Technological Forecasting & Social Change 146 (2019) 133–147

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Identification of the unique attributes and topics within Smart Things Open Innovation Communities

M. Olmedilla[a], H. Send[b], S.L. Toral[c,*]

[a] Léonard de Vinci Pôle Universitaire, Research Center, 92916 Paris, La Défense, France
[b] Alexander von Humboldt Institute for Internet and Society (HIIG), Französische Straße 9, 10117 Berlin, Germany
[c] Escuela Técnica Superior de Ingeniería, Camino Descubrimientos, Isla Cartuja, 41092 Sevilla, Spain

ARTICLE INFO

Keywords:
Text mining
Unique innovations
Open communities
Unique attributes
Co-occurrence analysis
Open innovation

ABSTRACT

One of the main challenges of open innovation communities is how to create value from shared content either by selecting those ideas that are worthy of pursuit and implementation or by identifying the users' preferences and needs. These tasks can be done manually when there is an overseeable amount of content or by using computational tools when there are massive amounts of data. However, previous studies on text mining have not dealt with the identification of unique attributes, which can be defined as those contributions that are inextricably linked with a specific tag or category within open innovation websites. The uniqueness of these ideas means that they can only be obtained through a selection of one choice among several alternatives. To obtain such unique ideas and thus to also obtain innovations, this paper proposes a novel methodology called co-occurrence differential analysis. The proposed methodology combines traditional co-occurrence analysis with additional statistical processing to obtain the unique attributes and topics associated with different alternatives. The identification of unique content provides valuable information that can reveal the strengths and weaknesses of several options in a comparative fashion.

1. Introduction

Global competition, shortening product life cycles and increasingly complex products are continuously increasing the importance of innovation management. Thus, to achieve effective and efficient product development, practitioners and research studies are focusing on the information exchanges and modes of collaboration with stakeholders (Bashir et al., 2017; Cooke and Buckley, 2008; Lee et al., 2012). Communication on the Internet is central to these efforts because of its unparalleled ability to reach large audiences, its low transaction costs and its provision of a great amount of independency from time and place.

Following the first introduction of open innovation (Chesbrough, 2003), most firms embraced the idea of using valuable contributions from outside of the firm and are now employing parts of the open innovation concept. Especially in the context of digitally enabled products and services, it is now commonly accepted that innovation is most powerful when it can rely on distributed knowledge and thereby rely on access to resources in networks and across dynamic knowledge domains (Nambisan et al., 2017). In innovation challenges, crowdsourcing input from an unknown public is currently a popular approach to gather outside ideas, although some researchers question the quality of ideas generated herein (Malhotra and Majchrzak, 2014). In most cases, firms find it difficult to actively obtain feedback on their products and services (Dahlander and Piezunka, 2014). Part of the problem is that asking users to actively engage in ideation or creative activities on behalf of a firm is a lot to ask in times of ever-shrinking attention spans and packed schedules. If users do engage in such endeavours, they rarely connect and refine existing ideas and designs but instead generate isolated ideas (Majchrzak and Malhotra, 2013), leaving the work of making connections between the ideas to the researchers and consultants.

1.1. User generated content and online communities

Against this backdrop, user-generated content has become an important source of information. The amount of useful voluntarily generated content is rising continually (Kim et al., 2017) but firms find it difficult to direct user contributions to specific topics and to manage and harness communities (Haefliger et al., 2011). Research is currently concerned with different organizing approaches to the production of user-generated input, the provision of incentives, and the governance of

* Corresponding author.
E-mail addresses: maria.olmedilla_fernandez@devinci.fr (M. Olmedilla), h.send@hiig.de (H. Send), storal@us.es (S.L. Toral).

https://doi.org/10.1016/j.techfore.2019.05.004
Received 18 May 2018; Received in revised form 3 May 2019; Accepted 4 May 2019

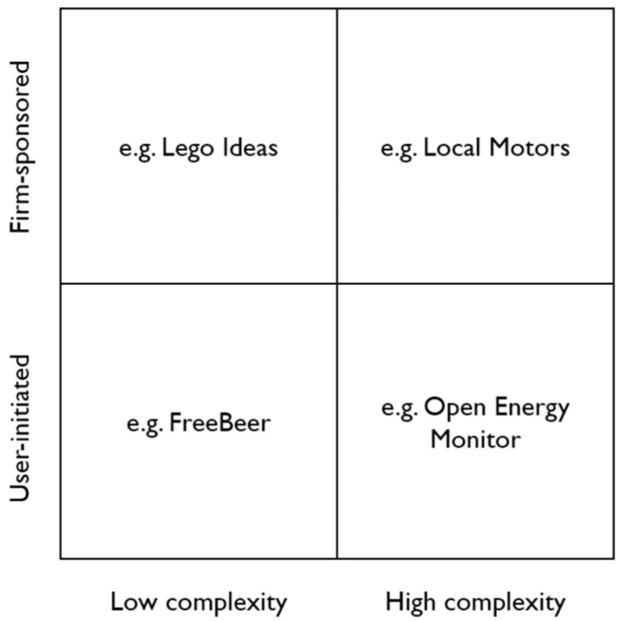


Fig. 1. User- and firm-initiated innovation communities.

the process (Majchrzak and Malhotra, 2013). The use of social media networks requires minimal investment, and efficient management of these digital tools is widely required to generate cyclical and continuous innovation (Scuotto et al., 2017). At this point, how firms can appropriate co-created assets still remains a central issue (Bashir et al., 2017; Bresciani et al., 2018; Haefliger et al., 2011).

Depending on how close a community is to a firm, the firm has more or less control over the community's activities (Fig. 1). Some examples of such firm-initiated communities are Lego Ideas, Dell Idea Storm created by Dell or MyStarbucksIdea created by Starbucks (Di Gangi and Wasko, 2009; Martínez-Torres et al., 2015). In all these cases, the communities are directly aligned towards each firm's products and services, and users contribute by proposing small innovations or just by posting their needs. There are also communities aimed at more professional users, such as Launch Forth by Local Motors, which deal with more complex problems such as designing a self-driving bus.

By contrast, open source communities are situated outside of a firm's control and initiated by user innovators (Von Hippel, 2001). We again can find instances of low complexity products such as the following: (1) electronic word-of-mouth communities, such as Amazon or Ciao (Arenas-Márquez et al., 2014; Ulloa et al., 2016), (2) a community centred on achieving a recipe for brewing beer, or (3) instances of very complex products, such as the Open Energy Monitor community to monitor energy consumption.

Finally, intermediary facilitators represent another possible solution. Such facilitators include service providers, such as InnoCentive, Kickstarter and Seedr, which connect firms with innovation solvers. Firms can post their projects on these platforms, and a call for proposals/solutions will be initiated to registered members. Typically, winning solutions receive a monetary reward (Allio, 2004).

Open innovation in this context is concerned with firms that actively seek out the users' contributions to innovations, while user innovation describes users that engage in innovation activities on their own and do not expect monetary remuneration (Bogers and West, 2012). Both fields overlap when firms access user innovation communities to obtain new knowledge for innovation (ibid.).

1.2. Analysing customer feedback

Practitioners face a trade-off regarding feedback. It can be sought in the form of qualitative or quantitative information. Structured quantitative data such as that derived from questionnaires and multiple-item scales are easy to process but do not provide the opportunity to express an opinion on unforeseen subjects. However, in product development, structured quantitative data is a novelty that we often use in order to generate new interesting features or products. Qualitative data is less structured but can be rich in context, can address completely unanticipated topics and is more suitable to capture the broad impression of users in the context of a product. However, the processing and analysis of qualitative data remains time intensive and thus costly (Walter and Back, 2013), and it is better to devote employee retention to foster knowledge specialization and fortification (Dezi and Del Giudice, 2014; Papa et al., 2019a, 2019b).

Researchers have therefore developed approaches for the automated analysis of large amounts of text by using text mining. Frequent applications of text mining include text categorization and extraction of topics, sentiment analysis, and summarizing documents. Approaches to topic extraction such as TF-IDF are useful when trying to classify documents or attempting to extract the main topic of several documents. Moreover, in many instances, humans can intuitively extract such information, and it will be useful for research or business applications: patents, documents and user-generated content can be adequately found, quickly grouped and assessed or stored by using the central topic. Accordingly, in everyday life, it is rather uncommon to just look for the unique features of a text or a piece of information. In a manner consistent with the topic extraction process, we classify the architecture of buildings and books by their main themes and by their prevalent combinations.

By contrast, people do not classify cooking recipes by the one ingredient that is only used in the focal recipe, buildings by the single feature that is only found in its architecture or books by the few pages on a distinct topic that sets the book apart from all others. Therefore, the algorithms that cater to this need have apparently played no role in research. However, in analysing the user-generated content that revolves around the implementation of a service for different operating systems, we seek to identify the unique terms for each of the implementations to identify the unique benefits and problems that each of these implementations brings.

The overall objective of this paper is to discover ideas and potential innovations within open innovation communities that are uniquely associated with one among several alternative topics. Being uniquely associated with one choice and no other alternatives could mean that the potential innovation is only affordable by using a specific technology; therefore, it could represent a future competitive advantage and could lead to a winner technology. To this end, a text mining method called co-occurrence differential analysis (CODA) is proposed. The main contribution of the proposed methodology is its ability when performing a comparative analysis, to capture those unique attributes specifically associated with a given tag.

The rest of the paper is organized as follows: the next section details relevant previous works in the field of open innovation from the knowledge flow and data analytics perspective. Section 3 presents the research setting, the main hypotheses and research questions. Section 4 introduces the methodology. Section 5 describes the case study and data collection; the results are described in Section 6. Along with the limitations and future work, Section 7 discusses the findings and implications. Finally, Section 8 concludes the paper.

2. Related work

This paper combines central ideas of open innovation with the data analytics perspective.

2.1. Open innovation and user integration

Chesbrough (2003) described how firms strategically allow knowledge flows across their boundaries to generate innovations more efficiently or to more broadly profit from their own research and development. The idea of open innovation describes the conscious opening of the innovation process and the company in the context of information flows into and out of the company. Thus, the approach can also be explained in the context of the research on knowledge spillovers (Chesbrough et al., 2013). Knowledge spillovers describe the unintentional migration of information from a company by exchanging employees with other firms and are therefore also seen from a cost perspective for the focal company. By their very nature, these spillovers are not subject to explicit management. Open innovation, on the other hand, reintroduces the idea of management to knowledge spillovers and describes channels, roles, processes and mechanisms to actively manage the knowledge flows across the boundary of the enterprise (Tucci et al., 2016). Previous research demonstrates that knowledge management system facilitates the creation of open and collaborative ecosystems, and the exploitation of internal and external flows of knowledge (Bresciani et al., 2018; Ferraris et al., 2017). Moreover, networks and collaborations improve innovation performance, following an open innovation logic (Santoro et al., 2018a). The open innovation literature discusses methods and approaches that can be used to more advantageously manage these information flows from and into the enterprise. Beyond the level of the firm, Wang et al. (2012) show that open innovation practices improve the effectiveness and reinforce the importance of innovation systems at the national level by moving the focus of innovation into a network of firms. A complete review about knowledge management practices that support open innovation activities can be found in Natalicchio et al. (2017).

The question of costs and benefits of an open innovation process is a central line of argument in open innovation research (Chesbrough et al., 2013; Laursen and Salter, 2006). Research addresses the cost of accessing talented employees, comparing the costs of internal or open development and the effective use of intellectual capital. Chesbrough et al. (2013)disagree with the perception that knowledge exchange primarily comes at costs for the focal company, as the knowledge spillover perspective suggests. Instead, Chesbrough (2003) describes that strong specialization, increasingly differentiated areas of knowledge, and globally distributed talent prohibitively increase the cost of hiring the best talent for a company. Laursen and Salter (2006) have shown that the search for external information still comes at a cost and that both breadth and depth of external search have an inverted U-shape relationship with firm performance, showing that more openness is not always better.

Research on user integration (Von Hippel, 1986) and open innovation research share the investigation of valuable, external information that users and customers contribute and that exists dispersed outside the company. Gassmann et al. (2010) therefore describe user innovation as a part of open innovation and note the intense research activity taking place in this area. Both research directions deal with users, but open innovation is a firm-centric paradigm concerned with the strategic and operational possibilities of the firm. However, user innovation is user-centric and more concerned with the possibilities and needs of users (Piller and West, 2014). Hippel has focused his research strongly on customers who are more demanding than other customers and that, additionally, have acted themselves to meet their needs for specific products or product features. He refers to customers who meet these two qualities as lead users. Lead users develop products to meet their own needs and to use them by themselves. In contrast, the researched companies in the field of open innovation have developed products with the intention of making a profit. Current research in this area also deals with the question of how communities can be designed to integrate lead users into the innovation process (Schemmann et al., 2016). Such communities can exist within the direct interactions of a group of individuals. However, a great deal of research is concerned with online communities that can exist at very low transaction cost and are much less restricted by the time and place of their members (Mahr and Lievens, 2012; Shirky, 2011). Therefore, online communities are intensely used in practice and research (Bayus, 2013; Di Gangi and Wasko, 2009; Füller et al., 2006; West and Bogers, 2014). The goal of accessing external knowledge in the form of ideas and feedback, especially from within online communities, has also been central to the open innovation research (Dahlander and Wallin, 2006; Füller et al., 2008; Mahr and Lievens, 2012). Online communities can contribute to activities along the complete innovation process by identifying needs, generating ideas, modifying concepts, developing prototypes and testing products (Füller et al., 2008). In the context of our paper, an important finding about user innovation is that the commercial products resulting from the ideas of users are often very successful and can even be more successful than the products developed for profit (Franke et al., 2006; Hyysalo and Usenyuk, 2015).

Chesbrough (2003) reports in his studies on the use of open innovation in a wide variety of industries, and current surveys show that open innovation has become an established instrument for many companies (Cricelli et al., 2016). Because the research on user innovation also shows very extensive involvement by a significant proportion of national populations, with, e.g., 2.9 million people involved in user innovation activities in the UK (Von Hippel et al., 2012), a hitherto low level of application of user innovation is particularly surprising in the context of the Internet of Things (IoT) (Kortuem and Kawsar, 2010). However, Santoro et al. (2018b) conclude that the IoT is meant to be a technology facilitating open innovation and the activities of knowledge management communities by enabling the connectivity of individuals and organizations from different sectors. Smart manufacturing has also been envisioned as a field for open innovation and skills development, facilitating sustainable real-time business improvements (Edgar and Pistikopoulos, 2018). Finally, prior literature on SmartThings has mainly focused on Smart Cities (Incki and Ari, 2018) or Smart Living conditions, i.e., health care devices (Papa et al., 2019b).

Some creative users will directly express valuable ideas and that the best of those can then be identified, transferred and put them to use by the hosting firm (Bullinger et al., 2010). However, the users of products or potential customers often possess implicit knowledge about a problem and lack the specific knowledge to express or frame their knowledge so that it can be easily transferred to a firm (Rohrbeck et al., 2009; Von Hippel, 1998). In these cases, current research argues that user innovation outside of the firm is more effective because the access to sticky information is at prohibitively high transaction cost (Bogers and West, 2012).

2.2. The data analytics perspective

The open approach to innovation means managing a huge amount of data generated by different sources in real time, being this point the logical conceptual connection between the open innovation paradigm and data analytics (Del Vecchio et al., 2018). The cost of accessing the information provided by many geographically dispersed individuals with rather diverse expertise has been considerably reduced as a result of the advances in the Internet and in the development of collaboration tools (Morgan and Wang, 2010). The development of such computer tools has promoted innovation through communities in which participants from different backgrounds, with different skills, experiences and perspectives can work together (Majchrzak et al., 2004). This phenomenon is also known as crowdsourcing.

Millions of people can freely post their ideas or solutions thus, firms must deal with such volume of information that needs to be processed effectively (the quantity of information processed in a certain amount of time) and efficiently (the quality of extracted information). Consequently, firms are increasingly interested in applying powerful computational techniques to the data to reveal trends and patterns and extract new insights (George et al., 2014). In the context of open innovation communities, users can share ideas or even score other users' ideas. One of the most challenging problems relates to distinguishing those ideas worth being implemented (Brem and Bilgram, 2015;

Martínez-Torres, 2013; Martínez-Torres and Olmedilla, 2016). Because this process is time consuming, and the size of the data are huge, it is needed a pre-filtering of ideas by using computational techniques.

The problem of identifying valuable shared ideas can be approached in two different ways: either by focusing on the authors who posted the ideas or by focusing on the content of the shared ideas (Trabucchi et al., 2018). The first approach consists of identifying the so-called lead users, who have been described as those that can anticipate needs that will later be experienced by other users in the market (Urban and Von Hippel, 1988; Von Hippel, 1986). Later studies have confirmed that products based on the lead users' ideas lead to a higher degree of novelty and fit customer needs much better (Gruner and Homburg, 2000; Lilien et al., 2002). Lead users have been identified by using virtual stock markets (Spann et al., 2009), social network analysis (Martínez-Torres, 2014), swarm intelligence (Martínez-Torres and Olmedilla, 2016) and digital anthropology principles (Somoza Sánchez et al., 2018). The second approach is to analyse the content of the shared innovations. As the number of shared innovations can be very large, text mining techniques can facilitate this task. In the context of product innovations, several text mining techniques, such as topic modelling and co-occurrence analysis, have been used to analyse the content of innovations. Both techniques rely on a set of predefined terms that are typically selected by using some attributes, such as their TF (Term Frequency) or their TF-IDF (Term-Frequency-Inverse Document Frequency) values (Youn Kim and Yoon, 2013). The TF value, a non-normalized value, measures the frequency of the appearance of terms in each document in the corpus of documents. The TF-IDF value normalizes the TF value by using the inverse of the number of documents containing the term. Documents are represented as vectors of words by using the TF or the TF-IDF values as attributes. By using the similarity of vectors as the distance metric of a clustering algorithm, we can extract the final topics of discussion. To reduce the dimensionally of the vector space model, the so-called latent semantic indexing process can be applied to a documents-term matrix, which has been applied to the identification of topics within Dell Idea Storm (Martínez-Torres, 2015). Co-occurrence analysis consists of analysing the relationships among the terms extracted by using their TF or TF-IDF attributes. Unlike topic modelling, where the number of topics is unknown a-priori, the classes in co-occurrence analysis are predefined (Fournier-Viger et al., 2014). In the context of open innovation, binary classes, i.e., “good ideas” and “bad ideas”, are typically used together with machine learning techniques such as classifiers. For instance, Christensen et al. (2018) make use of support vector machines for an automatic idea detection system for an online home brewing community. The results show the correspondence between the items of several basic categories, the closer the items are to each other, the higher their similarity (Whitlark and Smith, 2001). Co-occurrence has been used to analyse patent information (Jeong and Kwon, 2014) and to obtain the structure of innovation systems (Lee and Su, 2010). In this study, we apply co-occurrence analysis, but instead of selecting a term by their TF or TF-IDF attributes, we propose to extract terms based on their uniqueness, that is, by the terms that are associated with one, and only one, of a set of pre-defined classes.

3. Research motivation and hypotheses

The proliferation of user-generated content through the Internet has promoted discussions and opinions over a wide variety of products and services. In many cases, forum and opinion websites organize the discussions in specific subjects. In general, the manual processing of data is unaffordable in many cases and can introduce some bias due to the subjectivity of the analysts. Therefore, the analysis of shared opinions requires the use of computational techniques, such as text mining techniques. For instance, given a query text, information retrieval provides the closest documents, and latent semantic analysis can discover the main topics of discussion within a given subject. This is useful when dealing with firm-initiated open innovation websites, where users can post their innovation ideas for a predefined set of topics, which are usually product lines or areas of interest of a company (Di Gangi and Wasko, 2009; Martínez-Torres et al., 2015). However, this is not the case for third-party open innovation websites, where different products competing against each other are reviewed and discussed or competing innovations are proposed (Sikdar and Vel, 2010). In this case, as the innovations are organized by products or choices, consumers post their innovations within the product where the innovation is best suited. One important issue when dealing with competing innovations is the discovery of the distinctive characteristics belonging to one choice compared to other related choices. For instance, there are many electronic word-of-mouth communities where users post opinions and reviews about different brands and models of smartphones. They are organized as different threads, so users that want to review a specific model post his or her opinion in that specific thread (Olmedilla et al., 2016b). In general, firm-initiated innovation websites are structured in several tags, usually provided by the firm itself, while user-initiated innovation websites are organized in several options or choices created either by the community manager or the users themselves. As innovations are already clustered in different classes, they can be further analysed as a classification problem. Obtaining the unique attributes belonging to each tag or option can provide their corresponding main attributes and topics, which is quite useful information for manufacturers. In the case of open innovation communities, the unique topics are the unique innovations associated with one choice compared to other alternatives. Hence, the following research questions are proposed:

RQ_1. In an open innovation community, what are the unique attributes that enable a user to differentiate among several alternative choices?

RQ_2. In third-party open innovation websites, can unique attributes be used to infer the unique differences between products or services?

Unique attributes refer to those attributes associated with a specific product or brand when that product or brand is compared with that of competitors (Toral et al., 2017). In the case of innovations, such unique attributes represent potential innovations not available in competing products. Generally, attributes in text mining are obtained by using the terms with higher TF-IDF values. Although the TF-IF value is useful for discriminating topics of discussion, it does not specifically distinguish the uniqueness of attributes when the number of classes is known a priori. However, the TF-IDF value emphasizes those words that are locally frequent, so they have a better chance to be discriminative words among the set of predefined classes. Therefore, we hypothesize the following:

H_1. Unique attributes are more likely to be found among the terms with the highest TF-IDF value.

Moreover, as the TF-IDF value decreases, the chances of being unique diminishes because lower TF-IDF values indicate that terms are more globally frequent, thereby reducing their discriminative power (Erra et al., 2015). It should be noticed that this value represents the inverse document frequency; therefore, as the terms are used in more documents, its value drops down non-linearly with the document frequency (with a factor given by $1/DF$, where DF represents the document frequency). Consequently, it is expected that the chances of obtaining unique attributes from the list of terms with the highest TF-IDF value also diminishes non-linearly. Hence, we hypothesize the following:

H_2. The number of unique attributes has a non-linear relationship with the number of attributes obtained based on their TF/IDF value.

4. Research methodology

The methodology for testing the previous hypotheses is depicted in

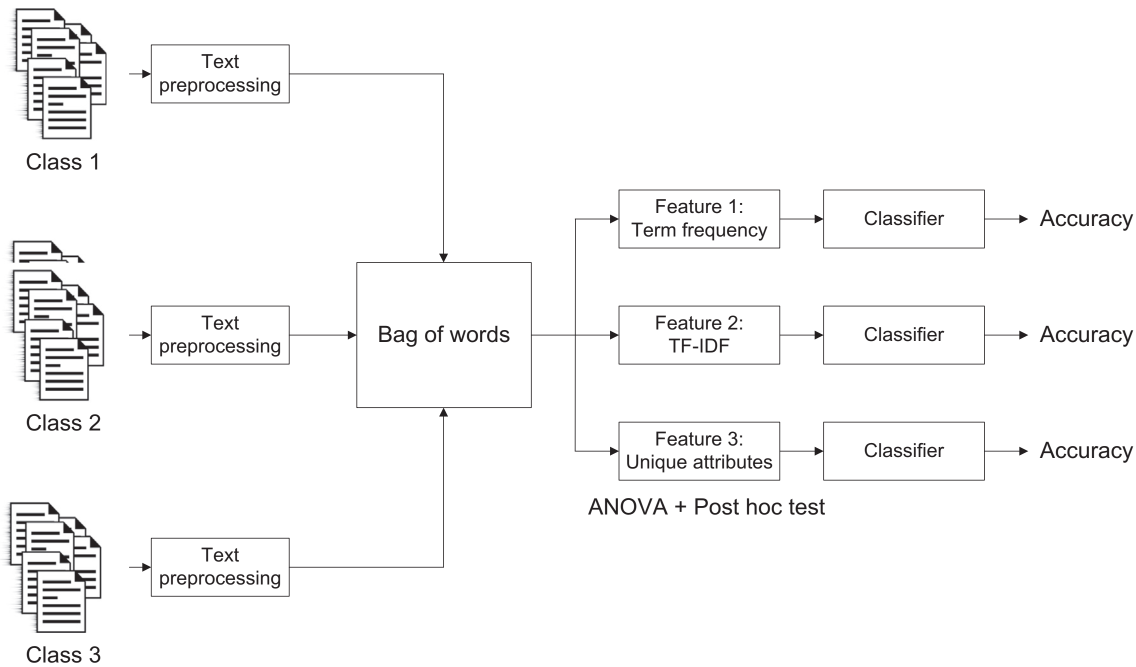


Fig. 2. Research methodology.

Fig. 2. For the sake of simplicity, this figure only shows the case of three classes, although the methodology could be extended to any number of classes.

Each class represent a homogeneous set of documents since each class contains documents addressing some specific subject. Text preprocessing consists of removing punctuation and stop words and conducting a lower-case conversion. Then, a bag of words is obtained as the features representative of the corpus of documents. Finally, three different approaches are used to classify the three predefined classes. The first one consists of using the term frequency, which basically is the number of times each word from the bag of words appears in each document. The TF-IDF (Term Frequency – Inverse Document Frequency) is a different metric that emphasizes important words, which means those words that appear frequently in some documents but rarely in corpus. The aim is to achieve a good balance between frequency and global rarity, avoiding those words that are so common that they are meaningless for distinguishing among classes. Finally, a new metric is proposed: the unique attributes, which emphasize the important words but emphasize them separately within each class. The two prior metrics have been frequently used in the literature. The latter is a novel contribution of this paper and consists of an ANOVA followed by a post hoc test. Whenever there are more than two groups, a significant ANOVA does not tell which groups are different from the others. Therefore, it is necessary to conduct post hoc paired comparisons to prevent excessive type 1 error. Among the different post hoc tests, Tukey's Honestly Significant Difference Test was selected. Only those terms where the null hypothesis is rejected in all the paired comparisons are selected as unique attributes. Hence, the unique attributes have the property of being important words for one out of the n classes of the study.

5. Case study and data collection

The case study is based on an online community, which constitutes a network of users in any virtual social space where they communicate with one another and make contributions through a discussion-thread structure (Lee et al., 2002; Preece, 2001; Williams and Cothrel, 2000). A user writes a post about a topic or writes a question, and then other users reply to either take part in the discussion or to answer the question written in the original post (Zhang et al., 2007). The contents or the topics discussed in the online community are driven by the participants (Lee et al., 2002). Online communities have become important places for people to seek and share expertise since they provide individuals with alternative methods for common interest information sharing and problem solving (Andrews, 2002; Zhang et al., 2007).

Data collection for this study has involved accessing data from the website SmartThings Community, which is an open innovation online community, where there are public forums for learning, helping, and sharing experiences through discussions threads that are categorized as SmartThings, SmartApps, the Internet of Things, and home automation. For this paper, we chose this third-party open innovation community because it provides a way to discuss a group of alternative products that are categorized by the SmartThings Community. Moreover, it is possible to use SmartThings to ask other users for help with solving business problems that a user is facing. Accordingly, the users can submit their ideas and suggested solutions to the problems defined to the community and can potentially receive a reward for their participation. Thus, our analysis is based on a rich data set of user reviews posted for innovations concerning not only a single product but also concerning several products.

The website SmartThings Community belongs to the technology company SmartThings that was bought by Samsung in August 2014. SmartThings' primary products include a free SmartThings app, a SmartThings Hub, as well as various sensors and smart devices. To use the services provided by the SmartThings community, the customers are required to sign up free of charge for an account in the website, provide a verifiable email address, provide their name, and select a password or PIN number as well as a user name. To create a topic, the users fill in some fields, such as the title of the topic and the user's thoughts about a problem, a product or maybe a solution to some issue. The standard topic also has a proper category and subcategory. This is actually the target data to be collected in this paper. Thus, taking into account these data of interest, this part of the web was crawled with an API (Application Programming Interface). To store all the data gathered online, a small relational database was designed, comprised of three tables: user, topic and reply. This provided the possibility of tracking everything stored inside and filtering the specific data needed for this study through SQL queries. The relational model is depicted in Fig. 3,

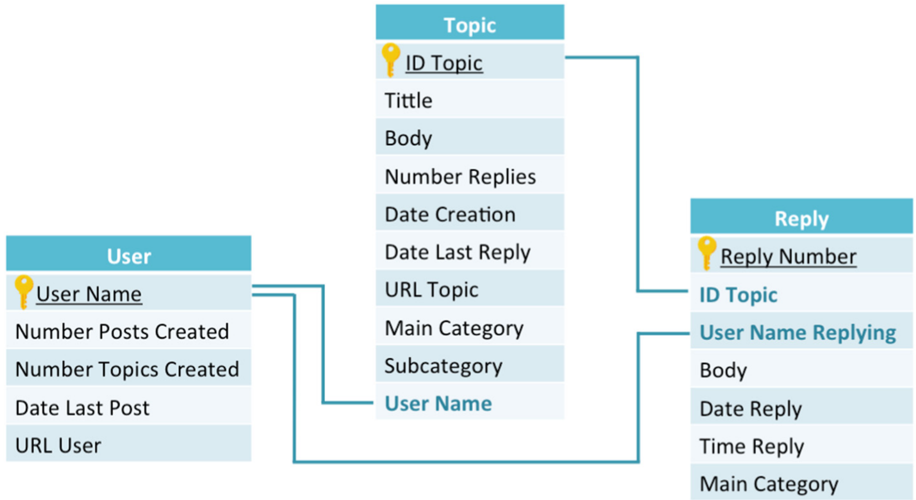


Fig. 3. Relational model of the gathered data.

which illustrates the database model based on all the gathered data with a representation in terms of tuples, grouped into relations.

This study is specifically focused on the Mobile Apps category, where customers discuss the various SmartThings mobile apps available at the following three main mobile platforms: the App Store, Google Play and the Windows Store. This category is subdivided into five subcategories: (1) mobile tips and tricks, (2) features and feedback, (3) iOS, (4) Android and (5) Windows. For the purpose of this study, the content of the subcategories of the main mobile platforms have been selected and analysed: the body (text) information has been stored in the tables *topic* and *reply* of the Mobile Apps subcategories iOS, Android and Windows.

6. Results

For this study, a sample of 336 reviews has been collected from the three Mobile App subcategories (Windows, Android and iOS). Although the subcategory Android as well as the subcategory iOS has more than 1500 texts (reviews), a sub-sample of the 336 reviews has been selected, so that all classes are balanced; otherwise, the classifiers could work poorly.

First, the proposed methodology was applied to obtain a set of unique attributes, considering different sizes of the bag of words. Fig. 4 shows the evolution of unique attributes as the size of the bag of words increases in steps of 100. These words were selected based on their TF/IDF value. The shape of the curve shows that the unique attributes are mainly found among those terms with the highest TF-IDF, as stated by H1. As the size of the bag of words increases, the term added has a lower TF/IDF value, and the number of unique attributes finally saturates. Moreover, Fig. 4 shows that the number of unique attributes grows faster at the beginning and then reaches a maximum value of 135. Therefore, there is a non-linear relationship between the number of attributes and the number of unique attributes, as stated by H2.

Table 1 details the unique attributes obtained for the three operating systems. For each unique attribute, Table 1 lists the corresponding operating system obtained by applying the ANOVA and the multi-comparison test, and the word count in each of the three classes represented by the three operating systems. It should be noticed that a term can be a unique attribute because it appears either many times or rarely in a given class. Although most of the terms in Table 1 are assigned as unique attributes of an operating system because their high frequency of appearance in a specific class, some of them are unique due to their rarity. Those terms are marked in Table 1 in blue with an (*), and they are associated with the Windows operating system. An inspection of these words (old, change, monitor, discussion, close) reveals that they are strong aspects of Windows (counted negatively)

Fig. 4. Variation of unique attributes as the Bag of Words increases.

Table 1
Unique attributes corresponding to the dimensions of the three subcategories (Android/iOS/Windows).

Word	OS	And/iOS/Win	Word	OS	And/iOS/Win	Word	OS	And/iOS/Win
Location	Android	57/20/18	Screen	iOS	19/60/25	App	Windows	254/318/591
View	Android	32/7/12	Widget	iOS	29/58/0	Phone	Windows	100/58/463
Battery	Android	32/5/0	Notification	iOS	6/54/3	Fix	Windows	27/12/214
Text	Android	29/6/3	iPhone	iOS	6/52/17	Issues	Windows	57/81/182
Used	Android	28/12/11	Section	iOS	8/42/3	Things	Windows	76/63/165
Nexus	Android	24/0/1	Feature	iOS	10/37/7	Support	Windows	48/48/104
Play	Android	20/1/2	Watch	iOS	1/36/1	Platform	Windows	9/7/58
Marshmallow	Android	20/0/0	Logging	iOS	3/32/3	Users	Windows	15/8/42
Screenshot	Android	19/0/0	Apple	iOS	1/29/1	Run	Windows	4/11/33
Original	Android	19/0/0	Center	iOS	5/28/0	Universal	Windows	1/1/28
Custom	Android	19/5/1	Request	iOS	7/26/1	Lumia	Windows	0/0/24
Icons	Android	18/5/2	Log	iOS	8/21/6	Weeks	Windows	6/6/21
Tasker	Android	17/2/0	Sunny	iOS	0/19/0	UWP	Windows	0/0/18
Google	Android	15/2/2	Discovery	iOS	0/12/0	PC	Windows	0/1/18
Scaling	Android	12/1/0	Steps	iOS	1/11/2	Dashboard*	Windows	57/75/17
Broke	Android	12/2/1	Slowness	iOS	0/11/0	Pin	Windows	0/3/16
Test	Android	11/2/2	Stay	iOS	2/9/2	Guid	Windows	0/0/10
Router	Android	11/0/0	Clicking	iOS	0/7/0	Cortana	Windows	0/0/10
Events	Android	10/0/1	Terrible	iOS	0/6/0	Xbox	Windows	0/0/9
Opening	Android	9/2/0	Perfect	iOS	0/5/0	Microsoft	Windows	2/0/9
Sharptools	Android	8/0/0	Glitches	iOS	0/4/0	Emulator	Windows	1/0/9
Cache	Android	8/0/0	Glimpse	iOS	0/4/0	Desktop	Windows	0/0/9
Types	Android	7/1/1	Downgrade	iOS	0/4/0	Unusable	Windows	1/0/8
Rooted	Android	7/0/0	Disable	iOS	0/4/0	Developing	Windows	0/0/8
Model	Android	7/1/1	–	–	–	Win10	Windows	0/0/7
S6	Android	6/0/0	–	–	–	Usable	Windows	1/0/7
Permission	Android	6/0/0	–	–	–	Old*	Windows	19/20/5
Messed	Android	6/0/0	–	–	–	Identified	Windows	0/0/5
Destroys	Android	6/0/0	–	–	–	Change*	Windows	40/29/5
Sms	Android	4/0/0	–	–	–	Navigate	Windows	0/0/4
Rendering	Android	4/0/0	–	–	–	Monitor*	Windows	28/24/3
Reducing	Android	4/0/0	–	–	–	Discussion*	Windows	72/61/3
Plugged	Android	4/0/0	–	–	–	Close*	Windows	11/10/0
Grid	Android	4/0/0	–	–	–	–	–	–
Cleared	Android	4/0/0	–	–	–	–	–	–

(*) Terms associated with a specific operating system due to their rarity.

because they are significantly different from words associated with iOS or Android.

To facilitate the interpretation of results, a correspondence analysis was performed to visualize the association between the terms and the operating systems (Greenacre, 2007). It mainly utilizes the coordinates on the bi-plot, which is the basic outcome of this analysis, showing the correspondence between the items of the unique attributes and the operating systems, according to their distance to each other. Fig. 5 illustrates the obtained result. As expected, the three operating systems are clearly separated, and surrounding each operating system, a set of unique attributes is displayed. The attributes highlighted in dark blue are those having the highest word count—up to 10—and those in light blue are appearing less. The unique attributes in green represent those unique attributes from Windows that are significantly different from those of iOS or Android and count negatively.

To test how representative the unique attributes of the three classes are, a classifier was trained by using 80% of the original sample and the three types of features: term frequency, TF/IDF and unique attributes. A K-nearest neighbour (kNN) classification approach was used because it has been widely used in various types of classification tasks related to text mining (Govindarajan and Chandrasekaran, 2010; Wan et al., 2012). A *k* value of 10 was considered, and the F-score was calculated as the performance metric of the classifier by using the test sample (the remainder 20% of the original sample).

Fig. 6 compares the performance of the three classifiers. Clearly, the features consisting of the unique attributes achieve better accuracy than the other two types of features but by using a smaller set of words. The numerical results are shown in Table 2. For instance, with a bag of words of size 1000, the TF and the TF/IDF features achieve an F-score of 0.70 and 0.75, respectively, while with only 113 unique attributes, the F-score reaches a 0.78 value. As the size of the bag of words increases, the number of unique attributes reaches a maximum value of 135, but still the accuracy of the unique attributes is better than that obtained from the TF and the TF/IDF value. Therefore, and answering RQ_1, we conclude that it is possible to distinguish posted innovations belonging to several choices by using the set of unique attributes.

Table 3 details an effective classification of the obtained unique attributes for the different classes (Android, iOS, Windows). This classification has been done according to their relation to the SmartThings App or to the operating system. On one hand, attributes classified as "Mobile Phone/OS Features" can be objectively associated with each operating system and have to do with the functioning and performance of the operating system itself or with the mobile phone. This means that they are attributes that can only belong to one operating system or mobile phone and are therefore not available from competitors. For instance, unique attributes, such as Nexus, Marshmallow or Google, can only be associated with Android, while the unique attributes of the iPhone and Apple can only be associated with iOS or the unique attributes of Microsoft, Lumia and Xbox are uniquely associated with Windows. On the other hand, there are unique attributes that are not intrinsically connected with Android, iOS or Windows but are associated with the SmartThings App. This second set of attributes can in turn be divided into two categories: (1) the SmartThings App's features and (2) the SmartThings App's issues or improvements.

The first category—SmartThings App features—refers to the App's technical characteristics and functionality that are uniquely associated with each of the operating systems. For example, the unique attribute *location* could be related to all operating systems; however, it only appears when associated with Android. Fig. 7 illustrates an extract of several reviews where location always appears related to mobile phones

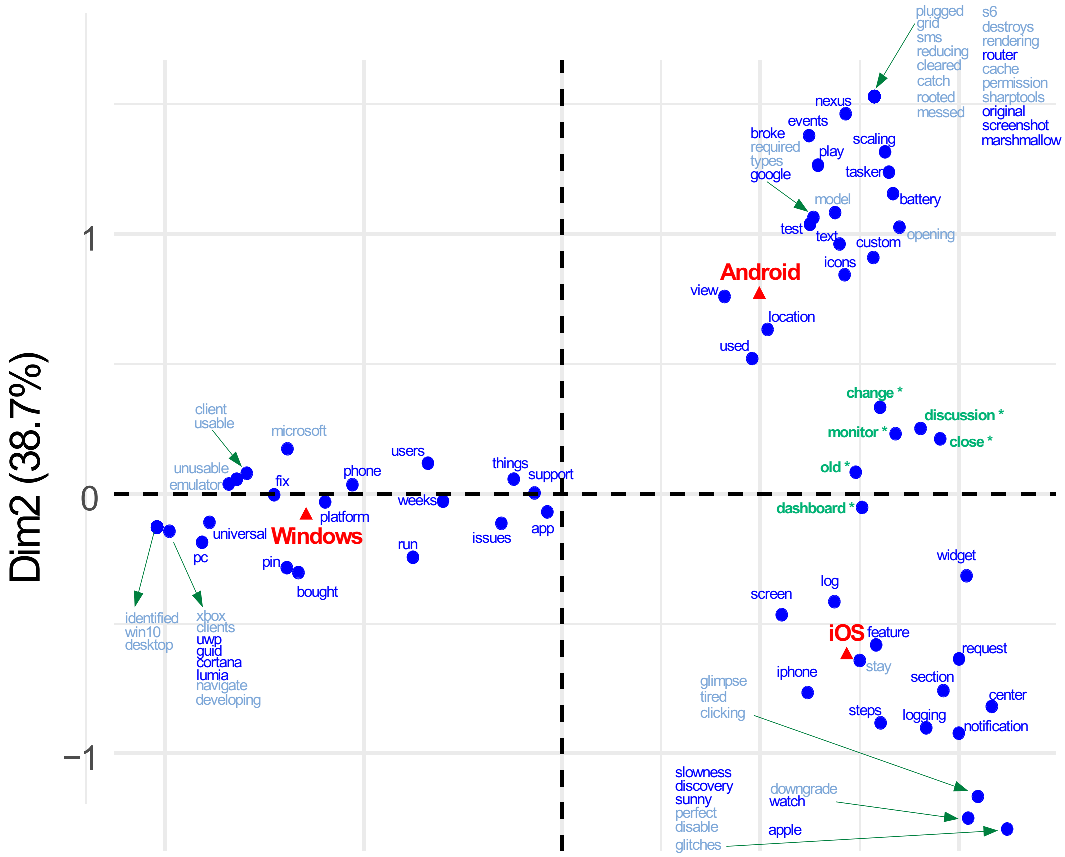


Fig. 5. Map of the unique attributes corresponding to the three dimensions (iOS, Android and Windows).

that only work with Android (Samsung Galaxy, Nexus 6, Galaxy S5…) or Android features (Marshmallow, Android ST…). Similarly, the unique attribute *widget* (see Fig. 8) appears exclusively in relation to iOS. Among these extracts of the online reviews, *widget* appears in threads that have to do indisputably with iOS (e.g., Apple watch or the iOS widget dashboard). The extracts of the Windows online reviews (see Fig. 9) reveal that the unique attribute *app* is exclusively associated with Windows through some of the users' comments comparing their Windows app issues with those of other operating systems: "*I tried it in my wife's Android phone and it works ok. I hope it can be fixed soon*" or "*when can we Windows Phone users expect a new version of the app, to match the latest iOS and Android release*?" Moreover, *app* appears among the reviews that are related to features uniquely offered by Windows (e.g., Windows Phone, PC, Microsoft store).

The second category—SmartThings App issues or improvements—refers to the users experience and focuses on the problems they experience when using the app or the solutions they give to other users having the same problem in a particular operating system. Fig. 10 shows that the unique attributes mentioned in the online reviews are exclusively associated with one of the operating systems. For example, the unique attribute *battery* is found along the Android thread, and it is possible to distinguish among the reviews other words that are intrinsically related to Android, such as Marshmallow or Android App. The unique attribute *logging* is related to issues of the new app versions launched just for iOS. Similarly, the unique attribute *universal* is enclosed in the thread for the operating system Windows. Here, users are demanding some improvements for the Windows SmartThings app by asking for a universal app compatible with Windows. Note that among these reviews, there are also words intrinsically related to Windows, such as Xbox, PC or Windows 10.

By taking a deeper look into the online reviews written by the users, it is possible to distinguish why SmartThings can also be considered an open innovation community. Among the online reviews from which unique attributes are classified as *SmartThings App issues or improvements*, there is collaboration taking place between users of the SmartThings open community. For instance, Fig. 11 shows an online review of a user who is responding to another user's Android app location issue and giving the user three possible solutions. Similarly, Fig. 12 details some extracts of the user's suggestions on how to improve the text (unique attribute from Android see Table 3) in the *Things Tiles View* of the Android SmartThings app.

The example displayed in Fig. 13 shows the users' reviews gathered from the iOS thread and in which the users request improvements for the Apple Watch, the Smart Home Monitor or the Notification Center Widget of SmartThings. Similarly, from the Windows thread, Fig. 14 shows a user online review suggesting the development of an app for Windows 10 and Xbox.

From the online reviews that include the unique attribute *sharptools*, which is exclusively associated with Android, it is possible to discern users who have criticized the SmartThings mobile application's built-in presence feature for randomly indicating that a person is no longer present in their home (when they are sleeping or sitting on the couch). To solve this issue, one SmartThings community user (see an example of his online review in Fig. 15) has provided an accurate and effective custom presence solution for SmartThings by developing the app *SharpTools* for the community. This application supports the following features: control of SmartThings automated home devices; control of a device through a simple interface; and the ability to use device commands that are automatically pulled from the SmartThings API (Application Programming Interface) or to send standard commands such as on and off.

This user identified a need expressed in the online users reviews within the online community and provided the community a solution, the app *SharpTools*. Therefore, in answering RQ_2, we found that among

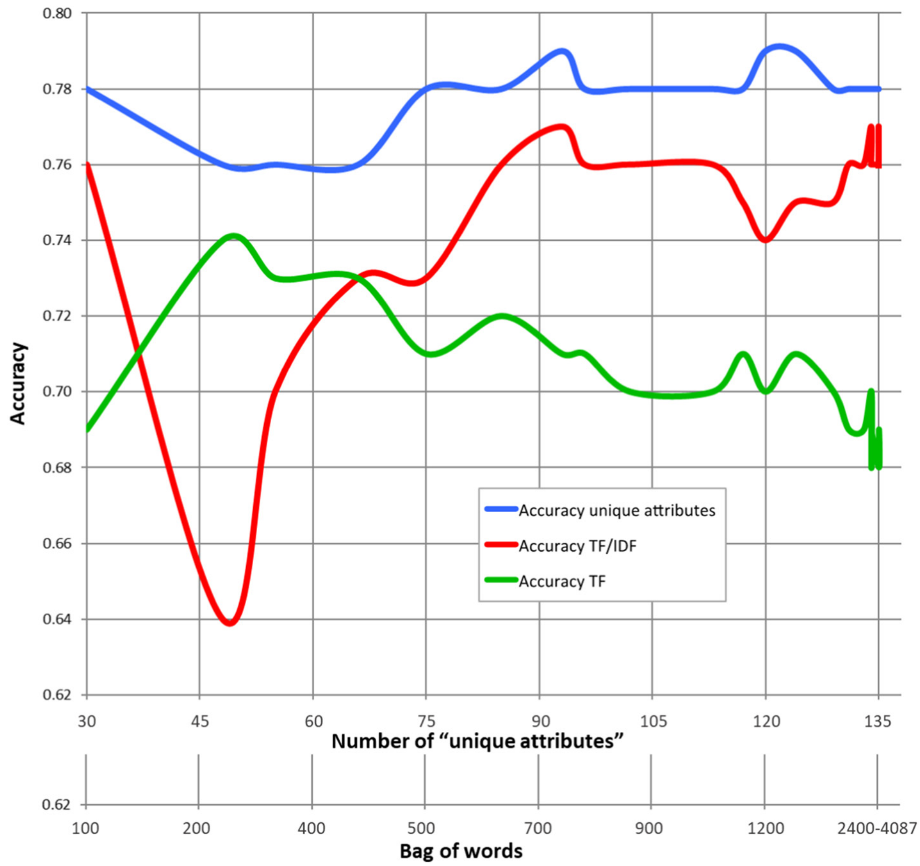


Fig. 6. F-score value for the three types of features considered with a 10-NN classifier.

the gathered online reviews of the open innovation community, it is possible to differentiate the attributes that are uniquely associated with each operating system.

7. Discussion and implications

In this paper, we contribute to the growing research on open innovation. More specifically, we use the large amount of available user-generated content to automatically extract specific relevant information about products, thereby offering a way for reducing the cost and time needed to analyse customer feedback and expanding the tools for practitioners that are seeking to understand customer feedback.

The value of user-generated information and customer feedback is recognized in the open innovation literature and in user innovation research. However, the extent to which such information is used must always be in proportion to the costs incurred. The existing methods for automated text mining already allow the fast completion of categorization and extraction of topics, sentiment analysis, and the summarization of documents. In product development, however, it is precisely the analysis of unique characteristics (or problems) of individual

Table 2
F-score results of the three classifiers for the three types of features.

Bag of words	Unique attr.	F-score TF	F-score TF/IDF	F-score unique	Bag of words	Unique attr.	F-score TF	F-score TF/IDF	F-score unique
100	30	0,69	0,64	0,76	2100	134	0,69	0,76	0,78
200	48	0,74	0,70	0,76	2200	134	0,70	0,76	0,78
300	55	0,73	0,73	0,76	2300	134	0,69	0,76	0,78
400	66	0,73	0,73	0,78	2400	135	0,68	0,76	0,78
500	75	0,71	0,76	0,78	2500	135	0,69	0,76	0,78
600	85	0,72	0,77	0,79	2600	135	0,68	0,76	0,78
700	93	0,71	0,76	0,78	2700	135	0,69	0,77	0,78
800	96	0,71	0,76	0,78	2800	135	0,68	0,77	0,78
900	102	0,70	0,76	0,78	2900	135	0,69	0,77	0,78
1000	113	0,70	0,75	0,78	3000	135	0,69	0,76	0,78
1100	117	0,71	0,74	0,79	3100	135	0,69	0,76	0,78
1200	120	0,70	0,75	0,79	3200	135	0,69	0,76	0,78
1300	124	0,71	0,75	0,78	3300	135	0,69	0,76	0,78
1400	129	0,70	0,76	0,78	3400	135	0,69	0,77	0,78
1500	131	0,69	0,76	0,78	3500	135	0,69	0,76	0,78
1600	133	0,69	0,77	0,78	3600	135	0,69	0,76	0,78
1700	134	0,70	0,76	0,78	3700	135	0,69	0,77	0,78
1800	134	0,68	0,76	0,78	3800	135	0,69	0,77	0,78
1900	134	0,69	0,76	0,78	3900	135	0,69	0,77	0,78
2000	134	0,69	0,76	0,78	4000	135	0,69	0,76	0,78

Table 3
Classification of unique attributes conforming to the three classes (Android, iOS and Windows)

Android			iOS			Windows		
SmartThings App features	Mobile phone/OS features	SmartThings App issues/ improvements	SmartThings App features	Mobile phone/OS features	SmartThings App issues/ improvements	SmartThings App features	Mobile phone/OS features	SmartThings App issues/improvements
Location	Nexus	Battery	Screen	iPhone	Feature	App	Phone	Fix
View	Play	Used	Widget	Watch	Logging	Things	Platform	Issues
Text	Marshmallow	Screenshot	Notification	Apple	Request	Dashboard*	Users	Support
Custom	Tasker	Original	Center	–	Log	Monitor*	Lumia	Run
Icons	Google	Scaling	Glimpse	–	Sunny	–	UWP	Universal
Sharptools	Rooted	Broke	–	–	Discovery	–	PC	Weeks
Grid	Model	Test	–	–	Steps	–	Guid	Pin
–	s6	Router	–	–	Slowness	–	Cortana	Unusable
–	SMS	Events	–	–	Stay	–	Xbox	Developing
–	–	Opening	–	–	Clicking	–	Microsoft	Usable
–	–	Cache	–	–	Terrible	–	Emulator	Old*
–	–	Types	–	–	Perfect	–	Desktop	Identified
–	–	Permission	–	–	Glitches	–	Win10	Change*
–	–	Messed	–	–	Downgrade	–	–	Navigate
–	–	Destroys	–	–	Disable	–	–	Discussion*
–	–	Rendering	–	–	–	–	–	Close*
–	–	Plugged	–	–	–	–	–	–
–	–	Cleared	–	–	–	–	–	–

(*) Terms associated with a specific operating system due to their rarity.

products or product groups that is important information for managers. For instance, topic-modelling approaches were used to identify user preferences from the MyStarbucksIdea website (Martínez-Torres et al., 2015). However, product innovation is the result of matching customer need with unique technology solutions, that is, for a given situation, determining the unique technology that enables the desired outcome and the satisfaction of a customer need (Fetterhoff and Voelkel, 2006). In this regard, our study has found evidence to identify the unique attributes associated with one choice in an open innovation community. The unique attributes' accuracy obtained by applying the ANOVA and the multicomparison test was better than that obtained from the application of the TF and the TF/IDF values that have been used in the existing literature to determine word relevance (Hu et al., 2012; Lee et al., 2018). Consequently, our study goes one step ahead of the previous studies focusing only on customer needs.

To illustrate the use of the CODA method, we have examined a set of user discussions on a Samsung open innovation platform for their IoT product line SmartThings. Specifically, we focused on the discussions of the company's mobile SmartThings app and grouped them according to the three different operating systems for which the app was developed. The results of this study reveal that the identified unique attributes can shed light on specific and exclusive differences between the operating systems. For instance, problems that were specific to Android are shown in discussions on several issues regarding *location* functionality or *scaling* associated with the Android implementation. Key features were identified for iOS, where the terms *widget* and *watch* were brought up in discussions and feature requests related to the unique functionality of iOS for the Apple Watch were posted.

As described in the lead user theory, the users in the investigated community not only expressed specific problems but also suggested suitable solutions (see Fig. 11, for example). In another product category, the type of information provided could be used to determine which product characteristics are perceived as unique within a category by users. On the basis of this information, practitioners can decide whether the feature can be used for communication, whether it should be strengthened at the product level, or whether it is associated with an unfavourable perception that must be restricted in future versions or in communication.

In this paper, we also contribute to the research in the field of crowdsourcing where the aggregation of distributed information is needed to derive meaningful conclusions from a large body of information (Surowiecki, 2004). In the context of quantitative data, it has enabled prediction markets (Plott and Chen, 2002) or in the context of qualitative data sentiment analysis, (Younis, 2015), it builds aggregation. To this field, for qualitative data, we contribute an aggregation mechanism that can be used to extract unique topics from different groups of content.

From a theoretical perspective, the study advances the company's justification of seeking different customization solutions. While researchers in the marketing domain focus on the value of product differentiation and are exploring new ways to achieve differentiation at the level of a product (Luchs et al., 2016) or at the level of

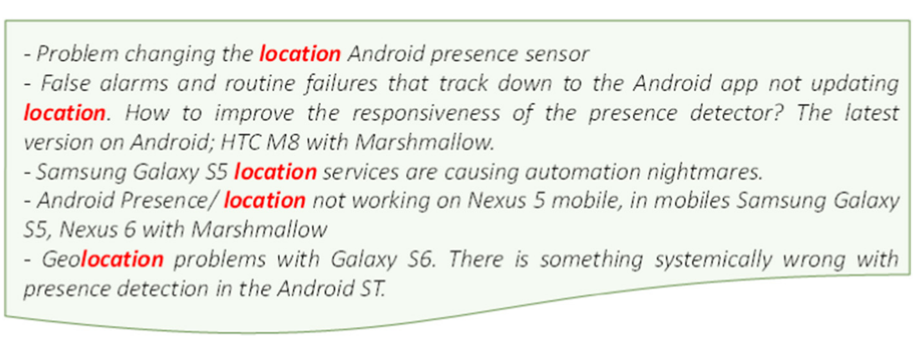
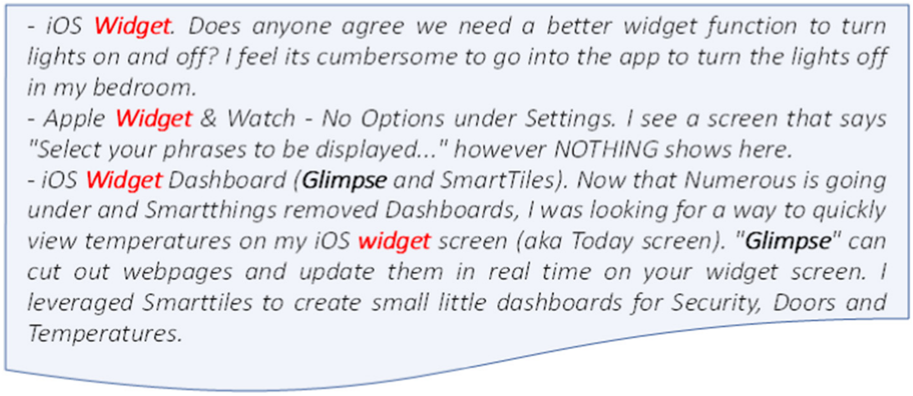
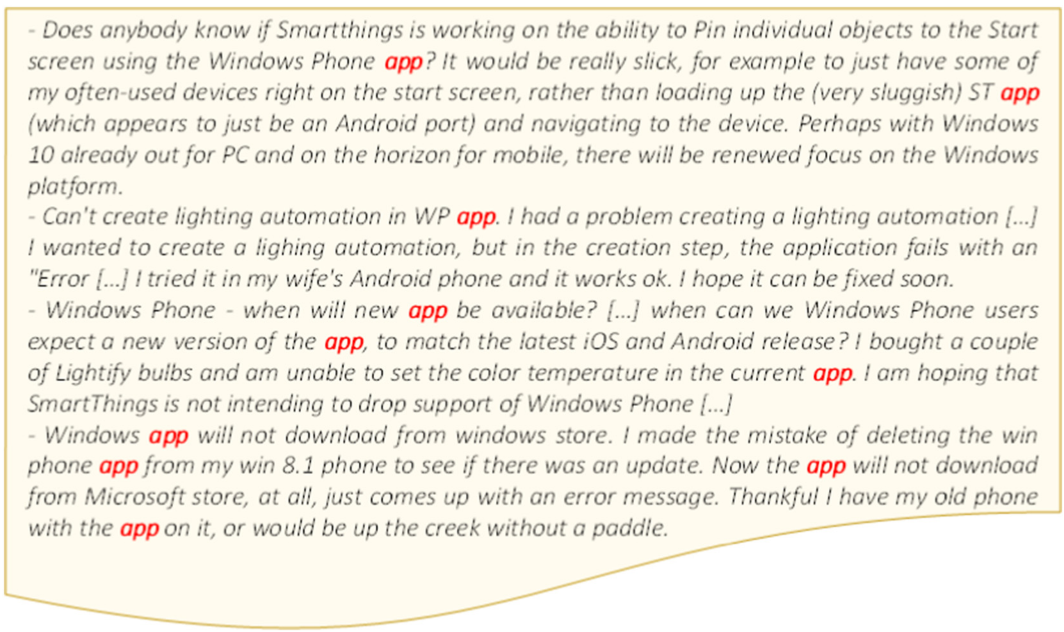


Fig. 7. For the Mobile Apps subcategory Android, examples of online reviews extracts highlighting the unique attribute “location”.



- iOS **Widget**. Does anyone agree we need a better widget function to turn lights on and off? I feel its cumbersome to go into the app to turn the lights off in my bedroom.
- Apple **Widget** & Watch - No Options under Settings. I see a screen that says "Select your phrases to be displayed..." however NOTHING shows here.
- iOS **Widget** Dashboard (**Glimpse** and SmartTiles). Now that Numerous is going under and Smartthings removed Dashboards, I was looking for a way to quickly view temperatures on my iOS **widget** screen (aka Today screen). "**Glimpse**" can cut out webpages and update them in real time on your widget screen. I leveraged Smarttiles to create small little dashboards for Security, Doors and Temperatures.

Fig. 8. For the Mobile Apps subcategory iOS, examples of online reviews extracts highlighting the unique attribute "widget".



- Does anybody know if Smartthings is working on the ability to Pin individual objects to the Start screen using the Windows Phone **app**? It would be really slick, for example to just have some of my often-used devices right on the start screen, rather than loading up the (very sluggish) ST **app** (which appears to just be an Android port) and navigating to the device. Perhaps with Windows 10 already out for PC and on the horizon for mobile, there will be renewed focus on the Windows platform.
- Can't create lighting automation in WP **app**. I had a problem creating a lighting automation [...] I wanted to create a lighing automation, but in the creation step, the application fails with an "Error [...] I tried it in my wife's Android phone and it works ok. I hope it can be fixed soon.
- Windows Phone - when will new **app** be available? [...] when can we Windows Phone users expect a new version of the **app**, to match the latest iOS and Android release? I bought a couple of Lightify bulbs and am unable to set the color temperature in the current **app**. I am hoping that SmartThings is not intending to drop support of Windows Phone [...]
- Windows **app** will not download from windows store. I made the mistake of deleting the win phone **app** from my win 8.1 phone to see if there was an update. Now the **app** will not download from Microsoft store, at all, just comes up with an error message. Thankful I have my old phone with the **app** on it, or would be up the creek without a paddle.

Fig. 9. For the Mobile Apps subcategory Windows, examples of online extracts of reviews highlighting the unique attribute "app".

customization for the customer (Piller et al., 2005), co-occurrence differential analysis (CODA) could complement the internal activities of a firm and suggest the identification of product differentiation features derived from an aggregated perspective of the customer. It can be used as a quantitative approach to document actual differentiation and to support the identification of truly differentiating features that are derived from customer experience and not based on the assumptions of marketing professionals. On the operational level, CODA relies on a set of unique attributes that have been demonstrated to be the best subset of predictors of given classes, despite being much lower in number than the attributes selected based on their TF or their TF-IDF value. This supports the first hypothesis (H1) revealing that the unique attributes are more likely to be found among those words showing a higher TF-IDF value.

Our approach has important practical implications for practitioners and entrepreneurs: unique attributes lead to the identification of unique topics that represent a competitive advantage over other choices or alternatives. In this case, the selection of the operating system can drive some innovations that cannot be achieved by using other operating systems. Therefore, our approach can help managers and entrepreneurs to differentiate their product from those of their competitors by implementing obtained unique innovations. Moreover, the distinction of unique innovations can also be used within a firm to obtain the complementarity of their product lines, so the weaknesses of some products can be reinforced by other complements (i.e., new batteries) to extend the autonomy of smart apps. From a methodological viewpoint, our approach also facilitates the realization of prospective studies through open-ended questions.

Our paper also has direct implications for market research. Market research has traditionally fulfilled the role of gathering relevant insights from stakeholders, especially consumers, and reporting those insights to marketing and product development. The professionalization of this function led to a streamlining and industrialization of market research (Cooke and Buckley, 2008). This approach is flawed for several reasons, and polling specifically has faced harsh criticism after producing consistently flawed results in the last US election (Skibba, 2016). Traditional market research techniques have only managed to skim the surface of user needs (Kristensson et al., 2004); therefore, now, it is increasingly used to harness relevant data generated by distributed consumers (Cooke and Buckley, 2008). Additionally, doing market research through gathering web content is non-intrusive and avoids direct interactions with subjects, as direct interaction can always create biases on the side of the respondents (Lu and Stepchenkova, 2015). However, dealing with user-generated content also means dealing with a large amount of data, which presents its own challenges regarding the collection, organization and analysis of such material in a quantifiable and

Android Threat

- Android App destroys **battery** and data - 13 gigs used in 15 days.
- Much higher **battery** drain and network usage using the new app (Android v 2.0)
- App crashes at launch on Android 5.1.1, 5.0.3, and 5.0 and it rapidly drains the **battery**.
- Android App 2.1.1 slightly reduced **battery** life
- Android Widget crushed under update Marshmallow. **Battery** going down quicker during the day

iOS Threat

- iOS app **logging** out consistently
- Ever since release 2.0.6 of the app I have had continuing problems with the activity log. Sometimes it's slow to refresh, no refresh at all or most times it's completely blank. Was hoping for some fixes with 2.0.7 release but no such luck. The constant **logging** in is getting old as well.
- I can't even log in to the 2.0.1 app. Deleted/reinstalled. Reset my password. Confirmed I know what my password is by **logging** in to the IDE. No joy. This update just keeps getting better and better.
- Try going to the Logs section of the IDE to start Live **Logging** and try changing the color of one of them to report any errors. Also, the on/off and dimming still works?

Windows Threat

- Also curious/looking towards support for this. I loaded the SmartThings app APK onto my Lumia 930 and it performed amazingly, save for services dependent on Google APIs such as GPS/Maps. Honestly, I would like to see a **universal** app that way they can support HoloLens, Xbox One, Windows 10 Mobile, and Windows 10. The incentive is there and I wonder if SmartThings will take advantage of it. P.S. If you (SmartThings) need beta testers for your Windows Phone or Windows 10 (fingers crossed) app, I would be more than happy to participated and provide feedback.
- I'm assuming at this point it would be a Windows 10 **universal** app. Would be nice to be able to use it on PC, tablet, and (hopefully) Xbox One.
- In all fairness, to follow your analogy, the Windows ecosystem is the entire Naval Fleet. Sure, Windows Phone is a small slice of the pie, but with the advent of **Universal** Apps for Windows, it would seem by your own argument that a decent Windows app for phone/tablet/laptop/PC/Xbox would be the way to go.
- One More Bump for a Windows **Universal** App, please!

Fig. 10. Online reviews extracts related to the unique attributes classified as "SmartThings App issues or improvements" for each operating system thread.

"I'm sitting literally 5 feet away from my wifi router and the SmartThings hub. But according to ST, I left 3 minutes ago The proximity to the hub doesn't make any difference with mobile presence as it is all about geo-location using cell towers, GPS, and wi-fi networks. I don't know what constitutes high accuracy settings but if you have wi-fi turned on it will be much better. If you have wi-fi on, I wonder if you are dropping from it for s split second from time to time. It has been talked about in many other threads but I see possible paths for improvement as 1) allowing presence based solely on wi-fi network 2) utilizing beacons or other BLU technology to determine location 3) wait for better geo-location mechanisms built into smart phones."

Fig. 11. For the Mobile Apps subcategory "Android", an example of an online review indicating an issue and a possible solution.

Threat: Better the original dashboard (Things Tile View).

User 1 *"[...] I like the idea of the text labels in the new app, but agree a quick view screen with more icons per page is needed."*

User 2 *"[...]Best of both worlds might be a screen that looks like the Routines screen where 6 icons can be displayed with small text describing each below. Being able to scroll up and down in this view would probably be very useful."*

User 3 *"[...] The old Text based UI of ST was better. If you make bad use of space for the sake of visual aesthetics, you will sacrifice real functionality."*

Fig. 12. Example of some users reviews suggesting improvements and included in the same thread in the Mobile Apps subcategory "Android".

time-efficient manner (Boyd and Crawford, 2012). To this end, many web extraction methods and machine-learning techniques have become popular during the last decade to answer such challenges (González-Rodríguez et al., 2016; Kim et al., 2017; Olmedilla et al., 2016a). Therefore, we propose to use further text mining methods and, in particular, to use co-occurrence differential analysis for the identification of differentiation characteristics and to further develop the method in the context of market research.

7.1. Limitations

Our approach is limited by shortcomings that are inherent to the method. Most prominently, we need to balance the number of documents from each class, as classifiers may suffer from biases towards the majority class when the assumption of balanced classes is not supported. In our case, this led us to omit a considerable amount of text from the biggest class. The approach is thus only viable in situations where the number of documents or reviews from each category is large

> **User 1** << I don't know about you, but I have many devices Zigbee, Z-Wave and WiFi and would love to see in the configuration for each a place to write down the type of battery it uses. As of now I have four different batteries. So it would be nice as my battery gets low if I look at the battery condition on the things screen it mentions battery type. I am not sure it could be done but I think it would be a nice addition to the app.>>
>
> **User 2** << A couple Apple Watch feature requests. Just tried the Apple Watch app for the first time, and it worked great. I'm impressed by how quickly the ST team put this together, so kudos! A couple requests to make this more useful: -Visual feedback to indicate if Hello, Home phrase was successfully triggered or not (similar to how the Widget will flash green or red). -Show me the current Mode at the top of the screen (would love if the iOS widget did this too!) Thanks! >>
>
> **User 3** << FEATURE Request: It would really be helpful to be able to swipe down the notification center to see the current Mode of the home and possibly a brief summary of the Smart Home Monitor status at the top of the SmartThings widget.>>
>
> **User 4** << Feature Request: iOS 8 Notification Center Widget. How about as a future improvement, a choice to put the lights and switches entries in the widget in addition to hello homes? >>

Fig. 13. For the Mobile Apps subcategory iOS, examples of user reviews requesting features for the SmartThings App

> In 8.1 "Help" Will give you the actions as well with Cortana. As far as a Beta Tester. Yea lets go. I would be even willing to help code the app. I got nothing better to do! I have seen the Live tile one time. it said something about my hub. Most of the time the Tile is a empty black tile though. Its really a bummer the windows phone is so far behind. I also agree a universal Windows 10 app would be killer! A Xbox One App? YES PLEASE! Who is are main developer in Windows anyways?

Fig. 14. For the Mobile Apps subcategory Windows, an example of an online review requesting some newness to SmartThings.

> I've been working on an Android app which provides widgets/shortcuts and a Tasker plugin for controlling SmartThings devices. The app is still in it's early stages, but my friends that have been testing it with me have been encouraging me to share the app with the SmartThings community. If you interested in alpha/beta testing the app, follow the instructions below: http://sharptools.boshdirect.com/installation-instructions I would note that while the widgets and Tasker plugin are free during the alpha testing period, I do plan on limiting the number of available widgets/Tasker tasks in the free version and having an IAP to unlock unlimited access once the app is released.

Fig. 15. For the Mobile Apps subcategory "Android", an example of an online review indicating an Open Innovation solution.

enough or is balanced. In addition, our approach is also applicable in those cases where clear differences among classes are expected. Finally, there is also a limitation in the number of classes considered. Finding unique attributes becomes harder as the number of classes increases. This happens because the condition of being unique requires that an attribute is significantly different in one class over the others.

7.2. Future research directions

The work presented in this paper can be extended by considering not only single words but also phrases that can help to contextualize the meaning of attributes. However, the main drawback of using phrases is that phrases make the matrices of vector space model even sparser, making the subsequent extraction of topic difficult.

Another possibility for contextualizing attributes consists of applying sentiment analysis, as it enables the determination of whether the attribute is mentioned in a positive, negative or neutral context. This may be interesting in studies such as ours, where people frequently report problems or app malfunctions.

8. Conclusion

We propose the use of text mining techniques in online user reviews to obtain unique attributes that can differentiate the optimal unique ideas or specific relevant information related to the main operating systems of the Mobile apps category within the open innovation community SmartThings. The results show that there are unique attributes intrinsically associated with each operating system and others referring to the technical characteristics and functionality of the app or the users' experiences when using the app. Hence, the proposed methodology can be used to identify the uniqueness in a growing volume of ideas generated by users within open innovation communities. Moreover, for marketers, the early prediction of certain ideas only associated with one choice among several alternatives can be useful for highlighting and finding relevant innovations and reporting them to marketing and product development before implementation. Additionally, this early prediction ability can be beneficial for researchers, who can document differentiation in a large number of products derived from user experience and not from the assumptions of marketers.

References

Allio, R.J., 2004. CEO interview: the InnoCentive model of open innovation. Strateg. Leadersh. 32 (4), 4–9.

Andrews, D.C., 2002. Audience-specific online community design. Commun. ACM 45 (4), 64–68.

Arenas-Márquez, F.J., Martínez-Torres, M.R., Toral, S.L., 2014. Electronic word-of-mouth communities from the perspective of social network analysis. Tech. Anal. Strat. Manag. 26 (8), 927–942.

Bashir, N., Papamichail, K.N., Malik, K., 2017. Use of social media applications for supporting new product development processes in multinational corporations. Technol. Forecast. Soc. Chang. 120, 176–183.

Bayus, B.L., 2013. Crowdsourcing new product ideas over time: an analysis of the Dell IdeaStorm community. Manag. Sci. 59 (1), 226–244.

Bogers, M., West, J., 2012. Managing distributed innovation: strategic utilization of open and user innovation. Creat. Innov. Manag. 21 (1), 61–75.

Boyd, D., Crawford, K., 2012. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf. Commun. Soc. 15 (5), 662–679.

Brem, A., Bilgram, V., 2015. The search for innovative partners in co-creation: identifying lead users in social media through netnography and crowdsourcing. J. Eng. Technol. Manag. 37, 40–51.

Bresciani, S., Ferraris, A., Del Giudice, M., 2018. The management of organizational ambidexterity through alliances in a new context of analysis: Internet of Things (IoT) smart city projects. Technol. Forecast. Soc. Chang. 136, 331–338.

Bullinger, A.C., Neyer, A.K., Rass, M., Moeslein, K.M., 2010. Community-based innovation contests: where competition meets cooperation. Creat. Innov. Manag. 19 (3), 290–303.

Chesbrough, H.W., 2003. Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business Press.

Chesbrough, H., Vanhaverbeke, W., West, J., 2013. Open Innovation Researching a New Paradigm. Oxford University Press, Oxford.

Christensen, K., Scholderer, J., Hersleth, S.A., Næs, T., Kvaal, K., Mollestad, T., Nina, V., Risvik, E., 2018. How good are ideas identified by an automatic idea detection system? Creat. Innov. Manag. 27 (1), 23–31.

Cooke, M., Buckley, N., 2008. Web 2.0, social networks and the future of market research. Int. J. Mark. Res. 50 (2), 267.

Cricelli, L., Greco, M., Grimaldi, M., 2016. Assessing the open innovation trends by means of the Eurostat Community Innovation Survey. Int. J. Innov. Manag. 20 (03), 1650039.

Dahlander, L., Piezunka, H., 2014. Open to suggestions: how organizations elicit suggestions through proactive and reactive attention. Res. Policy 43 (5), 812–827.

Dahlander, L., Wallin, M.W., 2006. A man on the inside?: unlocking communities as complementary assets. Res. Policy 35, 1243–1259.

Del Vecchio, P., Di Minin, A., Petruzzelli, A.M., Panniello, U., Pirri, S., 2018. Big data for open innovation in SMEs and large corporations: trends, opportunities, and challenges. Creat. Innov. Manag. 27 (1), 6–22.

Dezi, L., Del Giudice, M., 2014. Economic performance and capital structure choices: a nonparametric analysis of the southern Italian textile-clothing industry. Int. Stud. Manag. Organ. 44 (1), 42–54.

Di Gangi, P.M., Wasko, M., 2009. Steal my idea! Organizational adoption of user innovations from a user innovation community: a case study of Dell IdeaStorm. Decis. Support. Syst. 48 (1), 303–312.

Edgar, T.F., Pistikopoulos, E.N., 2018. Smart manufacturing and energy systems. Comput. Chem. Eng. 114, 130–144.

Erra, U., Senatore, S., Minnella, F., Caggianese, G., 2015. Approximate TF–IDF based on topic extraction from massive message stream using the GPU. Inf. Sci. 292, 143–161.

Ferraris, A., Santoro, G., Dezi, L., 2017. How MNC's subsidiaries may improve their innovative performance? The role of external sources and knowledge management capabilities. J. Knowl. Manag. 21 (3), 540–552.

Fetterhoff, T.J., Voelkel, D., 2006. Managing open innovation in biotechnology. Res. Technol. Manag. 49 (3), 14–18.

Fournier-Viger, P., Gomariz, A., Campos, M., Thomas, R., 2014. Fast vertical mining of sequential patterns using co-occurrence information. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, pp. 40–52.

Franke, N., Von Hippel, E., Schreier, M., 2006. Finding commercially attractive user innovations: a test of lead-user theory. J. Prod. Innov. Manag. 23 (4), 301–315.

Füller, J., Bartl, M., Ernst, H., Mühlbacher, H., 2006. Community based innovation: how to integrate members of virtual communities into new product development. Electron. Commer. Res. 6 (1), 57–73.

Füller, J., Matzler, K., Hoppe, M., 2008. Brand community members as a source of innovation. J. Prod. Innov. Manag. 25 (6), 608–619.

Gassmann, O., Enkel, E., Chesbrough, H., 2010. The future of open innovation. R&D Manag. 40 (3), 213–221.

George, G., Haas, M.R., Pentland, A., 2014. Big data and management. Acad. Manag. J. 57 (2), 321–326.

González-Rodríguez, M.R., Martínez-Torres, R., Toral, S., 2016. Post-visit and pre-visit tourist destination image through eWOM sentiment analysis and perceived helpfulness. Int. J. Contemp. Hosp. Manag. 28 (11), 2609–2627.

Govindarajan, M., Chandrasekaran, R.M., 2010. Evaluation of k-nearest neighbor classifier performance for direct marketing. Expert Syst. Appl. 37 (1), 253–258.

Greenacre, M., 2007. Correspondence Analysis in Practice. vol. 2 Chapman & Hall CRC, New York.

Gruner, K.E., Homburg, C., 2000. Does customer interaction enhance new product success? J. Bus. Res. 49 (1), 1–14.

Haefliger, S., Monteiro, E., Foray, D., Von Krogh, G., 2011. Social software and strategy. Long Range Plan. 44 (5–6), 297–316.

Hu, N., Bose, I., Koh, N.S., Liu, L., 2012. Manipulation of online reviews: an analysis of ratings, readability, and sentiments. Decis. Support. Syst. 52 (3), 674–684.

Hyysalo, S., Usenyuk, S., 2015. The user dominated technology era: dynamics of dispersed peer-innovation. Res. Policy 44 (3), 560–576.

Incki, K., Ari, I., 2018. Model-based runtime monitoring of smart city systems. Proc. Comput. Sci. 134, 75–82.

Jeong, D.H., Kwon, Y.I., 2014. Analysis on convergence in green technology field using patent information. In: Applied Mechanics and Materials. vol. 548. Trans Tech Publications, pp. 1981–1993.

Kim, K., Park, O.J., Yun, S., Yun, H., 2017. What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. Technol. Forecast. Soc. Chang. 123, 362–369.

Kortuem, G., Kawsar, F., 2010, May. User innovation for the internet of things. In: Proceedings of the Workshop What can the Internet of Things do for the Citizen (CIoT) at The Eighth International Conference on Pervasive Computing (Pervasive 2010), Helsinki, Finland.

Kristensson, P., Gustafsson, A., Archer, T., 2004. Harnessing the creative potential among users. J. Prod. Innov. Manag. 21 (1), 4–14.

Laursen, K., Salter, A., 2006. Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. Strateg. Manag. J. 27 (2), 131–150.

Lee, P.C., Su, H.N., 2010. Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. Innovation 12 (1), 26–40.

Lee, F.S., Vogel, D., Limayem, M., 2002, January. Virtual community informatics: What we know and what we need to know. In: HICSS. Proceedings of the 35th Annual Hawaii International Conference on System Sciences. IEEE, pp. 2863–2872.

Lee, S.M., Olson, D.L., Trimi, S., 2012. Co-innovation: convergenomics, collaboration, and co-creation for organizational values. Manag. Decis. 50 (5), 817–831.

Lee, H., Choi, K., Yoo, D., Suh, Y., Lee, S., He, G., 2018. Recommending valuable ideas in an open innovation community: a text mining approach to information overload problem. Ind. Manag. Data Syst. 118 (4), 683–699.

Lilien, G.L., Morrison, P.D., Searls, K., Sonnack, M., Hippel, E.V., 2002. Performance assessment of the lead user idea-generation process for new product development. Manag. Sci. 48 (8), 1042–1059.

Lu, W., Stepchenkova, S., 2015. User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software. J. Hosp. Mark. Manag. 24 (2), 119–154.

Luchs, M.G., Swan, K.S., Creusen, M.E., 2016. Perspective: a review of marketing research on product design with directions for future research. J. Prod. Innov. Manag. 33 (3), 320–341.

Mahr, D., Lievens, A., 2012. Virtual lead user communities: drivers of knowledge creation for innovation. Res. Policy 41 (1), 167–177.

Majchrzak, A., Malhotra, A., 2013. Towards an information systems perspective and research agenda on crowdsourcing for innovation. J. Strateg. Inf. Syst. 22 (4), 257–268.

Majchrzak, A., Cooper, L.P., Neece, O.E., 2004. Knowledge reuse for innovation. Manag. Sci. 50 (2), 174–188.

Malhotra, A., Majchrzak, A., 2014. Managing crowds in innovation challenges. Calif. Manag. Rev. 56 (4), 103–123.

Martínez-Torres, M.R., 2013. Application of evolutionary computation techniques for the identification of innovators in open innovation communities. Expert Syst. Appl. 40 (7), 2503–2510.

Martínez-Torres, M.R., 2014. Analysis of open innovation communities from the perspective of social network analysis. Tech. Anal. Strat. Manag. 26 (4), 435–451.

Martínez-Torres, M.R., 2015. Content analysis of open innovation communities using latent semantic indexing. Tech. Anal. Strat. Manag. 27 (7), 859–875.

Martínez-Torres, M.R., Olmedilla, M., 2016. Identification of innovation solvers in open innovation communities using swarm intelligence. Technol. Forecast. Soc. Chang. 109, 15–24.

Martínez-Torres, M.R., Rodríguez-Piñero, F., Toral, S.L., 2015. Customer preferences versus managerial decision-making in open innovation communities: the case of Starbucks. Tech. Anal. Strat. Manag. 27 (10), 1226–1238.

Morgan, J., Wang, R., 2010. Tournaments for ideas. Calif. Manag. Rev. 52 (2), 77–97.

Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M., 2017. Digital innovation management: reinventing innovation management research in a digital world. MIS Q. 41 (1).

Natalicchio, A., Ardito, L., Savino, T., Albino, V., 2017. Managing knowledge assets for open innovation: a systematic literature review. J. Knowl. Manag. 21 (6), 1362–1383.

Olmedilla, M., Martínez-Torres, M.R., Toral, S.L., 2016a. Harvesting big data in social science: a methodological approach for collecting online user-generated content. Computer Standards & Interfaces 46, 79–87.

Olmedilla, M., Martínez-Torres, M.R., Toral, S.L., 2016b. Examining the power-law distribution among eWOM communities: a characterisation approach of the long tail. Tech. Anal. Strat. Manag. 28 (5), 601–613.

Papa, A., Dezi, L., Gregori, G.L., Mueller, J., Miglietta, N., 2019a. Improving innovation performance through knowledge acquisition: the moderating role of employee retention and human resource management practices. J. Knowl. Manag. https://doi.org/10.1108/JKM-09-2017-0391.

Papa, A., Mital, M., Pisano, P., Del Giudice, M., 2019b. E-health and wellbeing monitoring using smart healthcare devices: an empirical investigation. Technol. Forecast. Soc. Chang. https://doi.org/10.1016/j.techfore.2018.02.018.

Piller, F., West, J., 2014. Firms, users, and innovation. In: New Frontiers in Open Innovation. 29(1).

Piller, F., Schubert, P., Koch, M., Möslein, K., 2005. Overcoming mass confusion: collaborative customer co-design in online communities. J. Comput.-Mediat. Commun. 10 (4), JCMC1042.

Plott, C.R., Chen, K.Y., 2002. Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem.

Preece, J., 2001. Sociability and usability in online communities: determining and measuring success. Behav. Inform. Technol. 20 (5), 347–356.

Rohrbeck, R., Hölzle, K., Gemünden, H.G., 2009. Opening up for competitive advantage–how deutsche Telekom creates an open innovation ecosystem. R&D Manag. 39 (4), 420–430.

Santoro, G., Bresciani, S., Papa, A., 2018a. Collaborative modes with cultural and creative industries and innovation performance: the moderating role of heterogeneous sources of knowledge and absorptive capacity. Technovation. https://doi.org/10.1016/j.technovation.2018.06.003.

Santoro, G., Vrontis, D., Thrassou, A., Dezi, L., 2018b. The internet of things: building a knowledge management system for open innovation and knowledge management capacity. Technol. Forecast. Soc. Chang. 136, 347–354.

Schemmann, B., Herrmann, A.M., Chappin, M.M., Heimeriks, G.J., 2016. Crowdsourcing ideas: involving ordinary users in the ideation phase of new product development. Res. Policy 45 (6), 1145–1154.

Scuotto, V., Del Giudice, M., della Peruta, M.R., Tarba, S., 2017. The performance implications of leveraging internal innovation through social media networks: an empirical verification of the smart fashion industry. Technol. Forecast. Soc. Chang. 120, 184–194.

Shirky, C., 2011. Cognitive Surplus?: How Technology Makes Consumers into Collaborators. Penguin Books, New York.

Sikdar, A., Vel, P., 2010. Getting the customer interested in your innovation: role of distribution and promotion strategies. Bus. Strat. Ser. 11 (3), 158–168.

Skibba, R., 2016. The polling crisis: how to tell what people really think. Nature 538, 304–306.

Somoza Sánchez, V.V., Giacalone, D., Goduscheit, R.C., 2018. Digital anthropology as method for lead user identification from unstructured big data. Creat. Innov. Manag. 27 (1), 32–41.

Spann, M., Ernst, H., Skiera, B., Soll, J.H., 2009. Identification of lead users for consumer products via virtual stock markets. J. Prod. Innov. Manag. 26 (3), 322–335.

Surowiecki, J., 2004. The Wisdom of Crowds. Doubleday, New York.

Toral, S.L., Martínez-Torres, M.R., Gonzalez-Rodriguez, M.R., 2017. Identification of the unique attributes of tourist destinations from online reviews. J. Travel Res. 0047287517724918.

Trabucchi, D., Buganza, T., Dell'Era, C., Pellizzoni, E., 2018. Exploring the inbound and outbound strategies enabled by user generated big data: evidence from leading smartphone applications. Creat. Innov. Manag. 27 (1), 42–55.

Tucci, C.L., Chesbrough, H., Piller, F., West, J., 2016. When do firms undertake open, collaborative activities? Introduction to the special section on open innovation and open business models. Ind. Corp. Chang. 25 (2), 283–288.

Ulloa, D., Saleiro, P., Rossetti, R.J., Silva, E.R., 2016. Mining social media for open innovation in transportation systems. In: Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on. IEEE, pp. 169–174.

Urban, G.L., Von Hippel, E., 1988. Lead user analyses for the development of new industrial products. Manag. Sci. 34 (5), 569–582.

Von Hippel, E., 1986. Lead users: a source of novel product concepts. Manag. Sci. 32 (7), 791–805.

Von Hippel, E., 1998. Economies of product development by users: the impact of 'sticky' local information. Manag. Sci. 44 (5), 629–644.

Von Hippel, E., 2001. Innovation by user communities: learning from open-source software. MIT Sloan Manag. Rev. 42 (4), 82.

Von Hippel, E., De Jong, J.P., Flowers, S., 2012. Comparing business and household sector innovation in consumer products: findings from a representative study in the United Kingdom. Manag. Sci. 58 (9), 1669–1681.

Walter, T.P., Back, A., 2013. A text mining approach to evaluate submissions to crowdsourcing contests. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE, pp. 3109–3118.

Wan, C.H., Lee, L.H., Rajkumar, R., Isa, D., 2012. A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine. Expert Syst. Appl. 39 (15), 11880–11888.

Wang, Y., Vanhaverbeke, W., Roijakkers, N., 2012. Exploring the impact of open innovation on national systems of innovation—a theoretical analysis. Technol. Forecast. Soc. Chang. 79 (3), 419–428.

West, J., Bogers, M., 2014. Leveraging external sources of innovation: a review of research on open innovation. J. Prod. Innov. Manag. 31 (4), 814–831.

Whitlark, D.B., Smith, S.M., 2001. Using correspondence analysis to map relationships. Mark. Res. 13 (3), 22.

Williams, R.L., Cothrel, J., 2000. Four smart ways to run online communities. MIT Sloan Manag. Rev. 41 (4), 81.

Youn Kim, H., Yoon, J.H., 2013. Examining national tourism brand image: content analysis of lonely planet Korea. Tour. Rev. 68 (2), 56–71.

Younis, E.M., 2015. Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study. Int. J. Comput. Appl. 112 (5).

Zhang, J., Ackerman, M.S., Adamic, L., 2007, May. Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th International Conference on World Wide Web, pp. 221–230.