Framework of early adopters’ incipient and innovative ideas and chance discovery

Chao-Fu Honga, Mu-Hua Linb, Pen-Choug Suna and Hsiao-Fang Yangc

aDepartment of Information Management, Aletheia University, New Taipei City, Taiwan; bInnovative DigiTech-Enabled Applications & Services Institute, Institute for Information Industry, Taipei, Taiwan; cDepartment of Accounting, College of Commerce, National Chengchi University, Taipei, Taiwan

ABSTRACT
The innovative diffusion theory indicates that the key to success of businesses is the innovative ideas of the early adopters. Furthermore, the early adopters’ documents on the Internet were extremely rare; the traditional associative analyses in text mining tend to ignore these useful ideas of the early adopters. In this study, a framework was proposed, which uses low-term frequency (TF), low-term frequency with inverse document frequency and low TF with the inverse clusters frequency, to acquire rare connections between low-frequency terms, to extract early adopters’ incipient and innovative ideas. This new proposed framework amplifies the rare chance to find potential terms which are valuable for businesses’ future. Finally, some observed data obtained from the passengers on airplanes or trains were used to extract the innovative ideas from early adopters. By putting the data into a business scenario, a case study was presented and the feasibility of the framework can therefore be checked by experts. A comparison has been made between the proposed framework and chance discovery. The experimental result evidences that the results in the new framework are more effective than the outcomes of chance discovery method to sift out incipient ideas.

1. Introduction
According to Rogers’ innovation diffusion model (IDM), early adopters make up merely 13.5% of the market. Moore (1991) discovered a bowling alley crossing the chasm between the early adopter and the early majority, which can be narrowed down by increasing the early majority’s awareness and level of acceptance of innovative products. Therefore, in order to bring a big business chance for enterprise, the issue of discovering ‘Early Adopters’ Incipient and Innovative Ideas’ needs to be addressed. There are two kinds of traditional analyses related to rare terms with high confidence. The first kind of analysis, which is called linking analysis and is effective in trend analysis, mainly focuses on high support terms. The second kind of analysis, which is widely used in Chance discovery, concentrates on low support terms with high confidence. However, these traditional
approaches can hardly explore any early adopters’ incipient innovative ideas, which can only be found in terms of low support and low confidence.

Hong, Lin, and Lin (2016) and Lin, Hong, and Lin (2017) used low-frequency terms with inverse clusters frequency and successfully discovered early adopters’ incipient and innovative ideas. Unfortunately, because they used the traditional logic and thinking to analyse problems related to rare terms with high confidence, there were still much redundant data existing in the extreme groups in their experimental results. There were two extreme groups of data, which appeared not to be early adopters’ incipient and innovative ideas. The first group of data dominating the values of low-term frequency (TF) found terms which were not so rare and failed to reveal early adopters’ incipient and innovative ideas. The other group of data dominating the values of inverse clusters frequency found only a fortuitous idea, which came from one early adopter.

This paper is an extended version of the article in ACIIDS 2016 (Hong et al., 2016). To target early adopters’ incipient and innovative ideas more precisely than previous works, a framework using hybrid methods was proposed in this paper. This new framework involves analyses of low TF, low-term frequency with inverse document frequency (TF-IDF) and low-term frequency with inverse clusters frequency (TF-ICF). In order to proof the terms extracted by the proposed framework to create incipiently innovative ideas, the framework need to be test by experts. Some travelling documents were collected using Google search. And then, by using the exact same data, the new proposed framework and the traditional chance discovery were put to run at the same time. The experimental results prove that some rare terms found by the framework could really estimate the experts to recognize incipiently innovative ideas. By comparing the results of both the approaches, the researchers evidenced that the new proposed framework is better than chance discovery in finding incipient and innovative ideas. Consequently, the innovative ideas created in the new framework effectively provided enterprise owners a future business chance to success.

In the beginning of the paper, a literature review is displayed. A proposed framework of incipient innovative ideas discovery is presented next. Finally, a case study is presented in a business scenario, in which there are some data collected from the passengers on airplanes or trains. The feasibility of the proposed framework to extract the innovative ideas provided by the early adopters can therefore be compared with chance discovery by experts.

2. Literature review

For a business to grow sustainably, it is better for an enterprise owner to start the business from neo-market, and then to niche market or mass-market, which are markets consisting of neo-tribe, niche-tribe and majority individually. As in previous discussion, the early adopters’ incipient ideas are the key factors to build neo-market. In this section, the contribution of the useful ideas of early adopters and how to process the business growth through gaining a bigger share of market will be discussed as follows.

2.1. Translation of early adopters’ incipient and innovative ideas

According to IDM, the early adopters or lead users are the few people who buy it in a short time and form a neo-market, when products are newly introduced to the market.
Robertson (1971) defined innovative use as ‘the degree to which an individual is relatively earlier in adopting new products than other members of his social system’. In addition, creative consumers may possess special skills and abilities required for using the product in a wide variety of ways to solve their problems and meet their needs (Hirschman, 1980; Price & Ridgway, 1982). Therefore, the results are like the first-stage translation of consumption-driven market emergence (CDME) (Martin & Schouten, 2014). The variations of an innovation are incipiently produced by special consumers such as early adopters.

Therefore, Shane and Venkataraman (2000) argued that individuals (early adopters or lead users) must possess prior knowledge to perceive the value of innovative products and to recognize their worth. The ability and sensitivity of those people, who compare ideas of new products as ‘business opportunity’, may lead to development of new markets (Shane, 2003). By further identifying key early adopters (neo-tribes) to explore their innovative ways of using new products, business may trigger social influence. The ways how early adopters influence innovation diffusion are discussed here in detail. Social influence, which is the interactions within individuals of a group, may serve as the fundamental role of a medium for spreading information, ideas and influence among its members (Agrawal, Imielinski, & Swami, 1993). In the second-stage translation of CDME, the incipient innovative ideas will be used to organize local meta-communities for formatting in a niche market.

Nowadays, benefiting from Web 2.0, the early adopters may easily put their comments about innovative use of a new product. In 2011, by using grounded theory (GT)-based co-occurrence analysis, Lin and Hong extracted innovative values of the related consumers’ data collected from the Google blog. They focused on the terms with low confidence and successfully explored the evidences of the majority influenced by the early adopters to accept a new product. Unfortunately, their research only evidenced that they can organize local communities to build a niche market, just like the second-stage translations of CDME.

In summary, if an enterprise can catch the first-stage innovative ideas of CDME, then it has a chance to entrance the second stage of CDME to influence more people in the markets. So, incipient and innovative ideas of the early adopters are important key factors for enterprise to run a new beneficial business. In this study, the first stage of CDME thinking is also adopted. The list of early adopters’ incipient innovative words is studied first, and then the potential word’s linkages with other words’ network was analysed. The higher potential words have more ability to organize a bigger market. This is different from traditional analysis methods, which extract rare but important words from words’ networks with low support and high confidence. More discussions in detail will be in the following section.

2.2. Extracting innovative ideas from the neo-market: rare frequency and rare connection with words’ clusters

As mentioned above, only lead users are able to create new ways of using innovative products that lead to the popularity of such products (Price & Ridgway, 1982). In Morrison’s experiment (2000), only about 26% of the consumers have the ‘leading edge status’ and in-house technical capabilities for them to freely share their innovations with
others. Therefore, the characteristics of these lead users are similar to those of the innovators and early adopters, who are mentioned by Rogers and are only about 15% of all the consumers (Rogers, 2003). When a company decides to involve in innovation/knowledge creation, how to select right lead users and bring them on board can be a critical point to the success.

The traditional predictive methods are usually adopted, which use associative analysis involving words with high support and high confidence to build strong words network. All the words in the network are well known to people and nearly impossible to find any incipient ideas; so this type of analysis is not practical to be used in this research. In order to extract early adopters’ incipiently innovative ideas, such an associative analysis is not going to be adopted in this study.

Two kinds of analyses are associative with rare words with high confidence. The first kind of analysis called linking analysis focuses on high-frequency words. It calculates the links of the words to low-frequency words with high confidence. It is effective in trend analysis and quite often to be used to show future directions to business. Just like shown in Figure 1(a), Chiu (2014) and Chiu and Hong (2015) successfully extracted near-future technical opportunity for business by using this type of mechanism.

On the other hand, the second kind of analysis concentrates on rare words with high confidence. It counts the links of these words to strong words’ networks, which are clusters of high-frequency words. The famous KeyGraph algorithm, which was proposed by Ohsawa et al. in 1998, belongs to this type of analysis. The algorithm calculates a node’s frequency. The co-occurrences of the node and other nodes in a shopping cart are also counted. By doing so, strong words’ clusters are to emerge as shown in Figure 1(b). Key-values of rare nodes are derived first, and then the nodes’ linkages to strong words’ clusters are calculated. Between the strong clusters, some rare nodes with high key-values may be found. In addition, researchers get clues from these rare nodes and the connections with strong words’ clusters. Some chances may be discovered through some scenarios generated by them. Ohsawa (2003) used such a method to predict a new style of cloth, which might be accepted by majority. This new cloth was made and put on the market. A big sale of the cloth proved that the method worked.

Wang and Ohsawa (2013) proposed a new way to discover innovative ideas. Their method, which originally came from Ohsawa’s KeyGraph mechanism, assisted enterprises

![Figure 1. The conceptual diagraph of (a) trend analysis and (b) chance node.](image)
to make good decisions. In their decision-making process, it clearly showed that the quality of decisions heavily depend on team members’ knowledge and experiences. The problem of this kind of analysis is that the selected rare words’ frequency is near the threshold frequency, which means that all these rare words will be soon in a strong words’ network and will be known to the majority. The purpose of this study was to try to discover the first stage of CDME, not the second stage of CDME, which Ohsawa’s Key-Graph mechanism mainly focuses.

The leader users/early adopters are very few in number; so it seems to be a rather difficult job for researchers to find them in the real world. Fortunately, in Web 2.0, the consuming values are easy to share by posting people’s ideas on the Internet (O’Reilly, 2007). A report on big data published in 2011 by the McKinsey Global Institute declared that there was an increase in data by 7000 PB (petabyte) in 2010. Big data include the data of browsing websites, virtual social communication, weblog articles, weather data and transaction data. Therefore, it is not hard to image that to extract early adopters’ useful innovative ideas from the Internet’s big data is definitely not an easy task, but their innovative ideas exactly exist on the Internet (Hong et al., 2016).

However, Shannon’s information theory pointed out that if a word appears in almost all of the documents, it is not valuable to reveal any information. For instance, in TF-IDF method, the word ‘the’, which appears in almost every document, is not a meaningful word in the process of analysis. On the contrary, if a word appears only in a few documents, it can be considered as an important word. In addition, at the early adopters’ initial creation stage, only few documents are supposed to be uploaded to the Internet. So, the traditional low-support but high-confidence mechanisms hardly discover incipient and innovative ideas. Because rare words have low frequency and appear only in few documents, their confidence is also very low. Therefore, owners of an enterprise want to find early adopters’ incipient and innovative ideas and attract more consumers to buy new products. First, they have to focus on words with low support and low confidence, which is impossible to deal with using any traditional analysis method. In order to explore innovative ideas, it is a critical task in this research to invent a new method capable of handling words with low support and low confidence. In the next section, our method is discussed.

3. Framework of incipient innovative ideas discovery

Most of the above models are methods associated with low support and high confidence. If innovative ideas come from rare words of which frequencies are very close to clusters’ threshold, unfortunately, the ideas will soon be known and used by some companies. In this case, these innovative ideas cannot be the real chance for companies to generate products and sell on the market. Analysis methods based on a dipolar thinking perspective are targeted on to find low-frequency words with low support and low confidence. In this paper, a new framework with a method using low TF, low TF-IDF and low TF-ICF was developed. The process to sort out innovative ideas is shown in Figure 2.

3.1. Develop concepts

According to Rogers’ IDM, the early majority has above-average social status, having well contact with early adopters. On the contrary, the late-majority is typically sceptical about
an innovation, having below-average social status and little financial liquidity, because they have little contact with other members in late-majority or early majority. Naturally, their consuming behaviours are very different from others. After collecting and clustering their consuming data, the results will form strong consuming words’ clusters (Price & Ridgway, 1982). In this section, the researchers cluster consuming words to identify the majority’s consuming concepts.

Phase 1: Prepare data and initialize clustering analysis.
Step 0: Data preprocess.
0-1) the researcher defines the domain and relevant key words he/she intends to study.
0-2) the researcher sorts out the data which correspond to keywords from the Internet.
0-3) based on his/her domain knowledge, the researcher interprets the texts, and at the same time, segments texts into words, removes useless words and marks meaningful words as conceptual labels.

Step 1: Word co-occurrence analysis to emerge the clusters.
1-1) use Equation (1) to calculate the association values of all words as below:

\[ \text{assoc}(w_i, w_j) = \sum_{s \in \text{all } D} \min(|w_i|_s, |w_j|_s), \]

whereas \( s \) stands for the co-occurrence of words in the sentence, and \( D \) stands for all the textual data.
1-2) decide on the threshold value of word frequency and link value to remove rare words and low linkage between words.
1-3) search for the linkage to each other words to let the clusters emerge.

3.2. Find rare concepts/terms
According to Rogers’ IDM, there are only few users of the innovative products in the incipient stage. The users are so-called early adopters whose innovative messages passing on to early majority will be even fewer. Furthermore, there are only extreme few useful ideas in a document. Shanon’s information theory pointed out that if a word appears in almost
all the documents, it is not valuable to reveal any information, such as the word ‘the’. In order to reveal incipient ideas of early adopters, a new framework is created. How the framework can effectively extract the externalizing incipient concepts is described in the following of this paper.

Phase 2: TFICF definition

As the first stage of innovation, only very few documents of early adopters are uploaded to the Internet. By putting proper key words in Google Blog, documents related to the subject can be easily collected and the authors of the documents can always be identified. Under this circumstance, rare words, which appear only in few documents and therefore have low frequency, may be important because some innovative ideas relevant to neo-market are hidden in these terms. Another essential feature is that these rare words can only connect to a small number of word clusters.

\[
\text{ICF}_i = \frac{M}{|\{m \in M; rw_i \in m\}|}
\]

\[
\text{TFICF}_i = \text{TF}_i \times \text{ICF}_i
\]

with \(M\) being total number of clusters.

Hence, \(|\{m \in M; rw_i \in m\}|\): number of clusters which rare word \(i\) connects with. If the rare word is not in the clusters, this rare word will be dropped.

3.3. Remove extreme rare concepts/terms

In Equation (2), it consists of TF (term frequency) and ICF (inverse cluster frequency). If term frequency or inverse clusters frequency of a word are extremely low, the value of TF-ICF is low.

Phase 3: Perform TF analysis

To further distinguish the words, we may count the number of times each word occurs in a document, which is its word frequency. In the case of the word frequency \(\text{tf}(i,d)\), the simplest choice is to use the raw frequency of a word in a document, that is, the number of times that word \(t\) occurs in documents. If we denote the raw frequency of \(t\) by \(f_{idr}\) then the simple tf scheme is \(\text{tf}(i,d) = f_{idr}\).

\[
Tf_i = \sum_D tf_{ij} = \sum_D \sum_k \frac{n_{ij}}{n_{kj}}
\]

In the above algorithm, \(n_{ij}\) stands for frequency of the word appears in the document \(d_j\), and the denominator is the total number of the word that appear in document \(d_j\).

Phase 4: Perform TF-IDF analysis

Because the word ‘the’ is so common, word frequency tends to incorrectly emphasize documents which happen to use the word ‘the’ more frequently, without giving enough weight to the more meaningful words. Hence, an inverse document frequency
factor is incorporated which diminishes the weight of words that occur very frequently in
the document set and increases the weight of words that occur rarely.

\[
\text{IDF}_i = \log \frac{D}{|\{d \in D; i \in d\}|}
\]

\[
\text{TfIDF}_i = \text{Tf}_i \times \text{IDF}_i
\]

with $|D|$: total number of documents, $|\{d \in D; t \in d\}|$: number of documents where
the word $t$ appears (i.e. $\text{tf}(t,d) \neq 0$). If the word is not in the corpus, this will lead to division
by zero. It is therefore common to adjust the denominator to $1 + |\{d \in D; t \in d\}|$.

### 3.4. Create innovations scenario

In this phase, the lists of lowTF and lowTF-IDF are used to remove useless words from the
list of lowTF-ICF, the remaining words in the list of lowTF-ICF can be used to estimate
researchers to create innovative business scenario.

**Phase 5: Process to find innovative words**

5-1) Sort all words with their TF values by descending and keep top 20 words with the
highest TF in the list of lowTF.

5-2) Sort all words with their TF-IDF values by descending and keep the top 20 words
with the highest TF-IDF in the list of lowTF-IDF.

5-3) Sort all words with their TF-ICF values by descending and keep the top 20 words
with the highest TF-ICF in the list of lowTF-ICF.

5-4) If any word in the lists of lowTF-IDF or lowTF-ICF happens to be in the list of lowTF, it
is removed from the first lists of lowTF-IDF or lowTF-ICF.

5-5) If any word in the lists of lowTF-ICF happens to be in the list of lowTF-IDF, it is
removed from the first lists of lowTF-ICF.

At last, the remaining words in the list of lowTF-ICF are the innovative words.

### 4. A case study

Going abroad to enjoy leisure life is becoming one of the most popular activities in Taiwan.
The number of international Taiwanese travellers has increased since 2006. Some ques-
tions therefore may arose – for instance, what services do airline travellers expect? And
what would be the important business problems for airlines? The main research
purpose of this paper was to explore effective services for airline travellers, and how air-
lines could supply theses good services to customers to win more business.

In this section, by putting data that were collected in 2008 on the Internet into a
business scenario, a case study was presented. There were two stages, Socialization and
Refinement, to learn the characteristics of Taiwanese passengers by performing clustering
analysis and to remove extreme words from the experiential data. Words extracted by the
proposed framework were presented here. Finally, the feasibility of the framework can
therefore be checked by experts.
4.1. Stage 1 socialization: perform clustering analysis

In this stage of socialization, a process using text mining is applied to learn the characteristics of Taiwanese passengers and discover their activities. The aim of this process is to ensure that people engaging in these activities behave in a way that is acceptable to Taiwanese society. In order to investigate major activities for travellers during transport time, six months after notebooks were newly introduced to the market, from 1 October 2008 to 31 October 2008, Keywords, travel + flight + train + kill-time, were searched in Google blogs (http://blogsearch.google.com/blogsearch). When 330 blog articles were collected and carefully read, some of them irrelevant to travel life have been removed. Eventually, there were only 48 articles remained. These articles were put into the co-occurrence analysis by running formula (Equation (1)). The majority’s consuming concepts were discovered, as shown in Figure 3.

The collected articles were posted by travellers who can be roughly divided into two groups. The first group, national passengers travelling by train, enjoyed video and audio entertainment on the trip. The other group is international passengers, who went abroad by airplane. When they were on the air, they also enjoyed video and audio entertainments just like the members in the first group. During the time of the long trip, they had food served by flight attendants, which the first group did not.

As shown in Figure 3, the high-frequency words relate to leisure are video, TV, food, meal, and drink. One particular high-frequency word (sleep) was a common activity for international travellers, when they had nothing to do on the plane. This will be a huge problem for service providers because no one desires for any new invented activity as long as people fall asleep. On the other hand, there may be a chance for enterprise owners, if they can provide a new service special for these international travellers while they sleep on the airplanes.

4.2. Stage 2 refinement: remove extreme words

In this stage of refinement, a process of removing unwanted terms was performed to make promising concepts clear. Therefore, researchers can more easily decide important words. The value of TF-ICF was calculated using Equation (2). The TF indicating frequency of a word was derived by using Equation (3). At the end, Equation (4) was used to compute TF-IDF. In Table 1, there are 20 words in each lists of lowTF, lowTF-IDF and lowTF-ICF. They were sorted by descending with TF, TF-IDF and TF-ICF individually. In text-mining analysis, the higher frequency words are the most important words. On the contrary, from the more future-oriented chance/neo-market perspective, the products relating to the higher frequency words have higher possibility to be quickly accepted by the majority. Therefore, companies have no time to use the idea to develop new products for the market. Consequently, the 20 words in the list of lowTF may not be more promising to bring good chance for businesses. And for this very reason, the words with high frequency were not in the lists of lowTF-IDF and lowTF-ICF.

Derived from Equation (4), the words with the low IDF were from only one document. That means these words were from just only one early adopter and only very few people in the early majority would actually perceive the value related to the words. So the words with lowest IDF must be removed from the list of lowTF-ICF.
As a consequence, the remaining words on the list of lowTF-ICF came from at least two documents. Although these low-frequency words in Table 1 connected with very few words’ clusters and people may argue that they cannot pass strong value to the people in the early majority, the rare concepts behind these words may be popular in a neo-market and full of potential to bring useful value to enterprise owners for creating new business chance in the future.

Finally, the feasibility of the proposed framework to extract the innovative ideas provided by the early adopters was checked by three experts, who have worked in the management, business and finance fields for over 10 years and were invited to take part in the research. After Figure 3, Table 1 and Table 2 were shown to these experts; they found some words indicate that the passengers used the entertaining equipment and ate foods on their trip. They also realized that some travellers enjoyed the entertainments on flight or train. For words such as PDA and IPOD, it was not easy to attract any enterprise owners to invest on any businesses related to them because a considerable amount of money is needed to develop such businesses and required PDA or IPOD technologies at that time.

Some unusual words, such as Skin, Eye, Air and Moisture Emulsion, were found in the list of lowTF-ICF. These words drew a lot of attention of, and stimulated, these experts, who believed not only females but also males having the problem of dry skin during travelling. It is easy to create a story: when female consumers want to keep skin moistened, they bring stylish emulsion to avoid the problem of dry skin. This led to a more interesting observation: eight articles about travelling male emulation were in 2008 on the Internet. Unlike female travellers, most of the male consumers did not have emulation using experiences, even though they knew their skins being dry and having trouble with it. Since it was
a great chance, the experts’ suggestions such as ‘producing special package of moisture emulsion for male passengers when they on the fight or train’ were sent as new business ideas to some companies.

The idea of moisture emulsion for male passengers was not common in 2008, which was the time when the experience took place. As time pass by, a wide variety of moisture emulsion products for male passengers during travelling are popular in the market nowadays. The results pointed to the fact that the new framework using text mining is able to find rare and useful words to be future innovation for business success in the market.

5. Discussion: comparison of the results with chance discovery’s results

The purpose of this research is to identify ideas which are incipient and innovative. Since incipient ideas only exist in the articles which are newly posted on the Internet, the frequency of the key words relating to the incipient ideas must be extremely low. As for innovative ideas, the key words of these innovative ideas have usually linked with strong clusters. According to the theory of chance discovery, the higher a word’s key-value is, the stronger its connected cluster is. The key-value for every word was calculated and ranked here. By doing so, the words connected to strong clusters can be found. The top 20 words with highest key-values are listed on the left-hand side of Table 2.

Most of the results of chance discovery, the top 20 words with highest key-value, were among the top 108 in the list of TF and made not much difference with the results of TF.
When comparing with the results of the proposed framework, although the outcomes of chance discovery method sifted out more words, only three words, Charger, Battery and Tram, ranking 66, 67 and 68 individually, can be classified in rare words. The first two words, Charger and Battery, have high key-values because they are computer peripheral products and strongly connected to computer-related words’ networks, such as entertainments’ cluster. And so is the word Tram, for it is strongly connected to travel clusters. Unfortunately, all the words extracted by chance discovery cannot be the candidates of future ideas for the market, because they are words with high frequency and will be known by the majority in the very near future.

On the contrary, in the proposed framework, there are four words: Skin, Moisture emulsion, Blanket and Air, ranking 83, 84, 86 and 89 individually, can be classified in rare words and may be used as creative ideas to develop new products for the market of male passengers. Words such as Eye, even only ranking 48, caught the experts’ eyes and was put into their design. The experimental result evidences that the results in the new framework are more effective than the outcomes of chance discovery method to sift out incipient ideas.

### 6. Conclusion

This study, adopting dipolar prospect in text mining, aims to explore incipient and innovative ideas, which will be applied to innovations for enterprise. While traditional approach pays attention to high-frequency words with low support and high confidence, this research developed a framework to emphasis terms with low support and low confidence. As long as proposed words are identified, the values of TF, TF-IDF and TF-ICF of each term are calculated and used to extract early adopters’ incipient and innovative ideas. In the case study, using Google search, the researchers discovered some useful words with features of low frequency, high inversed documents and high inversed clusters. Inspired by these words, some innovative ideas were generated and provided enterprise owners a future business chance to success.

### Table 2. Comparison of results from chance discovery.

| Outcomes of chance discovery | Key-value | Ranking | TF Ranking | Results of the proposed framework | TF Ranking |
|-----------------------------|-----------|---------|------------|-----------------------------------|------------|
| Media entertainment         | 0.1320    | 1       | 11         | Eye                               | 48         |
| Ticket                      | 0.1217    | 2       | 18         | Skin                              | 83         |
| Flight                      | 0.1200    | 3       | 4          | Moisture emulsion                 | 84         |
| Space                       | 0.1131    | 4       | 17         | Blanket                           | 86         |
| Magazine                    | 0.1123    | 5       | 21         | Air                               | 89         |
| Charger                     | 0.1113    | 6       | 76         | PDA                               | 55         |
| Tram                        | 0.1098    | 7       | 78         | IPOD                              | 58         |
| Windows                     | 0.1092    | 8       | 43         |                                   |            |
| Film                        | 0.1090    | 9       | 29         |                                   |            |
| Wireless                    | 0.1089    | 10      | 13         |                                   |            |
| Battery                     | 0.1037    | 11      | 77         |                                   |            |
| Food                        | 0.0979    | 12      | 5          |                                   |            |
| Computer                    | 0.0906    | 13      | 44         |                                   |            |
| Newspaper                   | 0.0860    | 14      | 20         |                                   |            |
| Trunk                       | 0.0833    | 15      | 8          |                                   |            |
| Japanese drama              | 0.0700    | 16      | 49         |                                   |            |
| Cards                       | 0.0693    | 17      | 49         |                                   |            |
| Inflight entertainment      | 0.0662    | 18      | 6          |                                   |            |
| APPLE IPOD TOUCH            | 0.0659    | 19      | 38         |                                   |            |
| Remote controller           | 0.0652    | 20      | 39         |                                   |            |
Disclosure statement
No potential conflict of interest was reported by the authors.

Notes on contributors

Chao-Fu Hong received a PhD degree in Electric and Information Science from Yokohama National University, Japan, in 1993. His research specialties are Evolutionary Computing Algorithms, Creative Design, Data Mining, Chance discovery, Multi-objective Decision, Value focused thinking, social network, and Innovative Service. He is Professor in Information Management, and from 2015 he became Dean of the College of Management, Aletheia University, Taiwan.

Mu-Hua Lin received a PhD degree in Management Information Systems from National Chengchi University, Taiwan in 2016. Her research interests are Discovering the Business Opportunities, and Innovative Service from web blogs. She is currently an analyst in the Institute for Information Industry.

Pen-Choug Sun received a degree in Computer Science from Tamkang University, Taiwan in 1988, an MSc degree in Computer-Based Information Systems from the University of Sunderland, U.K. in 1991, and a PhD degree in Computer Science from Coventry University, U.K. in 2012. In 1983, he joined the Computer Centre of Tamsui Oxford College, Taiwan, as a Technician, and became a Lecturer and an Associate Professor at Aletheia University, Taiwan, in 1992 and 2013, respectively. Since August 2014, he has been the chair of the Department of Information Management at the University. His current research interests include Innovation Diffusion Model, Chance Discovery, Text Mining, E-Commerce, Group-Trading System, Broking System and Big Data Analytics.

Hsiao-Fang Yang was a postdoctoral researcher in the Department of Accounting at the National Chengchi University, Taiwan, from November, 2014 to August, 2017. She received her PhD degree from the Department of Management Information Systems of the National Chengchi University, Taiwan, in 2014. Her research interests are in the areas of human–computer interaction in product design, recommender mechanism in electronic commerce, and data/text mining of knowledge discovery in the capital market.

References

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). Mining association rules between sets of items in large databases. In SIGMOD conference, 207–216. doi:10.1145/170035.170072

Chiu, T. F. (2014). A proposed IPC-based clustering method for exploiting expert knowledge and its application to strategic planning. Journal of Information Science, 40(1), 50–66. doi:10.1177/0165551513507404

Chiu, T. F., & Hong, C. F. (2015). Recognizing and evaluating the technology opportunities via clustering method and Google Scholar. In N. Nguyen, B. Trawiński, & R. Kosala (Eds.), Lecture notes in artificial intelligence, 9012 (pp. 159–169). doi:10.1007/978-3-319-15705-4_16

Hirschman, E. C. (1980). Innovativeness, novelty seeking and consumer creativity. Journal of Consumer Research, 7(3), 283–295. doi:10.1086/208816

Hong, C. F., Lin, M. H., & Lin, W. T. (2016). The framework of discovery early adopters’ incipient innovative ideas. In N. Nguyen, B. Trawiński, H. Fujita, & T. P. Hong (Eds.), Intelligent information and database systems, ACIIDS 2016, lecture notes in computer science (Vol. 9622, pp. 319–327). Berlin: Springer. doi:10.1007/978-3-662-49390-8_31

Lin, M. H., Hong, C. F., & Lin, W. T. (2017). Innovative chance discovery for finding innovative useful idea in travelling. International Journal of Intelligent Information and Database Systems, 9(3–4), 269–288. doi:10.1504/IJIIDS.2016.081601

Martin, D. M., & Schouten, J. W. (2014). Consumption-driven market emergence. Journal of Consumer Research, 40(5), 855–870. doi:10.1086/673196

Moore, G. A. (1991). Crossing the chasm – marketing and selling technology products to mainstream customers. New York, NY: HarperCollins.
Morrison, P. D., Roberts, J. H., & von Hippel, E. (2000). Determinants of user innovation and innovation sharing in a local market. *Management Science, 46*(12), 1513–1527. doi:10.1287/mnsc.46.12.1513.

O’Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies, 65*, 17–37. http://www.oreilly.com/pub/a/web2/archive/what-is-web-2.0.html

Ohsawa, Y. (2003). Chance discovery: The current states of art. In Y. Ohsawa & P. McBurney (Eds.), *Chance discovery* (pp. 3–20). Berlin: Springer Verlag.

Ohsawa, Y., Benson, N. E., & Yachida, M. (1998). KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor. Paper presented at the ADL ‘98: Proceedings of the advances in digital libraries conference, IEEE Computer Society, Washington DC, 12–18. doi:10.1109/ADL.1998.670375

Price, L. L., & Ridgway, N. M. (1982). Use innovativeness, vicarious exploration and purchase exploration: Three facets of consumer varied behaviour. In Bruce Walker (Ed.), *American marketing association educator’s conference proceedings* (pp. 56–60). Chicago, IL: American Marketing Association.

Robertson, T. S. (1971). *Innovative behavior and communication*. New York, NY: Holt, Rinehart and Winston.

Rogers, E. M. (2003). *Diffusion of innovations*. New York, NY: Free Press.

Shane, S. (2003). *A general theory of entrepreneurship: The individual-opportunity nexus*. Cheltenham: Edward Elgar.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review, 25*(1), 217–226. doi:10.5465/AMR.2000.2791611

Wang, H., & Ohsawa, Y. (2013). Idea discovery: A scenario-based systematic approach for decision making in market innovation. *Expert Systems with Applications, 40*(2), 429–438. doi:10.1016/j.eswa.2012.07.044