Data-Driven Study on Temporal Characteristics of WeChat Users' Information Behavior: Postprint

Authors: Zhang Dayong, Kong Hongxin, Xu Lei, Jingdong

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Compared with traditional information behavior analysis, data-driven information behavior research emphasizes more on the externality and objectivity of data, enabling results that offer a more comprehensive understanding of the essential characteristics of user information behavior. [Method/Process] Through a self-developed APP for collecting records of WeChat users' sharing and reading behaviors, and for systematically analyzing the temporal characteristics of WeChat users' information behavior. [Results/Conclusion] Results indicate that: WeChat users' daily information behavior demonstrates a significant holiday effect; however, the distribution of time intervals between information behaviors exhibits clear heavy-tail phenomena and strong burstiness, indicating that WeChat users' information behavior has high complexity and uncertainty, making effective prediction of its generation process impossible; furthermore, the content shared by WeChat users has strong timeliness, with most content being disseminated promptly within WeChat, but the length of the dissemination chain is significantly influenced by the topic of the shared content.

Full Text

Introduction

Data-driven research on human information behavior statistical characteristics began in 2005 with Professor A.-L. Barabási at Northeastern University [18], who analyzed email data from 3,188 users and found that user behavior deviated from Poisson processes in temporal distribution—a phenomenon that traditional research paradigms could not explain. Subsequent studies on online movie streaming [19], web browsing [20], music downloads [21], and community interactions [22] have found similar non-Poisson characteristics, suggesting that non-uniform and unstable spatiotemporal distribution of user information behavior may be a universal phenomenon. To address this, K.-I. Goh and A.-L. Barabási proposed detection methods for anomalous behavior in temporal and spatial dimensions [23].

This approach, which reveals users' daily information application patterns through human dynamics, differs from traditional qualitative or semi-quantitative analysis. It uses non-intrusive objective data from an external observer perspective, following a process of "observation—data acquisition and analysis—statistical pattern mining—modeling to reproduce data patterns" [24].

In summary, the research scope of information behavior is extensive, covering the entire process from information needs recognition and expression, searching, selection, storage, processing, to user interaction. Research methodologies have gradually shifted from qualitative or semi-quantitative descriptions to data-driven information behavior mining. Data-driven information behavior research focuses on behavioral traces left during natural interaction between users and information system platforms, with data being external, non-interventional, and objective. This provides new research ideas and implementation approaches for comprehensively understanding and optimizing user information behavior, understanding user behavior characteristics and habits, and revealing the complexity of human information behavior.

However, since WeChat does not provide open data APIs, obtaining individual user behavior information is impossible. Existing studies mostly rely on interviews and questionnaires to understand WeChat user behavior characteristics, patterns, and influencing factors [25-28]. These methods have significant limitations: first, they cannot objectively and accurately describe the spatiotemporal distribution characteristics of user behavior; second, they cannot quantitatively analyze user behavior correlations and information sharing trajectories between users. These limitations greatly restrict research on WeChat user information behavior.

Data Acquisition and Processing

To overcome these limitations, we developed an APP called "At HIT" (@HIT) and embedded a reading log tracking program in its articles. When users share articles from within the APP to WeChat Moments, friends, or groups, the reading and resharing logs of these articles are recorded on the server. The tracking program records the sharing user ID, reading user ID, sharing time, reading time, sharing scenario, article ID, article topic, and topic category, thereby enabling the collection of each user's sharing and reading logs.

From August 31, 2015, to January 5, 2018, we collected 247,711 WeChat user sharing records involving 200,023 users and 2,128 articles. To protect user privacy, all collected data were anonymized and used solely for scientific research, not for commercial purposes. Sharer IDs and reader IDs were renumbered after anonymization. Data sources indicate the origin of shared articles, including mobile terminals (APP), web pages (Device), and WeChat (WX). Reading type (to_{type}) specifically refers to articles shared to WeChat and read by WeChat users. Sharing time (sharetime) and reading time (timestamp) are marked in timestamp format. Sharing scenarios refer to WeChat destinations, including friend messages (singlemessage), Moments (timeline), and group chats (groupmessage). Article ID is the unique identifier of shared articles. Article topic (title) refers to the title provided by the sharer. Topic category (subscribe_{name}) refers to the source of shared articles. See Figure 1 [Figure 1: see original paper] for details.

3 Temporal Characteristics Analysis of WeChat User Information Behavior

Human behavior temporal characteristics focus on the statistical patterns exhibited when individuals repeatedly engage in a specific event along the timeline. We analyze WeChat user information behavior temporal characteristics from two aspects: (1) daily temporal distribution of WeChat user information behavior, and (2) inter-event time distribution of continuous information behavior.

3.1 Daily Information Behavior Temporal Distribution

WeChat users' daily information behaviors include sharing (share), reading (stamp), and forwarding (reshare). We analyze the temporal characteristics of these three behaviors. Sharing time refers to when users obtain articles from APP, web pages, or WeChat and forward them to friends, Moments, or groups via mobile or fixed terminals. Reading time refers to when users read articles through friends, Moments, or groups. Forwarding time refers to when users read articles and reshare them.

First, we analyze WeChat users' daily usage habits. Due to the large time span of the data, we extracted the original temporal data by hour and weekday to identify daily patterns. Figure 2 [Figure 2: see original paper] shows the 24-hour distribution of sharing, reading, and forwarding behaviors. From the daily usage distribution, user activity experiences two peaks and one trough: usage begins increasing from 4 AM, reaching the first peak at 9-10 AM, followed by a slight decline at 10-11 AM, a rebound during lunch break at 12 PM, then a decline for three hours. After 3 PM, usage increases gradually, reaching the day's second peak around 10 PM, then declining rapidly to the lowest point around 3 AM.

The usage peaks are no longer limited to rest periods; WeChat usage actually increases during work or study hours, indicating deep integration into users' daily work and learning. Furthermore, user activity significantly increases during holidays, showing a notable weekend effect. As shown in Figure 3 [Figure 3: see original paper], usage on Saturdays and Sundays is significantly higher than on weekdays. During the sample collection period, holiday sharing accounted for 46.7% of total sharing, reading accounted for 60.6% of total reading, and forwarding accounted for 56.3% of total forwarding. This demonstrates that as a social application, WeChat not only meets daily information needs but also occupies users' overall rest time, indicating increased dependence on WeChat for social interaction. However, this also suggests that WeChat excessively occupies users' rest time, and dependence on the network will inevitably reduce real-life interpersonal interaction time and opportunities, potentially leading to decreased interpersonal skills and emotional deficiency.

3.2 Continuous Information Behavior Inter-event Time Distribution

Traditional human behavior temporal characteristic research often used Poisson distribution, which can be viewed as a counting process with negative exponential intervals. Therefore, inter-event times for similar behaviors were approximately uniform, and the probability of dense occurrences over long or short periods was nearly zero. However, recent studies show that many human behaviors, including communication, shopping, and transactions, deviate from Poisson processes [19-22]. The most typical non-Poisson process exhibits burstiness and memory in behavioral time series.

We construct user behavior time sequences based on each user's behavior occurrence times to analyze inter-event time distribution characteristics and the time intervals experienced from initial sharing to reading and forwarding, thereby assessing information timeliness in WeChat.

First, we calculate the inter-event time distribution of WeChat user information behavior to determine whether burstiness and memory exist. Burstiness refers to the non-uniform frequency of events—sometimes many events concentrate in short periods, while intervals can be very long. Memory refers to correlations between consecutive time intervals: if long intervals follow long intervals or short intervals follow short intervals, the sequence exhibits memory.

For burstiness and memory in human activities, we use the anomaly detection metrics proposed by K.-I. Goh and A.-L. Barabási [23]. The burstiness detection coefficient B = $(\sigma\tau$ - $m\tau)/(\sigma\tau$ + $m\tau)$ focuses on statistics of standard deviation $\sigma\tau$ *and mean* $m\tau$ of inter-event time sequences, where B (-1, 1). If B > 0, the time series has burstiness effects, and B closer to 1 indicates more uneven distribution and stronger burstiness. B = 0 indicates a Poisson distribution where mean equals standard deviation, representing a neutral sequence.

As shown in Figure 4(a) [Figure 4: see original paper], overall, WeChat user information behavior inter-event time distribution shows significant heavy-tail deviation from exponential distribution. The distribution curve P(t) = 1.004x^{-1.532}, R^2 = 0.964 follows a power-law distribution, indicating extremely uneven temporal distribution. Most behavior intervals are short (within 2 minutes), accounting for 83.78% of total behaviors, while a small portion have long intervals—those exceeding 1 hour account for 0.58% of total behaviors, with the maximum interval reaching 5,826 minutes.

From an individual user perspective, the frequency of user information behavior is also extremely uneven. As shown in Figure 4(b), only 10 users had more than 50 behavior occurrences during the sample period, while 84.31% of users had only one occurrence, with the maximum being 573 times. Both the inter-event time distribution and individual user behavior frequency exhibit heavy-tail patterns distinct from Poisson distribution's bell curve, indicating that high-frequency information activities occur within small scopes (limited participants and limited time intervals). These high-frequency activities play crucial roles in promoting information dissemination and improving propagation efficiency, warranting attention in daily WeChat management.

Figure 5 [Figure 5: see original paper] shows user information behavior occurrence density at different time granularities from the first minute to 100 hours of the sample collection period. Using hour-level granularity as an example: one behavior occurred in the first hour, followed by nearly 24 hours of silence, then two consecutive behaviors at 25-26 hours, then 7 hours of silence, followed by three consecutive behaviors at 34-36 hours, then a long gap before entering a dense period. Overall, most inter-event times differ significantly from their mean. In minutes, the standard deviation $\sigma\tau$ *is 145.39 while the mean* $m\tau$ is only 5.32, giving a burstiness coefficient B = 0.929, close to 1, indicating extremely uneven sequences with strong burstiness.

By sharing scenario (see Table 1), at smaller time granularities, burstiness coefficients in Moments (timeline) and group chats (groupmessage) are higher than in friend messages (singlemessage). As time granularity increases, burstiness coefficients decrease, particularly significantly in Moments.

Memory concerns correlations between consecutive time interval sequences. K.-I. Goh and A.-L. Barabási proposed the memory correlation coefficient M = (τi *- m_1)(*τ{i+1} - m_2) in [23], where n_τ is the number of intervals in user τ's behavior time sequence, m_1 and m_2 are means of two sampled sequences τi *and* τ{i+1}, and σ_1 and σ_2 are corresponding standard deviations. Positive M indicates memory effects; negative M indicates anti-memory effects.

Calculating memory correlation requires grouping user sharing behavior sequences, with each group of n_τ elements requiring n_τ + 1 sharing behaviors. Too few samples per group compromise behavioral continuity. Therefore, we selected the top 50 high-frequency users as samples, with maximum behavior occurrences of 573 and minimum of 19. Calculations show sample users' memory correlation coefficient M ranges from -0.13 to 0.14, with average M = -0.03, approximately zero, indicating no significant memory effects for most individuals during the statistical period.

As shown in Figure 6 [Figure 6: see original paper], each point represents a qualifying sample individual. Except for one user with over 500 occurrences, all sample users had fewer than 100 occurrences. Memory coefficients are symmetrically distributed around zero, while burstiness coefficients are all positive with mean 0.4. This indicates that high-frequency WeChat users' information behavior exhibits significant burstiness but no memory effects in inter-event time distribution.

Information dissemination in WeChat relies on interpersonal relationship chains driving information propagation chains. Information only spreads effectively through layers of user sharing and forwarding. Analyzing time intervals between sharing-reading and reading-forwarding helps understand information timeliness in WeChat. Generally, shorter intervals between information sending from the source and reading indicate more timely adoption and stronger timeliness. Longer propagation chains reflect more forwarding layers, strengthening effective reach rates and user participation depth.

We define timeliness as C'*k = (1/w_k) Σ a^*k{ij}, where a^k_{ij} represents the time interval between user i and user j's reading and sharing for article k, and w_k represents article k's reading count. Average timeliness is C' = (1/N) Σ C'*k. Propagation chain length is defined as D_k = max(d^*k{ij}), where d^k_{ij} represents the maximum forwarding depth between any two users i and j in article k's propagation chain.

Figure 8 [Figure 8: see original paper] shows the propagation chain length for individual articles in WeChat.

Conclusion

This study follows the process of "observation—data acquisition and analysis—pattern mining and reproduction." First, we collected WeChat users' sharing and reading behavior records through a self-built mobile APP platform, including temporal information, sharing relationships, scenarios, and topics. The revelation of WeChat user information behavior temporal characteristics focused on two aspects: (1) Analysis of daily information behavior temporal distribution characteristics shows that WeChat usage peaks are no longer limited to rest periods but have deeply integrated into users' daily work and life. User activity significantly increases during holidays, showing a notable holiday effect. (2) Analysis of inter-event time distribution characteristics shows obvious heavy-tail phenomena and strong burstiness in WeChat user continuous information behavior intervals. Smaller time granularities yield stronger burstiness, particularly significant in WeChat Moments. However, correlation analysis of high-frequency users' continuous information behavior shows no significant memory effects during the statistical period.

The strong burstiness and absence of memory in WeChat user information behavior indicate high complexity and uncertainty, making effective prediction of generation processes difficult. Additionally, timeliness analysis shows that most information spreads timely in WeChat with high efficiency in information transmission and utilization. However, influenced by content themes, most content fails to gain sustained user attention or generate forwarding value.

References

[1] OHO, EOMC, RA OHR. Role of social media in social change: analysis of collective sensemaking during the 2011 Egypt revolution [J]. Information systems research, 2015, 26(1): 210-223. [2] Li D, Xu ZM, Li S, et al. Information diffusion in online social networks [J]. Chinese Journal of Computers, 2014, 37(1): 189-206. [3] LIU L, QU B, CHEN B, et al. Modelling of information diffusion on social networks with applications to WeChat [J]. Physica A, 2018, 496: 318-329. [4] ZHANG YX, LI ZQ, GAO CL, et al. Mobile social big data: WeChat moments dataset, network applications, and opportunities [J]. IEEE network, 2018, 32(3): 146-153. [5] XU X, YAO Z, SUN Q. Social media environment effect on perceived interactivity: an empirical investigation from WeChat moments [J]. Online information review, 2019, 43(2): 239-255. [6] GAN C. Understanding WeChat users' liking behavior: an empirical study in China [J]. Computers in human behavior, 2017, 68: 30-39. [7] Li J, Ren JL, Liu X, et al. Research on users' continuance intention of WeChat public platform [J]. Information Science, 2016(10): 26-33, 55. [8] Li L, Ding N. Research progress on user information behavior of mobile social applications at home and abroad [J]. Library and Information Service, 2015, 59(10): 137-144. [9] BELKIN N. The cognitive viewpoint in information science [J]. Journal of information science, 1990, 16(1): 11-15. [10] Wang Y, Deng XZ. Discussion on basic issues of network user information behavior [J]. Library and Information Service, 2009, 53(16): 35-39. [11] Ye PH, Zhang LY. Comparative analysis of information behavior research at home and abroad from 2002 to 2011 [J]. Information Science, 2013(4): 149-154. [12] KOSINSKI M, BACHRACH Y, KOHLL P, et al. Manifestations of user personality in website choice and behaviour on online social networks [J]. Machine learn, 2014, 95(3): 357-380. [13] PHAM T, SHERIDAN P, SHIMODAIRA H. Joint estimation of preferential attachment and node fitness in growing complex networks [J]. Scientific reports, 2016, 6: 1-13. [14] ZHANG Z, LIU L, WANG H, et al. Collective behavior learning by differentiating personal preference from peer influence [J]. Knowledge-based systems, 2018, 159: 233-243. [15] GARG S, KAUR K, KUMAR N, et al. Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: a social multimedia perspective [J]. IEEE transactions on multimedia, 2019, 21(3): 566-578. [16] SEOL S, LEE H, YU J, et al. Continuance usage of corporate SNS pages: a communicative ecology perspective [J]. Information & management, 2016, 53(6): 740-751. [17] Li YL, Zhang XT. Review and prospect of data-driven information behavior research [J]. Journal of Information Resources Management, 2018(2): 13-27. [18] BARABÁSI A-L. The origin of bursts and heavy tails in human dynamics [J]. Nature, 2005, 435(7039): 207-211. [19] DEZSO Z, ALMAAS E, LUKACS A, et al. Dynamics of information access on the web [J]. Physical review E, 2006, 73(6): 066132. [20] ZHOU T, KIET HAT, KIM BJ, et al. Role of activity in human dynamics [J]. Europhysics letters, 2008, 82(2): 28002. [21] HU HB, HAN DY. Empirical analysis of individual popularity and activity on an online music service system [J]. Physica A, 2008, 387(23): 5916-5921. [22] GRABOWSKI A, KRUZEWSKA N, KOSINSKI RA. Dynamic phenomena and human activity in

an artificial society [J]. Physical Review E, 2008, 78(6): 066110. [23] GOH KI, BARABÁSI A-L. Burstiness and memory in complex systems [J]. Europhysics letters, 2008, 81(4): 48002. [24] Zhou T, Han XP, Yan XY, et al. Statistical mechanics of spatial-temporal characteristics of human behavior [J]. Journal of University of Electronic Science and Technology of China, 2013, 42(4): 481-539. [25] Xue Y, Xu ZL. Analysis and model construction of influencing factors of user information behavior in WeChat marketing environment: Based on the perspective of flow theory [J]. Information Studies: Theory & Application, 2016, (6): 104-109. [26] CHEN L, CHIN FG, SUN Y, et al. Integrating guanxi into technology acceptance: an empirical investigation of WeChat [J]. Telematics and informatics, 2017, 34(7): 1125-1142. [27] Lv YL, Huang C, Zhou H. Research on health information behavior of WeChat platform users and its influencing factors [J]. Journal of Medical Informatics, 2018, 39(3): 77-80. [28] WANG G, ZHANG W, ZENG RX. WeChat use intensity and social support: the moderating effect of motivators for WeChat use [J]. Computers in human behavior, 2019, 91: 244-251.

Author Contributions

Zhang Dayong: Proposed research ideas, designed research plan, and wrote the paper; Kong Hongxin: Participated in data collection and processing; Xu Lei: Participated in partial paper writing; Jing Dong: Participated in paper revision.

Abstract: [**Purpose/Significance**] Compared with traditional information behavior approaches, data-driven information behavior research pays more attention to the externality and objectivity of data, enabling more comprehensive understanding of user information behavior characteristics. [**Method/Process**] This paper realizes the collection of WeChat users' sharing and reading behavior records through a self-built APP and systematically analyzes the temporal characteristics of WeChat users' information behavior. [**Result/Conclusion**] Results show that WeChat users' daily information behavior has significant holiday effects. The time interval distribution shows obvious fat-tail phenomenon and strong burstiness effects, indicating high complexity and uncertainty. Most content can be timely disseminated in WeChat, but the length of the dissemination chain is significantly affected by the theme of shared content. This study provides a reference for revealing the complexity of human information behavior.

Keywords: WeChat; information behavior; complexity; temporal characteristics

Note: Figure translations are in progress. See original paper for figures.