IIDM$_{CC}$: An Innovation Idea Discovery Model Using Online Customers' Complaint Messages

Shu-Chen Kao*

Department of Information Management, Kun Shan University, Taiwan
kaosc@mail.ksu.edu.tw

Abstract

Online customers' complaints have attracted increasing attention to innovation developers. By applying text mining and classification-oriented data mining techniques, an innovation idea discovery model using online customers' complaint messages (IIDM$_{CC}$) was proposed and implemented in this article. Methods included text mining to derive bags of words, sparsity exclusion to produce a term matrix, and supervised classification data mining to reveal decision rules. The IIDM$_{CC}$ showed 90.63% prediction accuracy based on 14720 complaint messages collected from official forum and online communities of a case company in the mobile phone sector from Taiwan. Validation of data inputs, method, and outputs was conducted via case company specialists. The article concludes that analyses of online complaint messages may potentially contribute to the exploration and discovery of innovation ideas. The paper demonstrates the use of mining open textual data in general and complaint messages in particular in the domain of knowledge discovery in databases.

Keywords: Complaint messages, Text mining, Data mining, Term matrix, Bag of words

1 Introduction

In innovation, idea discovery is the key determinant of value creation for products and services [1]. Knowledge acquisition has been a relevant approach that involves acquiring and applying pre-existing knowledge to develop and create innovative ideas. Despite the various proposed mechanisms, enterprises are still actively searching for new knowledge sources and discovery mechanisms to maintain competitiveness. In this context, recent focus on open innovation has attracted increasing attention to the role of end users in exploring innovative ideas [2-5]. User innovation, a proposed user-centered mechanism based on analyses of opinions, comments or suggestions, was introduced as a possible way of improving products and services [6].

Advanced social media provides efficient channels such as forums, online communities and fan pages that allow online users to share their views, thereby providing crucial feedback for companies. For example, the programs of Dell IdeaStorm and Samsung Galaxy Note 5 (PD-SGN5) were created to stimulate users to freely share new ideas via social media, and have played an active role in the formulation and adoption of innovations [6-7].

Previous studies indicated that extracting clues from online communities can assist in idea discovery [1, 8]. In particular, customers' complaints have received extensive attention by providing information implicitly reflecting their evaluation of services or products. Despite their predominantly negative character, complaints offer useful insights into user problems or expectations, such as what problems users encounter and need to be solved [1], how existing service processes can be improved [9], how to increase organizational learning performance [10], and how stronger customer relationships can be maintained or enhanced [11]. However, there are still rooms to be improved from the perspective of open innovation based on web materials, such as how online complaint messages can be explored to help discover innovative product and service ideas.

Based on text mining and classification-oriented mining techniques to derive decision trees and rules, a knowledge discovery model (IIDM$_{CC}$) is proposed that analyzes online customers' complaint messages to explore innovative ideas. The model inputs, operations (idea discovery mechanism), and outputs were presented in detail. The proposed model was tested and evaluated by using online complaint messages from a mobile phone case company from Taiwan.

2 Literature Review

2.1 Innovation Resources

In order to increase innovation capability, enterprises have made use of internal or external resources to perform innovation tasks [12]. In particular, explicit knowledge is beneficial for innovation for being easily encoded, stored and transmitted, and characteristics thought to accelerate problem solving and innovation [13]. Social media is a rich source of explicit information about user behaviors, opinions, thoughts, and complaints [14]. The customer-generated, large-scale, and unregulated views on products and services can be used to develop knowledge conducive to innovative product and service opportunities [1, 5, 7-8, 12]. Compared to user's opinions, complaints are directly related to practical experiences and feedback, since they directly address problems or suggestions that provide suitable materials for the development of innovative ideas [1, 10].

2.2 Innovative Ideas Hidden in Customers' Complaints

*Corresponding Author: Shu-Chen Kao; E-mail: kaosc@mail.ksu.edu.tw
DOI: 10.53106/160792642022032302002

Online complaint messages in the form of text are mostly posted to (1) vent emotions resulting from purchase experiences; (2) search for solutions to unexpected problems; and (3) offer personal comments on unsatisfactory products or services [14]. For example, a message may suggest that online stores should allow gift senders to record a video to be watched upon delivery by the recipient scanning a QR code. This message expresses an additional need from a customer and may be the source of an innovative idea [14]. Complaints more often refer to problems faced by customers [10-11], and exploring their potential for idea discovery requires an understanding of (1) suitable techniques to extract knowledge, (2) the processes of knowledge discovery from customers' complaints, (3) ways to present discovered knowledge, and (4) the implications of discovered knowledge to idea development.

2.3 Knowledge Discovery in Customers' Complaints

Current techniques to elicit knowledge in text-form documents include the techniques of text mining combined with classification [1, 5, 8, 12, 15-19], topic (or theme) modeling [4, 7], and Latent Dirichlet Allocation (LDA) [2, 20]. Text mining is a well-known technique that has been successfully applied to article parsing. Unstructured documents can be transformed into a structured format expressed by a matrix of terms and term frequencies suitable to analysis. By adopting classification technique, it has been successfully applied to generating creative topics from user opinions [7, 14], identifying innovative thinking for new product development [1, 7-8, 12], and detecting emerging product and service ideas [4-5, 14]. For example, the valuable ideas are developed in the datasets collected from MyStarbucksIdea.com by employing the techniques of text mining, sentiment analysis, and machine learning [5]. However, verification of output still needs to be addressed.

When analyzing large online textual datasets, understanding classification rules is also crucial [5, 19]. For example, in a term matrix, there are relationships between specific terms ("update" and "suggestion") and potential idea categories (product and service). Since the matrix contains the frequency of terms appearing in the complaint messages, if a certain number of online complaint messages contain the terms "suggestion" and "update" linked to product or service problems, an innovative idea is more likely to be found. To address this issue, the innovative idea discovery model (IIDMcc) is proposed to discover innovative ideas by analyzing online customers' complaint messages.

3 IIDM$_{CC}$ Framework

The IIDM$_{CC}$ procedure and design features are presented in Figure 1 and Table 1. It consists of three parts: Input, Process, and Output. Model input consists of customers' complaint messages collected from official forums and online social communities. Process is divided into three stages. Process A or data preparation performs data cleaning and conversion. Process B applies text mining to the obtained dataset in Process A in order to turn textual messages into structural forms. The outcome is a bag of words and a term matrix. Process C performs model training and prediction testing utilizing a classification-oriented algorithm. Output is the presentation of outcomes in the form of decision tree and decision rules related to idea development. Details are presented below.

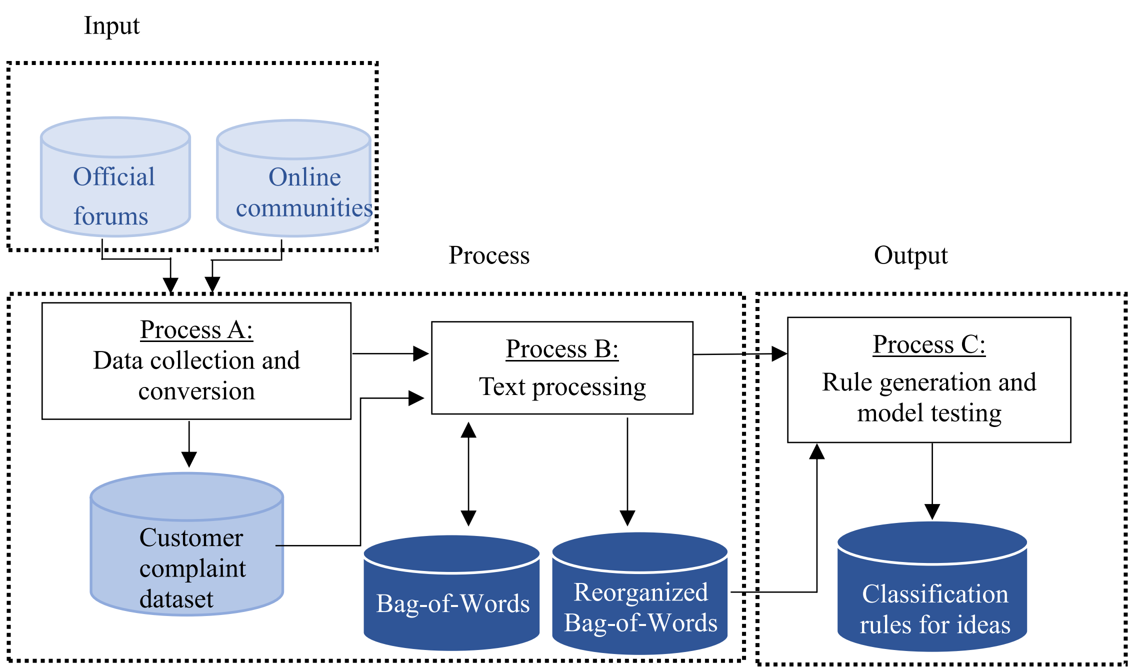

Figure 1. IIDM$_{CC}$ Framework

Table 1. Design features

Features	Description
Objectives	(1) Applicability of mining techniques on idea development (2) Reutilization of customers' complaints for innovative idea discovery (3) Proposing a feasible approach to ideation
Data sources	(1) Complaint messages from official forum of the case company (2) Complaint message from online communities of the case company
Data preprocessing	(1) Combine datasets collected from official forum and online communities (2) Remove messages containing missing data and unnecessary data attributes (3) Label the class attribute with values of 1(product creativeness), or 2 (service creativeness), or 0 (neither) (4) Term matrix is produced and sparsity threshold was set at 60% (5) Frequency of each term is discretized with label 'No', 'Low' or 'High'
Mining mechanism	(1) Text mining (2) Classification-oriented data mining
Mining output	(1) Classification rules as ways to present discovered knowledge (2) Themes associated with complaint messages
Validation and implication	(1) Accuracy rate (2/3 for training and 1/3 for testing) (2) Simplicity of classification rules (3) Implications of the themes derived from the generated rules

3.1 Inputs

Customers' complaint messages related to specific products or services were collected from online communities of the case company and stored in a dataset. Data attributes (or features) include poster name, posting date, textual contents and responses from followers of a specific topic. Each complaint message was stored as a data record.

3.2 Process

3.2.1 Process A: Data Collection and Conversion

The dataset was cleaned by removing complaint messages that contained missing data, and unnecessary data attributes such as poster name and posted date. Each record was then assigned a unique number with its corresponding complaint message content. In order to assess whether complaint messages included any innovative ideas, they were reviewed by three specialists. Class label 1 was assigned to a message if it contained potential ideas for products, class label 2 for services, and class label 0 if neither. The final class label for each message was determined by the majority of the three reviewers, and if not possible, an additional specialist was invited.

The resulting pre-processed CCD (Customer Complaint Dataset) included the three attributes Message ID, Message Contents, and Class Label (Table 2). For example, message M00001 was given the class label 1 for containing an innovative suggestion for product improvement (adding a face+voice recognition function). Message M00002 (class level 0) only shared an experience of fixing a cover glass, whereas M00003 was given class level 2 for suggesting a possible new service idea.

3.2.2 Process B: Text Processing

To transform each complaint message into a structural format, textual contents in CCD were processed using a text mining technique in R language. First, all punctuation marks and numbers were removed. Second, frequent words with no discriminative value were deleted ('I', 'you', 'has', 'have', 'and', 'a', among others). The remaining textual contents were extracted which contained verbs, nouns, adjectives, or adverbs. However, only are nouns and verbs, such as "problem", "charge", "update", "battery", were selected since they represent specific objects, places, stakeholders, emotions, or human values [8].

The outcome of text mining is the CTD (Complaint Terms Dataset) containing mined terms, or its attributes (Table 3). For example, message M00001 contains 23 terms ("want", "activate", "mobile", "sensing", "side frame", "phone", "needed", "unlock", "action", among others). A Bag-Of-Words (BOW) is formed from the mined terms, followed by an Empty Term Matrix (ETM) generated based on the BOW and containing Message ID, the number of terms in BOW, and class level.

Records are then added to the ETM in two steps. The first step appends a new empty record in which Message ID and class level from CTD were generated. The second step counts each term in the message content and adds the result into the ETM. The outcome is the 'developed term matrix' (DTM; Table 4). For example, message M00001 shows the attributes "phone" once and "recognition" twice. The third term "call" does not appear in M00001 but twice in message M00002.

Due to the large number of attributes in the DTM, terms that do not appear very frequently are removed based on a sparsity threshold defined as the ratio of number of non-zero records to total records. Term coverage is then computed and the DTM is sorted in descending order. The granulation was implemented to convert numeric value (frequency of terms) into linguistic value ('no', 'low', or 'high') in the DTM, which is required by the classification-oriented data mining technique. The granulated value is 'high' if a term appears in a complaint message at least three times, 'low' if once or twice, and 'no' if empty. The outcome is the PTM dataset (Produced Term Matrix; Table 5). For example, Term 1 occurs in message M00001 once, and is therefore assigned a linguistic value 'low', and over three times in both M00002 and M00004 where it is assigned a 'high' value. There is no unique way to granulate the DTM as it depends on the real context.

Table 2. Part of customer complaint dataset

Message ID	Message Contents	Class Label
M00001	If I want to activate the mobile phone by sensing the side frame, it is needed to unlock the action before starting the app. Currently, only fingerprint and face recognition is available. I wish company could add the function of face and voice recognition.	1
M00002	My phone cover glass was broken, changing the screen takes more than 3 days. I called the service center sending me a message at the time my phone is fixed. But the phone is on the way, what a stupid call I made!	0
M00003	My mobile phone battery is not normal. I want to send it to the service center for repair. Is there any extra phone for me to use during the fixing period?	2

Table 3. Some mined terms from Table 2

Message ID	Complaint Terms	Class Label
M00001	want, activate, mobile, sensing, side, frame, phone, needed, unlock, action, starting, app, currently, fingerprint, recognition, available, wish, voice, company, add, function, face, recognition	1
M00002	phone, cover, glass, broken, changing, screen, take, asked, service, center, call, phone, send, back, phone, process, delivery, answer, call	0
M00003	mobile, phone, battery, abnormal, want, send, service, center, repair, extra, phone, fix, period	2

Table 4. Excerpt from developed term matrix (DTM)

Message ID	phone	recognition	call	want	activate	app	battery	cover	deliver	…	Class label
M00001	1	2	0	1	1	1	0	0	0	…	1
M00002	3	0	2	0	0	0	0	1	1	…	0
M00003	2	0	0	1	0	0	1	0	0	…	2
…										…	…
M0000m										…	…

m is the size of DTM

Table 5. Excerpt from granulated produced term matrix (PTM)

Message ID	phone	recognition	call	want	Activate	App	Battery	Cover	Deliver	…	Label
M00001	low	high	no	low	low	low	no	no	no	…	1
M00002	high	no	high	no	no	no	no	low	low	…	0
M00003	high	no	no	low	no	no	low	no	no	…	2
…										….	….
M0000m										…	….

Notes: n is the number of terms generated; m is the size of PTM

3.2.3 Process C: Rule Generation and Model Testing

The classification-oriented data mining algorithm ID3 [21] was utilized to extract classification rules from the PTM. The entropy (information gain) of each attribute in its class was calculated and used for the selection of branches of a decision tree. The ID3 algorithm is described below as:

$$I(nc_1, nc_2, \ldots, nc_n)=((-\frac{nc_1}{N}\boldsymbol{log_2}\frac{nc_1}{N})+(-\frac{nc_2}{N}\boldsymbol{log_2}\frac{nc_2}{N}) + \ldots + (-\frac{nc_n}{N}\boldsymbol{log_2}\frac{nc_n}{N})) \quad \textbf{(1)}$$

nc_i: number of records belonging to class c_i，$i = 1,2\ldots n$
N: total number of records

$$\mathrm{E}(Att)= \sum_{i=1}^{m}(\frac{n_{vi}}{N})\, I(n_1v_1c_1, n_2v_2c_2, \ldots, n_nv_nc_n) \quad \textbf{(2)}$$

E(Att): information gain of attribute Att
m: number of outcomes of attribute Att
n_{vi}: number of records of attribute Att belonging to v_i
$n_iv_ic_i$: number of records of attribute Att belonging to v_i and class belonging to c_i

$$\text{Gain } (Att) = (1) - (2) \quad \textbf{(3)}$$

Gain (*Att*): entropy of attribute *Att*

Based on the decision tree, *if-then* rules can be derived such as '*if* term A = high and term B = low, *then* innovativeness type = 1'. Although similar conditions should lead to the same conclusion, when similar conditions result in different classes, the rule generation mechanism may face difficulties. To deal with this issue, the research adopted the strategy of choosing the rules with the highest number of records (defined as the value of the attribute Support). To test the classification model, the research randomly selected 70% of PTM dataset being trained to develop decision rules, whereas the remaining 30% was used to test prediction accuracy.

3.3 Outputs

Output consisted of the generated rules presented in a structured form and stored in the rule base (Table 6) containing Rule identification number (Rule#), Terms, Innovativeness type, and Support. The Terms attribute refers to keywords developed from the PTM. The number of Term attributes depends on the contents of PTM. Innovativeness type is the class level of the corresponding rule, with values 1 (product), 2 (service), or 0 (irrelevant). The Support attribute refers to the rule strength. The rule base can be regarded as an updatable knowledge base that can be provided to companies and used as a robust foundation for idea discovery.

Table 6. Sample of rule base

Rule#	$Term_1$	$Term_2$	$Term_3$	$Term_4$	…	$Term_n$	Innovativeness	Support
R1	high	low	low	no	…	no	1	12
R2	low	no	no	no	…	high	2	9
R3	high	low	no	high	…	low	0	3
…	…	…	…	…	…	…	…	….

4 Model Testing and Evaluation

4.1 Data Collection and Pre-processing

The web crawling techniques were applied to collect large-scale social media data from the case company's official forum (https://community.htc.com/tw/chat.php? type=product) and other online communities (https://www.mobile01.com/topiclist.php?f=566). The collected dataset contained 14,746 textual messages (complaints and comments) from 2016/02/18 to 2019/8/31. It was pre-processed by removing message records that contained missing data and unnecessary data items, such as poster's name and posted date, resulting in 14,720 message records in total. Each message record was assigned a unique number, reviewed by specialists and given an innovativeness type class level (Table 2). The review task took around three months to complete. The size of collected dataset was properly used to demonstrate and test the proposed IIDM$_{CC}$ model. A practical survey from the case company was conducted to verify the proposed model since outcome discovered from outdated complaint data may not be able to properly fit to the actual needs of the case company.

4.2 Text Processing

Text mining using R software turned each complaint message in the CCD into a structural textual content to generate the PTM. To maintain term majority to simplify the classification task, the sparsity threshold was set at 60%, so that terms appearing in complaint message records less than 8832 times were removed from the PTM. After thresholding, 846 terms remained in the updated PTM and were used in further analyses.

4.3 Rule Generation and Model Testing

To implement rule generation, discretization was performed to convert the term frequency (numeric value) into linguistic values 'no', 'low' or 'high'. The 70-30% criterion for database division was applied to randomly select 10,304 complaint messages for rule generation (training), and the remaining 4416 to test the model.

The generated decision rules reveal that 'suggest', 'resolve', 'wish', 'launch', among others hold the highest power to discriminate class levels and therefore to predict innovativeness. Other terms ('problem', 'repair', 'staff', 'complain', 'physical', 'sales', 'end', 'price', 'entertainment', 'disappoint', 'update') also predicted innovativeness. An example of generated *if-then* rules was "*If* Suggest is [High] and Customer is [High] *then* Innovativeness type is [Service creativeness]".

A total of 395 rules were generated and stored in the rule base. They were tested with the remaining 4416 (30%) messages in the dataset. The testing result is presented in Table 7 that was derived from the outcome of software used. Of the 4416 testing samples, the corrected samples for three classes (i.e., No_creativeness, Product_ creativeness, and Service_creativeness) are 3845, 139, and 18, respectively, which produce a 90.63% prediction accuracy (i.e., (3845+139+18)/4416=0.9063). Despite that the proposed IIDM$_{CC}$ model does not present a perfect prediction accuracy, it is considered not unacceptable.

4.4 Model Evaluation

IIDM$_{CC}$ was verified in two steps, usefulness of the generated rules and applicability of the model by linking the model to the complaint messages. Both verification tasks were carried out by case company specialists.

First, generated rules with descending support were summarized and verified (Figure 2). For example, rule #52 has the highest support from complaint messages (154 relevant message records regarding products), followed by rule #11 supported by 85 message records. Second, rules were summarized into themes and evaluated by case company specialists for relevance to idea discovery [7]. As an example,

Figure 3 shows summarized themes and their associated complaint messages for rule #11 (*if* [Suggestion] in [Low] and [Wish] in [Low] *then* Class in [Product creativeness]).

In Figure 3, theme #1 ("System update issues") derived from six complaint messages (Messages #77, 6453, 9379, 9428, 10264, 12331), reveals four main issues: (1) security patches not being updated, (2) problems caused by system updates, such as redundant replies, candidate words flashing, Sonybravia75 linking failure after 10 minutes, message delay in LINE, and unstable voice volume, (3) restore boost+ function able to lock immediately right after depart, and (4) camera update failure. Company specialists then provided comments on themes and issues revealed by the rules to verify whether $IIDM_{CC}$ outputs could support innovative idea development.

To be more focused, only nine rules with support level of at least 8 (Rule #52, 11, 6, 37, 8, 34, 5, 6, and 9) were selected to develop review outline used for reviewers to provide comments about rule generation process (i.e., input, method, and output) and rule usefulness (i.e., representation and theme associations with idea development). Review outline also contained interview purpose and background information of the proposed $IIDM_{CC}$, classification-oriented idea exploration, and rule generation.

Invited specialists reviewed and marked each question as recognized or not recognized, and wrote a concise report for each rule. Backward thinking was applied to assess innovativeness since collected complaint messages usually represent the post-use experiences with regard to products and services. Evaluation feedback is summarized and discussed.

5 Discussion and Implications

First, the reviewers supported the research argument that model outputs may show stronger associations with products and services than possible in the case of general messages, although limited to incremental innovation ideas. Some specialists praised the selected input data and proposed model as sources of innovative ideas. However, one specialist reported that cost of collecting data may be an issue for companies interested in learning about concerns by customers.

Second, the use of text mining and classification techniques to generate rules from complaint messages was generally supported as an efficient way of detecting latent innovation ideas. It can be expected to help enterprises to mitigate the problems of information overloading. However, a major drawback identified is that rules with conditions and consequences may not fully represent the meaning of original customers' messages, especially given that the model attempts to extract majority rules from highly unstructured online complaint messages.

From the systematic viewpoint, defining processes leading from data collection, pre-processing and rule generation to theme summarization would be of great value, but theme summarization was not covered by the proposed model. This effort should take the shape of embedding the proposed model in a theme exploration and idea generation system, which would specify a function to access generated rules and their associated original complaint messages and summarizing themes.

Third, specialists accepted the output of the proposed model including decision tree and *if-then* rules. However, their feedback also revealed that outputs were in practice weakly linked to the generation of innovative ideas. To reply to the feedback, the research pruned the generated decision trees so that users can concentrate on the rules with strongest support. Since it is hard to interpret the innovation value of a rule only considering the ordinal values 'no', 'low', 'high' designed to represent importance levels of a term, rules should also be explained in the light of more detailed interpretation that addresses type of innovation (service or product), associated terms and their ordinal levels (customer, "high"), and linked messages. This procedure may turn implicit machine presentation into explicit meaning that people can more easily grasp.

Finally, the proposed $IIDM_{CC}$ can be recognized with the capability of supporting idea development. Moreover, a review comment stated that "*Looking back to the past, our product development and improvement were highly associated with the discovered themes, such as suggestion of additional function, camera functions, and hardware enhancement*". However, some reviews stated that model-generated rules are still far from being able to produce actual innovative ideas. A specialist claimed that generated rules can be at best a complementary mechanism allowing companies to filter out most concerns and suggestions from customers. In this sense, discovering sources for innovative idea development will always be a way the proposed $IIDM_{CC}$ can play a supportive role for the innovative idea development teams.

Practically, for any company who plans to utilize the proposed $IIDM_{CC}$ to explore innovative ideas from complaint messages should follow the steps: data collection from online complaint messages, category assignment for complaint message records, text mining, term matrix development, data mining (training and testing), validation, and overall rule generation. The detection mechanism embedded with the generated decision rules can be developed and put in service online. The research demonstrates the value of discovering innovative ideas in abundant volume of online complaint messages.

Table 7. Confusion matrix of model test

	No_creativeness	Product_creativeness	Service_creativeness
No_creativeness	3845	86	24
Product_creativeness	233	139	12
Service_creativeness	35	23	18

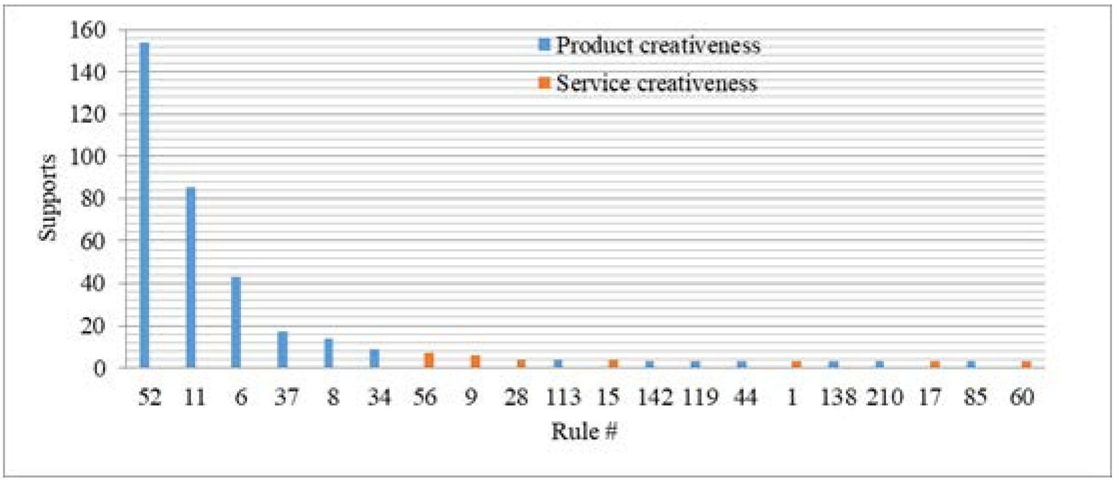

Figure 2. Part of generated rules with high number of supports

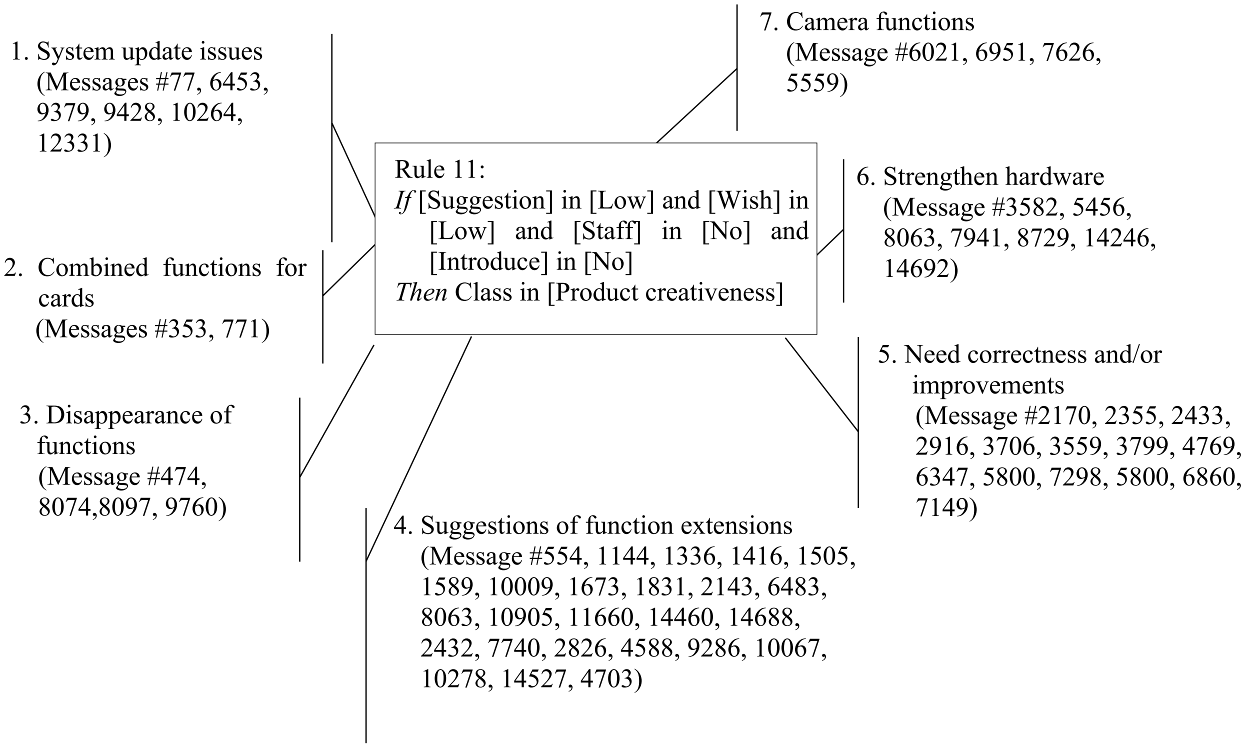

Figure 3. The themes developed from Rule #11

6 Conclusion

Drawing from online customers' complaint messages, the study proposed an innovative idea discovery model (IIDM$_{CC}$) to assist in creative ideas development. Based on text mining and a classification-oriented data mining technique the model showed a 90.63% prediction accuracy. In practice, themes derived from extracted rules must be linked to original complaint messages to facilitate idea generation. The article concludes that use of large online data sources, data pre-processing, knowledge extraction mechanisms, outcome interpretation with structured links to original data, and practical validation of solutions constitute a promising strategy for the generation of product and service innovation when open innovation is concerned.

However, as pointed out by some specialists, the proposed model requires further verification of the practical usefulness of its outputs, and this issue will be addressed in follow-up studies. Furthermore, the study focused on a supervised-oriented learning mechanism to test the proposed model, which involves a high cost in terms of labor required to process and categorize original complaint messages. Future research may consider the use of unsupervised mechanisms such as hierarchical clustering or summarization models.

Acknowledgements

The author thanks Taiwan's Ministry of Science and Technology for partially financial support under the Grant of NSC 108-2410-H-168-004.

References

[1] K. Christensen, S. Nørskov, L. Frederiksen, J. Scholderer, In search of new product ideas: Identifying

ideas in online communities by machine learning and text mining, *Creativity and Innovation Management*, Vol. 26, No. 1, pp. 17-30, March, 2017.
[2] M. Grimaldi, V. Corvello, V. De Mauro, E. Scarmozzino, A systematic literature review on intangible assets and open innovation, *Knowledge Management Research & Practice*, Vol. 15, No. 1, pp. 90-100, December, 2017.
[3] Y. Shin, C. Im, H. Oh, J. Kim, Design for experience innovation: understanding user experience in new product development, *Behaviour & Information Technology*, Vol. 36, No. 12, pp. 1218-1234, August, 2017.
[4] W. S. Lee, S. Y. Sohn, Discovering emerging business ideas based on crowdfunded software projects, *Decision Support Systems*, Vol. 116, pp. 102-113, January, 2019.
[5] H. Lee, K. Choi, D. Yoo, Y. Suh, S. Lee, G. He, Recommending valuable ideas in an open innovation community: A text mining approach to information overload problem, *Industrial Management & Data Systems*, Vol. 118, No. 4, pp. 683-699, May, 2018.
[6] M. Li, A. Kankanhalli, S. H. Kim, Which ideas are more likely to be implemented in online user innovation communities? An empirical analysis, *Decision Support Systems*, Vol. 84, pp. 28-40, April, 2016.
[7] B. Jeong, J. Yoon, J. M. Lee, Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis, *International Journal of Information Management*, Vol. 48, pp. 280-290, October, 2019.
[8] K. Christensen, K. H. Liland, K. Kvaal, E. Risvik, A. Biancolillo, J. Scholderer, S. Nørskov, T. Næs, Mining online community data: The nature of ideas in online communities, *Food Quality and Preference*, Vol. 62, pp. 246-256, December, 2017.
[9] N. Yan, X. Xu, T. Tong, L. Huang, Examining consumer complaints from an on-demand service platform, *International Journal of Production Economics*, Vol. 237, https://doi.org/10.1016/j.ijpe.2021.108153, July, 2021.
[10] C. Yilmaz, K. Varnali, B. T. Kasnakoglu, How do firms benefit from customer complaints? *Journal of Business Research*, Vol. 69, No. 2, pp. 944-955, February, 2016.
[11] A. Ghazzawi, B. Alharbi, Analysis of customer complaints data using data mining techniques, *Procedia Computer Science*, Vol. 163, pp. 62-69, 2019.
[12] Y. S. Chen, J. R. Chang, C. J. Huang, Strategic Decision-making Processes of NPD by Hybrid Classification Model Techniques, *Journal of Internet Technology*, Vol. 21, No. 6, pp. 1635-1646, November, 2020.
[13] E. Ode, R. Ayavoo, The mediating role of knowledge application in the relationship between knowledge management practices and firm innovation, *Journal of Innovation & Knowledge*, Vol. 5, No. 3, pp. 210-218, July-September, 2020.
[14] T. Hellebrandt, I. Heine, R. H. Schmitt, Knowledge management framework for complaint knowledge transfer to product development, *Procedia Manufacturing*, Vol. 21, pp. 173-180, 2018.
[15] V. Kayser, K. Blind, Extending the knowledge base of foresight: The contribution of text mining, *Technological Forecasting and Social Change*, Vol. 116, pp. 208-215, March, 2017.
[16] H. Kwon, Y. Park, Y. Geum, Toward data-driven idea generation: Application of Wikipedia to morphological analysis, *Technological Forecasting & Social Change*, Vol. 132, pp. 56-80, July, 2018.
[17] X. Wang, K. Zhao, S. Cha, M. S. Amato, A. M. Cohn, J. L. Pearson, G. D. Papandonatos, A. L. Graham, Mining user-generated content in an online smoking cessation community to identify smoking status: A machine learning approach, *Decision Support Systems*, Vol. 116, pp. 26-34, January, 2019.
[18] S. Deng, Y. Zhou, P. Zhang, A. Abbasi, Using discussion logic in analyzing online group discussions: A text mining approach, *Information & Management*, Vol. 56, No. 4, pp. 536-551, June, 2019.
[19] J. L. Chen, Y. W. Ma, S. Y. Tsai, Intelligent Classifier for Identify Reliable On-Demand Messages, *Journal of Internet Technology*, Vol. 21, No.7, pp. 1993-1997, December, 2020.
[20] J. F. Yeh, Y. S. Tan, C. H. Lee, Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet allocation, *Neurocomputing*, Vol. 216, No. C, pp. 310-318, December, 2016.
[21] J. R. Quinlan, Induction of decision trees, *Machine Learning*, Vol. 1, No. 1, pp. 81-106, March, 1986.

Biography

Shu-Chen Kao received the MS degree in Information Management from National Chiao Tung University and PhD degree in Business Administration from Nation Cheng Kung University in 1994, 2004 respectively. Her current research interests include knowledge management, intelligent system and social media mining.