I^3 : Sensing Scrolling Human-Computer Interactions for Intelligent Interest Inference on Smartphones

LI LU, Shanghai Jiao Tong University, China

JIADI YU*, Shanghai Jiao Tong University, China

YINGYING CHEN, Rutgers University, USA

YANMIN ZHU, Shanghai Jiao Tong University, China

MINGLU LI, Shanghai Jiao Tong University, China

XIANGYU XU, Shanghai Jiao Tong University, China

The scrolling interaction is a pervasive human-computer interaction on smartphones, which can reflect intrinsic characteristics during dynamic browsings. Different from extrinsic statistical measures like frequency of visits and dwell time, intrinsic features underlying scrolling interactions reveal fine-grained implicit feedbacks about user interests. Toward this end, we explore user interest inference by extracting efficient browsing features from scrolling human-computer interactions on smartphones. In this paper, we first analyze browsing traces of 40 volunteers, and find two intrinsic browsing features underlying scrolling interactions, i.e., *browsing velocity stability* and *browsing velocity sequence*, which are tightly related to user interests. Inspired by the observation, we propose an *Intelligent Interest Inference* system, I^3 , which infers user interests through sensing scrolling interactions during browsings. Specifically, I^3 first extracts the two intrinsic browsing features from users' scrolling interactions. Then, I^3 applies a Naive Bayesian-based approach to construct an interest discriminator for coarse-grained user interest (i.e., like-preferred or dislike-preferred) inference. Furthermore, we develop a deep learning-based approach for I^3 to train rating classifiers for fine-grained rating inference in like-preferred and dislike-preferred browsings respectively. Finally, I^3 utilizes the interest discriminator and rating classifiers to infer exact user ratings about browsing contents on smartphones. Experimental results under browsing traces of 46 volunteers present that I^3 achieves 92.4% overall accuracy in interest inference.

CCS Concepts: • **Information systems** → **Personalization**; • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: Interest inference, scrolling interactions, smartphone, recommendation

*Jiadi Yu is the corresponding author, Email: jiadiyu@sjtu.edu.cn

Authors' addresses: Li Lu, Shanghai Jiao Tong University, Department of Computer Science and Engineering, 800 Dongchuan Rd, Shanghai, China, 200240, luli_jtu@sjtu.edu.cn; Jiadi Yu, Shanghai Jiao Tong University, Department of Computer Science and Engineering, 800 Dongchuan Rd, Shanghai, China, 200240, jiadiyu@sjtu.edu.cn; Yingying Chen, Rutgers University, WINLAB and Department of Electrical and Computer Engineering, New Brunswick, NJ, USA, 08854, yingche@scarletmail.rutgers.edu; Yanmin Zhu, Shanghai Jiao Tong University, Department of Computer Science and Engineering, 800 Dongchuan Rd, Shanghai, China, 200240, yzhu@sjtu.edu.cn; Minglu Li, Shanghai Jiao Tong University, Department of Computer Science and Engineering, 800 Dongchuan Rd, Shanghai, China, 200240, mlli@sjtu.edu.cn; Xiangyu Xu, Shanghai Jiao Tong University, Department of Computer Science and Engineering, 800 Dongchuan Rd, Shanghai, China, 200240, chillex@sjtu.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Association for Computing Machinery.

2474-9567/2019/9-ART97 \$15.00

<https://doi.org/10.1145/3351255>

ACM Reference Format:

Li Lu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Minglu Li, and Xiangyu Xu. 2019. I^3 : Sensing Scrolling Human-Computer Interactions for Intelligent Interest Inference on Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 97 (September 2019), 22 pages. <https://doi.org/10.1145/3351255>

1 INTRODUCTION

As the popularity of personalized services[46], user interest inference becomes essential for many applications. User interest inference helps personalized services to provide specific information and services that users are interested in. For example, there are over 70,000 items about the ‘camera’ product in Amazon[1], which are almost impossible for a user to browse all webpages of these items. To help users find the target product from enormous candidates with satisfactory experience, Amazon infers the user’s interest about camera products by exploring existing browsing records of the user, and provides personalized recommendation about the specific camera product for the user. Similar situations are widely existed in various field, such as news browsing, point-of-interest (POI) finding, etc. Therefore, it is necessary to infer user interest for the various applications in our daily life.

Typically, the best approach of extracting user interests is to request explicit ratings for browsing contents from users. Unfortunately, few users are willing to explicitly provide the ratings after browsings. For example, a webpage in Quora[39] is viewed by 2,328 users, but only 9 users give ratings to the webpage. Thus, it is necessary to infer user interests about contents through implicit feedbacks during users’ browsings. Some existing works focus on user interest inference based on human-computer interactions during users’ browsings, i.e., implicit feedbacks, such as the frequency of visits and dwell time. Núñez-Valdéz et al.[37] apply the frequency of webpage visits to infer a user’s interest, where a low frequency of a user’s visits to a webpage depicts a low interest of the user, and vice versus. On the other hand, Yin et al.[51] utilize users’ browsing features, i.e., the dwell time, which is intrinsically the average browsing velocity, to infer user interests. However, both the frequency of visits and dwell time are extrinsic statistical measures, which are susceptible to feature outliers during users’ browsings, and further unsatisfactory for user interest inference.

Recently, as the development of Internet and popularity of smartphones, more people employ smartphones for browsing and shopping online instead of laptops or desktops. Since the screen sizes of smartphones are usually limited, the content of a webpage needs to be presented on multiple screens. Users usually browse the whole content of a webpage through scrolling interactions on smartphones. Hence, these human-computer interactions can reflect users’ browsing features. This triggers our research in this work to extract intrinsic browsing features from scrolling human-computer interactions on smartphones for inferring user interests during browsings. To realize the interest inference leveraging scrolling interactions on smartphones, we face several challenges. First, we need to extract efficient browsing features from the scrolling interactions during a browsing. Second, we should exploit the relationship between intrinsic browsing features and user interests during browsings for the interest inference. Finally, the solution should be robust to different impact factors, such as the length of browsing contents and screen size of smartphones.

In this paper, we first study the relationship between scrolling interactions and user interests based on browsing traces of 40 volunteers, and find that the average browsing velocity is not satisfactory for user interest inference. This is because users’ browsings are usually dynamic, which cannot be revealed by a statistical measure, i.e., the average browsing velocity. Through further analysis on the same traces, we explore two browsing features underlying dynamic scrolling velocities, i.e., *browsing velocity stability* and *browsing velocity sequence*, which are tightly related to user interests. Based on the analysis above, we propose an *Intelligent Interest Inference* system, I^3 , which infers user interests about contents through sensing scrolling human-computer interactions on smartphones. I^3 first senses scrolling interactions during users’ browsing in real time, and extracts two browsing features, i.e., browsing velocity stability and browsing velocity sequence, from scrolling interactions. Based on extracted features, I^3 builds a silent rating inference model, including an interest discriminator and rating

classifiers, to model the relationship between scrolling interactions and user interests. Specifically, based on the browsing velocity stability, I^3 first constructs an interest discriminator through a Naive Bayesian-based approach to infer coarse-grained user interest (i.e., like-preferred, neutral-preferred or dislike-preferred) about browsing contents. Furthermore, to infer the fine-grained user rating about browsing contents under the coarse-grained user interest, I^3 utilizes a deep learning-based approach, i.e., Convolutional Neural Networks (CNN), to train two rating classifiers based on the browsing velocity sequence for rating inference in like-preferred and dislike-preferred browsings respectively. Finally, through the interest discriminator and rating classifiers, I^3 can infer exact user ratings about browsing contents.

We highlight our contributions as follows.

- We propose an intelligent interest inference system, I^3 , which infers user interests about contents through sensing scrolling interactions during users' browsing.
- We explore two intrinsic browsing features underlying scrolling interactions, i.e., the browsing velocity stability and browsing velocity sequence, which are tightly related to user interests during browsings.
- We construct an interest discriminator and two rating classifiers through Naive Bayesian-based and deep learning-based approach respectively, to model the relationship between scrolling interactions and user interests.
- We conduct experiments to evaluate the performance of I^3 under browsing traces of 46 volunteers in real environments, and results show that I^3 can achieve an overall accuracy of 91.6% in user interest inference.

The rest of this paper is organized as follows. We first review related works in Section 2. Section 3 presents the analysis of relationship between user interests and scrolling human-computer interactions. The system design of I^3 is described in Section 4. Next, we present the evaluation results in Section 5. Finally, we discuss some practical issues and limitations, as well as make a conclusion in Section 6 and 7.

2 RELATED WORK

In this section, we review related works which focus on user interest inference.

Since few users provide their interests after browsing content in real application scenarios, many researchers infer user interests not only based on explicit feedbacks, but also leveraging implicit feedbacks, for satisfactory content recommendation[25]. The frequency of visits is one of the most widely adopted implicit feedbacks. Existing works utilize the frequency of visits on different types of tasks, e.g., visits of product information[37], visits of webpages[27], and watch times of TV programs[53], to improve the accuracy of recommendations and further enhance the user experience of personalized services. Other works concentrate on adopting the dwell time (i.e., time-spent-reading) as implicit feedbacks to infer user interests. The dwell time is intrinsically the average browsing velocity. Morita et al.[36] conduct the first research about the usefulness of dwell time as implicit feedback of user interests. Inspired by the research, Yin et al.[51] model the relationship between the dwell time and user interests from a psychology perspective, and Kim et al.[19] study the dwell time of each click during searching processes to infer users' searching result preference. However, both the frequency of visits and dwell time are extrinsic statistical measures, which are easily affected by feature outliers, and thus cannot reveal dynamic properties during users' browsings.

To reveal dynamic properties underlying browsings, recent works exploit intrinsic measures from human-computer interactions during browsings of users. On the one hand, some works explore cognitive and psychology theory to investigate the browsing process. Early work[5] studies different kinds of browsings on hypertext webpages from the perspective of cognitive science. Following works[14, 29] investigate the psychological and cognitive factors for the information retrieval during a browsing process. Even eye-tracking techniques are employed to study the relationship between user interest and human-computer interactions[42, 44]. However, these works focus on investigating the browsing process without studying on intrinsic measure for interest

inference. On the other hand, some works conduct experimental studies to reveal the relationship between user interest and intrinsic measures underlying human-computer interactions, such as social network-related interaction[6, 32, 55], speaking voice[50] and eye focus[7, 48], etc. However, all of them involve sensitive and private information of users, which arises users' privacy concerns. Moreover, both voice sensing and eye tracking require to enable specific hardware on mobile devices (e.g., microphones, cameras), which introduce significant power consumption of mobile devices. Even worse, the integrated hardware of commercial smartphones cannot achieve satisfactory performance, due to the limited resolution.

To release the concern of the privacy and power consumption, other works[8, 9, 18] explore the scrolling, as well as other mouse and keyboard interactions as implicit feedbacks for desktop browsings. But the results present that only utilizing these interactions is unsatisfactory for interest inference of desktop browsings. This is because the screen sizes of desktops are usually large enough to present the whole content of a page, which does not require many interactions during browsings. Nowadays, mobile and smart devices are increasingly involved in our life, such as sensing driving conditions[49], adapting frame rate[52], authenticating user identity[30], and even issuing a side-channel attack[31]. Different from desktop browsings, on these mobile and smart devices, users usually utilize scrolling human-computer interactions during browsings, due to the limited size of these devices' screen. Hence, scrolling interactions during mobile device browsings can provide more intrinsic information about user interests. Early work[17] reveals that scrolling interactions reflect a user's browsing features, and can be used to provide automatic zooming so that the user can more efficiently browse content. Recent studies[24, 26] further demonstrate that scrolling interactions are related to user browsing features, and can be utilized to improve information retrieval performance. However, these works only reveal the relationship between scrolling interactions and browsing behaviors, while not exploit that between scrolling interactions and user interests.

Our work is to exploit intrinsic characteristics from scrolling human-computer interactions during dynamic browsings on smartphones for inferring user interests about contents, without introducing significant power consumption and the requirement of additional infrastructures.

3 MOTIVATION STUDY

In this section, through the collected data from 40 volunteers in real environments, we first study the relationship between user interests and average scrolling velocity, and find the reasons of interest inference performance degradation through the average scrolling velocity. Then, from results of extensive studies, we discover intrinsic browsing features underlying scrolling human-computer interactions, which are tightly related to user interests.

3.1 Data Collection

In order to collect data, we recruit 40 volunteers as browsing users, whose ages are from 19 to 53. Among them, 23 volunteers are male, while the rest 17 are female. In order to ensure an indicative result, these volunteers are involved in different occupations. Specifically, the 40 volunteers consist of 6 faculties and 8 staffs of a university, 7 software engineers, 4 industry researchers, and 15 students with different majors (e.g., computer science, math, history, etc.). The volunteers are recruited through the online recruitment method, and most of them share similar university education background. After the experiments, we provide a small souvenir of the university for each volunteer as the compensation. Since news is one of the most prevalent browsing contents for different people (across different genders, ages, occupations, etc.), we randomly fetch 1,000 news webpages from USA Today[43], Voice of America[38], BBC News[2] and CNN[10] as browsing contents to perform a generalized study. The traced news webpages include different kinds of webpages, e.g., text-only webpages, text-figure hybrid webpages, etc., and involve various topics, such as technique, science, business, health, etc. The 40 volunteers are required to randomly browse the webpages, and each volunteer only reads the same webpage once. We concentrate on inferring user interest of the whole webpage with the scrolling interactions during the browsing. Note that the

Fig. 1. CDFs of average browsing velocities of browsings under five rating levels.

volunteers can browse the webpages without any constraints. Specifically, we do not request volunteers to browse webpages in a fixed time or browse a fixed number of webpages, and volunteers can skip to the next webpage at any point. After a browsing finishes, the volunteer is required to give an explicit rating from five levels (i.e., 1~5-star, more stars depict a user prefers to like the content) about the webpage as ground truth. The explicit ratings given by volunteers measure the user interest about the whole webpage during the browsing.

Meanwhile, we develop an application to trace scrolling interactions of users during browsings. When a user browses a webpage on a smartphone, his/her finger scrolls on the screen, which leads to the scrolling of webpage. After the finger leaves the screen, the webpage continues to scroll due to the inertia. Thus, the scrolling of webpage contents can reflect users' browsing features, while the scrolling of user fingers cannot. After a user scrolls to the end of the webpage, the rating window pops up for the user to provide the explicit rating.

In each browsing, our application traces the scrolling of a webpage every Δt . Specifically, the pixels of a webpage can be extracted into a pixel sequence line by line. Hence, at time t , all pixels of the webpage displayed on screen can be expressed as a pixel sequence, and a certain pixel line of the webpage can be expressed as a subsequence. By matching the subsequence in the pixel sequence of the whole screen, we can obtain the position of the subsequence in the screen pixel sequence at time t , i.e., $d(t)$. After the trace interval Δt , the subsequence has forwarded (or backwarded) for a distance, its new position in the screen pixel sequence $d(t + \Delta t)$ can be obtained at time $t + \Delta t$. Then, we can derive the scrolling velocity $v(t)$ from time t to $t + \Delta t$ as follows,

$$v(t) = \frac{d(t + \Delta t) - d(t)}{\Delta t}. \quad (1)$$

We trace the scrolling interactions with $\Delta t = 100ms$.

3.2 Relationship between the Average Browsing Velocity and User Interests

Dwell time[51] is an implicit feedback in interest inference, which is intrinsically the average velocity during a browsing. A low average browsing velocity depicts that there is a high probability that a user likes the webpage, and vice versus. Thus, we first analyze the relationship between the average browsing velocity and user interests on smartphones.

Since users usually adopt scrolling interactions to browse webpages on smartphones, the scrolling velocity can be used to express the browsing velocity. For a browsing, the average browsing velocity is $\bar{v} = \frac{1}{T} \sum_{t=0}^{T-1} v(t)$, where T is the number of sample points during the browsing. Fig. 1 shows CDFs of average velocity of browsings under five rating levels respectively. It can be observed that CDFs of average browsing velocity under the five kinds of browsings are quite similar to each other. The result indicates that it is difficult to distinguish the five kinds of browsings from each other through the average browsing velocity. This is because the scrolling velocity is usually dynamic during a browsing, e.g., a user can first browse a content quickly, then slowly, and finally in a stable velocity. Such a dynamic characteristic is actually an intrinsic feature of browsing. On the contrary,

Fig. 2. Scrolling velocity sequences and distributions under sample browsings of five rating levels.

the average browsing velocity is an extrinsic feature, which cannot reveal the intrinsic feature of a browsing. Moreover, the extrinsic feature is a statistical measure, which is easily affected by the outliers in the raw browsing data[15]. For example, during the browsings, a user may suddenly increase/decrease the browsing velocity to an extremely high/low value (i.e., an outlier), which significantly affects the extrinsic measures. Therefore, the average browsing velocity cannot reflect users' browsing features.

It is necessary to extract efficient browsing features from the dynamic scrolling velocities. To exploit the relationship between dynamic scrolling velocity and user interests, we analyze the scrolling velocity sequences of different browsings. Fig. 2(a) shows scrolling velocity sequences of five sample browsings under different rating levels respectively. We can observe that, although the average scrolling velocities are similar (i.e., all around 650 *pixels/ms*) for the five browsings, the ranges of scrolling velocities during the two browsings are quite different, i.e., most scrolling velocities of the 1-star and 2-star (i.e., like-preferred) browsings are in the range of [0, 4000] *pixels/ms*, while that of the 5-star and 4-star (i.e., dislike-preferred) browsings are in the range of [100, 1500] *pixels/ms*. This example indicates that the browsing stabilities are significantly different between like-preferred and dislike-preferred browsings, but the average browsing velocity cannot reflect such a browsing feature.

3.3 Relationship between the Browsing Velocity Stability and User Interests

In this section, we study the relationship between the browsing velocity stability and user interests based on browsing traces of the 40 volunteers.

To represent the stability of a browsing, we formulate the browsing velocity stability with the absolute frequency of scrolling velocity distribution's mode. The mode of scrolling velocity distribution is a scrolling velocity which appears the most frequently during a browsing. The larger the value of mode's absolute frequency is, the larger the percentage of modal scrolling velocity during a browsing is, which depicts a more stable browsing. Therefore, we utilize the absolute frequency of mode as the *stability degree* to model the browsing velocity stability. As shown in Fig. 2(b), the browsing velocity stability of like-preferred (i.e., 5-star and 4-star) browsings is significantly distinguishable with that of dislike-preferred (i.e., 1-star and 2-star) browsings. We further derive the stability degrees of the five sample browsings. Fig. 2(b) shows scrolling velocity distributions of the five sample browsings

Fig. 3. CDFs of stability degrees of browsings under five rating levels.

respectively. We can find that the stability degrees are both above 10 for 5-star and 4-star browsings, while that are both less than 5 for 1-star and 2-star browsings, which is consistent with Fig. 2(a).

We then analyze the relationship between the stability degree and user interests. Fig. 3 shows CDFs of stability degrees of browsings under five rating levels respectively. It can be seen that 5-star and 4-star browsings are distinguishable with 3-star browsings as well as 1-star and 2-star browsings, which indicates that the browsing velocity stability is feasible to distinguish like-preferred, neutral-preferred and dislike-preferred browsings. Except for the maximum absolute frequency of the mode, many other measures (e.g., variance and entropy) may be the alternatives to represent the browsing velocity stability. However, through empirical studies, these alternative measures can only achieve a low interest inference accuracy (i.e., $\sim 55\%$). Instead, the maximum absolute frequency in the scrolling velocity distribution surprisingly exhibits significant differences for browsings with different coarse-grained user interests (i.e., like and dislike). And such a feature exhibits the browsing velocity stability. Hence, the maximum absolute frequency in the scrolling velocity distribution is finally selected as the stability measure.

However, it can be also observed from Fig. 3 that, in like-preferred browsings, stability degree of 5-star browsings is quite similar to that of 4-star browsings. Similarly, in dislike-preferred browsings, stability degree of 1-star browsings is quite similar to that of 2-star browsings. This is because the stability degree only considers the global feature during a browsing, while local features are not considered. We further investigate the local features during browsings to achieve the fine-grained interest inference. From like-preferred browsings in Fig. 2(a), it can be seen that scrolling velocities in the two browsings are not always in a particular range. Specifically, scrolling velocities in the 4-star browsing present significant range variations over time, while that in the 5-star browsing show little. The two dislike-preferred browsings in Fig. 2(a) depict similar results. Therefore, the sequential relationship between scrolling velocities can represent local features during browsings, which may be an effective feature to further identify the fine-grained rating after extracting the coarse-grained user interest (i.e., like-preferred, neutral-preferred or dislike-preferred) about browsing content.

3.4 Relationship between the Browsing Velocity Sequence and User Interests

In this section, we study the relationship between the browsing velocity sequence and user interests based on browsing traces of the 40 volunteers.

To exploit the sequential relationship between scrolling velocities, we formulate the browsing velocity sequence through Markov chain[11]. A Markov chain is formulated as $x = \langle S, x^{(0)}, Q \rangle$, where S is a state space, $x^{(0)}$ is an initial probability vector, and Q is a state transition matrix. Assuming there are M different scrolling velocities in a browsing, the state space is $S = \{v_0, v_1, \dots, v_{M-1}\}$. Usually, the initial scrolling velocity $v_0 = 0$ in a browsing. Hence, the initial probability vector is $x^{(0)} = (1, 0, \dots, 0)^T$. The number of unique scrolling velocities in data collection is limited. This is because the application traces the scrolling velocities during a browsing in a discrete

Fig. 4. Dimension-reduced features of state transition matrices under browsings of four rating levels.

manner. Hence, the unique scrolling velocities can be regarded as the state space of Markov chain. Each entry of the state transition matrix Q represents the frequency of transitions between different scrolling velocities. For example, if a user first browses at a scrolling velocity v_i then at another scrolling velocity v_j during a browsing, the $(i, j)^{th}$ entry q_{ij} of the state transition matrix Q is $q_{ij} = P(v_j|v_i) = \frac{1}{K}$, where K is the number of all transitions during the browsing.

To investigate the feasibility of fine-grained user rating inference based on the browsing velocity sequence, we study the state transition matrix (i.e., the browsing velocity sequence) for browsing classification. For the sake of simplicity, we assume the state transition matrix is:

$$Q = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad (2)$$

where a , b , c and d are the numbers of transitions between slow scrollings, from slow to fast scrolling, from fast to slow scrolling, and between fast scrollings respectively.

Intuitively, a user may prefer to like a browsing content if the user browses content more slowly (i.e., the value of a is larger), while dislike a content if he/she browses content under a larger browsing velocity (i.e., the value of d is larger). However, it is difficult to investigate the impact of slow to fast scrolling (i.e., b) and fast to slow scrolling (i.e., c) on user ratings about contents. If a user's interest of a browsing is regarded as like-preferred in advance, a larger value of a indicates a higher rating level for the browsing in the like-preferred domain. On the other hand, if a user's interest is regarded as dislike-preferred in advance, a larger value of d leads to a lower rating level in the dislike-preferred domain. Fortunately, we can infer the coarse-grained user interest of a browsing through the browsing velocity stability. Therefore, after the coarse-grained user interest is given, it is feasible to infer fine-grained user ratings about browsing contents in like-preferred and dislike-preferred domains respectively based on the browsing velocity sequence.

To validate the analysis, we extract a Markov chain for each browsing from traces of the 40 volunteers, and then adopt Principal Component Analysis (PCA)[45] technique to derive the first two principal components (PCs) of the state transition matrix. Fig. 4(a) shows dimension-reduced features of state transition matrices during 5-star, 4-star, 2-star, and 1-star browsings. We can see that the four kinds of browsings are indistinguishable. However, for like-preferred browsings shown in the top part of Fig. 4(b), it can be observed that most of 5-star browsings are distinguishable from 4-star browsings through the solid line. Dislike-preferred browsings depict similar results.

These results demonstrate that the browsing velocity sequence is feasible to infer users' fine-grained ratings about browsing contents after extracting the coarse-grained user interests about contents.

3.5 Underlying Connection between User Interest and Browsing Velocity Stability and Sequence

Through empirical studies, we find that the two intrinsic features, i.e., browsing velocity stability and browsing velocity sequence, can exhibit user interest during a browsing underlying the scrolling interactions. From the perspective of psychology, there is the underlying connection between user interest and the two features. In psychology, if a user has more interest in an activity/a behavior, he/she would pay more attention to the specific activity/behavior, which indicates a tight relationship between user interest and attention degree during a browsing. In cognitive load theory[22], there are two kinds of mechanisms that reveal the user attention degree during a behavior. The first mechanism is the perceptual selection mechanism[23], in which an individual can ignore irrelevant information when he/she is under high perceptual load (i.e., large amount of information absorption by the user). Another one is the active attentional control mechanism[22], in which the individual can actively reject irrelevant information even under low perceptual load. These two mechanisms indicate a user's cognitive attention degree on a specific behavior, and such a cognitive attention degree can be revealed with the scrolling human-computer interactions. When a user adopts the perceptual selection mechanism, i.e., under a higher attention degree, in a browsing, he/she should focus more on the content, and the scrolling interactions would be stable and repetitive, which leads to a high browsing velocity stability and concentrated browsing velocity sequence. Therefore, the two intrinsic features are tightly related to a user's interest during a browsing.

From the analysis above, we explore the intrinsic features, i.e., browsing velocity stability and browsing velocity sequence, to depict the fine-grained browsing process through scrolling interactions. Such features can reveal more high-level and objective characteristics of scrolling interactions during a browsing.

4 SYSTEM DESIGN

To precisely infer user interests about browsing contents, we propose an *Intelligent Interest Inference* system, I³, which leverages scrolling human-computer interactions during users' browsings on smartphones to infer interests of browsings.

4.1 Overview

Fig. 5 shows the architecture of I³. *Building Silent Rating Inference Model* is the core part of I³. I³ first extracts browsing features from scrolling interactions, i.e., browsing velocity stability and browsing velocity sequence, in *Feature Extraction*. After that, in *Model Training*, I³ constructs an interest discriminator through a Naive Bayesian-based approach to infer the coarse-grained user interest about browsing contents. And then based on the browsing velocity sequence, I³ trains rating classifiers through a deep learning-based approach to further infer a user's fine-grained rating of a browsing under like-preferred or dislike-preferred domain. Finally, the silent rating inference model, including the interest discriminator and rating classifiers, is built and stored in a database.

Based on the built silent rating inference model, I³ infers a user's interest based on scrolling human-computer interactions during the user's browsing in *Inferring Users' Ratings about Contents*. When a user browses a webpage on a smartphone, I³ first extracts the browsing velocity stability and browsing velocity sequence from scrolling interactions of the user in *Feature Extraction*. In *User Rating Inference*, with the built silent rating inference model, I³ first inputs the browsing velocity stability into the interest discriminator to infer the user's coarse-grained interest about the browsing content. Then, I³ feeds the browsing velocity sequence during the browsing into the rating classifiers for inferring the user's fine-grained rating about the browsing content.

Fig. 5. Architecture of I^3 .

4.2 Building Silent Rating Inference Model

In this section, we describe the silent rating inference model building process leveraging intrinsic features of scrolling interactions during browsings, which is the core part of I^3 .

4.2.1 Feature Extraction. The browsing velocity stability and browsing velocity sequence are basic browsing features from scrolling interactions of users. To build the silent rating inference model, I^3 first needs to extract browsing features from scrolling interactions.

For the browsing velocity stability, I^3 extracts the stability degree during a browsing to represent the stability of the browsing. Mathematically, we denote different scrolling velocities during a browsing i as $v^{(i)} = [v_0^{(i)}, v_1^{(i)}, \dots, v_{M-1}^{(i)}]$. The absolute frequencies of these scrolling velocities during the browsing i are $f^{(i)} = [f_0^{(i)}, f_1^{(i)}, \dots, f_{M-1}^{(i)}]$. Then, the stability degree of the browsing i can be represented as $stab^{(i)} = \max\{f_0^{(i)}, f_1^{(i)}, \dots, f_{M-1}^{(i)}\}$.

For the browsing velocity sequence, I^3 applies Markov chain $x = \langle S, x^{(0)}, Q \rangle$ to represent the sequential relationship between scrolling velocities during a browsing. The state space S of Markov chain contains all unique scrolling velocities during a browsing. Since the number of unique scrolling velocities M during different browsings are different with each other, the sizes of state spaces S in Markov chains are different. To make Markov chains of different browsings comparable, I^3 needs to align the sizes of state spaces. I^3 first centers all scrolling velocities to the range of $[-1, 1]$ to obtain the relative scrolling velocities, and then discretizes the relative scrolling velocity space $[-1, 1]$ into N states, which constructs the state space $S = \{v_0, v_1, \dots, v_{N-1}\}$. For example, during a user's browsing, the unique scrolling velocities are $\{0, 20.45, 370.96, 400.01\}$ pixels/ms, while that during another browsing are $\{0, 37.01, 45.96, 100.22, 104.71, 120.34\}$ pixels/ms. I^3 first centers scrolling velocities of the two browsings as $\{0, 0.05, 0.93, 1\}$ and $\{0, 0.31, 0.38, 0.83, 0.87, 1\}$. And then I^3 discretizes the relative scrolling velocities to N (e.g., $N = 8$) states, which constructs the state space $S = \{v_0, v_1, \dots, v_7\}$. Based on the state space S including N states (i.e., N different scrolling velocities), I^3 further derives the $N \times N$ -dimension state transition matrix Q and N -dimension initial probability vector $x^{(0)}$ respectively.

4.2.2 Model Training. Based on extracted features, I^3 trains the silent rating inference model.

Naive Bayesian-based Interest Discriminator Construction. I^3 first constructs an interest discriminator based on the extracted browsing velocity stability from user browsing traces through a Naive Bayesian-based

approach. The interest discriminator can be used to infer coarse-grained user interests (i.e., like-preferred, neutral-preferred or dislike-preferred).

Assume there are n rating levels (i.e., 1~ n -star) for users to express interests about browsing contents in the traces. As mentioned in Section 3, the browsing velocity stability can be used to express the coarse-grained user interests (i.e., like-preferred, neutral-preferred, or dislike-preferred). Hence, we first relabel the browsing traces with the coarse-grained user interests. Specifically, if n is an odd,

- $(n + 1)/2 + 1 \sim n$ -star are replaced with *like-preferred*;
- $(n + 1)/2$ -star is replaced with *neutral-preferred*;
- $1 \sim (n + 1)/2 - 1$ -star are replaced with *dislike-preferred*.

On the other hand, if n is an even,

- $n/2 + 1 \sim n$ -star are replaced with *like-preferred*;
- $1 \sim n/2$ -star are replaced with *dislike-preferred*.

Based on the relabeled traces, we further train the interest discriminator through Naive Bayesian approach[40].

Since the browsing velocity stability is formulated with the stability degree, i.e., a one-dimension feature, we consider to utilize a light-weight classifying approach to infer coarse-grained user interests. Naive Bayesian (NB) classifier is a basic probabilistic classifier based on Bayes' theorem, which can be executed within linear computational complexity. Hence, we train the interest discriminator based on NB method.

Given a stability degree $stab^{(i)}$ of browsing i , we intend to observe the conditional probability $p(c_k|stab^{(i)})$, i.e., the probability that the coarse-grained user interest about browsing i is c_k (like-preferred, neutral-preferred, or dislike-preferred) under the stability degree of $stab^{(i)}$. Hence, c_k with the largest $p(c_k|stab^{(i)})$ could be regarded as the coarse-grained user interest c about browsing i , i.e.,

$$c = \arg \max_{c_k} p(c_k|stab^{(i)}). \quad (3)$$

To construct the NB classifier, it is necessary to derive conditional probability $p(c_k|stab^{(i)})$. Based on Bayes' theorem,

$$p(c_k|stab^{(i)}) = \frac{p(c_k)p(stab^{(i)}|c_k)}{p(stab^{(i)})}, \quad (4)$$

where $p(c_k)$ is the prior probability of the coarse-grained user interest c_k , $p(stab^{(i)})$ is the evidence of the feature $stab^{(i)}$, and $p(stab^{(i)}|c_k)$ is the likelihood of the feature $stab^{(i)}$ under the user interest of c_k . Since the evidence $p(stab^{(i)})$ remains the same for all browsings,

$$p(c_k|stab^{(i)}) \propto p(c_k)p(stab^{(i)}|c_k). \quad (5)$$

For the prior probability, since the coarse-grained user interest about a browsing is usually random, we consider $p(c_k)$ is equal for all potential user interests. Moreover, the likelihood $p(stab^{(i)}|c_k)$ can be derived based on the distribution of browsing traces[34].

Through the approach above, I^3 can construct an NB-based interest discriminator based on the browsing velocity stability. Based on the interest discriminator, I^3 can infer the coarse-grained user interest about the browsing (i.e., like-preferred, neutral-preferred, or dislike-preferred).

Deep Learning-based Rating Classifiers Training. In order to infer fine-grained ratings of browsings, I^3 further trains rating classifiers based on the extracted browsing velocity sequence from user browsing traces through a deep learning-based approach for like-preferred and dislike-preferred browsings respectively.

In Section 4.2.1, we extract the browsing velocity sequence through Markov chain $x = < S, x^{(0)}, Q >$. Since the state space S and initial probability vector $x^{(0)}$ are all the same for different browsings, we only utilize the state transition matrix Q of Markov chain as input to train rating classifiers. Although the state transition matrix Q of

Fig. 6. Illustration of the convolutional neural network.

Markov chain can be directly adopted to train rating classifiers, it suffers the curse of dimensionality[13]. Thus, it is necessary to further extract dimension-reduced features from the matrix Q for rating classifier training.

Traditional feature extraction methods abstract features by observing unique patterns manually. Features extracted by these methods usually have redundant information, and are poor in robustness. Although existing linear feature extraction approaches (e.g., PCA or LDA) can achieve preferable features by generating linear decision boundaries[45], the state transition matrix of Markov chain are usually non-linear separated. To extract efficient and reliable features from the state transition matrix and further infer users' fine-grained ratings about browsing contents, we develop a deep learning-based approach, i.e., Convolutional Neural Network. Convolutional Neural Network (CNN)[21] can exploit in-depth temporal and spatial characteristics underlying the original input, which is consistent with the characteristic of browsing velocity sequence.

As shown in Fig. 6, the proposed CNN consists of three layers as hidden layers, i.e., a convolutional layer, a pooling layer, and a fully-connected layer. The convolutional layer is the core of CNN, which intrinsically extracts the sequential relationship between scrolling velocities and filter other redundant information. The convolutional layer first abstracts in-depth features from the state transition matrix, and then the pooling layer reduces the dimension of extracted in-depth features from the convolutional layer. Finally, the fully-connected layer maps the reduced features to inferred fine-grained user ratings.

First, the convolutional layer conducts convolution operation between the input (i.e., the state transition matrix Q) and a 3×3 -dimension kernel $\Theta = [\theta_{ij}]_{3 \times 3}$, and outputs the feature map F_1 , i.e.,

$$F_1^{(m,n)} = f_c(q) = \text{ReLU}\left(\sum_{i,j} \theta_{(3-i)(3-j)} q_{ij} + b\right), \quad (6)$$

where q is a sub-matrix of the state transition matrix Q , i.e.,

$$q = \begin{bmatrix} Q_{(m,n)} & Q_{(m+1,n)} & Q_{(m+2,n)} \\ Q_{(m,n+1)} & Q_{(m+1,n+1)} & Q_{(m+1,n+2)} \\ Q_{(m+2,n)} & Q_{(m+2,n+1)} & Q_{(m+2,n+2)} \end{bmatrix},$$

$\text{ReLU}(x) = \max(0, x)$ is Rectified Linear Units[12] as the activation function, b is bias. Assume the dimension of state transition matrix is 4×4 , as shown in the convolutional layer in Fig. 6. The convolutional layer first selects top-left 3×3 sub-matrix as q , and then conducts convolution operation $f_c(\cdot)$ on sub-matrix q and kernel Θ to obtain the $(1, 1)^{th}$ entry of feature map $F_1^{(1,1)}$. With similar operations, the convolutional layer can finally output a 2×2 feature map F_1 . Through the layer, state transition matrix Q can be transformed to the in-depth feature map F_1 with the convolutional kernel. Next, to avoid overfitting problem[21], in the pooling layer, a max pooling approach is used for dimension reduction of the feature map. Mathematically, the pooling layer extracts

Fig. 7. Illustration of rating inference.

the maximum value in a particular range of the feature map through a 2×2 pooling filter $f_p(g)$ as follows:

$$F_2^{(m,n)} = f_p(g) = \max\{g_{11}, g_{12}, g_{21}, g_{22}\}, \quad (7)$$

where g is a sub-matrix of the feature map F_1 , i.e., $g = [F_1^{(2m-1, 2n-1)} \ F_1^{(2m-1, 2n)}; F_1^{(2m, 2n-1)} \ F_1^{(2m, 2n)}]$. Assume the dimension of extracted feature map F_1 is 4×4 , as shown in the pooling layer of Fig. 6. The pooling layer first selects the top-left 2×2 sub-matrix as g , and then applies the max pooling filter $f_p(\cdot)$ on the sub-matrix g to obtain the $(1, 1)^{th}$ entry of the reduced feature $F_2^{(1,1)}$. With similar operations, the pooling layer finally outputs a 2×2 -dimension reduced feature F_2 . Through the pooling layer, the in-depth feature map F_1 can be transformed to the reduced feature F_2 with the max pooling approach. Finally, in the fully-connected layer, I^3 adopts the softmax function[3] to map the reduced feature F_2 to inferred fine-grained user ratings.

Through the approach above, I^3 trains two CNN models (i.e., two rating classifiers), to infer fine-grained user ratings for like-preferred and dislike-preferred browsings respectively.

4.3 Inferring Users' Ratings about Contents

Based on the silent rating inference model, I^3 can infer user ratings about browsing contents. When a user browses a new webpage as the browsing content on a smartphone, I^3 first senses scrolling human-computer interactions during the browsing, and then obtains relative scrolling velocities during the browsing based on Eq. (1). Through analyzing scrolling velocities during the browsing, I^3 extracts browsing features underlying scrolling interactions, i.e., the browsing velocity stability and browsing velocity sequence.

After extracting browsing features, I^3 infers user ratings about contents during browsings, as shown in Fig. 7. To infer the user's rating, I^3 first inputs the extracted browsing velocity stability into the constructed NB-based interest discriminator for obtaining the coarse-grained user interest of the browsing (i.e., like-preferred, neutral-preferred, or dislike-preferred). To further obtain the fine-grained rating (i.e., 1~ n -star) about the browsing content, I^3 feeds the browsing velocity sequence into the trained CNN-based like-preferred rating classifier or dislike-preferred rating classifier depending on the inferred coarse-grained user rating from the NB-based interest discriminator.

Since the silent rating inference model is already built before the user interest inference, we can infer user interests about browsing contents within limited computational workload. Therefore, the user interest inference of I^3 is lightweight and computationally efficient for smartphones.

5 EVALUATIONS

In this section, we evaluate the performance of I^3 under real traces collected from 46 volunteers.

Fig. 8. Accuracy of I^3 with only browsing velocity (BV) stability, only browsing velocity (BV) sequence and both features under different numbers of rating levels and webpage types respectively.

5.1 Experiment Setup and Methodology

We implement I^3 on a Galaxy S3 with 4.8-inches screen, a Nexus 6P with 5.7-inches screen, a Huawei Honor X2 with 7.0-inches screen and a Galaxy Tab S2 with 9.7-inches screen respectively. To evaluate the performance of our system, we randomly fetch another 2,000 webpages from USA Today, VOA, BBC and CNN, which include three different kinds of webpages, i.e., text-only webpages, text-figure hybrid webpages, and figure only webpages. We also recruit another 46 volunteers, whose ages are in the range of [17, 53], to browse these webpages. Among them, 30 volunteers are male and the rest 16 are female. The volunteers are also involved in different occupations. Similar to the setup in motivation study, each volunteer is required to randomly browse the webpages, and only reads the same webpage once. And the volunteers can browse the webpages without any constraints, i.e., we do not request volunteers to browse webpages in a fixed time or browse a fixed number of webpages, and volunteers can skip to the next webpage at any point. During each browsing, our application traces the volunteer's scrolling interactions every 100ms. After each browsing, the volunteer gives an explicit rating from 3, 5, 7, 9 different rating levels for the webpage as the ground truth. The number of webpages browsed by each volunteer is different from each other. In total, we collected 5,139 browsing records for the evaluation. We conduct the statistical significance test[35] on the collected data to validate the significance of the collected data for experiments. Such a test derives a *p-value* to measure the significance of a data set. When $p < 0.05$, the data set can be regarded as statistically significant[35]. Specifically, we perform the one-way analysis of variance (ANOVA)[33] to obtain the *p-value* for the statistical significance test. The result shows that the *p-value* of the test is $p = 0.021 < 0.05$, which indicates that the collected data is statistically significant. For each experiment, we divide the collected data into 8 groups and conduct 8-fold cross validation for the evaluation.

We use several metrics to evaluate the performance of I^3 .

- **Accuracy:** the probability that an interest inference is correctly inferred with the explicit rating, i.e.,

$$Accuracy = \frac{\sum_{i=1}^n \rho_{ii}}{\sum_{j=1}^n \sum_{i=1}^n \rho_{ij}}.$$

- **F1-score:** a metric combining both precision and recall, i.e.,

$$F1-score_k = 2 \times \frac{P_k \times R_k}{P_k + R_k},$$

Fig. 9. F1-score of I^3 under different rating levels and different screen sizes.

where P_k and R_k are the precision and recall of the interest inference for a k -star browsing respectively. *Precision* of the interest inference for a k -star browsing is defined as the probability that the interest inference for a k -star browsing is exactly the k -star browsing in ground truth, i.e., $P_k = \rho_{kk} / \sum_{i=1}^n \rho_{ik}$. On the other hand, *Recall* of the interest inference for a k -star browsing is defined as the probability that a k -star browsing in ground truth is inferred as the k -star browsing, i.e., $R_k = \rho_{kk} / \sum_{i=1}^n \rho_{ki}$.

5.2 Performance of Interest Inference

Accuracy. Fig. 8(a) and 8(b) show the accuracy of I^3 with only browsing velocity (BV) stability, browsing velocity (BV) sequence and both features under different number of rating levels and webpage types respectively. We can see that the accuracies of I^3 with both features under different numbers of rating levels and webpage types are all around 90%, and that are similar to each other, which indicates that I^3 can achieve a satisfactory performance in interest inference under different numbers of rating levels and webpage types. Also, accuracies for figure-only webpages are a little lower than that of other two kinds of webpages. This is because a figure is usually presented with more screen spaces, which leads to quick and similar scrolling interactions during different browsings. Moreover, we can find the accuracies of I^3 with only one feature are all less than that of I^3 with both features. Specifically, the overall accuracies of I^3 with only BV stability and sequence are around 60% and 40% respectively. This result indicates both browsing velocity stability and sequence contribute to interest inference of I^3 , which is consistent with our analysis in Section 3.

F1-score. We further evaluate F1-score of I^3 under different rating levels to demonstrate the efficiency of user interest inference. Fig. 9 shows F1-score of I^3 under different rating levels respectively. We can observe that the overall F1-scores under four different rating levels are all above 90%, which indicates that I^3 achieves a satisfactory performance. For different rating levels, I^3 can achieve similar performance of interest inference, which is consistent with the result of accuracy. It can be also observed that the F1-scores of interest inference under different screen sizes are similar, which indicates that I^3 is insensitive to screen sizes. Although the length of contents decreases as the screen size increases, scrolling interactions would not decrease under smartphones with larger screen sizes during browsings. This is because users usually browse webpages with their eyes in a fixed range of the screen. Hence, during a browsing on a smartphone with a larger screen size, the user would continuously scroll the screen to make the contents in the fixed range for browsing. Therefore, interest inference performances of I^3 under different screen sizes are similar.

Performance Comparison. We also compare the performance of I^3 with that of dwell time[51] and frequency of visits[27] methods, as well as average velocity-based and random approaches. Fig. 10 shows accuracies of I^3 , dwell time, frequency of visits, average velocity-based and random approaches respectively. It can be seen that the overall accuracy of I^3 is 92.4%, and that of dwell time and frequency of visits are 69.5% and 20.0% respectively,

Fig. 10. Accuracies of I^3 , dwell time, frequency of visits, average velocity-based and random approaches.

Infering Group	G_1	G_2	G_3	G_4	G_5	G_6
G_1	0.917	0.894	0.863	0.892	0.874	0.857
G_2	0.884	0.912	0.863	0.862	0.909	0.851
G_3	0.866	0.883	0.911	0.852	0.873	0.899
G_4	0.904	0.870	0.858	0.925	0.896	0.861
G_5	0.894	0.919	0.851	0.900	0.933	0.863
G_6	0.858	0.871	0.907	0.862	0.882	0.920
Training Group	G_1	G_2	G_3	G_4	G_5	G_6

Fig. 11. Accuracy matrix of I^3 under different user types.

Fig. 12. Accuracy of I^3 under different sizes of training sets.

while that of average velocity-based and random approaches are only 61.3% and 20.1% respectively. Since dwell time and average velocity reveal similar browsing characteristics (i.e., the velocity for a whole browsing), the accuracy of interest inference under these two approaches tend to be similar. On the other hand, due to most browsings only having one visit, the accuracy of frequency of visits is similar with that of random approach. This result further demonstrates that I^3 can achieve satisfactory performance in the interest inference.

5.3 Impact of User Types

We evaluate the performance of I^3 under different user types. We divide the 46 volunteers into 6 groups based on their ages and genders, i.e., G_1 (male under 17-28 years old), G_2 (male under 29-45 years old), G_3 (male under over 45 years old), G_4 (female under 17-28 years old), G_5 (female under 29-45 years old), and G_6 (female under over 45 years old). For each group, we divide the collected data into 8 groups, and conduct 8-fold cross validation to evaluate the accuracies of I^3 under the six groups. Fig. 11 shows the accuracy matrix of I^3 under different user types. It can be observed that the differences of accuracies between different user types are insignificant and all of the accuracies are above 85%, which indicates that I^3 can tolerant the difference between various types of users, and accurately infer user interests. We also evaluate the performance of I^3 under different user occupations. But the results present that the accuracies of interest inference under different user occupation are similar, which indicates user occupation has insignificant impact on interest inference.

5.4 Impact of Training Set Size

The size of training set is proportional to times of users' browsings. Hence, we evaluate the performance of I^3 under different sizes of training sets. Fig. 12 shows the accuracy of I^3 under different sizes of training sets. We can see that the accuracy of I^3 first increases and then goes stable as the size of training set increases. Specifically, the overall accuracy is stable after the size of training set increase to 825 samples. After this size of training set, the increase of training set sizes cannot lead to higher accuracy. This result indicates that I^3 can achieve satisfactory performance with an acceptable size of training sets. Moreover, it can be observed that the size of training set for stable accuracy increases as the number of rating levels increases. This is because the increasing number of labels (i.e., the number of rating levels) for a classification problem requires more training data for an accurate interest inference. Also, we can see that the differences among training set sizes for stable accuracies under different number of rating levels are not significant. Specifically, the size of training set for stable accuracy under 3 rating levels is 510, while that under 9 rating levels is 920.

Fig. 13. Accuracy of I^3 under different lengths of contents with different smartphones.

Fig. 14. CDF of scrolling time in collected browsing data.

Fig. 15. Accuracy of I^3 under different scrolling time.

5.5 Impact of Content Length

I^3 utilizes scrolling interactions to infer user interests about contents. Due to few scrolling interactions under short webpages, I^3 may fail to infer user interests accurately. To measure the impact of content length on I^3 , we define the content length based on the screen of smartphones, i.e., the content length is the number of smartphone screens for a webpage. We divide the browsing traces into five groups based on the content length, i.e., 2 screens, 3 screens, 4 screens, 5 screens, as well as 6 screens and above, and then analyze accuracies of I^3 under the five groups respectively. Fig. 13 shows the accuracy of I^3 under different lengths of webpages with different smartphones. We can see that overall accuracies of I^3 are 64.9% and 76.3% under 2 screens and 3 screens contents respectively. This is because a shorter length of a webpage leads to fewer scrolling interactions during a browsing, and further leads to the performance degradation of I^3 . As the length of a content increases to 4 screens and above, accuracies of I^3 are all above 85% under four screen sizes.

In the experiment, we only evaluate the performance of I^3 under content lengths larger than 2 screens, without taking contents with a length of 1 screen into consideration. This is because there is no scrolling interaction when a user browses a content with the length of 1 screen, so that I^3 cannot infer a user's interest during a browsing with a 1 screen content. Fortunately, there are few webpages with 1-screen length in real environments. Through an analysis on collected 2,000 webpages, we find that only 0.25% of the 2,000 webpages are 1 screen contents under the 4.8-inches smartphone, and 1.1% of the 2,000 webpages are 1 screen contents under the 9.7-inches smartphone. For 1 screen contents, I^3 utilizes the dwell time of the browsing to infer user interests.

Moreover, we find that 80.4% of the 2,000 webpages' lengths are 4 screens and above under the 4.8-inches smartphone, and 76.3% of that are 4 screens and above under the 9.7-inches smartphone. Thus, I^3 can accurately infer user interests through scrolling interactions in real environments.

5.6 Impact of Scrolling Time

Scrolling time depicts the percentage of browsing time with scrolling interactions among the whole browsing, i.e., the scrolling time $ST = t_{scroll}/T$, where t_{scroll} is the browsing time with scrolling interactions, and T is the time of a whole browsing. Fig. 14 shows the CDF of scrolling time in the collected browsing data. It can be observed that the data is approximately linearly distributed in the scrolling time of [20, 80]%, and there is only a small amount data ranging in the scrolling time of [80, 100]%. To validate the impact of scrolling time on I^3 , we further evaluate the performance of I^3 under different scrolling time. Fig. 15 shows the accuracy of I^3 under different scrolling time. We find that the accuracy under the scrolling time ranging in [20, 40]% (i.e., few scrolling interactions) is lower than that under other scrolling time, which indicates performance degradation under fewer

Fig. 16. Performance of typical recommendation algorithms combined with I^3 , dwell time, Freq-Visits and OPT.

scrolling interactions. But as the scrolling time increases to 40% and larger, the accuracies are all above 90%, which indicates a robust performance of I^3 under different scrolling time.

5.7 Impact of I^3 on Content Recommendation

Recommendation is one of the most widely used personalized services. I^3 can infer user interests about contents and improve recommendation performance for users. To validate the improvement of I^3 for typical recommendation algorithms, we conduct extensive experiments with three baselines, i.e., the dwell time[51] and frequency of visits (Freq-Visits)[27] and optimal rating (OPT, i.e., every rating is set as the ground truth), to evaluate the improvement to typical recommendation algorithms through I^3 .

We adopt two typical recommendation algorithms, i.e., ICF[41] and SVD++[20], which are two variants of the collaborative filtering[4], to conduct the experiment. In the collaborative technique, the user-item matrix represents the relationships among all users on the basis of their ratings to items (i.e., contents). A user-item matrix is first constructed from various browsing traces, whose $(i, j)^{th}$ entry means the rating of user i to content j . The inferred silent ratings from the interest inference are also used to enrich the user-item matrix. The collaborative technique utilizes the ratings to recognize similar users and browsing contents for recommendation.

We use two metrics, i.e., the hit ratio and hit rank[47], to evaluate the improvement for typical recommendation algorithms through I^3 .

- **Hit Ratio** measures the improvement of proposed method over random recommendation. The performance of a random recommendation[54] is represented as $\frac{|M|}{|C| \cdot |U|}$, where C and U are the collections of candidate contents and users respectively, M is the set of the rating to contents from users. Thus, the hit ratio is defined as:

$$ratio^N = \frac{N_h}{N} \times \frac{|C| \cdot |U|}{|M|},$$

where N_h is the number of hits, and N is the length of the recommendation list.

- **Hit Rank** measures the closeness of rank results between the proposed recommendation and ideal recommendation. We assume $\langle r_1, \dots, r_m \rangle$ are the ranks of hits. Ideally, the rank of recommendation should be $r_i = i$. Thus, we define the hit rank as:

$$rank^N = \sum_{i=1}^N r_i / \frac{N(N+1)}{2},$$

where N is the length of recommendation list. The smaller value of hit rank indicates better performance of the recommendation algorithm.

Fig. 16(a) and 16(b) show hit ratios of ICF and SVD++ combined with I^3 , dwell time, Freq-Visits and OPT respectively. We can observe that both ICF and SVD++ combined with I^3 benefit more improvements than

that with dwell time and Freq-Visits. From Fig. 16(a), we see that the average hit ratio of ICF combined with I^3 is 16.4% and 50.7% larger than that with dwell time and Freq-Visits respectively. From Fig. 16(b), we find that the average hit ratio of SVD++ combined with I^3 is 20.4% and 42.1% larger than that with dwell time and Freq-Visits respectively. Also, we can observe from Fig. 16(b) that the hit ratios of I^3 -ICF and OPT-ICF are close, so as hit ratios of I^3 -SVD++ and OPT-SVD++. This result validates I^3 can efficiently improve the performance of recommendation algorithms.

Except for the hit ratio, we also utilize the hit rank to evaluate the performance improvement that I^3 makes for recommendations. Fig. 16(c) and 16(d) show hit ranks of ICF and SVD++ combined with I^3 , dwell time, Freq-Visits and OPT respectively. Similar to the result of hit ratios, we can see that both algorithms combined with I^3 achieve better performance than that with dwell time and Freq-Visits methods. The average hit rank of ICF combined with I^3 is 27.3% and 40.3% lower than that with dwell time and Freq-Visits respectively. The average hit rank of SVD++ combined with I^3 is 14.7% and 33.9% lower than that with dwell time and Freq-Visits respectively. Additionally, the hit ranks of both algorithms combined with I^3 are close to that with OPT. This result further validates that I^3 can accurately infer user interests and improve the performance of typical recommendation algorithms.

6 DISCUSSION

Forward Scrolling vs. Backward Scrolling. In real environments, users browse webpages with not only forward scrolling, but also backward scrolling. Thus, it is necessary to consider both scrolling velocities in the design of interest inference system. In the design of I^3 , we actually take both forward and backward scrolling velocities into consideration on building the interest inference model. Specifically, we construct the interest classifier and rating classifiers based on the browsing velocity stability and browsing velocity sequence respectively. We set the velocity of forward scrolling as positive value, and that of backward scrolling as negative value. Both positive and negative velocities are involved to derive the browsing velocity stability and browsing velocity sequence. Therefore, the designed system not only considers the value of scrolling velocities, but also is related with the direction of scrolling velocities.

Impact of User-related Factors on I^3 . Different from the extrinsic statistical features (e.g., the average velocity, dwell time, etc.), the browsing velocity stability and sequence are the intrinsic features. The intrinsic features can exhibit in-depth browsing characteristics underlying the scrolling interactions, which are not easily affected by various user-related factors theoretically[15], such as search skill, knowledge background, etc. If a user is interested in a webpage, the browsing process should be still stable and concentrated, i.e., the value of browsing velocity stability is large and the matrix of browsing velocity sequence is concentrated, regardless of the search skill or familiarity. Hence, the interest inference based on the two intrinsic features is not affected by user-related factors.

Comparison with Existing Solutions. I^3 provides an intelligent interest inference which outperforms existing interest inference solutions. Many existing works[19, 27, 37, 51] explore the frequency of visits and dwell time to infer user interest during the browsings. However, both of the two measures are extrinsic statistical measures, which are easily affected by feature outliers. To reveal dynamic browsing properties, recent works[7, 48, 50] exploit features from human-computer interactions (e.g., tracking eye movements) during browsings of users. However, these solutions either arouse privacy concerns or induce extra power consumptions. Different from all of these solutions, I^3 explores the intrinsic feature underlying the scrolling interactions, which achieves robust interest inference performance while induce no privacy and power consumption concerns.

Limitations. This work is a research prototype, instead of a mature industry product. Hence, there are several limitations. i). Browsing Scenario: in this work, we concentrate on inferring user interest about the whole webpage, instead of parts in the webpage. However, for other kinds of browsing tasks, the scrolling interactions may be

quite different. For example, if a user is looking for something important or relevant during a browsing, I^3 cannot accurately infer the user interest with scrolling interactions. Such a situation is beyond the goal of this work, which is a limitation of this paper. But there are some existing works[16, 26, 28] investigating the impacts of task type and search intentions on users' browsing strategies. The early work[16] explores the cursor interactions can be used to understand and improve search behaviors. Another work[26] studies the impacts of various task types on scrolling interactions. Recent work[28] even discusses the connections between task, query-level search intentions, and web search behaviors. Along this direction, we expect to extend our work to investigate the impact of these factors on scrolling interactions and further on scrolling-based interest inference in the future. ii). Collected Data: in the data collection of motivation study and evaluation, the volunteers are recruited through an online recruitment method, and most volunteers share similar university education background. Due to the difficulty of recruiting volunteers outside the university, the studies under these volunteers are more representative for university users. To extend the system to common users, we need to further collect more data from various people with different backgrounds.

7 CONCLUSIONS

In this paper, we first analyze real browsing traces collected from 40 volunteers, and find two intrinsic browsing features, i.e., the browsing velocity stability and browsing velocity sequence, which are tightly related to user interests. Inspired by the observation, we propose an intelligent interest inference system, I^3 , which infers user interests through sensing scrolling human-computer interactions on smartphones during browsings. First, I^3 extracts the intrinsic browsing features from scrolling interactions during browsings. Based on extracted features, I^3 then builds a silent rating inference model including an interest discriminator and rating classifiers, to model the relationship between scrolling interactions and user interests. Finally, I^3 infers a user's interest during a browsing based on the silent rating inference model. Experimental results show that I^3 can accurately infer user interests, and is feasible as well as efficient to improve the performance of recommendation. Moving forward, we are interested in integrating both scrolling interactions and eye-tracking techniques to further improve the performance of interest inference on mobile devices.

ACKNOWLEDGMENTS

This research is supported in part by NSFC (No. 61772338, 61772341, 61472254) and STSCM (No. 18511103002). This work is also supported in part by the NSF grants CNS1801630, CNS1820624, CNS1814590 and Army Research Office W911NF-18-1-0221. This research is also supported by the China Scholarship Council, Program for Changjiang Young Scholars in University of China, the Program for China Top Young Talents and SJTU Global Strategic Partnership Fund (2019 SJTU-HKUST).

REFERENCES

- [1] Amazon. 2019. Amazon.com: Online Shopping for Electronic, Apparel, Computers, Books, DVDs & more. [Online]. Available: <https://www.amazon.com>.
- [2] BBC. 2019. Home - BBC News. [Online]. Available: <http://www.bbc.com/news>.
- [3] Christopher M. Bishop. 2006. Pattern recognition. *Machine Learning* 128 (2006).
- [4] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370.
- [5] E. Carmel, S. Crawford, and H. Chen. 1992. Browsing in hypertext: a cognitive study. *IEEE Transactions on Systems, Man, and Cybernetics* 22, 5 (1992), 865–884.
- [6] Giuliana Carullo, Aniello Castiglione, Alfredo De Santis, and Francesco Palmieri. 2015. A triadic closure and homophily-based recommendation system for online social networks. *World Wide Web* 18, 6 (2015), 1579–1601.
- [7] Sylvain Castagnos, Nicolas Jones, and Pearl Pu. 2010. Eye-tracking Product Recommenders' Usage. In *Proc. ACM Recsys'10*. Barcelona, Spain, 29–36.

- [8] M. Claypool, D. Brown, P. Le, and M. Waseda. 2001. Inferring user interest. *IEEE Internet Computing* 5, 6 (2001), 32–39.
- [9] Mark Claypool, Phong Le, Makoto Wased, and David Brown. 2001. Implicit Interest Indicators. In *Proc. ACM IUT'01*. Santa Fe, NM, USA, 33–40.
- [10] CNN. 2019. CNN - Breaking News, U.S., World, Weather, Entertainment & Vedio News. [Online]. Available: <http://edition.cnn.com>.
- [11] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. 1995. *Markov chain Monte Carlo in practice*. CRC press.
- [12] Xavier Glorot, Antoine Bordes, Yoshua Bengio, Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2012. Deep Sparse Rectifier Neural Networks. In *Proc. AISTATS'12*. La Palma, Canary Islands.
- [13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (2 ed.). Springer-Verlag New York Inc.
- [14] Jannica Heinström. 2006. Psychological factors behind incidental information acquisition. *Library & Information Science Research* 28, 4 (2006), 579–594.
- [15] Victoria Hodge and Jim Austin. 2004. A survey of outlier detection methodologies. *Artificial intelligence review* 22, 2 (2004), 85–126.
- [16] Jeff Huang, Ryen W. White, and Susan Dumais. 2011. No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search. In *Proc. ACM CHI'11*. Vancouver, BC, Canada, 1225–1234.
- [17] Takeo Igarashi and Ken Hinckley. 2000. Speed-dependent Automatic Zooming for Browsing Large Documents. In *Proc. ACM UIST'00*. San Diego, California, USA.
- [18] Diane Kelly and Nicholas J. Belkin. 2001. Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback. In *Proc. ACM SIGIR'01*. New Orleans, Louisiana, USA, 408–409.
- [19] Youngho Kim, Ahmed Hassan, Ryen W. White, and Imed Zitouni. 2014. Modeling Dwell Time to Predict Click-level Satisfaction. In *Proc. ACM WSDM '14*. New York, NY, USA.
- [20] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proc. ACM SIGKDD'08*. Las Vegas, Nevada, USA.
- [21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Proc. MIT Press NIPS'12*. Lake Tahoe, Nevada, USA.
- [22] Nilli Lavie. 1995. Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance* 21, 3 (1995), 451.
- [23] Nilli Lavie, Aleksandra Hirst, Jan W De Fockert, and Essi Viding. 2004. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General* 133, 3 (2004), 339.
- [24] Byungjoo Lee, Olli Savisaari, and Antti Oulasvirta. 2016. Spotlights: Attention-Optimized Highlights for Skim Reading. In *Proc. ACM CHI'16*. San Jose, CA, USA.
- [25] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. 2018. xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems. In *Proc. ACM KDD'18*. London, United Kingdom, 1754–1763.
- [26] Chang Liu, Jiqun Liu, and Yiming Wei. 2017. Scroll up or down? Using Wheel Activity as an Indicator of Browsing Strategy across Different Contextual Factors. In *Proc. ACM CHIIR'17*. Oslo, Norway.
- [27] Jiahui Liu, Peter Dolan, and Elin Ronby Pedersen. 2010. Personalized News Recommendation Based on Click Behavior. In *Proc. ACM IUT'10*. Hong Kong, China.
- [28] Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, and Chirag Shah. 2019. Task, Information Seeking Intentions, and User Behavior: Toward A Multi-level Understanding of Web Search. In *Proc. ACM CHIIR'19*. Glasgow, Scotland UK, 123–132.
- [29] J. Lo and Y. Chan. 2008. Relationships between User Cognitive Styles and Browsing Behaviors of an Online Learning Web Site. In *Proc. IEEE CW*. Hangzhou, China, 51–57.
- [30] L. Lu, J. Yu, Y. Chen, H. Liu, Y. Zhu, Y. Liu, and M. Li. 2018. LipPass: Lip Reading-based User Authentication on Smartphones Leveraging Acoustic Signals. In *Proc. IEEE INFOCOM'18*. Honolulu, HI, USA, 1466–1474.
- [31] L. Lu, J. Yu, Y. Chen, Y. Zhu, X. Xu, G. Xue, and M. Li. 2019. KeyLiSterber: Inferring Keystrokes on QWERTY Keyboard of Touch Screen through Acoustic Signals. In *Proc. IEEE INFOCOM'19*. Paris, France, 775–783.
- [32] Xiao Lu, Peng Li, Hongyuan Ma, Shuxin Wang, Anying Xu, and Bin Wang. 2014. Computing and applying topic-level user interactions in microblog recommendation. In *Proc. ACM SIGIR'14*. Gold Coast, Australia, 843–846.
- [33] MathWorks. 2019. One-Way ANOVA - MATLAB & Simulink - MathWorks. [Online]. Available: <https://ww2.mathworks.cn/help/stats/one-way-anova.html>.
- [34] Thomas M. Mitchell. 2003. *Machine learning*. China Machine Press.
- [35] Douglas C Montgomery, George C Runger, and Norma F Hubele. 2009. *Engineering statistics*. John Wiley & Sons.
- [36] Masahiro Morita and Yoichi Shinoda. 1994. Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In *Proc. ACM SIGIR'94*. Dublin, Ireland.
- [37] Edward Rolando Núñez-Valdéz, Juan Manuel Cueva Lovelle, Oscar Sanjuán Martínez, Vicente García-Díaz, Patricia Ordoñez De Pablos, and Carlos Enrique Montenegro Marín. 2012. Implicit feedback techniques on recommender systems applied to electronic books. *Computers in Human Behavior* 28, 4 (2012), 1186–1193.

- [38] Voice of America(VOA). 2019. VOA - Voice of America English News. [Online]. Available: <https://www.insidevoa.com/>.
- [39] Quora. 2019. Quora. [Online]. Available: <https://www.quora.com/>.
- [40] Mehran Sahami. 1996. Learning Limited Dependence Bayesian Classifiers. In *Proc. ACM KDD'96*. Portland, Oregon, USA.
- [41] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proc. ACM WWW'01*. Hong Kong, China.
- [42] Madelijn Strick, Rob W. Holland, Rick van Baaren, and Ad Van Knippenberg. 2009. Humor in the Eye Tracker: Attention Capture and Distraction from Context Cues. *The Journal of General Psychology* 137, 1 (2009), 37–48.
- [43] USA Today. 2019. USA TODAY: Latest World and US News - USATODAY.com. [Online]. Available: <https://www.usatoday.com/>.
- [44] Qiuzhen Wang, Sa Yang, Manlu Liu, Zike Cao, and Qingguo Ma. 2014. An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems* 62 (2014), 1–10.
- [45] Xuechuan Wang and Kuldip K Paliwal. 2003. Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern recognition* 36, 10 (2003), 2429–2439.
- [46] Wikipedia. 2019. Personalization - Wikipedia. [Online]. Available: <https://en.wikipedia.org/wiki/Personalization>.
- [47] Liang Xiang, Quan Yuan, Shiwan Zhao, Li Chen, Xiatian Zhang, Qing Yang, and Jimeng Sun. 2010. Temporal recommendation on graphs via long- and short-term preference fusion. In *Proc. ACM SIGKDD'10*. Washington, DC, USA.
- [48] Songhua Xu, Hao Jiang, and Francis C.M. Lau. 2008. Personalized Online Document, Image and Video Recommendation via Commodity Eye-tracking. In *Proc. ACM Recsys'08*. Lausanne, Switzerland, 83–90.
- [49] X. Xu, J. Yu, Y. Zhu, Z. Wu, J. Li, and M. Li. 2018. Leveraging Smartphones for Vehicle Lane-Level Localization on Highways. *IEEE Transactions on Mobile Computing* 17, 8 (2018), 1894–1907.
- [50] Longqi Yang, Michael Sobolev, Christina Tsangouri, and Deborah Estrin. 2018. Understanding User Interactions with Podcast Recommendations Delivered via Voice. In *Proc. ACM RecSys'18*. Vancouver, BC, Canada, 190–194.
- [51] Peifeng Yin, Ping Luo, Wang-Chien Lee, and Min Wang. 2013. Silence is also evidence: interpreting dwell time for recommendation from psychological perspective. In *Proc. ACM SIGKDD'13*. Chicago, IL, USA.
- [52] J. Yu, H. Han, H. Zhu, Y. Chen, J. Yang, Y. Zhu, G. Xue, and M. Li. 2015. Sensing Human-Screen Interaction for Energy-Efficient Frame Rate Adaptation on Smartphones. *IEEE Transactions on Mobile Computing* 14, 8 (2015), 1698–1711.
- [53] J. Zhang, Y. Li, M. Chen, and L. You. 2016. An implicit feedback integrated LDA-based topic model for IPTV program recommendation. In *Proc. IEEE ISCIT'16*. Qingdao, China.
- [54] Tao Zhou, Zoltán Kucsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107, 10 (2010), 4511–4515.
- [55] Xiangmin Zhou, Lei Chen, Yanchun Zhang, Dong Qin, Longbing Cao, Guangyan Huang, and Chen Wang. 2017. Enhancing online video recommendation using social user interactions. *The VLDB Journal* 26, 5 (2017), 637–656.