

How Do Consumers Behave in Social Commerce? An Investigation Through Clickstream Data

Qican Gu, Qiqi Jiang^(✉), and Hongwei Wang

School of Economics and Management, Tongji University, Shanghai, China
{1531066_guqican, jiangqq, hwwang}@tongji.edu.cn

Abstract. The social commerce has received considerable attentions in both academia and practitioners in last decade. However, most of current studies investigated such topic from consumers' psychological impetus, but lack of the objective evidence. In this work, we employed the clickstream data analysis to depict online consumers' cross-site browsing behaviors in the context of social commerce. Four prominent clusters depicting distinctive consumers' online browsing behaviors are found. Additionally, the consumers' online behaviors characterized by the browsing patterns are also unveiled and discussed.

Keywords: Social commerce · Clickstream data · Cluster analysis · Chinese consumers

1 Introduction

Social commerce, as an emerging business model that combines social network and e-commerce [1], has received considerable attentions in recent years. Comparing with the conventional e-commerce, social commerce facilitates buying and selling products by using social interaction and user contributions on social media. Various topics related to social commerce have been discussed in previous studies [2, 3]. Such findings make us spontaneously raise the research question: How will consumers behave in social commerce context?

To answer the question above, we collaborated with a leading marketing research agency in Mainland China and attempted to apply the clickstream data to investigate the consumers' online behaviors in social commerce. In particular, the clickstream data included the browsing histories (in the URL form) generated by 2000 randomly selected consumers from December 2014 to January 2015. Besides the individual's online browsing trajectories (the URL records), the demographics of the focal 2000 consumers including gender, age, education and area were also collected. After a series of sophisticated data analysis, several key findings are obtained. First, four groups were highlighted from the cluster analysis after several attempts, which were named as width browsing, depth browsing, goal-oriented browsing, and hedonic browsing. Second, a series of regression models were employed to understand the consequence on purchase commitment in each group. We first unveiled that higher likelihood of being guided into e-commerce websites from social media in the depth browsing, goal-oriented

browsing, and hedonic browsing clusters. Besides, we also unveiled that the goal-oriented group had the highest propensity to the purchase commitment, which was consistent with the findings in prior literatures [4].

For the remainder of this work, an introduction and a succinct literature reviews on social commerce research are presented in Sect. 2. The research proposition and data analysis are given in Sect. 3. We concluded the study in Sect. 4.

2 Social Commerce

The term “social commerce” was first created by Yahoo! on 2005 with the introduction of Shoposphere and Pick Lists, i.e. the two social tools assisting consumers for online shopping [5]. After that, social commerce evolved rapidly from traditional e-commerce with the support of emerging technologies associated with Web 2.0. In line with Liang and Turban [6], we considered social commerce as an environment in which social interaction and user-generated content assist the acquisition of products and services.

The social commerce was composed of two fundamental elements, i.e. social media and commercial activities [6]. From Social perspective, consumers in social commerce can interact with others in shopping activities. For example, they can search product information and share with friends, or aggregate products and make collaborative decision [7]. These social features such as wish lists, chat rooms, tagging, ranking tools and blogs carry unique and interesting capabilities for online shopping [8]. Curty and Zhang [9] analyzed 42 social features found in the top 5 e-commerce websites and categorized them into 4 groups, all of them are found to have the goal of promoting social interactions and exchanges among consumers thus improving their shopping experience. From Commercial perspective, social commerce websites are expected to have ecommerce functions to help consumers accomplishing shopping activities after selecting products. However, for current social commerce websites, few of them provide consumers with tools such as shopping cart, payment zone and confirmation to finish the whole shopping process. Others provide users with product descriptions, price comparison and the link to a third party to accomplish the rest shopping transaction [10].

In the past decade, social commerce has been studied from different angles with diverse methods. Some academic studies were found to investigate the design features and their impacts on perceived usefulness and enjoyment of consumers [8]. Some marketing scholars observed the factors like loyalty [11], social influence [12] and network ties [13] and examined their influence on marketing strategy. For examples, Curty and Zhang [10] studied the framework of social commerce and categorized social commerce websites into two groups: direct sales and referrals. Direct sales refer to websites that contains a full-transaction platform such as Amazon, while referrals provide users with an external site to complete their transactions, which refer to a social commerce process.

Differing from prior studies employing the self-reported data for studying social commerce, we collaborated with a leading marketing agency in China for collecting the clickstream data from 2,000 real users. After a series of sophisticated data analysis, several key findings are obtained. The details of the analysis are given below.

3 Proposition and Data Analysis

3.1 Data Description

With the help with a leading marketing research agency in Mainland China, we randomly collected the clickstream data (browsing histories in the URL form) generated by 2000 consumers, including 1120 males and 880 females, which is consistent with the gender ratio in CNNIC's report of China's netizens [14], from December 2014 to January 2015. For these clickstream data, we firstly removed the extreme observations such as the top one percentage of users who viewed the most pages and we finally got 7,560,000 records.

Afterwards, the cleaned dataset was aggregated into sessions. The session denotes a sequential series of queries submitted by a user when he/she is seeking for certain information during a period of time [15], which was mainly employed to study online consumer behaviors in prior literatures [16]. In this study, we set 30 min as the interval threshold of the visiting times in order to segment the clickstream data (the URL form) into the respective sessions of each user. Finally, 240,000 sessions were obtained. Next, two actions were made to enable these sessions to characterize the social commerce. First, only the sessions containing the browsing histories of both e-commerce websites and social network websites were kept. Second, we removed the sessions in which e-commerce sites were viewed prior to visiting social media websites. The description and descriptive analysis¹ of the variables (with 13412 observations) of each session are given in Table 1.

Table 1. Description of key variables in sessions

Variable	Mean	Min	Max	Description
TOTAL_PAGES	-2.46e-17	-1.012	10.587	Total number of pages viewed in each session
TOTAL_AVG_DURATION	8.82e-15	-1.052	22.062	Average time spent per page
EC_PAGES	1.06e-17	-0.004	0.007	% of pages that were e-commerce pages
EC_AVG_DURATION	4.95e-15	-0.379	45.712	Average time spent per e-commerce page

(Continued)

¹ The standard deviation is not listed due to the same value after standardizing the variables.

Table 1. (Continued)

Variable	Mean	Min	Max	Description
EC_DIFFSITE	1.21e-17	-0.894	9.617	Number of unique e-commerce sites in this session
C_SEARCHPAGE	3.06e-15	-0.461	10.122	% of pages that were e-commerce search pages
EC_PRODUCTPAGE	3.12e-15	-0.723	6.948	% of pages that were e-commerce product pages
EC_CHANNELPAGE	-5.97e-16	-0.343	17.062	% of pages that were e-commerce channel pages
EC_ACTIVITYPAGE	-5.46e-16	-0.247	21.885	% of pages that were e-commerce activity pages
EC_CARTPAGE	-8.28e-16	-0.287	16.464	% of pages that were e-commerce cart pages
SNS_PAGES	1.17e-17	-0.444	29.317	% of pages that were social media pages
SNS_AVG_DURATION	3.03e-15	-0.422	35.170	Average time spent per social media page
SNS_DIFFSITE	-1.08e-16	-.5134445	10.20456	Number of unique social media sites in this session

3.2 Cluster Analysis

After cleaning and preprocessing the raw session data, we applied cluster analysis with K-means algorithm to segment consumers' trajectories in social commerce context. K-means algorithm aims to segment the observations into k clusters in which each observation belongs to the cluster with the nearest mean [17], i.e. the center of the cluster, and the center serves as the average level of the cluster.

Notably, in K-means clustering process, we did not know the exact number of clusters beforehand, so we tried each time with different number of groups with the use of “distance”, denoting the sum of distance between each point and its center, to measure the satisfaction of the result. The “distance” decreased when the number of groups increased because the points were closer to their center. However, less distance means more clusters but too many clusters cannot reflect the real pattern in consumer behavior. In this way, we consider both the number of cluster and their distance and finally got a solution with 5 clusters, the centers of dimensions in each cluster were presented in Table 2.

Table 2. Description of distinguished clusters

Cluster	Width browsing (Cluster 1)	Depth browsing (Cluster 2)	Goal-oriented browsing (Cluster 3)	Hedonic browsing (Cluster 4)	Shallow
N	2025	244	2151	1028	7962
TOTAL_PAGES	216.943	28.450	89.635	70.743	65.722
TOTAL_AVG_DURATION	36.030	234.144	48.850	37.500	47.610
EC_PAGES	0.322	0.322	0.323	0.322	0.322
EC_AVG_DURATION	31.322	394.010	44.710	28.084	38.340
EC_DIFFSITE	4.144	1.401	2.604	3.141	1.914
EC_SEARCHPAGE	0.021	0.016	0.123	0.010	0.011
EC_PRODUCTPAGE	0.062	0.058	0.230	0.055	0.035
EC_CHANNELPAGE	0.005	0.002	0.004	0.064	0.002
EC_ACTIVITYPAGE	0.004	0.003	0.004	0.024	0.002
EC_CARTPAGE	0.003	0.001	0.022	0.003	0.002
SNS_PAGES	24.850	3.662	4.770	3.823	5.477
SNS_AVG_DURATION	48.580	536.287	62.960	42.888	60.264
SNS_DIFFSITE	2.350	1.106	1.183	1.188	1.225

Cluster 1 consists of the sessions containing massive viewed pages with great variety, with a cluster average of 216.943 pages viewed (TOTAL_PAGES) and 4.144 unique e-commerce websites (or 2.350 unique social media websites). The consumers in this cluster are found to spend little time on each page (36.030 TOTAL_AVG_DURATION), which depicted a width-browsing pattern.

In Cluster 2, consumers were found to spend a significant amount of time on viewing each page with high level of TOTAL_AVG_DURATION, EC_AVG_DURATION and SNS_AVG_DURATION, which depicted a deep involvement of users. Furthermore, the number of unique websites consumers viewed (low EC_DIFFSITE and SNS_DIFFSITE value) indicates that users in this cluster visited websites with specific destinations like men’s online stores or baby products websites to obtain target information. Thus, Cluster 2 was named as “Depth Browsing”.

The sessions in Cluster 3 are distinctive in the large amount of searching pages ($EC_SEARCHPAGE = 0.123$), product pages ($EC_PRODUCTPAGE = 0.230$) and cart pages ($EC_CARTPAGE = 0.022$), exhibiting a focused goal-driven behavior. In this cluster, consumers were found to purposively retrieve the information for subsequently purchase. Thus, Cluster 3 was named as “Goal-oriented Browsing”.

Cluster 4 was denoted as hedonic browsing, in which high ratio of activity pages and channel pages with relatively low duration of visiting time were found. An alternative explanation of such patterns depicted in this group is that the consumers were driven by the stimulus like the activities listed on the homepage, the products recommended by system or the advertisements encountered during the visits.

Moreover, the fifth cluster contained sessions that had few pages and less visiting time, and these sessions are named “Shallow” to represent a type of visitors who may visit the site just to see what the site is. This kind of behavior is common in web environment because the Internet inflows users who are exploring different and new features of websites [18].

3.3 Investigation of Post-hoc Behaviors

To investigate the Post-hoc behaviors in each cluster, two additional analyses were made. We first investigate whether the browsing behaviors in the different clusters will lead to different extent of subsequent click-out, denoting clicking a link to e-commerce website after visiting a social media site. Next, we delve into how the probability of purchase commitment in different clustered groups.

For investigating the click-out, a logistic regression model was employed. In particular, a binary dependent variable was set to denote the action of whether the click-out was made ($INTRODUCE = 1$) or not ($INTRODUCE = 0$). In the regression model, the predictor is the categorical variable of each cluster, and the click-out is the dependent variable. In addition, the demographics of the focal 2000 consumers including gender, age, education and area were also collected and included. The descriptive statistics of such control variables are given in Table 3 below.

Table 3. Description of demographic variables

Variable	Observations	Mean	Std. Dev.	Min	Max
INTRODUCE	13412	0.323	0.468	0	1
GENDER	13412	0.407	0.491	0	1
AGE	13412	32.788	8.232	5	73
AREA	13412	2.795	1.724	1	7
EDUCATION	13412	3.536	1.108	0	5

Table 4 showed the results of the logistic regression with the reference group of “Width Browsing”. A significant likelihood ratio test (Log likelihood = -7839.347) was yielded from the model, which implies that our model fits better than an empty model as a whole. According to the coefficient, cluster 2, 3 and 4 performed better than

cluster 1 in clicking-out to e-commerce websites (coefficient > 0) and the results were statistically significant at 0.05 and 0.001 level. Therefore, we can conclude that the online consumers behaving the browsing patters in the groups of depth browsing, goal-oriented browsing and hedonic browsing have higher probability to be redirected to e-commerce website.

Table 4. Results of Click-out behaviors

Independent variables	Coefficient	Z	p-value
Depth browsing (Cluster 2)	0.309	2.200	0.028 [*]
Goal-oriented browsing (Cluster 3)	0.576	4.510	0.000 ^{***}
Hedonic Browsing (Cluster 4)	0.361	8.901	0.000 ^{***}
GENDER	0.309	8.080	0.000 ^{***}
AGE	-0.126	-5.300	0.000 ^{***}
AREA	-0.045	-4.061	0.000 ^{***}
EDUCATION	0.420	2.450	0.014 [*]

N = 13412, Log likelihood = -7839.347, *p < 0.05; ** p < 0.01; ***p < 0.001

Following the result, we also tested the difference among cluster 2, 3 and 4 in clicking-out performance. In this regard, we attempt to unveil how these three clusters differentiate in the purchase commitment. Similarly, a binary variable PURCHASE is set to present whether consumers commit purchase (PURCHASE = 1) or not (PURCHASE = 0), then the logistic regression was adopted and the result was showed in Table 5 below.

Table 5. Results of purchase commitment

Independent variables	Coefficient	Z	p-value
Depth browsing (Cluster 2)	-1.832	-5.310	0.000 ^{***}
Hedonic Browsing (Cluster 4)	-1.231	-8.790	0.000 ^{***}
GENDER	-0.295	-2.860	0.004 ^{**}
AGE	-0.001	0.250	0.840
AREA	0.079	2.670	0.007 ^{**}
EDUCATION	0.070	1.520	0.127

N = 13412, Log likelihood = -7839.347, *p < 0.05; ** p < 0.01;***p < 0.001

According to the results presented in Table 5, we could conclude that consumers behaving goal-oriented browsing patterns (Cluster 3) showed higher purchase commitment than those who behaved in depth browsing (Cluster 2) and hedonic browsing (Cluster 4). Such finding is consistent with the prior literatures in e-commerce research [4, 18].

4 Conclusion

Although the social commerce has been aroused plenty of attentions in the past decade, the multifaceted research on social commerce is still limited. In this work, we employed the clickstream data analysis to depict online consumers' cross-site browsing behaviors in the context of social commerce. In particular, we first applied the cluster analysis with K-means algorithm to segment consumers' behaviors and obtained 4 clusters, i.e. width browsing, depth browsing, goal-oriented browsing, and hedonic browsing. Second, we unveiled how various browsing behaviors depicted in each cluster influence consumers' post-hoc behaviors, i.e. visiting e-commerce site and making the purchase commitment. This study affords several key contributions to both theoretical and practical implications. For researchers, our study provides a better explanation of variation in consumers' behaviors in social commerce context. For practitioners, the segmentation in our findings conduces to a more accurate personalized recommendation, which is expected to bring a higher conversion rate.

Acknowledgement. This work was partially supported by Program for Young Excellent Talents in Tongji University (2014KJ002), the Fundamental Research Funds for the Central Universities (2850219028), and the National Natural Science Foundation of China (NSFC 71532015).

References

1. Zhou, L., Zhang, P., Zimmermann, H.D.: Social commerce research: an integrated view. *Electron. Commer. Res. Appl.* **12**(2), 61–68 (2013)
2. Kim, S., Park, H.: Effects of various characteristics of social commerce (s-commerce) on consumers' trust and trust performance. *Int. J. Inf. Manage.* **33**(2), 318–332 (2013)
3. Liang, T.P., Ho, Y.T., Li, Y.W., et al.: What drives social commerce: the role of social support and relationship quality. *Int. J. Electron. Commer.* **16**(2), 69–90 (2011)
4. Phang, C.W., Kankanhalli, A., Ramakrishnan, K., et al.: Customers' preference of online store visit strategies: an investigation of demographic variables. *Eur. J. Inf. Syst.* **19**(3), 344–358 (2010)
5. Rubel, S.: Trends to watch part II: social commerce. *Micro Persuasion Vom* **2006**, 23 (2005)
6. Liang, T.P., Turban, E.: Introduction to the special issue social commerce: a research framework for social commerce. *Int. J. Electron. Commer.* **16**(2), 5–14 (2011)
7. Shen, J., Eder, L.B.: An examination of factors associated with user acceptance of social shopping websites. In: Mesquita, A. (ed.) *User Perception and Influencing Factors of Technology in Everyday Life* 28–45. Information Science Reference, Hershey (2012)
8. Grange, C., Benbasat, I.: Online social shopping: the functions and symbols of design artifacts. In: 2010 43rd Hawaii International Conference on System Sciences (HICSS), pp. 1–10. IEEE (2010)
9. Curty, R.G., Zhang, P.: Website features that gave rise to social commerce: a historical analysis. *Electron. Commer. Res. Appl.* **12**(4), 260–279 (2013)
10. Curty, R.G., Zhang, P.: Social commerce: looking back and forward. *Proc. Am. Soc. Inf. Sci. Technol.* **48**(1), 1–10 (2011)
11. Huang, Z., Benyoucef, M.: From e-commerce to social commerce: a close look at design features. *Electron. Commer. Res. Appl.* **12**(4), 246–259 (2013)

12. Amblee, N., Bui, T.: Harnessing the influence of social proof in online shopping: the effect of electronic word of mouth on sales of digital microproducts. *Int. J. Electron. Commer.* **16**(2), 91–114 (2011)
13. Stephen, A.T., Toubia, O.: Deriving value from social commerce networks. *J. Mark. Res.* **47**(2), 215–228 (2010)
14. CNNIC. Statistical Report on Internet Development in China (2015)
15. Huntington, P., Nicholas, D., Jamali, H.R.: Website usage metrics: a re-assessment of session data. *Inf. Process. Manage.* **44**(1), 358–372 (2008)
16. Elbaum, S., Karre, S., Rothermel, G.: Improving web application testing with user session data. In: *Proceedings of the 25th International Conference on Software Engineering*, pp. 49–59. IEEE Computer Society (2003)
17. Ghosh, S., Dubey, S.K.: Comparative analysis of k-means and fuzzy c-means algorithms. *Int. J. Adv. Comput. Sci. Appl.* **4**(4), 34–39 (2013)
18. Moe, W.W.: Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. *J. Consum. Psychol.* **13**(1), 29–39 (2003)