First International Conference on Advanced Research Methods and Analytics, CARMA2016
Universitat Politècnica de València, València, 2016
DOI: http://dx.doi.org/10.4995/CARMA2016.2016.3642

Identification of Influencers in eWord-of-Mouth communities using their Online Participation Features

Olmedilla, M.[a]; Arenas-Marquez, F. J.[b]; Martinez-Torres, M. R.[a] and Toral, S. L.[c]
[a]Departamento de Administración de Empresas y Comercialización e Investigación de Mercados (Marketing), Universidad de Sevilla, Spain, [b]Departmento de Economía Financiera y Dirección de Operaciones, Universidad de Sevilla, Spain [c]Departmento de Ingeniería Electrónica, Universidad de Sevilla, Spain.

Abstract

The identification of influencers in any type of online social network is of paramount importance, as they can significantly affect consumers' purchasing decisions. This paper proposes the utilization of a self-designed web scraper to extract meaningful information for the identification of influencers and the analysis of how this new set of variables can be used to predict them. The experimental results from the Ciao UK website will be used to illustrate the proposed approach and to provide new insights in the identification of influencers. Obtained results show the importance of the trust network, but considering the intensity and the quality of both trustors and trustees.

Keywords: *e-word of mouth; influencers; Social Network Analysis; virtual communities.*

1. Introduction

The emergence of user-generated content has facilitated the interactions among users, so they can easily share opinions and exchange experiences. In this regard, the electronic interactions are complementing traditional word of mouth (WOM). The importance of WOM is widely accepted in traditional marketing research (Lee et al., 2008) and it is usually considered to be a very effective marketing tool with major repercussions on consumer behavior. However, it has evolved to a more impersonal but more pervasive form of WOM, the so-called electronic worth-of-mouth (eWOM), which is based on technology information advances and the growing access to the Internet (Law et al., 2014). eWOM is also providing an alternative and effective marketing channel to firms, which does not require huge investments in advertising (Ku et al., 2012). As a result, the identification of possible influencers is of great interest to business given the importance and impact that their reviews can cause on other consumers' purchase intentions. Marketing information can be propagated faster and promoted better via recommendations by influencers to their followers and peers (Cheung & Thadani, 2012). Previous approaches for the identification of influencers have been mainly focused on the idea of trust (Kim & Tran, 2013) and the degree of expertise in a specific domain (Ku et al., 2012). However, modern computational techniques can collect much more information about the social networking practices of users within these communities. For instance, the reputation of users can be measured using the ratings that their reviews receive from the rest of the community. Popularity is another feature of users that can also be measured using several metrics such as the number of comments or the number of readings received.

In this paper, we propose using a combination of reputation and popularity. Collected information can also enrich the dependent variables of the study. Typically, the trust network was included in previous studies by considering the size of the trust network. However, more recent works propose studying the trust network as a 2-hop network, considering also the quality of trustors (Kuk et al., 2012). Moreover, the social networking practices of users also include the possibility of scoring other posted reviews and trusting other users. This information is also publicly available in many eWOM websites. Finally, the domain in which users post their reviews can also be collected, considering different domain levels. All these new variables will be considered

The remainder of this paper is organized as follows. Section 2 details the related work about eWOM, the collection of information and the identification of influencers. The proposed methodology for collecting information and the definition of collected variables are detailed in section 3. Section 4 presents the empirical work and reports the evaluation results. The last section concludes the paper by summarizing the most important features of the proposed approach and by suggesting future research directions.

2. Related work

eWOM websites provide tools for consumers to discuss products and learn from other customer how to better use them (King et al., 2014). Among others, the online reviews usually include aspects such as a main text with the comments about the product, a general rating and the scoring of certain attributes and key phrases related to the product's perceived weaknesses and strengths. Additionally, some consumer-opinion websites include mechanisms that report reviewers' reputation (e.g. ratings received from other consumers) and allow members to add other members to a trust network (Ku et al., 2012).

Influencers are usually early adopters in markets, have multiple interests and are trusted by other consumers in a wide social network (Kiss & Bichler 2008). One major challenge of eWOM research consists in determining the characteristics that are more suitable for identifying influencers. Reviewer's exposure in the eWOM community (usually measured by how many times a user posts reviews on the website) is an important magnitude in previous studies. Hu et al. (2008) state that consumers pay more attention to reviewers with high exposure and their reviews are more likely to change consumers' uncertainties and transaction costs for buying a product. Lu et al. (2010) indicate that the number of reviews contributed by focal members positively correlates with the helpfulness of their reviews. Meanwhile, Huang et al. (2010) state that when a user has a great expertise in a field, she or he often writes more reviews on that specific field.

A number of papers suggest that reviewers' degree of expertise positively relates to their reputation and is likely manifested in their review behaviour. From this point of view, probably a high-level reviewer is a very active contributor in a certain product category or domain (Ku et al., 2012; Martínez-Torres & Diaz-Fernandez 2013). Arenas-Marquez et al. (2014) conclude that influencers usually review a wider range of products (i.e. products of different brands, technical features or benefits), which reflects their greater expertise with regard to a certain domain. Hung & Yeh (2014) state that influencers often post useful and knowledgeable contents. Therefore, these authors propose a text mining-based approach to evaluate features of quality of information and to identify influencers.

Finally, other existing works are based mainly on social network analysis. These papers study the topological features of the network formed by registered users within the consumer platform to identify influencers. Luarn et al. (2014) examine the influence of Facebook user's networks on the dissemination of information. They conclude that users with high network degree (more connections) and high clustered connections (frequency of information dissemination) have a greater influence on the dissemination process.

3. Methodology

After developing a set of tools for crawling the eWOM website, the gathered information must be transformed into a structured data format. The aim is obtaining a set of metrics representing the social networking practices of users from the collected information, which are going to be used for modelling and analysis in subsequent stages. The website includes some structured statistical information such as the number of reviews written, the review rating values or the number of reads received that are directly obtained from the programmed crawler. The user trust relationships (number of users trusted and number of users trusted-by) were obtained from the circles of trust information. Table 1 lists the variables considered in this study.

Table 1. Metrics describing the social networking practices of users.

Variable	Description
CritCap	∑ rating scores given per user
Tint	Size of the trust network
AvTInt	∑ network size of trustors / Size of the trust network
Tint-by	∑ users trusted-by
AvTInt-by	∑ network size of trustors / ∑ users trusted-by
ExpertCat	∑ categories of posted reviews per user
ExpertSubCat	∑ subcategories of posted reviews per user
MaxExpertCat	Maximum number of reviews in one category

- Critical capacity (CritCap). This variable is obtained as the sum of the rating scores given by each user.
- Trust intensity (Tint): Number of members who trust a given user (the size of his circle of trust).
- Average trust intensity of trustors (AvTInt): It is the average trustworthiness of all the trustors of a given user (i.e., average trust intensity of members who trust this user).
- Trust-by intensity (Tint-by): Number of users trusted by a given member (i.e., this user is included in other circles of trust)
- Average trust intensity of trustees: (AvTInt-by) It is the average trustworthiness of all the users trusted-by a given user.
- Level of expertise per category (ExpertCat): Total number of distinct categories in which a reviewer has written.
- Level of expertise per subcategory (ExpertSubCat): Total number of distinct subcategories in which a reviewer has written.
- Maximum level of expertise (MaxExpertCat): Maximum number of reviews written by a reviewer in a particular category.

4. Results

A crawler that follows the hyperlink structure of the users' webpages at Ciao has been developed using Scrapy with Python. As a result, the whole website Ciao.co.uk was crawled gathering information from about 45 thousand registered users within Ciao UK.

Although the number of registered users at Ciao UK is about 45 thousands, only a fraction, 12886 users, has posted at least one review. This is the typical participation inequality exhibited in many virtual communities (Martinez-Torres, 2013). However, the number of users posting only one review is still quite high. Therefore, in this study, we have filtered the original data and we have only considered those users posting more than one review. The number of users accomplishing this condition is 3158.

The condition of being an influencer can be defined in terms of reputation and popularity, following the studies by Kuk et al. (2012) and Arenas-Marquez et al. (2014). The reputation was measured by the average value of the received rating scores and the popularity by the average value of comments received. In this study we have considered two different thresholds given by the percentiles 90 and 95. More specifically, two definitions of influencers will be considered, as given by equations (1) and (2):

$$Infl90 = Reputation90 \,\&\, Popularity90 \qquad (1)$$

$$Infl95 = Reputation95 \,\&\, Popularity95 \qquad (2)$$

Infl_{90} considers as influencers those users located in the percentile 90 of both reputation and popularity, while Infl_{95} considers the percentile 95. Note that in both cases is a dichotomous variable, which takes the value 1 when the double condition is accomplished and 0 otherwise. The number of obtained influencers is 190 in the case of percentile 90 and 76 in the case of percentile 95. As we have a dichotomous variable, a binary logistic regression is appropriate to determine the variables that characterize the behaviour of influencers. However, obtained results show that influencers only represent a small fraction of community users. That means that the dependent variable contains a high number of zeros (which is the value for non influencers) and a low number of ones (which is the value for influencers). This kind of problems, where the dependent variable contains a disproportionally high number of zeros, are known as zero inflated problems, and they can lead to biased/inconsistent parameter estimates, inflated standard errors and invalid inferences (Lee et al., 2006). A possible alternative consists in considering generalized linear modelling with Poisson distribution. However, generalized linear modelling with Poisson distribution has problems with overdispersion (Hinde & Demetrio, 1998). The model with negative binomial distribution is an alternative way to fix over-dispersion problem in Poisson distribution (Hinde & Demetrio, 1998), as the variance and mean are not assumed to be equal. This is the chosen regression model for this study. Obtained results for the two definitions of influencers are detailed in Table 2.

Table 2. Negative binomial regression results for the influenc-ers measured with the percentile 90 and 95.

	Dependent variable:	
	Infl_{90}	Infl_{95}
CritCap	-0.001^{***}	-0.000
	(0.000)	(0.000)
Tint	0.01^{***}	0.03^{***}
	(0.002)	(0.002)
AvTInt	0.1^{***}	0.1^{***}
	(0.004)	(0.01)
Tint-by	0.03^{***}	0.04^{***}
	(0.004)	(0.005)
AvTInt-by	0.01^{***}	0.01^{***}
	(0.002)	(0.004)
ExpertCat	0.1^{***}	0.1^{***}
	(0.01)	(0.02)
ExpertSubCat	-0.01^{**}	-0.03^{***}
	(0.003)	(0.01)
MaxExpertCat	0.004^{***}	-0.002
	(0.001)	(0.003)
Constant	-4.9^{***}	-6.0^{***}
	(0.1)	(0.2)
Observations	3,158	3,158
Log Likelihood	-434.1	-214.0
Akaike Inf. Crit.	886.1	446.1
Bayesian information criteria (BIC)	940.6	500.6
Precision	0.415	0.621
Recall	0.513	0.237
Overall	0.971	0.978
McFadden R^2	0.553	0.520
Note:	**p<0.1; **p<0.05; ***p<0.01*	

Results from Table 2 show that the critical capacity is not an important feature of influencers. Only in the case of Influ_{90} definition there is a significant negative relationship, but with a very low coefficient. The negative relationship is the one expected, as the influencers are supposed to have a good knowledge about the reviews they are scoring. More important is the influence of the circles of trust. The same than previous studies (Ku et al., 2012; Liu et al., 2015), both the trust intensity and the average trust intensity of trustors have a significant and positive influence over the condition of being an influencer. However, and as a novel contribution of this paper, the trust-by intensity and the average trust intensity of trustees also show this significant positive relationship, although at a lower level. Therefore, it is not only important the size of the circle of trust, but also the quality of this circle of trust, which means that people trusting an influencer also have a high circle of trust.

Table 2 shows a positive and significant relationship with the number of categories where the reviewer posts his or her reviews, but a negative relationship with the number of subcategories. According to previous studies, influencers exhibit a high level of expertise, which means they should focus on specific categories. Therefore, a negative relationship with the number of categories (ExpertCat), subcategories (ExpertSubCat) and the maximum number of reviews (MaxExpertCat) is expected. However, Table 2 only shows a negative relationship with the number of subcategories. This result can be explained by the way categories and subcategories are defined at Ciao. Ciao establishes 28 categories and the subcategories are then defined by reviewers. That means the main categories have a wide scope, so it is easy that a reviewer posts reviews belonging to several main categories.

4. Conclusions

This paper proposes a methodology for collecting user-generated content within eWOM websites in order to extend the number of variables usually considered for the identification of influencers. The data collection is based on the design of a self-programmed crawler to access the meaningful information related to the social networking practices at eWOM. Obtained results show the importance of the trust network, but considering the intensity and the quality of both trustors and trustees. We have also confirmed the low relevance of the critical capacity and the specialization of influencers but considering the level of subcategories rather than the level of the main categories.

References

Arenas-Márquez, F. J., Martínez-Torres, M. R., Toral, S. L. 2014. Electronic word-of-mouth communities from the perspective of social network analysis, Technology Analysis & Strategic Management, 26 (8): 927-942.

Cheung, C.M.K., Thadani, D.R. 2012, The impact of electronic word-of-mouth communication: A literature analysis and integrative model, Decision Support Systems, 54(1):461–470.

Hinde, J., and Demetrio, C. 1998, Overdispersion: Models and Estimation, Comput. Statistics and Data Analysis, 27:151-170.

Hu, N., Liu, L. and Zhang, J.J. 2008. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. Information Technology and Management 9(3):201−214.

Huang, S., Shen, D., Feng, W., Baudin, C., Zhang, Y. 2010. Promote product reviews of high quality on e-commerce site, Pacific Asia Journal of the Assoc. for Information Systems 2(3):51-71.

Hung, C. Yeh, P.W. 2014. Identification of opinion leaders using text mining technique in virtual community. 1st Symp. on Information Management and Big Data, Cusco (Peru), 1318:8-13.

Kim, Y. S., & Tran, V. L. 2013. Assessing the ripple effects of online opinion leaders with trust and distrust metrics. Expert Systems with Applications, 40(9): 3500-3511.

King, R.A., Racherla, P., Bush, V.D. 2014. What we know and don’t know about online word-of-mouth: a review and synthesis of the literature. Journal of Interactive Marketing 28:167-183.

Kiss, C. and Bichler, M. 2008. Identification of influencers - Measuring influence in customer networks. Decision Support Systems 46:233–253.

Ku, Y.C., Wei, C.P. and Hsiao, H.W. 2012, To whom should I listen? Finding reputable reviewers in opinion-sharing communities, Decision Support Systems, 53: 534–542.

Law, R., Buhalis, D. and Cobanoglu, C. 2014, Progress on information and communication technologies in hospitality and tourism, International Journal of Contemporary Hospitality Management, 26(5): 727-750.

Lee, A.H., Wang, K., Scott, J.A., Yau, K.K., McLachlan, G.J. 2006. Multi-level zero-inflated poisson regression modelling of correlated count data with excess zeros, Statistical Methods in Medical Research, 15:47–61.

Liu, S., Jiang, C., Lin, Z., Ding, Y., Duan, R., & Xu, Z. 2015. Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach. Information Sciences, 306: 34-52.

Lu, Y., Tsaparas, P., Ntoulas, A. Polanyi, L. 2010. Exploiting social context for review quality prediction. International Conference on World Wide Web.

Luarn, P., Yang, J,C., Chiu, Y.P. 2014. The network effect on information dissemination on social network sites. Computers in Human Behavior 37:1–8.

Martínez-Torres, M. R. 2013. Application of evolutionary computation techniques for the identification of innovators in open innovation communities. Expert Systems with Applications, 40(7): 2503-2510.

Martínez-Torres, M.R., Díaz-Fernandez, C. 2013. Current Issues and Research Trends on Open Source Software Communities. Technology Analysis & Strategic Management 26(1):55-68.