A QUANTITATIVE MODEL FOR ANALYSIS OF COLLECTIVE OPINIONS

Jiang-Liang Hou, Chia-Lin Han

National Tsing Hua University, Hsinchu (300), Taiwan

Abstract

This research develops a platform for opinion analysis and user profile determination. Visualization of the relationship between user profiles and opinions can improve reading efficiency and effectiveness of users. On the whole, the proposed methodology and platform that automatically converts free-form opinions into structured opinion components and visualize the relationship between user profiles and opinions, then people can make quick decision about the products and services based on the appropriate opinions of others.

Key words: *opinions, decision support system*

1. INTRODUCTION

People get opinions from customers and evaluate the values of customer's opinions by themselves. However, people must spend much time and efforts to get representative opinions. People may possibly ignore the opinions which are more important and they cannot acquire opinions completely. Moreover, it is difficult for people to analyze various opinions and evaluate the values of distinct opinions, so people cannot make an appropriate consuming decision. The product/service-consuming model is illustrated in Fig 1.

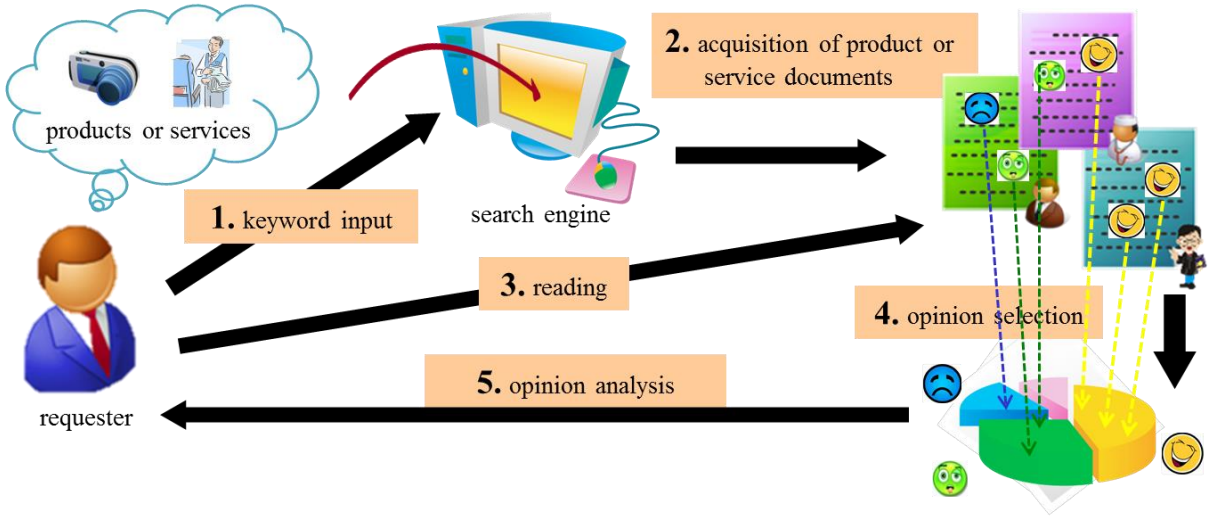


Fig 1. As-is Model.

Our research develops a model for analysis and visualization of user profiles and opinions. The model automatically converts free-form opinions into structured opinion components and calculate the dimension value of opinion tendency, user profile and importance. Then, the model can visualize the relationship between user profiles and opinions. Finally, people can find customers with similar profiles and acquire their opinion quickly. The to-be model in this study is illustrated in Fig 2.

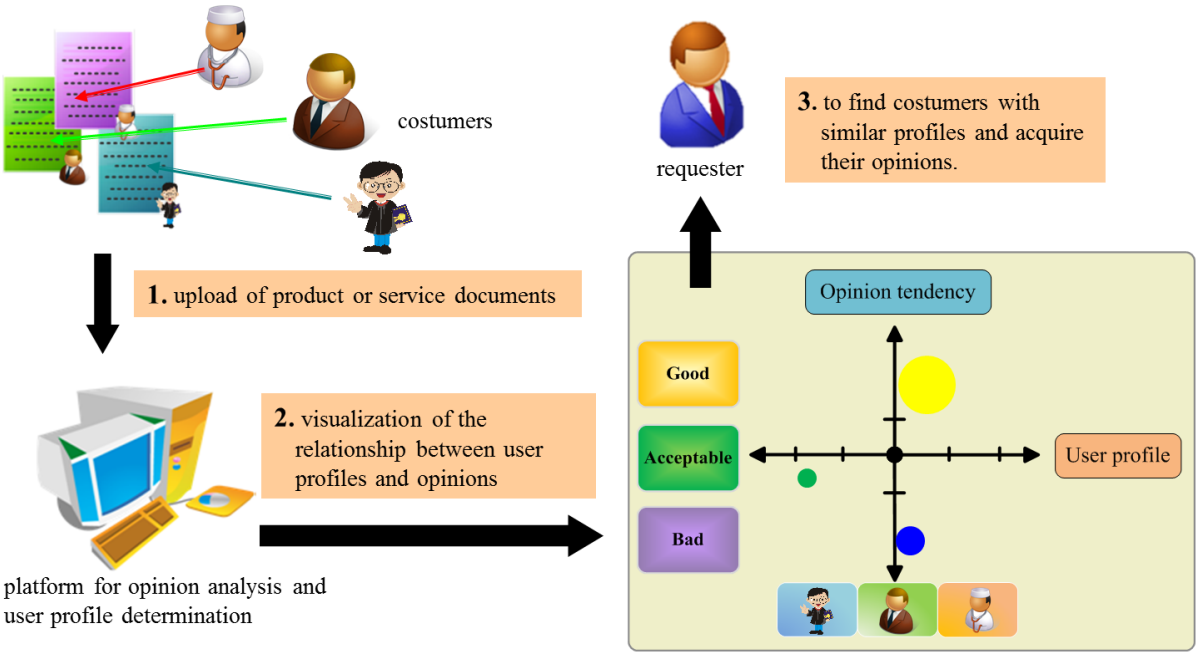


Figure 2. To-be Model

2. A REVIEW OF RELATED WORKS

The research subjects included decision support systems and information visualization technology. The related works about these subjects are reviewed as follows:

2.1. Decision Support Systems

Decision support systems (DSS) often analyze the data which are related to the decision-making and produce the opinions tendency. Then decision-makers can get the recommendations of decision-making and they can make an appropriate decision probably. Moreover, the decision support systems can divide into two main parts based on the different type of decision-maker. There are individual DSS and group DSS. In the individual DSS, the related researches usually analyze and generalize a large number of decision-making data (Gruber (2008), Boulos and Wheeler (2007)). Then the individual decision-maker can make decision based on the analyzed data. In group DSS, the related researches usually integrate the tendency of opinions and show the final result. Therefore, the group decision-maker can make final decision-making (Masona and Rennieb (2007), Rasmussen et al. (2007)).

2.2. Information Visualization Technology

Information visualization technology usually analyzes and generalizes a large number of data. Moreover, the information visualization technology visualizes the key point of data, so people can read data quickly. The information visualization technology can divide into four main parts based on the different development. There are principles for information visualization, tools and techniques for information visualization, performance of information visualization and applications of information visualization. Firstly, in principles for information visualization, the related researches usually concluded the principles of showing data (David (1999), Chittaro et al. (2001)). In tools and techniques for information visualization, the related researches usually developed 2D or 3D techniques to visualize original data and then people can read data effectively (George et al. (1993), Montero et al. (2006)). In performance of information visualization, some related researches often evaluated the performance of information visualization and the others usually concluded the benefit of information

visualization (Kobsa (2001), Tegarden (1999)). Finally, in applications of information visualization, the related researches usually visualized the information of life, security and art (Lange et al. (2003), Heer et al. (2005)).

By comparing with the previous research, previous researches usually do not think about user profiles. However, our research mainly focuses on analyzing user profiles. On the other hand, previous researches usually attended to visualize the quantitative data. Rather, our research mainly analyzes the characteristic of free-form opinions, and visualize the relationship between user profiles and opinions. Therefore, people can find customers with similar profiles and make a decision quickly.

3. ANALYSIS OF OPINIONS CONTENTS

In order to understand components of free-form opinions and the characteristics of these components, this research acquired many product or service documents firstly. After that, this research extracted the sentences which have judgment concept, and modified judgment sentences. Then, this research generalized components of free-form opinions and analyzed the characteristics of these components. There are "extraction of components" and "analysis of component characteristics", the following sections will show details.

3.1. Extraction of components

The product or service documents were acquired from the bulletin board system (BBS). Then this research extracted judgment sentences from product or service documents and generalized components of free-form opinions. The process of component extraction is shown in Fig 3. In this case, people usually report the comments of product or service, and also make a description of reference to support their comments. In conclusion, opinions consist of three main components including opinion tendency, user profile and reference. Furthermore, the three components can be further divided into eleven items.

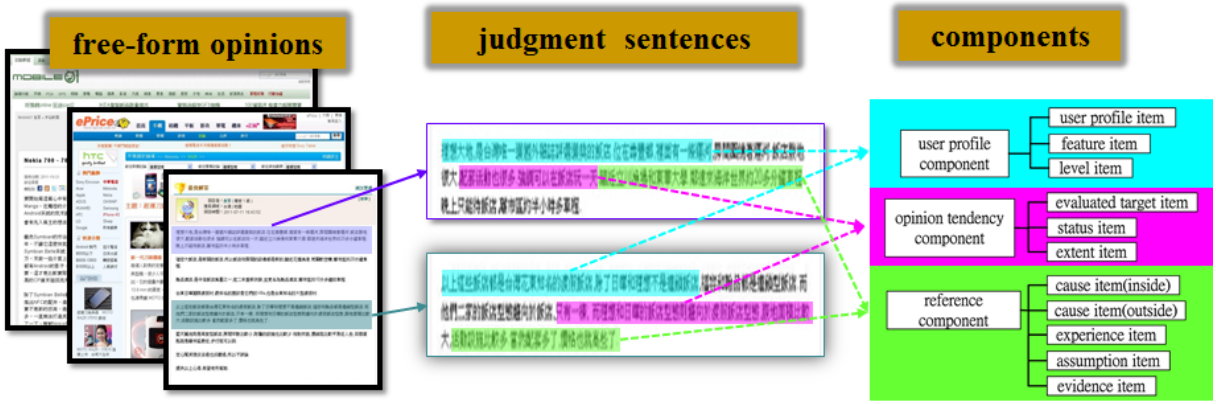


Fig 3. The process of component extraction

3.2. Analysis of component characteristics

After understudied the components of free-form opinions, this research analyzed three main components and got "component structure" and "properties of words". In this case, user profile component usually uses "specific words" to describe user's age, occupation or gender. Opinion tendency component also use "specific words" to describe tendency, such as good, perfect or bad. The reference component possible uses many "combination words" to describe reason of opinion tendency, and usually uses some "initiation words" to describe reference of opinion.

Moreover, this research analyzed properties of words, including status/feature word and grade word.

The status/feature word which can express opinion tendency or user profile has four properties, including feeling, level, measurement or item. The grade word which can change opinion tendency or user profile has three properties, including strengthen, opposite or attenuation.

4. OPINION FORMATTING

In the opinion formatting stage, the free-form opinions would be broken up to target sentences by sign mapping. After that, the methodology determines target sentences whether have judgment concept or not, and extracts the judgment sentences. Then the methodology uses three main mapping methods to get components of free-form opinions. Moreover, the methodology gets location of components and establishes the formatted matrix of free-form opinions.

Step (A1): Target Sentences Getting

In order to get target sentences from free-form opinions, this step use sign mapping method to get target sentences. Target sentences usually use sign to present the completely concept, so this step get target sentences by mapping sign.

Step (A2): Extraction of Judgment Sentences

When the step (A1) get target sentences form free-form opinions, this step determines each target sentences whether have judgment concept. The judgment sentence usually uses status words to describe the opinion tendency of product or service, such as good, bad, or perfect. Similarly, the judgment sentence usually uses feature words to describe the user profile, such as age, gender, or occupation. Therefore, this step determines the target sentence whether have status /feature words to extract the judgment sentence.

Step (A3): Extraction of Components

In order to extract components of free- form opinions, this step use three main methods to extract components. There are specific words mapping, initiation words mapping and combination words mapping.

Step (A3-1): Extraction of Components Based on Specific Words Mapping

This step uses specific words mapping method to extract user profile item, feature item, level item, evaluated target item, status item and extent item. These items usually use specific words to describe their representative meaning, so this step maps specific words with judgment sentence to extract components.

Step (A3-2): Extraction of Components Based on Initiation Words Mapping

This step uses initiation words mapping method to extract experience item, assumption item and evidence item. These items usually use some initiation words to lead assumption, consumer experience or market research, so this step maps initiation words with judgment sentence to extract components.

Step (A3-3): Extraction of Components Based on Combination Words Mapping

This step uses combination words mapping method to extract cause item (inside) and cause item (outside). These items usually use two words to describe customer’s conclusion or consultation of

people, so this step maps combination words (including opinion provider word, reference word or cognition word) with judgment sentence to extract components.

Step (A4): Establishment of Formatted Matrix

After the step (A3) extracts components, this step integrates components to establish the formatted matrix. This step gets the location of component and establishes the formatted matrix of free-form opinions.

5. VISUALIZATION OF USER PROFILES AND OPINIONS

This research's purpose is to visualize the relationship between user profiles and opinions, and use icon's size to present the importance of opinions. Therefore, this stage uses component of opinion tendency, user profile and importance to visualize this information. Firstly, this stage determines what kind of property does status/feature item of formatted matrix have and assign status/feature value to status/feature item. Secondly, this stage uses grade item of formatted matrix to calculate the grade value. Then, this stage calculates the values corresponding to opinion tendency/user profile dimension based on the status/feature value and grade value. After that, this stage determines the values belong opinion tendency or user profile. On the other hand, this stage uses reference of formatted matrix to calculate the importance value. Finally, this stage visualizes the relationship between user profile and opinions, and visualizes the importance of opinion based on the value corresponding to three dimensions.

Step (B1): Decision of Property Corresponding to Status/feature Item

Status/feature item have tour properties, so this step mainly determines property corresponding to status/feature item. Firstly, this step gets the status/feature item $SF(CF_i,k)$ from formatted matrix CF_i and maps the status/feature item $SF(CF_i,k)$ with status/feature word DS_{1j} which is from status/feature set DS_SET(1). When the status/feature item $SF(CF_i,k)$ in accordance with status/feature word DS_{1j}, the property value $T(DS_{1j})$ of status/feature word DS_{1j} can be the judgment value $Judge(SF(CF_i,k))$ of status/feature item $SF(CF_i,k)$.

Step (B2): Decision of Value Corresponding to Status/feature Item

In order to using quantization to present the property of status/feature item, this step determines status/feature value corresponding to status/feature item based on property of status/feature item. When judgment value $Judge(SF(CF_i,k))$ of status/feature item $SF(CF_i,k)$ is 1. 2 or 3(this means that property of status/feature item is felling, level or item), the status/feature value $TQV(SF(CF_i,k))$ equals to first value $CH(SF(CF_i,k),1)$. (equation 2) On the other hand, When judgment value $Judge(SF(CF_i,k))$ of status/feature item $SF(CF_i,k)$ is 4(this means that property of status/feature item is measurement), and second value $CH(SF(CF_i,k),2)$ is accordance with unit word MEA_q, first value $CH(SF(CF_i,k),1)$ multiplied by unit ratio is status/feature value $TQV(SF(CF_i,k))$.

Step (B3): Decision of Value Corresponding to Grade Item

In order to using quantization to present the property of grade item, this step determines grade value corresponding to grade item based on property of grade. Firstly, this step calculates contribution value corresponding to grade word of grade item and uses these contribution value do multiplication to get grade value of grade item.

Consequently, when grade item $LEW(CF_i,r)$ includes grade words ,this step maps these grade word $DV(LEW(CF_i,r),w)$ with grade word DS_{2hp} which are from grade word set DS_Set(2). Then this step

can get property index CA(DV(LEW(CF_i,r),w)) and contribution index INF(DV(LEW(CF_i,r),w)) corresponding to grade word DV(LEW(CF_i,r),w).

Furthermore, when grade item LEW(CF_i,r) do not include grade words DV(LEW(CF_i,r),w) ,then the grade value W(LEW(CF_i,r)) corresponding to grade item LEW(CF_i,r) equals to "1". On the other hand, if grade item includes grade words, then execute "calculation of contribution value corresponding to grade word" and "calculation of grade value corresponding to grade item" steps.

Step (B3-1): Calculation of Contribution Value Corresponding to Grade Word

This step calculates contribution value corresponding to grade word. When property index corresponding to grade word DV(LEW(CF_i,r),w) is "1", "2" or "3" (this means that property of grade word is strengthen, opposite or attenuation), then this step can calculates influence index IL(DV(LEW(CF_i,r),w)) corresponding to grade word DV(LEW(CF_i,r),w).

Step (B3-2): Calculation of Grade Value Corresponding to Grade Item

This step gets grade value W(LEW(CF_i,r)) corresponding to grade item LEW(CF_i,r) and the product of these influence index IL(DV(LEW(CF_i,r),w)) is grade value W(LEW(CF_i,r)).

Step (B4): Calculation of User profile/Opinion tendency Value

This step calculates user profile/opinion tendency value based on status/feature value and grade value. In order to present grade item LEW(CF_i,r) modify status/feature item SF(CF_i,k), status/feature value TQV(SF(CF_i,k)) multiplied by grade value W(LEW(CF_i,r)) and this step get user profile/opinion tendency value CO(CF_i,k). (equation 9)

Step (B5): Decision of User Profile/Opinion Tendency Corresponding to User Profile/Opinion Tendency Value

This step mainly determines user profile/opinion tendency value what dimension does it belong. Moreover, this research found that user profile item and evaluated target item are usually in front of status/feature item. Besides, the status/feature item often describes the user profile/evaluated target item which is most close to status/feature item. Therefore, this step calculates the distance between status/feature item and user profile/evaluated target item, and determines user profile/opinion tendency corresponding to user profile or opinion tendency value.

Step (B6): Calculation of Importance Value Corresponding to Formatted Matrix

When people report the comments of product or service, they usually make a description of reference to support their comments. A number of reference is more, the comment is more importance. Therefore, this step calculates importance value corresponding to formatted matrix based on a number of reference.

Related to cause item, a number of cause item RF(CF_i,1) divided by the maximum value of cause item RF(CF_i,1) number. After that, the quotient multiplied by weight value is contribution importance corresponding to cause item(indise) RF(CF_i,1). In the same way, this step can get contribution importance value corresponding to cause item(outdise) RF(CF_i,2), experience item RF(CF_i,3), assumption item RF(CF_i,4) and evidence item RF(CF_i,5). Finally, importance value calculated by adding together several contribution importance values. (equation 11)

Step (B7): Visualization of Relationship between User Profiles and Opinions

This step's purpose is to visualize the relationship between user profile and opinion and visualize the opinion's importance. Therefore, this step gets a coordinates based on user profile value, opinion tendency value and importance value. Then coordinates can be signed in picture to present the relationship between user profile and opinion and the coordinates' size presents the opinion's importance (the size bigger, the opinion more important).

6. CONCLUSIONS

In this study, our research generalizes components of free-form opinions and analyzes the characteristics of these components. Then our research develops a methodology and system that automatically converts free-form opinions into structured opinion components and visualize the relationship between user profiles and opinions.

REFERENCES

Boulos, K. M. N. and Wheeler, S., 2007, "The emerging Web 2.0 social software: An enabling suite of sociable technologies in health and health care education," *Health Information & Libraries Journal*, Vol. 24, No. 1, pp. 2-23.

Chittaro, L., 2001, "Information visualization and its application to medicine," *Artificial Intelligent in Machine*, pp. 81-88.

David, A., 1999, "Guidelines for designing information visualization applications," *Proceedings of the 1999 Ericcson Conference on Usability Engineering*, pp. 1-7.

George, G. and Jock, D., 1993, "The document lens," *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, pp. 101-108.

Gruber, T., 2008, "Collective knowledge systems: Where the social web meets the semantic web," *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 6, No. 1, pp. 4-13.

Heer, J. and Boyd, D., 2005, "Vizster: Visualizing online social networks," *Proceedings of IEEE Symposium on Information Visualization*, pp. 33-40.

Kobsa, A., 2001, "An empirical comparison of three commercial information visualization systems," *Proceedings of IEEE Conference on Visualization*, pp. 1-8.

Lange, M., Hjalmarsson, J., Cooper, M. and Ynnerman, A., 2003, "3D visualization and 3D and voice interaction in air traffic management," *The Annual SIGRAD Conference*, pp. 40-45.

Masona, R. and Rennieb, F., 2007, "Using Web 2.0 for learning in the community," *The Internet and Higher Education*, Vol. 10, No. 3, pp. 196-203.

Montero, Y. and Solona, V., 2006, "Improving tag-clouds as visual information retrieval interfaces," *International Conference on Multidisciplinary Information Sciences and Technologies*, pp. 25-28.

Rasmussen, S., Mangalagiu, D., Ziock, H., Bollen, J. and Keating, G., 2007, "Collective intelligence for decision support in very large stakeholder networks: The future US energy system," *Proceedings of the 2007 IEEE Symposium on Artificial Life*, pp. 468-475.

Tegarden, P., 1999, "Business information visualization," *Communications of AIS*, pp. 1-38