on ontology–QFD–CBR

Jin Zheng[1,2], Yong-Hai Li[1]* and Zhi-Ping Fan[3,4]

[1]School of Management, Henan University of Technology, Zhengzhou, China, [2]School of Business, Liaoning University, Shenyang, China, [3]School of Business Administration, Northeastern University, Shenyang, China, [4]National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China

EDITED BY
Helena Martín,
Universitat Politecnica de Catalunya,
Spain

REVIEWED BY
Gerard Escudero,
Universitat Politecnica de Catalunya,
Spain
Juan José Mesas,
Universitat Politecnica de Catalunya,
Spain

*CORRESPONDENCE
Yong-Hai Li,
yonghai.li@163.com

SPECIALTY SECTION
This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research

RECEIVED 01 June 2022
ACCEPTED 24 October 2022
PUBLISHED 09 January 2023

CITATION
Zheng J, Li Y-H and Fan Z-P (2023), Demand-driven NEV supplier selection: An integrated method based on ontology–QFD–CBR. *Front. Energy Res.* 10:958885.
doi: 10.3389/fenrg.2022.958885

With the rapid development of new energy vehicles (NEVs), the market competition in the NEV industry is becoming increasingly fierce. Selecting the right supplier has become a critical aspect for NEV manufacturers. Therefore, based on the user's demand information, selecting a suitable NEV supplier to support the NEV manufacturer's management decision is a noteworthy research problem. The purpose of this study is to develop an integrated method for demand-driven NEV supplier selection based on ontology–quality function deployment (QFD)–case-based reasoning (CBR). The method is composed of three parts: 1) construction of domain ontology of NEV component supplier selection criteria based on text information mining; 2) extraction of demand attributes and determination of their weight based on latent Dirichlet allocation (LDA) and Kano model, as well as determination of expected attributes and their weights based on QFD; and 3) selection of an NEV component supplier based on CBR. To illustrate the use of the proposed method, an empirical study on the supplier selection of the XP NEV manufacturer is given. This method is helpful in selecting the most suitable component supplier for NEV manufacturers and relevant decision-makers.

KEYWORDS

new energy vehicle, supplier selection, ontology, quality function deployment, case-based reasoning, users' demands

1 Introduction

The transportation sector accounts for 24% of global CO_2 emissions, while fuel vehicles are an important source of greenhouse gases and pollutants (Wang et al., 2017; Luo et al., 2021). Compared with the fuel vehicle, the new energy vehicle (NEV) has significantly higher fuel efficiency and lower or even zero CO_2 emissions (Teixeira and Sodre, 2018; Liu et al., 2021). Studies have shown that plug-in hybrid electric vehicles and hybrid electric vehicles can reduce CO_2 emissions by about 30%, while in areas with a high proportion of hydro power, pure electric vehicles can reduce CO_2 emissions by 90% (Wang et al., 2022; Yu et al., 2022). NEVs have the advantage of being green, low carbon, energy saving, and convenient (Cano et al., 2018; Xie et al., 2020; Hao et al., 2021). As an effective solution to face the global challenges of environmental pollution and energy shortages caused by traditional fuel vehicles, it has attracted the attention of governments all over the world (He 2016; Bastida-Molina et al., 2021). Governments have proposed various policies and invested a lot of money to support the promotion and application of NEVs (Adnan et al., 2017; Tan and Lin 2019). For example, the Chinese government has attached importance to the development of NEVs since the 1990s and listed NEVs as one of the strategic emerging industries in 2010 (Gong et al., 2013; Li and Jing, 2019; Tan et al., 2021). With the support of various policies of the Chinese government (such as purchase subsidies, special energy conservation, and emission reduction fund), the territory of the NEV industry has continued to expand, and 28 provinces in China have joined the ranks of NEV manufacturing (Gong et al., 2013; Lu et al., 2017; He et al., 2018; Kendall 2018). In 2020, China added more than 68,000 NEV-related enterprises, an increase of 85% compared with 2019. Meanwhile, NEV sales in China have also grown rapidly. In 2014, NEV sales have quadrupled compared with 2013. In 2020, NEV sales were 18.2 times those of 2014, accounting for 41% of the global NEV sales (Hao et al., 2021; Liu et al., 2021). With the rapid growth of NEVs, the competition in the NEV industry is becoming increasingly fierce, and NEV manufacturers are facing more and more challenges. In this case, it is difficult for NEV manufacturers to survive and develop entirely on their own research and development (R&D) investments and innovation capabilities (Huth et al., 2015; Lu et al., 2020). Therefore, the development of cooperation between NEV manufacturers and NEV component suppliers has become an inevitable trend of development (Fan et al., 2020). However, some procurement departments of NEV manufacturers have not fully introduced the supplier competition mechanism, and there are many problems such as an imperfect supplier evaluation system, strong subjectivity, and lack of consideration of user demands. For this, how to select the appropriate supplier from a large number of NEV component suppliers with varying capabilities and levels of the market has become an urgent problem for NEV manufacturers.

As a starting point in the NEV industry supply chain, the selection of an NEV component supplier directly determines the cost and quality of the whole supply chain and even the benefits of the enterprise (Hosseini and Sarder, 2019). Choosing a suitable NEV component supplier can effectively improve the performance, customer satisfaction, and overall benefits of NEV manufacturers; while the improper selection of a component supplier may lead to the loss of time, cost, and market share of NEV manufacturers (Büyüközkan and Çifçi, 2011; Ayağ and Samanlioglu, 2014; Eydi and Fazli, 2019; Hosseini and Sarder, 2019; Wang et al., 2021). For example, in 2016, due to potential safety hazards in Samsung SDI batteries, JAC had to temporarily suspend production of its pure electric vehicle IEV6S, resulting in JAC's sales of this vehicle being significantly lower than expected. In 2021, due to the risk of fire in the power battery provided by Funeng Technology, BAIC recalled more than 30,000 pure electric vehicles, EX360 and EU400, and suffered huge economic losses. Thus, for NEV manufacturers, it is very important to choose a suitable component supplier from multiple component suppliers, which is the guarantee of their success (Khan et al., 2016; Liu et al., 2019). However, NEV manufacturers are different from traditional automobile manufacturers in terms of technology, sales service, corporate culture, and marketing mode, and their selection requirements and criteria for component suppliers are also different from those of traditional automobile manufacturers. Compared with the selection of a traditional automobile component supplier, NEV manufacturers pay more attention to the component suppliers' R&D capabilities and technology, environmental protection capabilities, and supply nodes (Liu 2016; Xiao 2020). Meanwhile, for NEV manufacturers, NEV component suppliers are not only simple suppliers but also mutually beneficial partners, sharing information and coordinated development with them. Given this, it is difficult to apply the component supplier selection method of traditional automobile manufacturers to solve the selection problem of NEV manufacturers. Thus, it is necessary to develop a targeted method to select a suitable component supplier for NEV manufacturers.

At present, the studies on the selection of an NEV component supplier are still scarce, but some related research results can be found. These research results can be roughly divided into two aspects: one is the selection or evaluation criteria of suppliers in the automotive industry; the other is the supplier selection method in the automobile industry. Determining the appropriate criteria is an important part of the supplier selection decision. The selection criteria of suppliers in the automotive industry have been widely concerned by scholars. For example, Dargi et al. (2014) developed a framework to support the selection of semi-assembly component suppliers for automobile manufacturers, and they extracted key criteria for evaluating suppliers. Azizi et al. (2015) introduced the most important criteria and sub-criteria for selecting the best supplier in the automotive industry and used Fuzzy technology and TOPSIS to select suitable suppliers. Li et al. (2015) constructed an evaluation index system of component suppliers and proposed a supplier selection model so as to choose the best combination of component suppliers for automobile manufacturers. Lima-junior and Carpinetti (2016) divided the selection criteria of automobile clutch suppliers into priority level, critical level, complementary level, and costly level based on the method of fuzzy quality function deployment and information acquisition difficulty evaluation procedures so as to guide the final selection decision. Manelloa and Calabrese (2019) conducted an ex-post analysis of the main factors influencing the supplier selection process in the automotive industry. They add new empirical evidence by investigating the impact of corporate reputation factors in the supplier selection criteria. For the selection of a headlamp supplier, Jain et al. (2018) determined the selection criteria through literature review and interviews with industry experts and ranked the suppliers using AHP and TOPSIS. Mathiyazhagan et al. (2018) investigated and ranked the environmental criteria for green supplier selection in the Indian automobile industry on the basis of expert judgment and ranked the supplier selection criteria. You et al. (2018) summarized the general evaluation criteria of automobile suppliers through literature screening and found out the categories and key criteria that automobile manufacturers need to consider when selecting component suppliers. Dai and Zhu (2019) established an auto component supplier evaluation system that considers the three dimensions of quality, business, and development and comprehensive capabilities. They used the TOPSIS method to determine the weight of each index and judged the classification of suppliers. Gupta et al. (2019) proposed a framework for the selection of green suppliers in the automotive industry and formulated six environmental standards and three conventional standards. This framework could help decision-makers distinguish the selection criteria of green suppliers. Choosing an appropriate method is very important for the supplier evaluation. Currently, scholars have put forward some relevant methods and a variety of combination methods to evaluate and select suppliers. For example, Park and Lee (2017) proposed a supplier selection method for automotive chassis parts through a hybrid method of DEA and AHP. Dweiri et al. (2016) proposed a supplier ranking model for the automobile industry based on AHP. This method provided managers with insights into various factors that needed to be considered when selecting suppliers. Mou et al. (2018) constructed a sustainable auto-part supplier selection model and proposed a three-stage decision-making method based on a probability distribution-hesitant fuzzy linguistic set and group decision-making theory. Memari et al. (2019) proposed an intuitionistic fuzzy TOPSIS method to select the appropriate automotive catalytic converter supplier. Liu et al. (2019) proposed a fuzzy three-stage integrated MCDM method for sustainable NEV part supplier selection. Zhou et al. (2019) proposed a multi-objective optimization model of auto-part suppliers considering customer complaints, and the multi-attribute utility theory and linear weighting method were used to obtain the optimization results. Wang et al. (2021) proposed a battery supplier selection framework for electric vehicles based on MCDM. In this framework, the MULTIMOORA method is used to sort the alternatives. Wu et al. (2020) proposed a hybrid framework that combines interval type 2 fuzzy sets, k-fuzzy measure, and Choquet integral operator to select the optimal green supplier in the electric vehicle industry. Liu et al. (2021) proposed a hybrid fuzzy symmetric MCDM model, integrating fuzzy linguistic set, the best and worst method, prospect theory, and VIKOR. The model could help NEV manufacturers select innovative power battery suppliers. Ilyas et al. (2021) used the integrated MCDM method based on BWM and FTOPSIS to consider the risks associated with COVID-19 so as to support the selection of suppliers in the automotive industry.

From the abovementioned literature review, we find that the related research provides some methods and ideas that can be used as a reference for this paper, such as the construction of a criteria system for NEV supplier selection. However, relevant methods and technologies in the existing research can only solve part of the problem in the selection of an NEV supplier, and there are still certain limitations, given as follows: 1) most of the existing studies focus on the selection of a traditional automobile component supplier, while the relevant methods and technologies to solve the problem of NEV component supplier selection are still obviously scarce (Galankashi et al., 2016; Hendiani et al., 2020). The method and criteria of component supplier selection for traditional automobile manufacturers are different from those for NEV manufacturers. For example, environmental protection capability and collaborative capability are less considered in the selection process of the traditional automobile supplier. In addition, most of the existing studies only focus on partial problems in the process of supplier selection, rather than systematically solving the problem of supplier selection, and select the best supplier based on the comprehensive performance of suppliers, without considering the indispensable subjective preference of decision-makers. 2) In terms of problem-solving methods, most of the existing studies solve the supplier selection problem based on the MCDM method. However, some common MCDM methods have certain limitations, such as the AHP method is highly subjective, the ranking has irregularities, and the use of additive aggregation may cause information loss; in the TOPSIS method, Euclidean distance is used to measure the distance between two schemes, and the correlation of decision criteria is not considered; some assumptions in DEA may not always be correct in reality; the weight of decision criteria in the VIKOR method is determined by the decision-maker subjectively, and there is no paired comparison between decision criteria and alternatives (Konidari and Mavrakis, 2007; Noori et al., 2021). 3) In terms of the supplier selection decision, most of the existing studies retrieve supplier case information through keyword matching and attribute similarity calculation so as to select component suppliers, which may result in problems such as low retrieval efficiency and low retrieval accuracy due to the lack of consideration of the semantic relations among keywords or attributes (Zhao and Yu 2011). 4) In terms of supplier selection criteria and their weight determination, most of the existing studies use the cited frequency or questionnaires to determine selection criteria, failing to fully consider the logic and semantic relationship of selection criteria (for example, product quality and quality are synonymous), which easily affects the efficiency of the supplier selection decision (Tavana et al., 2021). In addition, the weights of selection criteria in the existing studies are mostly derived from the subjective preferences of decision-makers and fail to fully consider the demands of users or objective data information. 5) In terms of acquiring and processing users' demands, most of the existing studies collect users' demands through questionnaires or structured scales. This easily restricts the expression of users' demands, leading to insufficient consideration of users' demands. In addition, the existing studies usually only focus on sentiment words of different sentiment polarities expressed in users' demands or the proportions of different sentiment polarities (Balazs and Velásquez 2016) and fail to fully consider the different strengths of sentiment polarities. It is easy to cause the loss of information and one-sided sentiment consideration. 6) In terms of transforming user demand attributes into expected attributes toward suppliers, most of the existing studies only rely on the domain knowledge and experience of experts (Yang and Chai, 2018; Deng et al., 2021) and fail to integrate objective information, such as academic literature, industry standards, etc., which are highly subjective.

To overcome these limitations, it is necessary to develop a novel method to select a suitable component supplier for NEV manufacturers so as to enrich the relevant theory and method. Specifically, the purpose of this study is to solve the following three questions:

1) How to intelligently construct the ontology of selection criteria of the NEV component supplier so as to support the selection decision of component supplier for NEV manufacturers?
2) How to organize and mine user demand information concerning NEV and incorporate it into the weights of the supplier selection criteria so as to avoid information loss and reduce the interference of subjective factors in the calculation process?
3) How to realize the case information retrieval of component suppliers and improve the retrieval efficiency under the condition of considering the expectations of a decision-maker?

To address the abovementioned questions, we propose an integrated method for demand-driven NEV supplier selection based on ontology–quality function deployment (QFD)–case-based reasoning (CBR). In the method, the literature on the selection of the NEV component supplier is first preprocessed. Second, the domain concepts are extracted using the latent Dirichlet allocation (LDA), and the hierarchical agglomerative clustering (HAC) algorithm and association rules are used to extract the relations between domain concepts so as to construct the domain ontology of selection criteria of the NEV component supplier. Then, demand attributes are extracted from the demand documents provided by users with different demands using LDA, and the initial weights of demand attributes are determined based on the sentiment strength analysis algorithm and information entropy method. On the basis of this, the categories of demand attributes and the final weights of demand attributes are determined based on the Kano model and the Better-Worse coefficient. Furthermore, the expected attributes and their weights of NEV component supplier selection are determined based on QFD, and the semantic similarity of ontology concepts and attribute similarity are calculated based on CBR and ontology theory. Based on the hybrid similarity, the NEV component supplier is selected. Obviously, the proposed method enriches and develops the existing benchmark methods (such as the traditional sentiment analysis method and the traditional hybrid similarity calculation method) and has distinct characteristics and advantages in the construction of selection criteria for the NEV component supplier, the processing of user demand information, and the retrieval of NEV component supplier cases, which can help improve the rationality and efficiency of the selection decision of the NEV component supplier and address other similar issues (e.g., product design scheme optimization and case knowledge retrieval).

The remainder of this paper is organized as follows. Section 2 gives a description of the problem. In Section 3, the architecture of the demand-driven NEV supplier selection system is designed. Section 4 presents a solution framework and a method for demand-driven NEV supplier selection. In Section 5, an empirical study on the selection of a component supplier for the XP NEV manufacturer is given to illustrate the use of the proposed method. Finally, the conclusions of this study and the directions for future research are presented in Section 6.

2 Problem description

To clearly describe the problem of NEV supplier selection, the notations are defined as shown in Table 1.

The problem concerned in this study is how to construct the domain ontology of the selection criteria for NEV component suppliers, determine $E_b, d_a, \omega_b^{P'}$, and ω_a^E and select an appropriate NEV component supplier from $Z_1, Z_1, \ldots, Z_{q_\mu}$ according to D_j, P_k, $z_{\mu a}^P$, and z_a^o so as to support the NEV manufacturer's procurement decision.

3 Demand-driven NEV supplier selection system

To address the above mentioned problem, the architecture of a demand-driven NEV supplier selection system is designed, as shown in Figure 1. The architecture of the system is divided into four layers: the user layer, the application layer, the business logic layer, and the data access layer. The user layer is the communication link between the user and the system, which is used to connect the main objects served by the system. Different objects can access different functional modules and

TABLE 1 Notations frequently used throughout this paper.

Notation	Explanation
$D = \{D_1, D_2, \ldots, D_m\}$	Set of domain literature, where D_j denotes the j-th domain literature and m denotes the number of domain literature, $j \in \{1, 2, \ldots, m\}$
$C = \{C_1, C_2, \ldots, C_n\}$	Set of domain concepts, where C_i denotes the ith domain concept and n denotes the number of domain concepts, $i \in \{1, 2, \ldots, n\}$.
$P = \{P_1, P_2, \ldots, P_v\}$	Set of demand documents, where P_k denotes the k-th demand document and v denotes the number of demand documents, $k \in \{1, 2, \ldots, v\}$
$E = \{E_1, E_2, \ldots, E_I\}$	Set of demand attributes, where E_b denotes the b-th demand attribute and I denotes the number of demand attributes, $b \in \{1, 2, \ldots, I\}$
$DA = \{d_1, d_2, \ldots, d_\delta\}$	Set of expected attributes, where d_a denotes the a-th expected attribute and δ denotes the number of expected attributes, $a \in \{1, 2, \ldots, \delta\}$
$\omega_b^{P'}$	Final weight of the demand attribute E_b, such that $\omega_b^{P'} \geq 0$ and $\sum_{b=1}^{I} \omega_b^{P'} = 1$. The value of $\omega_b^{P'}$ can be obtained by adjusting ω_b^{P} (the weight of E_b), $b \in \{1, 2, \ldots, I\}$
ω_a^E	Weight of the expected attribute d_a, such that $\omega_a^E \geq 0$ and $\sum_{a=1}^{\delta} \omega_a^E = 1$
Z_μ	μ-th historical case, which is also the μ-th alternative NEV component supplier, $\mu \in U$, $U = \{1, 2, \ldots, q_\mu\}$
Z^o	Target case, which is also the expectation of the decision-maker concerning NEV component suppliers
$z_{\mu a}^P$	Attribute value of historical case Z_μ concerning the expected attribute d_a, $\mu \in U$ and $a \in \{1, 2, \ldots, \delta\}$
z_a^o	Attribute value of target case Z^o concerning the expected attribute d_a, $a \in \{1, 2, \ldots, \delta\}$

participate in the requirement survey and the NEV supplier selection and evaluation process. The application layer consists of eight critical functional components: domain concept extraction module, domain ontology management module, user demand acquisition module, QFD module, case retrieval module, system management module, model and information management module, and case management module. This layer provides a series of key procedures that enable the system to support the NEV supplier selection decision. The business logic layer mainly provides the interface for system operation, allowing different users to complete the corresponding operation of the system. This layer uses Protégé to define the domain ontology, extracts the abstract semantic elements of the domain ontology to represent the cases, and implements case retrieval by considering user demands to obtain the ranking results of the suppliers. The data access layer uses MySQL as the main database management system to realize the storage of data by category.

The brief explanation of each functional component is given as follows:

- Domain concept extraction module: the main function of domain concept extraction is to select relevant domain literature from the literature database and extract domain concepts and relationships between domain concepts based on text mining. Since the existing literature on NEV supplier selection and evaluation has the characteristics of large quantity, knowledge dispersion, and multi-source heterogeneity, an effective method to solve the sharing, exchange, and integration of heterogeneous information, that is, ontology, is adopted. According to relevant scientific literature, the selection and evaluation criteria of suppliers are determined based on ontology. Applying knowledge management practices to supply chain management may greatly improve organizational performance.
- Domain ontology management module: the main function of domain ontology management is to model and maintain the domain knowledge about NEV supplier selection criteria so as to make the constructed domain ontology more comprehensive and complete. The quality of the domain ontology affects the determination of the expected attributes and the selection decision of the NEV supplier, so the setup of the domain ontology management module is necessary.
- User demand acquisition module: the main function of user demand acquisition is to collect users' demands in the form of network documents, extract demand attributes from the demand documents, and use user demands as reference factors for NEV supplier selection. For NEV manufacturers, focusing on user demands can improve their competitive and strategic advantages. Therefore, the acquisition and analysis of user demands are important in supplier selection. In addition, collecting user requirements in the form of network documents can enable users to fully express their demands and make the obtained demand information more comprehensive.
- QFD module: the main function of the QFD module is to investigate and analyze user demands for NEV through a systematic approach and convert them into expected attributes concerning supplier selection. Full consideration of user demands is an effective way to rationally select NEV suppliers. It is helpful for NEV manufacturers to improve product quality and increase customer satisfaction by considering user demands to make an NEV supplier selection decision.
- Case retrieval module: the main function of case retrieval is to search the case base by using the similarity calculation method according to the relevant description of the alternative NEV supplier and then obtain the ranking of the alternative NEV suppliers based on the calculation results. Case retrieval is the most important step in the process of CBR, which can realize the real-time retrieval and related queries of cases in the case base.
- System management module: the main function of system management is user management and system settings. The system divides users into five categories based on permissions. An ordinary user can participate in market

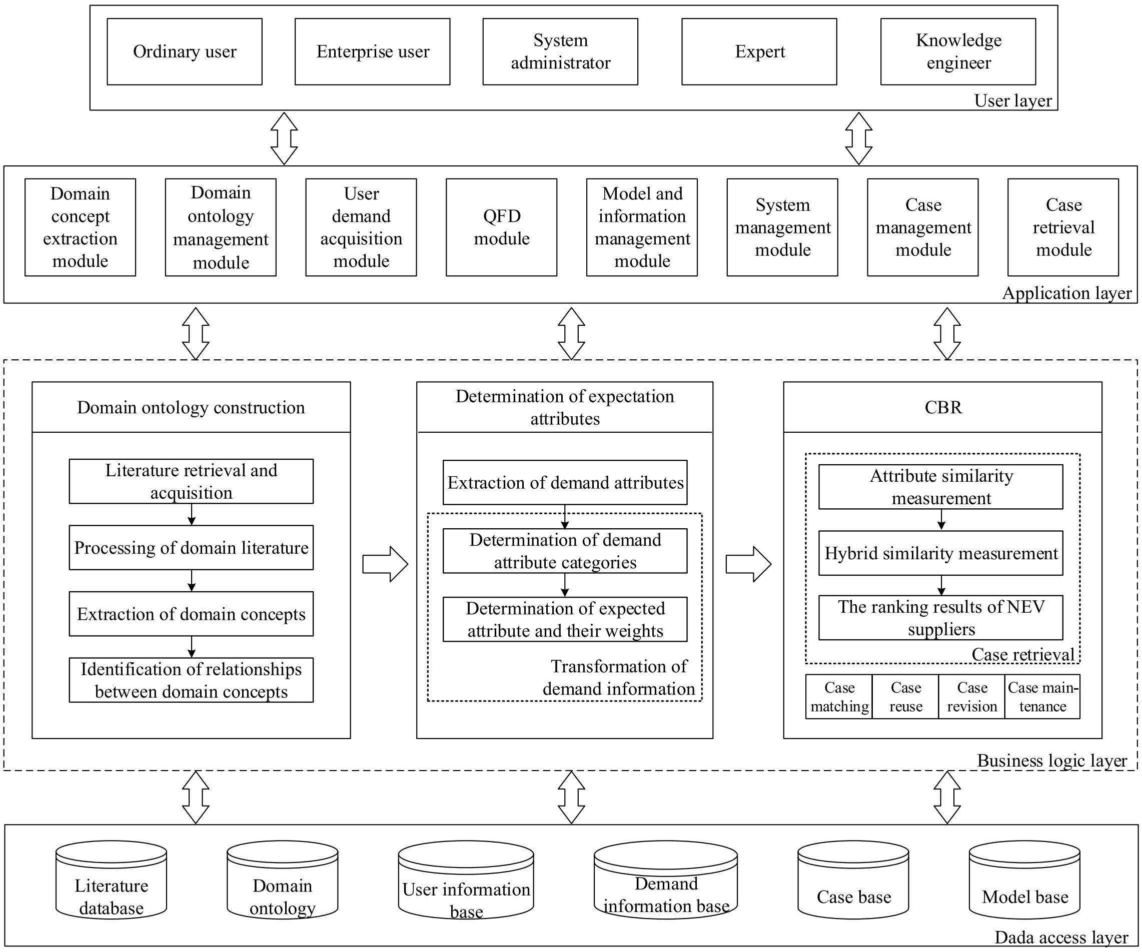

FIGURE 1
Architecture of the demand-driven NEV supplier selection system.

research and provide demand documents. An enterprise user can determine the case retrieval conditions according to the terms in the domain ontology and the current case, search the case library, and obtain the ranking results of the NEV supplier. Experts and knowledge engineers can assist in building domain ontology based on the processing results of text information, building case bases according to domain ontology and user demands, and managing and maintaining case bases. The system administrator is responsible for managing user permissions.

- Model and information management module: the information in the databases is the basis for implementing the system functions. The main functions of model and information management are to construct and update various calculation formulas in the model base and update and maintain all types of data in the databases.
- Case management module: the main function of case management is to build the case representation and retrieval mechanism and manage and maintain the case base. Since the case base is the basis for realizing case reasoning, the effectiveness of a supplier selection decision depends on the quality of cases, the representation method of case knowledge, and the retrieval mechanism to a large extent. Therefore, the setting of the module is necessary to ensure the continuous updating of the case base and better serve case retrieval.

4 Method for demand-driven NEV supplier selection

Based on the architecture of demand-driven NEV supplier selection system, we propose a resolution process, as shown in Figure 2. The resolution process is composed of three parts, that is, 1) construction of domain ontology of selection criteria for the NEV component supplier; 2) determination of demand attributes and expected attributes as well as their weights; and 3) selection of the NEV component supplier based on CBR. In the first part,

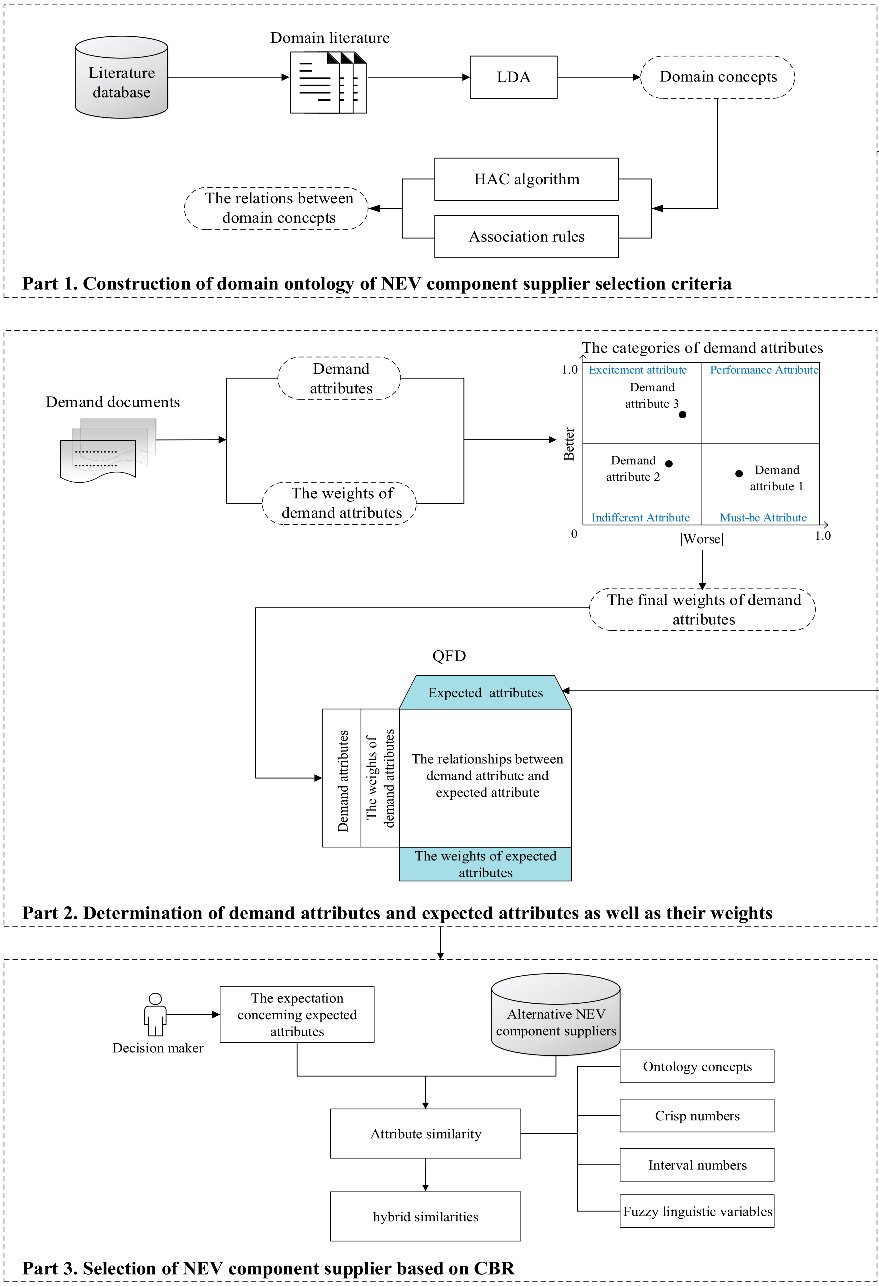

FIGURE 2
Resolution process for the problem of demand-driven NEV supplier selection.

the literature on the selection of the NEV component supplier is preprocessed, and the domain concepts are extracted using LDA. Then, the relations between domain concepts are extracted based on the HAC algorithm and association rules. Furthermore, the domain ontology of the selection criteria for NEV component suppliers can be constructed. In the second part, demand attributes are extracted from the demand documents provided by users with different demands using LDA, and the initial weights of demand attributes are determined based on a sentiment strength analysis algorithm and the information entropy method. Then, the categories of demand attributes and the final weights of demand attributes are determined based on the Kano model and the Better-Worse coefficient. Furthermore, the expected attributes of NEV component supplier selection are determined based on QFD. On the basis of this, the relationship matrix between the demand attributes and the expected attributes is constructed, and the weights of the expected attributes are determined. In the final part, based on CBR and ontology theory, the attribute

10.3389/fenrg.2022.958885

similarities and hybrid similarities between the alternative cases (i.e., NEV component suppliers) and the target case (i.e., the expectation of decision-makers) are calculated, and the NEV component supplier with the largest hybrid similarity is selected.

According to the resolution process shown in Figure 2, a description of the proposed method for demand-driven NEV supplier selection will be given in this section. The detailed descriptions of each part are, respectively, illustrated in Section 4.1, Section 4.2, and Section 4.3.

4.1 Construction of domain ontology of selection criteria for the NEV component supplier

In this section, the construction process of the domain ontology of selection criteria for the NEV component supplier includes four aspects: 1) representation of case knowledge based on ontology, 2) acquisition and preprocessing of text information, 3) extraction of domain concepts based on LDA, and 4) extraction of relationships between domain concepts.

4.1.1 Representation of case knowledge based on ontology

The ontology representation of case knowledge is the basis for realizing case knowledge retrieval, and it is also the basis for the selection of the NEV component supplier in this paper. In ontology, a framework system is used to describe the relationship between objective concepts, and the semantic association rules between concepts are defined to realize the communication and sharing between knowledge systems (Yoo and No, 2014; Lee et al., 2015). The ontology-based case knowledge representation method is easy to deal with structured knowledge, which can ensure the uniqueness of knowledge understanding. It is suitable for knowledge retrieval with complex semantic relations. The integration of ontology and CBR can improve the efficiency and accuracy of case retrieval (Küçük, 2015).

In this paper, the domain ontology of selection criteria for the NEV component supplier can be represented by the following tuple:

$$O = \{C, A, R, M\}, \tag{1}$$

where C is a concept set of cases; A is an attribute set of each concept; R is a relationship set between concepts; and M is a set of instances.

There are three main steps to construct the domain ontology: 1) obtain domain concepts; 2) obtain the relationship between concepts; and 3) construct the ontology of tree structure (Guo and Zhou, 2017). Therefore, it is necessary to acquire domain concepts before constructing a domain ontology.

4.1.2 Acquisition and preprocessing of text information

To ensure the standardization and professionalism of domain concepts, domain concepts are extracted from academic literature in this paper. The academic literature is a type of knowledge resource with high professional value, and it is a relatively standard text form (Liu et al., 2011; Ren, 2012). The domain academic literature contains authoritative data such as domain ontology terms, concepts, and concept relationships. Due to its accessibility and high coverage in the research field, it is possible to construct a relatively complete domain ontology based on academic literature, and the domain ontology constructed is normative and professional (Tang et al., 2020). Therefore, academic literature is used as the source of domain concepts. The detailed description of the acquisition and preprocessing of academic literature is as follows.

First, "NEV," "selection and evaluation of supplier," and "automotive industry" are used as keywords, and "2015–2021" was used as search years to search databases, such as China National Knowledge Infrastructure (CNKI), WanFang database, ScienceDirect, and Web of Science. Manually search and download the literature on the selection criteria of the NEV component supplier, and remove duplicate literature in the retrieved literature.

Then, according to the set standards, the source literature of a domain concept is selected from the collected literature, that is, 1) for papers with multiple publications or citations, only one paper is selected; 2) for the review literature, the original literature in the reference list within the retrieval period is taken as the statistical literature; and 3) literature from various core journals or senior authors. For papers that meet the standards, the titles, keywords, abstracts, and text contents are saved, and they are regarded as domain literature. For the convenience of description, let $D = \{D_1, D_2, \ldots, D_m\}$ be a set of domain literature, where D_j denotes the j-th domain literature and m denotes the number of domain literature, $j \in \{1, 2, \ldots, m\}$.

Furthermore, the domain literature is preprocessed, which includes two steps, that is, 1) word segmentation and part-of-speech (POS) tagging, and 2) text data filtering. The details are given below.

4.1.2.1 Word segmentation and POS tagging

The Chinese Lexical Analysis System (ICTCLAS) and the Natural Language Toolkit (NLTK) are adopted to carry out word segmentation and POS tagging in domain literature D_j, $j \in \{1, 2, \ldots, m\}$. To improve the accuracy, a secondary tagging method is used for POS tagging, that is, not only the verbs and nouns but also the verbs or adjectives with noun functions and proper nouns are tagged.

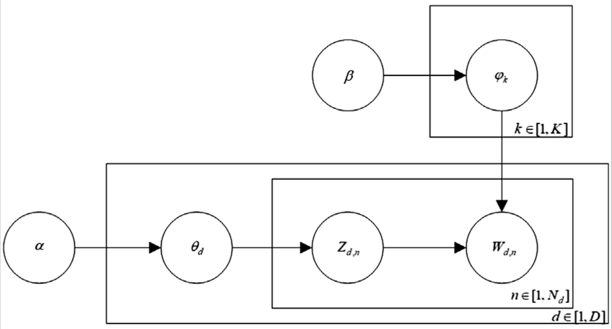

FIGURE 3
Probability map of the topic extraction model.

4.1.2.2 Text data filtering

To improve the training efficiency of the LDA topic model and the quality of the acquired domain concepts, it is necessary to sequentially perform paragraph filtering, stop word deletion, and POS filtering on domain literature D_j, $j \in \{1, 2, \ldots, m\}$. The details are described as follows:

1) Paragraph filtering. The text content of each domain literature D_j contains a large amount of text content that has nothing to do with the selection of an NEV component supplier, $j \in \{1, 2, \ldots, m\}$. Therefore, through keyword tagging, the paragraphs irrelevant to the selection criteria of an NEV component supplier are filtered out.

2) Stop word deletion. Stop words are words that appear frequently, but they have no practical meaning and cannot reflect domain-specific knowledge. To improve efficiency, the Chinese stop word list (CSWL) and English stop word list (ESWL) are used to delete the stop words in domain literature D_j, $j \in \{1, 2, \ldots, m\}$. Based on the results of word segmentation and POS tagging, D_j is compared with the stop words in the CSWL and ESWL. If a word in D_j is the same as the word in CSWL or and ESWL, then it should be deleted; if not, it should be kept, $j \in \{1, 2, \ldots, m\}$.

After preprocessing the domain literature D_j, the preprocessed domain literature set can be obtained, noted as $D^{pro} = \{D_1^{\text{pro}}, D_2^{\text{pro}}, \ldots, D_m^{\text{pro}}\}$, where D_j^{pro} denotes the preprocessed domain literature corresponding to D_j, $j \in \{1, 2, \ldots, m\}$. Also, the initial concept set can be noted as $WD_j = \{d_1^j, d_2^j, \ldots, d_{q_j}^j\}$, where d_Q^j denotes the Q-th initial concept in domain literature D_j^{pro} and q_j denotes the number of initial concepts in domain literature D_j^{pro}, $Q \in \{1, 2, \ldots, q_j\}$, $j \in \{1, 2, \ldots, m\}$.

4.1.3 Extraction of domain concepts based on LDA

In this section, a brief introduction of the LDA topic model is first given. Then, the process of extracting domain concepts based on LDA is illustrated.

4.1.3.1 LDA topic model

LDA is a three-layer Bayesian probability model that contains a three-layer structure of words, topics, and documents (Blei et al., 2003). It can be used to identify the topic information hidden in large-scale document sets or corpora and to effectively extract the key information from the text (Tirunillai and Tellis, 2014; Guo et al., 2017). The LDA model is a typical directed probability graph model, which is determined by the hyper parameters α and β, where α denotes the Dirichlet prior parameter of the multinomial distribution of a topic in any document and β denotes the Dirichlet prior parameter of the multinomial distribution of the words under this topic. The probability map of the topic extraction model is shown in Figure 3.

In Figure 3, θ_d denotes the topic probability distribution in the d-th document, and φ_k denotes the probability distribution of words under the k-th topic. $W_{d,n}$ and $Z_{d,n}$ denote the nth word in the d-th document and the topic of the nth word in the d-th document, respectively. D denotes the total number of documents, K denotes the total number of topics, and N_d denotes the number of words in the d-th document.

The topic is regarded as the middle layer. Through the probability of a certain topic in the document and the probability of a certain word in this topic, the probability of the word in the document can be obtained (Qian et al., 2016; Bi, Liu, Fan, and Cambria 2019). It can be expressed as a probability formula, that is,

$$p(word|doc) = \sum_{topic} p(word|topic) \times p(topic|doc), \quad (2)$$

where doc denotes a document, topic denotes a topic, and word denotes a word. $p(topic|doc)$ and $p(word|topic)$ can be calculated using θ and φ, respectively.

4.1.3.2 The process of extracting domain concepts based on LDA

Domain concept extraction is an important part of domain ontology construction. Usually, domain concepts are those terms that are widely distributed and used more frequently in domain-related data and are less used in non-domain-related data (Zheng et al., 2019). Existing studies show that extracting domain concepts from academic literature conforms to the characteristics of LDA training text. Meanwhile, the LDA model has many advantages, such as flexibility and easy expansion, and has good domain portability (Feng and Zhang, 2017). Therefore, the LDA model is used to extract domain concepts in this paper. The details are given as follows.

Step 1. Let $WD = \{w_1, w_1, \ldots, w_R\}$ denote the set of words concerning the initial concept, where w_r denotes the r-th word in WD and R denotes the number of words, $r \in \{1, 2, \ldots, R\}$. Then, the term frequency of each word in the word set, tf_r^j, is calculated. The calculation formulas of WD and tf_r^j can be represented, respectively, as follows:

$$WD = WD_1 \cup WD_2 \cup \cdots \cup WD_j \cup \cdots \cup WD_m,\ j \in \{1, 2, \ldots, m\}, \tag{3}$$

$$tf_r^j = \frac{n_r^j}{\sum_{r=1}^{R} n_r^j}, r \in \{1, 2, \ldots, R\}, j \in \{1, 2, \ldots, m\}, \tag{4}$$

where n_r^j denotes the number of occurrences of the word w_r in WD_j, $r \in \{1, 2, \ldots, R\}$, $j \in \{1, 2, \ldots, m\}$.

Step 2. Use Python and the LDA package in Sklearn to train the LDA model. In general, the hyper parameters α and β of the LDA model are set to 50/K and 0.01, respectively (Blei et al., 2003). K denotes the number of topics, and it is determined according to perplexity. The model perplexity is an index used to evaluate the language model. The lower the perplexity is, the stronger the predictive ability of the model will be, and the better the LDA model established will be. The calculation formula of perplexity is given by

$$perplexity(D^{text}) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(W_d)}{\sum_{d=1}^{M} N_d} \right), d \in \{1, 2, \ldots, l\}, \tag{5}$$

where $D^{\text{text}} = \{D_1^{\text{text}}, D_2^{\text{text}}, \ldots, D_1^{\text{text}}\}$ denotes a test set of domain literature, W_d denotes the word in the domain literature D_d^{text}, N_d denotes the number of words in the domain literature D_d^{text}, and $p(W_d)$ denotes the probability of word W_d appearing in D_d^{text} and $D^{\text{text}} \subseteq D^{\text{pro}}$, $d \in \{1, 2, \ldots, l\}$.

Step 3. Solve the trained LDA model to obtain the domain topic distribution. The candidate concept set is composed of all domain topics, noted as $C^{\text{all}} = \{C_1^{\text{all}}, C_2^{\text{all}}, \ldots, C_g^{\text{all}}\}$, where C_e^{all} denotes the e-th candidate concept and g denotes the number of candidate concepts, $e \in \{1, 2, \ldots, g\}$.

Step 4. Calculate the term frequency–inverse document frequency (TF-IDF) of the word w_r (Chen et al., 2008; Ray and Chandra, 2012). According to the results of TF-IDF, sort the set of candidate concepts obtained in step 3, screen high-frequency domain topics, and obtain the final domain concept set $C = \{C_1, C_2, \ldots, C_n\}$, where C_i denotes the i th domain concept and n denotes the number of domain concepts, $i \in \{1, 2, \ldots, n\}$. The calculation formula of TF-IDF of w_r is given by

$$TF-IDF_{rj} = tf_r^j \times \log\left(\frac{|m|}{\left|\{j: w_r \in WD_j\}\right| + 1}\right), r \in \{1, 2, \ldots, R\}, j \in \{1, 2, \ldots, m\}, \tag{6}$$

where $|\{j: w_r \in WD_j\}|$ denotes the number of initial concept sets containing the word w_r (i.e., the number of sets with $n_r^j \neq 0$).

4.1.4 Extraction of relationships between domain concepts

The extraction of relationships between domain concepts is the most important step in domain ontology construction, which determines the final domain ontology structure (Zheng et al., 2019). There are two kinds of relations between domain concepts, namely, taxonomic relation and non-taxonomic relation. The taxonomic relation refers to the typical classification structure between domain concepts, such as "kind of." The non-taxonomic relation refers to domain concepts that do not have a typical classification structure between concepts but have certain connections between concepts, such as "synonymous of." In this paper, the HAC method and association analysis method are used to obtain the relationships between domain concepts (Han et al., 2019). The details are described as follows.

4.1.4.1 Mining taxonomic relation between domain concepts

Due to the variety of words, it is difficult to extract the relationship between many domain concepts based on linguistics, while clustering method can make up for this deficiency (Li et al., 2018). Therefore, this study uses the HAC method to mine the taxonomic relation between domain concepts.

First, the vector space model is used to construct the concept–document matrix (Castells et al., 2007). Since TF-IDF can measure the importance of a word to a document set, the vector space model of a domain concept can be constructed based on TF-IDF so as to describe the vector of a domain concept more comprehensively. The vector of domain concept, $\vec{C}_i$, is represented as follows:

$$\vec{C}_i = \left(TF-IDF_{i1}, TF-IDF_{i2}, \ldots, TF-IDF_{ij}, \ldots, TF-IDF_{im}\right), i \in \{1, 2, \ldots, n\}, j \in \{1, 2, \ldots, m\}, \tag{7}$$

where $TF-IDF_{ij}$ is obtained through Eq. 6, $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$.

Then, the cosine distance method is used to calculate the similarity $\text{sim}(\vec{C}_i, \vec{C}_h)$ between $\vec{C}_i$ and $\vec{C}_h$ (Liao and Xu, 2015), where the cosine value of two concept words is larger, the angle between the two vectors will be smaller, and the similarity between the two domain concepts will be greater, $i, h \in \{1, 2, \ldots, n\}$, $i \neq h$. The calculation formula of $\text{sim}(\vec{C}_i, \vec{C}_h)$ is given by

$$sim(\vec{C}_i, \vec{C}_h) = \frac{\sum_{j=1}^{m} TF-IDF_{ij} \times TF-IDF_{hj}}{\sqrt{\sum_{j=1}^{m} TF-IDF_{ij}^2 \sum_{j=1}^{m} TF-IDF_{hj}^2}}, \quad i, h \in \{1, 2, \ldots, n\}, i \neq h, j \in \{1, 2, \ldots, m\}. \tag{8}$$

Furthermore, the semantic similarity matrix $\vec{S}_{ih} = [sim(\vec{C}_i, \vec{C}_h)]_{n\times n}$ between the domain concepts can be obtained, that is,

$$\vec{S}_{ih} = \begin{array}{c} \\ \vec{C}_1 \\ \vec{C}_2 \\ \vec{C}_3 \\ \vdots \\ \vec{C}_n \end{array} \begin{array}{c} \begin{array}{ccccc} \vec{C}_1 & \vec{C}_2 & \vec{C}_3 & \cdots & \vec{C}_n \end{array} \\ \begin{bmatrix} 1 & sim(\vec{C}_1,\vec{C}_2) & sim(\vec{C}_1,\vec{C}_3) & \cdots & sim(\vec{C}_1,\vec{C}_n) \\ & 1 & sim(\vec{C}_2,\vec{C}_3) & \cdots & sim(\vec{C}_2,\vec{C}_n) \\ & & 1 & \cdots & sim(\vec{C}_3,\vec{C}_n) \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} \end{array},$$

$$i,h \in \{1,2,\ldots,n\},\ i \neq h;\ j \in \{1,2,\ldots,m\}. \quad (9)$$

The specific clustering process is given as follows.

Step 1Construct the vector $\vec{C}_i$ base on the TF-IDF and use each vector as a cluster to form a cluster set $\bar{A} = \{A_1, A_2, \ldots, A_n\}$, $i \in \{1,2,\ldots,n\}$.

Step 2Using average-linkage (AL) hierarchical clustering as the distance between A_i and A_h (Xu et al., 2021), and the calculation formula is given by

$$Dis(A_i, A_h) = \frac{\sum_{\vec{C}_a \in A_i, \vec{C}_b \in A_h} sim(\vec{C}_a, \vec{C}_b)}{|A_i||A_h|},\ i,h \in \{1,2,\ldots,n\},\ i \neq h, \quad (10)$$

where $\vec{C}_a$ and $\vec{C}_b$ denote the vectors in cluster A_i and cluster A_h, respectively; $|A_i|$ and $|A_h|$ denote the number of vectors in the clusters A_i and A_h, respectively.

Step 3Select the two clusters A_i and A_h with the highest average similarity. If $Dis(A_i, A_h)$ is greater than the preset threshold δ (usually, $\delta = 0.5$), then merge the two clusters into a new cluster A_{ih}, add A_{ih} to the cluster set $\bar{A}$, and remove A_i and A_h from cluster set $\bar{A}$, $i,h \in \{1,2,\ldots,n\}$, $i \neq h$. Otherwise, go to step 5.

Step 4If the number of clusters in cluster set $\bar{A}$ is greater than 1, then go to step 2. Otherwise, go to step 5.

Step 5Clustering is completed, output cluster set $\bar{A}$.

After clustering, the domain concepts can be divided into multiple clusters, but the division of parent concepts and sub-concepts within a cluster needs to be further defined. In this study, the average similarity within the cluster is used to extract the parent concept in the cluster. The calculation formula of average similarity within the cluster of domain concepts C_i can be given by

$$sim_{avg}(\vec{C}_i) = \frac{\sum_{h=1}^{n'} sim(\vec{C}_i, \vec{C}_h)}{n'},\ i,h \in \{1,2,\ldots,n\},\ i \neq h, \quad (11)$$

where n' denotes the number of domain concepts within the cluster A_i. If the average similarity within the cluster of C_i is greater, it means that C_i is most closely related to other domain concepts in the cluster and has a broader semantics, so it can be used as the parent concept of other concepts in the cluster, $i \in \{1,2,\ldots,n\}$.

4.1.4.2 Mining non-taxonomic relation between domain concepts

The traditional association rules can only obtain the non-taxonomic relation between domain concepts but cannot obtain the specific relationship name (Agrawal et al., 1996). Usually, the non-taxonomic relation between domain concepts is composed of a verb and a pair of domain concepts, so this structure is used as the main extraction object, and verbs are used as the non-taxonomic relation between domain concepts to improve the traditional association rules in this study (Dong et al., 2013). The related domain concept pair set can be represented as $R = (C_i, C_h, V_t)$, where V_t denotes the non-taxonomic relation between domain concepts C_i and C_h, $i, h \in \{1,2,\ldots,n\}, i \neq h, t \in \{1,2,\ldots,T\}$. The analysis of the association rules process is described as follows.

Step 1. Randomly select two concepts C_i and C_h without association rule analysis from the domain concept set $C = \{C_1, C_2, \ldots, C_n\}$, $i, h \in \{1,2,\ldots,n\}$, $i \neq h$. If there is no such domain concept pair, go to step 6.

Step 2. Calculate the support Sup and the confidence Con (Le and Lo 2015). Sup and Con can be, respectively, calculated by

$$Sup(C_i \to C_h) = \frac{\sum_{j=1}^{m} f\left(C_{i,j}, C_{h,j}\right)}{m},$$
$$i,h \in \{1,2,\ldots,n\},\ i \neq h,\ j \in \{1,2,\ldots,m\}, \quad (12)$$

$$Con(C_i \to C_h) = \frac{\sum_{j=1}^{m} f\left(C_{i,j}, C_{h,j}\right)}{\sum_{j=1}^{m} f\left(C_{i,j}\right)},$$
$$i,h \in \{1,2,\ldots,n\},\ i \neq h,\ j \in \{1,2,\ldots,m\}, \quad (13)$$

where $f(C_{i,j}, C_{h,j})$ denotes the frequency of co-occurrence of concepts C_i and C_h in the domain literature D_j^{pro}. $f(C_{i,j})$ denotes the frequency of concepts C_i in the domain literature D_j^{pro}, $i,h \in \{1,2,\ldots,n\}$, $i \neq h$, $j \in \{1,2,\ldots,m\}$.

Step 3. Set the threshold of support and confidence, that is, $Sup_{\min}$ and $Con_{\min}$. If $Sup(C_i \to C_h) \geq Sup_{\min}$ and $Con(C_i \to C_h) \geq Con_{\min}$, then C_i and C_h have a non-taxonomic relation, $i,h \in \{1,2,\ldots,n\}$, $i \neq h$, and go to step *4*. Otherwise, go to step 1.

Step 4. Count all the verbs $V_1, V_2, \ldots, V_T$ that appear between C_i and C_h in the domain literature D_j^{pro} and their frequency $FV_1, FV_2, \ldots, FV_T$, $i,h \in \{1,2,\ldots,n\}$, $i \neq h$, $j \in \{1,2,\ldots,m\}$.

Step 5. If the frequency of V_t is the largest and greater than the threshold F_{min}, then the verb V_t is defined as the non-taxonomic relation between C_i and C_h and saved in the domain concept pair set R; go to step 1, $i,h \in \{1,2,\ldots,n\}$, $i \neq h$, $t \in \{1,2,\ldots,T\}$.

Step 6. After manually checking all the verbs, output the domain concept pair set R.

According to the obtained domain concepts and the relationship between domain concepts, a domain ontology of tree structure for the selection criteria of NEV component suppliers can be constructed.

4.2 Determination of demand attributes and expected attributes as well as their weights

Based on the constructed domain ontology, demand attributes and expected attributes as well as their weights are determined, including the following three aspects: 1) determination of demand attributes and their weights, 2) determination of demand attribute categories and expected attributes, and 3) construction of a relation matrix and determination of expected attribute weights.

4.2.1 Determination of demand attributes and their weights

In this section, the process of extracting demand attributes from demand documents based on LDA is first illustrated. Then, the process of determining demand attribute weights is given. The detailed descriptions are given as follows.

4.2.1.1 Extracting demand attributes from demand documents based on LDA

The textual contents in demand documents for NEV component suppliers provided by users not only contain the words concerning demand attributes but also contain a lot of noisy and irrelevant words. To improve the effectiveness and efficiency of extracting demand attributes, demand documents are first preprocessed. Let $P = \{P_1, P_2, \ldots, P_v\}$ denote the set of demand documents, where P_k denotes the k-th demand document and v denotes the number of demand documents, $k \in \{1, 2, \ldots, v\}$. Similar to the preprocessing of domain literature D_j, ICTCLAS is first used to process word segmentation and POS tagging on the demand document P_k, and then the word set concerning P_k can be obtained by removing the stop words, noted as $WD_k^P = \{WD_{k1}^P, WD_{k2}^P, \ldots, WD_{kq}^P\}$, where WD_{kc}^P denotes the cth word in WD_k^P and q denotes the number of words in WD_k^P, $j \in \{1, 2, \ldots, m\}$, $k \in \{1, 2, \ldots, v\}$, and $c \in \{1, 2, \ldots, q\}$.

Similar to domain concept extraction, topics can be extracted from preprocessed demand documents based on steps 2–4 in the process of extracting domain concepts. Since there may be some noisy words and topics with similar meanings in the extracted topics, in order to get more reasonable results, the noise words can be filtered manually, the topics with similar meanings can be merged, the important topics can be selected, and each topic can be labeled. Following Poria et al. (2016) and Bi et al. (2019), each extracted topic can be regarded as a demand attribute. Then, the set of labeled topics (i.e., demand attributes) and the set of words concerning each topic can be determined. The determined demand attribute E_b can be denoted as $E_b = \{\mathrm{word}_{b1}, \mathrm{word}_{b2}, \ldots, \mathrm{word}_{bJ_b}\}$, where word_{bJ} denotes the J-th frequent word in the b-th demand attribute and J_b denotes the number of frequent words in E_b, $b \in \{1, 2, \ldots, I\}$ and $J \in \{1, 2, \ldots, J_b\}$.

4.2.1.2 Determining the weights of demand attributes

Usually, the demand information reflected in the demand document corresponds to multiple demand attributes. Hence, it is necessary to identify the demand information with respect to each demand attribute so as to determine the weight of each demand attribute. In accordance with the obtained word set WD_k^P and the extracted demand attribute E_b, demand information concerning E_b can be extracted, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. Let WD_{kb}^P be the demand information concerning E_b in WD_k^P, it can be identified by comparing the words in WD_k^P with the demand attribute word E_b, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. Specifically, if $E_b \in WD_k^P$, then the verbs, adverbs, and adjectives between two adjacent punctuations in the demand information, including demand attribute word E_b, are extracted from WD_k^P (Huang and Cheng, 2015); if $E_b \notin WD_k^P$, then $WD_{kb}^P = ''-''$, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. Thus, WD_{kb}^P can be obtained, that is, $WD_{kb}^P = \{WD_{kb}^{P1}, WD_{kb}^{P2}, \ldots, WD_{kb}^{Pu_b}\}$, where WD_{kb}^{Pu} and u_b denote the u-th word and the number of words in WD_{kb}^P, respectively, $k \in \{1, 2, \ldots, v\}$, $b \in \{1, 2, \ldots, I\}$, and $u \in \{1, 2, \ldots, u_b\}$.

Based on the demand information concerning demand attributes, the multi-granularity sentiments with respect to demand attributes reflected by users in their demand documents can be identified. Generally, the word sets of sentiment strength with respect to different demand attributes may be different. Therefore, in order to improve the accuracy of sentiment analysis, a sentiment dictionary with respect to demand attributes needs to be established. Let $WR = \{WR_1, WR_2, \ldots, WR_{h_{view}}\}$ be the demand preference word set of users with respect to demand attributes, where WR_f and h_{view} denote the f-th word and the number of words in WR, respectively, $f \in \{1, 2, \ldots, h_{view}\}$. Also, the word set WR can be represented by

$$WR = WD_{11}^P \cup WD_{12}^P \cup \cdots \cup WD_{kb}^P \cup \cdots \cup WD_{vI}^P. \quad (14)$$

To identify the sentiment polarity of users concerning demand attributes, HowNet (http://www.keenage.com/) was introduced to establish positive and negative sentiment dictionaries with respect to demand attributes. Let WR_{HowNet}^{+} and WR_{HowNet}^{-} be the positive and negative evaluation word sets that are commonly used in HowNet, respectively. Then, in accordance with WR_{HowNet}^{+}, WR_{HowNet}^{-}, and WR, the positive and negative sentiment dictionaries with respect to demand attributes, WR_H^{+} and WR_H^{-}, can be constructed, that is,

$$WR_H^{+} = WR_{HowNet}^{+} \cap WR, \quad (15)$$

$$WR_H^{-} = WR_{HowNet}^{-} \cap WR. \quad (16)$$

For some words in WR that may belong to neither WR_{HowNet}^{+} nor WR_{HowNet}^{-}, the manual identification method is adopted to determine the sentiment dictionary to which the corresponding words belong (Liu et al., 2017). Thus, the positive sentiment dictionary and negative sentiment dictionary with respect to demand attributes, WR^{+} and WR^{-}, are finally formed.

Let s_{kb}^P be the sentiment strength of WD_k^P concerning E_b, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. In this study, the five-granularity sentiment strengths of users are considered based on actual demands. In particular, if the set WD_k^P does not contain demand information concerning E_b, that is, $WD_{kb}^P = ''-''$, then $s_{ik}^P = ''-''$, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. Thus, s_{kb}^P has six possible values, that is, $s_{kb}^P \in \{-2, -1, 0, 1, 2, ''-''\}$, where -2 (2) and -1 (1), respectively, represent the "very" and "slightly" degree of negative (positive) sentiment strengths and 0 represents the neutral sentiment strength, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. Let WR_{neg} be the set of negative words, WR_d be the degree word set in HowNet, WR_d^+ be the set of enhanced sentiment words in WR_d, and WR_d^- be the set of weakened sentiment words in WR_d. Therefore, the value of s_{kb}^P can be obtained by comparing the words in set WD_{kb}^P with the words in sets WR^+, WR^-, WR_{neg}, and WR_d (Li, Zheng, Yue, et al., 2021), $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. Let s_{kb}^{P+}, s_{kb}^{P-}, s_{kb}^{Pneg}, and s_{kb}^{Pd} be the indicator variables of intersections of WD_{kb}^P and WR^+, WD_{kb}^P and WR^-, WD_{kb}^P and WR_{neg}, and WD_{kb}^P and WR_d, respectively. Then, the specific process of the multi-granularity sentiment analysis algorithm can be described as follows:

Step 1. If $WD_{kb}^P = ''-''$ or $WD_{kb}^P = \varnothing$, then $s_{kb}^P \leftarrow ''-''$, else go to step 2.

Step 2. If $WD_{kb}^P \cap WR^+ \neq \varnothing$, then $s_{kb}^{P+} \leftarrow 1$, else $s_{kb}^{P+} \leftarrow 0$.

Step 3. If $WD_{kb}^P \cap WR^- \neq \varnothing$, then $s_{kb}^{P-} \leftarrow 1$, else $s_{kb}^{P-} \leftarrow 0$.

Step 4. If $WD_{kb}^P \cap WR_{neg} \neq \varnothing$, then $s_{kb}^{Pneg} \leftarrow 1$, else $s_{kb}^{Pneg} \leftarrow 0$.

Step 5. If $WD_{kb}^P \cap WR_d = \varnothing$, then $s_{kb}^{Pd} \leftarrow 0$; else if $WD_{kb}^P \cap WR_d^+ \neq \varnothing$, then $s_{kb}^{Pd} \leftarrow 1$; else $s_{kb}^{Pd} \leftarrow -1$.

Step 6. If $s_{kb}^{P+} = s_{kb}^{P-} = 0$ or $s_{kb}^{P+} = s_{kb}^{P-} = 1$, then $s_{kb}^P \leftarrow 0$; when $s_{kb}^{Pd} = 1$, if $s_{kb}^{P+} = 1$, $s_{kb}^{P-} = 0$, $s_{kb}^{Pneg} \leftarrow 1$ or $s_{kb}^{P+} = 0$, $s_{kb}^{P-} = 1$, $s_{kb}^{Pneg} \leftarrow 0$, then $s_{kb}^P \leftarrow -1 - s_{kb}^{Pd}$, else $s_{kb}^P \leftarrow 1 + s_{kb}^{Pd}$; when $s_{kb}^{Pd} = 0$ or -1, if $s_{kb}^{P+} = 1$, $s_{kb}^{P-} = 0$, $s_{kb}^{Pneg} \leftarrow 1$ or $s_{kb}^{P+} = 0$, $s_{kb}^{P-} = 1$, $s_{kb}^{Pneg} \leftarrow 0$, then $s_{kb}^P \leftarrow -1$, else $s_{kb}^P \leftarrow 1$.

Based on the determined results of s_{kb}^P, the weights of demand attributes can be obtained by further integrating the users' demand preferences. In order to avoid information loss, this study uses an indicator vector to express sentiment strength information. Let $T_{kb}^P = (T_{kb}^I, T_{kb}^{II}, T_{kb}^{III}, T_{kb}^{IV}, T_{kb}^V)$ be the preference indicator vector of sentiment strength s_{kb}^P, where $T_{kb}^I, T_{kb}^{II}, T_{kb}^{III}, T_{kb}^{IV}, T_{kb}^V = 0$ or 1 and $T_{kb}^I + T_{kb}^{II} + T_{kb}^{III} + T_{kb}^{IV} + T_{kb}^V \le 1$, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$. In accordance with the value of s_{kb}^P, T_{kb}^P can be determined. Specifically, if $s_{kb}^P = -2$, then $T_{kb}^P = (1, 0, 0, 0, 0)$; if $s_{kb}^P = -1$, then $T_{kb}^P = (0, 1, 0, 0, 0)$; if $s_{kb}^P = 0$, then $T_{kb}^P = (0, 0, 1, 0, 0)$; if $s_{kb}^P = 1$, then $T_{kb}^P = (0, 0, 0, 1, 0)$; if $s_{kb}^P = 2$, then $T_{kb}^P = (0, 0, 0, 0, 1)$; especially, if $s_{kb}^P = ''-''$, then $T_{kb}^P = (0, 0, 1, 0, 0)$, this is because most users will not express his/her demand preferences for a demand attribute when he/she holds a neutral sentiment toward the demand attribute in reality (Li, Zheng, Fan, et al., 2021), $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$.

Next, the attribute value of WD_k^P concerning demand attribute E_b in the form of sentiment strength distribution, $P_{kb}(x)$, can be determined, $k \in \{1, 2, \ldots, v\}$ and $b \in \{1, 2, \ldots, I\}$, $x \in \Omega$, $\Omega \in \{-2, -1, 0, 1, 2\}$, that is,

$$P_{kb}(x) = \begin{cases} T_{kb}^I, & x = -2 \\ T_{kb}^{II}, & x = -1 \\ T_{kb}^{III}, & x = 0 \\ T_{kb}^{IV}, & x = 1 \\ T_{kb}^V, & x = 2 \end{cases}, k \in \{1, 2, \ldots, v\}, b \in \{1, 2, \ldots, I\}. \tag{17}$$

Furthermore, the expected values of WD_k^P concerning demand attribute E_b can be calculated by

$$e_{kb} = \sum_{x=-2}^{2} P_{kb}(x) \cdot x, k \in \{1, 2, \ldots, v\}, b \in \{1, 2, \ldots, I\}, x \in \Omega. \tag{18}$$

On the basis of this, the information entropy method is adopted to calculate demand attribute weights (Chen 2020). Generally, the smaller the information entropy of E_b is, the more information it will provide, that is, the greater the role it plays in the selection of NEV component suppliers, the greater its weight ω_b^P will be, and *vice versa* (Du and Gao, 2020; Li, Zheng, Yue, et al., 2021). The weight of the demand attribute E_b, ω_b^P, can be calculated, that is,

$$\omega_b^P = [1 - IE(\bar{e}_b)] \Big/ \sum_{b=1}^{I} [1 - IE(\bar{e}_b)], b \in \{1, 2, \ldots, I\}, \tag{19}$$

where $IE(\bar{e}_b)$ represents the information entropy concerning E_b, and it can be calculated by

$$IE(\bar{e}_b) = -\sum_{k=1}^{v} (\hat{e}_{kb} \cdot \ln \hat{e}_{kb}) \Big/ \ln v, k \in \{1, 2, \ldots, v\}, \\ b \in \{1, 2, \ldots, I\}. \tag{20}$$

In Eq. 20, $\hat{e}_{kb}$ can be calculated by Eq. 21. In particular, $0 \cdot \ln(0) = 0$.

$$\hat{e}_{kb} = |e_{kb}| \Big/ \sqrt{\sum_{k=1}^{v} e_{kb}^2}, k \in \{1, 2, \ldots, v\}, b \in \{1, 2, \ldots, I\}. \tag{21}$$

4.2.2 Determination of demand attribute categories and expected attributes

The analysis of demand attributes can promote the determination of expected attributes. In this section, the Kano model is used to further analyze the demand attributes, which helps to determine the categories and priority of demand attributes so as to determine the expected attributes of NEV component suppliers efficiently (Kano et al., 1984).

The Kano category of each demand attribute can be determined by designing a two-factor questionnaire and counting the results of the questionnaire (Matzler and Hinterhuber, 1998). In the designed questionnaire, there are five response options for each demand attribute, and the questions are in the form of forward and reverse, as shown in Table 2, that is, the satisfaction of users when NEV component

TABLE 2 Kano categories concerning demand attributes.

		With the demand attribute E_b				
		Like	**Deserved**	**Indifferent**	**Tolerable**	**Unlike**
Without the demand attribute E_b	Like	Q	A	A	A	O
	Deserved	R	I	I	I	M
	Indifferent	R	I	I	I	M
	Tolerable	R	I	I	I	M
	Unlike	R	R	R	R	Q

suppliers have the demand attribute E_b and the dissatisfaction of users when NEV component suppliers do not have the demand attribute E_b, $b \in \{1, 2, \ldots, I\}$. According to the results of the questionnaire, the attribute category with the largest number corresponding to each demand attribute is regarded as the final category of the demand attribute. Usually, there are five attribute categories (Xu et al., 2009), namely, must-be demand attributes (M), one-dimensional (or performance) demand attributes (O), attractive demand attributes (A), indifferent demand attributes (I), and reverse demand attributes (R).

However, the traditional Kano model takes "the choice of the majority of users" as the final category of each demand attribute, which fails to consider the influence of the distribution of other attribute categories. In addition, there are some unavoidable problems in the process of data statistics. For example, there are many demand attributes in the same Kano category, and the priority cannot be determined. To solve these problems, the Better-Worse coefficient is calculated based on the traditional Kano model (Berger et al., 1993), and various categories of demand attributes are comprehensively considered. Specifically, according to the questionnaire survey results, the percentages of demand attribute categories can be counted (A_b, O_b, M_b, I_b, and R_b), and then the satisfaction coefficient and the dissatisfaction coefficient of users toward demand attribute E_b, $Better_b$ and $Worse_b$, can be calculated, that is,

$$Better_b = \frac{A_b + O_b}{A_b + O_b + M_b + I_b}, \ b \in \{1, 2, \ldots, I\}, \tag{22}$$

$$Worse_b = -\frac{O_b + M_b}{A_b + O_b + M_b + I_b}, \ b \in \{1, 2, \ldots, I\}, \tag{23}$$

where $Better_b \in (0, 1)$, and the greater the $Better_b$ is, the greater the impact of demand attribute E_b on users' satisfaction will be; $Worse_b \in (-1, 0)$, and the smaller the $Worse_b$ is, the greater the impact of demand attribute E_b on users' dissatisfaction will be. Therefore, according to the calculation results of $Better_b$ and $Worse_b$, the final categories of demand attributes can be determined, as shown in Figure 4, $b \in \{1, 2, \ldots, I\}$. The method of combining the Kano model with the Better-Worse coefficient is more convenient to mine the potential demands of users than the traditional Kano model and helps to further determine the expected attributes.

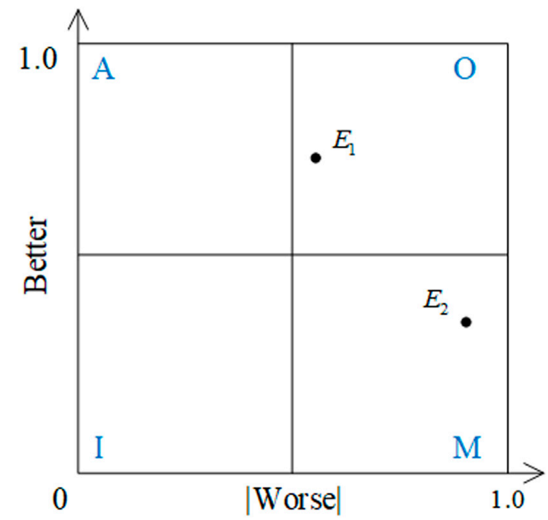

FIGURE 4
Better-worse coefficient quadrant chart.

Furthermore, the final weight of the demand attribute, $\omega_b^{P'}$, can be obtained by adjusting the weight of E_b (Högström et al., 2010), that is,

$$\omega_b^{P'} = \frac{\omega_b^P \cdot k_b}{\sum_{b=1}^{I} (\omega_b^P \cdot k_b)}, \ b \in \{1, 2, \ldots, I\}, \tag{24}$$

where k_b denotes the adjustment coefficient; if E_b is divided into M, O, A, or I, then k_b is 2, 1.5, 1, or 0.5, respectively.

Based on the domain ontology constructed in Section 4.1 and the categories of demand attributes determined earlier, the demand attribute E_b can be transformed into an expected attribute d_a by an expert group, and each expected attribute can meet certain users' demands, which makes the selection of expected attributes targeted, $b \in \{1, 2, \ldots, I\}$ and $a \in \{1, 2, \ldots, \delta\}$.

TABLE 3 Relationship matrix between E_b and d_a.

Demand attribute	Weight of demand attributes	Expected attribute			
		d_1	d_2	$\cdots$	d_δ
E_1	$\omega_1^{P'}$	R_{11}	R_{12}	$\cdots$	$R_{1\delta}$
E_2	$\omega_2^{P'}$	R_{21}	R_{22}	$\cdots$	$R_{2\delta}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
E_I	$\omega_I^{P'}$	R_{I1}	R_{I2}	$\cdots$	$R_{I\delta}$
Importance weights of expected attributes		$I\omega_1^E$	$I\omega_2^E$	$\cdots$	$I\omega_\delta^E$
Weights of expected attributes		ω_1^E	ω_2^E	$\cdots$	ω_δ^E

4.2.3 Construction of the relationship matrix and determination of expected attribute weights

Based on the determined demand attributes and expected attributes, the degrees of correlation between demand attributes and expected attributes can be evaluated by the expert group. Let R_{ba} denote the degree of correlation between the demand attribute E_b and expected attribute d_a, where $R_{ba} \in \{1, 3, 5, 7, 9\}$, and 1, 3, 5, 7, and 9, respectively, denote the "very weak," "weak," "general," "strong," and "very strong" degrees of correlation, $b \in \{1, 2, \ldots, I\}$ and $a \in \{1, 2, \ldots, \delta\}$.

Then, the importance weight and the weight of the expected attributes d_a, $I\omega_a^E$, and ω_a^E, can be obtained. $I\omega_a^E$ and ω_a^E are, respectively, expressed as follows:

$$I\omega_a^E = \sum_{b=1}^{I} R_{ba} \times \omega_b^{P'},\ b \in \{1, 2, \ldots, I\},\ a \in \{1, 2, \ldots, \delta\}, \quad (25)$$

$$\omega_a^E = \frac{I\omega_a^E}{\sum_{a=1}^{\delta} I\omega_a^E},\ a \in \{1, 2, \ldots, \delta\}. \quad (26)$$

Furthermore, the relationship matrix $\tilde{R} = [R_{ba}]_{I\times\delta}$ between E_b and d_a can be constructed as shown in Table 3.

4.3 Selection of the NEV component supplier based on CBR

Case retrieval is usually regarded as the most important step in CBR (Tadrat et al., 2012). The target of this process is to retrieve those cases in the case base that are closest to the new problem. To achieve this aim effectively, the similarities between the target case and the stored cases are measured (Homem et al., 2020). In this section, the expectation given by the decision-maker is taken as the target case, and the alternative NEV component suppliers are taken as the historical cases. By calculating the attribute similarity and hybrid similarity between the target case and the historical cases, the NEV component supplier that meets the expectations of the decision-maker is selected.

4.3.1 Attribute similarity measure for each format of attribute value

In case retrieval, there are usually multiple attributes involved, which may have different attribute value formats. In this study, four common attribute value formats are considered in NEV component supplier selection: ontology concepts, crisp numbers, interval numbers, and fuzzy linguistic variables. The attribute values of historical cases and target cases of the same expected attribute are expressed in the same format. Let Z_μ be the μth historical case, $\mu \in U$, $U = \{1, 2, \ldots, q_\mu\}$. Z^o denotes a target case, which is also the expectation of the decision-maker concerning NEV component suppliers. Let $z_{\mu a}^P$ be the attribute value of the historical case Z_μ concerning the expected attribute d_a, $\mu \in U$ and $a \in \{1, 2, \ldots, \delta\}$. Let z_a^o be the attribute value of the target case Z^o concerning the expected attribute d_a, $a \in \{1, 2, \ldots, \delta\}$. For the four attribute value forms considered, let $D^I = \{d_1, d_2, \ldots, d_{L_1}\}$, $D^{II} = \{d_{L_1+1}, d_{L_1+2}, \ldots, d_{L_2}\}$, $D^{III} = \{d_{L_2+1}, d_{L_2+2}, \ldots, d_{L_3}\}$, and $D^{IV} = \{d_{L_3+1}, d_{L_3+2}, \ldots, d_\delta\}$ denote the subsets of attributes concerning the formats of ontology concepts, crisp numbers, interval numbers, and fuzzy linguistic variables, respectively, where $D^I \cup D^{II} \cup D^{III} \cup D^{IV} = D^A$, $D^I \cap D^{II} \cap D^{III} \cap D^{IV} = \varnothing$, and $D^A = \{d_1, d_2, \ldots, d_\delta\}$. Accordingly, the subscripts of these four subsets can be represented as $\Omega^I = \{1, 2, \ldots, L_1\}$, $\Omega^{II} = \{L_1 + 1, L_1 + 2, \ldots, L_2\}$, $\Omega^{III} = \{L_2 + 1, L_2 + 2, \ldots, L_3\}$, and $\Omega^{IV} = \{L_3 + 1, L_3 + 2, \ldots, \delta\}$, respectively, where $\Omega^I \cup \Omega^{II} \cup \Omega^{III} \cup \Omega^{IV} = \Omega^A$, $\Omega^I \cap \Omega^{II} \cap \Omega^{III} \cap \Omega^{IV} = \varnothing$, and $\Omega^A = \{1, 2, \ldots, \delta\}$.

Considering the four attribute value formats, the method of measuring expected attribute similarity for each attribute value format is given as follows.

4.3.1.1 Attribute similarity measurement for ontology concept format attribute values

For $d_a \in D^I$, attribute values $z_{\mu a}^P$ and z_a^o are the ontology concepts. In the ontology of tree structure, the closer the distance between two nodes is, the greater the similarity between the concepts corresponding to the two nodes will be. Therefore, the semantic similarity can be measured according to the path distance between two concept nodes (Huang and Zhou 2007). The semantic distance similarity between $z_{\mu a}^P$ and z_a^o can be represented as $Sim_c(z_{\mu a}^P, z_a^o)$, then the calculation formula of $Sim_c(z_{\mu a}^P, z_a^o)$ is given by

$$Sim_c\left(z_{\mu a}^P, z_a^o\right) = \frac{1}{1 + Dis_c\left(z_{\mu a}^P, z_a^o\right)}, \mu \in U, a \in \Omega^I, \tag{27}$$

where

$$Dis_c\left(z_{\mu a}^P, z_a^o\right) = \sum\nolimits_{s=1}^{\xi_a} Len(L_s), \mu \in U, a \in \Omega^I. \tag{28}$$

In Eq. 28, $Dis_c(z_{\mu a}^P, z_a^o)$ denotes the semantic distance between $z_{\mu a}^P$ and z_a^o, $Len(L_s)$ denotes the length of the s-th side in the shortest path between $z_{\mu a}^P$ and z_a^o, and ξ_a denotes the total number of sides included in the shortest path between $z_{\mu a}^P$ and z_a^o.

However, the similarity algorithm based on semantic distance does not fully consider the factors that affect the similarity between ontology concepts, which may affect the accuracy of the case retrieval results (Wen et al., 2017). To improve the accuracy of the case retrieval results, the two factors that affect the similarity are considered, that is, the node depth and node density. The concept node depth is the number of sides included in the shortest path between the concept node and the root node. The semantic similarity of ontology concepts can be calculated using the node depth (Zhang et al., 2013), that is,

$$Sim_d\left(z_{\mu a}^P, z_a^o\right) = \frac{Depth\left(Con\left(z_{\mu a}^P, z_a^o\right)\right)}{\max\left(Depth\left(z_{\mu a}^P, z_a^o\right)\right)}, \mu \in U, a \in \Omega^I, \tag{29}$$

where $Depth(Con(z_{\mu a}^P, z_a^o))$ denotes the depth of the nearest common parent node of $z_{\mu a}^P$ and z_a^o and $\max(Depth(z_{\mu a}^P, z_a^o))$ denotes the maximum depth of $z_{\mu a}^P$ and z_a^o in the ontology of tree structure.

Furthermore, the semantic similarity of ontology concepts can be calculated using the node density (Wen et al., 2011), that is,

$$Sim_w\left(z_{\mu a}^P, z_a^o\right) = \frac{Degree\left(Con\left(z_{\mu a}^P, z_a^o\right)\right)}{Max\left(Degree\left(C_i\right)\right)}, \mu \in U, a \in \Omega^I, i \in \{1, 2, \ldots, n\}, \tag{30}$$

where $Degree(Con(z_{\mu a}^P, z_a^o))$ denotes the number of direct child nodes contained in the nearest common ancestor node of $z_{\mu a}^P$ and z_a^o and $Max(Degree(C_i))$ denotes the maximum number of direct child nodes contained in each concept node of the domain ontology. The greater the number of direct child nodes of a concept node is, the greater the distribution density of the concept node will be, and the greater the semantic similarity between direct child nodes will be.

By comprehensively considering the influence of various factors of the concept semantic similarity, the concept semantic similarity between the historical case Z_μ and the target case Z^o concerning the expected attribute d_a, $Sim_a(Z^o, Z_\mu)$, can be calculated as follows:

$$Sim_a\left(Z^o, Z_\mu\right) = \alpha' \cdot Sim_c\left(z_{\mu a}^P, z_a^o\right) + \beta' \cdot Sim_d\left(z_{\mu a}^P, z_a^o\right) + \gamma' \cdot Sim_w\left(z_{\mu a}^P, z_a^o\right), \mu \in U, a \in \Omega^I, \tag{31}$$

where α', β', and γ' are the adjustment coefficients, $\alpha', \beta', \gamma' \in (0, 1]$, and $\alpha' + \beta' + \gamma' = 1$. Herein, according to Wen et al. (2017), they are set as 0.9, 0.05, and 0.05, respectively.

4.3.1.2 Attribute similarity measurement for crisp number format attribute values

For $d_a \in D^{II}$, attribute values $z_{\mu a}^P$ and z_a^o are the crisp numbers. These values can be viewed as two points in a continuous value range of an expected attribute d_a, $a \in \Omega^{II}$. The closer these two points are, the more similar the history case Z_μ and the target case Z^o concerning d_a will be. Therefore, the distance-based method is used to calculate the attribute similarity in the form of a crisp number (Fan et al., 2014). Also, the attribute similarity between Z_μ and Z^o concerning d_a can be calculated as follows:

$$Sim_a\left(Z^o, Z_\mu\right) = \exp\left[-Dis_{cn}\left(z_{\mu a}^P, z_a^o\right)\right], \mu \in U, a \in \Omega^{II}, \tag{32}$$

where

$$Dis_{cn}^P\left(z_{\mu a}^P, z_a^o\right) = \frac{\sqrt{\left(z_{\mu a}^P - z_a^o\right)^2}}{\max\left\{\sqrt{\left(z_{\mu a}^P - z_a^o\right)^2} \middle| \mu \in U\right\}}, \mu \in U, a \in \Omega^{II}. \tag{33}$$

4.3.1.3 Attribute similarity measurement for interval number format attribute values

For $d_a \in D^{III}$, attribute values $z_{\mu a}^P$ and z_a^o are the interval numbers, that is, $z_{\mu a}^P = [z_{\mu a}^{P+}, z_{\mu a}^{P-}]$ and $z_a^o = [z_a^{o+}, z_a^{o-}]$. Similar to the attribute similarity measurement of crisp numbers, the distance-based method is employed. Then, the attribute similarity between Z_μ and Z^o concerning d_a can be calculated as follows:

$$Sim_a\left(Z^o, Z_\mu\right) = \exp\left[-Dis_{in}\left(z_{\mu a}^P, z_a^o\right)\right], \mu \in U, a \in \Omega^{III}, \tag{34}$$

where

$$Dis_{in}\left(z_{\mu a}^P, z_a^o\right) = \frac{\sqrt{\left(z_{\mu a}^{P+} - z_a^{o+}\right)^2 + \left(z_{\mu a}^{P-} - z_a^{o-}\right)^2}}{\max\left\{\sqrt{\left(z_{\mu a}^{P+} - z_a^{o+}\right)^2 + \left(z_{\mu a}^{P-} - z_a^{o-}\right)^2} \middle| \mu \in U\right\}}, \mu \in U, a \in \Omega^{III}. \tag{35}$$

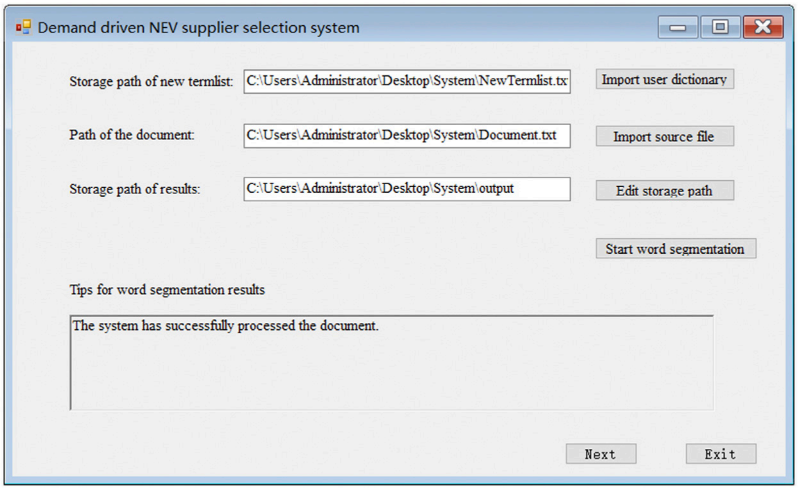

FIGURE 5
Text data word segmentation interface.

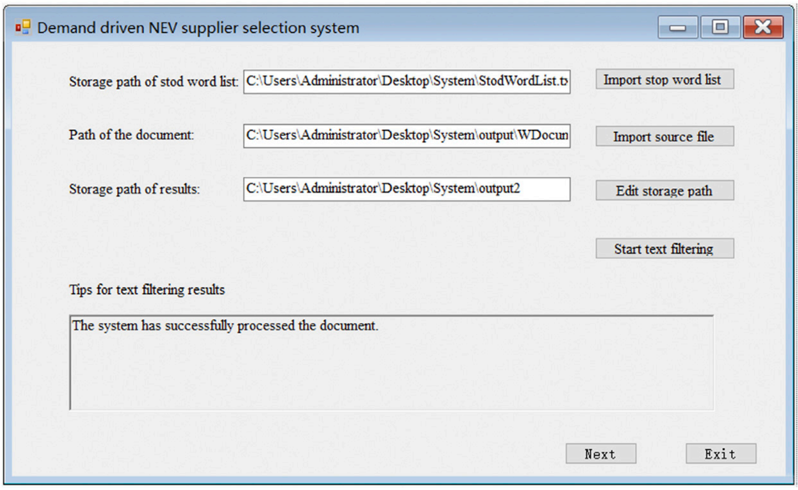

FIGURE 6
Stop word deletion interface.

4.3.1.4 Attribute similarity measurement for fuzzy linguistic variable format attribute values

For $d_a \in D^{IV}$, attribute values $z^P_{\mu a}$ and z^o_a are the fuzzy linguistic variables, where $z^P_{\mu a} = \tilde{z}^P_{\mu a}$ and $z^o_a = \tilde{z}^o_a$, $\tilde{z}^P_{\mu a}, \tilde{z}^o_a \in \Lambda$. In accordance with Jiang et al. (2008), the fuzzy linguistic variable F_V ($F_V \in \Lambda$) can be expressed by the triangular fuzzy number $\tilde{N}_V = (N^r_V, N^s_V, N^t_V)$, that is,

$$\tilde{N}_V = (N^r_V, N^s_V, N^t_V) = (\max((V-1)/T, 0), V/T, \min((V+1)/T, 1)), V \in \{0, 1, \ldots, T\}, \tag{36}$$

where N^r_V, N^s_V, and N^t_V are all the real numbers, $N^r_V \geq N^s_V \geq N^t_V \geq 0$. For instance, if $\Lambda = \{F_0 = DB$: definitely bad, $F_1 = VB$: very bad, $F_2 = B$: bad, $F_3 = M$: medium, $F_4 = G$: good, $F_5 = VG$: very good, $F_6 = DG$: definitely good $\}$, then, using Eq. 36, the fuzzy linguistic variables $(F_0, F_1, F_2, F_3, F_4, F_5, F_6)$ can be expressed by the corresponding triangular fuzzy numbers, that is, $\tilde{N}_0 = (0, 0, 0.17)$, $\tilde{N}_1 = (0, 0.17, 0.33)$, $\tilde{N}_2 = (0.17, 0.33, 0.5)$, $\tilde{N}_3 = (0.33, 0.5, 0.67)$, $\tilde{N}_4 = (0.5, 0.67, 0.83)$, $\tilde{N}_5 = (0.67, 0.83, 1)$, and $\tilde{N}_6 = (0.83, 1, 1)$.

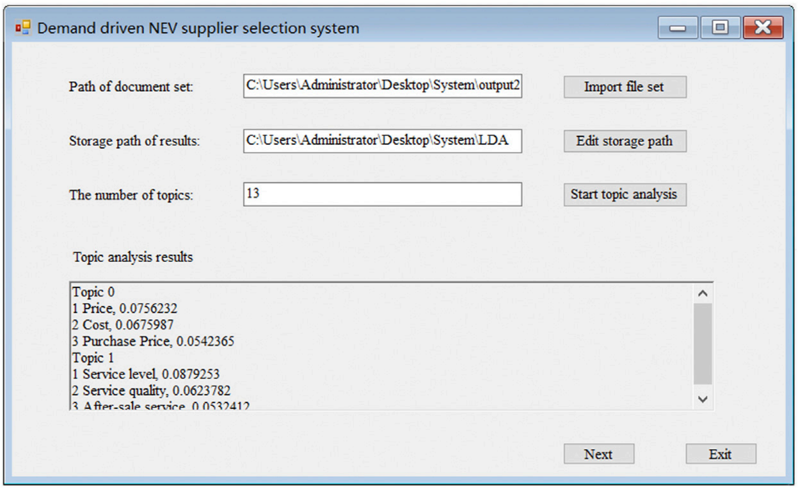

FIGURE 7
Topic analysis result interface.

TABLE 4 Extracted domain topics based on LDA.

Domain topic	Domain topic
Price (C_1)	Logistics cost (C_7)
Product quality (C_2)	Delivery date (C_8)
Service level (C_3)	Information integration capability (C_9)
Technical level (C_4)	Supply/production capacity (C_{10})
Product competitiveness (C_5)	After-sales service (C_{11})
Rate of qualified products (C_6)	R&D capability (C_{12})

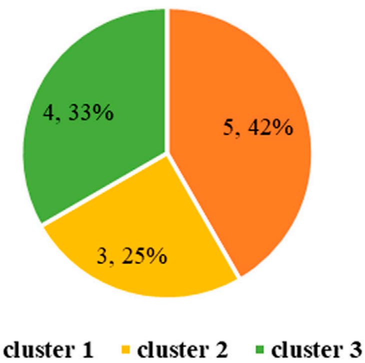

FIGURE 8
Clustering results of domain concepts.

Thus, $z^P_{\mu a}$ and z^o_a can be represented by $\tilde{z}^P_{\mu a} = (\tilde{z}^P_{\mu a}, z^{P-}_{\mu a}, \overline{z}^P_{\mu a})$ and $\tilde{z}^o_a = (\tilde{z}^o_a, z^{o-}_a, \overline{z}^o_a)$, respectively. Let $Dis_{fl}(\tilde{z}^P_{\mu a}, \tilde{z}^o_a)$ be the difference degree between $\tilde{z}^P_{\mu a}$ and $\tilde{z}^o_a$, then its calculation formula can be given by

$$Dis_{fl}\left(\tilde{z}^P_{\mu a}, \tilde{z}^o_a\right) = \frac{\sqrt{\left(\tilde{z}^P_{\mu a} - \tilde{z}^o_a\right)^2 + \left(z^{P-}_{\mu a} - z^{o-}_a\right)^2 + \left(\overline{z}^P_{\mu a} - \overline{z}^o_a\right)^2}}{\max\left\{\sqrt{\left(\tilde{z}^P_{\mu a} - \tilde{z}^o_a\right)^2 + \left(z^{P-}_{\mu a} - z^{o-}_a\right)^2 + \left(\overline{z}^P_{\mu a} - \overline{z}^o_a\right)^2}\,\middle|\,\mu \in U\right\}}, \mu \in U, a \in \Omega^{IV}. \tag{37}$$

Next, the attribute similarity between Z_μ and Z^o concerning d_a can be calculated, that is,

$$Sim_a\left(Z^o, Z_\mu\right) = \exp\left[-Dis_{fl}\left(\tilde{z}^P_{\mu a}, \tilde{z}^o_a\right)\right], \mu \in U, a \in \Omega^{IV}. \tag{38}$$

4.3.2 Hybrid similarity measurement

To retrieve and find the suitable NEV component supplier, the hybrid similarity needs to be measured. In accordance with the simple additive weighting method, the hybrid similarity between Z_μ and Z^o can be calculated based on the attribute similarities of different attribute value formats, that is,

$$Sim\left(Z^o, Z_\mu\right) = \sum\nolimits_{a=1}^{\delta} \omega^E_a \times Sim_a\left(Z^o, Z_\mu\right), \mu \in U, a \in \Omega^A. \tag{39}$$

Obviously, $Sim(Z^o, Z_\mu) \in [0, 1]$. The greater the $Sim(Z^o, Z_\mu)$ is, the more similar Z_μ with Z^o will be. Thus, according to $Sim(Z^o, Z_\mu)$, the ranking results of the NEV component suppliers can be obtained, so that the appropriate NEV component supplier can be selected.

5 Empirical study

An empirical study on the supplier selection of the XP NEV manufacturer is given to illustrate the use of the proposed method.

TABLE 5 Non-taxonomic relations of domain concepts.

Domain concept	Relation	Domain concept
Price (C_1)	Embody	Product competitiveness (C_5)
Technical level (C_4)	Ensure	Product quality (C_2)
Technical level (C_4)	Dominate	Supply/production capacity (C_{10})
Rate of qualified products (C_6)	Reflect	Product quality (C_2)
Rate of qualified products (C_6)	Reflect	Supply/production capacity (C_{10})
Supply/production capacity (C_{10})	Influence	Delivery date (C_8)

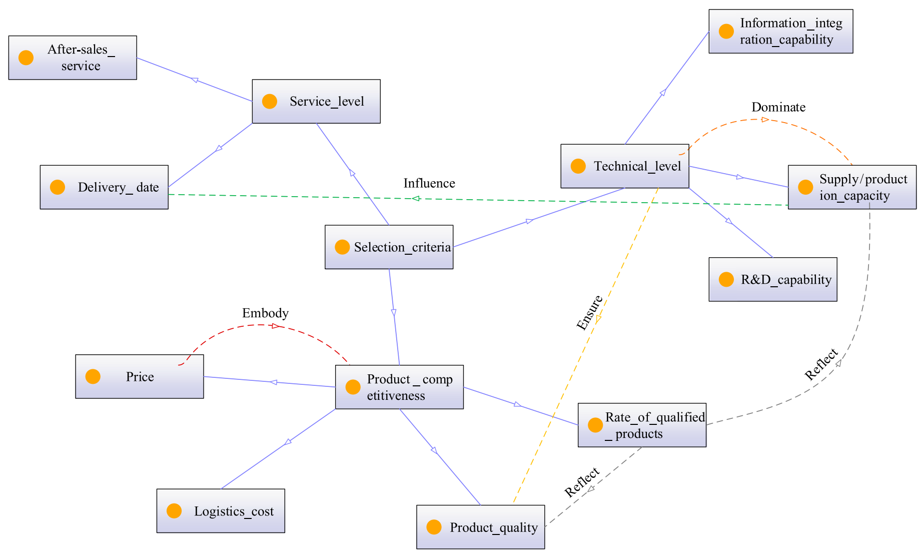

FIGURE 9
Initial domain ontology for NEV component supplier selection criteria.

5.1 Case description

The XP NEV manufacturer was founded in 2014 and is headquartered in Guangzhou. It is mainly engaged in NEV products such as electric vehicles and hybrid electric vehicles. XP manufacturers started the X project in 2019 and set up an expert group consisting of three sales managers and three senior product design engineers. The group is mainly responsible for the development and design of the EV3 series electric vehicles and plans to mass-produce them in 2022. In the stage of product design, XP manufacturers plan to purchase some required components through outsourcing, such as a power battery, motor, motor controller, and sensor. For this, four component suppliers ($Z_1, Z_2, Z_3,$ and Z_4) with relevant qualifications, production, and service experience were selected through online bidding, and the information about these suppliers was collected. In addition, XP manufacturers invited 500 users to experience EV3 series electric vehicle products, and users were asked to post their feelings and demands online in the form of documents. A total of 416 valid demand documents were collected through the network, that is, $P_1, P_2, \ldots, P_{416}$. Through statistical analysis of these demand documents, the expert group can integrate user demands into the product conceptual design so as to improve user satisfaction and product competitiveness.

5.2 Selection process of the NEV component supplier

To help an XP manufacturer select a suitable component supplier to carry out the X project smoothly, the method proposed in this paper was carried out. Some key calculation processes and results are described as follows.

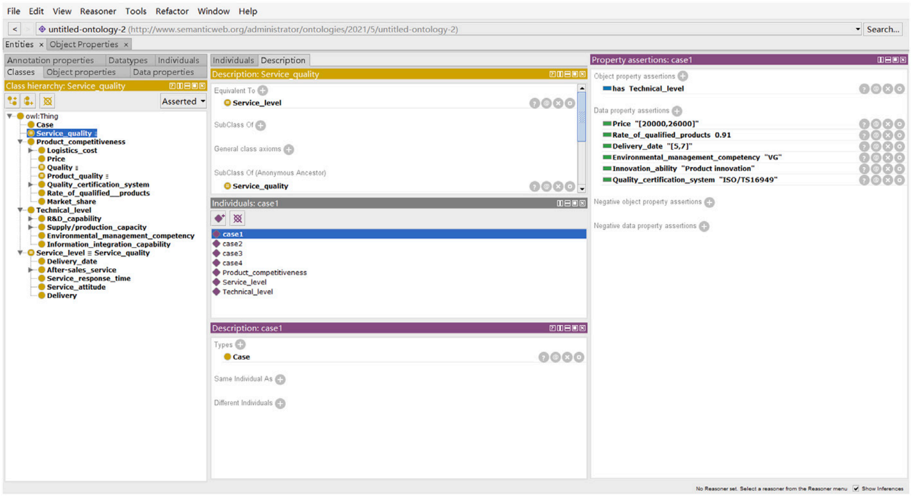

FIGURE 10
Domain ontology case for NEV component supplier selection criteria.

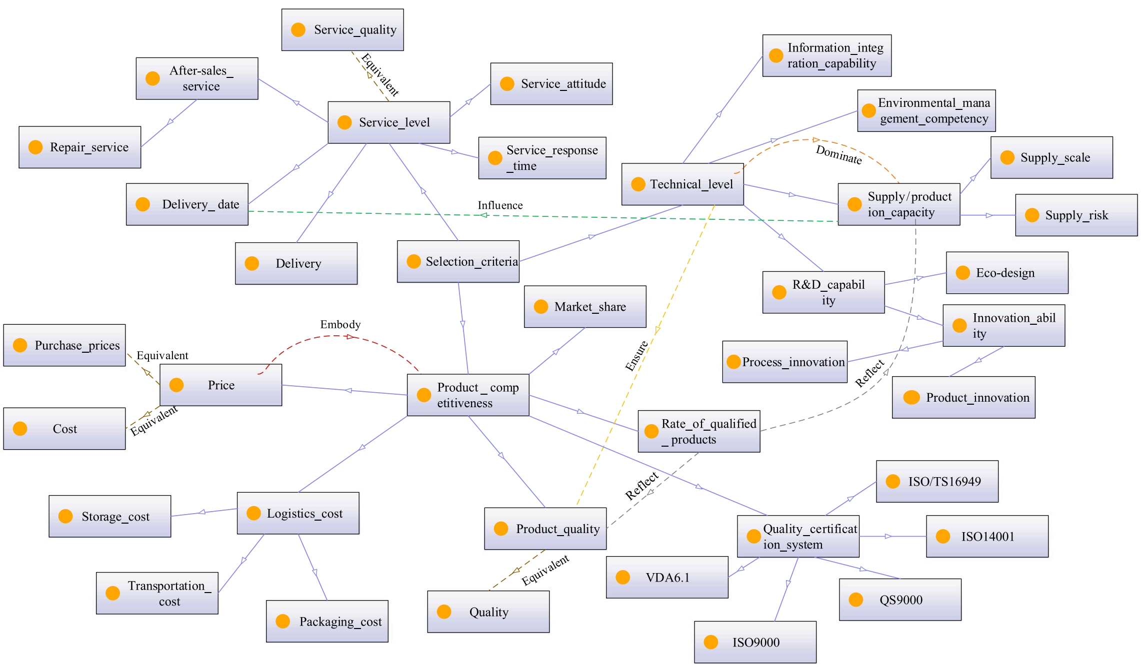

FIGURE 11
Domain ontology for NEV component supplier selection criteria.

5.2.1 Construction of domain ontology

First, "NEV," "selection and evaluation of supplier," and "automotive industry" were used as keywords and "2010–2021" was used as search years, and relevant references on the selection and evaluation of NEV suppliers were retrieved from the CNKI, WanFang, ScienceDirect, and Web of Science databases. The duplicate references were removed, and the set of domain literature was obtained, that is, $D = \{D_1, D_2, \ldots, D_{63}\}$.

TABLE 6 Processing results of the demand documents of users.

Demand document	Processing result (WD_k^P)
P_1	动力/n很/d给力/a,/wd提速/v很d快/a,/wd烧/vi油/n车/n完全/ad两/m个/q层次/n,/wd速度/n杠杠/n。/wj操控/v不错/a,/wd动力/n充沛/a,/wd自动/d泊/vg入/v车位/n很/d方便/a,/wd方向盘/n手感/n很/d棒/a,/wd刹车/n很/d脚/n感/vg,/wd加速/vi平顺/ns流畅/a。/wj电耗/n一般/a电车/n水平/n,/wd13.5/m kwh/100km/xu左右/m。/wj价格/n性价比/n绝对/d无敌/a,/wd很/d划算/a。/wj (The power is great, and the speed is very fast. It is completely at two levels with the fuel-burning car, and the speed is fast. The handling is good, and the power is abundant. It is very convenient to automatically park in the parking space, the steering wheel feels great, the brake feels good, and the acceleration is smooth. The power consumption is about 13.5 kwh/100 km, which is the average level of electric vehicles. The price and cost performance are absolutely invincible and very cost-effective.)
P_2	动力/n很/d足/a,/wd车速/n很/d快/a。/wj刹车/n一般/uyy,/wd总体/n可以/v,/wd真/a很/d省心/a,/wd操控/v非常/d棒/a,/wt电耗/n还/d可以/v,/wd11.2/m左右/m,/wd充电/vi很/d不错/a,/wd速度/n快/a,/wd很/d稳定/a,/wd安全系数/nl极/d高/a。/wj价格/n配置/vn比较/d厚道/a,/wd同/p价位/n性价比/n突出/a。/wj (The power is sufficient, and the speed is very fast. The brake is average, generally OK, and really worry-free, and the handling is great! The power consumption is okay, about 11.2, the charging is very good, the speed is fast and very stable, and the safety factor is extremely high. The price configuration is more kind, and the price-performance ratio is outstanding at the same price.)
⋮	⋮
P_{416}	动力/n可以/v,/wd开着/v还/d行/vi,/wd起步/vi轻盈/a,/wd挺/d快/a一/m款/q车/n。/wj当前/t车/n操控/v一般/a水准/n,/wd没有/d留下/v太/d深刻/a印象/n。/wj电耗/n觉得/v还/d可以/v,/wd有点/d担心/v总/b续航/vn,/wd总/d怕/v会/v突然/ad之间/f掉/v一/m大/a块/q电/n。/wj性价比/n还/d可以/v,/wd没有/d购置/v税/n,/wd养/v车/n压力/n比较/d小/a,/wd日常/b代/n步/qv挺/d不错/a。/wj (The power is good, the driving is OK, the start is light, and it is a fast car. The current handling of this car did not impress me too much, that is, the average level. The power consumption is okay, but I am little worried about the battery life, and I am afraid that a large amount of power will suddenly be lost. The cost performance is fairly good, there is no purchase tax, the pressure to maintain a car is relatively small, and the daily transportation is quite good.)

TABLE 7 Extracted demand attributes based on LDA.

Demand attribute	Frequent word
Power (E_1)	Start, acceleration, mode ⋯
Handling (E_2)	Control, steering, intelligence ⋯
Power consumption (E_3)	Charging, battery life, energy saving ⋯
Cost performance (E_4)	Price, configuration, cost ⋯

Then, as shown in Figures 5, 6, the domain literature $D = \{D_1, D_2, \ldots, D_{63}\}$ could be preprocessed, and the preprocessed domain literature set was obtained, that is, $D^{pro} = \{D_1^{pro}, D_2^{pro}, \ldots, D_{63}^{pro}\}$.

According to the process of extracting domain concepts proposed in this paper, LDA was used to extract domain topics from the preprocessed domain literature, where the parameters of LDA were set as $\alpha = 0.1$, $\beta = 0.01$, $K = 13$, and the number of iterations = 2,000. After filtering the noisy words in each domain topic, merging topics with similar meanings, and assigning a label to each domain topic, 12 domain topics were obtained, as shown in Table 4 and Figure 7.

According to the extracted domain concepts, the relations between domain concepts were extracted by the HAC algorithm and association rules. The results are shown in Figure 8 and Table 5 respectively. Then, the initial domain ontology of the NEV component supplier selection criteria could be obtained, and Protégé 5.2.0 software was used to visualize the initial domain ontology, as shown in Figure 9.

To further enrich the initial domain ontology, NVivo 11 software was used to code and analyze the set of domain literature $D = \{D_1, D_2, \ldots, D_{63}\}$ to deeply mine the relevant vocabulary of each concept in the initial domain ontology. The initial domain ontology was expanded, and the domain ontology case and the domain ontology are shown in Figures 10, 11, respectively.

5.2.2 Determination of expected attributes and their weights

By preprocessing 416 valid demand documents $P_1, P_2, \ldots, P_{416}$, the demand document word sets $WD_k^P = \{WD_{k1}^P, WD_{k2}^P, \ldots, WD_{kq}^P\}$ were obtained, $k \in \{1, 2, \ldots, 416\}$. The obtained demand document word sets are shown in Table 6.

Furthermore, the demand attribute E_b was extracted from the demand document word sets based on LDA, $k \in \{1, 2, \ldots, 416\}$ and $b \in \{1, 2, 3, 4\}$. After filtering the noisy words in each topic, merging topics with similar meanings, and assigning a label to each topic, four demand attributes were obtained, as shown in Table 7.

TABLE 8 Main sentiment words included in WR^+ and WR^-.

WR^+	WR^-
给力 (awesome), 不错 (good), 方便 (convenient), 流畅 (smooth), and 可靠 (reliable)	贵 (expensive), 垃圾 (trash), 虚电 (virtual electricity), 逊色 (inferior), and 迟钝 (dull)

TABLE 9 Attribute value in the form of multi-granularity sentiment strength distribution.

Demand document	Demand attribute			
	E_1	E_2	E_3	E_4
P_1	$P_{11}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$	$P_{12}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 1, & 1\le x<2 \\ 0, & x\ge 2 \end{cases}$	$P_{13}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 1, & 0\le x<1 \\ 0, & 1\le x<2 \\ 0, & x\ge 2 \end{cases}$	$P_{14}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$
P_2	$P_{21}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$	$P_{22}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$	$P_{23}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$	$P_{24}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$
⋮	⋮	⋮	⋮	⋮
P_{416}	$P_{416,1}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 1, & 1\le x<2 \\ 0, & x\ge 2 \end{cases}$	$P_{416,2}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 1, & 0\le x<1 \\ 0, & 1\le x<2 \\ 0, & x\ge 2 \end{cases}$	$P_{416,3}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$	$P_{416,4}(x)=\begin{cases} 0, & x<-2 \\ 0, & -2\le x<-1 \\ 0, & -1\le x<0 \\ 0, & 0\le x<1 \\ 0, & 1\le x<2 \\ 1, & x\ge 2 \end{cases}$

TABLE 10 Kano category of demand attribute E_b.

Demand attribute	Statistical results of questionnaires					Kano category	$Better_b$	$Worse_b$	Adjusted category
	A_b	O_b	M_b	I_b	R_b				
E_1	32.98%	15.95%	12.77%	38.30%	0	I	48.93%	−28.72%	I
E_2	23.40%	45.74%	19.15%	11.71%	0	O	69.15%	−64.90%	O
E_3	37.23%	20.21%	17.03%	25.53%	0	A	57.44%	−37.24%	I
E_4	40.42%	22.35%	9.57%	27.66%	0	A	62.77%	−31.92%	A

Based on WD_k^P and E_b, the demand information concerning E_b in each demand document, that is, WD_{kb}^P, was extracted by identifying the demand information, $k \in \{1, 2, \ldots, 416\}$ and $b \in \{1, 2, 3, 4\}$. Herein, the demand document P_1 was taken as an example to illustrate the extraction process. From Table 5, it was easy to see that P_1 involved the demand information concerning the four demand attributes, that is, power (E_1), handling (E_2), power consumption (E_3), and cost performance (E_4). After extracting the verbs, adverbs, and adjectives between two adjacent punctuations containing the words "power" (E_1), "handling" (E_2), "power consumption" (E_3), and "cost performance" (E_4), which were used to describe demand attributes, WD_{1b}^P could be determined, that is, $WD_{11}^P =$ {很/d, 给力/a, 充沛/a} (very/d, awesome/a, and abundant/a), $WD_{12}^P =$ {不错/a} (good/a), $WD_{13}^P =$ {一般/a} (average/a), and $WD_{14}^P =$ {绝对/d, 无敌/a} (absolutely/d and invincible/a).

Using Eqs 14–16, the positive sentiment dictionary and negative sentiment dictionary concerning demand attributes were established, that is, WR^+ and WR^-, as shown in Table 8.

Using the proposed sentiment analysis algorithm, the sentiment strength of the demand document concerning each demand attribute, that is, s_{kb}^P, was calculated, $k \in \{1, 2, \ldots, 416\}$ and $b \in \{1, 2, 3, 4\}$. Herein, WD_{11}^P was taken as an example to illustrate the process of determining sentiment strength value, where $WD_{11}^P =$ {很/d, 给力/a, 充沛/a} (very/d, awesome/a, and abundant/a). Obviously, there were an adverb "很" (very) and two positive sentiment words "给力" (awesome) and "充沛" (abundant) in WD_{11}^P, which meant $WD_{11}^P \cap WR^+ \neq \varnothing$,

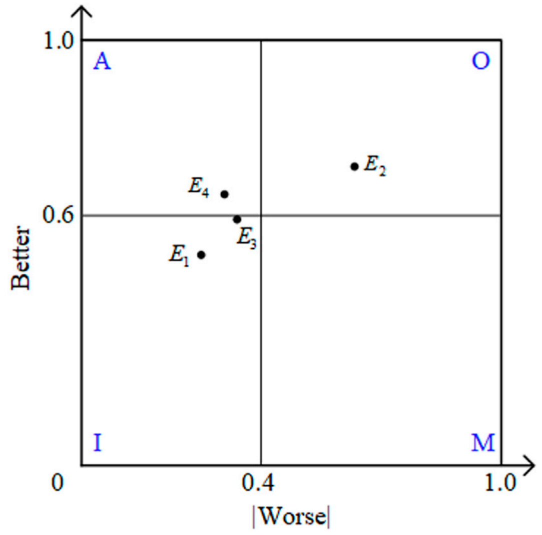

FIGURE 12
Better-Worse coefficient quadrant chart.

$WD_{11}^{P} \cap WR^{-} = \varnothing$, $WD_{11}^{P} \cap WR_{neg} = \varnothing$, and $WD_{11}^{P} \cap WR_{d} \neq \varnothing$. According to steps 2–5 of the sentiment analysis algorithm, the values of indicator variables were determined, that is, $s_{11}^{P+} = 1$, $s_{11}^{P-} = 0$, $s_{11}^{Pneg} = 0$, and $s_{11}^{Pd} = 1$. Then, according to step 6, the value of s_{11}^{P} was calculated, that is, $s_{11}^{P} = 2$. Thus, the sentiment strengths concerning demand attributes could be obtained.

Furthermore, the value of s_{kb}^{P} was transformed into an indicator vector T_{kb}^{P}, $k \in \{1, 2, \ldots, 416\}$ and $b \in \{1, 2, 3, 4\}$. Also, using Eq. 17, the attribute value in the form of multi-granularity sentiment strength distribution, that is, $P_{kb}(x)$, was determined, as shown in Table 9, $k \in \{1, 2, \ldots, 416\}$ and $b \in \{1, 2, 3, 4\}$.

Using Eqs 18–21, the weights of demand attributes were obtained, that is, $\omega_1^{P} = 0.3787$, $\omega_2^{P} = 0.2052$, $\omega_3^{P} = 0.1500$, and $\omega_4^{P} = 0.2661$. Then, 130 questionnaires were sent out to users who have experienced EV3 series electric vehicle products, and 94 valid questionnaires were collected. Using Eqs 22–23, the Kano category of each demand attribute and the satisfaction and dissatisfaction coefficients of users concerning demand attributes were obtained. The results are shown in Table 10 and Figure 12 respectively.

Using Eq. 24, the adjusted demand attribute weights were obtained, that is, $\omega_1^{P'} = 0.2259$, $\omega_2^{P'} = 0.3672$, $\omega_1^{P'} = 0.0894$, and $\omega_1^{P'} = 0.3175$.

According to the constructed domain ontology and the Kano category of each demand attribute, the demand attributes were transformed into the expected attributes by an expert group, and the relevant description of the expected attribute is shown in Table 11. The correlation degree between the demand attribute and the expected attribute was evaluated by expert group discussion, that is, R_{ba}, $b \in \{1, 2, 3, 4\}$ and $a \in \{1, 2, \ldots, 6\}$. Then, the weight of each expected attribute was calculated, as shown in Table 12.

5.2.3 Selection of the NEV component supplier

According to d_a, the expectation toward the supplier Z^o (i.e., target case) was determined by the expert group through the investigation and analysis of the production plan and actual demand of X project, as well as the relevant information of the component suppliers that XP manufacturer has successfully cooperated with. Meanwhile, the relevant information of alternative component suppliers concerning expected attributes was sorted out and regarded as historical cases, that is, Z_1, Z_2, Z_3, and Z_4. The expected attribute values involved in historical cases and target cases are shown in Table 13.

Using Eqs 27–38, the similarity between the historical case and the target case concerning d_a was calculated, that is, $Sim_a(Z^o, Z_\mu)$, $\mu \in \{1, 2, 3, 4\}$ and $a \in \{1, 2, \ldots, 6\}$, as shown in Table 14.

Using Eq. 39, the hybrid similarity between the historical case and the target case, that is, $Sim(Z^o, Z_\mu)$, could be calculated, as shown in Figure 13. The computation results were $Sim(Z^o, Z_1) = 0.5945$, $Sim(Z^o, Z_2) = 0.5206$, $Sim(Z^o, Z_3) = 0.8122$, and $Sim(Z^o, Z_4) = 0.6370$. Also, a ranking order of the

TABLE 11 Expected attributes and related descriptions.

Expected attribute	Expected attribute value format	Description of the desired attribute
Environmental management competency (d_1)	Fuzzy linguistic variables	Economic benefits and environmental achievements obtained through environmental management, energy conservation, CO_2 emission reduction, and other activities
Price (d_2)	Interval numbers	Wholesale price of components
Rate of qualified products (d_3)	Crisp numbers	Percentage of the number of components meeting the quality standard in the total number products
Innovation ability (d_4)	Ontology concepts	Innovative activities to achieve the organizational goals using the scientific knowledge system and advanced technology
Delivery date (d_5)	Interval numbers	Time from order placement by the manufacturing company to receipt of parts provided by the supplier
Quality certification system (d_6)	Ontology concepts	Quality system certification obtained by the supplier

TABLE 12 Relationship matrix between demand attributes and expected attributes.

Demand attribute	Demand attribute weight	Expected attribute					
		Environmental management competency (d_1)	Price (d_2)	Rate of qualified products (d_3)	Innovation ability (d_4)	Delivery date (d_5)	Quality certification system (d_6)
E_1	0.2259	5	5	5	7		5
E_2	0.3672		5	5	7		5
E_3	0.0894	9	5	5	7		5
E_4	0.3175	7	9	7	5	5	7
Importance weights of expected attributes		4.7916	6.2700	5.6350	6.3650	1.5875	5.6350
Expected attribute weights		0.1582	0.2070	0.1861	0.2102	0.0524	0.1861

TABLE 13 Expected attribute values involved in historical cases and target case.

Historical case and target case	Expected attribute					
	d_1	d_2	d_3	d_4	d_5	d_6
Z_1	VG	(20000, 26000)	91%	Product innovation	(5,7)	ISO/TS16949
Z_2	M	(16500, 19000)	79%	Process innovation	(4.5,6)	QS9000
Z_3	G	(18500, 23700)	86%	Process innovation	(4,7.5)	ISO/TS16949
Z_4	G	(15500, 24500)	87%	Product innovation	(3.5,8)	VDA6.1
Z^o	G	(15000, 24000)	9%	Process innovation	(5,6)	ISO/TS16949

TABLE 14 Computation results of $Sim_a(Z^o, Z_\mu)$.

Historical case	Similarity					
	$Sim_1(Z^o, Z_\mu)$	$Sim_2(Z^o, Z_\mu)$	$Sim_3(Z^o, Z_\mu)$	$Sim_4(Z^o, Z_\mu)$	$Sim_5(Z^o, Z_\mu)$	$Sim_6(Z^o, Z_\mu)$
Z_1	0.3679	0.3679	0.9131	0.3396	0.6703	0.9875
Z_2	0.3679	0.3793	0.3679	0.9875	0.8187	0.3489
Z_3	1	0.5208	0.6951	0.9875	0.4862	0.9875
Z_4	1	0.8769	0.7613	0.3396	0.3679	0.3489

four alternative component suppliers could be determined, that is, $Z_3 \succ Z_4 \succ Z_1 \succ Z_2$. Therefore, compared with other suppliers, Z_3 was more in line with the expectations of the expert group and should be selected as the component supplier of XP manufacturer.

5.3 Comparative analysis

In order to further analyze the advantages of the proposed method, we compare it with a similar method that was proposed by Yang and Chai (2018). Based on the case background and original data in this study, the method in this literature (Yang and Chai 2018) was used, and the ranking results of the alternative NEV component suppliers were $Z_3 \succ Z_1 \succ Z_4 \succ Z_2$, as shown in Figure 14. As can be seen from Figure 14, the ranking results obtained by our method ($Z_3 \succ Z_4 \succ Z_1 \succ Z_2$) were generally consistent with those obtained by Yang and Chai (2018), that is, Z_3 was the most appropriate of the four alternative suppliers and Z_2 was the last alternative supplier in the ranking. However, the ranking order of Z_1 and Z_4 was different in the results obtained by using the two methods. The main reason is that

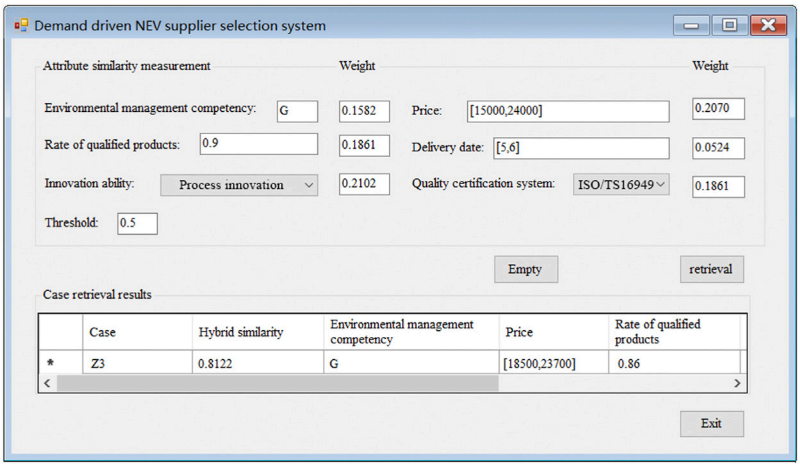

FIGURE 13
Screenshot of retrieval results.

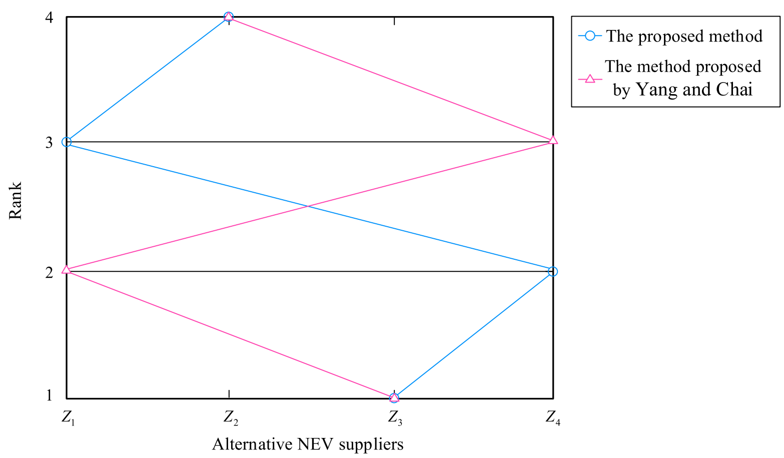

FIGURE 14
Ranking of the alternative NEV suppliers obtained by different methods.

Yang and Chai (2018) only considered the two attribute value forms (i.e., crisp numbers and fuzzy linguistic variables) in the process of hybrid similarity calculation. Multiple forms of attribute values are considered in the proposed method and integrate semantic similarity into the hybrid similarity calculation for retrieval, which improves the accuracy of case retrieval.

In addition, in terms of acquiring demand attributes, Yang and Chai (2018) obtain demand attributes through questionnaires, which cannot fully reflect the complex thoughts and opinions of users on a demand and are costly and expensive, while demand documents are collected through the network in this study and demand attributes are obtained by mining demand documents. This makes the user's demand expression unconstrained, thereby improving the accuracy and comprehensiveness of the demand information obtained. In determining the expected attributes of suppliers, Yang and Chai (2018) adopt the Delphi method to determine the expected attributes, which is highly subjective. However, this study determines expected attributes based on the mined user demands and the constructed domain ontology of NEV component supplier selection criteria, which is significantly more objective and realistic.

6 Conclusion and future work

This paper presents an integrated method for demand-driven NEV supplier selection based on ontology–QFD–CBR. In the method, the literature on the selection criteria of an NEV component supplier is first preprocessed. Second, the domain concepts are extracted using LDA, and the HAC algorithm and association rules are used to extract the relations between domain concepts so as to construct the domain ontology of NEV component supplier selection criteria. Then, demand attributes and their weights are determined based on LDA, the information entropy method, and the Kano model. On the basis of this, the expected attributes and their weights are determined based on QFD. Furthermore, the attribute similarities and hybrid similarities between the alternative cases and target cases are calculated based on ontology theory and CBR, and the most suitable NEV component supplier is selected. The four major contributions of this study are as follows.

First, a new solution framework for demand-driven NEV supplier selection based on ontology–QFD–CBR is proposed. Compared with the existing ones, free-form documents are used in an information acquisition way instead of traditional ways such as scales and questionnaires so as to enable users to express their demands more freely and improve the accuracy of obtaining demand information. Moreover, the framework integrates the merits of ontology, QFD, and CBR. To our knowledge, no previous studies have investigated the problem of NEV component supplier selection with this kind of integrated method.

Second, in the construction of supplier selection criteria and the determination of their weight, the domain ontology of NEV component supplier selection criteria is constructed based on text information mining, and the logical and semantic relationships of selection criteria are fully considered. In addition, the weights of selection criteria are determined based on QFD, which considers the demands of users and ensures objectivity.

Third, in the supplier selection decision, a concept semantic similarity calculation method is proposed based on semantic distance, node depth, and node density, and the semantic similarity is integrated into the calculation of hybrid similarity for retrieval, which is an extension to the traditional hybrid similarity calculation method and can improve the accuracy and efficiency of supplier case retrieval and provide good scalability and sharing for domain knowledge. In addition, the alternative suppliers are compared with the decision-maker's expectations, and multiple forms of attribute value formats (i.e., crisp numbers, interval numbers, fuzzy linguistic variables, and ontology concepts) are considered comprehensively. The supplier that meets the decision-maker's expectations is selected, which is closer to the reality and considered more comprehensively.

Fourth, in the processing of demand information and the determination of the weights of demand attributes, two improvements have been made. One is the multi-granularity sentiment analysis algorithm, which is developed to capture different sentiment strengths contained in different sentiment polarities. It is an extension to the traditional sentiment strength method and can effectively avoid the loss of information caused by only considering the positive and negative sentiment polarities. The other is to determine the weights of demand attributes using entropy weight and considering the Kano category of demand attributes, which fully embodies the idea of "letting the data speak for themselves," ensures objectivity and authenticity, and avoids the influence of subjective factors to a large extent.

In terms of future work, three interesting directions can be considered. First, to help NEV manufacturers manage component suppliers more effectively, a decision support system embedded in the proposed method can be developed. Second, intelligent algorithms such as deep learning will also be combined to rank a large number of potential suppliers so as to select suitable suppliers. Third, the application of the proposed method will be extended to solve the problems of supplier selection in other sectors, such as the oil industry and the paper industry.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by JZ, Y-HL, and Z-PF. The first draft of the manuscript was written by JZ, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was partly supported by the National Science Foundation of China (grant number 72031002), the China Postdoctoral Science Foundation (grant number 2019M653794), the Philosophy and Social Science Innovative Talent Support Program of Henan (grant number 2022-CXRC-27), the Training Program for Young Key Teacher of Higher Education Institutions in Henan Province (grant number 2021GGJS060), the Logistics Research Center Project of the Key Research Base of Humanities and Social Sciences in Henan Province (grant number 2015-JD-04) and the 111 Project of China (grant number B16009).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adnan, N., Nordin, S. M., Rahman, I., and Amini, M. H. (2017). A market modeling review study on predicting Malaysian consumer behavior towards widespread adoption of PHEV/EV. *Environ. Sci. Pollut. Res.* 24 (22), 17955–17975. doi:10.1007/s11356-017-9153-8

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. I. (1996). "Fast discovery of association rules," in *Advances in knowledge discovery and data mining* (California, USA: AAAI/MIT Press), 307–328.

Ayağ, Z., and Samanlioglu, F. (2014). An intelligent approach to supplier evaluation in automotive sector. *J. Intell. Manuf.* 27, 889–903. doi:10.1007/s10845-014-0922-7

Azizi, A., Aikhuele, D. O., and Souleman, F. S. (2015). A fuzzy TOPSIS model to rank automotive suppliers. *Procedia Manuf.* 2, 159–164. doi:10.1016/j.promfg.2015.07.028

Balazs, J. A., and Velásquez, J. D. (2016). Opinion mining and information fusion: A survey. *Inf. Fusion* 27, 95–110. doi:10.1016/j.inffus.2015.06.002

Bastida-Molina, P., Hurtado-Pérez, E., Pérez-Navarro, Á., and Alfonso-Solar, D. (2021). Light electric vehicle charging strategy for low impact on the grid. *Environ. Sci. Pollut. Res.* 28 (15), 18790–18806. doi:10.1007/s11356-020-08901-2

Berger, C., Blauth, R., Bolster, C., Burchill, G., Dumouchel, W., Pouliot, F., et al. (1993). Kano's methods for understanding customer-defined quality. *Cent. Qual. Manag. J.* 2 (4), 3–36.

Bi, J. W., Liu, Y., Fan, Z. P., and Cambria, E. (2019a). Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *Int. J. Prod. Res.* 57 (22), 7068–7088. doi:10.1080/00207543.2019.1574989

Bi, J. W., Liu, Y., Fan, Z. P., and Zhang, J. (2019b). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tour. Manag.* 70, 460–478. doi:10.1016/j.tourman.2018.09.010

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.

Büyüközkan, G., and Çifçi, G. (2011). A novel fuzzy multi-criteria decision framework for sustainable supplier selection with incomplete information. *Comput. Industry* 62 (2), 164–174. doi:10.1016/j.compind.2010.10.009

Cano, Z. P., Banham, D., Ye, S., Hintennach, A., Lu, J., Fowler, M., et al. (2018). Batteries and fuel cells for emerging electric vehicle markets. *Nat. Energy* 3 (4), 279–289. doi:10.1038/s41560-018-0108-1

Castells, P., Fernandez, M., and Vallet, D. (2007). An adaptation of the vector-space model for ontology-based information retrieval. *IEEE Trans. Knowl. Data Eng.* 19 (2), 261–272. doi:10.1109/tkde.2007.22

Chen, P. (2020). Effects of the entropy weight on TOPSIS. *Expert Syst. Appl.* 168 (8), 114186. doi:10.1016/j.eswa.2020.114186

Chen, R. C., Liang, J. Y., and Pan, R. H. (2008). Using recursive ART network to construction domain ontology based on term frequency and inverse document frequency. *Expert Syst. Appl.* 34 (1), 488–501. doi:10.1016/j.eswa.2006.09.019

Dai, Q., and Zhu, J. (2019). Research on the classification method of auto parts suppliers based on the typical sample set. *J. Nanjing Univ. Sci. Technol.* 43 (4), 527–534.

Dargi, A., Anjomshoae, A., Galankashi, M. R., Memari, A., and Tap, M. B. M. (2014). Supplier selection: A fuzzy-ANP approach. *Procedia Comput. Sci.* 31, 691–700. doi:10.1016/j.procs.2014.05.317

Deng, X., Liu, L., and Zhang, R. (2021). Optimal design of fitness game system based on Kano-QFD. *Packag. Eng.* 42 (14), 148–154.

Dong, L., Hu, Y., and Zhang, X. (2013). Method for non-taxonomical relations from domain concepts. *Comput. Eng. Appl.* 49 (4), 157.

Du, Y. W., and Gao, K. (2020). Ecological security evaluation of marine ranching with AHP-entropy-based TOPSIS: A case study of yantai, China. *Mar. Policy* 122, 104223. doi:10.1016/j.marpol.2020.104223

Dweiri, F., Kumar, S., Khan, S. A., and Jain, V. (2016). Designing an integrated AHP based decision support system for supplier selection in automotive industry. *Expert Syst. Appl.* 62, 273–283. doi:10.1016/j.eswa.2016.06.030

Eydi, A., and Fazli, L. (2019). A decision support system for single-period single sourcing problem in supply chain management. *Soft Comput.* 23 (24), 13215–13233. doi:10.1007/s00500-019-03864-0

Fan, Z. P., Chen, Z. W., and Zhao, X. (2020). Battery outsourcing decision and product choice strategy of an electric vehicle manufacturer. *Int. Trans. Oper. Res.* 29, 1943–1969. doi:10.1111/itor.12814

Fan, Z. P., Li, Y. H., Wang, X. H., and Liu, Y. (2014). Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion. *Expert Syst. Appl.* 41 (5), 2526–2534. doi:10.1016/j.eswa.2013.09.051

Feng, J., and Zhang, Y. (2017). Research on the method of detecting and analyzing scientific fronts based on LDA and ontology. *Inf. Studies:Theory Appl.* 40 (8), 49–54.

Galankashi, M. R., Helmi, S. A., and Hashemzahi, P. (2016). Supplier selection in automobile industry: A mixed balanced scorecard–fuzzy AHP approach. *Alexandria Eng. J.* 55 (1), 93–100. doi:10.1016/j.aej.2016.01.005

Gong, H., Wang, M. Q., and Wang, H. (2013). New energy vehicles in China: Policies, demonstration, and progress. *Mitig. Adapt. Strateg. Glob. Chang.* 18 (2), 207–228. doi:10.1007/s11027-012-9358-6

Guo, Y., Barnes, S. J., and Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tour. Manag.* 59, 467–483. doi:10.1016/j.tourman.2016.09.009

Guo, Y., and Zhou, J. (2017). Data mining modeling and its application in CAPP by integrating ontology and CBR. *Mech. Sci. Technol. Aerosp. Eng.* 36 (4), 579–585.

Gupta, S., Soni, U., and Kumar, G. (2019). Green supplier selection using multi-criterion decision making under fuzzy environment: A case study in automotive industry. *Comput. Industrial Eng.* 136, 663–680. doi:10.1016/j.cie.2019.07.038

Han, P., Li, Y., Yin, Y., and An, N. (2019). Cham, 34–45.Ontology construction for eldercare services with an agglomerative hierarchical clustering method*Proc. Hum. Aspects IT Aged Popul. Des. Elder. Technol. Accept.*

Hao, Z., Li, Z., Ni, H., Lv, S., Wang, X., and Zhu, Y. (2021). Present situation and prospect of new energy vehicle industry in China. *IOP Conf. Ser. Earth Environ. Sci.* 791, 012153. doi:10.1088/1755-1315/791/1/012153Qingdao, China

He, J. K. (2016). Global low-carbon transition and China's response strategies. *Adv. Clim. Change Res.* 7 (4), 204–212. doi:10.1016/j.accre.2016.06.007

He, X., Zhan, W., and Hu, Y. (2018). Consumer purchase intention of electric vehicles in China: The roles of perception and personality. *J. Clean. Prod.* 204, 1060–1069. doi:10.1016/j.jclepro.2018.08.260

Hendiani, S., Mahmoudi, A., and Liao, H. (2020). A multi-stage multi-criteria hierarchical decision-making approach for sustainable supplier selection. *Appl. Soft Comput.* 94, 106456. doi:10.1016/j.asoc.2020.106456

Högström, C., Rosner, M., and Gustafsson, A. (2010). How to create attractive and unique customer experiences. *Mark. Intell. Plan.* 28 (4), 385–402. doi:10.1108/02634501011053531

Homem, T. P. D., Santos, P. E., Reali Costa, A. H., da Costa Bianchi, R. A., and Lopez de Mantaras, R. (2020). Qualitative case-based reasoning and learning. *Artif. Intell.* 283, 103258. doi:10.1016/j.artint.2020.103258

Hosseini, S., and Sarder, M. D. (2019). Development of a Bayesian network model for optimal site selection of electric vehicle charging station. *Int. J. Electr. Power & Energy Syst.* 105 (4), 110–122. doi:10.1016/j.ijepes.2018.08.011

Huang, G., and Zhou, Z. (2007). Research on domain ontology-based concept semantic similarity calculation. *Comput. Eng. Des.* 28 (10), 2460–2463.

Huang, S. L., and Cheng, W. C. (2015). Discovering Chinese sentence patterns for feature-based opinion summarization. *Electron. Commer. Res. Appl.* 14 (6), 582–591. doi:10.1016/j.elerap.2015.08.007

Huth, C., Kieckhäfer, K., and Spengler, T. S. (2015). Make-or-buy strategies for electric vehicle batteries—A simulation-based analysis. *Technol. Forecast. Soc. Change* 99, 22–34. doi:10.1016/j.techfore.2015.06.027

Ilyas, M., Carpitella, S., and Zoubir, E. (2021). Designing supplier selection strategies under COVID-19 constraints for industrial environments. *Procedia CIRP* 100, 589–594. doi:10.1016/j.procir.2021.05.128

Jain, V., Sangaiah, A. K., Sakhuja, S., Thoduka, N., and Aggarwal, R. (2018). Supplier selection using fuzzy AHP and TOPSIS: A case study in the Indian automotive industry. *Neural comput. Appl.* 29 (7), 555–564. doi:10.1007/s00521-016-2533-z

Jiang, Y. P., Fan, Z. P., and Ma, J. (2008). A method for group decision making with multi-granularity linguistic assessment information. *Inf. Sci.* 178 (4), 1098–1109. doi:10.1016/j.ins.2007.09.007

Kano, N., Seraku, N., Takahashi, F., and Tsuji, S. (1984). Attractive quality and must-be quality. *J. Jpn. Soc. Qual. Control* 14 (2), 147–156.

Kendall, M. (2018). Fuel cell development for New Energy Vehicles (NEVs) and clean air in China. *Prog. Nat. Sci. Mater. Int.* 28 (2), 113–120. doi:10.1016/j.pnsc.2018.03.001

Khan, S. A. K., Dweiri, F., and Jain, V. (2016). Integrating analytical hierarchy process and quality function deployment in automotive supplier selection. *Int. J. Bus. Excell.* 9 (2), 156–177. doi:10.1504/ijbex.2016.074851

Konidari, P., and Mavrakis, D. (2007). A multi-criteria evaluation method for climate change mitigation policy instruments. *Energy Policy* 35 (12), 6235–6257. doi:10.1016/j.enpol.2007.07.007

Küçük, D. (2015). A high-level electrical energy ontology with weighted attributes. *Adv. Eng. Inf.* 29 (3), 513–522. doi:10.1016/j.aei.2015.04.002

Le, T-D. B., and Lo, D. (2015). Montreal, QC, Canada, 331–340.Beyond support and confidence: Exploring interestingness measures for rule-based specification mining, Proceedings of the 2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)

Lee, C-H., Wang, Y-H., and Trappey, A. J. C. (2015). Ontology-based reasoning for the intelligent handling of customer complaints. *Comput. Industrial Eng.* 84, 144–155. doi:10.1016/j.cie.2014.11.019

Li, L., and Jing, X. (2019). Xiamen, China, 315–320.Analysis of tax policy for promoting the development of China's new energy vehicles industry, Proceedings of the Proceedings of the 2019 4th International Conference on Humanities Science and Society Development (ICHSSD 2019)

Li, S., Yan, Z., Jian, L., and Xu, J. (2015). Leicester, UK, 521–527.Study on auto parts suppliers composition selection based on adaptive genetic algorithm, Proceedings of the 2015 IEEE International Conference on Grey Systems and Intelligent Services (GSIS)

Li, Y-H., Zheng, J., Fan, Z-P., and Wang, L. (2021a). Sentiment analysis-based method for matching creative agri-product scheme demanders and suppliers: A case study from China. *Comput. Electron. Agric.* 186, 106196. doi:10.1016/j.compag.2021.106196

Li, Y-H., Zheng, J., Yue, S-T., and Fan, Z-P. (2021b). Capturing and analyzing e-WOM for travel products: A method based on sentiment analysis and stochastic dominance. *K.* (In press). doi:10.1108/k-02-2021-0146

Li, Z., Li, D., and Zhao, P. (2018). Research on automatic extraction of ontology concept and its relation in e-commerce. *Inf. Sci.* 36 (7), 85–90.

Liao, H., and Xu, Z. (2015). Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSs and their application in qualitative decision making. *Expert Syst. Appl.* 42 (12), 5328–5336. doi:10.1016/j.eswa.2015.02.017

Lima-Junior, F. R., and Carpinetti, L. C. R. (2016). A multicriteria approach based on fuzzy QFD for choosing criteria for supplier selection. *Comput. Industrial Eng.* 101, 269–285. doi:10.1016/j.cie.2016.09.014

Liu, A., Xiao, Y. X., Lu, H., Tsai, S-B., and Song, W. (2019). A fuzzy three-stage multi-attribute decision-making approach based on customer needs for sustainable supplier selection. *J. Clean. Prod.* 239, 118043. doi:10.1016/j.jclepro.2019.118043

Liu, D., Xu, L., Sadia, U. H., and Wang, H. (2021a). Evaluating the CO2 emission reduction effect of China's battery electric vehicle promotion efforts. *Atmos. Pollut. Res.* 12 (7), 101115. doi:10.1016/j.apr.2021.101115

Liu, G., Fan, S., Tu, Y., and Wang, G. (2021b). Innovative supplier selection from collaboration perspective with a hybrid MCDM model: A case study based on NEVs manufacturer. *Symmetry* 13 (1), 143. doi:10.3390/sym13010143

Liu, H., Zhang, D. X., and Shi, F. M. (2016). First record of sonotrella gorochov (orthoptera: Gryllidae) from Laos, with description of a new species. *Zootaxa* 36 (24), 189–194. doi:10.11646/zootaxa.4114.2.8

Liu, P., Hu, Y., Wang, X., and Liu, K. (2011). Shanghai, China, 1–5.A methodology for domain ontology construction in information science, Proceedings of the 2011 International Conference on E-Business and E-Government (ICEE)

Liu, Y., Bi, J. W., and Fan, Z. P. (2017). A method for ranking products through online reviews based on sentiment classification and interval-valued intuitionistic fuzzy TOPSIS. *Int. J. Info. Tech. Dec. Mak.* 16 (6), 1497–1522. doi:10.1142/s021962201750033x

Lu, C., Liu, H. C., Tao, J., Rong, K., and Hsieh, Y. C. (2017). A key stakeholder-based financial subsidy stimulation for Chinese EV industrialization: A system dynamics simulation. *Technol. Forecast. Soc. Change* 118, 1–14. doi:10.1016/j.techfore.2017.01.022

Lu, C., Tao, J., An, Q., and Lai, X. (2020). A second-order cone programming based robust data envelopment analysis model for the new-energy vehicle industry. *Ann. Oper. Res.* 292 (1), 321–339. doi:10.1007/s10479-019-03155-9

Luo, Y., Chen, X., Fang, F., Zhang, X., and Guo, N. (2021). A trust-enhanced and preference-aware collaborative method for recommending new energy vehicle. *Environ. Sci. Pollut. Res.* 28 (7), 7901–7917. doi:10.1007/s11356-020-10834-9

Manello, A., and Calabrese, G. (2019). The influence of reputation on supplier selection: An empirical study of the European automotive industry. *J. Purch. Supply Manag.* 25 (1), 69–77. doi:10.1016/j.pursup.2018.03.001

Mathiyazhagan, K., Sudhakar, S., and Bhalotia, A. (2018). Modeling the criteria for selection of suppliers towards green aspect: A case in Indian automobile industry. *OPSEARCH* 55 (1), 65–84. doi:10.1007/s12597-017-0315-8

Matzler, K., and Hinterhuber, H. H. (1998). How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. *Technovation* 18 (1), 25–38. doi:10.1016/s0166-4972(97)00072-2

Memari, A., Dargi, A., Akbari Jokar, M. R., Ahmad, R., and Abdul, A. R. A. R. (2019). Sustainable supplier selection: A multi-criteria intuitionistic fuzzy TOPSIS method. *J. Manuf. Syst.* 50, 9–24. doi:10.1016/j.jmsy.2018.11.002

Mou, N., Chang, J., and Chen, Z. (2018). Sustainable supplier selection based on PD-HFLTS and group decision-making theory. *Comput. Integr. Manuf. Syst.* 24 (5), 1261–1278.

Noori, A., Bonakdari, H., Salimi, A. H., and Gharabaghi, B. (2021). A group Multi-Criteria Decision-Making method for water supply choice optimization. *Socio-Economic Plan. Sci.* 77, 101006. doi:10.1016/j.seps.2020.101006

Park, S. C., and Lee, J. H. (2017). Supplier selection and stepwise benchmarking: A new hybrid model using DEA and AHP based on cluster analysis. *J. Operational Res. Soc.* 69 (3), 449–466. doi:10.1057/s41274-017-0203-x

Poria, S., Cambria, E., and Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Syst.* 108, 42–49. doi:10.1016/j.knosys.2016.06.009

Qian, S., Zhang, T., Xu, C., and Shao, J. (2016). Multi-modal event topic model for social event analysis. *IEEE Trans. Multimed.* 18 (2), 233–246. doi:10.1109/tmm.2015.2510329

Ray, S., and Chandra, N. (2012). Domain based ontology and automated text categorization based on improved term frequency - inverse document frequency. *Int. J. Mod. Educ. Comput. Sci.* 4 (4), 28–35. doi:10.5815/ijmecs.2012.04.04

Ren, F. (2012). A cheap domain ontology construction method based on graph generation and conversion method. *J. Inf. Comput. Sci.* 9 (18), 5823–5830.

Tadrat, J., Boonjing, V., and Pattaraintakorn, P. (2012). A new similarity measure in formal concept analysis for case-based reasoning. *Expert Syst. Appl.* 39 (1), 967–972. doi:10.1016/j.eswa.2011.07.096

Tan, R., and Lin, B. (2019). Public perception of new energy vehicles: Evidence from willingness to pay for new energy bus fares in China. *Energy Policy* 130 (4), 347–354. doi:10.1016/j.enpol.2019.04.010

Tan, Z., Yang, Y., Wang, P., and Li, Y. (2021). Charging behavior analysis of new energy vehicles. *Sustainability* 13 (9), 4837. doi:10.3390/su13094837

Tang, L., Guo, C., Chen, J., and Sun, L. (2020). Learning concept hierarchies from Chinese academic literature for domain ontology construction. *J. China Soc. Sci. Tech. Inf.* 39 (4), 387–398.

Tavana, M., Shaabani, A., Santos-Arteaga, F. J., and Valaei, N. (2021). An integrated fuzzy sustainable supplier evaluation and selection framework for green supply chains in reverse logistics. *Environ. Sci. Pollut. Res.* 28, 53953–53982. doi:10.1007/s11356-021-14302-w

Teixeira, A. C., and Sodre, J. (2018). Impacts of replacement of engine powered vehicles by electric vehicles on energy consumption and CO2 emissions. *Transp. Res. Part D Transp. Environ.* 59, 375–384. doi:10.1016/j.trd.2018.01.004

Tirunillai, S., and Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Mark. Res.* 51 (4), 463–479. doi:10.1509/jmr.12.0106

Wang, R., Li, X., and Li, C. (2021). Optimal selection of sustainable battery supplier for battery swapping station based on Triangular fuzzy entropy -MULTIMOORA method. *J. Energy Storage* 34, 102013. doi:10.1016/j.est.2020.102013

Wang, R., Song, Y., Xu, H., Li, Y., and Liu, J. (2022). Life cycle assessment of energy consumption and CO2 emission from HEV, PHEV and BEV for China in the past, present and future. *Energies* 15 (18), 6853. doi:10.3390/en15186853

Wang, Y., Yu, Y., Huang, K., Chen, B., Deng, W., and Yao, Y. (2017). Quantifying the environmental impact of a Li-rich high-capacity cathode material in electric vehicles via life cycle assessment. *Environ. Sci. Pollut. Res.* 24 (2), 1251–1260. doi:10.1007/s11356-016-7849-9

Wen, J., Guo, W., and Shao, H. (2017). Case retrieve methodology based on domain ontology and case-based reasoning. *Comput. Integr. Manuf. Syst.* 23 (7), 1377–1385.

Wen, Y., Gao, C. M., and Pan, H. W. (2011). Research on concept semantic similarity computation based on ontology. Proceedings of the 2011 IEEE 2nd International Conference on Computing, Control and Industrial Engineering, 284–287.

Wu, Y., Xu, C., Huang, Y., and Li, X. (2020). Green supplier selection of electric vehicle charging based on Choquet integral and type-2 fuzzy uncertainty. *Soft Comput.* 24 (5), 3781–3795. doi:10.1007/s00500-019-04147-4

Xiao, G. (2020). Dalian, China, 451–454.Research on key manufacturing technologies of new energy vehicles based on artificial intelligenc, Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)

Xie, J., Li, J., Yang, F., and Liang, L. (2020). Decision-making and coordination optimized for multi-stage closed-loop supply chain of new energy vehicle. *J. Industrial Eng. Eng. Manag.* 34 (2), 185–198.

Xu, N., Finkelman, R. B., Dai, S., Xu, C., and Peng, M. (2021). Average linkage hierarchical clustering algorithm for determining the relationships between elements in coal. *ACS Omega* 6 (9), 6206–6217. doi:10.1021/acsomega.0c05758

Xu, Q., Jiao, R. J., Yang, X., Helander, M., Khalid, H. M., and Opperud, A. (2009). An analytical Kano model for customer need analysis. *Des. Stud.* 30 (1), 87–110. doi:10.1016/j.destud.2008.07.001

Yang, D., and Chai, H. (2018). Research on green product design selection based on QFD and case-based reasoning. *Sci. Technol. Manag. Res.* 38 (16), 251–259.

Yoo, D., and No, S. (2014). Ontology-based economics knowledge sharing system. *Expert Syst. Appl.* 41 (4), 1331–1341. doi:10.1016/j.eswa.2013.08.031

You, X., Lei, X., and Shi, Y. (2018). Criteria of green supplier selection for lightweight design in automotive industry. *Sci. Technol. Manag. Res.* 38 (16), 45–51.

Yu, R., Cong, L., Hui, Y., Zhao, D., and Yu, B. (2022). Life cycle CO2 emissions for the new energy vehicles in China drawing on the reshaped survival pattern. *Sci. Total Environ.* 826, 154102. doi:10.1016/j.scitotenv.2022.154102

Zhang, X., Hou, X., Chen, X., and Zhuang, T. (2013). Ontology-based semantic retrieval for engineering domain knowledge. *Neurocomputing* 116, 382–391. doi:10.1016/j.neucom.2011.12.057

Zhao, K., and Yu, X. (2011). A case based reasoning approach on supplier selection in petroleum enterprises. *Expert Syst. Appl.* 38 (6), 6839–6847. doi:10.1016/j.eswa.2010.12.055

Zheng, S., Huang, Q., Zhang, G., Li, Y., and Chen, X. (2019). A ontology construction method for user generated content. *Inf. Sci.* 37 (11), 43–47.

Zhou, F., Wang, X., Zhou, L., He, Y., Ni, L., and Yang, H. (2019). Multi objective optimization model of outsourcing supplier portfolio selection for automotive industry based on lifecycle quality of economics. *Comput. Integr. Manuf. Syst.* 25 (5), 1259–1271.