Expert Systems With Applications 193 (2022) 116472

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Problem-oriented CBR: Finding potential problems from lead user communities

Mintak Han [a,1], Youngjung Geum [b,*,2]

[a] Department of Industrial & Management Engineering, Myongji University, 116, Myongji-ro, Cheoin-gu, Yongin-si, Gyeonggi-do, Republic of Korea
[b] Department of Industrial Engineering, Seoul National University of Science and Technology (SeoulTech), 172 Gongreung 2-dong, Nowon-gu, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Technology-intensive product innovation
Lead user
CBR
SAO analysis
Sentiment analysis

ABSTRACT

Online communities, where lead users openly share their experiences and knowledge on product and technology in the form of a post, have become a fruitful source of innovation. While efforts have been made to provide a way of data-driven case-based reasoning (CBR), existing studies have limitations in reflecting lead users' characteristics into the CBR process. Current research has emphasized the retrieval and adaptation phase only, which retrieves, reuses, and revises cases. However, what is at the core of lead user characteristics is to find out the problem, and solve the problem by themselves before mass customers. This means CBR needs to focus on uncovering and defining problems and finding relevant solutions for designated problems. In response, this research suggests a novel approach for *problem-oriented* CBR approach to reflect lead user characteristics. First, this study defines problems, by extracting problem–solution sets related to the specific function using sentiment analysis. Second, this study improves case representation and case retrieval using subject-action-object (SAO) analysis and technology tree respectively. This study demonstrates the approach through a case of drone technology using lead user communities (diydrones.com), and our findings suggest that the approach can help firms broaden the knowledge of existing products to make an improvement.

1. Introduction

According to Chesbrough (2003), the innovation paradigm has been shifted from closed to open innovation. With the rise of open innovation, firms have to incorporate creative ideas from external sources. What has been dominant among many different sources of innovation is evidently the customer. This is especially true in a recent environment where technology changes so fast and customers become smart due to the explosively generated information from the web. They share information on product and technology and their experience in online communities such as websites, forums, and blogs (Franke & Piller, 2004; Lüthje, 2004; Marchi, Giachetti, & De Gennaro, 2011), thus have been increasing bargaining power based on well-developed knowledge. For this reason, firms to develop new products are trying to collaborate with these innovative users in the value co-creation processes (Marchi et al., 2011). Among many types of customers, what is the especially valuable group is lead users (Von Hippel, 1986).

According to several works, lead users can be defined based on the following two characteristics: (1) they generally face some strong needs before the mass customers encounter them (2) they solve their problems themselves (Marchi et al., 2011; Tuarob & Tucker, 2015; Von Hippel, 1986). Taken together, the main characteristics of lead users are simple: they are ahead of the market, anticipating the general needs of the mass market and seeking the solution in advance (Marchi et al., 2011). Due to this singularity, the lead user technique that incorporates lead users' opinions into the innovation process has been widely used in practice in order to help firms overcome the uncertainties of technological development (Roy, 2018).

However, in practice, the lead user technique is hard to implement on account of the time, effort, and cost for identifying and maintaining lead user groups. Therefore, recent studies try to focus on utilizing massive online data for understanding lead user groups (Aral & Walker, 2012; Marchi et al., 2011; Tuarob & Tucker, 2015; Vaughan, Seepersad, & Crawford, 2014) and identifying relevant solutions (Geum, Noh, &

* Corresponding author at: Department of Industrial & Information Systems Engineering, Seoul National University of Science and Technology, 172 Gongreung 2-dong, Nowon-gu, Seoul 139-746, Republic of Korea.
E-mail address: yjgeum@seoultech.ac.kr (Y. Geum).
[1] ORCID: 0000-0003-4124-8275.
[2] ORCID: 0000-0001-7346-2060.

https://doi.org/10.1016/j.eswa.2021.116472
Received 2 September 2021; Received in revised form 9 November 2021; Accepted 25 December 2021
Available online 5 January 2022

Park, 2016; Kim & Park, 2019; Tuarob & Tucker, 2015). A data-driven approach for replacing lead user techniques has been actively employed, by leveraging ideas from lead user communities (Kim & Park, 2019).

Especially, the use of case-based reasoning (CBR) fits the purpose of utilizing the data-driven lead user technique, because the core essence of lead users is directly related to the problem-solving processes of CBR. Case-based reasoning is a decision support technique to solve new problems based on existing problem-solving of similar cases (He, 2013; Liu & Ke, 2007; Mustapha, 2018; Richter & Weber, 2013). As many works noted (Kim & Park, 2019; Marchi et al., 2011; Von Hippel, 1986), lead users are innovative customer groups that face the needs and problems prior to the mass customers and find available solutions by themselves. This means both problems and solutions - which can be effectively utilized as important data for CBR - are expressed in the lead user communities. However, even if CBR is very helpful to utilize lead users' opinions in practice, how the contents of the lead user community can be applied to the CBR still remains a void in the literature. Even if thousands of posts are created from online lead user communities, it is difficult to monitor all posts on the web. This leads to the necessity of organized guidelines for utilizing such fruitful content generated by lead users (Füller, Jawecki, & Mühlbacher, 2007).

To address the necessities of models for utilizing online data to CBR, several approaches have been employed to provide a way of textual CBR (Geum et al., 2016; Goel & Diaz-Agudo, 2017; He & Wang, 2016; He, 2013; Kim & Park, 2019; Mustapha, 2018; Plaza, 2009; Shen, Yan, Fan, Wu, & Zhang, 2017; Sizov, Öztürk, & Aamodt, 2015; Weber, Aha, Sandhu, & Munoz-Avila, 2001; Zhou, Jiao, & Linsey, 2015; Zhu, Hu, Qi, Ma, & Peng, 2015). Ceausu and Despres (2007) suggested how to extract lexical patterns in CBR, such as Noun-Preposition-Noun or Verb-Preposition-Noun. Liu and Ke (2007) also employed text mining to CBR in order to extract the key concepts of situations and actions, and discover helpful knowledge from historical problem-solving logs. They employed similarity matching to the situation-action identification. Plaza (2009) tried to characterize a form of experience for textual CBR systems, which can be represented as (situation, outcome). Since how many different forms of experience are there is a critical question to the textual CBR, many different forms (such as How-To) should be considered and prepared (Plaza, 2009). Sizov et al.(2015) suggested a novel case retrieval method for textual CBR that employs evidence-driven retrieval using a text reasoning graph. Reuß, Stram, Juckenack, Althoff, Henkel, Fischer, and Henning (2016) also suggested FEATURE-TAK, which is a framework to employ natural language processes and some data mining techniques to the textual CBR.

Despite the contribution of previous works, most previous works focused on case retrieval and adoption, which deals with how properly retrieve relevant cases, and adapt the retrieved cases for the problem-solving process. This means most previous works neglected to identify proper problems, which is a great starting point for innovation. From a practical perspective, what is the most difficult task in practice is to find *uncovered problems*. This is especially true in a recent environment where the technology complexity has increased (Maine, Thomas, & Utterback, 2014). In addition, it is a challenging issue to fully understand cause and effect chains among the components from the technology-intensive product (Köhler & Som, 2014). This makes the problem statement, the most important first step of the CBR process, obscure. Therefore, the process of searching for existing or even, potential problems based on the technological structure of the product should be preceded, which can promote more concrete and proactive decision-making.

Especially, finding uncovered problems is worth the effort in lead user communities, because customers in lead user communities are actively communicating with one another by sharing their opinions including problems and their own solutions. The core essence of lead user communities – finding problems and solutions in advance – fits the purpose of CBR. This means the use of lead user communities can be the best remedy for addressing the limitation which neglects the identification of potential "problems" and finding relevant solutions.

In response, this study suggests a novel approach to the textual CBR, suggesting a problem-oriented CBR approach. The lead user communities are suitable for this process, since lead users generally find the problem by themselves, ahead of other normal customers, and fix the problem in their own way. From lead user communities, this study explores substantial problems from the database and suggests how solutions can be derived from the problem set. To identify potential problems, this study integrates SAO analysis and sentiment analysis and defines the problem statement. This study focuses on the fact that customers generally express their problem with a negative sense, or with specific phrases that require some modifications. It has also been proved that customers tend to make product-related attributions for negative reviews and non-product attributions for positive reviews (Qiu, Pang, & Lim, 2012; Sen & Lerman, 2007), which implies that negative reviews can provide innovation chances by revealing specific problems. After the problem statement, the technology tree and co-occurrence analysis are conducted for the case retrieval. This study is differentiated from previous works by not only preparing the procedure to consider the context of technology-intensive product innovation but also suggesting how to identify problems and provide relevant solutions from the CBR process, by leveraging the lead user communities systematically.

The remainder of this paper is organized as follows. The related works section deals with the background of online lead user communities and textual CBR in innovation research, which are the key theme of this paper. Then, the concept of problem-oriented CBR is described. The research framework addresses the proposed approach and its detailed procedures. The subsequent case study presents how the proposed approach can be applied. Finally, our discussion and concluding remarks are provided.

2. Related works in innovation research

2.1. Online lead user communities

As Chesbrough (2003) noted, innovation no longer happens as a closed form. Rather, it happens based on the active incorporation of external knowledge. Among many sources of innovation, customer knowledge has been considered as a prominent source of innovation. As noted in many previous studies, recent innovation paradigm has shifted from innovating *for* customers to innovating *with* customers (Desouza et al., 2008; Nambisan, 2002). This means customers are actively involved as a part of knowledge creation process (Desouza et al., 2008).

Actually, customers have long been considered to be a great source of innovation. Von Hippel (1986) introduced lead user approach that actively incorporate the customers' idea to the new product development process. Lead users are defined as customers who face some strong needs beyond what is currently available in the general market, and solve the problem by themselves (Marchi et al., 2011; Tuarob & Tucker, 2015; Von Hippel, 1986). Because they anticipate the general needs of the mass market and seek the solution in advance (Marchi et al., 2011), they are generally creative, helpful for firms that aims to develop new products.

According to Eisenberg (2011), there are some differences between lead user approach and traditional approach. Lead user approach focuses on the needs of leading-edge users, and tries to seek not only customer needs, but also innovation chances or solutions from those leading-edge users (Eisenberg, 2011). In addition, lead users generally find solution in adjacent, nonobvious, and analogous markets, not only focusing on the target market (Eisenberg, 2011).

Due to the distinct characteristics of lead users, there have been substantial number of previous works to deal with lead user approach. Especially, recent studies have emphasized the role of online communities (Marchi et al., 2011; Sawhney & Prandelli, 2000). Online communities generally create various themes and ideas, and attract innovative consumers (Sawhney & Prandelli, 2000). Customers freely communicate one another for a specific topic, actively exchanging their ideas. From the online communities, studies to understand lead user groups (Aral & Walker, 2012; Marchi et al., 2011; Tuarob & Tucker, 2015; Vaughan et al., 2014) or identify relevant solutions (Geum et al., 2016; Kim & Park, 2019; Tuarob & Tucker, 2015) have been extensively suggested. Especially, how to leverage the ideas from lead user communities has also been suggested (Kim & Park, 2019). The lead user communities are defined as “distributed groups of individuals focused on solving a general problem and/or developing a new solution supported by computer-mediated communication” (Dahlander & Wallin, 2006, p. 1246).

In particular, the rise of ‘Web 2.0′ has enabled new ways to apply the lead user approach: innovation by members of online communities (Dahlander, Frederiksen, & Rullani, 2008). The users in these communities tend to openly share their information, knowledge, and even technologies with other members, exchanging their experiences and developing innovative ideas (Kim & Park, 2019; Kozinets, Hemetsberger, & Schau, 2008). Open source software is probably the most well-known example, where geographically dispersed individuals collectively develop new software and product innovation (Von Krogh, Spaeth, & Lakhani, 2003).

The contribution of these online lead user communities can be further enhanced in the process of developing technology-intensive products. Particularly, in the case of products such as VR (virtual reality) and drone, there can be a wide variety of problems that lead users encounter. In terms of technology diffusion, these problems prevent the spread to the early majority of a market, so that it is buried in a “chasm” for a long time (Moore, 1991; Rogers, 2010). As discussed in many studies, a scientific approach to solve the discovered problems and identify potential needs preemptively has been regarded as a key strategy to overcome the chasm (Clarysse, Wright, Bruneel, & Mahajan, 2014; Denning, 2001; Gombault, Allal-Chérif, & Décamps, 2016; Malhotra, 2001; Maurer & Melnik, 2007; Moore, 1991). Thus, the distinctive characteristics of the lead user and online lead user community are appropriate for this strategic context. That is, a shared post that represents a specific problem or a problem-solving case can serve as a starting point for developing improvements to existing products or entirely new products. Despite these high utilization of the online lead user communities in the innovation process of technology-intensive products, there is a lack of research on the application of these textual contents. It leads to the necessity of an organized approach to reflect the problem-solving context, which is one of the main characteristics of posts in lead user communities.

2.2. Textual CBR

CBR is a problem-solving approach that relies on similar prior cases to find a solution (Kolodner, 2014; Richter & Weber, 2013). A case is basically an experience of a solved problem, and a case base is a collection of such cases (Richter & Weber, 2013, p. 17). Based on those case bases, proper approximate reasoning is conducted to solve a specific problem.

The conventional CBR process is represented in Fig. 1. Research on CBR in product innovation has emphasized the adaptation phase, which reuses and revises retrieved cases (Verhaegen, D’hondt, Vandevenne, Dewulf, & Duflou, 2011). CBR has been widely used for various kinds of decision-making in a wide range of applications (Chen et al., 2021; Geetha, Narayanamoorthy, Manirathinam, & Kang, 2021; Jin, Cho, Hyun, & Son, 2012; Okudan, Budayan, & Dikmen, 2021; Torrent-Fontbona, Massana, & López, 2019). Jin et al. (2012) employed multiple regression analysis to the CBR model to improve the cost prediction. Bach (2013) developed a SEASALT (Sharing Experience using an Agent-based System Architecture LayouT) architecture for extracting experiences from web sources and applied those to the CBR system. Geum et al. (2016) employed CBR to the generation of new service ideas, by reflecting the heterogeneity of service characteristics. Torrent-Fontbona et al. (2019) suggested a recommender system using a case-based maintenance system that combines case-based redundancy reduction and attributes weight learning. Bentaiba-Lagrid, Bouzar-Benlabiod, Rubin, Bouabana-Tebibel, and Hanini (2020) suggested a CBR system for supervised classification in the medical fields, presenting a new amplification technique. Okudan et al. (2021) employed CBR to the knowledge-based risk management system for construction projects. Geetha et al. (2021) also employed a fuzzy CBR approach for finding COVID-19 patents priorities by considering all relevant factors.

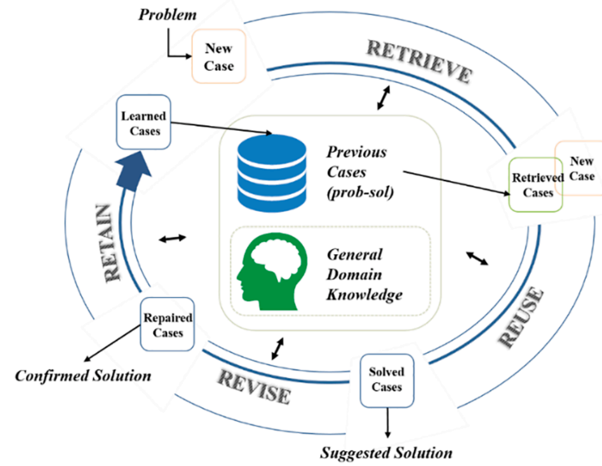

Fig. 1. Conventional CBR process ().
adopted from Althoff, 2001

With the trends of big data analytics, data-driven CBR becomes popular, thus efficient case collection, effective case representation, and accurate case retrieval are becoming increasingly important steps in a CBR system (Kim & Park, 2019; Wu, Lo, & Hsu, 2006). Especially, textual CBR is a subfield of CBR concerned with research and implementation on case-based reasoners where some or all of the knowledge sources are available in textual format. This is because a vast amount of experiences are now recorded on the web, as a form of text or document, and we try to use the web to find relevant information to take reasonable decisions (Plaza, 2009). It aims to use these textual knowledge sources in an automated or semi-automated way for supporting problem-solving through case comparison (Weber, Ashley, & Brüninghaus, 2005).

According to Goel and Diaz-Agudo (2017), textual CBR is one of the major improvements in CBR research. Weber et al. (2001) introduced a framework to manipulate knowledge artifacts, using both elicitation and extraction tool. They employed an adaptation of template mining and identified to interactively elicit knowledge from users. In particular, they identified “where to search, what to search, and, when requirements are met, what to extract” (Weber et al., 2001, p.4). Ceausu and Despres (2007) suggested a two-step approach to discover semantic traces in textual CBR. They developed a pattern recognition algorithm that detects several lexical patterns such as Noun-Preposition-Noun or Verb-Preposition-Noun, defining nominal lexical patterns and verbal lexical patterns. Plaza (2009) also tried to characterize a form of experience for textual CBR systems, which can be represented as (situation, outcome). Since how many different forms of experience are there is a critical question to the textual CBR, many different forms (such as How-To) should be considered and prepared (Plaza, 2009).

Recent studies have employed more advanced techniques for the textual CBR. Sizov et al.(2015) suggested a novel case retrieval method for textual CBR, which more focuses on retrieving adaptable cases. They called it evidence-driven retrieval (EDR). The text reasoning graph (TRG) is used for the representation in this study, addressing the gap between retrieval and reuse. Reuß et al. (2016) suggested a framework for textual CBR, which consists of the data layer, agent layer, CBR layer, NLP layer, and interface layer. They extracted phrases and keywords from the unstructured data and extended those keywords using synonyms and hypernyms. After extending, various techniques have been used such as association rule mining, and clustering.

Over the years, there has been significant progress addressing the threshold challenge facing textual CBR: how to bring textual knowledge sources to bear in supporting reasoning with cases. Specifically, the research has addressed the following questions: (1) how to map from texts to structured case representations; (2) how to assess the similarity between textually represented cases.

Most of the studies applying textual CBR (Geum et al., 2016; Kim & Park, 2019; Plaza, 2009; Reuß et al., 2016; Weber et al., 2001) use natural language processing techniques like text mining to make unstructured data coherent. Case representation and case retrieval, i.e., input and output of CBR, are based on the keyword vector. Thus, the retrieval method was forced to adopt a simple approach of measuring a similarity score between keyword vectors. However, this has the following critical limitations. First, loss of information that does not preserve the full meaning of the document or sentence occurs. Even if it consists of the same keywords, there are various situations depending on the context. Moreover, if a similar case is retrieved, the keywords themselves cannot be a genuine solution. A document (case) generally consists of sentences describing either a problem or a solution. Therefore, it is required to distinguish individual sentences within a document into problems or solutions. To do this, a method should be devised to represent sentences within a range that minimizes the loss of information. Second, in the case of technology-intensive products, problem-solving is usually a knowledge-dependent task because their components are highly complex, and thus knowledge support is inevitable to identify adequate solutions. For example, changing the specification of a particular component to solve a problem may entail changes to another associated component. Thus, it is essential that the relationships among the components should be ascertained proactively. In other words, experts' knowledge must be incorporated into the process of textual CBR, i.e. case retrieval, to identify the exact problems and solutions.

2.3. Sentiment analysis

Sentiment analysis is defined as the task of finding the opinions of authors in a specific document, by distinguishing the positive and negative attitudes towards some topics (Feldman, 2013; Liu, 2012). It has been widely used for many different applications, such as opinion question answering, opinion summarization, opinion retrieval (Montoyo, Martínez-Barco, & Balahur, 2012). Especially, sentiment analysis is actively applied for analyzing social media or customer preferences (Fiok, Karwowski, Gutierrez, & Wilamowski, 2021; Wan, Xu, Zhuang, & Pan, 2021).

Sentiment analysis can be conducted at many different levels: document-level, sentence-level, or aspect-level (Berka, 2020; Feldman, 2013). The document-level analysis is the basic form of sentiment analysis, which classifies the document into several groups, such as positive or negative. Considering that the document is composed of different opinion and topics, sentence-level analysis can be conducted, which make each sentiment be analyzed from the perspective of positive or negative. More extended from the document- or sentence-level analysis, aspect-level sentiment analysis defines some aspects for products or services, and applied sentiment analysis to the aspects, which are generally well-defined characteristics of products or services (Berka, 2020).

Albeit infrequent, sentiment analysis and case-based reasoning have been integrated in previous studies. Xin et al. (2012) suggested a quick emergency response model (QERM) for opinion mining, integrating sentiment analysis and case-based reasoning. Ohana, Delany, and Tierney (2012) suggested a case-based approach to sentiment classification of customer reviews, storing sentiment lexicons for later retrieval and reuse on similar documents. Dong, Schaal, O'Mahony, McCarthy, and Smyth (2013) suggested a novel way for a case-based product recommendation, by analyzing the similarity and sentiment of customer reviews and setting the basis for the further recommendation. Zhou et al. (2015) proposed a two-layer model. The first layer focused on SVM models for predicting sentiments, whereas the second layer implemented the use case analogical reasoning to identify implicit characteristics of customer needs. Berka (2020) focused on sentiment analysis and discussed its application to complementary approaches – rule-based reasoning and case-based reasoning.

Even if sentiment analysis has been applied to the case-based reasoning for many different purposes, most of the existing studies have applied sentiment analysis and CBR independently or in parallel within the framework. Unlike previous works, this study employs sentiment analysis to find the uncovered problems in CBR processes, focusing on the problem statement task which has been neglected.

3. Proposed approach: A problem-oriented CBR

This paper proposes a problem-oriented CBR approach to extend lead users' knowledge for supporting the development of technology-intensive products. In our approach, the nature of the lead user community and the context of technology-intensive product innovation are considered. Unlike the conventional CBR, the goal of our process is to identify not only the solution for the existing problems but also potential problems and issues that should be resolved.

For these reasons, the following Table 1 shows the differences between the current textual CBRs and our problem-oriented CBR. First of all, our approach focuses on not only finding solutions but also identifying problems from the lead user communities, considering the characteristics of lead users that find problems in advance of others and solve themselves.

On the procedural side, it starts with a case collection, as shown in Fig. 2. Due to the nature of the lead user community, a post has several descriptions of problems and solutions. Therefore, it is impossible to map a single post to a single problem or solution. If a post is used as a case itself, it cannot be a logical process to compare with the problem statement. Because unnecessary terms are included in the comparison with the problem statement when a post is expressed as a keyword vector. To address this issue, SAO structures in a post that minimize loss of information are classified into problem (P-SAO) or solution (S-SAO) based on semantic and sentiment analysis. For next, a main functional keyword at the heart of the problem is selected as an input for constructing mediating keywords in the case retrieval stage. These keywords are derived by complementary use of the quantitative approach (i.e. co-occurrence analysis) and the qualitative approach (i.e. technology tree). Finally, all the retrieved problems and solutions including mediating keywords are evaluated through the following two steps: identification of the critical functions and features associated with the main function, followed by a judgment of whether the solution is feasible with the experts. Detailed descriptions of these processes are discussed in the next section.

3.1. Step1: Case collection

The main objective of this step is preparing cases of 'knowledge: problems and solutions.' The most important thing to do is to select an appropriate lead user community. This is because data-driven approaches depend heavily on the quality of input data. Even more simply, bad inputs will lead to bad outputs, which is so-called 'Garbage in, Garbage out (GIGO)'. When applied to this study, a sufficient amount of data with high quality can improve the output quality for both aspects: explicit problem (i.e. same experience of most people) and latent problem (i.e. novel experience of only a few people). Therefore, there should be sufficient cases dealing with various issues about the target technology-intensive product in a lead user community.

Since the cases are collected from websites, web crawlers can be developed and utilized for automatic collection. However, they consist

Table 1
Comparison of previous CBR and our problem-oriented CBR approach.

Process stage	Current CBR	Our problem-oriented CBR
Main purpose of CBR	Finding solutions	*Identifying Problems*Finding relevant solutions for problems
Data source	Service or technology database (Geum et al., 2016; Kim & Park, 2019; Xiao, Skitmore, & Hu, 2017)	Lead user community
Case representation	A keyword vector with a value of tf-idf or frequency	A Subject-Action-Object (SAO) structure with a sentiment score
Case retrieval	Calculating similarity score	1) Constructing a mediating keywords (extended input) through hybrid approach of technology tree and co-occurrence analysis and2) Deriving all the SAO structures containing the mediating keywords

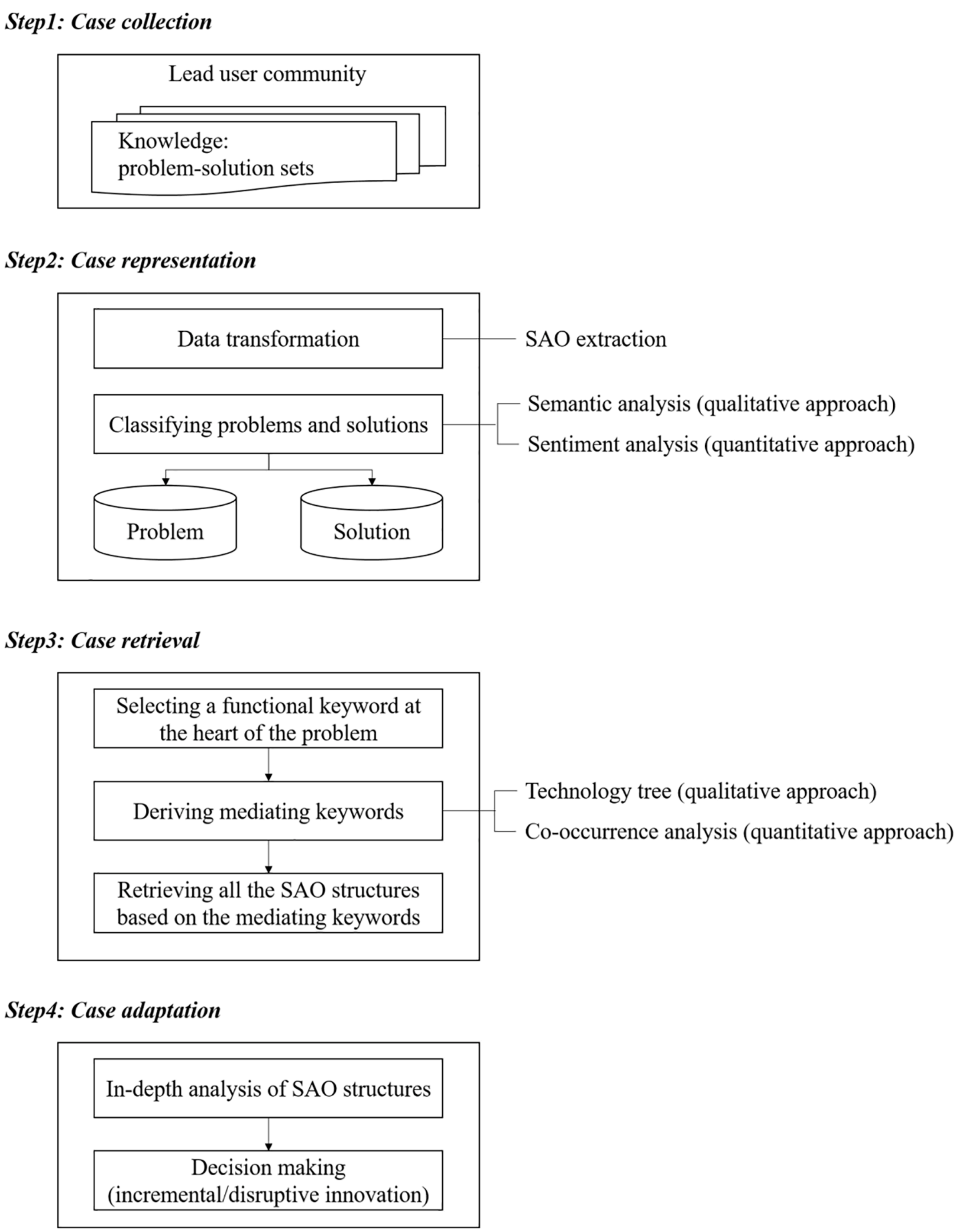

Fig. 2. Overall process of our problem-oriented CBR approach.

of unstructured text containing superfluous information such as email address, website address, and etc. to be eliminated. Thus it is inevitable to conduct pre-processing such as stemming, stop-word removal (e.g. a, the, of, etc.), and number removal. As a result, each post (case) is collected in the form of a pre-processed document.

3.2. Step2: Case representation: Describing problems and solutions

The case representation stage is a transformation of collected cases to suit the textual CBR format. This paper focuses on SAO structures in documents, which are the syntactically ordered structure of subject (noun phrase), action (verb phrase), and object (noun phrase). While SAO is not a deep representation of the sentence, it is more expressive than a keyword vector, offering potential improvements in the performance of interpreting meaning. For this reason, SAO analysis has been used primarily in the handling of technological documents (Guo, Wang, Li, & Zhu, 2016; Yoon, Park, & Kim, 2013). This structure clearly provides a relationship between components that appear in a technological text: subjects and objects may refer to components of a system, and actions may refer to functions performed by a certain component (Cascini, Fantechi, & Spinicci, 2004; Choi, Park, Kang, Lee, & Kim, 2012; Choi, Kang, Lim, & Kim, 2012; Yoon & Kim, 2011). That is, SAO structures are fundamentally related to the concept of function, which is defined as "the action changing a feature of an object" (Savransky, 2000). To conduct this, a modified version of the Java-based program called 'ReVerb' (Fader, Soderland, & Etzioni, 2011) which shows remarkable performance in the OpenIE field, is utilized. Its accuracy level has been proven enough in the recent relation extraction area such as Knowledge Graph (Muhammad, Kearney, Gamble, Coenen, & Williamson, 2020) and Question Answering (Wu, Zhu, Zhang, Kang, & Liu, 2021). It takes raw text as input, and outputs (argument1 = subject, relation phrase = action, argument2 = object) triple with a confidence score, which is the probability of whether the extraction is correct. For example, given the sentence "Lead user community is a potential source of innovation," ReVerb extracts the triple (lead user community, is a potential source of, innovation) with a 0.98 confidence score. In this study, only SAO structures with a high confidence score were filtered.

After extracting the SAO structure, what is important is to distinguish problems and solutions. An SAO structure of patent documents can be organized in a problem–solution format if the action-object (AO) forms the problem and the subject (S) states the solution. However, posts of the lead user communities have a different nature from the patent document. While the patent documents focus on what the technology provides to solve the problem, in general, the lead user's experience-based posts focus on the description of the problem situation encountered, and suggest the process of their own conceptual solution. Thus, this study utilizes two distinct types of SAO structures, depending on whether they represent a problem situation or a solution, instead of using an SAO itself as a problem–solution format. Especially, the semantic analysis with sentiment score fits this approach due to the following reasons: from the semantic relationship of the SAO elements and emotional expressions including positive or negative words, it can be inferred whether it represents a problem situation or a solution. Thus, a system is developed in this study to automatically classify a vast amount of SAO structures as Problem SAO (P-SAO) and Solution SAO (S-SAO) in the case representation stage.

Table 2 shows distinctive characteristics of SAO structures describing a problem or solution. In general, nouns and adjectives are included in "subject" and "object", on the other hand, verbs and adverbs are mixed within "action" (Benamara & Moriceau, 2006). This structure educates us on the features like 'who (subject) does what action (verb) to whom (object)' (Weber et al., 2005) and hence, the opinion or action of the subject on the object can be identified. Especially, the verb with adverbs describes the feeling the subject has on the object (Chaitanya, Harika, & Prabadevi, 2017). Moreover, through several posts randomly selected from lead user communities, representative terms describing problem situations and solutions, are derived respectively. A list of synonyms is extracted based on WordNet, which has been widely used in text mining.

Based on these characteristics, the semantic analysis, which focuses on the verbs contained in the action, is utilized as the primary filter, then the sentiment analysis for adjectives and adverbs is used as the secondary filter. The SAO structures are classified into problems and solutions depending on whether or not synonyms are included in the action phrase. In the case of not including synonyms, the calculated sentiment score is utilized as a classification index. The analyst should set the cut-off value based on the statistics for the sentiment score of the SAO structures and then perform the classification process. Note that, the important thing in this step is not to improve the accuracy of the classification. The eventual objective is to eliminate unnecessary sentences in a document and increase the probability of leaving useful sentences to be analyzed.

3.3. Step3: Case retrieval

After constructing the two databases, a case retrieval process is applied. For case retrieval, this study uses the concept of the technology tree. A technology tree is a branching diagram that expresses relationships among product components, technologies, or functions in a specific technology area (Choi et al., 2012; Choi et al., 2012).

Most technology tree diagrams are created by reflecting the opinions of domain experts in a qualitative manner. However, in the case of technology-intensive products, constructing the technology tree has become more difficult as technological development speed is very fast and the system components are becoming diverse and complex. Therefore, the co-occurrence analysis is used as a complementary tool. Two keywords with frequent co-occurrence are likely to have a deep relationship in a specific context. In the context of the posts describing the experience that users encounter in the middle of using technology-intensive products, it can be interpreted as follows. Frequently emerging keywords related to the main functional keyword (i.e. problem) are likely to be technically related: (1) They are core elements to improve the function or (2) other operational issues that can be caused during the problem-solving process. Through the hybrid approach of technology tree and co-occurrence analysis, this study defines the 'mediating keywords' set. It is utilized as an extended input to the case retrieval process to broaden the scope of the case search and consequently enrich the problems and solutions.

Based on this consideration, first of all, a main functional keyword

Table 2
Distinctive characteristics of SAO structures describing problems and solutions.

	Element	Problem description	Solution description
Semantic perspective	verbs in 'action'	Pattern of sentence structure Subject – Synonyms – Function/Feature/SpecificationSynonyms: need, require, seek, necessitate, want, lack, shortage, be inconvenient, cause, be insufficient, unsatisfied	Subject – Synonyms – Solution/FunctionSynonyms: can, could, enable, use, utilize, apply, devise, accelerate, adapt, address, employ, contribute, develop, design, invent, inspire, accomplish, enhance, overcome, solve, help, improve, propose, suggest, create, provide, support
Emotional perspective	adjectives and adverbs in 'SAO'	Negative adjectives and adverbs	Positive adjectives and adverbs

describing the problem situation most precisely is determined as an input. This paper defines the main functional keyword as 'Root function'. Then, a technology tree is created as shown in Fig. 3. Considering the functional relationship, the key functions affecting the 'Root function' are placed in the lower layer, and related features are listed sequentially in the substructure of the function layer. As a result, the sub-features in the bottom layer of the technology tree are selected as an experts-based mediating keywords list.

In co-occurrence analysis, 'Dice coefficient' and 'Log-Likelihood' are the widely used statistical indices to judge the existence of the relationship. Based on these indices, keywords with low co-occurrence frequency are filtered and eliminated. Note that in extracting and determining a data-driven keywords list, only the keywords that are not identified in the technology tree are selected. Through this, mediating keywords are constructed by hybrid use of the experts-based approach and data-driven approach. Next, all the SAO structures containing at least one of the mediating keywords are retrieved from two databases. What is important at this step is to assign which function and feature each SAO structure corresponds to, based on the technology tree, as demonstrated in Table 3. This task enhances the convenience of exploring and interpreting the exact solution corresponding to a specific problem.

3.4. Step4: Case adaptation

Generally, the adaptation is to compensate for the differences between an old situation and a new one, trying to adapt an old solution to a new problem (Avramenko & Kraslawski, 2008). However, the ultimate goal of this study is not to find the most similar case to the problem statement. It is to identify the various issues to be considered in the problem-solving process and to acquire the knowledge of lead users' experiences to address the potential issues proactively. It means that a guideline is required to determine critical issues from the derived P-SAOs and verify the ideas from S-SAOs proposed to resolve the issues.

Hence, an in-depth analysis of the retrieved SAO structures is conducted for supporting experts' judgment. To improve a "root function", there may be a solution to various features, however, a solution to the specific feature may affect another function or related features. The more a feature that is associated with other features, the more likely it is at the center of the problem-solving process. Thus, understanding these relationships can help to determine which functions and features should be solved preferentially. On the other hand, the applicability of solutions is judged by the experts based on their domain knowledge, economical efficiency, and the firm's current technical skills. The suggested guidelines are summarized in Table 4.

Table 3
An example of retrieved results with assigned function and feature.

Doc #	Problems	Relatedfunction-feature	Solutions	Relatedfunction-feature
1	(P-SAO) 1–1	FU1 – FE1	(S-SAO) 1–1	FU3 – FE2
	(P-SAO) 1–2	FU2 – FE3	(S-SAO) 1–2	FU2 – FE1
			(S-SAO) 1–3	FU4 – FE2

Table 4
Guidelines for in-depth analysis of the retrieved SAO structures.

Guidelines	Description
Function level	How often does this function cause a problem?The ratio of the number of P-SAOs assigned to a particular function to the total number of P-SAOs
Feature related keywords level	Could there be a causal relationship between features?Co-occurrence frequency of the feature within the same functionCo-occurrence frequency of the feature between different functions
Applicability	Checklist:Is it a solution that provides functionality meeting the concept of product?Have we tried the similar solutions? (Is it a completely new approach?)Could we gain a competitive advantage over other products through utilizing the solution?Is it reasonable in terms of product development cost?Is it accessible with our internal technological capability?

4. Case study: Commercial drone

4.1. Background

In order to illustrate the applicability of the proposed approach, a case study of a technology-intensive product-UAV, or so-called drone was conducted. A drone is a perfect example since it contains numerous technologies such as sensors, networks, batteries, software modules, and so on. Small changes in such technologies can have a direct impact on the performance of drones. Moreover, the drone itself holds a high level of both uncertainty and opportunity as it is still in the early stage of development (Kwon, Kim, & Park, 2017). Therefore, not only identifying various technological issues and potential needs but also searching for a suitable solution might lead to success in the market. In this section, the lead user community associated with drone is investigated, and the proposed approach is conducted in a consecutive manner.

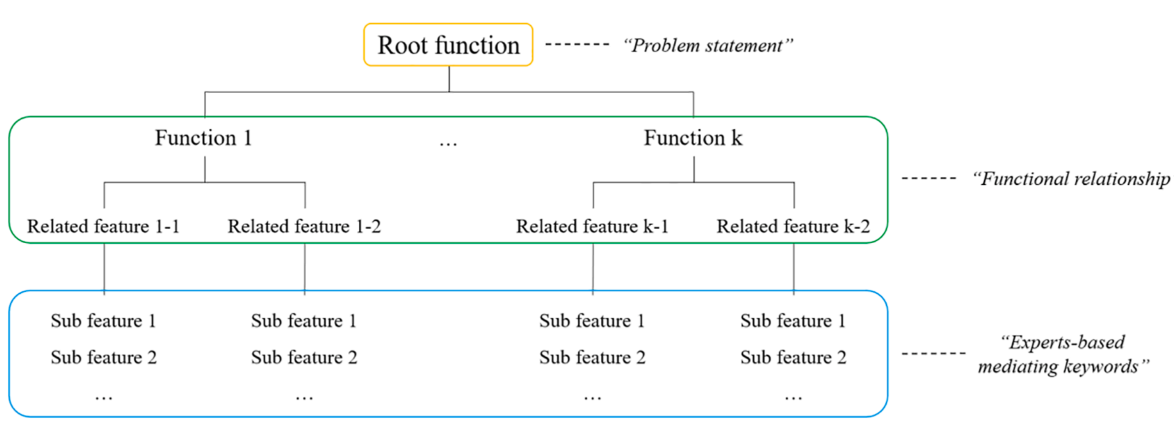

Fig. 3. A technology tree adopted in this study.

4.2. Process and results

4.2.1. Case collection

This study chose ‘diydrones.com’ among communities related to drone, based on the following criteria: sufficiency, diversity, and suitability. It is the leading community for personal UAVs and Robotics with over 40,000 members, where they can freely communicate their opinions. There are more than 10,000 posts that cover various topics such as aerial photography, aircraft platform, design, and software. However, it is meaningless without a text that reflects the characteristics of lead users. A typical article is shown in the following Fig. 4. Each article contains mainly its own problem with appropriate solutions and other useful information such as title, author, and postdate. Thus, the community is suitable for performing the revised textual CBR suggested in this study.

Because the number of articles is too large to be collected manually, a Java-based web-crawler was developed to download all the blog posts automatically. As a result, a total of 10,024 articles were collected with the reference period from April 2011 to September 2018. Before moving to the next step, natural language toolkits (NLTK), which is widely used in performing the NLP, is used. This means that pre-processing was conducted to remove the stop words such as articles, prepositions, and conjunctions. Moreover, the collected documents were parsed to separate out the parts into title, contents, author, and postdate.

4.2.2. Case representation

Pre-processed documents are transformed into SAO structures. The total number of derived SAO structures is 385,872. According to the above-mentioned filtering process, semantic analysis and sentiment analysis are conducted to assign a certain SAO structure to a problem or a solution. Fig. 5 illustrates an example that classifies SAO structures extracted from a certain document as problem, solution, and neutral. First, the existence of synonyms is confirmed by extracting the keyword from the ‘Action’ phrase. For example, the SAO structure, “I – was inspired by – the recent solarpowered launch system”, can be regarded as a concept of the solution. Because as the first filter, the verb, “inspire” is included in the ‘Action’ phrase. Second, a sentiment score is calculated focusing on adverbs and adjectives for SAO structures that cannot be classified through the first filter. “a gust of wind – easily push – the copter offcourse” is the SAO structure representing a problem situation since it records a negative sentiment score of ‘-0.44’ less than the cut off value determined by analysts. Consequentially, 17,093P-SAO structures and 59,700 S-SAO structures, labeled with document and structure index, formed each database.

4.2.3. Case retrieval

A major issue in drone control is the difficulty of landing the vehicle on a platform (Hérissé, Hamel, Mahony, & Russotto, 2010). Moreover, autonomous landing is one of the must-have operations of drones (Barták, Hraško, & Obdržálek, 2014). Autonomous landing, which guarantees simplicity, accuracy, and stability, requires a variety of requirements. However, autonomous landing is an open problem not yet definitively solved in every scenario (Cocchioni, Mancini, & Longhi, 2014). Thus, landing is selected as a “root function” of a problem statement.

The mediating keywords are derived through two steps to identify the overall technological and even non-technological features associated with landing, as described in Section 4. Firstly, a technology tree related to the root function (i.e. landing) is written by a group of experts. The group comprised five experts having experience in drone development from two years to five years. As shown in Fig. 6, landing is implemented by functions of mobility, sensing and perception, control systems, supporting devices, and others. As a result, keywords are assigned to 8 related features: power unit, driving device, aerial sensor, perception algorithm, auto control, manned control, supporting device, and etc.

Secondly, an implementation in R with “tm” and “openNLP”

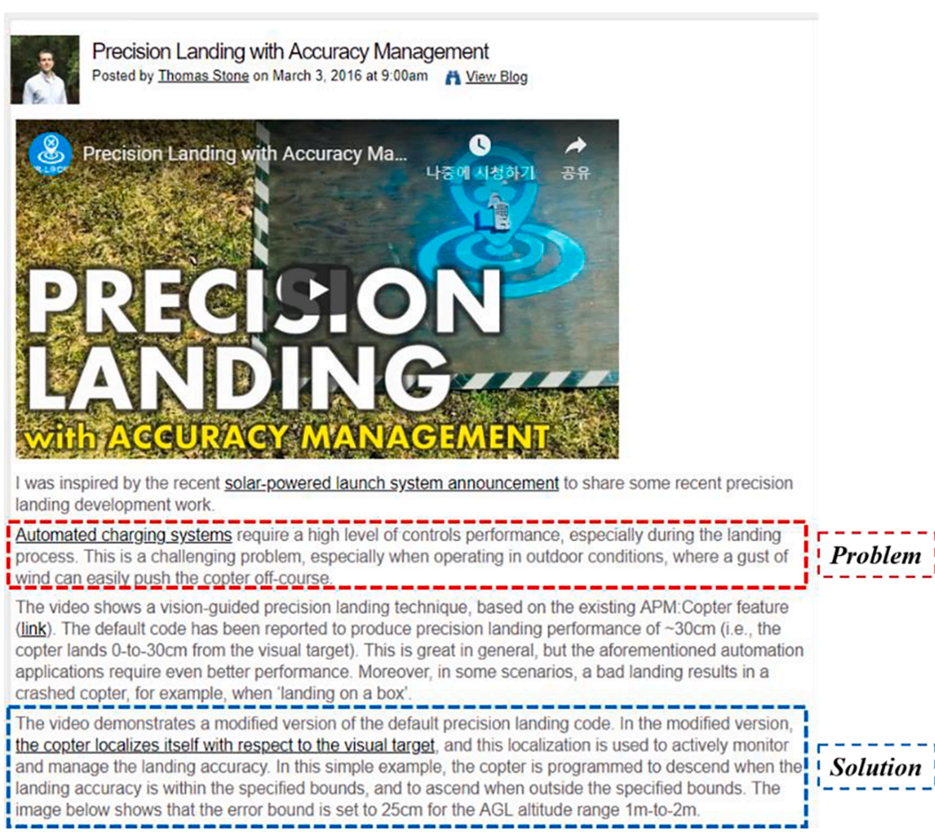

Fig. 4. The post of lead user’s experience at ‘diydrones.com’

Document No.	Structure No.	Subject - Action - Object	Syntactic (1st filter)	Emotional (2nd filter)	P/S
1522	1	I - was inspired by - the recent solarpowered launchsystem	inspire		S
1522	2	Automated charging systems - require - a high level of controls performance	require		P
1522	3	This - is - a challenging problem	-	-0.2	-
1522	4	a gust of wind - easily push - the copter offcourse	-	-0.44	P
1522	5	The video - shows - a visionguided precision landing technique	-	0.12	-
1522	6	The default code - has been reported to produce - precision landing performance of 30cm	-	0.16	-
1522	7	the aforementioned automation applications - require even - better performance	require		P
1522	8	The video - demonstrates a modified version of - the default precision landing code	demonstrates		S
1522	9	the copter - localizes - itself with respect to the visual target	-	0.08	-
1522	10	this localization - manage - the landing accuracy	-	0.33	S
1522	11	the landing accuracy - is within - the specified bounds	-	0.24	-
1522	12	This approach - can be extended and customized in - a variety of ways	can		S
1522	13	the default precision landing feature - is enabled by - default	enable		S

Fig. 5. An illustrative result of the case representation.

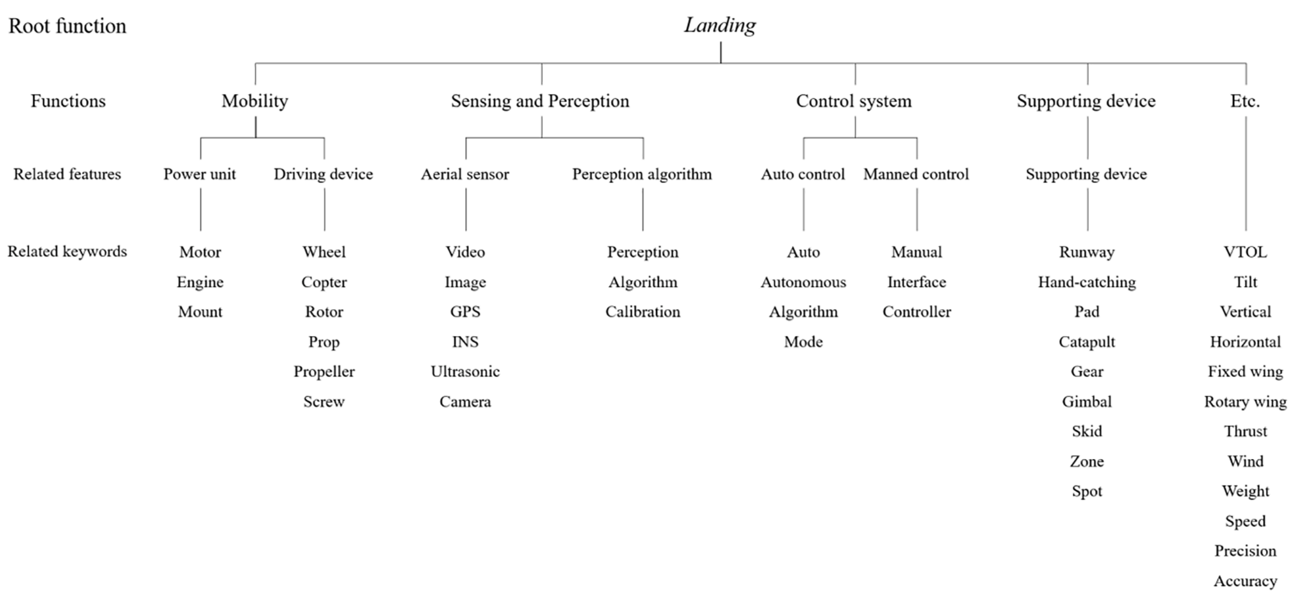

Fig. 6. A technology tree related to the landing function.

packages for Dice coefficient and Log-Likelihood yields the statistical extraction of co-occurrence terms. The result is sorted so that the most significant co-occurrences are the first ranks of the list. Due to the page limit, only the top 20 terms are entered in Table 5.

Table 5
Top 20 terms extracted from co-occurrence analysis.

Landing	Dice-Terms	Dice	LL-Terms	LL
1	Gear	0.2771	Gear	3151.50
2	Takeoff	0.0979	Takeoff	754.14
3	Precision	0.0601	Precision	349.86
4	Auto	0.0516	Auto	264.73
5	Autonomous	0.0419	Retractable	207.90
6	Ground	0.0381	Skids	207.02
7	Land	0.0367	Legs	164.04
8	Legs	0.0337	Ground	157.53
9	Vertical	0.0331	Parachute	152.65
10	Parachute	0.0317	Land	144.84
11	Site	0.0285	Vertical	143.47
12	Arms	0.0276	Pad	133.21
13	Altitude	0.0263	Arms	89.94
14	Camera	0.0257	Emergency	86.38
15	Motor	0.0237	Grass	85.79
16	Skids	0.0230	Height	82.21
17	Height	0.0225	Strip	76.03
18	Zone	0.0224	Frame	73.81
19	Crash	0.0217	Altitude	70.22
20	Control	0.0213	Camera	65.84

Finally, the mediating keywords list that integrates the keywords obtained from the textual data of lead users and the technology tree created by experts is shown in Table 6. A total of 64 keywords were determined excluding duplicate keywords.

In the next step, all SAO structures containing at least one of the mediating keywords are derived. A certain context can be inferred through a set of problems and solutions. As a result, from 1,488 documents including ‘landing’, 3,829P-SAO structures and 13,122 S-SAO structures are assigned to five functions and eight features. A partial result of problems assigned to the function of sensing and perception

Table 6
Mediating keywords for case retrieval.

Mediating keywords	Experts-based keywords	Motor, Engine, Mount, Wheel, Copter, Rotor, Prop, Propeller, Screw, Video, Image, GPS, INS, Ultrasonic, Camera, Perception, Algorithm, Calibration, Auto, Autonomous, Mode, Manual, Interface, Controller, Runaway, Hand-catching, Pad, Catapult, Gear, Gimbal, Skid, Zone, Spot, VTOL, Tilt, Vertical, Horizontal, Fixed wing, Rotary wing, Thrust, Wind, Weight, Speed, Precision, Accuracy
	Data-driven keywords	Gear, Ground, Legs, Parachute, Arms, Height, Crash, Emergency, Control, Waypoint, Hover, Climb, Sequence, Throttle, Transition, Switch, Flap, Position, Frame, Site

with corresponding solutions is described in Table 7. A certain context within a document can be inferred through a set of problems and solutions. With the experts' knowledge, this process determines whether the document contains ideas related to radical innovation or incremental innovation. Radical innovations are generally "new to the world or exceptionally different from existing products and processes" (Schilling, 2017, p,47). However, the incremental innovation "might have been previously known to the firm or industry, and involve only a minor change from existing practices" (Schilling, 2017, p,47). The two extremes of innovations are thus defined based on the level of the newness, differentness, and sometimes risk.

4.2.4. Case adaptation

Having retrieved problems and solutions with assigned function and feature, an in-depth analysis was conducted. The approximate importance is calculated at the function and feature level according to the guidelines suggested in Section 4 as follows. The number of P-SAO structures assigned to the five functions is shown in Table 8. Landing is basically conducted based on landing point acquisition and obstacle detection. Thus, it can be interpreted that the importance of the function, 'sensing and perception' in software aspect and the importance of the function, 'supporting device' in hardware aspect are emphasized.

In more detail, Table 9 shows the results of an in-depth analysis of feature-related keywords to determine which features have a crucial role in the 'sensing and perception' function.

Table 8
In-depth analysis of 'Landing' (function level).

Function level	Mobility	Sensing and Perception	Control system	Supporting device	Etc.
Number of assigned P-SAO structures	719	1147	270	932	761

Table 9
In-depth analysis of 'Sensing and Perception' (feature level).

Sensing and Perception (function)	Features and related keywords		
Co-occurrence frequency within same function	Aerial sensor	Video	152
		Image	293
		GPS	92
		INS	7
		Ultrasonic	11
		Camera	507
	Perception algorithm	Perception	42
		Algorithm	156
		Calibration	78
Co-occurrence frequency between different functions	Aerial sensor	Video	12
		Image	43
		GPS	172
		INS	3
		Ultrasonic	7
		Camera	266
	Perception algorithm	Perception	23
		Algorithm	177
		Calibration	31

5. Discussion

5.1. Incorporation of the lead user knowledge and the effectiveness of proposed approach

This paper proposed a novel approach to the textual CBR to identify problems and explore feasible solutions from knowledge in lead user communities, addressing a lack of approaches to leverage the big textual data therein. It embraces innovation from outside the boundaries of the firm, especially lead user innovations, into their own innovation strategies in two ways: incremental innovation and radical innovation. First, incremental innovation can be achieved by improving the performance of existing products based on the lead users' uncomfortable experiences. Second, totally new needs can be utilized as a starting point for radical innovation.

Table 7
A partial result of problems and solutions assigned to sensing and perception.

Reference (documents)	Requirement issue(problems)	Related features	Alternative(solutions)	Related function and feature	Innovationtype
443	professional aerial photographers want to obtain detailed pictures and videos of buildings or landscapes	Aerial sensor	The dust proof coating and propulsion system protects it from external electrical components The retractable landing skids make it easier to transport the drone		Incremental
1241	serious real time video processing power is required on the order of significant multi GPU based system	Aerial sensor	The Exo360 brings together the latest in drone and VR technology The drone can capture video in both linear and spherical modes	Sensing and Perception – Aerial sensor	Radical
1682	Most camera will not operate for more than 2 h without external power supplyThe camera will need external power for the long flight duration	Aerial sensor	The major breakthrough is the new Lithium Ion battery pack rated at 3C continuous discharge the new long range conservation drone open up the possibility to map or video area	Sensing and Perception – Aerial sensor	Radical
3045	The crucial component for geotagging images is the GPS receiver	Aerial sensor	images georeferenced using common GPS modules are sufficientimage matching techniques might be appliedthe camera settings have to be adjusted for every single image		Incremental
	the problem is a proper calibration of the gimbal	Perception algorithm	NADIR gimbals provided an easy way to reduce perspective distortions The offset can easily be determined using Google Earth	Supporting device –Supporting device	
3863	my photos came out blurry	Aerial sensor	The landing gear also allows you to take off like a traditional plane Canon Hack Development Kit allow you to use intervalometer scripts to trigger camera	Supporting device –Supporting device Sensing and Perception – Perception algorithm	Incremental
3963	you want to see a really impressive use of image processing		It used the simulated copters position and altitude	Etc	Incremental

In particular, the case representation and case retrieval process, which have been emphasized in textual case-based reasoning, are further improved in this study. In the process of conducting our case study, the effectiveness of the proposed approach is confirmed, as shown in Table 10. First, the classification of SAO structures in a document into the problem, solution, and neutral statement facilitates the interpretation of case retrieval results. The context can be grasped through SAO structures beyond the keyword-based approaches. Second, the experts' domain knowledge and the lead users' knowledge based on their experience are used complementarily in the form of mediating keywords. It extends the scope of the retrieval process exploring various solutions related to the problematic root function.

In order to verify the effectiveness, the comparison of the retrieved results is conducted by performing the traditional keyword-based textual CBR on the same database. The problem statement, 'autonomous landing accurately and reliably on a moving platform', is converted to a keyword vector as an input of textual CBR. The input keywords are autonomous, landing, takeoff, accurate, precise, reliable, moving, platform, ground, and station. Then, the collected documents are transformed to term-document matrix for calculating the cosine similarity. As a result, the keywords list, which represents the documents in the order where the similarity score is high, is retrieved, as shown in Table 11. It is inevitable to review the document to confirm the exact solution since the keywords list alone could not deliver even the concept of a solution. In addition, it is nearly impossible to explore issues comprised of keywords other than the input keywords.

Table 10
The effectiveness of the proposed approach compared with the keyword-based approach.

Illustrative document:Drones have already been in use in the security and surveillance industry, bringing a significant change in how the operations are carried out. However, most current aerial security and surveillance systems are either tied to a particular drone hardware, or need significant manual intervention during operation. These solutions lack critical features and software capabilities, such as, AI and machine learning for automated alerts, automatic mission scheduling, compatibility with wide-range of drone hardware, etc. This makes it expensive, and often infeasible, to deploy the drone-based security/surveillance solutions at scale. Drones have already established the value that they bring to the table, in terms of mobility, unrestricted bird's eye view and accessibility. The focus is now on efficiencies and realising a meaningful return on investment for wide commercial adoption. This calls for integration of "intelligence" and "connectivity" with drones, to build completely automated and integrated workflows. FlytSecurity offers a plug-and-play, drone-agnostic, SaaS platform to quickly deploy and scale drone-based automated security operations. This significantly cuts down the cost of development and time to market, translating into an attractive ROI for the drone security service providers. With a wide range of features, like, 4G/LTE connectivity over unlimited range, live video, control and telemetry, fleet management (for simultaneous coverage of a large, distributed facilities), AI/ML for automated alerts, automated mission schedules, FlytSecurity enables fully-automated 24 × 7 operations at scale. Compatibility with any drone hardware, further makes FlytSecurity easy to adapt to variety of customer requirements (large/small drone, long/short endurance, quad-planes/multicopters, thermal/RGB sensor, etc.), and makes it easy to upgrade hardware at any time.	
SAO-based approach	Keyword-based approach
Problem/Solution/Neutral SAO structures- most current aerial security and surveillance system are either tied to a particular drone hardware (N)- a particular drone hardware need significant manual intervention (P)- These solutions lack critical features and software capabilities (P)- This makes it expensive and infeasible (P)- The focus is now on efficiencies (N)- This calls for integration of intelligence and connectivity (P)- FLytSecurity scale drone-based automated security operations (N)- This significantly cuts down the cost of development and time (S)	DocumentTermMatrix (TermFrequency)- drone (9), automate (5), security (5), operation (4), hardware (4), scale (3), surveillance (3), access (2), alert (2), compatibility (2), connectivity (2), deploy (2), integration (2), mission (2), range (2), solution (2), mobility (1), plug (1), telemetry (1)DocumentTermMatrix (Tf-idf)- FlytSecurity (0.3572), automate (0.1295), security (0.1156), surveillance (0.0894), alert (0.07767), AIML (0.0715), connectivity (0.0661), operation (0.0639)

Table 11
A keywords list of the top 10 documents in the order where the similarity score is high.

Document no	Similarity score	Keywords
9283	0.2108	landing, speed, camera, height, helicopter, object, platform, surface, vision, chip, hover
5406	0.2107	autonomous, landing, compass, gps, multipath, back, ground, throttle, wind, runway
4904	0.2107	landing, car, roof, quad, antenna, moving, gear, airbone, speed, control, technology
4822	0.2107	landing, flap, fuselage, canopy, gear, skid, gimbal, grass, wing
7475	0.2013	landing, gear, camera, polycarbonate, mount, platform, camera, smart, simulator
2546	0.2009	autonomous, landing, plugin, platform, infrastructure, LTE, solution, environment, scale
1407	0.2009	landing, code, copter, rangefinder, altitude, precision, accuracy, platform
1164	0.2009	autonomous, video, Bluetooth, autopilot, data, online, accurate, cloud, device
395	0.2003	landing, wing, Bluetooth, autopilot, , wind, perching, algorithm, bird, disaster
1828	0.1924	autonomous, landing, data, time, altitude, thrust, algorithm, solution, mapping, sensor

5.2. Combination of other methods to utilize lead user communities

This study contributes to the field in that it presents a new data source for supporting the development of a technology-intensive product. As shown in the case study, there are heterogeneous problems and adequate conceptual solutions in lead user communities. The use of CBR is clearly a methodology that reflects the distinctive characteristics of lead user communities. However, it is also important to select the prerequisite issues and explore decent solutions where the technology-intensive product is in the early stage of growth. In this case, as the case adaptation step becomes more important as in the conventional CBR approach, a framework that focuses more on finding similar problems from experience datasets should be developed. For this reason, it seems to be required to combine a certain method with other powerful methods. For example, it should be noted that other methods in Table 12 can be employed as a remedy to develop quantitative indicators for the case adaptation guideline suggested in this paper.

6. Conclusion

This paper focuses on the necessities of finding problems from lead

Table 12
Possible methods applicable to the case adaptation stage.

Method	Applicability for development of case adaptation guideline
Topic modeling	Topic modeling technique such as Latent Dirichlet Allocation (LDA) defines topics based on the distribution of keywords from a large corpus and classifies documents into each topic (Blei, Ng, & Jordan, 2003). Specifically, it is possible to calculate the importance of features through the value (i.e. beta in LDA) of the influence of the keyword on the topic. Thus, it helps to judge the technical features that are central to a particular problem.
Network analysis	A network can be constructed based on the co-occurrence matrix between keywords. In order to observe the changes of a certain keyword in the network, degree centrality, closeness centrality, and betweenness centrality indicating the structure of the network can be utilized (Choi & Hwang, 2014; Knoke & Kuklinski, 1982). They can serve as indicators for evaluating the importance of technical features.
MCDA techniques	The process of selecting the problem to be solved should take into consideration several criteria rather than one. The multi-criteria decision analysis techniques fit this situation. Typically, AHP and ANP (Satty, 2004) can reflect the firm's strategic objectives in the case adaptation process.

user communities and proposes a problem-oriented CBR approach for technology-intensive product development. Highlighting the importance of lead user communities, this research focuses on the first phases of CBR, how to define problems and solutions in textual CBR, suggesting problem-oriented CBR. When problems are identified from the lead user communities, case retrieval and adaption are conducted using technology tree and co-occurrence analysis.

This paper contributes to research in three ways. First, from the theoretical perspective, this study proposes a differentiated approach reflecting the characteristics of the lead user, focusing on the characteristics of lead users that identify potential problems in advance to normal users and find the problem by themselves. This study employs sentiment analysis to distinguish the problems from the lead user communities, using the SAO structure instead of keyword-level analysis. The use of a technology tree is also useful to articulate the technological structure and compositions which is generally very complex in technology-intensive products. Third, from the managerial perspective, this study provides a well-organized and systematic approach for embracing lead user innovation from outside the boundaries of the firm into their own innovation strategies, supporting the open innovation paradigm. Since our approach can both efficiently identify the problems from lead users' experiences and also suggest ways to find out related solutions based on their ideas, firms can benefit by incorporating it into their R&D efforts.

Despite this contribution, there are some limitations to be addressed in future studies. First, in the case adaptation stage, this study provides only rough guidelines to find critical functions and features to be addressed. The applicability of solutions is evaluated through the suggested checklist. So another fruitful avenue for future research would be to employ relevant methodologies to automate the evaluation process. Second, even if lead users are generally helpful, the impact of each individual might vary. Therefore, identifying lead users with higher impact is another important issue to be addressed. Interaction among lead users occurs in the form of discussion and user comments. Utilizing these interactions in the network analysis with quantitative indicators such as the number of views and comments on the post, impactful lead users can be identified. Third, this research partially succeeded in converting raw textual data into intelligence knowledge in the case representation stage. It seems necessary to apply state-of-the-art NLP techniques by narrowing the scope of the study to the 'development of an algorithm that increases the accuracy of identifying and classifying problem statements in a specific document.' Last but not least, this study employs a single case study to illustrate our approach. However, this study can be extended to the different industries because what we suggest is the general process and procedure. Therefore, conducting the case study for different industries and comparing the industrial differences might be great future work. How customers express their problems and ideas can be different according to the industries because sometimes lead users might not actively contribute in some industries. Thus, analyzing industrial differences can be another good research direction.

CRediT authorship contribution statement

Mintak Han: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Youngjung Geum:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by Korea government (MOE) (NRF-2020R1I1A2070429).

References

Althoff, K. D. (2001). Case-based reasoning. In Handbook of Software Engineering and Knowledge Engineering: Volume I: Fundamentals (pp. 549-587).

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science, 337*(6092), 337–341. https://doi.org/10.1126/science.1215842

Avramenko, Y., & Kraslawski, A. (2008). Case-Based Reasoning Approach. *Case Based Design: Applications in Process Engineering, 51–70*. https://doi.org/10.1007/978-3-540-75707-8_3

Bach, K. (2013). *Knowledge acquisition for case-based reasoning systems*. Hut: Verlag Dr.

Barták, R., Hraško, A., & Obdržálek, D. (2014, May). A controller for autonomous landing of AR. Drone. In The 26th Chinese Control and Decision Conference (2014 CCDC) (pp. 329-334). IEEE. 10.1109/CCDC.2014.6852167.

Benamara, F., & Moriceau, V. (2006). *Prepositions in Cooperative Question-Answering Systems: A Preliminary Analysis. In Syntax and semantics of prepositions* (pp. 307–330). Dordrecht: Springer.

Bentaiba-Lagrid, M. B., Bouzar-Benlabiod, L., Rubin, S. H., Bouabana-Tebibel, T., & Hanini, M. R. (2020). A case-based reasoning system for supervised classification problems in the medical field. *Expert Systems with Applications, 150*, Article 113335. https://doi.org/10.1016/j.eswa.2020.113335

Berka, P. (2020). Sentiment analysis using rule-based and case-based reasoning. *Journal of Intelligent Information Systems, 1–16*. https://doi.org/10.1007/s10844-019-00591-8

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Cascini, G., Fantechi, A., & Spinicci, E. (2004). September). Natural language processing of patents and technical documentation. In *International Workshop on Document Analysis Systems* (pp. 508–520). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-28640-0_48.

Ceausu, V., Despres, S. (2007) A semantic case-based reasoning framework for text categorization. In: The Semantic Web, Lecture Notes in Computer Science. pp. 736–749 10.1007/978-3-540-76298-0_53.

Chaitanya, T. S. K., Harika, V. S., & Prabadevi, B. (2017). A sentiment analysis approach by identifying the subject object relationship. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 62–68). IEEE.

Chen, W., Wang, X., Wang, W., Zhu, Y., Cai, Z., & Yang, S. (2021). A heterogeneous GRA-CBR-based multi-attribute emergency decision-making model considering weight optimization with dual information correlation. *Expert Systems with Applications, 182*, Article 115208. https://doi.org/10.1016/j.eswa.2021.115208

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Choi, J., & Hwang, Y. S. (2014). Patent keyword network analysis for improving technology development efficiency. *Technological Forecasting and Social Change, 83*, 170–182.

Choi, S., Kang, D., Lim, J., & Kim, K. (2012). A fact-oriented ontological approach to SAO-based function modeling of patents for implementing Function-based Technology Database. *Expert Systems with Applications, 39*(10), 9129–9140. https://doi.org/10.1016/j.eswa.2012.02.041

Choi, S., Park, H., Kang, D., Lee, J. Y., & Kim, K. (2012). An SAO-based text mining approach to building a technology tree for technology planning. *Expert Systems with Applications, 39*(13), 11443–11455. https://doi.org/10.1016/j.eswa.2012.04.014

Clarysse, B., Wright, M., Bruneel, J., & Mahajan, A. (2014). Creating value in ecosystems: Crossing the chasm between knowledge and business ecosystems. *Research Policy, 43*(7), 1164–1176. https://doi.org/10.1016/j.respol.2014.04.014

Cocchioni, F., Mancini, A., & Longhi, S. (2014, May). Autonomous navigation, landing and recharge of a quadrotor using artificial vision. In 2014 international conference on unmanned aircraft systems (ICUAS) (pp. 418-429). IEEE. 10.1109/ICUAS.2014.6842282.

Dahlander, L., Frederiksen, L., & Rullani, F. (2008). Online communities and open innovation. *Industry and Innovation, 15*(2), 115–123. https://doi.org/10.1080/13662710801970076

Dahlander, L., & Wallin, M. W. (2006). A man on the inside: Unlocking communities as complementary assets. *Research Policy, 35*(8), 1243–1259. https://doi.org/10.1016/j.respol.2006.09.011

Denning, P. J. (2001). The profession of it: Crossing the chasm. *Communications of the ACM, 44*(4), 21–25. https://doi.org/10.1145/367211.367284

Desouza, K. C., Awazu, Y., Jha, S., Dombrowski, C., Papagari, S., Baloh, P., & Kim, J. Y. (2008). Customer-driven innovation. *Research-Technology Management, 51*(3), 35–44. https://doi.org/10.1080/08956308.2008.11657503

Dong, R., Schaal, M., O'Mahony, M. P., McCarthy, K., & Smyth, B. (2013, July). Opinionated product recommendation. In International conference on case-based reasoning (pp. 44-58). Springer, Berlin, Heidelberg. 10.1007/978-3-642-39056-2_4.

Eisenberg, I. (2011). Lead-user research for breakthrough innovation. *Research-Technology Management, 54*(1), 50–58. https://doi.org/10.1080/08956308.2011.11657673

Fader, A., Soderland, S., & Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 1535–1545).

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82–89. https://doi.org/10.1145/2436256.2436274

Fiok, K., Karwowski, W., Gutierrez, E., & Wilamowski, M. (2021). Analysis of sentiment in tweets addressed to a single domain-specific Twitter account: Comparison of model performance and explainability of predictions. *Expert Systems with Applications, 186*, Article 115771. https://doi.org/10.1016/j.eswa.2021.115771

Franke, N., & Piller, F. (2004). Value creation by toolkits for user innovation and design: The case of the watch market. *Journal of Product Innovation Management, 21*(6), 401–415. https://doi.org/10.1111/j.0737-6782.2004.00094.x

Füller, J., Jawecki, G., & Mühlbacher, H. (2007). Innovation creation by online basketball communities. *Journal of Business Research, 60*(1), 60–71. https://doi.org/10.1016/j.jbusres.2006.09.019

Geetha, S., Narayanamoorthy, S., Manirathinam, T., & Kang, D. (2021). Fuzzy case-based reasoning approach for finding COVID-19 patients priority in hospitals at source shortage period. *Expert Systems with Applications, 178*, 114997.

Geum, Y., Noh, E., & Park, Y. (2016). Generating new service ideas: The use of hybrid innovation tools to reflect functional heterogeneity of services. *R&D Management, 46*, 736–748. https://doi.org/10.1111/radm.12118

Goel, A., & Diaz-Agudo, B. (2017, February). What's hot in case-based reasoning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

Gombault, A., Allal-Chérif, O., & Décamps, A. (2016). ICT adoption in heritage organizations: Crossing the chasm. *Journal of Business Research, 69*(11), 5135–5140. https://doi.org/10.1016/j.jbusres.2016.04.093

Guo, J., Wang, X., Li, Q., & Zhu, D. (2016). Subject–action–object-based morphology analysis for determining the direction of technological change. *Technological Forecasting and Social Change, 105*, 27–40. https://doi.org/10.1016/j.techfore.2016.01.028

He, W. (2013). Improving user experience with case-based reasoning systems using text mining and Web 2.0. *Expert Systems with Applications, 40*(2), 500–507. https://doi.org/10.1016/j.eswa.2012.07.070

He, W., & Wang, F. K. (2016). Integrating a case-based reasoning shell and Web 2.0: Design recommendations and insights. *World Wide Web, 19*(6), 1231–1249. https://doi.org/10.1007/s11280-015-0380-y

Hérissé, B., Hamel, T., Mahony, R., & Russotto, F. X. (2010). A terrain-following control approach for a vtol unmanned aerial vehicle using average optical flow. *Autonomous Robots, 29*(3–4), 381–399. https://doi.org/10.1007/s10514-010-9208-x

Jin, R., Cho, K., Hyun, C., & Son, M. (2012). MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Systems with Applications, 39*(5), 5214–5222. https://doi.org/10.1016/j.eswa.2011.11.018

Kim, J., & Park, Y. (2019). Leveraging ideas from user innovation communities: Using text-mining and case-based reasoning. *R&D Management, 49*(2), 155–167. https://doi.org/10.1111/radm.12292

Knoke, D., & Kuklinski, J. H. (1982). *Network analysis*. Beverly Hills: SAGE Publishing.

Köhler, A. R., & Som, C. (2014). Risk preventative innovation strategies for emerging technologies the cases of nano-textiles and smart textiles. *Technovation, 34*(8), 420–430. https://doi.org/10.1016/j.technovation.2013.07.002

Kolodner, J. (2014). *Case-based reasoning*. Morgan Kaufmann.

Kozinets, R. V., Hemetsberger, A., & Schau, H. J. (2008). The wisdom of consumer crowds: Collective innovation in the age of networked marketing. *Journal of Macromarketing, 28*(4), 339–354. https://doi.org/10.1177/0276146708325382

Kwon, H., Kim, J., & Park, Y. (2017). Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. *Technovation, 60*, 15–28. https://doi.org/10.1016/j.technovation.2017.01.001

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Liu, D. R., & Ke, C. K. (2007). Knowledge support for problem-solving in a production process: A hybrid of knowledge discovery and case-based reasoning. *Expert Systems with Applications, 33*(1), 147–161. https://doi.org/10.1016/j.eswa.2006.04.026

Lüthje, C. (2004). Characteristics of innovating users in a consumer goods field: An empirical study of sport-related product consumers. *Technovation, 24*(9), 683–695. https://doi.org/10.1016/S0166-4972(02)00150-5

Maine, E., Thomas, V. J., & Utterback, J. (2014). Radical innovation from the confluence of technologies: Innovation management strategies for the emerging nanobiotechnology industry. *Journal of Engineering and Technology Management, 32*, 1–25.

Malhotra, Y. (2001). Expert systems for knowledge management: Crossing the chasm between information processing and sense making. *Expert Systems with Applications, 20*(1), 7–16. https://doi.org/10.1016/S0957-4174(00)00045-2

Marchi, G., Giachetti, C., & De Gennaro, P. (2011). Extending lead-user theory to online brand communities: The case of the community Ducati. *Technovation, 31*(8), 350–361. https://doi.org/10.1016/j.technovation.2011.04.005

Maurer, F., & Melnik, G. (2007, May). Agile methods: Crossing the chasm. In 29th International Conference on Software Engineering (ICSE'07 Companion) (pp. 176-177). IEEE. 10.1109/ICSECOMPANION.2007.18.

Montoyo, A., Martínez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems, 53*(4), 675–679. https://doi.org/10.1016/j.dss.2012.05.022

Moore, G. A. (1991). *Crossing the chasm*. Harper Business Essentials.

Muhammad, I., Kearney, A., Gamble, C., Coenen, F., & Williamson, P. (2020, September). Open Information Extraction for Knowledge Graph Construction. In International Conference on Database and Expert Systems Applications (pp. 103-113). Springer, Cham. 10.1007/978-3-030-59028-4_10.

Mustapha, S. S. (2018). Case-based reasoning for identifying knowledge leader within online community. *Expert Systems with Applications, 97*, 244–252. https://doi.org/10.1016/j.eswa.2017.12.033

Nambisan, S. (2002). Designing virtual customer environments for new product development: Toward a theory. *Academy of Management review, 27*(3), 392–413. https://doi.org/10.5465/amr.2002.7389914

Ohana, B., Delany, S. J., & Tierney, B. (2012, September). A case-based approach to cross domain sentiment classification. In International Conference on Case-Based Reasoning (pp. 284-296). Springer, Berlin, Heidelberg. 10.1007/978-3-642-32986-9_22.

Okudan, O., Budayan, C., & Dikmen, I. (2021). A knowledge-based risk management tool for construction projects using case-based reasoning. *Expert Systems with Applications, 173*, Article 114776. https://doi.org/10.1016/j.eswa.2021.114776

Plaza, E. (2009). On Reusing Other People's Experiences. *künstliche intelligenz, 23*(1), 18–23.

Qiu, L., Pang, J., & Lim, K. H. (2012). Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity: The moderating role of review valence. *Decision Support Systems, 54*(1), 631–643. https://doi.org/10.1016/j.dss.2012.08.020

Reuß, P., Stram, R., Juckenack, C., Althoff, K. D., Henkel, W., Fischer, D., & Henning, F. (2016, October). FEATURE-TAK-framework for extraction, analysis, and transformation of unstructured textual aircraft knowledge. In International Conference on Case-Based Reasoning (pp. 327-341). Springer, Cham. 10.1007/978-3-319-47096-2_22.

Richter, M. M., & Weber, R. O. (2013). *Case-based reasoning: A textbook*. Berlin An: Springer-Verlag.

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Roy, R. (2018). Role of relevant lead users of mainstream product in the emergence of disruptive innovation. *Technological Forecasting and Social Change, 129*, 314–322. https://doi.org/10.1016/j.techfore.2017.09.036

Satty, T. (2004). Fundamentals of the analytic network process-Dependence and feedback in decision-making with a single network. *Journal of Systems Science and Systems Engineering, 13*(2), 129–157.

Savransky, S. D. (2000). *Engineering of creativity: Introduction to TRIZ methodology of inventive problem solving*. CRC Press.

Sawhney, M., & Prandelli, E. (2000). Communities of creation: Managing distributed innovation in turbulent markets. *California management Review, 42*(4), 24–54. https://doi.org/10.2307/41166052

Schilling, M. A. (2017). *Strategic Management of Technological Innovation* (Fifth Edition). New York, NY: McGraw-Hill Irwin.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing, 21*(4), 76–94. https://doi.org/10.1002/dir.20090

Shen, L., Yan, H., Fan, H., Wu, Y., & Zhang, Y. (2017). An integrated system of text mining technique and case-based reasoning (TM-CBR) for supporting green building design. *Building and Environment, 124*, 388–401. https://doi.org/10.1016/j.buildenv.2017.08.026

Sizov, G., Öztürk, P., & Aamodt, A. (2015, September). Evidence-driven retrieval in textual CBR: bridging the gap between retrieval and reuse. In International Conference on Case-Based Reasoning (pp. 351-365). Springer, Cham. 10.1007/978-3-319-24586-7_24.

Torrent-Fontbona, F., Massana, J., & López, B. (2019). Case-base maintenance of a personalised and adaptive CBR bolus insulin recommender system for type 1 diabetes. *Expert Systems with Applications, 121*, 338–346. https://doi.org/10.1016/j.eswa.2018.12.036

Tuarob, S., & Tucker, C. S. (2015). Automated discovery of lead users and latent product features by mining large scale social media networks. *Journal of Mechanical Design, 137*(7), Article 071402. https://doi.org/10.1115/1.4030049

Vaughan, M. R., Seepersad, C. C., & Crawford, R. H. (2014, August). Creation of empathic lead users from non-users via simulated lead user experiences. In ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. V007T07A048-V007T07A048). American Society of Mechanical Engineers. 10.1115/DETC2014-35052.

Verhaegen, P. A., D'hondt, J., Vandevenne, D., Dewulf, S., & Duflou, J. R. (2011). Identifying candidates for design-by-analogy. *Computers in Industry, 62*(4), 446–459. https://doi.org/10.1016/j.compind.2010.12.007

Von Hippel, E. (1986). Lead users: A source of novel product concepts. *Management Science, 32*(7), 791–805. https://doi.org/10.1287/mnsc.32.7.791

Von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: A case study. *Research policy, 32*(7), 1217–1241. https://doi.org/10.1016/S0048-7333(03)00050-7

Wan, Q., Xu, X., Zhuang, J., & Pan, B. (2021). A sentiment analysis-based expert weight determination method for large-scale group decision-making driven by social media data. *Expert Systems with Applications, 185*, Article 115629. https://doi.org/10.1016/j.eswa.2021.115629

Weber, R. O., Aha, D., Sandhu, N., & Munoz-Avila, H. (2001). A textual case-based reasoning framework for knowledge management applications. In *In: Proceedings of the ninth german Workshop on Case-Based Reasoning* (pp. 244–253).

Weber, R. O., Ashley, K. D., & Brüninghaus, S. (2005). Textual case-based reasoning. *The Knowledge Engineering Review, 20*(3), 255–260. https://doi.org/10.1017/S0269888906000713

Wu, M. C., Lo, Y. F., & Hsu, S. H. (2006). A case-based reasoning approach to generating new product ideas. *The International Journal of Advanced Manufacturing Technology, 30*(1–2), 166–173. https://doi.org/10.1007/s00170-005-0048-3

Wu, W., Zhu, Z., Zhang, G., Kang, S., & Liu, P. (2021). A reasoning enhance network for muti-relation question answering. *Applied Intelligence, 1–10*. https://doi.org/10.1007/s10489-020-02111-6

Xiao, X., Skitmore, M., & Hu, X. (2017). Case-based reasoning and text mining for green building decision making. *Energy Procedia, 111*, 417–425. https://doi.org/10.1016/j.egypro.2017.03.203

Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics, 88*(1), 213–228. https://doi.org/10.1007/s11192-011-0383-0

Yoon, J., Park, H., & Kim, K. (2013). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics, 94*(1), 313–331. https://doi.org/10.1007/s11192-012-0830-6

Zhou, F., Jiao, R. J., & Linsey, J. S. (2015). Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews. *Journal of Mechanical Design, 137*(7), Article 071401. https://doi.org/10.1115/1.4030159

Zhu, G. N., Hu, J., Qi, J., Ma, J., & Peng, Y. H. (2015). An integrated feature selection and cluster analysis techniques for case-based reasoning. *Engineering Applications of Artificial Intelligence, 39*, 14–22. https://doi.org/10.1016/j.engappai.2014.11.006