Technological Forecasting & Social Change 174 (2022) 121266

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Converting consumer-generated content into an innovation resource: A user ideas processing framework in online user innovation communities

Jie Lin [a], Chao Wang [a], Lixin Zhou [b], Xiaoyan Jiang [a,*]

[a] School of Economics and Management, Tongji University, Shanghai, China
[b] Business School, University of Shanghai for Science and Technology, Shanghai, China

ARTICLE INFO

Keywords:
Online user innovation community
User ideas
Text clustering
Organizational information processing theory

ABSTRACT

An Online User Innovation Community (OUIC) is a space for consumers to share product usage experiences and put forward product improvement suggestions. However, as an increasing number of consumers post content in OUICs, companies face information processing challenges. Based on Organizational Information Processing Theory (OIPT), this study proposes a User Ideas Processing Framework (UIPF) to help enterprises efficiently process user ideas in OUICs and then applies it to a sample of 5,889 ideas from the Salesforce Idea Exchange. The case study results show that a UIPF can solve the information overload problem. Specifically, in Part 1 of the UIPF, we propose a new IDEA vectorization method and use it to cluster user ideas. Then, theme analysis is conducted on clusters to summarize the idea content in OUICs. This step gives us an overview of the information in OUICs. Compared with the standardized methods, our IDEA vectorization method can obtain better clustering results. Then, Part 2 of the UIPF builds a logistic regression model to identify innovative ideas from clusters. Compared with the famous "3C" method, the innovative ideas selected by the UIPF are more suitable for consumer requirements. In conclusion, the UIPF can help enterprises process information efficiently in OUICs.

1. Introduction

The Online User Innovation Community (OUIC) is a software platform for interactive innovation between consumers and enterprises, and it is a space for consumers to propose suggestions to improve products and services (Füller et al., 2014). However, the number of user ideas in OUICs far exceeds the information processing capabilities of enterprises (Frey and Lüthje, 2011; Li et al., 2016; Zheng et al., 2018), and information overload hinders interactive innovation. Therefore, enterprises need an efficient method to transform consumer-generated content in OUICs into innovative resources.

Many scholars have proposed approaches to filter valuable user ideas and alleviate information overload problems in OUICs (Bayus, 2013; Beretta, 2019; Di Gangi and Wasko, 2009; Hoornaert et al., 2017; Li et al., 2016; Ma et al., 2019; Zheng et al., 2018; Julia et al., 2015). The "3C" method has been frequently used in previous studies. Specifically, this method is used to predict the probability of an idea being adopted by a company through the characteristics of the content of the ideas, the contributor of ideas, and the crowd's feedback. User ideas that are usable, feasible, and novel can be adopted by companies for innovation (Thomas Mack, 2018), while others do not. User ideas usually refer to those that address certain product defects, and even if enterprises cannot adopt some ideas without innovative value, the defects described in the ideas should not be ignored by the company. In other words, user ideas that companies have not adopted are not valueless pieces of information but can be used as significant information resources for discovering consumer preferences (Martínez-Torres et al., 2015). In short, user ideas in OUICs are not only candidates for enterprise innovation plans but also the vane of enterprise innovation. Therefore, the utilization of information in OUICs consists of two parts: the first is refining idea content contained in user-generated ideas to discover what consumers discuss; and the second is selecting innovative ideas for product innovation. Scholars previously only focused on the second aspect of OUICs, ignoring the first. Idea content containing product use feedback, however, provides directional guidance for company innovation.

Companies need to process the information in OUICs wisely to convert user-generated content into innovative resources for the company. For the information processing challenges faced by companies in OUICs, Organizational Information Processing Theory (OIPT) provides reasonable solutions (Galbraith, 1974). OIPT suggests that an organization's Information Processing Capabilities (IPCs) should match the organization's Information Processing Needs (IPNs) for optimal

* Corresponding author.
E-mail addresses: linjie@tongji.edu.cn (J. Lin), chaoswang@foxmail.com (C. Wang), lxzhou@usst.edu.cn (L. Zhou), jiangxiaoyan@tongji.edu.cn (X. Jiang).

https://doi.org/10.1016/j.techfore.2021.121266
Received 30 October 2020; Received in revised form 16 August 2021; Accepted 5 October 2021
Available online 13 October 2021
0040-1625/

performance. However, the IPNs of companies has completely exceeded the IPC for open innovation in OUICs. There are strategies for meeting this challenge: creating slack resources or self-contained tasks, investing in Information System (IS), and creating lateral relations (Galbraith, 1974, Haussman et al., 2012).

We attempt to apply these four strategies to address the information processing challenges in the interactive innovation process. If companies solve this problem by creating slack resources, it requires hiring more product experts to review ideas in OUICs or adopting more user ideas to carry out product design experiments, in which case the cost may be greater than the benefit of open innovation. If companies build lateral relations between consumers and employees, employees will be overworked by having to interact with thousands of members in OUICs. Therefore, creating self-contained tasks or investing IS would be reasonable solutions. Natural language processing, deep learning, and computational science have been well developed and widely used. We combine these methods to propose a User Ideas Processing Framework (UIPF) for automatic analysis and mining consumer-generated content in OUICs. The UIPF can be integrated with OUICs to make the collection, analysis, and selection of ideas a self-contained task, thus reducing the information processing workload. Alternatively, it can also serve as an independent IS, allowing enterprises to analyze user-generated content according to their needs. Therefore, guided by OIPT, we propose a UIPF to enhance the IPCs of enterprises and meet their IPNs in open innovation.

To meet IPN for interactive innovation, the UIPF should include two sub-functions. The UIPF can first refine idea content from thousands of user-posted ideas, a new exploration of information value in OUICs, and then select innovative ideas from the content that users are concerned about in OUICs. The UIPF contributes to solving information processing challenges in OUICs, which helps to efficiently and cost effectively convert user-generated content into innovation resources. Customer-oriented product design and innovation can achieve better market performance. Moreover, capturing and understanding user preferences is a prerequisite for such innovation. The UIPF can complete this task.

This article is organized as follows: the second section reviews related works and introduces the theoretical background; the third section elaborates on the operation principle of the UIPF; the fourth section applies the UIPF to a real case; and the fifth section discusses the findings and limitations.

2. Conceptual background

2.1. Information overload and innovative idea identification in OUICs

There are two valuable resources in OUICs: consumer-generated content and gathered consumers. These two properties can bring two business values to companies: collecting user ideas for product innovation and introducing new products to consumers in OUICs (Dong and Wu, 2015; von Hippel, 2005). Our research focuses on the former. OUICs can be seen as an internet-based idea management system. As an increasing number of consumers join the community and post ideas, information overload has become a hindrance for companies to leverage consumer intelligence (Julia K. Froehlich, 2016; Eric et al., 2006; Rietzschel et al., 2010). Requiring employees or product experts to read and analyze ideas in OUICs line by line manually is a labor-intensive, costly, and time-consuming activity. Enterprises and academia have made some efforts to solve the information overload problem.

As OUIC initiators, companies lock in valuable ideas by encouraging users to comment, vote, and rate ideas in communities (Beretta, 2019; Schemmann et al., 2016). Companies believe that crowd wisdom foretells the future market for their products. This setting in OUICs and other public-facing scoring systems are open evaluations, which have an auxiliary effect on the selection and filtering of ideas in the community (Vivek et al., 2015). However, there are usually some differences between the user's vote and the company's choice. The user's vote can reflect consumer preferences, but the company's final choice is always related to the brand image (Martínez-Torres et al., 2015). Therefore, companies should consider various factors when they design open evaluation systems. After performing a state-of-the-art analysis on the existing open evaluation system, Jörg BA Haller identified 32 elements and element parameters for open evaluation system design (Joerg et al., 2017), thereby improving the usefulness of the open evaluation process.

Scholars have also proposed the "3C Method" to filter ideas in OUICs (Hoornaert et al., 2017). "3C" refers to the content of ideas, the contributor of ideas, and the crowd's feedback on ideas. Based on the characteristic of "3C", scholars use mathematical models (e.g., linear discriminant analysis, regularized logistic regression) to predict the likelihood of idea adoption by companies. Idea characteristics include length (Li et al., 2016) (Ma et al., 2019), supporting evidence of ideas (Li et al., 2016), and the number of videos (Ma et al., 2019). Contributor characteristics include prior participation (Bayus, 2013; Hoornaert et al., 2017; Li et al., 2016; Zheng et al., 2018) and the prior implementation rate (Li et al., 2016; Liu et al., 2020). Crowd feedback characteristics include popularity (Li et al., 2016; Schemmann et al., 2016), comments (Beretta, 2019; Hossain and Islam, 2015; Schemmann et al., 2016), and vote numbers (Beretta, 2019; Schemmann et al., 2016).

In summary, firms and scholars have proposed methods to select innovative consumer ideas in OUICs for product innovation, and these methods assist in addressing the information overload problem in open innovation. However, there are some problems with these methods. Typically, an idea with usability, feasibility, and novelty is more likely to be adopted by an enterprise (Thomas Mack, 2018), but the other ideas not adopted by companies are not necessarily invalid. For instance, users always state their preferences and identify the defects of products to develop improvement suggestions (Vivek et al., 2015) in ideas. The above methods only consider user ideas in OUICs as candidates for the enterprise's innovation plans but ignore discovering consumer preferences described in user ideas. Solving the information overload problem (Sloep, 2011) in this way does not take full advantage of information resources in OUICs. To fill this gap, our research proposes a User Ideas Process Framework to convert consumer-generated content into innovation resources.

2.2. Idea mining

Idea mining refers to mining new and innovative ideas automatically from unstructured text by computer methods (ALKSHER et al., 2016). There are two types of idea mining in the internet environment. The first is to identify text containing creative ideas from various web user-generated content (e.g., online product reviews, tweets). The second is the analysis of the content, structure, and type of ideas so that enterprises can use them for product innovation. The purpose of both types of idea mining is to transform big internet data into innovation resources for enterprises by using technologies such as natural language processing techniques, machine learning techniques, and ontologies.

Some scholars have proposed methods to identify ideas from internet big data. Chiu & Lin merge text mining and Kansei Engineering to analyze consumer reviews to extract customer preferences and product improvement ideas and then use them in the conceptual design stage of new products (Chiu and Lin, 2018). Thorleuchter introduces a new web mining approach that enables the automated identification of new technological ideas extracted from internet sources that can solve a given problem (Thorleuchter and Van den Poel, 2013). Olmedilla et al. propose the Co-occurrence Differential Analysis to discover ideas and potential innovations on the internet that are uniquely associated with one of several alternative topics (Olmedilla et al., 2019). Kasper Christensen uses a linear support vector machine to detect the pattern of ideas written as text and detect innovative ideas from 2803 texts in an online community (Kasper et al., 2016). Two experts from the company evaluated the novelty, feasibility, and value of ideas identified by the automatic idea detection system and concluded that machine learning and text mining are sufficiently valid for idea mining (Christensen et al., 2018).

There are also some methods to automatically analyze and present the semantic content and structure of ideas. Donghee Yoo builds an ontology-based cocreation approach to enhancing the system by transforming customers' ideas into semantic data in RDF format and then predicting the adaptability of each idea (Donghee Yoo, 2015). In addition, this system can also provide idea navigation services and semantic extraction services. Westerski et al. analyze the structure of ideas and suggest that an idea contains the description of "Proposed", "triggered", "innovation" and "object". According to the four characteristics, ideas can be classified into four categories: solutions, requests, problem reports, and suggestions (Westerski et al., 2013). Based on semantic technologies, Mackeprang et al. developed Kaleidoscope to describe ideas in a graph-based representation that is better able to explore and annotate existing ideas interactively (Mackeprang et al., 2018).

2.3. Organizational information processing theory

Galbraith introduced Organizational Information Processing Theory (OIPT) for the issue of information processing in organizations (Galbraith, 1974). There are three essential concepts of OIPT: Information Processing Needs (IPNs), Information Processing Capabilities (IPCs), and the fit between the two to obtain optimal performance (Galbraith, 1974). Companies face a significant challenge of information processing when they need to handle massive consumer-generated content in OUICs for open innovation. From the OIPT perspective, this problem can be explained and solved.

Since Galbraith proposed OIPT, some scholars have enriched and improved the theory (Haussmann et al.,2012). Advanced OIPT is as follows: interpersonal characteristics, interdependence and interdepartmental and interorganizational differences, and analyzability and a variety of tasks determine the equivocality and uncertainty of information processing. When there is a mismatch between an enterprise's IPNs and its IPCs, OIPT provides four methodological strategies to address this issue: by creating slack resources or self-contained tasks, time and job rotation, investing in IS, and creating lateral relations (Randolph B. Cooper, 2005; Burke et al., 2001; Chidambaram, 1996; Galbraith, 1974).

OIPT is widely used in business practice, specifically in the architecture of IS, Enterprise Resource Planning (ERP) systems, external technology integration (Haussmann et al.,2012), and organizational subunit process design (Zelt et al., 2018). According to OIPT, the architecture of IS needs to meet the enterprise's overall strategy, and the matching of technology and tasks is the key to efficient IS (Murugan Anandarajan, 1998). The ERP system is a typical representative of utilizing IT to enhance enterprises' IPCs, which combines marketing and manufacturing and solves the problem of information flow in the entire product line (Gattiker, 2007). OIPT can also explain technological uncertainty when integrating new external technologies (Stock and Tatikonda, 2008).

In short, with the blurring of organizational boundaries, OPIT has expanded from focusing only on information processing within the organization to simultaneously focusing on both its internal and external dimensions (Murugan Anandarajan, 1998). At the same time, introducing IT or IS for information processing may lead to an information overload problem(Haussmann et al.,2012; Galbraith, 1974). Therefore, the solution proposed by Galbraith becomes a case of an organizational design problem on its own (Haussmann et al.,2012). In fact, a decision support system for information processing rather than an information system is a better solution to the mismatch of IPNs and IPCs (Zmud, 1979).

In the process of interactive innovation, OIUCs are information systems introduced for the efficient collection of consumer feedback. OUICs can be seen as applying IS in the enterprise or as a new organizational subunit design for open innovation. The use of OIUCs instead of questionnaires and interviews has dramatically reduced the time and economic costs of collecting consumer information, but it has also led to the information overload problem. In other words, with thousands of people joining OUICs, the enterprise IPNs have completely exceeded their IPCs. Our research addresses the information processing challenges in open innovation from an OIPT perspective.

3. Method

The User Idea Processing Framework (UIPF) is designed to efficiently solve the challenge of information processing in OUICs and then convert user-generated content into innovation resources. UIPF (Fig. 1) includes two central parts: (1) Part 1: Discover Idea Content, and (2) Part 2: Select User Ideas. Part 1 uses a User Idea Cluster Algorithm to group user ideas into several clusters and conduct theme analysis to summarize the idea content of each cluster. Part 1 allows companies to know what consumers are concerned about in OUICs. Part 2 calculates the idea adoption rates of clusters and builds a logistic regression model to identify innovative ideas from clusters with low adoption rates.

The two parts of the UIPF correspond to two values of consumer-generated content in OUICs, and Part 2 is the continuation of Part 1. Part 1 provides an overview of the information in OUICs and helps companies grasp consumer preferences and the direction of innovation. Part 2 selects valuable user ideas targeting the idea content summarized from Part 1.

3.1. Part1: discover idea content

User-generated content in OUICs is users' dissatisfaction or suggestions for improving products, which is massive, complex, and uncertain. Companies need a tool to capture the big pictures of information in OUICs and determine the topics of greatest concern and most widely discussed by consumers.

3.1.1. User Idea Cluster Algorithm (UICA)

The User Idea Cluster Algorithm (UICA) is a text clustering algorithm that can group ideas into several subsets according to the semantic similarity of ideas contained in user-generated content. The uncertainty and complexity of ideas in a subset are less than those in all OUICs. Therefore, the cognitive load and selection complexity of companies for evaluating ideas in a single subset is significantly reduced.

Many cluster ideologies can be applied to text clustering, but we need to choose the appropriate ideology to build the UICA according to enterprises' information processing requirements. To match the innovation process in companies, the text cluster algorithm should meet the following two criteria: (1) The text cluster algorithm does not require initializing the cluster number because companies do not know in advance how many categories of consumer-generated content can be classified in OUICs. Therefore, the number of clusters cannot be one of the initialization parameters. (2) The result of the text cluster algorithm should show a hierarchical structure. Hierarchical results give companies the flexibility to view the results at different levels, allowing them to obtain an overview of the data or drill down to obtain detailed information.

We choose the hierarchical clustering ideology to construct the UICA for the above reasons. There are two steps in the UICA: (1) cluster preparation work and (2) cluster iteration.

(1) Cluster Preparation Work

Cluster preparation work includes three steps. ① Clean user-generated content: tokenize text into words list, steam words, and remove stop words and punctuation. ② Generate word vector: We use the standard vector space model, C-BOW, to obtain the vector representation of the word contained in the user-generated content. This step yields the word vectors. ③ Generate the IDEA vector and similarity

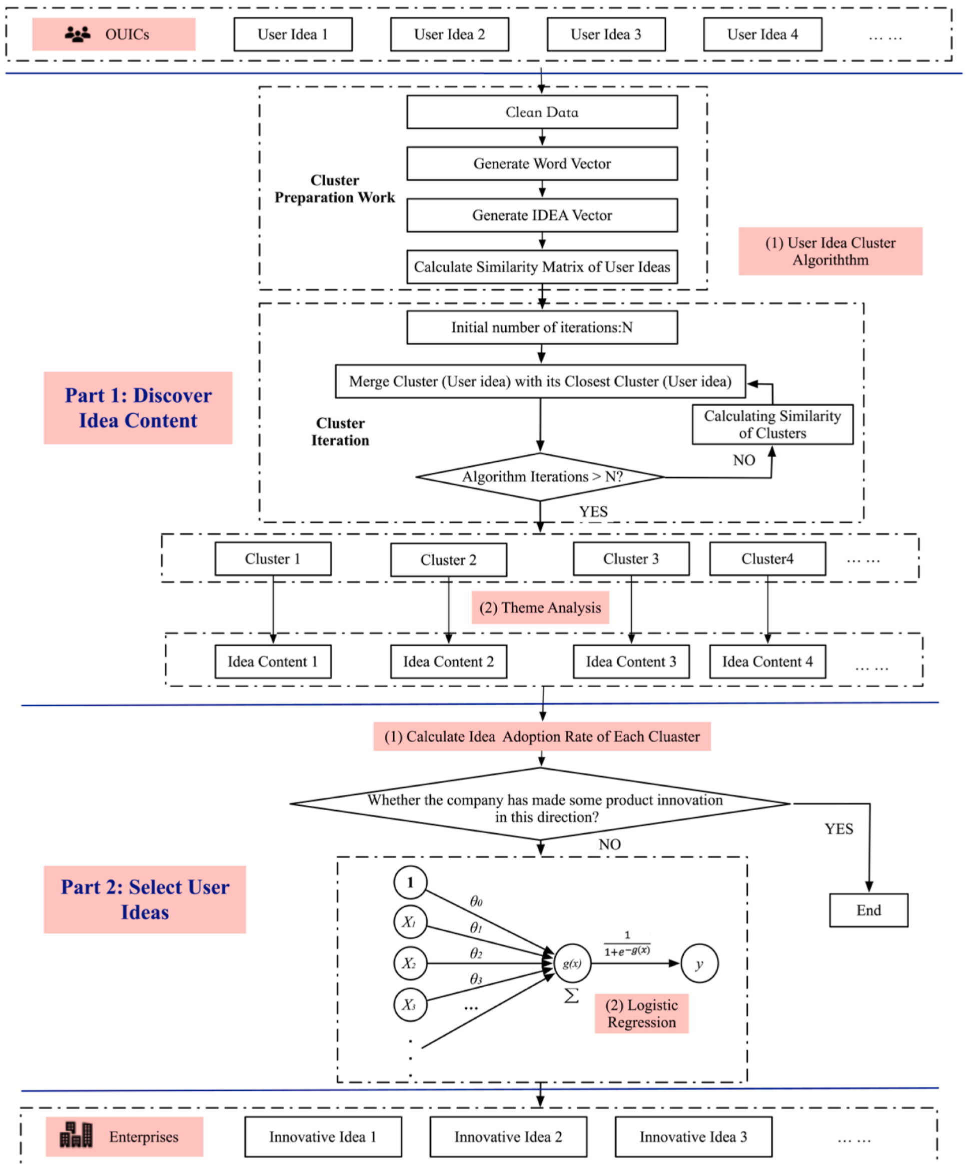

Fig. 1. User idea processing framework (UIPF) in OUICs.

matrix: calculate the relevance of words to idea expression, and then weighted average word vectors based on the relevance to obtain the IDEA vector. The similarity matrix is obtained by calculating the distance of the IDEA vector.

In step ③ generating the IDEA vector, we propose a new IDEA vectorization method to obtain the vectorized representation of the idea semantics contained in the text, rather than the vectorized representation of the complete text. Usually, in user-generated content, only some words are related to innovative ideas, and many words are not. UICA aims to cluster texts containing similar innovative ideas into one class. Therefore, obtaining the vectorized representation of the idea semantics contained in the user-generated content is the key to UICA. According to human writing habits, the title is usually the central concept of the whole text, and the body of the article is an extension around the title. Therefore, by calculating the similarity between the vector of words in the body and the TITLE vector, we can know whether the word is related to the idea description. Considering this similarity as a weight, the word vectors in the body are weighted and averaged to obtain the IDEA vector. Then, we calculate the distance of IDEA vectors to obtain the similarity matrix.

The IDEA vector is calculated as follows. First, get the TITLE vector. Calculate the TF-IDF value of the words in the title as the weight of words, and then weight average of the word vectors to obtain the TITLE vector (Eqn 1). Second, by calculating the similarity between the word vector in the body and the TITLE vector, we obtain the relevance between word and idea description (Eqn 2). Third, softmax all relevance of words in a text, and use the result to weight average of all words in a text to obtain the IDEA vector (Eqn 3). Finally, we calculate the distance between the idea vectors to obtain the idea similarity of the user-generated content for generating a similarity matrix (Eqn 4).

$$V_T = \frac{1}{n}\sum_{i}^{n} w_{(tf-idf)_i} * v_i, \tag{1}$$

where v_i is the vector of word i in the title, $w_{(tf-idf)_i}$ is the TF-IDF value of word, i, V_T is the TITLE vector, and n is the number of words in the title.

$$w_j = \frac{V_T \cdot v_j}{\| V_T \| \| v_j \|}, \tag{2}$$

where $\mathrm{v_j}$ is the vector of word j in the body and $\mathrm{w_j}$ represents the relevance of word j to the idea description.

$$V_I = \sum \frac{e^{w_j}}{\sum_k e^{w_k}} * v_j, \quad (3)$$

where $\frac{e^{w_j}}{\sum_k e^{w_k}}$ is the softmax value of w_j and V_I is the IDEA vector.

$$\text{Similarity}(V_{I_n}, V_{I_m}) = distance(V_{I_n}, V_{I_m}), \quad (4)$$

where V_{I_n} and V_{I_m} are two IDEA vectors.

(1) Cluster Iteration

Based on the similarity matrix of ideas, UICA starts the clustering iteration.① Initial iteration times N. Clustering is the process of merging nodes or clusters with high similarity, and N is the number of times to perform the above process. Companies can set a large N to overview the UGC or a small N to understand the UGC in detail. ② Merge clusters or nodes: traverse the similarity matrix, and merge node (cluster) with the node (cluster) that has the highest similarity to it. The algorithm can optionally set a minimum similarity threshold and never merge two clusters when their similarity is below this threshold. The similarity between clusters is calculated by Eqn 5. ③ Check whether the number of iterations reaches N. If it is satisfied, the iteration is ended; if not, return to ② and continue to merge clusters. Fig. 2 shows the process of the clustering iteration.

$$\text{Similarity } (c_p \ , \ c_q) = \frac{1}{y * k} \sum \text{Similarity}(V_{I_n,p}, V_{I_m,q}), \quad (5)$$

$$V_{I_n,p,} \in c_p \ , \ n \in \{1, 2, 3 \ldots\ldots y\} \ ,$$

$$V_{I_m,q} \in c_q \ , m \in \{1, 2, 3 \ldots\ldots k\} \ ,$$

where c_p , c_q are two independent clusters, $V_{I_n,p}$ is an idea contained in c_p, and $V_{I_m,q}$ is an idea contained in c_q.

3.1.2. Theme analysis

Theme analysis refers to abstracting the central concept in a large number of texts. UICA clusters UGC with a high degree of idea similarity into a cluster, and the text in the cluster usually describes the same product feature, product preference, or product usage feel. By understanding the content in a cluster, companies can grasp the direction of a product improvement. However, the number of texts in a single cluster can still be considerable. Therefore, the UIPF needs to include a function that automates the overview of the text in a cluster.

We accomplished automated theme analysis by using Python's 'textrank4zh' package. Considering the text in the cluster as elements in the list, we input the list into the textRank4Sentence function, consequently obtaining the abstract sentences of this cluster. By this method, the enterprise can quickly conduct an overview of the content in the cluster.

Using Part 1 to process UGC, the company can obtain a big picture of information and determine the idea content discussed by users in OUIC. This section indicates a direction for product improvement and innovation, and it explores the value of information in OUICs. Through Part 1, the IPC of enterprises is improved.

3.2. Part 2: select user ideas

Enterprises can consider the idea content obtained from Part 1 as the directions for product innovation. However, companies can only choose several directions to improve products because they have limited resources. Part 2 helps companies choose the innovation direction and select the most valuable ideas in that direction. Therefore, Part 2 should include the following steps: identifying the innovation direction, training the idea selection model, and selecting innovative ideas. Fig. 3 shows the process of selecting user ideas.

3.2.1. Identify innovation direction

The ideas in one cluster with high idea similarity can be summarized as idea content, which is considered a product innovation direction. Companies need to allocate innovation resources in these directions to produce products that consumers prefer. Part 2 helps companies select product innovation directions by calculating the adoption rate of user ideas in each cluster.

Of the thousands of ideas in OUICs, some are adopted for product development, some are rejected, and some are never reviewed by the company. When user ideas are grouped into different clusters, the adoption rate of each cluster varies. The adoption rate indicates the proportion of user ideas adopted by companies in a cluster, calculated by Eqn 6.

$$adoption\ rate = \frac{number\ of\ \ adopted\ ideas}{number\ of\ ideas}, \quad (6)$$

A high adoption rate of a cluster indicates that the enterprise has made multiple improvements toward this direction, while a low adoption rate of a cluster means that the enterprise has attempted to make improvements in this direction fewer times. In interactive innovation, a company should focus on those clusters with low adoption rates because the idea content of these clusters should be examined, but never has

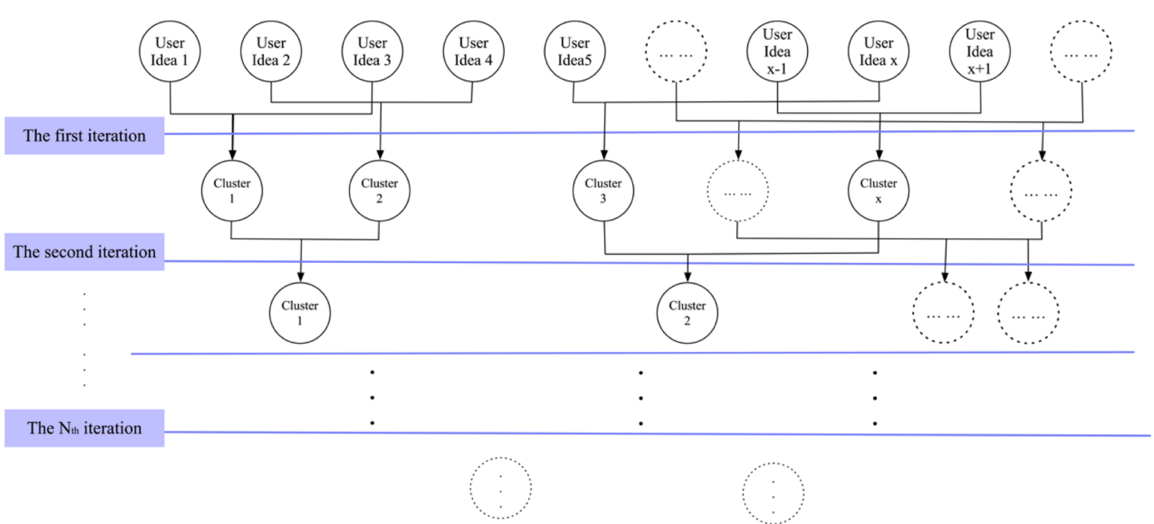

Fig. 2. The process of cluster iteration.

been.

3.2.2. Innovative ideas selection

Once the direction of innovation is selected, we need to select an idea from this cluster. A logistic regression model is designed and trained to distinguish valuable user ideas among the numerous ideas. Building a logistic regression model requires selecting and quantifying variables and calculating variable coefficients by historical data. Finally, the newly constructed model will be used to select user ideas.

(1) Variables and measurement

Previous research predicts the likelihood of an idea being adopted by a firm mainly based on the “3C” characteristics of the idea (Hoornaert et al., 2017): the content of ideas, the contributor of ideas, and the crowd’s feedback on ideas. To develop the logistic regression model, we integrate the variables in many studies (Bayus, 2013; Hoornaert et al., 2017; Li et al., 2016; Ma et al., 2019; Zheng et al., 2018). Variables and their measurements are interpreted in Table 1.

(1) Logistic regression model

It is a binary classification problem whether an idea is adopted by the company or not. Therefore, the logistic regression algorithm is suitable to solve this problem. Based on the provided variables, we formulate the logistic regression model as follows(Eqn 7):

$$y = \frac{1}{1+e^{-g(x)}}, \tag{7}$$

$$g(x) = \theta_0 + \theta_{1a}\cdot x_{1a} + \theta_{1b}\cdot x_{1b} + \theta_{1c}\cdot x_{1c} + \theta_{2a}\cdot x_{2a} + \theta_{2b}\cdot x_{2b} + \theta_{2c}\cdot x_{2c} + \theta_{2d}\cdot x_{2d} + \theta_3\cdot x_3 + \varepsilon.$$

Table 1
Variables and their Measurements.

Features	Measurements
Idea adoption	$y = 1$, *if idea i is adopted by companies*; $y = 0$, *otherwise*.
Content Characteristics	
- User Idea length (Li et al., 2016; Ma et al., 2019)	x_{1a} = *number of words in idea i*
Complexity of Content (Ahmed and Fuge, 2017)	x_{1b} : Complexity means the difficulty of understanding the text. This study uses the Probabilistic Language Model to calculate the text generation probability. The higher the probability, the lower the complexity.
Readability of Content (Ahmed and Fuge, 2017)	x_{1c}: Readability refers to how easy the text is to read. We calculate the Flesch Reading Ease (Palotti et al., 2015), reflecting the readability of content.
Contributors Characteristics	
-Contributor’s Prior Participation (Li et al., 2016; Schemmann et al., 2016)	x_{2a} = *question number* + *answer number* + *idea number* + *vote number* + *comment number* + *group number*
-Contributor’s Influence (Zheng et al., 2018)	x_{2b} = *badges point* + *certification number* + *employees* + *followers*
-Contributors’ Prior Adopted Rate (Li et al., 2016; Zheng et al., 2018)	$x_{2c} = \frac{\text{number of adopted ideds}}{\text{number of proposed ideas}}$
-Participants’ Diversity (Beretta, 2019)	x_{2d} = *location number* + *language number* + *industry number* + *job occupation number*
Crowd’s Feedback Characteristics	
- User Idea Popularity (Li et al., 2016; Zheng et al., 2018)	x_3 = *vote point* + 10 ∗ *merged number* + *comment number*

where θ_0 is the intercept, θ_i (from θ_1 to θ_3) is the regression coefficient of independent variables x_i (from x_1 to x_3), and ε is the error term. The maximum likelihood estimation method is used to examine this model. For the specific research question, the model is trained using historical data to determine the coefficients of each variable. Finally, a model with classification prediction capability can be obtained.

(1) Identify innovative user ideas

For a cluster with a low adoption rate, a company can use the logistic regression model to select innovative user ideas to improve the product. If companies hire product design experts to select ideas, it will cost significant human, time, and economic resources. Automatically selecting user ideas through mathematical models improves the IPC of enterprises, contributing to meeting their IPN. If the selected idea does not meet the innovation requirements, enterprises can improve a product by traditional innovation methods, such as recruiting experts to brainstorm.

4. Results

This section applies the UIPF to a case study and discusses the experimental results. Then, for idea clustering, the UICA proposed in this study is compared with the standard text clustering method; for identifying innovative ideas, the UIPF is compared with the ’3C’ method, a method recognized by many scholars for identifying innovative ideas in OUICs.

Section 4.1: provides a dataset characterization for more insight about it.

Section 4.2: shows the results of Part 1 of the UIPF.

Section 4.3: shows the results of Part 2 of the UIPF.

Section 4.4: compares the effect of our IDEA vectorization method with other baseline IDEA vectorization methods on idea clustering.

Section 4.5: compares the innovative idea identification results of the UIPF and "3C" methods.

4.1. Dataset characteristics

The dataset used in this case study includes 5889 ideas and 44,215 comments from the Salesforce Idea Exchange. Companies have processed some of the ideas, but not others. There are 5594 user ideas in Open status (not yet adopted), 293 ideas in Delivered status, 1 in Development status, and 1 in Not Planned status. All data are crawled by the Python crawler from the website (Salesforce.com), and the data fields are described in Table 2.

4.2. Result of part 1 – discover idea content

In Part 1, we apply the UICA and theme analysis on 5889 ideas. We randomly chose 38 as the number of iteration times in this case. In this experiment, we set the minimum similarity threshold to the 20th percentile of idea similarity. During each iteration, clusters with similarities below this threshold are not merged. The results of UICA contain 20 clusters.

Furthermore, we conduct a theme analysis of each cluster. The result

Table 2
Data field description.

Data fields	Description	Data type
IdeaId	The id of the idea	String
IdeaTitle	The title of the idea	String
IdeaBody	The content of the idea	String
UserId	The id of the idea’s publisher	String
IdeaComment	The comments of idea	String
IdeaStatus	The state of idea’s adoption	int

Table 3
Result of part 1 in UIPF.

	Number of elements	Theme of cluster
Cluster 0	273	Bundle product should get auto select in quote line editor and display the features and options related to the particular bundle
Cluster 1	307	This feature is available in Salesforce Classic, it would be nice to have in Lightning.
Cluster 2	393	It would be great to provide an option to log the email manually through Salesforce Inbox to the appropriate record's activity history
Cluster 3	214	When a user goes into an Opportunity Record and clicks on the Product Related List and clicks on the Products and then click 'Edit' they are permitted to update the Sales Price, Quantity, Date and Line Description Fields
Cluster 4	53	It would be helpful if there is a recall email option in Salesforce
Cluster 5	154	One feature of Salesforce Inbox is the ability to add new Accounts, Contacts, Leads or Opportunities to Salesforce from your email (Gmail in our case)'
Cluster 6	416	Our users wonder how to create a new account when they are on a custom list view in lightning
Cluster 7	347	Would be great to either be able to add the field to the page
Cluster 8	387	This is a must have feature in Salesforce
Cluster 9	131	Email and events are associated with related business account
Cluster 10	477	I would like to see Salesforce add a button in Campaigns entitled 'Launch Campaign Now'
Cluster 11	204	Filter the list of opportunities based on the account the contact is related to (the user creates the activity from a contact record and wants to also associate the activity to an opportunity
Cluster 12	332	Please consider this request to add 'Case' to the list of available 'Related Objects' when creating a Lookup Relationship custom field on the Opportunity object
Cluster 13	308	In the Winter '14 release, the Send and Add button was added to the Salesforce for Outlook panel when users send an email
Cluster 14	43	I really don't like this feature every time it sends me a reminder to log an internal email into Salesforce
Cluster 15	390	In reporting a Case to Salesforce Support, it appears that Lightning does not allow for any Read Only fields like the Phone and Email to be included in this view
Cluster 16	445	Calendar Syncing of Salesforce and Outlook.
Cluster 17	296	Authentication, supporting operation system, version, and API of salesforce app.
Cluster 18	504	Consistency and completeness of employee, contact, and record information retrieval.
Cluster 19	215	It would be great if users could set up a default sales team that depended on the record type of the opportunity.

of Part 1 is shown in Table 3. Fig. 4 shows a visualization of the 20 clusters. We reduced the dimensionality of the IDEA vector from 300 to 2 by the TSNE algorithm to present them in a two-dimensional space. In Fig. 4, different clusters are represented using points of different colors. The boundaries of each category can be clearly seen in the figure.

According to Table 3, we obtain a big picture of 5889 user ideas and summarize the idea content discussed by consumers in OUICs. Compared with all the information in the community, user ideas in a cluster have higher semantic similarity and a smaller amount, significantly reducing information uncertainty, reviewer's cognitive load, and choice complexity. In addition, Part 1 is mainly achieved by the computer method, significantly improving the enterprise's IPC.

4.3. Result of part 2 – select user ideas

With the overview of Idea content in OUICs, we need to choose the direction of product innovation and select the innovative ideas in the corresponding clusters.

(1) Identify innovation direction

In the experimental sample, some ideas were adopted by the companies, and some were not. The idea adoption rate is calculated for each cluster to learn the company's improvement frequency in this innovation direction. Table 4 shows the result.

The average adoption rate of clusters is 0.048. When the idea adoption rate of a cluster is lower than the average, it indicates that the firm may have neglected the content in this cluster in its previous interactive innovation. From the case study results, it is clear that Cluster 0, Cluster 2, Cluster 4, Cluster 7, Cluster 8, Cluster 10, Cluster 12, Cluster 14, Cluster 15, Cluster 17, and Cluster 18 have lower adoption rates than average. What is discussed in these clusters can be considered a future innovation direction for companies.

(1) Construct logistic regression model

We use some sample data to build the logistic regression model. There are 763 historical samples, 472 samples in Group 0 (unimplemented ideas), and 291 samples in Group 1 (implemented ideas). Their descriptive statistics are shown in Table 5. Moreover, the confusion matrix of the logistic regression model is shown in Table 6.

Table 6 shows that the model's precision is 94.8%, which is higher than 80%, a generally accepted standard precision. In the Salesforce Idea Exchange case, the logistic regression model is as follows:

$$y = \frac{1}{1 + e^{-(-4.548-0.001\cdot x_{1a}+0.001\cdot x_{1b}+0.003\cdot x_{1c}+0.001\cdot x_{2a}+\cdot 0.001x_{2b}+0.001\cdot x_{2c}+0.092\cdot x_{2d}+0.001\cdot x_3)}}$$

(1) Select innovative ideas

Among the 11 clusters with lower adoption rates in Table 4, Cluster 2, Cluster 8, Cluster 10, and Cluster 18 contain more user ideas. Therefore, we consider these four clusters as the directions for innovation. We randomly choose Cluster 2, Cluster 8, Cluster 10, and Cluster 18 as the directions for innovation. Then, this study uses the logistic regression model to select ideas in these four clusters; ideas with higher predicted scores are shown in Table 7.

In Table 7, ideas with a high likelihood of being adopted will be recommended to enterprises. If these ideas are practical and innovative, companies can utilize them for product innovation. At this stage, the use of mathematical models improves the company's IPC.

4.4. Comparison of IDEA vectorization methods and two standard method

In the clustering analysis, this study proposes a new IDEA vectorization method, which improves the performance of idea clustering. We compare our method with two standard text vectorization methods: the first uses the average of all word vectors in the UGC as the IDEA vector, and the second calculates the TF-IDF values of words in UGC as word weights and weighted averages the word vectors to obtain the IDEA vectors. This paper compares the influence of three IDEA vectorization methods on the clustering results, as shown in Table 8. The model proposed by us achieves the best results, with an F1 value of 85.41% and a Calinski Harabaz score (CH) of 336.46.

Fig. 5 shows the influence of the three IDEA vectorization methods on the cluster results. As shown in the figure, our model significantly outperforms the TF-IDF Weighted Model and is slightly better than the Average Model.

4.5. Comparison of UIPF and "3C" methods

"3C" methods are widely used in identifying innovative ideas in OUICs. For identifying innovative ideas, the difference between the UIPF

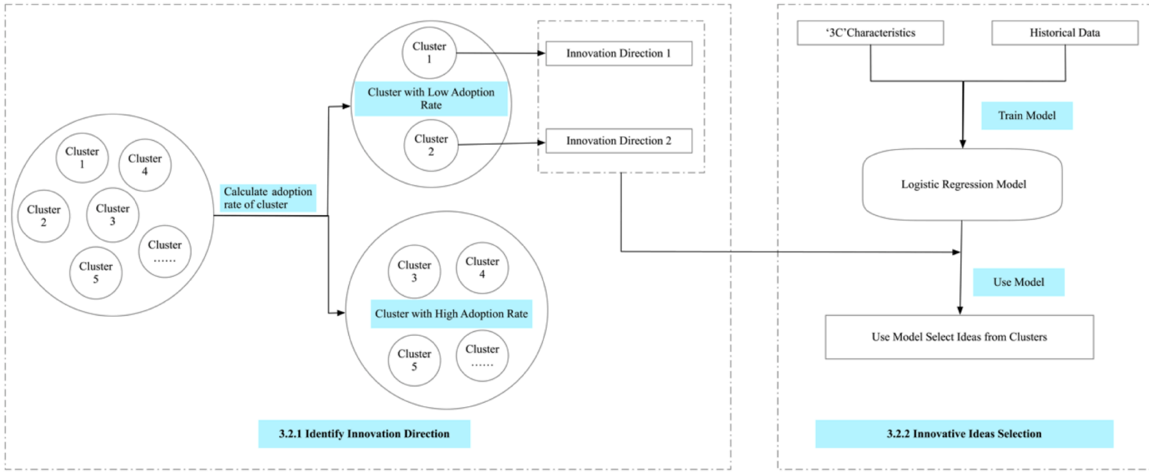

Fig. 3. The process of selecting user ideas.

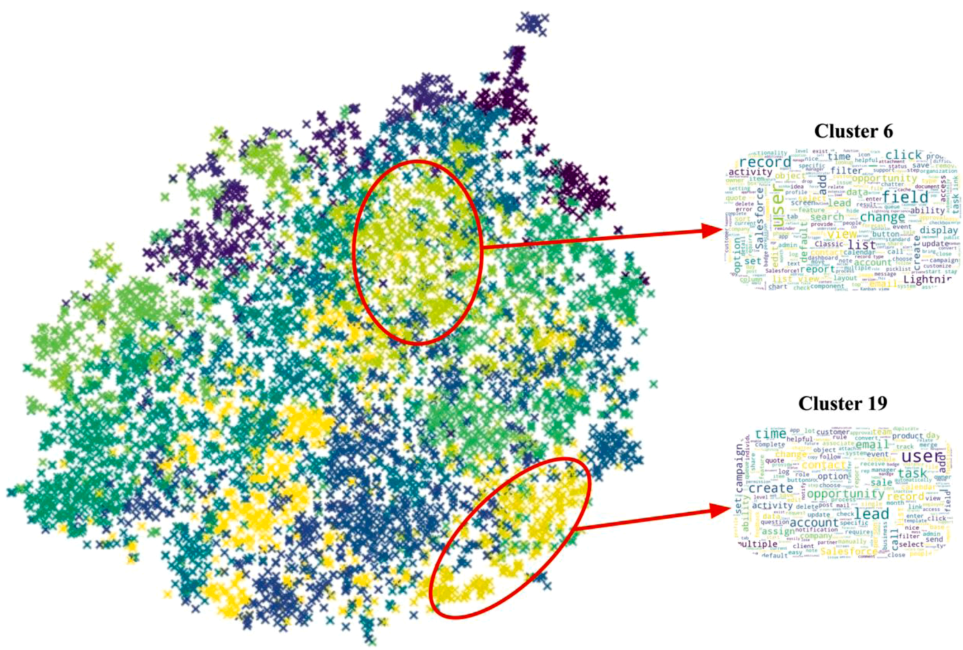

Fig. 4. The visualization of part 1 in UIPF.

Table 4
Adoption rate of user ideas.

Cluster No.	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Adoption rate	0.044	0.118	0.033	0.051	0.028
Cluster no.	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Adoption rate	0.056	0.061	0.0	0.031	0.082
Cluster no.	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14
Adoption rate	0.037	0.077	0.045	0.052	0.018
Cluster no.	Cluster 15	Cluster 16	Cluster 17	Cluster 18	Cluster 19
Adoption rate	0.034	0.057	0.044	0.023	0.082

and the "3C" method is whether the ideas are selected from specific innovation directions. The UIPF first clusters user ideas into N clusters and then selects ideas from clusters with lower adoption rates. The "3C" method directly identifies innovative ideas from all user ideas in OUICs. "3C" means the content of ideas, the contributor of ideas, and the crowd's feedback on ideas. The "3C" method refers to analyzing the "3C" characteristics of user ideas and predicting whether the company can adopt the ideas. Applying the "3C" method, we construct a logistic regression model to directly predict the likelihood of adopting all the ideas in the OUICs. The 12 ideas with the highest predictive value were selected.

Table 5
Descriptive statistics.

	N	Mean	Std. Dev.	Min	Max
Ideas implemented	763	0.38	0.49	0.00	1.00
User idea length	763	487.11	325.17	29.00	2784.00
Complexity of content	763	100.50	7196.86	135.20	445.36
Readability of content	763	−19.30	121.22	64.22	13.78
Contributor's prior participation	763	63.18	187.34	0.00	1365.00
Contributor's influence	763	0.39	10.45	0.00	215.00
Contributors' prior implementation rate	763	0.03	0.33	0.00	1.00
User idea popularity	763	493.26	4329.781	0.00	52,942.00
Feedback participants' diversity	763	5.82	21.80	0.00	613.00

Table 6
Confusion matrix of logistic regression model.

Confusion matrix		Predicted status 0	1	precision
True status	0	459	13	97.2%
	1	27	264	90.7%
(Total)				94.8%

Table 7
Ideas with high predicted scores.

	Tile of user idea
Cluster 2	l Make person accounts compatible with lightning l Make sales path available to web version of salesforce
Cluster 8	l Redirect to final merged account after merge accounts wizard l Mark complete button on tasks l Send an email: make CC and BCC fields lookup fields for any contact
Cluster 10	l Lightning console: close all tabs option l Lightning - increase contrast between text and background l Display custom objects in salesforce for outlook side panel l Ability to merge leads within lightning
Cluster 18	l Campaign history tracking l Ability to customize account hierarchy view l Allow 'add to campaign' for custom reports based on person accounts

Table 8
Comparison of clustering results.

Model	F1	Calinski Harabaz score
Our Model	85.41%	336.46
TF-IDF weighted model	80.26%	308.05
Average model	82.53%	333.32

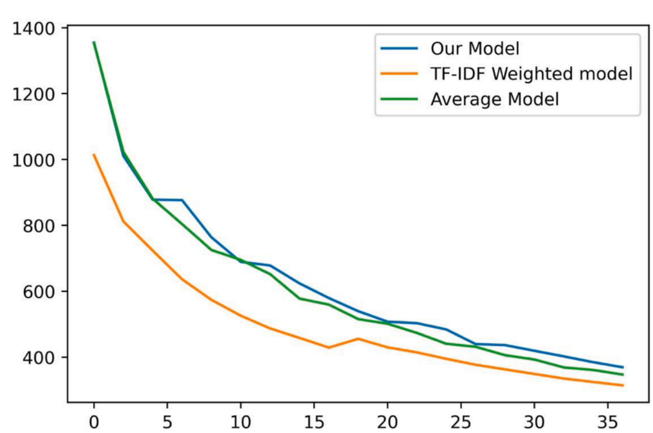

Fig. 5. Calinski Harabaz score of clustering algorithm.

In the process of product innovation, whether an idea can address consumer concerns is an essential indicator of the idea's value. Usually, the more elements contained within a cluster, the more users are concerned about this idea content; the lower the adoption rate of a cluster, the fewer product improvements are being made by the company in this direction. In other words, clusters with high user attention and low adoption rates are urgent issues for companies to improve. The ideas selected by the "3C" method belong to six clusters, and the adoption rates and the number of elements of these clusters are shown in Table 9. Table 10 shows ideas identified by the UIPF. By comparing Table 9 and Table 10, we find that the average adoption rate of the UIPF is lower than that of the "3C" method, and the number of elements contained in the clusters of the UIPF is higher than that of the '3C' method. Therefore, we conclude that the UIPF identifies more ideas oriented to those issues that consumers care about but and perhaps are being ignored by companies.

Table 9
Ideas identified by "3C" methods.

Cluster no	Number of ideas	Adoption rate	User attention(Number of elements)
Cluster 0	2	0.044	273
Cluster 9	3	0.082	131
Cluster 11	1	0.077	204
Cluster 12	3	0.045	332
Cluster 13	2	0.052	308
Cluster 17	1	0.044	296
Average	2	0.057	257.33

Table 10
Ideas identified by UISF.

Cluster no	Number of ideas	Adoption rate	User attention(Number of elements)
Cluster 2	2	0.033	393
Cluster 8	3	0.031	387
Cluster 10	4	0.037	477
Cluster 18	3	0.023	504
Average	3	0.031	440.25

5. Discussion

The vast amount of information in OUICs is an innovative resource for companies. However, companies need to have efficient IPC for utilizing this information. The UIPF is designed to solve this problem and then convert user-generated content into innovation resources. Part 1 of the UIPF summarizes the idea content to understand what consumers care about in the OUICs, and Part 2 of the UIPF identifies innovative ideas from the clusters corresponding to the issues that consumers are concerned about in OUICs. The two parts of the UIPF correspond to the two values of UGC in OUICs.

5.1. Implication for research

The UIPF makes some theoretical contributions to idea mining research in OUICs. First, the study identifies a new value of UGC in communities, discovering consumer concerns by summarizing idea content from the vast amount of information in OUICs. Previous studies (Beretta, 2019; Di Gangi and Wasko, 2009; Li et al., 2016; Ma et al., 2019; Zheng et al., 2018) have emphasized that OUIC is a source of innovative ideas and ignores the valuable information in ideas that enterprises do not adopt. Clustering user ideas and extracting cluster themes before identifying innovative ideas enables an overview of UGC in OUICs. This finding is instructive for scholars exploring how to convert UGC into business value at a low cost.

Second, the UIPF is an application of the Organization Information Processing Theory (OIPT) in the context of interactive innovation. As organizational boundaries become increasingly blurred, OIPT has evolved to focus increasingly on the external information of the enterprise. In the era of information explosion, proper handling of consumer information is an inaccessible link in production operations and innovation. The UIPF was designed to enhance a company's IPCs) to meet the IPNs of interactive innovation in OUICs. First, the UIPF splits user ideas into subsets, reducing information uncertainty, cognitive load, and complexity. Second, using computer-aided methods instead of manual work reduces the cost and improves information processing efficiency. The UIPF can be used as a stand-alone information system used by employees to process user ideas in OUICs according to companies' needs. It can also be integrated with OUICs to become an intelligent system that makes idea collection, analysis, and selection a self-contained task. The development of IT brings information processing challenges for enterprises, and OIPT will guide enterprises to develop more solutions to solve this problem.

Third, we propose a new IDEA vectorization method and apply it in text clustering. It can well transform the idea semantics contained in UGC into numerical vectors. When analyzing UGC, a standard method is to average or weighted average (e.g., using the TF-IDF value as the weight) all word vectors to obtain a vectorized representation of UGC. Usually, companies or academic research care only about target-related expressions in UGC, not all vocabularies. Therefore, the specific IDEA vectorization method can precisely extract the semantics of ideas instead of the entire content. Using the specific IDEA vector for text analysis, we can obtain better creative clustering results.

5.2. Implication for practice

For companies, the UIPF is a useful tool for discovering Idea Content and identifying innovative ideas in OUICs. Part 1 of the UIPF uses the UICA to group user ideas into several clusters and conducts theme analysis to summarize the clusters' main ideas. By this part, companies learn an overview of information and discover what consumers care about in OUICs. This tool can partially replace previous market research methods, such as questionnaires and interviews, to help companies obtain information on customer preferences, product usage feedback, and others at a lower time and economic cost. After calculating the adoption rate of user ideas in clusters, Part 2 of the UIPF built a logistic regression model to predict the likelihood of new idea adoption in a specific innovation direction. With the UIPF, companies can efficiently utilize information in OICUs for corporate innovation.

5.3. Limitations and future research

OUICs have brought good commercial value to enterprises, and an increasing number of enterprises are adopting this open and innovative way of updating and developing products. Information overload is a challenge of online communities, but if companies cannot efficiently process UGC, a large amount of valuable information may be ignored. The processing of information in communities is an inevitable act of utilization of crowd wisdom, which toolkits or employees can realize in enterprises (Parmentier, 2015). Today, when the cost of human resources is relatively expensive, computer-assisted methods are an excellent way to access ideas with high efficiency and low cost. Therefore, an intelligent UGC processing method is an indispensable tool for enterprises to use OUICs for open innovation.

In the future design of user idea processing systems in OUICs, a product function-oriented information filtering mechanism may be practical. For example, if an enterprise wants to improve function A of a product, it can only distinguish information related to function A through the system and then improve it. This design's realization can be achieved with machine learning and natural language processing technology (Christensen et al., 2018). Compared with the framework proposed in this article, the specific function-oriented information filtering mechanism only needs to process part of the information in OUICs, not all information. When an enterprise already has a specific innovation direction, the function-oriented information filtering system allows the enterprise to seize innovative ideas more proactively. This system may be more practical.

In addition, the UICA proposed in this paper can be further used for community detection of consumers by combining the behavior information of those who publish ideas. UICA divides UGCs in OUICs into clusters based on idea similarity, and consumers posting similar ideas within each cluster can be considered a group that focuses on the identical product innovation direction. This community detection approach can be extended from interactive innovation to other domains, such as academic networks and communities. Mercorio et al. proposed Discovery Information using Community detection (DICO), a framework that jointly documents metadata, such as authorships and citations, with semantic information to model the interactions between authors and finally identifies overlapping communities of authors (Mercorio et al., 2019). The methodology proposed by Mercorio et al. can also be used in the field of interactive innovation. The joint analysis of textual, behavioral, and interpersonal interaction information in OUICs enables community detection of consumers, which is of great value for interactive innovation. For example, when a company selects an innovation direction, it can interact with corresponding consumer groups to jointly establish product improvement goals, work together on product design, and try out and improve new products. This study focuses on uncovering the value of UGC in OUICs. Subsequent studies can focus on exploring the value of users in OUICs.

Author statement

We would like to submit the enclosed manuscript entitled “Converting Consumer-Generated Content into an Innovation Resource: A User Ideas Processing Framework in Online User Innovation Communities”, which we wish to be considered for publication in “Technological Forecasting & Social Change”. I would like to declare on behalf of my co-authors that the work described was original research that has not to be published previously, and not under consideration for publication elsewhere, in whole or in part. No conflict of interest exists in the submission of this manuscript, and the manuscript is approved by all authors for publication.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by the National Key Research and Development Program of China(2018YFC0830400), National Natural Science Foundation of China (71672128) and Postdoctoral Science Foundation Funded of China (2021M692135).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.techfore.2021.121266.

References

AHMED, F., FUGE, M., 2017. Capturing winning ideas in online design communities. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.

ALKSHER, M., AZMAN, A., YAAKOB, R., KADIR, R.A., MOHAMED, A.H., ALSHARI, E.M, et al., 2016. A review of methods for mining idea from text. 2016 Third International Conference On Information Retrieval and Knowledge Management. IEEE.

BAYUS, B.L., 2013. Crowdsourcing new product ideas over time: an analysis of the dell ideastorm community. Manage. Sci. 59, 226–244.

BERETTA, M., 2019. Idea selection in web-enabled ideation systems. J. Prod. Innov. Manag. 36, 5–23.

BURKE, K., AYTES, K., CHIDAMBARAM, L., 2001. Media effects on the development of cohesion and process satisfaction in computer-supported workgroups - An analysis of results from two longitudinal studies. Inf. Technol. People 14, 122–141.

CHIDAMBARAM, L., 1996. Relational development in computer-supported groups. MIS Q 20, 23.

CHIU, M.-.C., LIN, K.-.Z, 2018. Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. Adv. Eng. Inform. 38, 826–839.

CHRISTENSEN, K., SCHOLDERER, J., HERSLETH, S.A., NAES, T., KVAAL, K., MOLLESTAD, T., VEFLEN, N., RISVIK, E., 2018. How good are ideas identified by an automatic idea detection system? Creativity Innov. Manag. 27, 23–31.

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, 303–312.

DONG, J.Q., WU, W., 2015. Business value of social media technologies: evidence from online user innovation communities. J. Strateg. Inf. Syst. 24, 113–127.

DONGHEE, Y.O.O., K, C., HANJUN, L.E.E., YONGMOO, S.U.H., 2015. An ontology-based co-creation enhancing system for idea recommendation in an online community. Data Base Adv. Inf. Syst. 43, 14.
ERIC, F, RIETZSCHEL, B.A.N., WOLFGANG, STROEBE, 2006. Productivity is not enough: a comparison of interactive and nominal brainstorming groups on idea generation and selection. J. Exp. Soc. Psychol. 42, 244–251.
FREY, K., LÜTHJE, C, 2011. Antecedents and consequences of interaction quality in virtual end-user communities. Creativity Innov. Manag. 20, 22–35.
FÜLLER, J., HUTTER, K., HAUTZ, J., MATZLER, K, 2014. User roles and contributions in innovation-contest communities. J. Manag. Inf. Syst. 31, 273–308.
GALBRAITH, J.R., 1974. Organization design: an information processing view. Interfaces (Providence) 4, 28–36.
GATTIKER, T.F., 2007. Enterprise resource planning (ERP) systems and the manufacturing–marketing interface: an information-processing theory view. Int. J. Prod. Res. 45, 2895–2917.
HAUSSMAN, C., D, Y.K., VENKITACHALAM, K., WILLIAMS, M.D., 2012. A summary and review of Galbraith's organizational information processing theory. In: SCHNEBERGER, Y.K.D.R.W.L. (Ed.), Information Systems Theory. Springer, New York, NY.
HOORNAERT, S., BALLINGS, M., MALTHOUSE, E.C., VAN DEN POEL, D., 2017. Identifying new product ideas: waiting for the wisdom of the crowd or screening ideas in real time. J. Prod. Innov. Manag. 34, 580–597.
HOSSAIN, M., ISLAM, K.M.Z, 2015. Ideation through online open innovation platform: dell ideastorm. J. Knowl. Econ. 6, 611–624.
JOERG, B.A, HALLER, V.K.V., DIRK, SCHNECKENBERG, KATHRIN, M.MOESLEIN, 2017. Exploring the design elements of open evaluation. J. Strategy Manag. 10, 26.
JULIA, K.FROEHLICH, M, H., MICHAEL, GIBBERT, 2015. Idea selection in suggestion systems a thematic similarity perspective. R&D Manage. 46, 13.
JULIA, K.FROEHLICH, M, H., MICHAEL, GIBBERT, 2016. Idea selection in suggestion systems: a thematic similarity perspective. R&D Manage. 46, 887–899.
KASPER, C., SLADJANA, N., LARS, F., JOACHIM, S., 2016. Search of new product ideas identifying ideas in onlinecommunities by machine learning andtext mining. Creativity Innov. Manag. 26, 14.
LI, M., KANKANHALLI, A., KIM, S.H, 2016. Which ideas are more likely to be implemented in online user innovation communities? An empirical analysis. Decis. Support Syst. 84, 28–40.
LIU, Q., DU, Q., HONG, Y., FAN, W., WU, S., 2020. User idea implementation in open innovation communities: evidence from a new product development crowdsourcing community. Inf. Syst. J. 30, 899–927.
MA, J., LU, Y., GUPTA, S., 2019. User innovation evaluation: empirical evidence from an online game community. Decis. Support Syst. 117, 113–123.
MACKEPRANG, M., STRAMA, J., SCHNEIDER, G., KUHNZ, P., BENJAMIN, J.J. & MÜLLER-BIRN, C. 2018. Kaleidoscope: an RDF-based exploratory data analysis tool for ideation outcomes. The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings.
MARTÍNEZ-TORRES, M.D.R., RODRIGUEZ-PIÑERO, F., TORAL, S.L, 2015. Customer preferences versus managerial decision-making in open innovation communities: the case of Starbucks. Technol. Anal. Strateg. Manag. 27, 1226–1238.
MERCORIO, F., MEZZANZANICA, M., MOSCATO, V., PICARIELLO, A., SPERLI, G., 2019. DICO: a graph-DB framework for community detection on big scholarly data. IEEE Trans. Emerg. Top. Comput. 1.
MURUGAN ANANDARAJAN, B.A, 1998. Matching client/server processing architectures with information processing requirements: a contingency study. Inf. Manag. 34, 10.
OLMEDILLA, M., SEND, H., TORAL, S.L, 2019. Identification of the unique attributes and topics within smart things open innovation communities. Technol. Forecast. Soc. Change 146, 133–147.
PALOTTI, J.R.D.M., ZUCCON, G., HANBURY, A., 2015. The influence of pre-processing on the estimation of readability of web documents. Proc. 24th ACM Int. Conf. Inf. Knowl. Manag.
PARMENTIER, G., 2015. How to innovate with a brand community. J. Eng. Tech. Manage. 37, 78–89.
RANDOLPH B.COOPER, R. A.W. 2005. Information processing model of information technology adaptation: an intra-organizational diffusion perspective. Data Base Adv. Inf. Syst., 36, 19.
RIETZSCHEL, E.F., NIJSTAD, B.A., STROEBE, W., 2010. The selection of creative ideas after individual idea generation: choosing between creativity and impact. Br. J. Psychol. 101, 47–68.
SCHEMMANN, B., HERRMANN, A.M., CHAPPIN, M.M.H., HEIMERIKS, G.J, 2016. Crowdsourcing ideas: involving ordinary users in the ideation phase of new product development. Res. Policy 45, 1145–1154.
SLOEP, R.L.L.S.M.B.-R.P.B, 2011. What's in it for me? Recommendation of peers in networked innovation. J. Univers. Comput. Sci. 17, 14.
STOCK, G.N., TATIKONDA, M.V, 2008. The joint influence of technology uncertainty and interorganizational interaction on external technology integration success. J. Oper. Manage. 26, 65–80.
THOMAS, M.A.C.K., C, L, 2018. Submission Quality in Open Innovation Contests an Analysis of Individual-Level Determinants of Idea Innovativeness, 3. R& D Management, p. 16.
THORLEUCHTER, D., VAN DEN POEL, D., 2013. Web mining based extraction of problem solution ideas. Expert Syst. Appl. 40, 3961–3969.
VIVEK, K, VELAMURI, D.S., JÖRG, B.A, HALLER, KATHRIN M, MOESLEIN, 2015. Open evaluation of new product concepts at the front end of innovation: objectives and contingency factors. R& D Manage. 21.
VON HIPPEL, E., 2005. Democratizing innovation: the evolving phenomenon of user innovation. J. Betriebswirtschaft 55, 63–78.
WESTERSKI, A., DALAMAGAS, T., IGLESIAS, C.A, 2013. Classifying and comparing community innovation in Idea Management Systems. Decis. Support Syst. 54, 1316–1326.
ZELT, S., RECKER, J., SCHMIEDEL, T., VOM BROCKE, J, 2018. Development and validation of an instrument to measure and manage organizational process variety. PLoS ONE 13, e0206198.
ZHENG, Q., GUO, W., AN, W., WANG, L., LIANG, R., 2018. Factors facilitating user projects success in co-innovation communities. Kybernetes 47, 656–671.
ZMUD, R.W., 1979. Individual differences and MIS success: a review of the empirical literature. Manage. Sci. 25, 966–979.

Jie Lin is a professor at the School of Economics and Management, Tongji University. His-research interests include supply chain simulation, big data analysis, interactive innovation, and lead user. Tongji Building A, Siping Road 1500, Shanghai 200,092, China.

Chao Wang is a graduate student at the School of Economics and Management, Tongji University. His-research interests include data mining, innovative behavior of social network users. Address: Tongji Building A, Siping Road 1500, Shanghai 200,092, China.

Lixin Zhou a post-doctor at the Business School, University of Shanghai for Science and Technology. His-research interests include open innovation, prediction of personality in social networks. Address: 334 JunGong Road, Yangpu District, Shanghai 200,000, China.

Xiaoyan Jiang is a doctoral student at the School of Economics and Management, Tongji University. Her research interests include text mining, open innovation, and social network analysis. Address: Tongji Building A, Siping Road 1500, Shanghai 200,092, China.