Information Extraction of Review Using LIWC

Ji Yeon Lim, Jae Yoel Yoon, Lee Joon Kim, and Ung Mo Kim

Abstract—**Social commerce market, which is promoted using a social network service, is getting bigger with the activation of a social network service. If a number of consumers are more than a number which the sellers decide before selling, the deal in social commerce market is completed. It's actually hard to guarantee money back regarding the reason, so reviews are very important to effect for purchasing the products in social commerce market. It may have different direction of information through background, taste, emotion of the reviewers. At the same time, if the meaning of text information can be extracted, we can have the correct information that we need. This paper shows that we extract the meaning by opinion mining and identifying the user's basic information and psychological state by LIWC. It gives the correct ratings to consumers who want to purchase products through social commerce market. Also it can be helpful to increase reliability of opinion mining on the other fields.**

Index Terms—**Social commerce, opinion mining, LIWC, ratings.**

I INTRODUCTION

In social commerce market, reviews are very important to effect for purchasing the products since it's hard to guarantee money back after completing payment. This paper describes an effective opinion mining method on reviews of social commerce with user's information extraction method which gives reliability of its result.
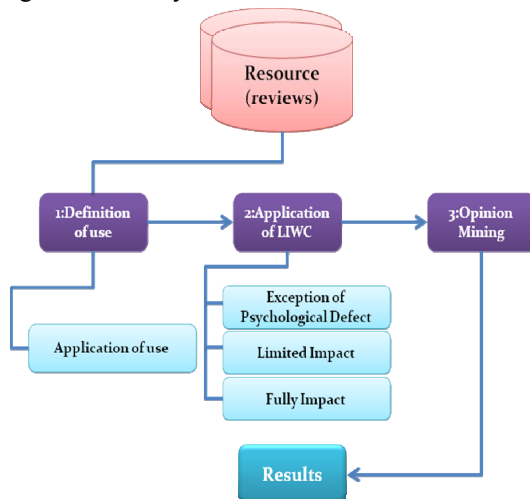

Fig. 1. Process of the method

Other users refer to the rating for choosing products which they prefer. Most people who already bought don't want to reveal their specific information, and consumers can't recognize whether it fits them or not. For this, the

psychological state or personal background of users who wrote reviews needs to be analyzed. For example, if the target of a product is 20s and the consumer who buys and writes is 50s, the review would be naturally bad. Also if a product for men, the review which women write isn't effective. In that case, It's helpful to give ratings considering the generation and gender of people who write the reviews. So we focused on opinion mining on social commerce, concerned about extraction of user's information for credibility of the extracted information.

The paper organized as follows. The following section reviews the related work. We describe the order to extract user's information in Section 3. We verify the method in the next section. Finally, we discuss our conclusions and future work.

II RELATED WORK

A. Social Commerce

By moving to the online network from realities, social network is developed and social commerce via word of mouth in social network is also booming. However, the social commerce doesn't work well because the marketing plan is so similar to joint purchase. We suggest that reviews which are written by a stranger help to ensure the reliability. By considering the numerical data extracted by contextual background of the individual, the user believes he can purchase in social commerce, and the seller also knows people who know a little more away from the joint purchase format at an affordable price so they can benefit from publicity.

Fig. 2. Resource for the case study

B. Opinion Mining

Even if a user doesn't directly write a rating, opinion mining is a numerically expressed technique which provides lots of information to customers who want to buy products or avail services. There are many kinds of method that the word of product can be extracted. We use a sentiment classifier technique [11]. By using an existing opinion mining, we can extract the user's opinion not indirectly connected area with the product such as blogs, social network. So the information, which we don't need, should be removed. There exist the technique for this [12], but the way to use statistics and trends cannot guarantee its effectiveness. The method that we propose can extract the opinion that fits with users through user's background and psychological information.

C. LIWC

LIWC (Linguistic Inquiry and word count) [1], [2] is, using "counting the number of words" used to analyze the text, an effective method to discover user's gender, age, other basic information and psychological state. The word count method is good to mash up with the opinion mining. If someone writes text with less long word which is composed of more than 6 character and articles who is probably a female [7], [8]. Besides, it would be a man who uses negative words and swear words. In case of age, most people who use a lot of positive, casual relationship, insight, present and future verb, and long words are seniors [9][15]. Causes words and insight are used by a man who is healthy. However, when lie, the first-person singular pronoun and the third-person pronoun are less used and lots of negative words and action verbs are used [6]. In this paper, using these characteristics, extracts the background to help determine the meaning of reviews on social commerce.

III INFORMATION EXTRACTION OF USERS

A. Definition of social commerce use

For the mining of reviews on social commerce, we need to define number of use at social commerce. Users who have which never used social commerce market could mention about bad aspect that is inevitable on social commerce, if mine including this aspect, the rating will be low. On the other hand, if too many times the reviews are written with lack of sincerity. So we divide the users into 2 groups by the use of social commerce, limiting the scope of actual users. On average, users who use it more than three times and less than 10 times write good quality of reviews compared to the existing services or product's quality. We apply users who have the 3 to 10 times of use.

B. Application of LIWC

Psychology can be revealed by LIWC which is made by analyzing between writing texts and their information. It counts a number of words and classifies into several categories. We focus on reliability of data which comes out from a writing habit by detecting a lie. Also, text which is inferred having all sorts of symptoms such as hypochondria is regarded as unreliable resource, so it will reflect its influence reduced to process. In addition, other properties like age and gender are extracted through the LIWC. It is for increasing reliability and extracting background information psychologically.

C. Extraction of Contents Meaning

By applying opinion mining, we will extract taste and thought of users through analyzing positive and negative emotions. The data is divided into data which is strongly used and not strongly used by the number of use which is defined at first. Than the extracted taste and through of users are applied with the psychological data which is extracted from the step of analyzing by the LIWC. If a healthy state can be guessed, the impact will be used fully. Less reliability of the data won't reflect at all or will have limited impact for service of rating on social commerce.

IV CASE STUDY

We found reviews in the Face book Payment as a social commerce and then applied definition of use. After applying the definition of use, we got a resource from users who use social commerce 3 times. Fig. 2 is the resource for the case study.

TABLE I: ANALYSIS OF REVIEWS OF A USER BY LIWC

LIWC dimension	Your data	Personal texts	Formal texts
Self-references (I, me, my)	8.47	11.4	4.2
Social words	3.39	9.5	8.0
Positive emotions	5.08	2.7	2.6
Negative emotions	0.00	2.6	1.6
Overall cognitive words	8.47	7.8	5.4
Articles (a, an, the)	6.78	5.0	7.2
Big words (> 6 letters)	20.34	13.1	19.6

Through the results in Table I, we can get the information of the user. The user has a positive emotion towards the product. It could be applied to the result by weight but we didn't consider it on the paper. The user has written many of self-references (I, me, my), cognitive words and less articles, big words. These aspects show that the user seems to be a female, so it can be presented to female. In this stage, if depression or defect on review has appeared, it is ignored on the rating.

TABLE II ANALYSIS OF A REVIEW SENTIMENT

No	Positive words	Negative words	Sentiment
1		not	-1
2	quickly		+1
3	excellent		+1
total			+1

```
integer i

if i < 3 or i > 10
    remove on list
else
    apply

j = examine i on LIWC

if j exists psychological defect
    k = 0
else if j = not sure
    k = 0.5
else
    k = 1

apply results of opinion mining * k
```

Fig. 3. Pseudo code of the method.

By using sentiment classifier, we divide the words into

positive and negative and then count those for total sentiments. According to Table 2, the rating is got one point on the product. It seems simple method but we think that it's very helpful for accuracy of opinion mining results. Fig. 3 shows the procedure by pseudo code of the method that we studied. Speaking of the procedure, we remove the review of reviewers who use less than 3 times and more than 10 times at first. We regard that the reviews which we removed have less words for this experiment. After removing it, if psychological defect exists, we applied zero point or if we are not sure about it, we applied 0.5 points or one point. Finally, we multiplied the number which we got in the last phase by a rating of opinion mining.

V Conclusion and Future Work

We have applied LIWC for increasing reliability of opinion mining of reviews in social commerce. In the experiment through the analysis of LIWC, reviews were enough to give credibility. On the other hand, there are not so many reviews in social commerce therefore we had some trouble analyzing it. We expect that the rating applied to our method and more analysis is shown for customers in social commerce market

However, it would be needed to experiment that the reliability is appropriate and other type of calculating the credibility should be studied for the development of social commerce and the rights of consumers to purchase in social commerce market.

References

[1] K. S. Cho, J. Y. Yoon, I. J. Kim, J. Y. Lim, and U. M. Kim, *Mining information of anonymous user on a social network service*, 2011.

[2] J. W. Pennebaker and R. J. Booth, and M. E. Francis, *Linguistic inquiry and word count (LIWC2007) Austin*, TX: LIWC (www. liwc. net), 2007.

[3] S. Rude, and E. M. Gortner, and J. Pennebaker, *Language use of depressed and depression-vulnerable college students, Cognition and Emotion*, vol. 18, no. 8, pp. 1121-1133, issn 0269-9931, Psychology Press, part of the Taylor and Francis Group, 2004

[4] S. W. Stirman, and J. W. Pennebaker, *Word use in the poetry of suicidal and nonsuicidal poets*, Psychosomatic Medicine, vol. 63, no. 8, pp. 517, 2001.

[5] J. W. Pennebaker, and A. Graybeal, *Patterns of natural language use:* Disclosure, personality, and social integration, Current Directions in Psychological Science, vol. 10, no. 3, pp. 90, 2001

[6] M. L. Newman, J. W. Pennebaker, and D. S. Berry, and J. M. Richards, Lying *words: Predicting deception from linguistic styles, Personality and Social Psychology Bulletin*, vol. 29, no. 5, pp. 665, 2003.

[7] M. R. Mehl and J. W. Pennebaker, *The sounds of social life: A* psychometric *analysis of students' daily social environments and natural conversations*, Journal of Personality and Social Psychology, vol. 84, no. 4, pp. 857–870, 2003.

[8] A. Mulac, J. J. Bradac, and P. Gibbons, *Empirical support for the gender-as-culture hypothesis*, Human Communication Research, vol. 27, no. 1, pp. 121-152, 2001.

[9] L. D. Stone, and J. W. Pennebaker, *Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana*, Basic and Applied Social Psychology, vol. 24, no. 3, pp. 173-183, 2002.

[10] S. M. Kim and E. Hovy, *Determining the sentiment of opinions*, Proceedings of the 20th international conference on Computational Linguistics, pp. 1367, 2004.

[11] B. Pang, and L. Lee, *Opinion mining and sentiment analysis*, Foundations and Trends in Information Retrieval, vol. 2, no 1-2, pp. 1-135, 2008.

[12] N. Jindal, and B. Liu, *Opinion spam and analysis, Proceedings of the international conference on Web search and web data mining*, pp. 219-230, 2008.

[13] S. Milgram, *The small world problem, Psychology today*, vol. 2, no. 1, pp. 60-67, 1967.

[14] M. S. Granovetter, *The strength of weak ties, The American journal of* sociology, vol. 78, no. 6, pp. 1360-1380, 1973.

[15] J. Golbeck and M. Rothstein, *Linking social networks on the web with FOAF: a semantic web case study*, Proceedings of the 23rd national conference on Artificial intelligence, pp. 1138-1143, 2008.