

Mining and Characterizing Opinion Leaders from Threaded Online Discussions

Naohiro Matsumura^{1,2} Yukio Ohsawa^{1,3} Mitsuru Ishizuka²

¹ *PRESTO, Japan Science and Technology Corporation,
2-2-11 Tsutsujigaoka, Miyagino-ku, Sendai, Miyagi, 983-0852 Japan*

² *Graduate School of Engineering, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan*

³ *Graduate School of Business Science, University of Tsukuba,
3-29-1 Otsuka, Bunkyo-ku, Tokyo, 112-0012 Japan*

Abstract. Activating communities on the Internet is catching attentions of web site's designers because the activity affects the growth of communities. In this paper, as a trigger of activation, we aim at mining and characterizing opinion leaders from threaded online discussions on the Internet. Then we try to understand the relations between opinion leaders and their characteristics by using correspondence analysis.

1 Introduction

Communication places on the Internet, such as BBS and chat rooms, are used for the sake of gathering people into particular web sites [2]. The aim of setting such places is to make communities where people share the common context [1] by activating the interaction among people. However, it is not always easy to activate in fact because of the lack of triggers of topics. In other words, we might control the activation of a community if we could throw fascinating topics into the community. Here we focus on “opinion leader” [4] who are sensitive to the trend and having a great influence on peoples' decision making. We believe that opinion leaders can provide fascinating topics which trigger the activation of the community. In this paper, we aim at mining and characterizing opinion leaders from threaded online discussions on the Internet. Then, we try to understand the relations between opinion leaders and their characteristics by using correspondence analysis.

2 Influence Diffusion Model

Influence Diffusion Model (IDM) is a method for discovering influential comments, people, and terms from threaded online discussions, such as BBS [3]. One of the features of threaded online discussions is that communications between people are done by exchanging comments. The first assumption of IDM is that the relations of comments, called *comment-chain*, show the flow of influence. For example, if comment C_y replies to comment C_x , it is considered that C_y is affected by C_x . Similarly, if person Y replies to a comment of person X , Y is considered to be affected by X . In these cases, the influence diffuses from C_x to C_y / from X to Y . In this way, the influence diffuses throughout the comment-chains. Another feature

of threaded online discussions is that comments are written by natural language composed of terms. The second assumption of IDM is that people’s idea is expressed and propagated by the medium of terms. Based on these assumptions, the process of diffusing influence is defined as follow.

Definition 1. *In text-based communication, influence diffuses along the comment-chains by medium of terms, i.e., words or phrases.*

According to **Definition 1**, the influence is defined by the degree of terms propagating throughout the comment-chains. For example, If C_y replies to C_x , the influence of C_x onto C_y , $i_{x,y}$, is defined as

$$i_{x,y} = \frac{|w_x \cap w_y|}{|w_y|}, \quad (1)$$

where w_x and w_y are the set of terms in C_x and C_y respectively, and $|w|$ denotes the count of w . In addition, if C_z replies to C_y , the influence of C_x onto C_z through C_y , $i_{x,z}$, is defined as

$$i_{x,z} = \frac{|w_x \cap w_y \cap w_z|}{|w_z|} \cdot i_{x,y}, \quad (2)$$

where w_z are the terms in C_z .

It is considered that the more a comment affects other comments, the more the influence increases. And the same can be applied to the influence of people/concepts. The influence of a subject (including a comment, person or a term) then comes to be measurable.

Definition 2. *The influence of a subject (a comment, person or a term) to the community is measured by the sum of influence diffused from the subject to all other members of the community.*

Applying **Definition 2** to C_x , the influence (here after, let us skip “to other members of the community”) is measured by the sum of influence diffused from C_x , i.e., $i_{x,y} + i_{x,z}$ if the community has three members x , y and z .

3 Mining Opinion Leaders

In IDM, the influence of a person X is defined as the sum of influence of X ’s comments. Here we apply IDM to Yahoo!Japan’s BBS discussing about the clothing of UNIQLO.com¹. The top 5 people in the order of values of diffusing influence (D_X) are listed in Table 1.

Table 1: The top 5 people in the order of diffusing influence.

Ranking	Member ID	D_X
1	M011	36.09
2	M002	3.949
3	M004	3.340
4	M010	2.985
5	M021	2.841

¹<http://www.uniqlo.com>

Let me introduce each people in Table 1. The top-rank people, M011, was a person that posted comment #1 which is no doubt the most influential comment because it is the beginning of the comment-chains in the BBS. The second-rank people, M002, and the fourth-rank people, M010, were the staffs of UNIQLO (M002 is male and M010 is female). They frequently offered the hot information about new clothing, hot-selling clothing or advices on dressing. The third-rank people, M004, posed comments which often raised argumentive topics. The fifth-rank people, M021 who had been an UNIQLO enthusiast for four years, had been also offering much information from customer's point of view. All of them actively offered influential comments that catch participants' interest and caused constructive discussions. Therefore, we believe that they were suitable for opinion leaders.

4 Characterizing Opinion Leaders

The characteristics of opinion leaders are different from each other. For example, one might have a great influence on the color of fleece, and another might have on the style of fleece. Here we regard one's influential terms as his/her characteristics. The influential terms for each opinion leaders in Table 1 are listed in Table 2.

Table 2: Opinion leaders and their characteristics (influential terms).

<i>Member ID</i>	<i>Characteristics (Influential Terms)</i>
M011	UNIQLO, cardigan, jam, advice, sneaker, scarf, return, wear, fleece, fashionable, jacket ...
M002	UNIQLO, wear, fitting, fleece, advice, shape, sleeve, sweater, design, color, catalogue ...
M004	touch, jeans, pants, color, UNIQLO, return, clerk, fitting, cheap, size, shape, fleece, skirt ...
M010	cute, fleece, coverall, slipper, UNIQLO, beige, T-shirt, beige, demin, size, pink, color ...
M021	fleece, spring, slipper, color, cheap, full zipper, winter, walk, the Internet, UNIQLO, bag ...

5 Correspondence Analysis

For understanding the relations between opinion leaders and their characteristics, we employ correspondence analysis [6] to visualize the relations as a two-dimensional positioning map. We skip the details of correspondence analysis because it is beyond the scope of the paper. Figure 1 shows the positioning map for the data in Table 2. By seeing Figure 1, we can clearly understand the characteristics of opinion leaders and their terms as follows:

- M002 and M004 have similar characteristics, which are price, jeans, catalogue, T-shirt, basic, etc.
- M021 has the characteristics, which are spring, winter, the Internet, weekend, etc.
- M010 has the characteristics, which are beige, color, pink, cheap, etc. Note that M010 is in the position among M002, M004, and M0021.
- M011 has the characteristics, which are jacket, scarf, cardigan, etc.
- M002, M004, M0010 and M0011 have common characteristics, which are return, advice, favorite, etc.
- All the opinion leaders have common characteristics, which are UNIQLO, fleece, etc.

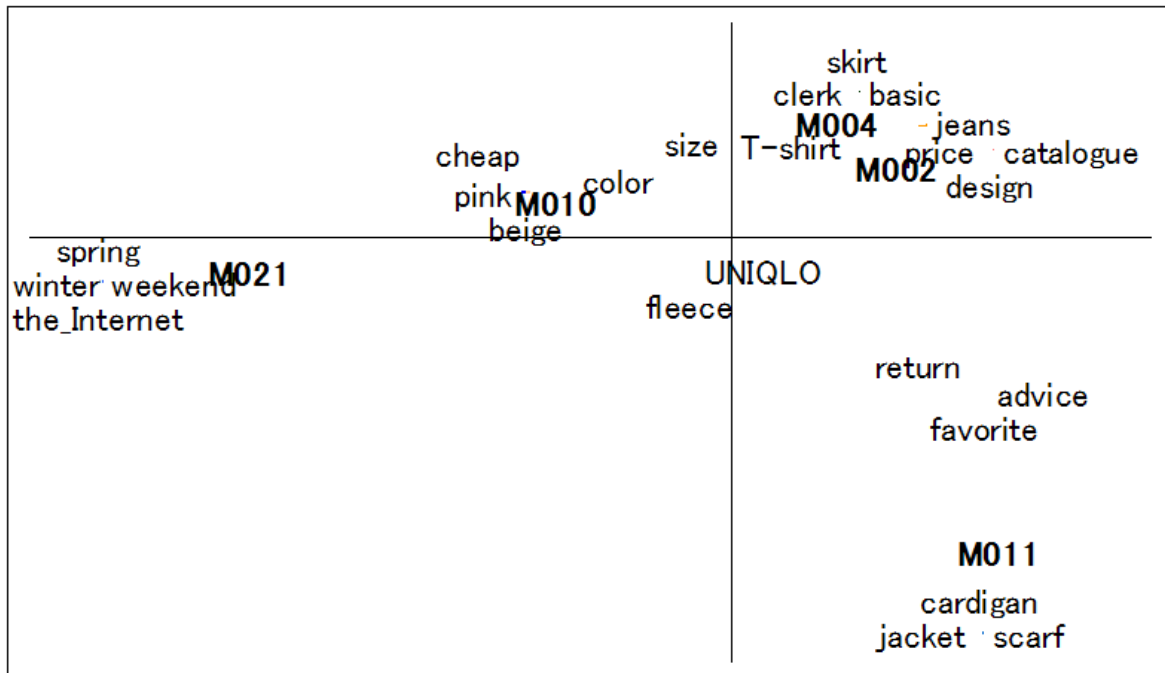


Figure 1: Positioning map of opinion leaders and their characteristic terms. For ease of understanding, moderate number of terms are shown.

6 Conclusions

The results mentioned in this paper help us understand the opinion leaders and their characteristics. As the future strategy for activating a community, for example, we plan to let a opinion leader whose characteristics are related to the community's current topic provide some topics for activating the community if the community becomes stagnant.

References

- [1] Yukio Ohsawa: Chance Discovery for Making Decision in Complex Real World, *New Generation Computing*, Vol. 20, No. 2, 2002.
- [2] Naoto Ishikawa: Internet Community Strategy, SoftBank Publishing, 2001. (in Japanese)
- [3] Naohiro Matsumura, Yukio Ohsawa, and Mitsuru Ishizuka: Influence Diffusion Model in Text-based Communication, *WWW02*, 2002. to appear
- [4] E.M. Rogers: Diffusion of Innovations, The Free Press, 1962.
- [5] S. Kiesler, J. Siegel, and T.W. McGuire: Social Psychoogical Aspects of Computer-Mediated Communication, *American Psychologist*, Vol. 39, pp. 1123–1134, 1984.
- [6] Miyagawa Masami: Graphical Modeling, Asakura Publisher, 1997 (in Japanese)