2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

Emotions in product reviews – Empirics and models

David Garcia
Frank Schweitzer
Chair of Systems Design
ETH Zurich
Kreuzplatz 5
8032 Zurich
Switzerland

***Abstract*—Online communities provide Internet users with means to overcome some information barriers and constraints, such as the difficulty to gather independent information about products and firms. Product review communities allow customers to share their opinions and emotions after the purchase of a product. We introduce a new dataset of product reviews from `Amazon.com`, with emotional information extracted by sentiment detection tools. Our statistical analysis of this data provides evidence for the existence of polemic reviews, as well as for the coexistence of positive and negative emotions inside reviews. We find a strong bias towards large values in the expression of positive emotions, while negative ones are more evenly distributed. We identified different time dynamics of the creation of reviews dependent on the presence of marketing and word of mouth effects. We define an agent-based model of the users of product review communities using a modeling framework for online emotions. This model can reproduce the scenarios of response to external influences, as well as some properties of the distributions of positive and negative emotions expressed in product reviews. This analysis and model can provide guidelines to manufacturers on how to increase customer satisfaction and how to measure the emotional impact of marketing campaigns through reviews data.**

I. Introduction

The recent increase of online commerce and marketing vividly demonstrates the importance of the Internet in human societies, also with respect to economic activities. Rather than in a traditional way, multinational and local companies buy and sell using the Internet. Also, customers find a faster and more independent way of informing themselves about products, purchasing them directly on the Internet. In this respect, online product reviews are a valuable resource of information, because they also allow for user feedback [1]. Such a feedback also allows independent recommendations without direct mediation of the manufacturer.

Market researchers and scientists started to analyse such third-party reviews of products[2], [3] with particular focus on their ratings regarding products and companies. Recently, a statistical model of movie ratings from `IMDB.com` [4] suggests the importance of social impact when rating. In fact, a written review transmits not only factual data and opinions, but also the user's feelings about the product and the brand. The impact of emotions and opinions in book and movie reviews has been analyzed in an individual way, [5], [6] but their collective features are still to be understood. Collective emotions regarding a product become key to predict and to optimize product acceptance. Thanks to review platforms and the way information is exchanged by user interaction, we can observe how certain products become famous, or "beloved" by means of the Internet.

II. Analysing online reviews

`Amazon.com` is not only a top selling platform, it also hosts the largest review community on the Internet, featuring more than 28 million products at the time of the analysis. According to `alexa`, it is the 16th most visited website, and it reached more than 4.5% of the Internet users, who can purchase and review products, especially books and media. Reviews are always accompanied by a star-rating of the product quality, so shoppers can get a third-person independent information of the product quality before buying it. By reporting about their experience online, shoppers also have an impact on the market. We have collected a massive dataset from `Amazon.com` in order to analyse the dynamics of emotions related to products.

A. Data retrieval from `Amazon.com`

During the month of November 2009 we retrieved information about reviews of products of `Amazon.com` through signed requests on its public API. We extracted a list of 16670 products from empty searches in the categories of books, music, DVDs, electronics and photography. After duplicate deletion, almost 1.8 million anonymised reviews were processed. A previous study on `Amazon.com` [7] only had a maximum of 20 reviews per product, while our work focuses on the whole set of reviews for the listed product, some of which received more than 1.000 reviews. Specifically, 42.4% and 91.4% of the reviews are concentrated in only a small set of products, known as the winner-takes-all effect [8] of cultural markets [9].

One of the best features of this dataset is the availability of counted positive and total votes for each review. When a user has read a review, he or she can vote it as helpful on unhelpful, providing a valuable feedback about the quality of the review. From this information, we calculated *helpfulness/unhelpfulness* as the amount of positive/negative votes and stored it together with the star-rating given by reviewer. During the retrieval, the text of each review was processed with emotion detection algorithms, explained below. We then

978-0-7695-4578-3/11 $26.00 © 2011 IEEE
DOI

also stored the positive and negative emotional scores of each review.

B. Emotional classification of reviews

From each review, information about the emotions expressed in its text was extracted by using a lexicon-based classifier called SentiStrength [1] [10]. It has been proven useful to classify emotions in written messages from Myspace and Twitter [11]. This technique uses a human-designed lexicon of emotional terms with a set of amplification, diminishing and negation rules, which are applied if the corresponding terms are detected inside the text. For each review, we extracted a positive and a negative score, which refer to the emotional content rather than to the star-rating. The positive score ranges from 1 (minimum) to 5 (maximum), the negative score ranges from -1 if no negative emotions are present to -5 as a maximum value.

The lexicon we used is an extension of the one used by SentiStrength, with the addition of emotional terms from the ANEW dataset [12]. In this dataset, laymen reported the valence perceived from a large set of words, on a scale of 1 to 9, being 5 neutral valence. We converted the real values of ANEW to the SentiStrength [-5,5] integer scale by means of a linear map, and we ignored the resulting 0 values. Our classification combines the accuracy of a lexicon-based classifier with an additional psychological support from the survey data of ANEW, as the validity of the classification output relies on the size of the lexicon used. The original lexicon contained 938 terms, and combined with the terms from ANEW, we can use a lexicon with 1430 terms.

C. Statistical analysis of collective emotions in reviews

For each review, we have the following variables available: (i) emotional variables, i.e. positive and negative score, (ii) the star-rating, and (iii) the (un)helpfulness of the review as rated by the users. As a first approach to understand the properties of emotions in product reviews, we have calculated the correlation of the different variables mentioned. The results are summarized in the following correlation matrix:

	helpful	unhelpful	rating	positive	negative
helpful	1.000	0.342	-0.047	0.040	0.060
unhelpful	0.342	1.000	-0.312	-0.023	0.071
rating	-0.047	-0.312	1.000	0.074	-0.186
positive	0.040	-0.023	0.074	1.000	0.251
negative	0.060	0.071	-0.186	0.251	1.000

TABLE I: Correlation between variables of each review (individual treatment).

We find that the amount of helpful and unhelpful votes are positively correlated, indicating that there are polemic reviews. This means that there is no unique criterion to evaluate the quality of a review, and the usefulness of the information contained in its text might not be the same for two different users. Interestingly, the amount of helpful votes is not correlated with the rating given in the review, while the amount of unhelpful votes is clearly negatively correlated. After an inspection of the conditional probabilities of unhelpfulness and rating, we have noticed that abnormally high or low rating reviews are more likely to be voted as unhelpful. Hence, votes about helpfulness are indeed useful as they are not given at random and are closely related to the content of the review.

We did not find any clear correlation between the emotional scores from the lexicon-based classifier and the helpfulness votes, which means that there is no trivial impact of the emotional content of a review and its usefulness. But the negative emotional score keeps some negative correlation to the rating value, i.e. the higher the negativity, the lower the rating, which is somehow expected. Additionally, positive and negative scores are positively correlated, even for the two separate lexicons used. This reflects the coexistence of positive and negative emotions in the same review text.

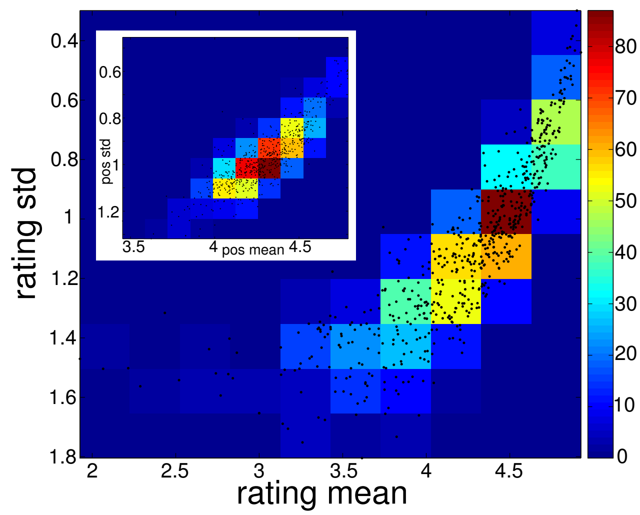

Fig. 1: Mean rating versus standard deviation of the rating for products with more than 50 reviews (collective treatment). Inset: Mean positive score versus standard deviation of the positive score for the same products. Color bins over the scatter plot show the amount of points inside. We note similar properties of aggregated star-ratings and aggregated positive scores.

Given the set of ratings of a product, we calculated the mean and standard deviation of the star-rating values. Their dependence can be seen in figure 1. We notice the concentration of a large amount of products close to a mean rating of 4.5 and a standard deviation of 1. The distribution of the mean positive score versus its standard deviation shows that global rating and positivity are similar, despite that individual review positive score and rating are uncorrelated. The range of values of both variables are similar, and their deviations are comparable, while this is not the case for the negative score. Figure 2 shows the same analysis for the negative score, revealing that the range of values for this aggregated measure

[1] http://sentistrength.wlv.ac.uk

is much wider, as well as its standard deviation.

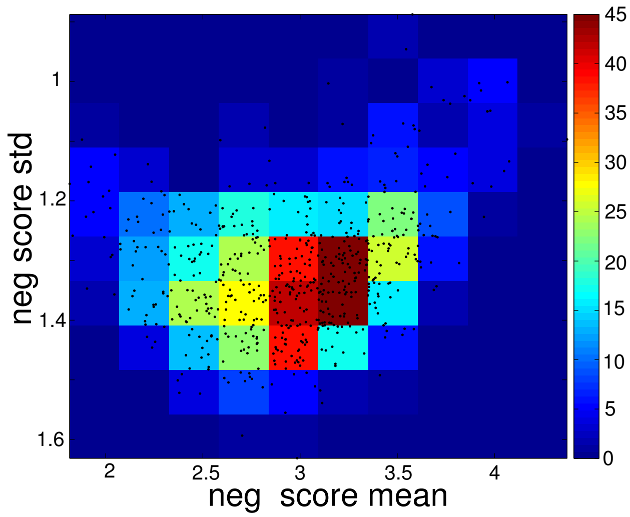

Fig. 2: Mean versus standard deviation of the negative score of products with more than 50 reviews.

To gain deeper insight into the emotional content, we studied the distributions of positive and negative scores of all reviews of some products. In general, for larger amount of reviews these distributions follow asymmetric patterns with respect to positive and negative emotions. Positive emotions show some bias to higher values, having a peak in most of the cases at the maximum value and much lower values for the rest. At the same time, negative emotions appear to be much more evenly distributed across possible values.

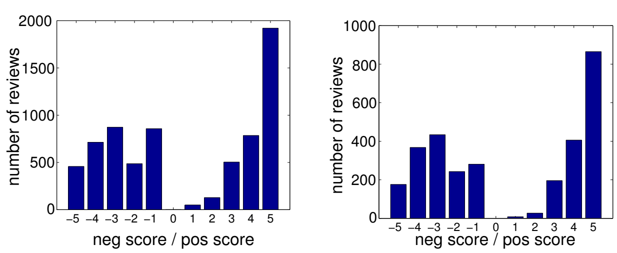

Fig. 3: Distributions of the emotional scores of reviews for “Harry Potter and The Deathly Hallows”, and “Marley and Me”.

The static properties of the collective emotions for products like “Harry Potter and The Deathly Hallows” and “Marley and Me” are very similar, but the dynamics of the user reviews can be very different. For certain high-impact products with more than 1000 reviews we notice the influence of the external media on the emotions of the users, which results into peaks when this exogenous information enters the system. This can be seen for example in the dynamics for “Harry Potter and the Deathly Hallows” a book that received a lot of attention in the mass media before its release. In the upper plot of Fig. 4 we see how this external influence created a strong initial amount of reviews that decreased fast in time.

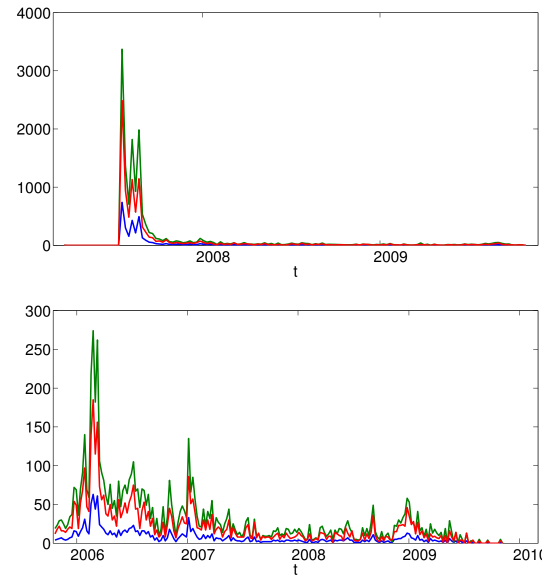

Fig. 4: Weekly statistics for ”Harry Potter and the Deathly Hallows” (top) and ”Marley and Me” (bottom). Amount of ratings (blue), total positive score (green) and total negative score (red) for each week date (x axis).

The review rate before and after the maximum value gives us an indication of the influence of mass media versus the socially mediated word of mouth, which both can result in peaks of reviews. I.e., emotions can grow in the Internet community without external input, as seen in the second example of fig. 4. The book “Marley and Me” was not subject of any important marketing or mass media campaign, but its readers spread information about the book through word of mouth and positive reviews. The increase to the peak is softer than in the case of the media-induced emotions, but also its decay is slower. This is a good example of emergent collective emotion created within the cyberspace, without much input from traditional media.

III. A MODEL FOR EMOTIONAL REVIEWS

We have designed an agent-based model for collective emotions which builds on the modeling framework given in [13]. As sketched in Fig. 5, the framework relates emotional expressions of users and exchange of emotional information with the dynamics of internal emotions, represented by two variables, arousal and valence [14]. Each agent is described by its individual valence v_i describing its pleasure and its arousal a_i describing its level of activity. When the arousal reaches a particular threshold $\mathfrak{T}_i$, the agent expresses its feelings in a written expression modeled by the variable s_i. The value of s_i depends on its valence as described in Table II. Each s contributes to a communication field $h_\pm$ dependent on its positive or negative value, i.e. h_+ “stores” the positive emotions and h_- the negative ones. Further $h_\pm$ can decay exponentially in time. To close the emotional communication, agents are able to perceive the emotions of others through $h_\pm$. We assume that the dynamics is continuous in time. The model we present here differs from the one studied in [13] in the fact that here we try to model collective emotions in product reviews, while the previous model aims at providing a general framework for any online social interaction. As explained later, we introduce particularities of emotional communication in reviews, customer preferences in valence dynamics, and a new activation rule that prevents an agent from creating more than one review per product.

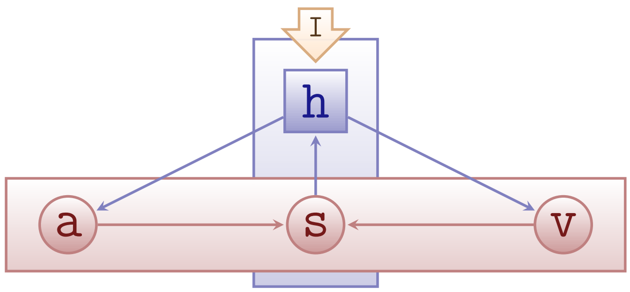

Fig. 5: Causation among the elements of the agent-based model of product reviews emotions.

A. Arousal dynamics

Arousal, i.e. the degree of activity associated with an emotional state, follows the dynamics:

$$\dot{a_i} = -\gamma_a\, a_i(t) + \mathcal{F}_a(h(t)) + A_a\, \xi_a(t) \tag{1}$$

The arousal decays exponentially with parameter γ_a, but is increased by a deterministic contribution $\mathcal{F}_a$, which depends on the emotional information $h_\pm$. The stochastic part $A_a\, \xi_a(t)$ represents the changes in the arousal due to random influences.

It is a particularity of product reviews that users review products only once (anomalous behavior excluded). To adapt the general model to this feature, we define the agents' threshold in a way that they express their emotions only once or never. When the arousal reaches the threshold, the following rule is applied:

$$\text{if} \quad a_i(t) > \mathfrak{T}_i \quad \Longrightarrow \quad \mathfrak{T}_i \leftarrow \infty \tag{2}$$

This way, for finite arousals, the agent will never contribute a review again after the first expression. The thresholds vary among agents in a way that they follow a normal distribution with mean μ and standard deviation σ.

The function $\mathcal{F}_a$ depends on the sum of both fields, using the assumption that emotional communication influences activity regardless of its valence sign:

$$\mathcal{F}_a \propto [h_+(t) + h_-(t)] \sum_{k=0}^{n} (d_0 + d_1 a + d_2 a^2(t)) \tag{3}$$

This way, the arousal is affected by an activity baseline d_0, a linear influence $d_1 a$, and a quadratic saturation if $d_2 < 0$. Previous analysis of the properties of this dynamics [13] showed that, for certain values of the parameters d, the expression dynamics had a one-peak behavior similar to the one showed in the time series of figure **??**.

B. Valence dynamics

The dynamics for arousal and valence we propose are supported by empirical studies in which both variables could be approximated with a stochastic process with exponential decay [15]. Therefore, the equation for the valence is of a similar form as equation 1. The influence of the field in the agent's valence $\mathcal{F}_v$ depends on whether its valence is positive or negative. It means that agents with negative experience of the product will likely develop negative feelings and pay less attention to the positive emotions expressed by the reviews, while agents with positive experiences focus more on the positive emotional information than on the negative one.

An exponential function with a cubic decay $b_2 v^3$, b_2 being negative, represents the asymmetry in the perception of the agent dependent on its valence. We assume the function $\mathcal{F}_v$ to be:

$$\mathcal{F}_v(h_+(t), h_-(t), v_i(t)) = \exp\left(\frac{h_+(t) - h_-(t)}{h_+(t) + h_-(t)} \cdot (b_1 v + b_2 v^3)\right) + b_0 \tag{4}$$

where the parameters have to satisfy $b_1 > 0$, $b_2 < 0$ and $b_0 < 0$ for the desired behavior. $\mathcal{F}_v$ shown in Fig. 6 describes the major contribution of these equation for changes in the valence. It is stronger in the positive v, and it diverges for extreme values of v, this way keeping the dynamics of v inside the interval $(-1, +1)$. This holds when the positive field is larger than the negative one, which is the case for marketing campaigns because of their positive impact. Hence, in our model, h_+ is larger than h_-.

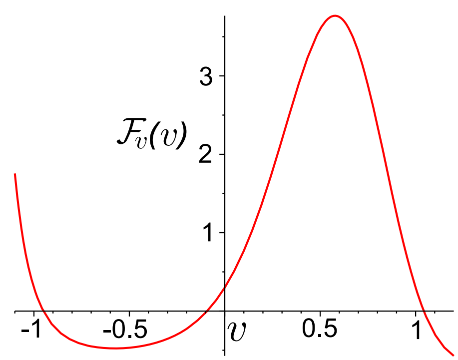

Fig. 6: $\mathcal{F}_v$ (y axis) versus v (z axis) as in equation (4), for $b_0 = -1$, $b_1 = 5$, $b_2 = -5$ under fields $h_+ = 40$ and $h_- = 10$

C. Customer satisfaction and frustration

User preferences and product properties play a key role in the expression of emotions when writing a review. Manufacturers have to deal with the fact that users are heterogeneous and their preferences vary a lot, so the properties of the product and the marketing campaigns give a starting point for the emotions of the reviewers. In the following, we will model the user preference as a agent internal variable u_i constant in time. To model the heterogeneity of agents, we assume that the preferences are uniformly distributed in the interval $[0, 1]$. Take note that preference is subjective in the sense that it simply determines what is preferred and not what is better or worse. The properties of the product are constant for all agents and are described by a parameter $q \in [0, 1]$ that distinguishes products from others.

Initially, an agent starts experiencing the product, which determines its initial valence v resulting from the difference between the agent's expectation u_i and the product property q:

$$v_i(0) = 1 - \frac{2}{max(q, 1-q)}|u_i - q| \qquad (5)$$

An agent that matches its preferences perfectly with the product has $|u_i - q| = 0$ and its initial valence will be extremely positive ($v_i(0) = 1$). On the other hand, if the product results to be completely opposite to what the agent was expecting, the value of the distance will be maximum (q or $1 - q$) and the initial valence will be -1.

D. Intensity of emotional expression

We assume that reviews with higher emotional content have a higher impact on the information field. A review with just factual information is assumed to just have influence on the opinion and information available to the agent, but the influence on the emotions are small. As humans show empathy, the more emotional the review, the larger its influence.

When the arousal of an agent reaches its threshold, the agent creates a review with an emotional content proportional to its valence. For this, we will set the s_i of the agent to a value between 1 and 5. Table II gives the resulting values dependent on the valence.

TABLE II: Expression intensity s given valence v.

v interval	s_-	v interval	s_+
$(-\infty, -0.8]$	5	$[0, 0.2)$	1
$(-0.8, -0.6]$	4	$[0.2, 0.4)$	2
$(-0.6, -0.4]$	3	$[0.4, 0.6)$	3
$(-0.4, -0.2]$	2	$[0.6, 0.8)$	4
$(-0.2, 0]$	1	$[0.8, +\infty)$	5

IV. Simulation of Emotions

A. Review rates

Given the initial value of the field and the dynamics of valence and arousal we can reproduce the two different scenarios in product reviews resulting from mass media versus word of mouth influence. The results are shown in Fig. 7. We notice that the presence of a strong initial input could create a high spike in the amount of reviews followed by a fast decay, as can be seen in Fig. 7 top. In absence of initial information, if the variance of the threshold distribution is large enough, there is a slower increase in the amount of reviews endogenously created within the community. The first agents to write a review are the few ones with a very low threshold, and their activities trigger the purchases and reviews of agents with higher thresholds. The lower diagram of Fig. 7 shows how the review frequency increases for this case.

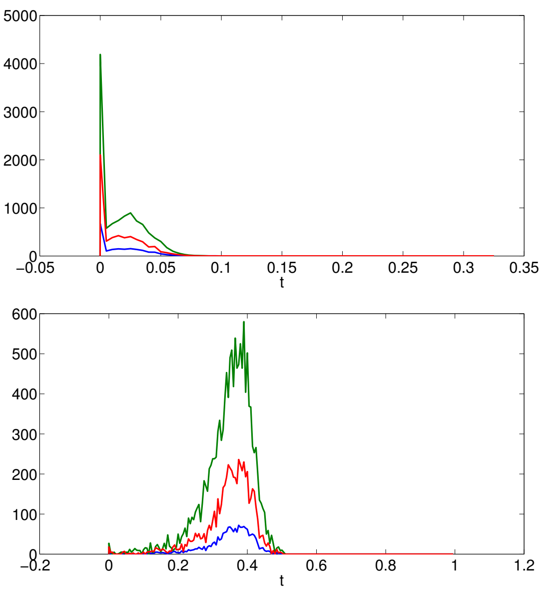

Fig. 7: Amount of ratings (blue), total positive expression (green) and total negative expression (red) for the simulated time. Rate of reviews and emotions for a strong media impulse (top) and when the emotions spread through the community (bottom).

B. Distributions of emotions

For certain parameter values, our model reproduces the distributions of emotions we measured in real-world review communities. This comparison is shown in Fig. 8. We find that the distribution of emotions in the simulations have the same bias in the positive part while they are more evenly distributed in the negative part. Our simulations indicate that there is an important herding effect present when perceiving positive emotional content in product reviews, while the expression and perception of negative reviews does not influence the agents' negative valences that much.

To conclude, in Fig. 8 we show for the distribution of emotions that the outcome of our model has the same macroscopic properties as found in the real world data. This model provides a phenomenological explanation based on psychological principles that links the dynamics of emotions with the collective behavior observed in product reviews. Here, we verified that the model can have the same qualitative properties as the reviewer community of `Amazon.com`, but its predictive power keeps to be explored. Our model allows for the exploration of a psychometric parameter space, from which

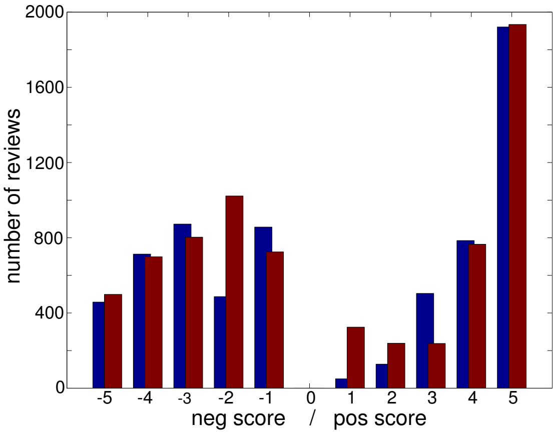

Fig. 8: Comparison between the emotional distribution of the reviews for "Harry Potter"(blue) and the simulation results (red).

the most likely values of the parameters can be found for each product. Following the appropriate approach, future analyses of this model can provide predictive results that could assist community managers and manufacturers in their decisions regarding customer reviews and satisfaction.

V. Conclusions

Thanks to a massive retrieval of product reviews available from `Amazon.com`, we were able to mine important information: helpfulness, rating and emotions. We quantified the evidence of polemic reviews, i.e. some users found them helpful and others not. We further identified the correlation between rating and helpfulness. We also found that positive an negative emotions can coexist in some reviews.

Generally, the collective expression of positive emotions in product reviews shows a strong bias towards positive values, while the negative content is more evenly distributed along its possible values. From the dynamics of the amount of reviews we identified different responses of the community dependent on the presence of marketing campaigns and word of mouth effects.

We modeled the behavior of users and product review communities using a modeling framework for emotions in online communities. This model can reproduce both possible scenarios in response to external influences. Further, the distributions of expressed emotions have similar properties in the simulations and in the data. Our model can be fitted to know the internal dynamics of the customers that reviewed them, as well as the properties of the social interaction they have online. Manufacturers can use this useful information to understand better the way their customers react to their products, and to derive norms and principles to follow in order to maximize customer satisfaction and sales. Our statistical analysis can serve as a comparison for community managers of reviews communities. In particular, understanding the dynamics of user arousal is key for encouraging participation. Our findings can serve as an example for other platforms in order to imitate `Amazon.com`'s success.

There is still the open question how this emotional online communication influences product sales, and whether the response of the community shows differences for successful products. It would be interesting to quantify media and marketing impact in the reviewer community, in order to evaluate the efficiency of marketing on consumer emotions. Our model provides a first step towards understanding online consumer interaction beyond simple ratings. This kind of understanding could help manufacturers to know how to satisfy their customers, as well as give insights to social scientists on how online emotional communication works.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7-ICT-2008-3 under grant agreement no 231323 (CYBEREMOTIONS).

References

[1] D. Wenjing, G. Bin, and A. B. Whinston, "Do online reviews matter? An empirical investigation of panel data," *Decision Support Systems*, vol. 45, no. 4, pp. 1007–1016, 2008.

[2] X. Li and L. M. Hitt, "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, vol. 19, pp. 456–474, Dec. 2008.

[3] F. Wu and B. A. Huberman, "Opinion formation under costly expression," *ACM Transactions on Intelligent Systems and Technology*, vol. 1, pp. 1–13, Oct. 2010.

[4] J. Lorenz, "Universality in movie rating distributions," *The European Physical Journal B*, vol. 71, pp. 251–258, Aug. 2009.

[5] D. Nahl and D. Bilal, *Information and emotion: The emergent affective paradigm in information behavior research and theory*, vol. 49. Asist, 2007.

[6] C. Dellarocas, X. M. Zhang, and N. F. Awad, "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," *Journal of Interactive Marketing*, vol. 21, no. 4, pp. 23–45, 2007.

[7] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, vol. 45, (Prague, Czech Republic), pp. 440–447, Association for Computational Linguistics, 2007.

[8] S. Rosen, "The Economics of Superstars," *The American Economic Review*, vol. 71, no. 5, pp. 845–858, 1981.

[9] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web*, vol. 1, p. 39, May 2007.

[10] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology*, vol. 61, pp. 2544–2558, Dec. 2010.

[11] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in Twitter events," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 2, pp. 406–418, 2011.

[12] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," tech. rep., The Center for Research in Psychophysiology, University of Florida., 1999.

[13] F. Schweitzer and D. Garcia, "An agent-based model of collective emotions in online communities," *The European Physical Journal B*, vol. 77, pp. 533–545, Oct. 2010.

[14] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[15] P. Kuppens, Z. Oravecz, and F. Tuerlinckx, "Feelings change: Accounting for individual differences in the temporal dynamics of affect," *Journal of personality and social psychology*, vol. 99, no. 6, pp. 1042–1060, 2010.