Are Influential Writers More Objective?
An Analysis of Emotionality in Review Comments

Lionel Martin   Valentina Sintsova   Pearl Pu

Human Computer Interaction Group
School of Computer and Communication Sciences
Swiss Federal Institute of Technology (EPFL)
CH–1015, Lausanne, Switzerland
{lionel.martin, valentina.sintsova, pearl.pu}@epfl.ch

ABSTRACT

People increasingly rely on other consumers' opinion to make online purchase decisions. Amazon alone provides access to millions of reviews, risking to cause information overload to an average user. Recent research has thus aimed at understanding and identifying reviews that are considered helpful. Most of such works analyzed the structure and connectivity of social networks to identify influential users. We believe that insight about influence can be gained from analyzing the *affective content* of the text as well as *affect intensity*. We employ text mining to extract the *emotionality* of 68,049 hotel reviews in order to investigate how those influencers behave, especially their choice of words. We analyze whether texts with words and phrases indicative of a writer's emotions, moods, and attitudes are more likely to trigger a genuine interest compared to more neutral texts. Our initial hypothesis was that influential writers are more likely to refrain themselves from expressing their sentiments in order to achieve a more perceived objectivity. But contrary to this initial assumption, our study shows that they use more affective words, both in terms of emotion variety and intensity. This work describes the first step towards building a helpfulness prediction algorithm using emotion lexicons.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models*

Keywords

Emotion analysis, review helpfulness, social influence, online product reviews

1. INTRODUCTION

Social media continue to grow and to play increasingly important roles both on the Internet and in our daily lives. They give users ever more possibilities to express their opinions and impressions about products, activities, businesses, and others. However, these platforms now display so many reviews, comments, and messages that an average user is unlikely to be able to process all of them. This information overload, unfortunately, hampers a user's ability to make sound judgments and take firm decisions. Meanwhile, only a few reviews or comments have a genuine impact on whether the user buys a certain product or stays in a particular hotel.

Understanding what the influential texts are in social media and who wrote them is of particular interest for Social Recommender Systems. An automatic tool to detect helpful reviews will allow busy users to get the most reliable information from amongst the numerous sources as quickly as possible (e.g. displaying influential reviews at the top of the product page to help users choose faster). At the same time, characterizing the content of such influence could shed light on how users may improve their writings of helpful comments.

In the last few years, researchers have become increasingly keen on finding means to identify these influential reviewers. Typical efforts involve analyzing the structure and connectivity of social networks [2, 4, 9, 10]. However, little attention has so far been paid to opinion leaders' actual textual content. Emotion analysis is one manner to study the content of reviews. Indeed, emotions are powerful tools for communication as they are most likely to evoke the feelings of others and engage their responses. They also drive people's actions and regulate their decision processes. On the other hand, emotional comments may be also related to reviewers' idiosyncratic experiences, thus risk being boring to other users. To the best of our knowledge, this work is the first in its kind to **examine** and **compare** how influential users and their counterparts (non-influential users) express their emotions in online review texts.

To be able to recognize emotions in text and analyze them is not as easy as it may seem. It requires an emotion model that is more elaborated than the commonly used polarity model (positive vs. negative emotions). This model needs to be supported by an emotion lexicon, i.e., a dictionary of words indicative of the emotions. Fortunately the sources of available information continue to multiply. Thus we can analyze the writing behavior of the users through their opinions and the content they exchange on review websites. We chose

thus to represent influencers as the users who have written comments that were later voted helpful by their community and conducted a content analysis of the review comments in order to identify the influencers with their use of emotional expressions. We attempted to retrieve and determine the specific characteristics of emotion reference that made opinions, comments, and reviews more helpful. We concentrate our retrieval efforts around 20 emotion categories, spanning a multitude of both positive and negative feelings for an optimal coverage. We hypothesized that the influential reviews would be the most descriptive ones providing information that can benefit to others and are not too much linked with the writer's personal thus those with fewer emotion signals. Our study shows the contrary, at least as far as hotel and restaurant reviews are concerned. We summarize the challenges faced by the study and the improvements that can be made in order to answer the following questions:

- How influential users behave in social media? What kinds of comments they are likely to construct? What is the emotionality of these comments?

- Beyond positive and negative emotionality, is there a more elaborate emotion model to further identify emotionality? What is the method to detect these emotions?

- What will be the major differences of emotionality between influential and non-influential users? Is our finding restricted to one type of dataset or can it be generalized to several others?

While the above research questions are concerned with constructing an emotion recognition model and the analysis of emotion content of reviews, the answers will help us reach our next-stage goal: the design of a helpful review prediction model using emotion lexicons. This would ultimately help consumers reach the best and most meaningful comments with little effort.

2. RELATED WORK

The study of reviewers' helpfulness scores has been addressed in several papers aimed at improving the visualization of comments [5, 8, 17]. These works mainly characterize helpful reviews and predict the future score of newly created comments to allow a classification that is not dependent of the time elapsed since the review was written using statistics of the text.

Kim et al. [11] identified the helpfulness to be affected mainly by three factors: the length of review, the product rating and TF-IDF scores of words used. This shows the impact of the content in the characterization of helpfulness even though the authors used only the statistical description of the content and not the actual meaning of the words.

A few works combine the study of structure and content to understand the influence of the users. Liu et al. [13] described and predicted the helpfulness of the reviews based on three classes of attributes. The first one regards content and computes statistics on the review text to determine the writing style of the reviewer. The second class is structural and considers the time elapsed since the review was written. Finally, the third one is a combination of the two first kinds: it represents the expertise of a user for a given class of items and links the structure and content.
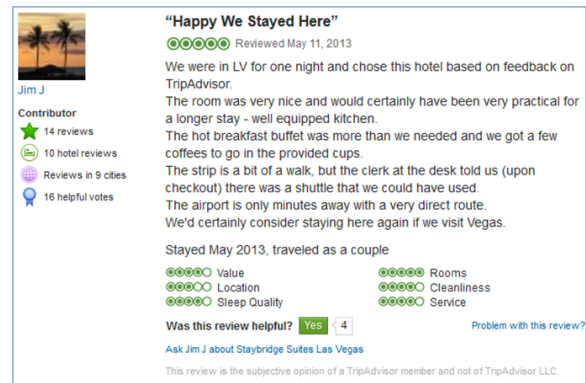

Figure 1: Example of a review on the Trip Advisor website indicating the rating and the comment of the reviewer as well as information about him and the helpfulness assessed by the community.

Unfortunately, none of the presented works formally studied the impact of the content of the reviews at its full potential. For this reason, we believe that text mining is an essential step towards the understanding of helpfulness given that it has been shown that content-based statistics are of importance.

Using sentiment analysis, Ahn et al. [1] selected the most important snippets of information to display for recommendation. They showed that those expressing a strong opinion were making people read more about a product compared to neutral texts. However, we are interested in our current study in a fine analysis of sentiments where we will not only observe the positivity but also to which emotion each word refers once its valence (i.e. positivity) is known. Such work is closely related to Mohammad's [14] which retrieved emotions in emails and books, using Plutchik's model with 8 different families (details are presented in section 3.2). He analyzed in his work the emotions present in different media (love letters, suicide notes, fairy tales, working emails, etc.) and presented a few visualization tools to track them. The author also introduced the concept of emotion word density with which he is able to cluster items with similar emotional features but also to sort these items with respect to this density. However, he did not investigate the impact that these emotions had on readers, e.g. by comparing the emotionality of best seller with average books.

3. OUR APPROACH

3.1 Dataset

We based our research on a dataset of 68,049 reviews from the Trip Advisor website,[1] which is one of the largest review websites for hotel reviews. Those reviews are anonymized versions of comments posted on the website between 2008 and 2011 about 216 different hotels in Las Vegas, and used in [16]. Each of them contains (a) an overall rating for the establishment giving information about the positivity of the review; (b) a review text and a review title from which we extracted the emotionality of the review; (c) different category ratings such as one for cleanliness, one for location, etc. providing information on the interests of the user when trav-

---

[1]http://www.tripadvisor.com

eling; and (d) the helpfulness score providing an estimate of its influence (see Figure 1).

3.2 Emotion Models

The classification of emotions has been an active area of research for more than 30 years. Even though most of the works agree on the fact that the most efficient way to classify emotions is to use families of feelings around one or several terms, the model of classification (i.e., the emotion model) by itself differs from one to the other.

Ekman [6] defined 6 and then 7 emotions (i.e., Happiness, Sadness, Surprise, Fear, Disgust, Anger, plus later Contempt) as a basis to express any more complex feeling. His work was centered on facial emotions, for which a single positive emotion was enough. However, the recognition of emotions in texts requires a model with balanced categories. This is the case of the work done by Plutchik [18] who associated emotions by pairs: Joy and Sadness, Trust and Disgust, Fear and Anger, Surprise and Anticipation. He also introduced the idea that these families could be expressed at different intensity defining three layers (e.g., Annoyance, Anger and Rage are different degrees in the Anger family) and that they could be mixed together (e.g, Joy and Trust combine to become Love). Nonetheless, retrieving only these 8 families of emotions is not precise enough to attempt to characterize entirely the emotions in our dataset of hotel reviews. For example, we want to be able to differentiate the reasons of Joy such as Amusement, Pride or Pleasure. Moreover, even though Plutchik pairs his emotions with opposite feelings, some of the pairs are composed of two negative emotions (e.g., Fear and Anger). Scherer [20] proposed a solution with 20 different categories, 10 of which are positive emotions and the other 10 are negative ones. This model also benefit from the works of Lang et al. [3, 12] known as the dimensional model. This different theory proposes to label emotions based on two features: valence and arousal. Scherer et al. link these dimensions to their categories and arrange them on a wheel separated in 4 quadrants describing the valence (positive or negative emotions) on the horizontal axis and the arousal (low-control or high-control) on the vertical axis, for each of them (Figure 2). This latter classification meets our willingness to precisely characterize the authors of the user-generated content and obtain a deep understanding of the influence process.

3.3 Emotion Recognition

We focused our approach on the affective content encountered in the reviews to characterize the emotions that the author tried to convey in his writing. As a pre-processing step, we estimated the writing style of the reviews in our dataset with the Flesch Reading Ease [7] but it appeared to be independent of the helpfulness scores ($|r| < 0.03$) and thus we kept the whole dataset for emotion extraction.

Given a review as an input, the text is first represented as a list of words, using the Bag of Words model. These words are matched one by one to a dictionary of emotional words, named GALC [20]. This lexicon is the first one to associate words to emotions of the Geneva Emotion Wheel model. Moreover, it is a general lexicon, i.e. it is independent of the domain of the texts, and thus allows our method to be generic. Words that are associated with an emotion are indexed, and counted. For each review text, our method outputs an emotional profile which consists of a vector con-
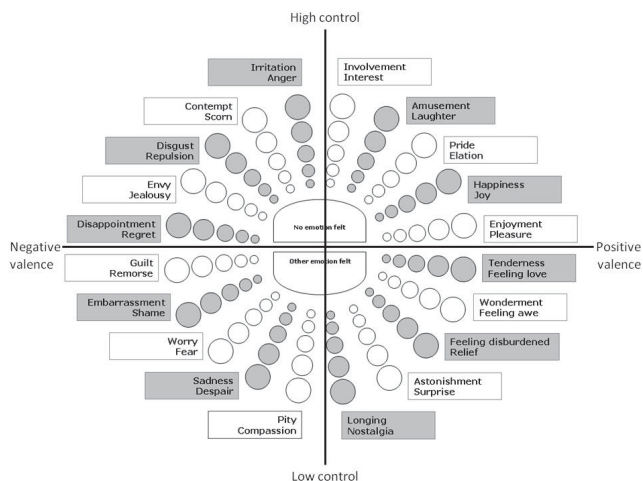

Figure 2: The second version of the Geneva Emotion Wheel [19] selected for its fine-grained classification and balanced choice of emotions represented. The upper half of the wheel describes high-control emotions, while the right half describes positive emotions.

taining the usage quantity of the 20 emotions of the Geneva Emotion Wheel.

3.4 Framework improvements

The sentence "I was absolutely not happy that I paid for a suite that wasn't available." contains the word "happy" that is associated with an emotion (Happiness in this case) but also a word that triggers negation ("not") and one for intensity ("absolutely"). This sentence exhibits the limitations of the simplistic approach of the Bag-of-Word model. Indeed, picking the words one by one doesn't contribute as efficiently to the detection of emotions as complex study of the full text at once.

In order to reach our willingness to fit reality as best as possible, on top of the use of a fine-grained emotion model, we decided to improve the straightforward counting technique with two additional checks: one for negation detection in the surroundings of emotion words, and one for intensity detection. The resulting emotion profiles would thus contain weighted sum of the affective occurrences, which would be written as:

$$\text{emotion}_i = \sum_{w \in \mathcal{R}} \text{int}(w) \cdot \delta_i(w)$$

where $\mathcal{R}$ is the set of words in the review, $\text{int}(w)$ represents the intensity attributed with the word w and $\delta_i(w)$ defines if w is associated with emotion i as:

$$\delta_i(w) = \text{em}_i(w) + \text{neg}(\mathcal{N}_w) \sum_{j \in \mathbb{E} \setminus i} \text{rules}(j \rightarrow i) \cdot \text{em}_j(w)$$

with $\mathcal{N}_w$ the neighborhood of word w, $\mathbb{E} \setminus i$ the list of emotion families except emotion i, and neg, rules and em respectively determining if negation is present in the surrounding, a rule says that emotion i is the converse of emotion j, and if word w is part of the lexicon of the given emotion.

Even though we can easily determine what emotion is not conveyed with the negation surrounding the word (e.g., Happiness in the previous example), it remains a complicated

task to associate the correct emotion in the GEW with such negated sentences. We would be very interested to derive generic rules for the negation of the words in this lexicon. Indeed, if we can show that a sentence containing a particular emotion matches always the same second emotion when it is negated, we could apply this model to any text.

We decided to create a set of rules between families of emotions disregarding symmetries, based on the other emotion words present in the surroundings of the negation term that we try to classify. With this set of rules, our method will look in the surroundings of each word belonging to the lexicon whether there is a word involving the negation of the feeling from a list of commonly used inverters from the NotLW category of the General Inquirer (e.g., "not", "no", "never", etc.). Although this principle worked quite well for the majority of emotion families, the relation between positive and negative emotions was not applicable for all of them. It happened a few times among the whole dataset that different occurrences of the same emotion family had their negations associated with two different emotion families. Thus, we decided to apply the association of positive and negative emotions but we kept in mind that rare cases could be improved with a different technique.

Concerning intensity, we chose to form a small lexicon of intensity keywords (such as "very", "extremely", "slightly", "quite", etc.) and to apply a similar procedure as we did with negation. The principle is to look for such words in the neighborhood of the emotion word but there are no need for inversion rules here. However, we need a scale of factors depending on the intensity word which is found close to the emotional term. For example, "We had very good weather during our stay" would not be detected since no emotional word is present, but "It was a very pleasant stay" would because "pleasant" is part of the GALC lexicon and thus the "very" which precedes it would modify the measure of this emotional term; instead of counting the presence of the emotion Enjoyment/Pleasure once, the method would increase a bit the value of this occurrence based on the intensity lexicon which contains every pair of keyword and amplification factor.

3.5 Review Text of Influential Users

Before discussing the behavior of authors and their review comments in more details, we first consider how to separate all of the reviews into two parts: those considered influential and their counterparts. One method in particular is mostly used in this field.

If a website allows users to vote on a review as either being helpful or not helpful, we can use the method described in [8]. Helpfulness scores in this context are computed as the number of positive votes h over the total number of votes v. A review is considered influential when h/v is greater than the threshold τ. It was shown that the best threshold is 0.6, then a review is considered helpful if more than 60% of the votes it received are helpful. Since the TripAdvisor dataset does not allow users to vote a review not helpful, we cannot apply this method in our case. In the first stage of our work, we thus decided to investigate a method on our own. We first plotted Figure 3, showing the average emotionality of helpful reviews (in blue) versus their counterparts (in red) using values ranging successively from zero to the maximum. By average emotionality, we mean the average number of emotional words found in a given review. This figure helps us
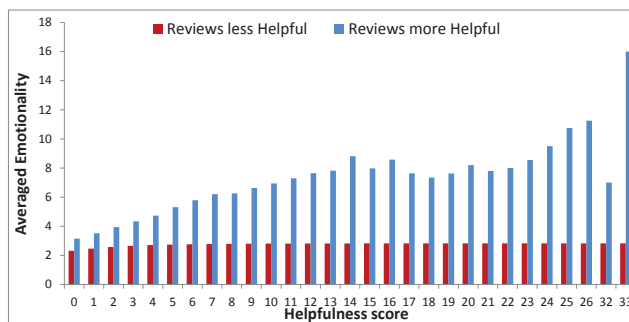

Figure 3: Expected emotionality of aggregated reviews. For each helpfulness score, the red bar represents the average for the reviews which are below this score and the blue one shows the same thing for those above.

see that even if we cut off at zero, i.e., receiving one helpful vote qualifies to be influential, helpful reviews contain significantly more emotions than their counterparts. This was already a surprise to our initial hypothesis.

We further realized that in our current work, we are only interested in the emotionality analysis of two kinds of reviews. Setting the threshold to a particular number will not affect our methods. Moreover we preferred to select a percentage of reviews to be consistent over different datasets, with different number of reviews and scores distributions. This is the reason we decided to cut off at 1%, meaning using the top 1% of all of the comments written for a hotel ranked by its helpfulness scores. This gives us almost exactly $\tau = 8$, meaning a review receiving a helpfulness score over 8 is considered helpful; otherwise, it is considered not helpful. With this threshold, the plot shows an increase of emotional words retrieved of 122% between non-influential and influential reviews as explained above.

4. DISCUSSION

4.1 Emotionality of Review Comments

We decided to represent on a radar plot the use of the 20 emotion categories of the Geneva Emotion Wheel in the reviews to detail the distribution of the emotions. Radar plots would help us detect rather all the emotions are more frequently used in influential comments or if a few emotion families in particular largely outclass non-influential comments while all the others possess a similar behavior for both kinds of reviews. Radar plots have been specially selected regarding their similarity with the GEW representation. The emotionality of the comments can be arranged around a circle keeping the principles of valance and control separation present in the emotion wheel. This visual tool is clear enough to be later presented to the readers as a summary of the review.

Since half of the emotions are positive and the other half are negative on the wheel, we separated the positive and the negative comments for the evaluation. We based our separation solely on the global rating that the reviewer provided: reviews with a rating less than 3 were labeled negative whereas those with a rating higher than 3 were labeled positive. For each of these two subsets (containing 10,649 and 46,249 reviews, respectively), we plotted the average number of emotional signals per comment (i.e., the retrieved emo-
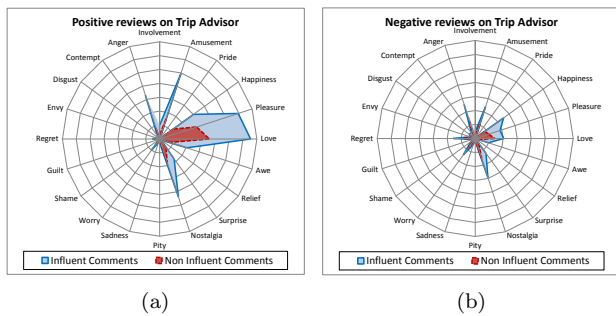
Figure 4: Results of the emotionality of review comments in the Trip Advisor dataset: positive comments are separated on the left part from the negative comments on the right .

tionality) in the plots of Figure 4a, 4b. We observed that for each emotion family, the influential comments are on average making more use of the associated emotion vocabulary, which completely contradicts our initial hypothesis. We also observed that the peaks of negative feelings are present in the negative reviews dataset and not in the positive review dataset. Furthermore, even in the negative dataset we found words associated to positive feelings which symbolize users' overall positivity. Numerically, we observed an increase of 144% of the emotional terms in positive reviews while there is a smaller increase for negative reviews since the quantity of words symbolizing emotion increases by 99% in this case.

4.2 The General Impact of Emotionality

This contradiction with our hypothesis is interesting to notice. However, it is specific to our dataset on TripAdvisor, with a particular emotion lexicon: GALC. To understand the generality of the method, we examined the emotionality of writers in different contexts to confirm our results.

The study of review titles was our first alteration. In our opinion, titles are short messages that are summarizing the whole idea of the review's comment below. We then thought that we would obtain a similar result by applying our method to the titles only. Namely, we thought that influential reviews would more frequently contain emotional words in titles than non-influential reviews. However, we noticed remarquable differences. Actually, positive and negative subsets of reviews are not supporting our hypothesis in the same way. In the positive dataset, we observe that non-influential reviews are using as many positive emotions in the titles than influential ones. However, in the negative dataset, the use of negative emotions is 36% more present in the influential reviews. We notice that the positive emotions have mostly disappeared from the negative dataset, which makes it a better estimator for sentiment analysis than the full review, and that the emotions used are even more targeted than in the graphs of the complete comments since it remains only Worry and Regret. These results made us derive the following initial conclusion. The excessive use of love-related words in review titles can be the reason why voters suspect a certain exaggeration in the reviews, thus leading them to ignore the reviews beginning with such titles.

Then, we changed the topic of the reviews and selected

a dataset from Yelp,[2] a social platform mostly containing restaurant reviews, among which 229,908 reviews were retrieved (extracted from the 2013 Recsys Challenge[3]) for our test.

Interestingly, reviewers on Yelp have the same behavior, namely the influential reviews are containing more emotional words and negative emotions appear more in negative reviews. Love, Pleasure, and Amusement are still the dominant emotions of the positive reviews even though Pleasure is much less frequent. On the contrary, Love remains remarkably present in negative reviews even though it is 40% less present than in the positive reviews. Also, the main negative emotion is the same as the one of the Trip Advisor dataset: Anger. We discovered that the amount of emotions doubled (increase of 100%) in positive reviews when comparing non-influential and influential comments and that negative reviews followed the same behavior with an increase of 116%.

Finally, we also challenged the GALC lexicon, which has been constructed for various applications. We claim that having a lexicon which is domain-specific could improve the emotionality perceived in the comments and the characterization of the influent comments but that GALC is overall a good lexicon for our purpose and allows reusing it to different datasets. We compared the efficiency of the aforementioned lexicon against one that was constructed to predict sentiment about hotel reviews [15].This alternative lexicon was used in a classifier of positivity achieving 95% of accuracy.

In order to compare the two lexicons, we merged all the positive families of the GEW into one and all the negative in a second one. By doing so, we observed that both contradict our initial hypothesis about the accentuated usage of emotional words in non-influential comments. Here again, an increase is visible for any possible threshold, similarly to Figure 3. In comparison with GALC, comments with a helpfulness score of 0 are 36% less emotional than the rest in average and the increase reaches 120% with a threshold of 8. This second lexicon is even better than GALC at recognizing negative feelings in text. However, we are not able to determine if this is Worry, Regret or something else that is retrieved with this analysis of polarity. We claim then that it would be optimal to adapt the GALC lexicon to the review domain (include indirect emotion statement) to obtain even better results about the characterization of negative emotions in hotel reviews.

5. CONCLUSION AND FUTURE WORK

This paper presented a method to characterize reviewers' emotionality at a fine-grained level using our own emotion extraction technique. We compared the emotionality retrieved with our technique to the influence that these reviews had on the community reading them using the helpfulness score of each review. Even though Bag of Word models are simplistic approaches of the extraction problem, we already showed evidence that contradicted our initial hypothesis. We were wrong to suppose that users would be more likely to be influenced by narratives containing fewer emotional terms, presenting the facts in a more objective fashion. We showed that the use of emotions in reviews had a positive impact on how much comments were influential.

---

[2]www.yelp.com

[3]www.kaggle.com/c/yelp-recsys-2013

We discovered that the most helpful comments contained an average of 122% more emotional words than the rest. The same results were observed on two different datasets, and using two different lexicons to prove the methodology's independence. In summary, the key contributions of this paper are:

- the design of a method for the fine-grained recognition of the emotionality expressed by authors in social media, based on a detailed model for the classification of emotions and existent resources;

- the discovery of the positive influence that reviews' emotionality has on readers;

- the highlighting that different parts of reviews do not follow the same writing style and that, for instance, titles and bodies of influential reviews don't possess the same emotion profiles.

The paper also discusses the classification of emotional terms followed or preceded by negative words (such as "no", "not", "never", etc.) and proposes a solution based on the association of positive and negative emotions in order to improve the results. In the future, we believe that it would be beneficial to confirm these associations via crowdsourcing. The most effective way to confirm our negation handling would be to carry out an online user study. During this experiment, we would ask people to give their opinions on the emotions which the author felt, as Sintsova et al. did for the construction of their domain-specific lexicon [21].

For the future, we plan to create a new domain-specific lexicon (or extend the GALC lexicon) and compare it with our current framework for the detection of influence in hotel reviews. We believe that specific lexicons will improve the detection of emotionality and thus will present a clearer picture of its influence.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] H.-i. Ahn, W. Geyer, C. Dugan, and D. R. Millen. "how incredibly awesome!"-click here to read more. In *ICWSM*, 2010.

[2] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec. Effects of user similarity in social media. In *Proc. International conference on Web search and data mining WSDM'12*, pages 703–712. ACM, 2012.

[3] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter?: The Million Follower Fallacy. In *Proc. International AAAI Conference on Weblogs and Social Media ICWSM'10*, pages 10–17, 2010.

[5] C. Danescu-Niculescu-Mizil, L. Lee, and J. Kleinberg. How Opinions are Received by Online Communities : A Case Study on Amazon.com Helpfulness Votes. In *Proc. International conference on World wide web WWW'09*, pages 141–150. ACM, 2009.

[6] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[7] R. Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[8] A. Ghose and P. G. Ipeirotis. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512, 2011.

[9] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proc. International conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[10] E. S. Kim and S. S. Han. An analytical way to find influencers on social networks and validate their effects in disseminating social games. In *Proc. International Conference ASONAM'09*, pages 41–46. IEEE, 2009.

[11] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti. Automatically assessing review helpfulness. In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 423–430. ACL, 2006.

[12] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[13] Y. Liu, X. Huang, A. An, and X. Yu. Modeling and predicting the helpfulness of online reviews. In *IEEE International Conference on Data Mining*, pages 443–452. IEEE, 2008.

[14] S. Mohammad. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proc. Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114. ACL, 2011.

[15] C. Morris. Comparing dictionary-based and corpus-based methods for sentiment analysis, 2011. Unpublished report.

[16] C.-C. Musat, Y. Liang, and B. Faltings. Recommendation using textual opinions. In *Proc. International Joint Conference on Artificial Intelligence IJCAI'13*, pages 2684–2690, 2013.

[17] M. P. O'Mahony and B. Smyth. Learning to recommend helpful hotel reviews. In *Proc. conference on Recommender systems, RecSys'09*, pages 305–308. ACM, 2009.

[18] R. Plutchik. A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*, 1(3):3–33, 1980.

[19] K. Scherer, V. Shuman, J. Fontaine, and C. Soriano. The grid meets the wheel: assessing emotional feeling via self-report. *Components of Emotional Meaning: A Sourcebook*, 2013.

[20] K. R. Scherer. What are emotions? and how can they be measured? *Social science information*, 44(4):695–729, 2005.

[21] V. Sintsova, C. Musat, and P. Pu. Fine-grained emotion recognition in olympic tweets based on human computation. *WASSA 2013*, page 12, 2013.