DETECTING MICROBLOGGING COMMUNICATION STRATEGIES USED BY ONLINE RETAILERS

Georg Lackermair
XT AG
E-mail: georg.lackermair@xt-ag.de

KEYWORDS

Microblogging, E-commerce, communication strategies;

ABSTRACT

The increasing popularity of the Social Web drives the transition of E-commerce to transactions mediated by social web services. This evolution is coined as Social Commerce. Due to its popularity, Twitter microblogging is an important channel for online retailers to communicate with customers. In order to determine the current state of adoption of Twitter, this paper introduces and evaluates a method for assigning an observed account with an E-commerce specific communication strategy.

1 INTRODUCTION

Since the term Web 2.0 was at first defined by O'Reilly (2005), the transformation of the Web of linked documents to an interactive communication platform has received a lot of attention. The popularity of services in the social web can be illustrated with impressive numbers: LinkedIn reached 225 million registered users in 2013 (see LinkedIn 2014); Facebook reports 1,230 million active users per month (see Sedghi 2014); Twitter states that 500 million message are delivered and 255 million users are logging in to the platform every month (see Twitter 2014a).

This development had major impacts on E-commerce. The user in the social web demands the availability of capabilities for quick and direct communication with the selling companies and with other consumers. As a consequence, retailers began to make efforts to become visible in the most popular networks. Those new requirements lead to an evolution from the E-commerce to the social commerce, which describes the emergence of social features on shopping-websites and the increasing commercial activities in online social networks.

Due to its popularity, Twitter is considered as a major platform for this change. Reasons for its success can be found in its simplicity, scalability, ubiquity and interactivity. Due to its publish/subscribe capabilities, traditional newsfeeds based on Rich Site Summary (RSS) or Atom are shifting gradually to Twitter. This development applies to E-commerce as well, as within this domain microblogging plays the role of a update notification capability for tools for personalization and direct customer interaction, e.g., discussion boards, weblogs or newsfeeds.

Twitter microblogging has received a lot of attention in the scientific community. In particular, the use of specific communication conventions, communication purposes and the network structure were subject of many works. Though, little is known about the use of Twitter microblogging within an E-commerce specific context. This paper introduces a method for analysis of the status messages of a given retailer in order to assign a basic communication strategy. An evaluation of the accuracy of the design is also shown. For future research the findings can be used to implement an observatory that enables the analysis of developments in a long-term examination.

This paper is structured as follows: Section 2 summarizes the theoretical background and describes the designed algorithm for classification of communication profiles. After that, the evaluation and it's results are summarized in Section 4. The paper concludes with an interpretation and contributions in Section 5.

2 BACKGROUND

2.1 Social Commerce

The main research areas addressed in this paper comprise the Social Web and online commerce. The emerging Web 2.0 has major impact on the E-commerce and shapes it in various ways (see Wigand, Benjamin, and Birkland 2008). This influence is often denoted as *Social Commerce* (see Bächle 2008; Richter, Koch, and Krisch 2007). Curty and Zhang (2011) define Social commerce as commerce that is mediated by social media . This definition reflects the growing integration of interactive media in the life-cycle of E-commerce transactions. According to Wang and Zhang (2012), customer generated reviews and ratings that are provided by the seller in the shopping websites were the first embodiment of this development.

This evolution of E-commerce is addressed by academic literature regarding the users, business strategies, information and technology (see Curty and Zhang 2011; Zhou, Zhang, and Zimmermann 2013). On user level, various works investigate the acceptance of and the attitudes concerning social commerce. Regarding business strategies, popular subjects of study are marketing strategies, revenue models and collective buying strategies. The information aspects focuses on the characteristics of *User Generated Content (UGC)* and analysis of the structure of social networks. The technology aspect considers social media tools, that provide various functionalities (see Huang, Yoon, and Benyoucef 2012): Social connection allows for creating links between E-commerce and social networking sites. Social community refers to communication capabilities, that allow user with similar interests to meet and form a group. Social media marketing provides features that support the dissemination of advertisements in social networks. Social application could help to integrate mobile commerce.

A design model for social commerce derived from design principles for Web 2.0 and E-commerce applications consists of the elements individual, conversation, community, and commerce (see Huang and Benyoucef 2013):

Individual The individual aspect aggregates features, that allow for the management of person-centric profiles, e.g., user profiles, or activity logs for a user. The latter is implemented on basis of a microblogging functionality (see Section 2.2).

Conversation Social commerce application provide various conversation features for the users. Important purposes for the use of those mechanisms are content creation and information sharing.

Community As basis for the effective use of provided communication channels, capabilities for managing the community are necessary. Those capabilities allow to establish connections between single users or groups of users. There are also various mechanisms to maintain those relationships.

Commerce The business functions could comprise of capabilities for collective buying, the management of social advertisements and feedback functionalities for participation in the value creation process.

2.2 Twitter Microblogging

Previous works have investigated that Twitter is used for various communication purposes. Kwak et al. (2010) categorize Tweets by the use of communication techniques: Retweet, Reply, Mention, and, Singleton . The Retweet technique is used to disseminate messages, written by others among the followers. Replies and mentions direct a message to one ore more named users. Mentions and replies can be identified by the use of the @-sign, which is called *User Mention (UM)* by Twitter. Although, this notation is also used for other purposes, e.g., referring to a place, the vast majority of @-signs are used for mentioning users in Twitter (see Honeycutt and Herring 2009). According to Edman (2010, p. 21–25) is this notation a measure of interactivity of an online relationship. If a Tweet is neither a Retweet or a Reply, Kwak et al. (2010) classify the message as singleton. Hong, Convertino, and Chi (2011) compare the use of those communication techniques by languages.

Westman and Freund (2010) and Paul, Hong, and Chi (2011) study the content of Twitter messages and derive various genres. Status messages can be classified in five different genres. Personal updates contain informations about what a person is doing, thinking or feeling at the moment. In directed dialogs, users embed UMs to direct messages to the named users. Real-time sharing describes the sharing of news and URLs among the user's followers. Business broadcasting describes the dissemination of advertising messages. Information seeking means the behaviour of asking questions in order to receive useful answers. Paul, Hong, and Chi (2011) found out, that the latter category plays only minor role in Twitter communication.

The recipients of a Tweet can be extended beyond following users by embedding a *Hashtag (HT)* in a status message. According to Yang et al. (2012) label HTs either the message's subject or a community. For an automatic classification of status messages, the Twitter platform extracts Hashtags, Usermentions and URLs from the message text and provides access to those entities in separate data fields (see Twitter 2013).

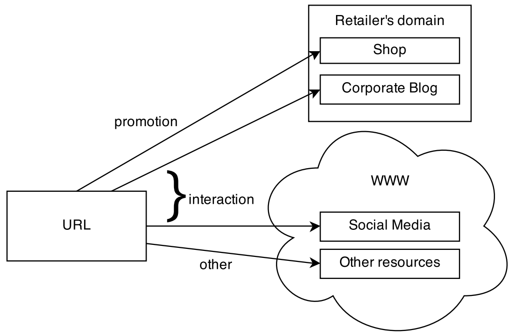


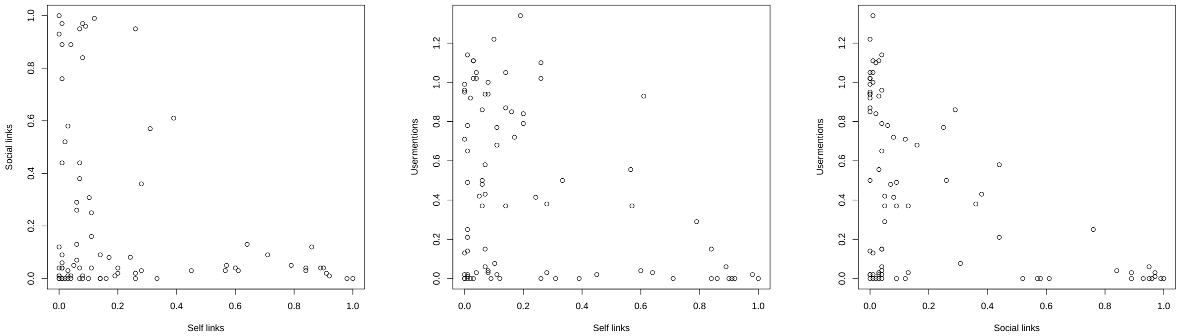
Figure 1: Classification of embedded URLs.

3 ASSIGNING STRATEGIES

Based on Twitters capabilities, UMs are embedded in order to address a single user in public. Thus, UMs are an indicator for dialogs between a company and a user. In frame of E-commerce transactions UMs are used by customers to express feedback about product quality, or to request information, e.g. about dates of delivery, etc.

Another interesting aspect about the communication on Twitter are URLs that are embedded in Tweets. URLs are by default shortened by the platform's own shortening service http://t.co (see Twitter 2014b). Figure 1 classifies the targets of those URLs from an E-commerce perspective. The links can be categorized as self-links, social media links and other external links. For the study presented, the former two contain interesting information. A *self-link* points to the own website of an online retailer and indicates the promotion of a product. *Social media* URLs direct users to discussions on other social networks, e.g., Facebook. This indicates a more community-centric activity than links to product pages.

For analysis, first, a classification is performed on level of a single message. For that purpose, the presence of an



(a) Social Links and Self Links. (b) Usermentions and Self Links. (c) Usermentions and Social Links.

Figure 2: Data distributions for UM, social links and self links ($n = 82$).

UM, as well as the classification of embedded URLs is captured by the following variables: M, P, S. M indicates the presence of a mentioned user in the status message. P captures self-links, and S URLs, that point to social media sites. For collection and data analysis, a script was implemented. The algorithm is structured in two subsequent steps. First, all necessary data is collected. The script fetches the timeline from Twitter's RESTful API for each account in the list of retailers. For this work, 82 Twitter-accounts where selected, that are owned by retailers in Germany and UK. Each of the companies is a top-selling retailer in the corresponding country (see Lackermair and Kailer 2014b; Lackermair and Kailer 2014a).

For each retrieved status message, the variables M, P, and S are determined. The latter require an analysis of the targets, where embedded URLs point to. For this purpose, the algorithm resolves shortened URLs to reveal the target web site. Antoniades et al. (2011) and Klien and Strohmaier (2012) claim, that shortened URLs can be detected by redirecting links that return the HTTP status codes 301, 302, or 307. The classification of self-links and social web links is then performed by several regular expressions.

In a second step, the algorithm performs the analysis and strategy assigment on retailer level. For this purpose, two subsequent steps are performed. First, the values for M, P and S are accumulated per retailer to a communication profile. Such a profile consists of the percentages of each variable. The corresponding values are captured in the variables P_M, P_P and P_S. The data distribution of these variables is shown in Figure 2. The distributions indicate a weak correlation between the variables. Thus, these values can be used to assign communication strategies to each profile in a second step. For this purpose, rules based on threshold values are applied. This work is focusing on the identification of *dominant* patterns. Thus, the threshold values .33 and .66 were selected to classify weak, medium and strong use of a specific communication technique. In order to label the communication patterns, basic categories are defined, that differentiate between *interactive* patterns and *advertising* activities. Interactive activities include the use of status messages for directed dialogs, as well as directing users to social media platforms that emphasize interactivity and social participation (Laudon, Laudon, and Schoder 2010, p. 388). Thus, the following strategies are defined:

- St_1 – *Twitter dialog:* The retailer predominantly directs messages to single users ($P_M \geq .66$).
- St_2 – *Direct to social media:* The status messages embed URLs that link to social media platforms ($P_S \geq .66$).
- St_3 – *Media and dialog:* If neither St_1 or St_2 is assigned, but $P_M + P_S$ dominate, a combined, interactive strategy is assigned ($P_M + P_S \geq .66$).
- St_4 – *Advertising:* Embedding of URLs that point to product sites in the online shop ($P_P \geq .66$).
- St_X – *No strategy:* None of P_M, P_P and P_S dominates in that way, that one of the aforementioned strategies could be assigned.

4 EVALUATION OF ASSIGNED STRATEGIES

4.1 Method

In order to test the described method of assigning strategies, status messages are manually labelled with a content category. Eight volunteers were recruited among colleagues, each of whom labelled 100 randomly selected status messages from the collected data set described in section 3. In order to correspond to the variables P, S, M, according E-commerce-specific content categories were predetermined. The variables are named accordingly P_L for promotional activities, S_L for linking social web platforms and D_L for Twitter dialogs.

For the further processing and analysis, a script was implemented using the R language R Core Team (2012). This script merges the data set created by the taggers with the data set created by the algorithm. Due to missing message ids, 772 out of 800 status messages could be merged, thus $n = 772$.

(a) Promotional

		P yes	P no
P_L	yes	523	35
	no	23	191

(b) Dialog

		D yes	D no
D_L	yes	506	3
	no	66	197

(c) Social Media

		M yes	M no
M_l	yes	468	64
	no	26	214

(d) Fields of contingency tables

		V_1 yes	V_1 no
V_2	yes	a	b
	no	c	d

Table 1: Contingency tables of message categories

For the test, for each category, a null hypothesis H0 and an alternative hypothesis H1 is defined. Since, the correlation between the values assigned by the algorithm and the labelled data set is used for testing significance, the hypotheses are formulated accordingly:

- $\mathrm{H0}_P$: The variables P and P_L do not correlate.
- $\mathrm{H1}_P$: The variables P and P_L correlate.
- $\mathrm{H0}_S$: The variables S and S_L do not correlate.
- $\mathrm{H1}_S$: The variables S and S_L correlate.
- $\mathrm{H0}_D$: The variables D and D_L do not correlate.
- $\mathrm{H1}_D$: The variables D and D_L correlate.

The defined hypotheses are tested on a significance level of 5 % ($\alpha = 0.05$). It has to be noticed, that those hypotheses are used for testing the correlation between attributes identified by the algorithm and human-assigned content categories on level of a single Tweet. The approach of aggregating the communication profiles is not tested by this design.

4.2 Results

In order to test the hypotheses, a measure of correlation between two dichotomous scales is needed. For this purpose χ^2 can be calculated on basis of the effect size ϕ (Bortz and Schuster 2010, p. 174-178). The ϕ coefficient is calculated as follows:

$$\phi = \frac{a \cdot d - b \cdot c}{(a + c) \cdot (b + d) \cdot (a + b) \cdot (c + d)} \quad (1)$$

The field names of the contingency tables used in this equation are shown in Table 1d. The ϕ coefficient expresses the effect size. In order to test an hypothesis on a given significance level, the significance value χ^2 can be calculated:

$$\chi^2 = n \cdot \phi \; (\mathrm{df} = 1) \quad (2)$$

The Tables 1a, 1b and 1c show the contingency tables for each category, which indicate already a strong correlation between the values assigned by the algorithm and the labelled data set. Table 2 summarizes the ϕ coefficient and χ^2 for each category. The ϕ values lie in the right sector of the value range $[0, 1]$. All of those values exceed 0.5 and, thus, indicate a large effect size (see Sheskin 2003, p. 535). For a selected α and $\mathrm{df} = 1$ the corresponding significance value results in $\chi^2_{1;95\%} = 3.84$ (see Bortz and Schuster 2010, p. 174). Since the χ^2 values exceed this value by many times, the null hypotheses $\mathrm{H0}_P$, $\mathrm{H0}_D$, $\mathrm{H0}_S$ are rejected and the alternative hypotheses $\mathrm{H1}_P$, $\mathrm{H1}_D$, $\mathrm{H1}_S$ are accepted. This means, that the results of the labelled data set correlate with the results produced by the algorithm. For the most cases, the algorithm would assign the same category as a human classificator would do. Thus, the performed experiment proves the correctness of the designed algormithm for each of the evaluated categories. The algorithm can be used to aggregate the activities of online retailers to basic communication profiles.

	ϕ	χ^2
P	0.82	519.09
M	0.80	494.08
D	0.80	494.08

Table 2: ϕ and χ^2 values for message categories

5 CONCLUSION

This work contains several contributions, mainly to the information systems related subjects of E-commerce and social media.

First, this paper describes a method of assigning a communication strategy to a given microblogging account in Section 3. This method was developed and used in frame of comparatative study of microblogging activities of retailers in Germany and the UK (see Lackermair and Kailer 2014a). In combination with the implemented data collection process used for the study, a continuous observation could be conducted in order to track the changes in communication pattern over a long time. With those results, different stages of adoption in microblogging in the online commerce and differences in the use of this technology by the consumers can be studied.

Second, this work evaluates the described approach in Section 4. By performing a significance test against a hand labelled control data set, the accuracy of the categories assigned by the algorithm is measured. The results show a very strong correlation for all the variables. This confirms the findings of previous studies, that examined the use Twitter microblogging by online retailers (see Lackermair and Kailer 2014b; Lackermair and Kailer 2014a).

In the next step, the results of this paper are used to implement an observatory that monitors and visualizes the Twitter activities of online retailers continuously. This could help to track changes in the usage of Twitter among consumers in a long term. Besides that, the observatory will help to study the question, whether changes of the communication pattern indicate special events.

REFERENCES

Antoniades, Demetris et al. "We.B: The Web of Short Urls". In: *Proceedings of the 20th International Conference on World Wide Web*. WWW '11. Hyderabad, India: ACM, 2011, pp. 715–724. ISBN: 978-1-4503-0632-4. DOI: 10.1145/1963405.1963505. URL: http://doi.acm.org/10.1145/1963405.1963505.

Bächle, Michael. "Ökonomische Perspektiven des Web 2.0 – Open Innovation, Social Commerce und Enterprise 2.0". German. In: *WIRTSCHAFTSINFORMATIK* 50.2 (2008), pp. 129–132. ISSN: 0937-6429. DOI: 10.1365/s11576-008-0024-2. URL: http://dx.doi.org/10.1365/s11576-008-0024-2.

Bortz, J. and C. Schuster. *Statistik für Human- und Sozialwissenschaftler. Lehrbuch mit Online-Materialien*. Springer-Lehrbuch. Springer, 2010. ISBN: 9783642127700.

Curty, Renata Gonçalves and Ping Zhang. "Social commerce: Looking back and forward". In: *Proceedings of the American Society for Information Science and Technology* 48.1 (2011), pp. 1–10.

Edman, Haley. "Twittering to the Top: A Content Analysis of Corporate Tweets to Measure Organization-Public Relationships". In: (2010). URL: http://etd.lsu.edu/docs/available/etd-04292010-162453/unrestricted/edmanthesis.pdf.

Honeycutt, Courtenay and Susan C. Herring. "Beyond Microblogging: Conversation and Collaboration via Twitter". In: (2009). URL: http://ella.slis.indiana.edu/~herring/honeycutt.herring.2009.pdf.

Hong, Lichan, Gregorio Convertino, and Ed H Chi. "Language Matters In Twitter: A Large Scale Study." In: *ICWSM*. 2011, pp. 518–521.

Huang, Zhao and Morad Benyoucef. "From e-commerce to social commerce: A close look at design features". In: *Electronic Commerce Research and Applications* 12.4 (2013). Social Commerce- Part 2, pp. 246 –259. ISSN: 1567-4223. DOI: http://dx.doi.org/10.1016/j.elerap.2012.12.003. URL: http://www.sciencedirect.com/science/article/pii/S156742231200124X.

Huang, Zhao, Seo Yeon Yoon, and Morad Benyoucef. "Adding Social Features to E-commerce". In: *Proceedings of the Conference on Information Systems Applied Research ISSN*. Vol. 2167. 2012, p. 1508.

Klien, Florian and Markus Strohmaier. "Short Links Under Attack: Geographical Analysis of Spam in a URL Shortener Network". In: *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*. HT '12. Milwaukee, Wisconsin, USA: ACM, 2012, pp. 83–88. ISBN: 978-1-4503-1335-3. DOI: 10.1145/2309996.2310010. URL: http://doi.acm.org/10.1145/2309996.2310010.

Kwak, Haewoon et al. "What is Twitter, a Social Network or a News Media?" In: *Proceedings of the 19th International Conference on World Wide Web*. WWW '10. Raleigh, North Carolina, USA: ACM, 2010, pp. 591–600. ISBN: 978-1-60558-799-8. DOI: 10.1145/1772690.1772751. URL: http://doi.acm.org/10.1145/1772690.1772751.

Lackermair, Georg and Daniel Kailer. "Comparing the Twitter Usage of Online Retailers in Germany and in the UK". In: *ICIW 2014, The Ninth International Conference on Internet and Web Applications and Services*. 2014, pp. 66–71. URL: http://www.thinkmind.org/download.php?articleid=iciw_2014_3_40_20051.

— "Twitter Usage of German Online Retailers". In: *ICCGI 2014, The Ninth International Multi-Conference on Computing in the Global Information Technology*. Vol. 7. URL: http://www.thinkmind.org/download.php?articleid=iccgi_2014_5_20_10116 [retrieved: July, 11th 2014]. 2014, pp. 95–100.

Laudon, Kenneth C., Jane P. Laudon, and Detlef Schoder. *Wirtschaftsinformatik – Eine Einführung*. 2. Auflage. Pearson Studium. Pearson Deutschland, 2010. ISBN: 9783827373489.

LinkedIn. *A Brief History of LinkedIn*. URL: http://ourstory.linkedin.com/ (visited on 07/14/2014).

O'Reilly, Tim. "What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software." In: (2005). URL: http://oreilly.com/web2/archive/what-is-web-20.html.

Paul, Sharoda A, Lichan Hong, and Ed H Chi. "Is Twitter a Good Place for Asking Questions? A Characterization Study." In: *ICWSM*. 2011, pp. 578–581.

R Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2012. URL: http://www.R-project.org/.

Richter, Alexander, Michael Koch, and Jochen Krisch. *Social commerce: eine Analyse des Wandels im E-commerce*. Fak. für Informatik, Univ. der Bundeswehr München, 2007.

Sedghi, Ami. *Facebook: 10 years of social networking in numbers*. [retrieved: 14th July 2014. 2014. URL:

http://www.theguardian.com/news/datablog/2014/feb/04/facebook-in-numbers-statistics.

Sheskin, D.J. *Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition*. Taylor & Francis, 2003. ISBN: 9781420036268.

Twitter. *About Twitter, Inc.* URL: https://about.twitter.com/company (visited on 07/14/2014).

— *About Twitter's link service (http://t.co)*. URL: https://support.twitter.com/entries/109623# (visited on 09/09/2014).

— *Tweet (v1.1)*. URL: https://dev.twitter.com/docs/platform-objects/tweets (visited on 05/29/2013).

Wang, Chingning and Ping Zhang. "The evolution of social commerce: The people, management, technology, and information dimensions". In: *Communications of the Association for Information Systems* 31.5 (2012), pp. 1–23.

Westman, Stina and Luanne Freund. "Information Interaction in 140 Characters or Less: Genres on Twitter". In: *Proceedings of the Third Symposium on Information Interaction in Context*. IIiX '10. New Brunswick, New Jersey, USA: ACM, 2010, pp. 323–328. ISBN: 978-1-4503-0247-0. DOI: 10.1145/1840784.1840833. URL: http://doi.acm.org/10.1145/1840784.1840833.

Wigand, Rolf T, Robert I Benjamin, and Johanna LH Birkland. "Web 2.0 and beyond: implications for electronic commerce". In: *Proceedings of the 10th international conference on Electronic commerce*. ACM. 2008, p. 7.

Yang, Lei et al. "We Know What @You #Tag: Does the Dual Role Affect Hashtag Adoption?" In: *Proceedings of the 21st International Conference on World Wide Web*. WWW '12. Lyon, France: ACM, 2012, pp. 261–270. ISBN: 978-1-4503-1229-5. DOI: 10.1145/2187836.2187872. URL: http://doi.acm.org/10.1145/2187836.2187872.

Zhou, Lina, Ping Zhang, and Hans-Dieter Zimmermann. "Social commerce research: An integrated view". In: *Electronic commerce research and applications* 12.2 (2013), pp. 61–68.