SENTIMENT ANALYSIS IN SOCIAL WEB ENVIRONMENTS ORIENTED TO E-COMMERCE

Luigi Lancieri
Lille 1 University
Cité Scientifique, Villeneuve d'Ascq, France
Luigi.lancieri@univ-lille1.fr

Eric Lepretre
Lille 1 University
Eric.lepretre@univ-lille1.fr

ABSTRACT

The purpose of this paper is to have a better understanding of users' behavior in a social web environment. The data used are collected from web sites containing reviews about web merchants. In such sites, users provide a short description of the quality of the merchant as well as a degree of satisfaction. The idea underlying this study is to explore the textual structure (e.g the reviews length) and the perceived ambiguity of the opinion expressed in the reviews. One of the results of this study is that there is a correlation between the degree of satisfaction and the length of the user contribution. The other result is that the ambiguity of the opinion appears to be different when the opinion is positive compared to a negative one. We analyze the meaning of these finding in the growing context of social web interactions.

KEYWORDS

Measure, ambiguity, opinion, sentiment.

1. INTRODUCTION

Social web environments are nowadays very common but are not always easy to understand. Platforms as Facebook, Twitter or sites dedicated to the reviews of commercial products have increasing influence in "real life" society. A study of Doubleclick shows that more than half (up to 75% for travel purchases) of the consumers look for information (reviews, ..) before making their online purchase (Qiang *et al*, 2009).

From a general point of view we may wonder how the content of such reviews can inform us about the way people expresses their feelings within social web environments. The content of reviews refers here to other thing than the strict meaning of words and refers more to the text structure or the indirect expression conveyed by the text. For example, in face to face interactions, the voice tone or the speech rhythm can often give us more information than the words. In the same way, it is well known that physical posture or facial expression (smiling, etc) or in one word the indirect forms of communication are also extremely expressive.

Even if social web does not transpose all aspects of real interactions, it's an interesting context in order to analyze indirect communications. This can be done thanks to the capacity of these environments to memorize traces of interactions. Computer scientists (among others) have studied the huge quantity of data accessible in such a milieu in order to compute behavioral indicators from raw data.

In this paper, we first discuss of the meaning of the length of users input when such a contribution convey sentiments, opinions or reputations. This part of the work is achieved through the analysis of reviews content descriptive statistics compared to the degree of satisfaction provided by the author of the review. Then we investigate the relation between the ambiguity of the review and the polarity (negative or positive) of the opinion. This part is achieved by comparing the degree of satisfaction expressed by the user with those obtained through a computer based sentiments classifier. We assume that greater the difference, the higher the ambiguity of the review is high. We analyze the result of both parts of this work as well as the link between them. First of all, it is important to note that some web social environments have high constraints regarding the size of the users' contribution. The best example is that of Twitter that is dedicated to very short texts. Others platforms (forums, blogs, ..), on which we focus, are more open.

This paper is organized as follow. First we describe our methodology and the results we obtain. Then we make a review of similar works and finally we discuss the outcome.

2. METHODOLOGY AND RESULTS

We begin by analyzing the statistics properties of the reviews including the relation between the lengths of the reviews according to the polarity (negative or positive) of the opinion expressed. Then we investigate the relation between the ambiguity of the reviews and their polarity.

2.1 Statistics of the reviews

First of all, we got a set of 5014 reviews including the degree of satisfaction (from 0 to 10) expressed by the reviewer. The reviews were collected from online web sites dedicated to consumer expression (e.g. rateitall.com). The average length of a review is of 66.7 words (std=66) and we can observe that 90 % of them have a length lower than the average plus one std. The total length of all reviews is of 335 246 words.

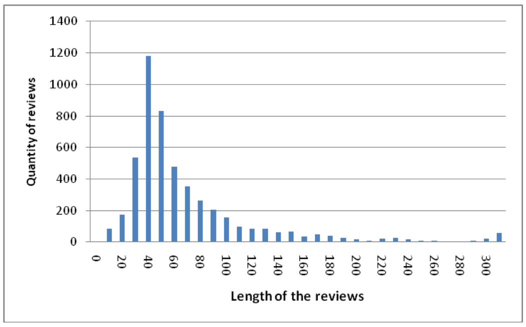

Figure 1. Quantity of reviews per class of length

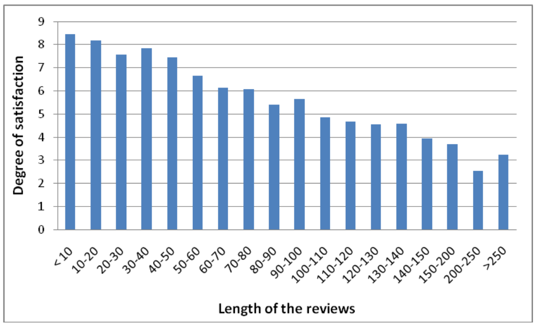

Figure 2. Degree of satisfaction versus length of the review

In order to show the relation between the length of the review and the degree of satisfaction of the user, we first rank all the reviews by class of length (see abscissa of figure 2). Then, we compute the average degree of all the reviews by class of length. The figure 2 shows this relation. For example, the first bar of the plot tells us that all the reviews that have under 10 words have an average notation of 8.5. This figure suggests that the less the reviewer is satisfied, the more he expresses himself. Indeed, satisfied users (notation near 10) have written shorter reviews (left of the plot) whereas unsatisfied users (notation near zero) have written longer reviews (right of the figure).

Such a remark may lead us to the conclusion that people express themselves mainly when they are unsatisfied. Actually this is not the case. It should be done a difference between the fact that someone expresses himself or not, and the length of his expression. This is confirmed by another view of our dataset. The figure 3 shows the distribution of the quantity of reviews per degree of satisfaction (from 1 to 10). This figure corroborate that most of the reviews are positive (1425 less than score 5, 3410 more than score 5).

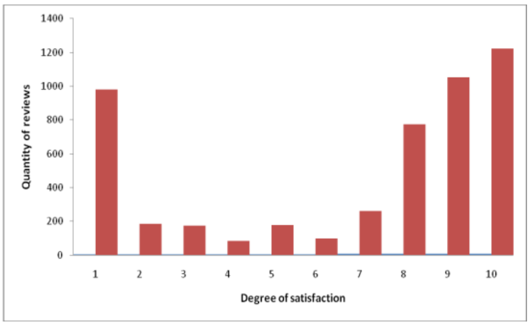

Figure 3. Distribution of reviews depending on the degree of satisfaction

We see that most of users (around 2/3) are satisfied. Furthermore, most of the unsatisfied users (around 2/3) are strongly unsatisfied (peak on notation 1). This is logical and consistent with the fact that those users continue to trust and use online commerce. On the other hand, the situations that make users unsatisfied are most of the time complexes and need long explanations contrary to situations where everything is fine and from which there is nothing special to say.

2.2 Measuring reviews ambiguity

Another question that we may ask highlights the relation between the review length or its polarity with the ambiguity of the content. The ambiguity can be defined as the fact that a text expresses clearly or not an opinion (negative or positive). In order to evaluate the ambiguity, we compare the degree of satisfaction given by the user with that found by an artificial sentiment classifier (that is supposed to be objective). The idea is that ambiguity is proportional to the level of disagreement between the classifier and the human. Both have rated each review between 1 (disappointment) and 10 (satisfaction).

For this study we used a sentiment classifier developed within our lab. It is out of this study to describe this classifier but we provide a set of references with detailed information about underlying techniques (Cui *et al*, 2006), (Gindl *et al*, 2008), (Qiang *et al*, 2009), (Xia *et al*, 2010). We may also say that our sentiment classifier has 89 % precision and 85 % of recall. These performances are consistent with this category of tools. Precision can be seen as a measure of exactness or of fidelity, whereas recall is a measure of completeness. First, we focus on the amplitude distribution of the disagreement which is simply the absolute value of the difference between both notations. In the figure 4 the abscissa represents this difference. For example the fourth bar means that classifier gives a notation of 5 and that the human gives 9 (or the opposite) in 9 % of the cases.

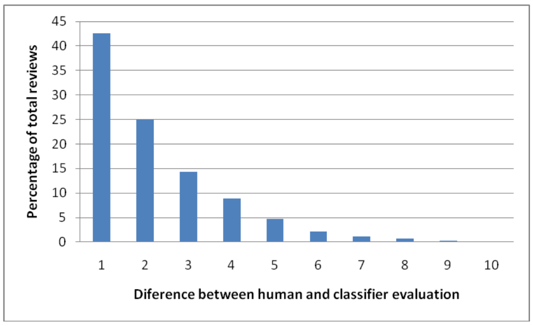

Figure 4. Percentage of reviews according to the difference between the classifier and the user notations.

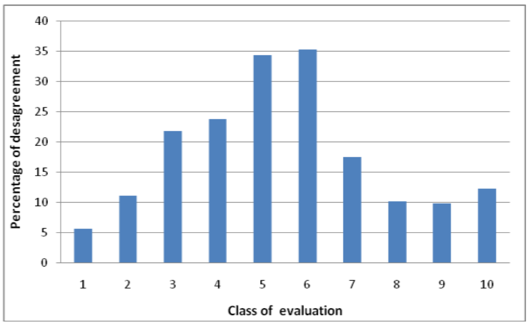

Figure 5. Percentage of disagreement within each class of opinion notation

First of all, we can see that most of the time the classifier agrees with users. Indeed more than 66 % of the reviews have only one or two points of difference between both notations. Let's see now how this difference is distributed over the classes of sentiments. The purpose is to see where the disagreement is the most visible. We consider here that for both the classifier and the human: if the notation is between 1 and 5, it expresses a sentiment of disappointment, otherwise a satisfaction. Then, for each class of sentiment level (1 to 10), we compute the ratio of reviews where there is a disagreement. For example in the figure 5, the bar 5 that has the value of 34 % means that the classifier disagrees with 34 % of the messages that were noted 5 by the user.

As we mentioned before, this level of disagreement is considered as a level of ambiguity. As expected, the bars 5 and 6 have the highest level of ambiguity that corresponds to the middle range position of these notations. On the other hand, as shows the table 1, we can notice that whereas the level of disagreement is globally equal (88%) for each polarity (negative, positive), the distribution is very different inside of each polarity class (figure 5). We observe that the level grows regularly from the notation 1 to 6 and then goes down quickly. This suggests that the positive opinion appears rapidly without ambiguity contrarily to the negative opinions. We may consider these relations as a dynamic model of ambiguity progression (ie from one class 1, 2… to another).

Table 1. Level of disagreement depending on the polarity of reviews

	Negative(1-5)	Positive (6-10)
Number of message	1604	3410
Number of disagreement	1412	3001
Percentage of disagreement	88%	88%

Let us remember our previous suggestion, that negative messages are the larger one due to of the complexity of the discourse necessary for expressing problems or reasons of unsatisfactions. In such case, it is not surprising to see that ambiguity disappears less rapidly in negative messages.

3. RELATED WORKS

The relation between the length of the expression and the satisfaction of customers has already been studied in relation to mouth-of-mouth. Anderson, for example, has stated that unsatisfied customers engage themselves in greater word-of-mouth than satisfied ones (Anderson, 1998).

In the context of collaborative work several studies observe a relation between the amounts of contributors in a group and the length of individual contribution. Probably due to different methodologies there is no unique conclusion regarding the orientation of the relation. Some works report that an increase of a group size tends to reduce individual participation. Other works find the opposite but it appears that the structure of the group has a crucial influence on the level of individual contribution (Lancieri, 2008). Avouris *et al* remark for example that larger groups produce better results and generate greater activity, but this activity is less homogeneously distributed between different members of the group (Avouris *et al*, 2004). Valacich *et al* reported an even more precise finding, observing that the activity rate per member increases with the size of the group when it is composed of members with diverse skills, whereas it decreases when the group is homogeneous (Valacich *et al*, 1993).

Apart the structure of the group, other researchers notice that the modality of the communication and especially the user interface also have an influence on the user contribution. C. Jensen, for example showed that the contribution of users to a group activity, such as an online game, was higher when the mode of communication was more evolved (voice, speech synthesis, text chat, no communications) (Jensen *et al*, 2000). The author points out that artificial speech generates more contributions than the equivalent text (chat), but still less than human conversation. It is also often observed that the length of the contribution depends on the modality of the interaction. Questions or answer, for example, does not have the same length. Although Twitter allows status updates to be up to 140 characters long, and Facebook up to 423, the collected questions had a mean length of only 75.1 characters (13.8 words). MR. Moris *et al* founds that question length influence response relevance. Indeed, questions that have fewer sentences receive more helpful responses than those with many sentences (r = -0.13)(Morris *et al*, 2010). A lot of work, especially in the language theory field, has for a long time analyzed the influences of text complexity and structure on its understanding. These works show that text length, linked with the lexical redundancy level, is directly related to the text comprehensibility. Details, discussion and references can be found in (Lancieri, 2008).

4. DISCUSSION

As we have seen, the reason explaining the length of a textual contribution may depend on several factors. First of all, there are personal reasons. The education or the psychological profile of an individual may lead him to express the less possible or at contrary to give his opinion spontaneously. Social factors are also influential as observed in face-to-face interactions. The opinion of others contributors or of the group often motivate personal expression.

Dislike to some web environments that exhibit communities and long-term interactions (hobbies, open source software, ..), e-commerce and reviews dedicated sites are more focussed on the expression of final opinion. In such environments, users rarely expect responses or long-term interactions. The specificity of this kind of context is important to have in mind while discussing the reasons or the implications of text length. If we exclude or limit the social influence, the direct relation between the length of the review and the disappointment degree of the users may come from two main raisons. The first one, as we said previously is linked with the fact that a problem is more complex to explain than a situation occurring as expected. The second reason may have a relation with the anger and the frustration generated by the feeling of disappointment leading to the need of expression.

Regarding the ambiguity of the reviews, it is clear and quite obvious for middle range opinions (degree 5 and 6). What is interesting is that in the positive opinions, the ambiguity disappears more rapidly than in negative opinion. We assumed that this was linked with the difficulty to explain situations leading to disappointment. In one word this would means that negative opinions are not only difficult to explain but also difficult to understand. This may explain also why conflicts linked to these situations are not easy to solve.

REFERENCES

Anderson EW,1998, Customer satisfaction and word of mouth - Journal of Service Research, 1998

Avouris N *et al*, 2004, The effect of group size in synchronous collaborative problem solving activities, *Proceedings AACE Conference ED-MEDIA* 2004, Lugano, June 2004.

Cui *et al*,2006, Comparative Experiments on Sentiment Classification for Online Product Reviews, *proceedings of the 21st national conference on Artificial intelligence* (2006)

Gindl *et al*, 2008, Evaluation of different sentiment detection methods for polarity classification on web-based reviews. *18th European Conference on Artificial Intelligence,* ECAI2008 Patras Greece ECAI Workshop on Computational Aspects of Affectual and Emotional Interaction p. 35–43

Jensen C *et al*,2000, The Effect of Communication Modality on Cooperation in Online Environments, *In Proceedings of CHI 2000*, The Hague, Netherlands March 2000. http://research.microsoft.com/scg/papers/dilemmachi2000.pdf,

Lancieri L, 2008, Relation between the Complexity of Individuals' Expression and Groups Dynamic in Online Discussion Forums, *The Open Cybernetics and Systemics Journal (TOCSJ)*, Volume 2. 2008;

Morris MR *et al*, 2010, What Do People Ask Their Social Networks, and Why? , Proceedings of the 28th ACM *International conference on Human factors in computing systems* (CHI10)

Qiang Y *et al*, 2009, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, *Elsiever Journal Expert Systems with Applications,* Volume 36, Issue 3, Part 2, April 2009, pp. 6527-6535. http://hci.ece.upatras.gr/pubs_files/c80_Avouris_Margaritis_Komis_2004_ED_MEDIA.pdf,

Valacich JS *et al*, 1993, Computer-mediated idea generation: the effects of group size and group heterogeneity, *Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences*, Volume iv, 5-8 Jan. 1993 Page(s):152 - 160 vol.4

Xia H *et al*, 2010, Sentiment Text Classification of Customers Reviews on the Web Based on SVM, 2010, *Sixth International Conference on Natural Computation* (ICNC 2010)