Investigating the Role of Eye Movements and Physiological Signals in Search Satisfaction Prediction Using Geometric Analysis

Yingying Wu
Department of Mathematics, The University of Texas at Austin, Austin, TX

Yiqun Liu*
Department of Computer Science and Technology, Tsinghua University, Beijing, China.
E-mail: yiqunliu@tsinghua.edu.cn

Yen-Hsi Richard Tsai
Department of Mathematics, The University of Texas at Austin, Austin, TX

Shing-Tung Yau
Department of Mathematics, Harvard University, Cambridge, MA

Two general challenges faced by data analysis are the existence of noise and the extraction of meaningful information from collected data. In this study, we used a multiscale framework to reduce the effects caused by noise and to extract explainable geometric properties to characterize finite metric spaces. We conducted lab experiments that integrated the use of eye-tracking, electrodermal activity (EDA), and user logs to explore users' information-seeking behaviors on search engine result pages (SERPs). Experimental results of 1,590 search queries showed that the proposed strategies effectively predicted query-level user satisfaction using EDA and eye-tracking data. The bootstrap analysis showed that combining EDA and eye-tracking data with user behavior data extracted from user logs led to a significantly better linear model fit than using user behavior data alone. Furthermore, cross-user and cross-task validations showed that our methods can be generalized to different search engine users performing different preassigned tasks.

Introduction

Neurophysiological (NP) signals convey a wide range of information, such as the level of cognitive effort, curiosity, arousal, and emotional states (Mostafa & Gwizdka, 2016). Compared with traditional user behavior signals adopted in web search-related research, such as click-throughs, NP signals may capture evidence that is not directly expressed or expressible (Arapakis, Bai, & Cambazoglu, 2014) and that can be collected free of response bias (Barreda-Ángeles, Arapakis, Bai, Cambazoglu, & Pereda-Baños, 2015). Plenty of existing work has tried to correlate search behavior signals, including clicking, mouse movement, and query reformulation, with search performance metrics such as satisfaction and success (Arapakis & Leiva, 2016; Chen, Liu, Zhang, & Ma, 2017). However, none of these efforts has taken NP signals into consideration, probably due to the difficulties in preprocessing and interpreting low-level features. Therefore, the need for adequate techniques to understand the mechanisms of underlying physiological and behavioral signals has been acknowledged (Prokasy, 2012).

User satisfaction measures users' subjective feelings about their interactions with modern search engines and can be defined as the fulfillment of a specified information requirement (Kelly, 2009). Since search satisfaction is important user feedback for search system optimization, predicting search satisfaction is one of the major concerns in search evaluation studies (Liu et al., 2015). User satisfaction assesses the overall impact of the cognitive and communicative aspects of user-intermediary interactions (Wu & Liu, 2011); therefore, we believe that NP signals may provide more precise predictions of search quality and deepen our

*Corresponding author

Received November 16, 2017; revised March 23, 2019; accepted April 8, 2019

© 2019 ASIS&T • Published online June 28, 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.24240

understanding of users' satisfaction perception mechanisms in the search interaction process, which offers potential for the advancement of new search tools.

In this study we considered a particular type of NP signal named electrodermal activity (EDA), which has been widely used since the turn of the century as a dependent variable in research related to psychophysiology, psychopathology, the detection of deception, and social psychopathology as an aspect of behavior that can be measured and quantified (Prokasy, 2012). Research on EDA dates back to 1849 (Veraguth, 1907). Since then, EDA has been used to measure emotion (Sokolov, 1963) and attention (Maltzman & Raskin, 1965) and to differentiate between positive and negative emotions (Boucsein, 2012). Flanagan (1967) found a positive correlation between the magnitude of EDA amplitude and attention (+.64) and emotion (+.32). Forbes and Bolles (1936) also reported a connection between EDA and emotional reactions, which was confirmed by later investigators (Uno & Grings, 1965; Wilcott, Darrow, & Siegel, 1957; Yokota, Takahashi, Kondo, & Fujimori, 1959).

Both EDA and eye movements are associated with attention, which is important for predicting satisfaction (Chuklin & de Rijke, 2016); therefore, the relationship between search satisfaction and users' eye movement behaviors is also investigated in this work. The fact that eye movements provide evidence of visual attention was discovered early (Von Helmholtz, 1867). According to the eye–mind link hypothesis (Just & Carpenter, 1980), eye-tracking corresponds to the location of attention. Therefore, eye-tracking can help researchers understand search process complexity by effort indicators such as fixation duration, the existence of regression fixations, the spacing of fixations, and reading speed (Rayner, Pollatsek, Ashby, & Clifton Jr, 2012). They may serve as effective predictors of user perception in search satisfaction prediction studies.

Furthermore, user behaviors recorded in user logs have been shown to be an effective predictor of search satisfaction (Ageev, Guo, Lagun, & Agichtein, 2011; Feild, Allan, & Jones, 2010; Guo, Lagun, & Agichtein, 2012; Guo, White, Zhang, Anderson, & Dumais, 2011; Hassan, Jones, & Klinkner, 2010; Liu et al., 2015; Su, He, Liu, Zhang, & Ma, 2018), and we adopted these behavior features as a baseline to show that neurophysiological data achieve a similar effect.

Although existing works have already shown that some extracted high-level EDA signals, such as increased EDA levels, can reveal certain kinds of human emotions that otherwise go unnoticed (Barreda-Ángeles et al., 2015), how to extract effective search satisfaction predictors from fine-grained EDA data remains underinvestigated. Researchers have adopted a number of dependent variables obtained from EDA recordings, such as the basal level, amplitude, and latency of specific responses (Flanagan, 1967) and the frequency of nonspecific responses (Forbes & Bolles, 1936). However, the shape of the EDA signals and eye movements has not been taken into consideration, and there is still a limited understanding of the effect of individual differences on eye gaze patterns in information searches (Wittek, Liu, Darányi, Gedeon, & Lim, 2016).

To understand the relationship between users' satisfaction feedback and physiological and behavioral signals, we conducted an empirical study with 40 search users with diversified backgrounds. Each user performed 12 search tasks with eye movements and EDA signals were recorded. Each task contained multiple search queries to complete, and users were asked to give self-reported satisfaction ratings after each query and each task.

Since orthogonal invariants describe the change in EDA levels and disregard the reference level, and slope and curvature are invariant under rigid transition, we expected certain invariant properties that describe EDA graph and eye movements to remain unchanged under certain transformations. Yau (2014) classified geometric invariants into orthogonal, affine, conformal, projective, and homological geometries. Applications of conformal invariants have been presented using computer graphics (Gu & Yau, 2008; Huang, Gu, Lin, & Yau, 2016; Huang et al., 2016; Wu, Su, Yueh, Lin, & Yau, 2015; Wu, 2012; Yau, Gu, & Wang, 2002), and research on homological invariants has suggested that distributions of persistent homology barcodes provide robust invariants of metric measure spaces (Blumberg, Gal, Mandell, & Pancia, 2014). Recent efforts have also shown the effectiveness of predicting product search satisfaction on mobile phones using geometric invariants of EDA data (Wu, Liu, Su, Ma, & Ou, 2017). Since the EDA graph and eye movements we investigated reside in Euclidean space, we used geometric invariants to bridge the gap between low-level EDA and eye movement data and high-level user perceived satisfaction.

The EDA signals we collected in the time window of interest were multiscale, so the data have important features at multiple scales of time and/or space. Following the methods of Engquist and Tsai (2005) for stiff ordinary differential equations with oscillatory solutions, we used a multiscale approach to model and filter the data to recover the properties that are intrinsic to the collected data. An example of the process of smoothing out the collected EDA signals is presented in Figure 1. The enlarged raw EDA graphs show that the data oscillate and contain spurious noise; in contrast, the smoothed EDA data reflect the true variation in a user's EDA level more accurately, with sampling noise filtered out. The multiscale model allows the computation of geometric invariants in a discretized scenario with the presence of periodic sampling noise; as a result, the geometric invariants extracted as machine learning features are robust against individual differences and tasks performed.

The major contributions of this work are therefore threefold:

- We introduced the multiscale framework, which facilitates the approximation of geometric invariants in a finite metric space; we proved the rate of convergence for the discretization schemes; and we showed the equivalence of different discretizations.
- We found that geometric features from visual fixation sequences and EDA signals effectively predict user satisfaction, and the bootstrap analysis shows that the linear model fitted by the

DOI: 10.1002/asi

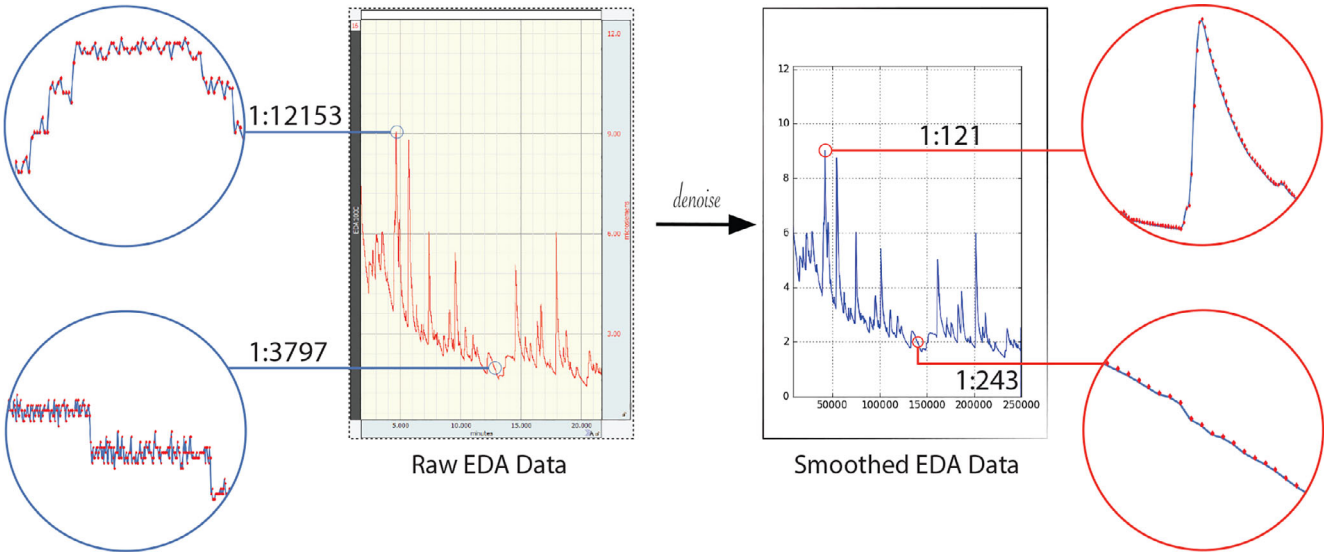

FIG. 1. Filtering out oscillation using a multiscale model. [Color figure can be viewed at wileyonlinelibrary.com]

physiological features and behavioral features better explains query-level user satisfaction than behavioral data alone.

- Empirical studies suggest that these geometric invariants are robust across users (Table 7), tasks (Table 8), and various environmental stimuli.

Related Work

Neurophysiological Signals for Information Retrieval (IR)

EDA refers to the conductivity of the skin, which varies according to the activation of the sympathetic branch of the autonomous nervous system (Boucsein, 2012). It can be very helpful in unveiling attentional and emotional reactions. Research on EDA by du Bois-Reymond dates back to 1849 (Veraguth, 1907). Since then, EDA has been used to measure stress (Lazarus, 1966; Lazarus & Opton Jr, 1966; Nomikos, Opton Jr, & Averill, 1968) and arousal (Duffy, 1972), especially for the lower arousal range, which reflects small variations (Boucsein, 2012). Researchers found that there is decreased skin conductance during pleasure (Stemmler, 1989) and increased skin conductance during fear (Ax, 1953; Boucsein, 2012; Stemmler, 1984, 1989; Wagner, 1989) and anger (Stemmler, 1989). Furthermore, studies incorporating EDA signals have observed that unconsciously perceived small web search latency increases could also affect the search experience (Barreda-Ángeles et al., 2015), contradicting the noticeable threshold of latency established by prior research (Arapakis et al., 2014). Using high-level EDA signals, Edwards and Kelly (2017) found that participants had the highest skin conductance when performing interesting tasks with normal search engine result page (SERP) quality and the second highest conductance level when performing uninteresting tasks with poor SERP quality. However, how to extract meaningful information to reveal the underlying information carried by EDA graphs in the presence of sampling noise and confounding factors remains under-investigated. An EDA graph is effective for predicting online shopping search satisfaction in mobile search scenarios (Wu et al., 2017), but whether utilizing EDA helps with understanding user satisfaction during general web searches remains unknown.

Because of the limited visual field that human eyes can see with full acuity, we need to move our eyes constantly. Therefore, the use of eye-tracking in information retrieval research under the hypothesis that visual fixation indicates attention was proposed over two decades ago. Researchers have been dedicated to using eye-tracking methodologies to explore the effects of changes in search result presentation (Cutrell & Guan, 2007; Granka, Joachims, & Gay, 2004), to study examination behavior (Tatler & Vincent, 2009; Xie et al., 2017), and to investigate the interaction between eye movements and visual saliency (Buscher, Cutrell, & Morris, 2009; Liu et al., 2016; Underwood & Foulsham, 2006). Gwizdka (2014) found that eye-tracking measures characterize relevance, which correlates with search satisfaction (Gluck, 1996; Huffman & Hochster, 2007; Mao et al., 2016). NP signals as indicators of human engagement and responses to search tasks come with the great promise of "lifting the lid" of the black box and may reveal what level of cognitive effort, anxiety, stress, curiosity, arousal, pleasure, and so forth, subjects experience while engaging in search tasks (Mostafa & Gwizdka, 2016). Moshfeghi, Triantafillou, and Pollick (2016) employed functional magnetic resonance imaging (fMRI) to study the connection between an information need and brain activity. These connections motivate us to extract geometric features that characterize eye-fixation as predictors of search satisfaction; however, combining eye-tracking information with EDA signals to understand the interactions between these signals remains unexplored.

DOI: 10.1002/asi

Relationship Between Emotion and the Autonomous Nervous System

For over a century, psychologists have tried to understand how the autonomic nervous system relates to emotions (Bernstein, Penner, Clarke-Stewart, & Roy, 2003). In the process, a number of theories have been developed. In the late 1800s, William James offered one of the first formal accounts of how physiological responses relate to emotional experience, known as the *peripheral theory* of emotion, which regards activity in the peripheral nervous system as the cause of emotional experience (James, 1890). A similar argument was given by Carl Lange; hence, James's view is also called the *James–Lange theory* of emotion. While James believed that the experience of emotion depends on feedback from physiological responses occurring outside the brain, Walter Cannon believed that the experience of emotion appears directly in the brain, with or without feedback from peripheral responses (Cannon, 1929). According to Cannon's central theory, also known as the *Cannon–Bard theory*, when the thalamus receives sensory information about emotional events, it sends signals to the autonomic nervous system and the cerebral cortex, where the emotion becomes conscious. Stanley Schachter suggested that emotions are shaped partly by how we interpret the arousal we feel. Schachter argued that emotions emerge from a combination of feedback from peripheral responses and cognitive interpretation based on the nature and cause of those responses, and interpretation occurs again when we identify feedback from those responses as a particular emotion (Schachter & Singer, 1962). His cognitive theory of emotion is known as the *Schachter–Singer theory.* Research on these theories suggests that both peripheral autonomic responses and the cognitive interpretation of those responses add to emotional experience, and the brain can also generate emotional experiences on its own, independent of physiological arousal.

Web Search User Satisfaction Analysis

Search satisfaction measures users' subjective feelings in regard to their interactions with the retrieval system and can be defined as the fulfillment of specified information needs (Kelly, 2009). Satisfaction as an evaluative criterion takes explicit account of the user and involves most of the components of IR interaction (Belkin & Vickery, 1985; Su, 2003; Wu & Liu, 2011). It is strongly correlated with certain evaluation metrics, such as cumulative gain (CG) and discounted cumulative gain (DCG) (Al-Maskari, Sanderson, & Clough, 2007); therefore, user satisfaction is sometimes referred to as the gold standard in search performance evaluation (Zhang et al., 2018). Research has focused on selecting effective mouse movement patterns (Guo et al., 2012; Liu et al., 2015) and query logs (Feild et al., 2010) to predict satisfaction. Lagun et al.'s study employed distance measures to improve result relevance estimation and reranking tasks (Lagun, Ageev, Guo, & Agichtein, 2014), and Liu et al.'s (2015) study computed the distribution difference to predict search satisfaction. However, satisfaction ratings are rather subjective: High satisfaction ratings are sometimes associated with low precision scores (Hitchingham, 1979), and searches with a large number of useful references are rated as less satisfying and vice versa (Tagliacozzo, 1977). Some studies have collected users' explicit feedback as the ground truth of satisfaction (Feild et al., 2010), and others have employed external assessors (Huffman & Hochster, 2007). However, significant differences between user feedback and annotator feedback have been found (Mao et al., 2016). Therefore, we studied users' search satisfaction and tried to find predictors that truly reflect their satisfaction.

Controlled User Study

User satisfaction measures users' subjective feelings and is defined as the fulfillment of a specified information requirement (Kelly, 2009). It is a positive emotional feeling related to the experience that one's goal has been fulfilled or is certain to be fulfilled in the future (Price, Barrell, & Barrell, 1985). Therefore, we believe that the perception of satisfaction may be closely related to certain neurophysiological signals. Specifically, we attempt to predict users' web search satisfaction with EDA signals and eye-tracking by using a controlled experiment in which we examine users' interactions during a web search experience. To this end, we collected the visual fixation sequence, EDA signals, and search logs to predict users' search satisfaction.

Experimental Design

The experiment used a repeated-measure design with user and search task as control variables. The independent variables were user behavior features extracted from search logs, descriptive features of the EDA signals, and eye-tracking—which are the minimum, maximum, average, median, variance, and standard deviation (SD) of EDA levels and eye movements—and geometric characterizations of eye-tracking and EDA signals proposed in this study. Each user performed 12 search tasks on the topics of TREC Session Track 2010–2014,[1] which were randomized by the Graeco-Latin square so that each task had the same opportunity to be shown to the users. We made minor modifications to the original TREC task descriptions based on the results from the pilot experiment to ensure that these tasks had appropriate complexity with no obvious different interpretations, as shown in Table 1. We also provided an initial query to reduce potential topic drifts.

Apparatus and platform. To obtain more natural and realistic interaction between the user and the experimental platform, we used a Tobii X2-30 (Stockholm, Sweden) remote eye-tracker to capture participants' eye movements. Eye fixations typically last ~200 to 250 ms, where 150 ms (Gwizdka, 2014) and 200 ms (Gerjets, Kammerer, & Werner, 2011) are

[1] https://trec.nist.gov/data/session.html

DOI: 10.1002/asi

TABLE 1. Task description.

Task	Description
1	The department recruited 10 new employees; please find a suitable icebreaker game for the new staff training.
2	Please choose a suitable travel destination for a week, considering the cost, value, and accessibility.
3	Please find fixed-gear bike price ranges, tips for purchasing, and available merchants.
4	Please find restrictions on the size of carry-on baggage for international flights.
5	Please find the premium and claim amounts of long-term care insurance and companies that provide such insurance.
6	Please find information about lupus erythematosus, such as treatment and precautions.
7	Please find the benefits of smoking cessation, side effects of smoking cessation, and effective ways to quit smoking.
8	Please find product features of Google Glass, the usage thereof, and prices.
9	Please find out whether Red Bull induces health concerns, which ingredients may lead to health problems, and which countries prohibit sales of Red Bull.
10	Please find iPhone prices and taxes in Hong Kong and whether duty-free iPhones are supported by domestic carriers.
11	Please find the location of Tsinghua University's swimming pool, hours, and student pricing.
12	Please find information about an aircraft carrier, including the water displacement, length, beam, number of crew, speed, weapons, and equipment.

typical thresholds of fixation duration; saccades indicate periods when eyes are moving, and they typically last 20 to 40 ms (Rayner, Juhasz, & Pollatsek, 2008). Tobii X2-30 has a sampling rate of 30 Hz, which implies an intermediary window of no sampling at 33.3 ms long. We detected fixations using built-in algorithms and all default parameters from Tobii Studio. A search system was deployed on a 17-inch LCD monitor with a resolution of 1366×768 pixels.

To analyze the emotional reactions of participants, we used a Biopac (Goleta, CA) MP-150 Data Acquisition System with a sampling rate of 200 Hz. Data collection devices with NP approaches may be intrusive; to minimize this concern, we collected EDA signals from participants' fingertips of the left index and middle fingers. AcqKnowledge 4.4 (Biopac) software was used for processing the data. In the background, an in-house application written in JavaScript was used to synchronize the physiological data with Unix timestamps.[2] We built an experiment platform so that when the user issues a query or clicks a pagination link, the system forwards the request to a commercial search engine and retrieves the corresponding SERP reformatted into a controlled experiment platform. A customized Google Chrome extension was injected to capture browser events including query reformulation, click, scrolling, and tab switching with Unix timestamps.

[2] A Unix timestamp is the number of seconds that have elapsed from 01-Jan-1970 0000 hrs to the second that the event has occurred (Sheikh, Wegdam, & Van Sinderen, 2007).

Participants. We recruited 40 participants of mixed ethnicities via online social networks; the first participant was recruited to examine the clarity of task statements and to test the system. All participants were familiar with the usage of web search engines. The age of the participants ranged from 18 to 33 years. Sixteen participants were female (40%). With regard to educational backgrounds, the participants ranged from current bachelor's degree students to postdoctoral researchers, and some had industrial work experience. Twenty-one of them wore corrective glasses during the experiment (52.5%). The native languages of the participants varied, but all participants had attained at least intermediate levels in English.

Procedure. The experiment was conducted in a separate room in a research lab. As illustrated in Figure 2, at the beginning of the experiment, each participant was asked to rest in the lab for 15 minutes to reduce the effects introduced to EDA levels due to prior activities. Then the fingertips of their left index and middle fingers were cleaned with 75% alcohol, and the residue was removed with saline wipes. Afterward, two electrodes were attached to each participant, and another 15 minutes were used to allow the electrodes to be completely settled on the fingertips. A total of 30 minutes in the lab allowed the physiological metrics to drop to a stable state, and then each participant was advised about the details of the experimental procedure. The baseline EDA level is generally considered to be the average level of an individual during resting conditions and in the absence of external stimuli. We collected the EDA baseline for each individual for 15 minutes in a resting state prior to performing search tasks, which is typical for baseline collection (Doberenz, Roth, Wollburg, Maslowski, & Kim, 2011).

After the eye-tracker was calibrated by a built-in program pre-established in Tobii Studio, the search tasks started. Each participant was asked to complete all 12 search tasks. For every search task, each participant first read the task description and clicked the start button. Then each participant was asked to briefly summarize the session task. Next, the initial SERP was presented, and each participant could click on the results to browse the landing pages or reformulate the query until she/he clicked the complete button at the bottom. In the post-session feedback stage, satisfaction for each search query submitted (including the initial query) was collected on a 4-point Likert scale (1: very unsatisfied, 2: somewhat satisfied, 3: fairly satisfied, 4: very satisfied) following Mao et al.'s treatment (2016). Similarly, overall satisfaction for each search task was collected. At the end of every search task, each participant was asked to answer a question related to the search task for quality control purposes. The average time each participant spent during the experiment was 178 minutes ($\sigma = 37.84$), out of which an average of 113.05 minutes ($\sigma = 30.24$) were spent on performing search tasks.

DOI: 10.1002/asi

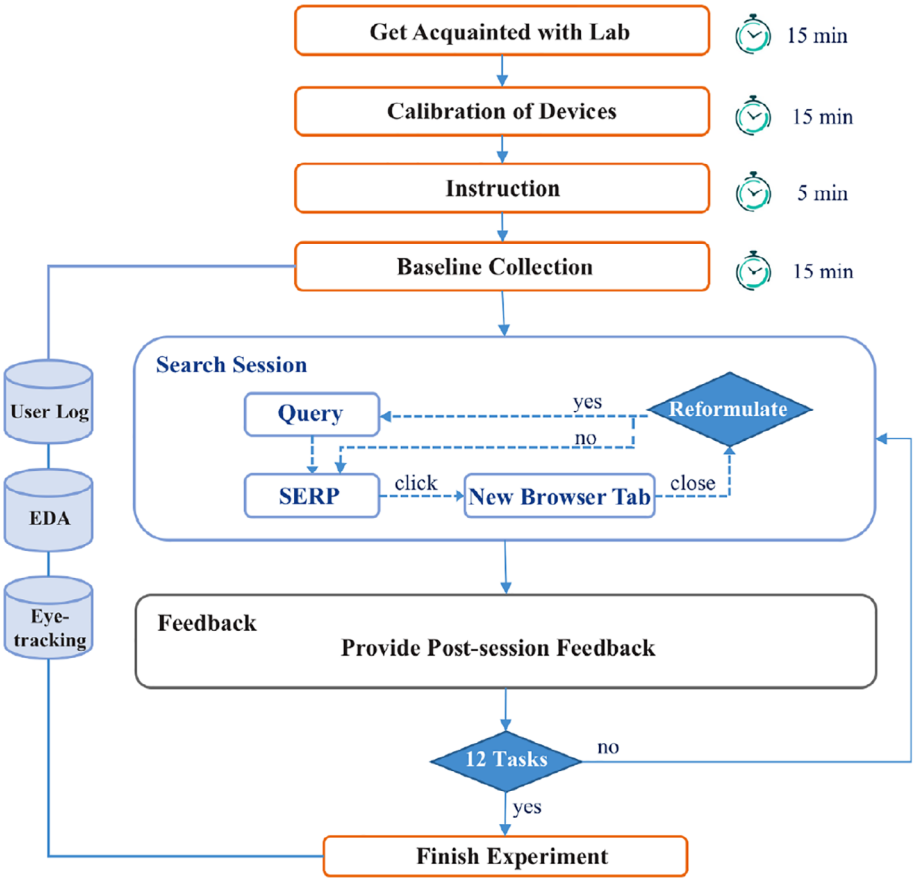

FIG. 2. Experimental procedure of predicting satisfaction using eye movements and physiological signals for general web searches. [Color figure can be viewed at wileyonlinelibrary.com]

Data

Figure 3 shows the ordinal satisfaction for queries and tasks collected in the experiment process. Out of 1,590 queries collected, a total of 439 queries were annotated with a satisfaction rating of 1 (28%); 254 queries were annotated with a satisfaction rating of 2 (16%); 466 queries were annotated with a satisfaction rating of 3 (30%); and 386 queries were annotated with a satisfaction rating of 4 (25%). The remaining 45 queries were not annotated. A total of 463 tasks were performed and annotated; there were 18 tasks annotated with a satisfaction rating of 1 (4%), 62 tasks annotated with a satisfaction rating of 2 (13%), 210 tasks annotated with a satisfaction rating of 3 (45%), and 173 tasks annotated with a satisfaction rating of 4 (37%). Five tasks performed by users were not rated.

The task satisfaction had higher ratings than the query satisfaction. We attribute this finding to the effectiveness of the commercial search engine, which is likely to return a query that addresses users' information needs precisely through query reformulations, therefore yielding higher session satisfaction. We also observe that simpler search tasks tend to receive better satisfaction ratings; for example,

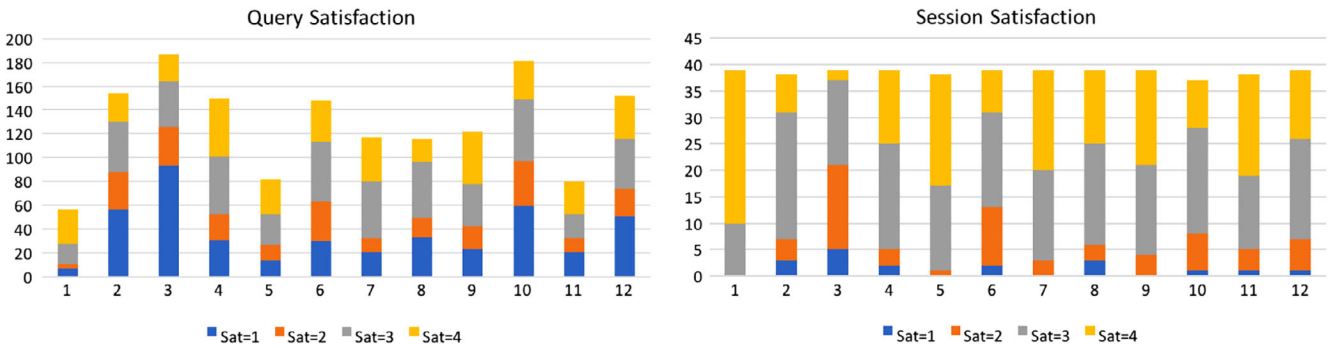

FIG. 3. Distribution of query satisfaction and task satisfaction. [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/asi

the average number of queries issued for each task was 3.30, and an average of only 1.43 queries were submitted for Task 1, "find an icebreaker game." Therefore, we consider this to have been an easy task. Although 50% of queries performed for Task 1 were rated as "very satisfied," only 25% of queries among all tasks were rated as "very satisfied." A similar situation occurred for Task 1 satisfaction, with 74% of Task 1 sessions rated as "very satisfied" by different users, while only 37% of sessions were rated as "very satisfied" for all sessions in different tasks. Conversely, for a difficult task such as Task 3, which asked a specific question on a comparatively obscure topic, "tips for purchasing fixed-gear bikes," the average number of queries submitted was 4.79, with 12% of queries for this task rated as "very satisfied" and 5% of sessions for this task rated as "very satisfied."

The Multiscale Framework

Two general challenges faced by data analysis are the existence of noise and the extraction of meaningful information from collected data. To address the first issue, we employed discretization schemes and the multiscale framework to filter out the sampling noise of EDA signals, and we proved their convergence to the true EDA graph, shown as Theorem 4.3.

To address the second issue, we constructed the feature space to predict satisfaction by eye-tracking sequences and EDA graphs using the geometric analogy of invariants in the finite metric spaces. Geometric invariants are natural characterizations of geometric objects in Euclidean space. The slope or gradient describes the direction and the steepness of a line, and the curvature is the amount by which a geometric object deviates from being flat. Figure 4 shows two examples of EDA graphs of our participants during the experiment.

However, because geometric invariants are defined for first- and second-order differentiable curves, we need to show that the discretized geometric properties converge to their smooth counterparts, presented as Theorem 4.11, which is the analytical equivalence of discretized curvatures. In the following sections, we will present four ways to discretize curvature. In particular, we introduce the method of moving frames that provides two ways to compute curvature via the Frenet frame and the Bishop frame.

Multiscale Modeling

In the time scale of interest, the EDA signals we collected were highly oscillatory in time. Since the EDA signals in the time window of interest are multiscale, such that the data have important features at multiple scales of time, we follow Enquist and Tsai's (2005) multiscale modeling approach and filter out the noise by smoothing out the data, as shown in Figure 1.

The way we transform the experimental data to a smoother curve for analysis is as follows: let $\left\{\left(t_j, f_j^\epsilon\right)\right\}_{j=1}^N$ be the collection of data points, and assume that f_j^ϵ are sampled from a function denoted by $f^\epsilon(t)$. More precisely, $f_j^\epsilon = f^\epsilon(t_j)$. The function $f^\epsilon(t)$ is highly oscillatory, but the amplitude of the fast oscillations is small. Hence, we use a small parameter epsilon to describe the period of the fast oscillations and write:

$$f^\epsilon(t) = \bar{f}(t) + g^\epsilon(t), \tag{1}$$

where $\bar{f}$ is smooth, its derivative is not large compared to $1/\epsilon$, and g^ϵ is our model for the spurious fast oscillations with small amplitudes in the observed data. The function $f^\epsilon(t)$ and error $g^\epsilon(t)$ can be expressed as the periodic functions:

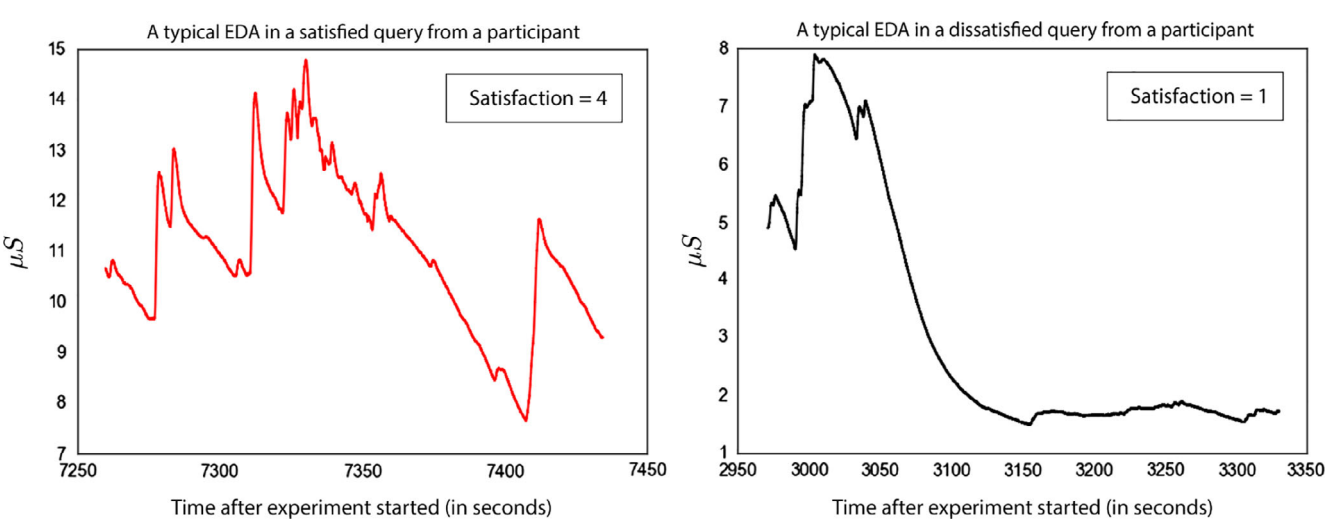

FIG. 4. Typical EDA graphs in a satisfied and dissatisfied query. [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/asi

$$f^{\epsilon}(t-t_j)=\sum_l\left(a_l\cos\frac{2\pi k_{jl}^1 t}{T}+b_l\sin\frac{2\pi k_{jl}^2 t}{T}\right),$$
$$g^{\epsilon}(t-t_j)=\delta\sum_l\left(c_l\cos\frac{2\pi k_{jl}^3 t}{\epsilon T}+d_l\sin\frac{2\pi k_{jl}^4 t}{\epsilon T}\right),$$

where $k_{jl}^1, k_{jl}^2, k_{jl}^3$ and k_{jl}^4 are integers, with $k_{jl}^1, k_{jl}^2 \ll \frac{1}{\epsilon}$, and k_{jl}^3 and k_{jl}^4 are not both equal to zero. By assumption, the amplitude of the oscillations represented by $g^{\epsilon}(t)$ is small; we define:

$$I(t,p,\epsilon)=\frac{1}{p}\int_{t_j}^{t_j+p} g^{\epsilon}(t)dt. \tag{2}$$

According to the definition of $g^{\epsilon}(t)$, for any fixed t and p and sufficiently small ϵ, the size of $I(t, p, \epsilon)$ is bounded above by some constant multiple of epsilon; that is,

$$|I(t,p,\epsilon)|\le C\epsilon, \text{for some constant } C.$$

We estimate the curvature of the graph of $\bar{f}$ by filtering out $g^{\epsilon}(t)$. We downsize the sampled data by averaging $p=100$ data points of $f^{\epsilon}(t)$ and obtain a function $\tilde{f}(t)$ with data points $\{(t_k, \tilde{f}_k)\}_{k=1}^{\tilde{N}}$, where $\tilde{N}\cdot 100=N$. Following Equation 2 and the Riemann sum approximation of $I(t, p, \epsilon)$ for $t=t_j$ and $p=100\cdot\Delta t$, the downsized error is:

$$\begin{aligned}\Phi_{100}\left[g^{\epsilon},t_j\right]&=\frac{1}{100\Delta t}\sum_{l=1}^{100}g^{\epsilon}\left(t_j+l\cdot\Delta t\right)\Delta t\\&=I\left(t_j,100\Delta t,\epsilon\right)+\mathcal{O}(\Delta t)=\mathcal{O}(\delta\cdot\epsilon)+\mathcal{O}(\Delta t).\end{aligned} \tag{3}$$

We assume that $\bar{f}$ varies slowly such that:

$$\left\|\frac{d}{dt}\bar{f}(t)\right\|_{\infty}\le C_2$$

for some constant C_2 independent of ϵ. The downsized $\bar{f}$ can be represented by:

$$\Phi_{100}\left(\bar{f}(t)\right)=\frac{1}{100\Delta t}\sum_{l=1}^{100}\bar{f}\left(t_j+l\cdot\Delta t\right)\Delta t=\bar{f}\left(t_j\right)+\mathcal{O}(\Delta t) \tag{4}$$

according to Taylor expansion. Combining Equations 1, 3, and 4, we have:

$$\Phi_{100}\left(f^{\epsilon}(t_j)\right)=\Phi_{100}\left(\bar{f}(t_j)\right)+\Phi_{100}\left(g^{\epsilon}(t_j)\right)=\bar{f}(t_j)+\mathcal{O}(\delta\cdot\epsilon)+\mathcal{O}(\Delta t). \tag{5}$$

Finite Difference Method

In this section, we employ finite difference methods to compute the slope and curvature as part of the feature space. A natural way to approximate the derivative of a function $f(x)$ on a given interval $[a, b]$ is to introduce $n+1$ nodes $\{x_0, \ldots, x_n\}$ with $x_0=a$, $x_n=b$ and $x_{k+1}=x_k+h$ for $k=0, \ldots, n-1$, where $h=\frac{b-a}{n}$. The discretized slope by centered finite difference is computed by the straight line passing through the points $(x_{i-1}, f(x_{i-1}))$ and $(x_{i+1}, f(x_{i+1}))$. The error induced by centered finite difference is second-order (Quarteroni, Sacco, & Saleri, 2007); that is, the error induced by the discretization is proportional to $\mathcal{O}(h^2)$. Rewriting the limit with the incremental ratio with h finite, the approximation u_i to $f'(x_i)$ by centered finite difference is:

$$u_i^{CD}=\frac{f(x_{i+1})-f(x_{i-1})}{2h}=f'(x_i)+\mathcal{O}\left(h^2\right), 0\le i\le n-1. \tag{6}$$

To estimate the curvature of the true curve with the downsized data points, we first consider the first-order derivative. Denote ΔT as the step size of downsized data, and then:

$$\begin{aligned}\Phi'_{100}\left(f^{\epsilon}(t_j)\right)&=\frac{\Phi_{100}\left(f^{\epsilon}(t_{j+1})\right)-\Phi_{100}\left(f^{\epsilon}(t_{j-1})\right)}{2\Delta T}\\&=\frac{\bar{f}(t_{j+1})-\bar{f}(t_{j-1})}{2\Delta T}+\frac{\mathcal{O}(\Delta t)+\mathcal{O}(\delta\cdot\epsilon)}{2\Delta T}\\&=\bar{f}'(t_j)+\mathcal{O}\left(\Delta T^2\right)+\mathcal{O}\left(\frac{\Delta t}{\Delta T}\right)+\mathcal{O}\left(\frac{\delta\cdot\epsilon}{\Delta T}\right).\end{aligned}$$

Finite difference approximations of higher-order derivatives of f can be constructed by a Taylor series. This gives the following centered finite difference scheme, which is a second-order approximation to $f''(x_i)$ with respect to h:

$$u_i''=\frac{f(x_{i+1})-2f(x_i)+f(x_{i-1})}{h^2}=f''(x_i)+\mathcal{O}\left(h^2\right),\quad 0\le i\le n-1.$$

Following Equation 5, the second-order derivative of the downsized data points is:

$$\begin{aligned}\Phi''_{100}\left(f^{\epsilon}(t_j)\right)&=\frac{\Phi_{100}\left(f^{\epsilon}(t_{j+1})\right)-2\Phi_{100}\left(f^{\epsilon}(t_j)\right)+\Phi_{100}\left(f^{\epsilon}(t_{j-1})\right)}{\Delta T^2}\\&=\frac{\bar{f}(t_{j+1})-2\bar{f}(t_j)+\bar{f}(t_{j-1})}{\Delta T^2}+\frac{\mathcal{O}(\Delta t)+\mathcal{O}(\delta\cdot\epsilon)}{\Delta T^2}\\&=\bar{f}''(t_j)+\mathcal{O}\left(\Delta T^2\right)+\mathcal{O}\left(\frac{\Delta t}{\Delta T^2}\right)+\mathcal{O}\left(\frac{\delta\cdot\epsilon}{\Delta T^2}\right).\end{aligned}$$

Definition 4.1. (Graph curvature). Consider a twice-differentiable graph $y=f(x)$, and denote the first- and second-order derivatives as $f'(x)$ and $f''(x)$. The graph curvature is:

$$\kappa(x)=\frac{f''(x)}{\left(1+f'(x)^2\right)^{3/2}}.$$

Definition 4.2. (Discrete graph curvature). Let $F(x_i)=y_i$ be the discretization of a twice-differentiable function f. The discrete graph curvature by second-order centered finite difference is defined by:

DOI: 10.1002/asi

$$\mathcal{K}(x_i) = \frac{u_i''}{\left(1 + (u_i^{CD})^2\right)^{3/2}}. \tag{7}$$

Theorem 4.3. *Consider a smooth curve in the Sobolev space $W^{2,\infty}(\Omega)$ of two derivatives: $W^{2,\infty}(\Omega) = \{f \in L_\infty(\Omega) : D^\alpha f \in L_\infty(\Omega)$ for all multi-indices α such that $|\alpha| \leq 2\}$, where the Lebesgue space $L_\infty(\Omega)$ is defined as the class of all measurable functions f such that:*

$$\|f(x)\|_\infty < \infty, \quad \Omega \subset \mathbb{R}^2.$$

Then the discretized curvature of the sampled curve downsized by the multiscale framework by centered finite difference approximates the curvature of the smooth curve with an error in the order of $\mathcal{O}(\Delta T^2) + \mathcal{O}\left(\frac{\Delta t}{\Delta T}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T}\right) + \mathcal{O}\left(\frac{\Delta t}{\Delta T^2}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T^2}\right)$.

Proof. By definition,

$$\mathcal{K}_{\Phi_{100}}(t_i) = \frac{\Phi_{100}''(f^\epsilon(t_j))}{\left(1 + \left(\Phi_{100}'(f^\epsilon(t_j))\right)^2\right)^{3/2}}$$
$$= \frac{\bar{f}''(t_j) + \mathcal{O}(\Delta T^2) + \mathcal{O}\left(\frac{\Delta t}{\Delta T^2}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T^2}\right)}{\left(1 + \left(\bar{f}'(t_j) + \mathcal{O}(\Delta T^2) + \mathcal{O}\left(\frac{\Delta t}{\Delta T}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T}\right)\right)^2\right)^{3/2}}.$$

We denote the error term of $\Phi_{100}''(f^\epsilon(t_j))$ by:

$$e_2 = \mathcal{O}(\Delta T^2) + \mathcal{O}\left(\frac{\Delta t}{\Delta T^2}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T^2}\right),$$

and the error term of $\Phi_{100}'(f^\epsilon(t_j))$ by:

$$e_1 = \mathcal{O}(\Delta T^2) + \mathcal{O}\left(\frac{\Delta t}{\Delta T}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T}\right).$$

Because $\bar{f}'$ is bounded, $(\bar{f}'(t_i) + e_1)^2 = \bar{f}'(t_i)^2 + e_1$; since $f \in W^{2,\infty}(\Omega)$, f' and f'' are bounded. Therefore, Taylor expanding $\left(1 + \bar{f}'(t_i)^2 + e_1\right)^{3/2}$ around $1 + \bar{f}'(t_i)^2$ gives:

$$\left(1 + \bar{f}'(t_i)^2 + e_1\right)^{-3/2} = \left(1 + \bar{f}'(t_i)^2\right)^{-3/2} + e_1.$$

Therefore,

$$\mathcal{K}_{\Phi_{100}}(t_i) = \left(\bar{f}''(t_i) + e_2\right)\left(\left(1 + f'(t_i)^2\right)^{-3/2} + e_1\right)$$
$$= \kappa(t_i) + e_1 + e_2$$
$$= \kappa(t_i) + \mathcal{O}(\Delta T^2) + \mathcal{O}\left(\frac{\Delta t}{\Delta T}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T}\right)$$
$$+ \mathcal{O}\left(\frac{\Delta t}{\Delta T^2}\right) + \mathcal{O}\left(\frac{\delta \cdot \epsilon}{\Delta T^2}\right).$$

□

Now, we consider the upper bound of the actual errors. In our experiment, $\Delta t = 0.005$, $\Delta T = 0.5$, $\epsilon = 0.02$, and $\delta = 0.001$, so $\Delta t \ll \Delta T < 1$. Then:

$$e_1 \leq C_1 \Delta T^2 + C_2\left(\frac{\Delta t}{\Delta T}\right) + C_3\left(\frac{\delta \cdot \epsilon}{\Delta T}\right) = 0.25C_1 + 0.01C_2 + 0.00004C_3,$$
$$e_2 \leq C_4 \Delta T^2 + C_5\left(\frac{\Delta t}{\Delta T^2}\right) + C_6\left(\frac{\delta \cdot \epsilon}{\Delta T^2}\right) = 0.25C_4 + 0.02C_5 + 0.00008C_6.$$

As a result, e_1 and e_2 are small; hence, $\mathcal{K}_{\Phi_{100}}(t_i)$ approximates the true curvature $\kappa(t_i)$.

The Method of Moving Frames

In this section, we derive formulae for geometric invariants of a discretized curve based on the methods of moving frames. All parallel vectors of the same length and orientation are identified in a vector space; however, in Euclidean space, the basic geometric object is a vector and its starting point. The notion of frame allows us to compute the vector plane phenomena of the Euclidean plane.

Definition 4.4. (Frame [Guggenheimer, 1963]). A frame is a vector $\{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ of mutually orthogonal unit vectors so that $\boldsymbol{e_2}$ is obtained from $\boldsymbol{e_1}$ by a rotation of $\pi/2$.

Definition 4.5. (Moving Frame [Guggenheimer, 1963]). Given a $\mathcal{C}^2$ curve in terms of its arc length, its tangent $\mathbf{t}(s) = x_1'\mathbf{e}_1 + x_2'\mathbf{e}_2$ is a unit vector. The normal $\mathbf{n}(s) = -x_2'\mathbf{e}_1 + x_1'\mathbf{e}_2$ is a unit vector obtained from $\boldsymbol{t}(s)$ by a rotation of $\pi/2$. $\{\boldsymbol{t}(s), \boldsymbol{n}(s)\}$ is the moving frame of the curve.

Definition 4.6. (Nondegenerate [Clelland, 2017]). A smooth, parametrized curve $\boldsymbol{x}$: $I \to \mathbb{E}^2$ that maps some open interval $I \subset \mathbb{R}$ into Euclidean space is regular if $\boldsymbol{x}'(s) \neq \mathbf{0}$ for every $s \in I$. $\mathbf{x}$ is nondegenerate if it is regular and $\mathbf{e}_1'(s) \neq \mathbf{0}$ for all $s \in I$.

Frenet frame. The Frenet frame of a nondegenerate $\mathcal{C}^3$ curve in Euclidean space may be the earliest example of a moving frame analyzing properties of the curve that are invariant under Euclidean motion. It consists of a triple of orthonormal vectors ($\mathbf{t}$, $\mathbf{n}$, $\mathbf{b}$) based at each point of the curve described (Frenet, 1852). In the Euclidean plane, a moving frame defines a Cartesian system of coordinates for any point on the curve such that $\mathbf{x}(s)$ becomes the origin, and the x_1 axis is tangent to the curve in the direction of increasing arc length. A moving frame is obtained from the fixed one by a rotation about θ between the direction of $\boldsymbol{e}_1$ and $\mathbf{t}$:

$$\{\mathbf{t}, \mathbf{n}\} = A(s)\{\mathbf{e}_1, \mathbf{e}_2\}, \text{where } A(s) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The variation of the moving frame along the curve is:

DOI: 10.1002/asi

$$\frac{d}{ds}\{\mathbf{t},\mathbf{n}\} = A'\{\mathbf{e}_1,\mathbf{e}_2\} = A'A^{-1}\{\mathbf{t},\mathbf{n}\} = \begin{pmatrix} 0 & -\frac{d\theta}{ds} \\ \frac{d\theta}{ds} & 0 \end{pmatrix}\{\mathbf{t},\mathbf{n}\}.$$

Definition. 4.7. (Frenet frame curvature). $\kappa_F(s) = \frac{d\theta}{ds}$ is the Frenet frame curvature of a nondegenerate parametrized curve $\mathbf{x}(s) \in \mathcal{C}^3$ in Euclidean space.

This gives the Frenet equation of plane differential geometry:

$$\frac{d\mathbf{t}}{ds} = \kappa_F(s)\mathbf{n}$$
$$\frac{d\mathbf{n}}{ds} = -\kappa_F(s)\mathbf{t}$$

Since $\kappa(s) \equiv \frac{d\theta}{ds}$, the changing of angle from a point s_0 to a point s_1 on the curve $\mathbf{x}$ is

$$\Delta\theta = \int_{s_0}^{s_1} \kappa(s)ds.$$

Therefore, the change in angle θ from a point s_0 to a point s_1 on the curve $\mathbf{x}$ is the integrated curvature to a domain $\mathfrak{D} = [s_0, s_1]$ of the curve corresponding to the integral of curvature over that domain. Dividing by the length $|\mathfrak{D}|$ of the domain of the integration recovers an integrated curvature to the pointwise curvature.

Definition. 4.8. (Discretized curvature from Frenet frame). Discretize a nondegenerate parametrized curve $\gamma(s) \in \mathcal{C}^3$ by piecewise linear segments $\mathbf{x}(s)$ with vertices $\mathbf{x}_i$. The discretized curvature from the Frenet frame at $\mathbf{x}_i$ is:

$$\kappa_F(s) = \frac{\Delta\theta}{l_l + l_r}, \tag{8}$$

for some point $\mathbf{x}_i^l, \mathbf{x}_i^r \in \mathbf{x}(s)$ on different sides of $\mathbf{x}_i$, and l_l, l_r are the distance from $\mathbf{x}_i^l, \mathbf{x}_i^r$ to $\mathbf{x}_i$ on $\mathbf{x}(s)$.

Bishop frame. The Bishop frame $\{\mathbf{t}(s), \mathbf{y}(s), \mathbf{v}(s)\}$ is another way to frame a curve. The evolution of the Bishop frame can be described in terms of the *Darboux vector* Ω(s):

$$\mathbf{t}^{i+1}(s) = \Omega \times \mathbf{t}^i(s), \mathbf{u}^{i+1}(s) = \Omega \times \mathbf{u}^i(s), \mathbf{v}^{i+1}(s) = \Omega \times \mathbf{v}^i(s).$$

The Darboux vector defines *parallel transport.* Parallel transporting a vector $\mathbf{x}$ corresponds to a rotation about the binormal that keeps the tangential component of $\mathbf{x}$ tangential. We denote by $\mathbf{x}_i$ the vertices of a discretized curve $\mathbf{x}$ and the piecewise linear segments between $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ as $\boldsymbol{e}^i$ with $\mathbf{t}^i = \boldsymbol{e}^i/|\boldsymbol{e}^i|$ being the *unit tangent* vector for edge $\boldsymbol{e}^i$, depicted in Figure 5.

The discrete parallel transport P_i is defined as:

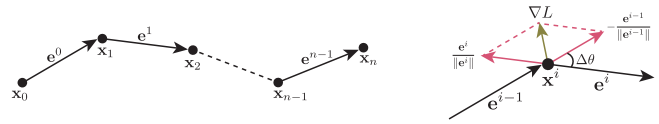

FIG. 5. Interpreting curvature in terms of a normal vector. [Color figure can be viewed at wileyonlinelibrary.com]

$$P_i(\mathbf{t}^{i-1}) = \mathbf{t}^i, P_i(\mathbf{t}^{i-1} \times \mathbf{t}^i) = \mathbf{t}^{i-1} \times \mathbf{t}^i,$$

where P_i is the identity if $\mathbf{t}^{i-1} = \mathbf{t}^i$, and P_i is undefined if $\mathbf{t}^{i-1} = -\mathbf{t}^i$. Consider the variation of two consecutive tangents $\mathbf{t}^{i-1}$ and $\mathbf{t}^i$ with variation parameter ϵ. Define the parallel transport from $\mathbf{t}^i(0)$ to the variation of a tangent vector $\mathbf{t}^i(\epsilon)$ by $P^i(\epsilon)$, where $P^i(0)$ is the identity and

$$P^i(\epsilon)(\mathbf{t}^i(0)) = \mathbf{t}^i(\epsilon), P^i(\epsilon)(\mathbf{t}^{i-1}(0) \times \mathbf{t}^i(\epsilon)) = \mathbf{t}^i(0) \times \mathbf{t}^i(\epsilon).$$

Define the concatenation of parallel transports by:

$$R^{i-1}(\epsilon) = P^{i-1}(\epsilon)^{-1} \circ P_i(\epsilon)^{-1} \circ P^i(\epsilon) \circ P_i(0).$$

The holonomy of the connection induced by parallel transport R^{i-1} is the change in angle $\psi_i(\epsilon)$. By the Ambrose–Singer theorem, the curvature gives the infinitesimal holonomy over the infinitesimal parallelogram. Therefore, according to the *writhe* of polygonal curves (de Vries, 2005), the variation of *holonomy* ψ_i is:

$$\delta\psi_i = \frac{-2\mathbf{t}_{i-1} \times \mathbf{t}_i}{1 + \mathbf{t}_{i-1} \cdot \mathbf{t}_i} \cdot \left(\frac{1}{2}\frac{\mathbf{x}_i - \mathbf{x}_{i-1}}{|\mathbf{e}^{i-1}|} + \frac{1}{2}\frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{|\mathbf{e}^i|} \right),$$

where $\frac{2\mathbf{t}_{i-1} \times \mathbf{t}_i}{1 + \mathbf{t}_{i-1} \cdot \mathbf{t}_i}$ is the curvature binormal $(\kappa\mathbf{b})_i$, and the second factor is the forward difference approximation of $\mathbf{x}_i$ (Bergou, Wardetzky, Robinson, Audoly, & Grinspun, 2008). Since $\mathbf{t}_i$ is a unit vector,

$$\kappa_i = \left| \frac{2\mathbf{t}_{i-1} \times \mathbf{t}_i}{1 + \mathbf{t}_{i-1} \cdot \mathbf{t}_i} \right| = \frac{2\sin\theta_i}{1 + \cos\theta_i} = 2\tan\frac{\theta_i}{2}, \tag{9}$$

according to the half-angle formula. Since this curvature discretization is induced by the Bishop frame, we denote by κ_B.

Curvature Normal

A physical interpretation of curvature is to view it as the magnitude of the rate of change in unit tangent vector $\mathbf{T}$.

Theorem 4.9. *Denote the unit tangent vector of a curve $\gamma(s)$ parametrized by arc-length s as $\mathbf{t}(s)$. Denote by $\Delta\theta$ the angle between $\mathbf{t}(s + \Delta s)$ and $\mathbf{t}(s)$; then,*

$$\kappa = \lim_{\Delta s \to 0} \left| \frac{\Delta\theta}{\Delta s} \right| = |\dot{\mathbf{t}}(s)|.$$

Proof. As Figure 6 shows, since $\mathbf{t}(s)$ is a unit vector,

DOI: 10.1002/asi

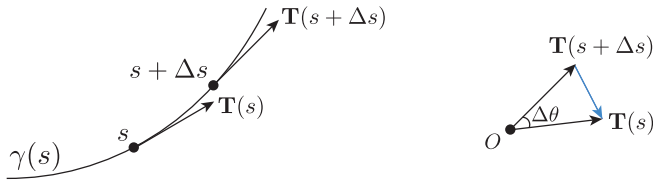

FIG. 6. The physical interpretation of the curvature. [Color figure can be viewed at wileyonlinelibrary.com]

$$|\,\mathbf{t}(s+\Delta s)-\mathbf{t}(s)\,| = 2\left|\sin\frac{\Delta\theta}{2}\right| = \kappa_N. \tag{10}$$

Therefore,

$$|\,\dot{\mathbf{t}}(s)\,| = \lim_{\Delta s\to 0}\frac{|\,\mathbf{t}(s+\Delta s)-\mathbf{t}(s)\,|}{|\,\Delta s\,|} = \lim_{\Delta s\to 0}\frac{|\,\Delta\theta\,|}{|\,\Delta s\,|} = \kappa. \qquad \square$$

Curvature also corresponds to the steepest descent of the length of a discretized curve $\mathbf{x}$ at $\mathbf{x}_i$.

Corollary 4.10. *The arc-length function L: $\mathbb{R}^2 \to \mathbb{R}$ with the displacement of $\mathbf{x}_i \in \mathbb{R}^2$ varies but $\mathbf{x}_k$ fixed for $k \neq i$ is defined as:*

$$\mathbf{x}_i \mapsto \sum_{k=0}^{n-1}\left\|\mathbf{e}^k\right\|,$$

where $\mathbf{e}_{i-1}$ and $\mathbf{e}_i$ depend on $\mathbf{x}_i$. Then $\nabla L = \kappa\mathbf{n}$.

Proof. As shown in Figure 5,

$$\nabla L = \frac{\mathbf{e}^{i-1}}{\|\mathbf{e}^{i-1}\|} - \frac{\mathbf{e}^{i}}{\|\mathbf{e}^{i}\|} = 2\cos\frac{\pi-\Delta\theta}{2}\mathbf{n} = 2\sin\frac{\Delta\theta}{2}\mathbf{n} = \kappa\mathbf{n}. \qquad \square$$

Theorem 4.11. (Analytical Equivalence of Discretized Curvatures). *Assuming the point cloud of eye-fixation sequences and EDA graphs is an approximation of smooth curves, we have the analytical equivalence of discretized curvatures:*

$$\kappa_F = \kappa_B = \kappa_N.$$

Proof. According to Equations 8, 9, and 10,

$$\kappa_F(s) = \frac{\Delta\theta}{l_l + l_r},\ \kappa_B = 2\tan\frac{\theta_i}{2},\ \kappa_N = 2\left|\sin\frac{\Delta\theta}{2}\right|.$$

By the convergence of trigonometry identities, we have:

$$\kappa = \lim_{l_l+l_r\to 0}\frac{\Delta\theta}{l_l+l_r} = \lim_{\Delta s\to 0}\frac{2\tan\Delta\theta/2}{\Delta s} = \lim_{\Delta s\to 0}\frac{2\sin\Delta\theta/2}{\Delta s};$$

hence the analytical equivalence of discretized curvatures in Equations 8–10. $\square$

Finite Metric Space Characterization

EDA signals can be regarded as a function of time; both EDA graphs and eye movement trajectories can be viewed as space curves. Since slope captures the direction and the steepness of the line and curvature is the amount by which a curve deviates away from being straight, they are natural characterizations of the EDA graphs and trajectories of eye movement as finite metric spaces. We first considered the descriptive features for the EDA graph and eye-tracking data and then combined their geometric characterizations to form a feature space. The multiscale model was utilized to compute the discretized geometric invariants in the presence of periodic sampling noise. The satisfaction was dichotomized with dimensions grouped as 1 and 2 totaling 693 instances of being not satisfied, and 3 and 4 totaling 852 instances of being satisfied. In this section, we report the *p*-values and odds ratios of features proposed previously predicting dichotomized satisfaction using the generalized linear mixed model with the binomial family.

Features Extracted From EDA Levels

In our experiment, a sequence of EDA levels as a discretization of the smooth EDA graph was collected for each query. The data were collected every 0.005 seconds, and we downsized the data by averaging every 100 data points. For each query, we first computed the descriptive features, including the minimum, maximum, average, median, variance, and SD of the EDA levels of each query. Then we took the height of the downsized EDA graph and computed the slope according to Equation 6 based on the centered finite difference method. Then we computed the slope at each vertex and formed a list with the absolute values of the slope, a second list with the positive slopes, and a third list with the negative slopes. We computed the curvature at each vertex and similarly created three lists that contained the absolute values of curvatures, positive curvatures, and negative curvatures using Equation 7 based on the centered finite difference method. We then computed the curvatures induced by the Frenet frame, Bishop frame, and curvature normal using Equations 8–10. For each characterization of EDA levels, we computed the minimum, maximum, mean, median, variance, and SD to be features of the EDA graph corresponding to each query. The classifiers were trained by all of the aforementioned features to predict user satisfaction, and here, we report a sample selection of features in Table 2 with *p*-values less than 0.001.

The odds ratio reflects the factor by which the odds of being satisfied versus unsatisfied change given a 1-unit increase in the given predictor. We find that the *minimum EDA level*, the *minimum negative slope* using the finite difference method, the *minimum pointwise curvature* using the Frenet frame, and the *minimum integrated curvature* using curvature normal are negatively associated with satisfaction. If their value increases by 1 unit, the odds of being satisfied will decrease by 19.7%, 30.3%, 19.1%, and 20.3%, respectively. Typical EDA graphs for search queries about which

DOI: 10.1002/asi

TABLE 2. Sample feature significance and odds ratio for EDA graph.

EDA Feature	Odds ratio
Min EDA level	0.803***
EDA level SD	1.376***
Max absolute slope by finite difference	1.392***
Max positive slope by finite difference	1.338***
Min negative slope by finite difference	0.697***
Slope SD by finite difference	1.264***
Max absolute curvature by finite difference	1.425***
Min pointwise curvature by Frenet frame	0.809***
Max pointwise curvature by Bishop frame	1.208***
Max integrated curvature by Frenet frame	1.218***
Min integrated curvature by curvature normal	0.797***

***$p < .001$.

users reported feeling satisfied and dissatisfied are illustrated in Figure 4.

The EDA graph of a query about which users reported being dissatisfied displays this tendency accordingly. For example, the mixed model indicates that the *minimum negative slope* negatively correlates with satisfaction. The increase in *minimum EDA levels* indicates an increase in EDA levels for a given period of time. Since lower EDA levels imply pleasure and increased EDA levels imply displeasure (Stemmler, 1989), this explains that the increase in *minimum EDA levels* is associated with a lower chance of being satisfied in our experiment. Similarly, the increase in the *minimum negative slope* means that the EDA graph drops more slowly, implying that it enters the lower EDA range more slowly. Hence, the change reflects a negative association with satisfaction.

The curvature is the magnitude of the rate of change in the unit tangent vector of a curve. If the *minimum curvature* increases, it implies that the EDA graph is becoming "less flat"; that is, the graph increases or decreases more rapidly. Since the EDA increases during anger (Stemmler, 1989), if a curvature increases when the graph increases, it may reflect that a user is becoming "impatient" in the context of a search experience. A larger curvature can also imply that the graph decreases more rapidly. Since a low EDA level signifies "pleasure," an event that triggers a "sudden pleasure" is less likely. The interpretation for features with odds ratios greater than 1 is similar; for example, the *maximum pointwise curvature* by Bishop frame has an odds ratio of 1.208, implying that if the value increases by one unit, the odds of being satisfied will increase by 20.8%.

Features Extracted From Eye-Tracking

We extracted the eye movement recordings when a user was viewing SERPs to form ordered fixation points on the 2D Euclidean space. We collected the fixation duration and location, and then we computed the geometric properties similar to those of EDA levels, including the saccadic direction, the slope between two fixations (which is the tangent of the saccadic direction), and the saccadic curvature. The detected fixation data are a sequence of fixation transitions detected at the sampling rate of 30 Hz. From each fixation point to the next, we computed the travel time, which is the time difference between these two detected fixations, and the perceptual span, which is the spacing of fixations in the reading sequence from one fixation point to another. For the fixation sequence associated with each search query, we computed the maximum, minimum, average, median, variance, and SD of the travel time and perception span between adjacent fixations. We also extracted features from the graph of perceptual span over time using the same method as we adopted for analyzing EDA graphs in the previous section.

We report some of the significant features in Table 3. The SD of the x coordinate has an odds ratio of 1.207, implying that if the SD of the x coordinate increases by 1 unit, the odds of being satisfied versus unsatisfied will increase by 20.7%. This implies that if the fixation sequence is more stretched along the horizontal direction, the user tends to be more satisfied with the content being viewed. Gwizdka (2014) examined the effects of relevance on reading versus perceptual span and found that the length of reading fixations was longer on relevant documents. Since document relevance is associated with users' search satisfaction (Gluck, 1996; Huffman & Hochster, 2007), the significance between perceptual span and satisfaction reported in Table 3 agrees with Gwizdka's result.

In our study, we found that a 1-unit increase in *average travel time* increases the odds of being satisfied by 353.1%, which agrees with prior work that a longer duration of reading implies that a document is topical or relevant (Gwizdka, 2014). A case study from one of our participants is presented in Figure 7. The graphs suggest that a satisfied query tends to have a larger perceptual span, agreeing with its coefficients in the mixed model (odds ratio = 1.197, $p < .001$); similarly, the satisfied query has more spanning eye movements and hence corresponds to larger saccadic curvature.

Features Extracted From Search Interactions

We consider coarse-grained features (Ageev et al., 2011; Feild et al., 2010; Guo et al., 2011; Hassan et al., 2010) and fine-grained features (Guo et al., 2012), and we report some of the significant features in Table 4. We find that the *average*

TABLE 3. Sample feature significance and odds ratio from fixation.

Eye-fixation feature	Odds ratio
Maximum *perceptual span*	1.197***
x coordinate SD	1.207***
Median of y coordinate	1.149**
(x, time) minimum curvature by finite difference	0.899*
Median of *slope* by finite difference	1.206**
Average travel time	4.531***
Max integrated curvature by Frenet frame	1.179**
Integrated curvature SD by Frenet frame	1.174**
Max integrated abs(curvature) by Frenet frame	1.201**
Max integrated curvature by Bishop frame	1.176**
Integrated curvature SD by Bishop frame	1.189**

*$p < .05$; **$p < .01$; ***$p < .001$.

DOI: 10.1002/asi

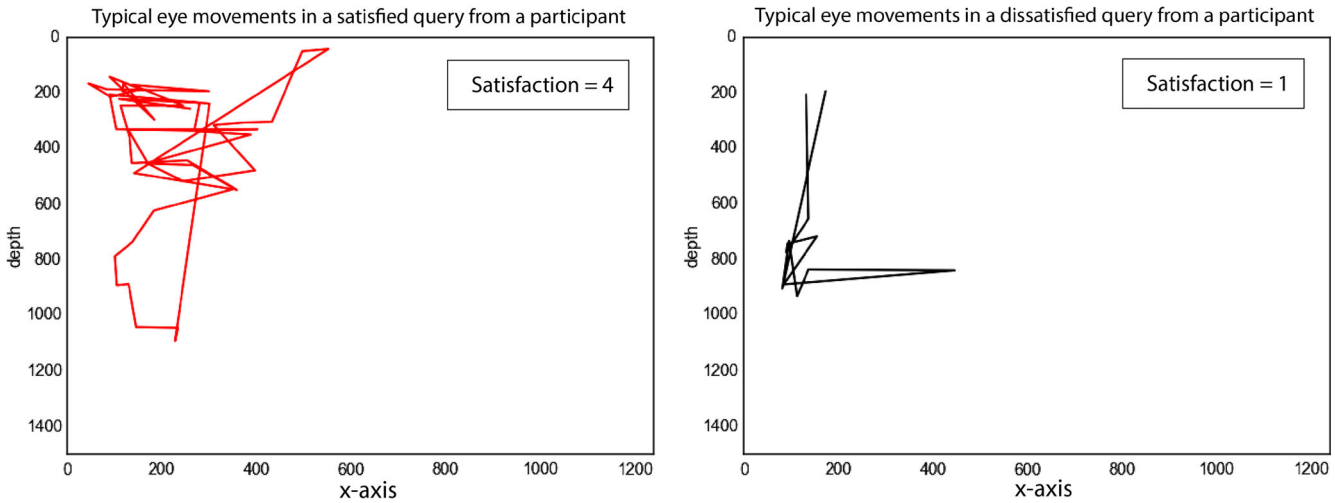

FIG. 7. Typical eye movements in satisfied and dissatisfied queries. [Color figure can be viewed at wileyonlinelibrary.com]

number of clicks in the session and the *average clicks* over the number of queries was positively associated with search satisfaction, agreeing with Ageev's results on search success (Ageev et al., 2011). In our study, the *ratio of clicks with a dwell time* ≥ 30 and the *median of the max y coordinates* were positively associated with success ratings, agreeing with Guo's findings (Guo et al., 2012). However, Guo found that the *average max y coordinates*, the *number of max y coordinates* < 400, and the *number of max y coordinates* < 800 *pixels* on SERPs were negatively related to search satisfaction, while we found that they were positively associated with search satisfaction. Guo and Feild also reported that the number of clicks was negatively associated with search success (Feild et al., 2010; Guo et al., 2012), but we found that the relation of number of clicks to search satisfaction was positive. We suspect that these differences might be due to the current users' tendency to abandon an unsuccessful retrieval; therefore, clicks and deep SERP *y*-coordinates reflect a satisfactory SERP, according to our study.

Experimental Results

With the above features, we use the gradient boosting decision tree (GBDT), random forest (RF), and support vector machine (SVM) with 5-fold cross-validation to predict user satisfaction using features extracted from eye-fixation sequences, EDA graphs, and behavioral information collected from user logs. We used classifiers from the scikit-learn package (Pedregosa et al., 2011). The experimental results on dichotomized query satisfaction are reported in Table 5, and Table 6 reports the mean squared error (MSE) predicting query satisfaction over a 4-point scale. We discuss the prediction results in terms of the area under the ROC curve (AUC), but the results of other evaluation metrics are consistent with those of AUC. In our extensive cross-validation experiments, we exhibited the robustness of our method: according to Tables 5 and 6, geometric invariants successfully predict user satisfaction regardless of gender, age, ethnicity, educational background, and work experience. Tables 7 and 8 show the cross-user and cross-task prediction results, suggesting that the proposed geometric methods are effective in the presence of confounding factors such as user and task. The cross-user validation has an average AUC of 0.707, and the cross-task validation achieves an average AUC of 0.712. In comparison, an average AUC of 0.756 is reported for standard cross-validation in Table 5.

TABLE 4. Sample feature significance and odds ratio from behavior.

Coarse-grained features	Odds ratio
Number of clicks in the session	2.509^{***}
Number of clicks over number of queries	2.869^{***}
Average dwell time on clicked landing pages	2.586^{***}
Ratio of clicks with dwell time ≥30 sec	2.155^{***}
Fine-grained features	
Average max *y* coordinates on SERP pages	1.236^{***}
Median of max *y* coordinates on SERP pages	1.142^{**}
Number of max *y* coordinates <400 pixels	1.120^{*}
Number of max *y* coordinates <800 pixels	1.119^{*}

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Comparison With Traditional Methods

We summarized coarse-grained features from several works (Ageev et al., 2011; Feild et al., 2010; Guo et al., 2011; Hassan et al., 2010), fine-grained features (Guo et al., 2012), and EDA features (Wu et al., 2017) as baseline methods. For NP signals, we first used descriptive features of EDA, which are the minimum, maximum, average, median, variance, and SD of the EDA levels of each query. Then we used descriptive features of eye movements, which are the minimum, maximum, average, median, variance, and SD of the *x*-axis, *y*-axis, and fixation time of each query; we then used the descriptive features of EDA and those of eye-tracking combined.

DOI: 10.1002/asi

TABLE 5. The AUC, precision (P), recall (R), F1-measure (F), and accuracy score (A) predicting dichotomized satisfaction in query level. Five-fold cross-validation average (*macro*: unweighted mean for each label; w*eighted*: weighted mean for each label).

Features	CLF	AUC	P_{macro}	P_{weight}	R_{macro}	R_{weight}	F_{macro}	F_{weight}	A
E[1]	GBDT	0.693	0.678	0.652	0.674	0.652	0.622	0.630	0.652
	RF	0.663	0.623	0.625	0.622	0.623	0.620	0.622	0.623
	SVM	0.679	0.659	0.658	0.640	0.651	0.636	0.641	0.651
E^{des}	GBDT	0.656	0.651	0.648	0.610	0.627	0.587	0.596	0.627
	RF	0.629	0.584	0.587	0.583	0.582	0.581	0.581	0.582
	SVM	0.640	0.645	0.642	0.602	0.620	0.575	0.584	0.620
I^{des}	GBDT	0.600	0.555	0.555	0.538	0.559	0.495	0.507	0.559
	RF	0.523	0.525	0.527	0.525	0.525	0.523	0.525	0.525
	SVM	0.556	0.539	0.540	0.532	0.547	0.517	0.525	0.547
E^{des}+I^{des}	GBDT	0.665	0.664	0.661	0.616	0.635	0.592	0.601	0.635
	RF	0.603	0.573	0.575	0.571	0.570	0.568	0.569	0.570
	SVM	0.602	0.575	0.577	0.571	0.577	0.568	0.572	0.577
Coarse[2]	GBDT	0.806	0.791	0.784	0.723	0.740	0.716	0.722	0.740
	RF	0.759	0.688	0.690	0.686	0.687	0.685	0.686	0.687
	SVM	0.793	0.758	0.754	0.727	0.738	0.725	0.729	0.738
Fine[3]	GBDT	0.561	0.627	0.622	0.538	0.569	0.448	0.467	0.569
	RF	0.545	0.537	0.540	0.537	0.537	0.535	0.537	0.537
	SVM	0.567	0.568	0.568	0.550	0.569	0.524	0.534	0.569
Majority		0.500	0.268	0.290	0.500	0.536	0.348	0.375	0.536
E	GBDT	0.697	0.685	0.681	0.641	0.657	0.627	0.634	0.657
	RF	0.672	0.626	0.629	0.624	0.625	0.622	0.623	0.625
	SVM	0.696	0.659	0.658	0.641	0.650	0.636	0.641	0.650
I	GBDT	0.747	0.713	0.714	0.709	0.713	0.709	0.711	0.713
	RF	0.688	0.647	0.650	0.646	0.645	0.644	0.645	0.645
	SVM	0.671	0.628	0.628	0.617	0.625	0.615	0.619	0.625
E + I	GBDT	0.756	0.708	0.708	0.701	0.707	0.702	0.705	0.707
	RF	0.692	0.649	0.651	0.649	0.649	0.648	0.649	0.649
	SVM	0.695	0.636	0.636	0.624	0.633	0.621	0.626	0.633
E + B	GBDT	0.808	0.791	0.783	0.724	0.740	0.716	0.722	0.740
	RF	0.765	0.688	0.690	0.686	0.688	0.685	0.687	0.688
	SVM	0.764	0.708	0.708	0.700	0.705	0.699	0.702	0.705
I + B	GBDT	0.810	0.789	0.782	0.723	0.739	0.716	0.721	0.739
	RF	0.763	0.695	0.697	0.693	0.694	0.692	0.693	0.694
	SVM	0.741	0.674	0.676	0.673	0.676	0.673	0.676	0.676
E + B + I	GBDT	0.810	0.789	0.782	0.723	0.739	0.716	0.721	0.739
	RF	0.749	0.688	0.690	0.687	0.688	0.686	0.687	0.688
	SVM	0.739	0.685	0.685	0.679	0.684	0.678	0.681	0.684

Note. [1](Wu et al., 2017), [2](Ageev et al., 2011; Feild et al., 2010; Guo et al., 2011; Hassan et al., 2010), [3](Guo et al., 2012). E = EDA, I = eye-tracking, B = behavior (Coarse[2] + Fine[3]), E^{des} = descriptive features for EDA, I^{des} = descriptive features for eye-tracking; GBDT = gradient boosting classifier, RF = random forest classifier, SVM = support vector machine.

The forward difference method was employed to extract geometric features including slope and curvature from the EDA graphs and achieved an AUC of 0.62 predicting search satisfaction for mobile shopping (Wu et al., 2017). In our study, we first downsized the sampled data using the multiscale framework and then used the centered difference to compute the slope and curvature of the EDA graph, along with three other formulae derived from the method of moving frames to compute the discrete curvature. According to the results reported in Table 5, with the GBDT classifier, we obtained an AUC of 0.656 predicting search satisfaction using descriptive features of EDA data; we obtained an AUC of 0.600 using the descriptive features of eye-tracking data; we received an AUC of 0.665 using these two combined; we obtained an AUC of 0.693 using Wu's method (Wu et al., 2017); we obtained an AUC of 0.806 using coarse-grained features collected from several works (Ageev et al., 2011; Feild et al., 2010; Guo et al., 2011; Hassan et al., 2010); and we obtained an AUC of 0.561 in our experiment, using fine-grained features (Guo et al., 2012).

Predicting Satisfaction With EDA Signals

We first investigated the effectiveness of features extracted from physiological signals. With a GBDT classifier, we first noticed that the feature space that includes the geometric invariants achieves an AUC score of 0.697, which is higher than that using descriptive features alone (AUC = 0.656). This suggests the effectiveness of geometric techniques. We also found that combining EDA and eye-tracking features obtained a higher AUC value at 0.756, showing that eye movement data complement EDA signals. Combining the descriptive features of EDA and

DOI: 10.1002/asi

TABLE 6. The MSE predicting four-scale satisfaction in query level. Five-fold cross-validation.

Classifiers	E[1]	Coarse[2]	Fine[3]	E	I	E + I	E + B	I + B	E + B + I
GBDT	1.273	1.003	1.473	1.263	1.246	1.157	0.988	1.027	1.001
RF	1.372	1.148	1.753	1.355	1.308	1.253	1.052	1.110	1.048
SVM	1.309	1.035	1.504	1.272	1.414	1.342	1.141	1.252	1.250

Note. [1](Wu et al., 2017), [2](Ageev et al., 2011; Feild et al., 2010; Guo et al., 2011; Hassan et al., 2010), [3](Guo et al., 2012). E = EDA, I = eye-tracking, B = behavior (Coarse[2] + Fine[3]), GBDT = gradient boosting classifier, RF = random forest classifier, SVM = support vector machine.

descriptive features of the eye-tracking data yielded an AUC value of 0.665. This observation showed that comprehensive eye movement features boost comprehensive EDA features more effectively (from 0.697 to 0.756) than descriptive eye movement features boost descriptive EDA features (from 0.656 to 0.665).

Predicting Satisfaction With Eye Movement

We found that features extracted based on geometric invariants improve the performance of classifiers. With eye movement data, predicting user satisfaction using a feature space that includes the geometric invariants has a much higher score (AUC = 0.747) than using the descriptive features alone (AUC = 0.600). According to our study, eye-tracking data predicted user satisfaction better than EDA data (AUC = 0.697), combining eye-tracking data with behavioral data resulted in an AUC of 0.810, and combining EDA data with behavioral data obtained an AUC of 0.808. Both values are greater than those obtained using behavioral data alone, since the AUC is 0.561 for the fine-grained feature group and 0.806 for the coarse-grained feature group. This phenomenon implies that EDA data and eye movements can boost user behavior data.

Cross-User Validation

Tables 5 and 6 report the results obtained by the GBDT classifier that is trained without user or task information in the feature space, suggesting the robustness of the features against user and task. To further validate the robustness of geometric invariants, we performed a cross-user validation predicting query-level satisfaction. The classifier was trained with the EDA and eye movement data of 37 participants and was tested with the remaining two participants disjoint from the 37 participants previously chosen to be tested. Since there were 39 participants, this procedure allowed 19 cross-user validations; the results are shown in Table 7. The performances of the prediction slightly decreased, with an average AUC of 0.707 ($\sigma = 0.057$), showing the robustness of the proposed method based on geometric analysis against individual differences.

Cross-Task Validation

This section presents the cross-task results predicting query satisfaction. The classifier was trained with EDA and eye-tracking data from 11 tasks and was tested on the remaining task, as shown in Table 8. The performances appeared to be effective, with an average AUC

TABLE 7. The AUC, precision, recall, F1-measure, and accuracy score predicting dichotomized satisfaction in query level. Cross-validation over users (*macro*: unweighted mean for each label; w*eighted*: weighted mean for each label).

User group	AUC	P_{macro}	P_{weight}	R_{macro}	R_{weight}	F_{macro}	F_{weight}	Accuracy
1	0.676	0.838	0.814	0.676	0.725	0.664	0.686	0.725
2	0.746	0.749	0.758	0.746	0.736	0.735	0.735	0.736
3	0.651	0.722	0.730	0.651	0.635	0.609	0.604	0.635
4	0.646	0.830	0.803	0.646	0.702	0.623	0.651	0.702
5	0.675	0.760	0.752	0.675	0.693	0.658	0.666	0.693
6	0.663	0.731	0.803	0.663	0.574	0.562	0.543	0.574
7	0.671	0.753	0.745	0.671	0.689	0.654	0.661	0.689
8	0.757	0.833	0.836	0.757	0.754	0.740	0.739	0.754
9	0.793	0.863	0.857	0.793	0.802	0.790	0.792	0.802
10	0.671	0.789	0.800	0.671	0.654	0.621	0.615	0.654
11	0.622	0.794	0.786	0.622	0.636	0.566	0.573	0.636
12	0.717	0.822	0.804	0.717	0.762	0.721	0.740	0.762
13	0.703	0.770	0.760	0.703	0.737	0.705	0.719	0.737
14	0.746	0.801	0.792	0.746	0.771	0.750	0.760	0.771
15	0.801	0.813	0.836	0.801	0.775	0.774	0.773	0.775
16	0.629	0.700	0.716	0.629	0.600	0.573	0.562	0.600
17	0.725	0.864	0.838	0.725	0.778	0.732	0.753	0.778
18	0.752	0.752	0.768	0.752	0.735	0.735	0.735	0.735
19	0.794	0.839	0.830	0.794	0.815	0.802	0.809	0.815
Mean	0.707	0.791	0.791	0.707	0.714	0.685	0.690	0.714
SD	0.057	0.049	0.040	0.057	0.068	0.077	0.081	0.068

DOI: 10.1002/asi

TABLE 8. The AUC, precision, recall, F1-measure, and accuracy score predicting dichotomized satisfaction in query level. Cross-validation over tasks (*macro*: unweighted mean for each label; w*eighted*: weighted mean for each label).

Task id	AUC	P_{macro}	P_{weight}	R_{macro}	R_{weight}	F_{macro}	F_{weight}	Accuracy
1	0.792	0.793	0.793	0.792	0.792	0.792	0.792	0.792
2	0.714	0.760	0.816	0.714	0.647	0.642	0.632	0.647
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.593	0.660	0.655	0.593	0.625	0.564	0.580	0.625
5	0.679	0.775	0.802	0.679	0.640	0.618	0.607	0.640
6	0.630	0.815	0.794	0.630	0.673	0.594	0.614	0.673
7	0.700	0.812	0.812	0.700	0.700	0.670	0.670	0.700
8	0.765	0.846	0.842	0.765	0.771	0.755	0.757	0.771
9	0.630	0.744	0.788	0.630	0.565	0.533	0.512	0.565
10	0.655	0.706	0.808	0.655	0.535	0.529	0.509	0.535
11	0.683	0.782	0.769	0.683	0.750	0.691	0.725	0.750
12	0.702	0.735	0.733	0.702	0.706	0.694	0.695	0.706
Mean	0.712	0.786	0.801	0.712	0.700	0.674	0.674	0.700
SD	0.107	0.084	0.079	0.107	0.122	0.132	0.136	0.122

of 0.712 ($\sigma = 0.107$). We notice that the variation of AUCs from the cross-task validation is slightly larger than that of cross-user validation ($\sigma = 0.057$). We attribute this to the inhomogeneity of task difficulties, which can be inferred from Figure 3. Nevertheless, the evaluation metrics for the prediction performances are sufficiently promising, showing the robustness of the proposed method against task differences.

Bootstrapping Linear Models

Bootstrap samples ($B = 1,\ 000$), in accordance with Efron (1987), were additionally drawn to compare linear models fitted by different feature groups. For each replicate, we sampled 1,590 data points with replacement from our data set. Adjusted R-squared was observed to be significantly different among different feature groups in the analysis. First, the analysis indicated a significant difference in adjusted R-squared between the linear model fitted by behavioral features and the full feature group extracted from behavioral data, eye movement data, and EDA signals (95% bootstrap confidence interval [CI]: [−0.1919, −0.1186]). This provides important insight into our study in that the NP data and eye-tracking data provide information not conveyed in user behavior features. Second, the analysis indicated a significant difference in adjusted R-squared between the linear model fitted by the feature group extracted from EDA signals and the full feature group (95% bootstrap CI: [−0.2155, −0.1372]). Third, the analysis indicated a significant difference between the linear model fitted by the feature group extracted from eye movements and the full feature group (95% bootstrap CI: [−0.1717, −0.1002]). Furthermore, the analysis indicated a significant difference between the eye movements feature group and the feature group combining EDA and eye movements (95% bootstrap CI: [−0.1141, −0.0579]) as well as a significant difference between the EDA feature group and the feature group combining EDA and eye movements (95% bootstrap CI: [−0.1569, −0.0896]).

Conclusion, Limitations, and Future Work

In this article, we adopted the multiscale framework to filter out sampling noise for EDA data, and we proposed discretization schemes to compute geometric invariants for finite metric spaces. We proved that the discretized curvature of a sampled curve downsized by the multiscale framework under the centered finite difference scheme approximates the curvature of the smooth curve, and we computed the order of the error. We proved the analytical equivalence of the proposed discretization schemes. Empirical studies suggest that both eye-tracking data and EDA signals extracted by these invariants are robust regardless of deviations in individual differences, the specific task performed, and environmental stimuli. According to bootstrap analysis comparing linear models fitted by different feature groups, we found that using both eye-tracking and EDA data explains the variability of the query-level search satisfaction significantly better than using either of them alone, and using NP data and eye-tracking data combined with behavioral data explained the variability of the response data significantly better than using the behavioral feature group alone. This implies that the physiological signals and eye movement data complement behavior-based signals in predicting search satisfaction.

This study has several limitations: we believe that the use of NP data can be further utilized; a higher eye-tracker sampling rate could be employed to measure saccades more accurately, such that the effectiveness of predictions could be further improved; and questions related to predicting search satisfaction with other physiological measurements, such as EGG and fMRI, also remain uninvestigated. From the analysis perspective, there are also classical measurements from the literature that can be added as features, such as pupil dilation, curve direction, visual angle, and area of interest analysis. Although we used a remote eye-tracker and collected EDA signals attached to two fingertips to minimize influences induced by data collection, ecological validity issues remain unresolved.

For future work, we will study the effectiveness of physiological methods in deep neural networks to predict

DOI: 10.1002/asi

search satisfaction. We will consider other evaluation metrics, such as document relevance and search effort, to understand the interplay between these metrics and the proposed method. We will also investigate other convolution kernels of the multiscale framework, such as the exponential kernel and the cosine kernel, and we will consider features extracted based on topological invariants.

Acknowledgments

The authors thank Dr. Karen Uhlenbeck, Dr. Andrew J. Blumberg, Dr. Clifford H. Taubes, Dr. Andrew C. Yao, Dr. Björn Engquist, Dr. Kui Ren, Dr. Mu-Tao Wang, Dr. Magdalena Czubak, Dr. Kefeng Liu, Dr. An Huang, Dr. Qiaozhu Mei, Dr. Sean P. Carney, and Dr. Lindsay Martin for discussions on geometric analysis and multiscale frameworks; Dr. Yong-Jin Liu, Dr. Xianfeng Gu, Dr. Shi-Min Hu, and Dr. Zipeng Ye for discussions on computer graphics; Dr. Tetsuya Sakai and Erika Hale for discussions on the statistical models; Dr. Ying-Hsang Liu for discussions on physiological signals; and Dr. Liang Ma for consultation on electrodermal activity. This work was supported by the Natural Science Foundation of China (Grant No. 6162 2208, 61732008, 61532011) and the National Key Basic Research Program (2015CB358700). Wu was supported in part by NIH grant 5U54CA193313, NIH grant GG010211-R01-HIV, and AFOSR research grant FA9550-15-1-0302.

References

Ageev, M., Guo, Q., Lagun, D., & Agichtein, E. (2011). Find it if you can: A game for modeling different types of web search success using interaction data. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 345–354). New York, NY: ACM.

Al-Maskari, A., Sanderson, M., & Clough, P. (2007). The relationship between IR effectiveness measures and user satisfaction. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 773–774). New York, NY: ACM.

Arapakis, I., Bai, X., & Cambazoglu, B. B. (2014). Impact of response latency on user behavior in web search. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 103–112). New York, NY: ACM.

Arapakis, I., & Leiva, L.A. (2016). Predicting user engagement with direct displays using mouse cursor information. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 599–608). New York, NY: ACM.

Ax, A.F. (1953). The physiological differentiation between fear and anger in humans. Psychosomatic Medicine, 15(5), 433–442.

Barreda-Ángeles, M., Arapakis, I., Bai, X., Cambazoglu, B.B., & Pereda-Baños, A. (2015). Unconscious physiological effects of search latency on users and their click behaviour. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 203–212). New York, NY: ACM.

Belkin, N.J. & Vickery, A. (1985). Interaction in information systems: A review of research from document retrieval to knowledge-based systems. (No. 025.04 BEL. CIMMYT.). United Kingdom: Information Processing & Management.

Bergou, M., Wardetzky, M., Robinson, S., Audoly, B., & Grinspun, E. (2008). Discrete elastic rods. In ACM Transactions on Graphics (Vol. 27, p. 63).

Bernstein, D., Penner, L., Clarke-Stewart, A., & Roy, E. (2003). Psychology (9th ed.). Belmont, CA: Wadsworth.

Blumberg, A.J., Gal, I., Mandell, M.A., & Pancia, M. (2014). Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. Foundations of Computational Mathematics, 144, 745–789.

Boucsein, W. (2012). Electrodermal activity. New York: Plenum.

Buscher, G., Cutrell, E., & Morris, M.R. (2009). What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 21–30). New York, NY: ACM.

Cannon, W.B. (1929). Bodily changes in pain, hunger, fear and rage. New York, NY: D. Appleton & Company.

Chen, Y., Liu, Y., Zhang, M., & Ma, S. (2017). User satisfaction prediction with mouse movement information in heterogeneous search environment. IEEE Transactions on Knowledge and Data Engineering, 29(11), 2470–2483.

Chuklin, A. & de Rijke, M. (2016). Incorporating clicks, attention and satisfaction into a search engine result page evaluation model. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (pp. 175–184). New York, NY: ACM.

Clelland, J.N. (2017). From Frenet to Cartan: The method of moving frames. Providence, RI: American Mathematical Society.

Cutrell, E. & Guan, Z. (2007). What are you looking for?: An eye-tracking study of information usage in web search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 407–416). New York, NY: ACM.

de Vries, R. (2005). Evaluating changes of writhe in computer simulations of supercoiled DNA. The Journal of Chemical Physics, 122(6), 064905.

Doberenz, S., Roth, W.T., Wollburg, E., Maslowski, N.I., & Kim, S. (2011). Methodological considerations in ambulatory skin conductance monitoring. International Journal of Psychophysiology, 80(2), 87–95.

Duffy, E. (1972). Activation. In N.S. Greenfield & R.A. Sternbach (Eds.), Handbook of psychophysiology (pp. 577–622). New York: Holt, Rinehart & Winston.

Edwards, A. & Kelly, D. (2017). Engaged or frustrated?: Disambiguating emotional state in search. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan (pp. 125–134). New York, NY: ACM.

Efron, B. (1987). Better bootstrap confidence intervals. Journal of the American Statistical Association, 82(397), 171–185.

Engquist, B., & Tsai, Y.-H. (2005). Heterogeneous multiscale methods for stiff ordinary differential equations. Mathematics of Computation, 74 (252), 1707–1742.

Feild, H.A., Allan, J., & Jones, R. (2010). Predicting searcher frustration. In International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland (pp. 34–41). New York, NY: ACM.

Flanagan, J. (1967). Galvanic skin response: Emotion or attention? In Proceedings of the Annual Convention of the American Psychological Association.

Forbes, T., & Bolles, M.M. (1936). Correlation of the response potentials of the skin with exciting and non-exciting stimuli. The Journal of Psychology, 2(2), 273–285.

Frenet, F. (1852). On curves with double curvature. Journal of Pure and Applied Mathematics, 437–447.

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. Learning and Instruction, 21(2), 220–231.

Gluck, M. (1996). Exploring the relationship between user satisfaction and relevance in information systems. Information Processing & Management, 32(1), 89–104.

Granka, L.A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, United Kingdom (pp. 478–479). New York, NY: ACM.

Gu, X.D., & Yau, S.-T. (2008). Computational conformal geometry. Somerville, MA: International Press.

Guggenheimer, H. (1963). Differential geometry (p. 18). New York: McGraw-Hill.

Guo, Q., Lagun, D., & Agichtein, E. (2012). Predicting web search success with fine-grained interaction data. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management (pp. 2050–2054). New York, NY: ACM.

Guo, Q., White, R.W., Zhang, Y., Anderson, B., & Dumais, S.T. (2011). Why searchers switch: Understanding and predicting engine switching rationales. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 335–344). New York, NY: ACM.

Gwizdka, J. (2014). Characterizing relevance with eye-tracking measures. In Proceedings of the 5th Information Interaction in Context Symposium (pp. 58–67). New York, NY: ACM.

Hassan, A., Jones, R., & Klinkner, K.L. (2010). Beyond DCG: User behavior as a predictor of a successful search. In Proceedings of the Third ACM International Conference on Web Search and Data Mining, New York, NY (pp. 221–230). New York, NY: ACM.

Hitchingham, E.E. (1979). A study of the relationship between the search interview of the intermediary searcher and the online system user, and the assessment of search results as judged by the user (ERIC Report ED 180 478). Rochester, MI: Kresge Library, Oakland University.

Huang, W.-Q., Gu, X.D., Lin, W.-W., & Yau, S.-T. (2016). Method for computing conformal parameterization (US Application US20170212868A1). Retrieved from https://patents.google.com/patent/US20170212868A1

Huang, W.-Q., Huang, T.-M., Lin, W.-W., Lin, S.-S., Gu, X.D., & Yau, S.-T. (2016). Method for computing spherical conformal and Riemann mapping (US Application US20170212867A1). Retrieved from https://patents.google.com/patent/US20170212867A1

Huffman, S.B. & Hochster, M. (2007). How well does result relevance predict session satisfaction? In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 567–574). New York, NY: ACM.

James, W. (1890). The principles of psychology (Vol. 1). New York: Holt.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 87(4), 329–354.

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. Foundations and Trends in Information Retrieval, 3 (12), 1–224.

Lagun, D., Ageev, M., Guo, Q., & Agichtein, E. (2014). Discovering common motifs in cursor movement data for improving web search. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining (pp. 183–192). New York, NY: ACM.

Lazarus, R.S. (1966). Psychological stress and the coping process. New York, NY: McGraw-Hill.

Lazarus, R.S., & Opton, E.M., Jr. (1966). The study of psychological stress: A summary of theoretical formulations and experimental findings. In C.D. Spielberger (Ed.), Anxiety and behavior (pp. 225–262). New York: Academic Press.

Liu, Y., Chen, Y., Tang, J., Sun, J., Zhang, M., Ma, S., & Zhu, X. (2015). Different users, different opinions: Predicting search satisfaction with mouse movement information. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 493–502). New York, NY: ACM.

Liu, Y., Liu, Z., Zhou, K., Wang, M., Luan, H., Wang, C. … Ma, S. (2016). Predicting search user examination with visual saliency. In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 619–628). New York, NY: ACM.

Maltzman, I., & Raskin, D.C. (1965). Effects of individual differences in the orienting reflex on conditioning and complex processes. Journal of Experimental Research in Personality, 1(1), 1–16.

Mao, J., Liu, Y., Zhou, K., Nie, J.-Y., Song, J., Zhang, M. … Luo, H. (2016). When does relevance mean usefulness and user satisfaction in web search? In International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 463–472). New York, NY: ACM.

Moshfeghi, Y., Triantafillou, P., & Pollick, F.E. (2016). Understanding information need: An fMRI study. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 335–344). New York, NY: ACM.

Mostafa, J. & Gwizdka, J. (2016). Deepening the role of the user: Neurophysiological evidence as a basis for studying and improving search. In Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (pp. 63–70). New York, NY: ACM.

Nomikos, M.S., Opton, E., Jr., & Averill, J.R. (1968). Surprise versus suspense in the production of stress reaction. Journal of Personality and Social Psychology, 8(2p1), 204.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. Journal of Machine Learning Research, 12(Oct), 2825–2830.

Price, D.D., Barrell, J.E., & Barrell, J.J. (1985). A quantitative-experiential analysis of human emotions. Motivation and Emotion, 9(1), 19–38.

Prokasy, W. (2012). Electrodermal activity in psychological research. New York, NY: Academic Press.

Quarteroni, A., Sacco, R., & Saleri, F. (2007). Numerical mathematics (2nd ed.). Berlin and Heidelberg: Springer.

Rayner, K., Juhasz, B.J., & Pollatsek, A. (2008). Eye movements during reading. In The science of reading: A handbook (pp. 79–97). Malden, MA: Wiley-Blackwell.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). Psychology of reading. New York: Psychology Press.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review, 69(5), 379–399.

Sheikh, K., Wegdam, M., & Van Sinderen, M. (2007). Middleware support for quality of context in pervasive context-aware systems. In Pervasive Computing and Communications Workshops, 2007 (pp. 461–466). New York: IEEE.

Sokolov, E.N. (1963). Perception and the conditioned reflex. New York: McMillan.

Stemmler, G. (1984). Psychophysiologische Emotionsmuster: ein empirischer und methodologischer Beitrag zur inter-und intraindividuellen Begründbarkeit spezifischer Profile bei Angst, Ärger und Freude. Lang.

Stemmler, G. (1989). The autonomic differentiation of emotions revisited: Convergent and discriminant validation. Psychophysiology, 26(6), 617–632.

Su, L.T. (2003). A comprehensive and systematic model of user evaluation of web search engines: II. An evaluation by undergraduates. Journal of the Association for Information Science and Technology, 54(13), 1193–1223.

Su, N., He, J., Liu, Y., Zhang, M., & Ma, S. (2018). User intent, behaviour, and perceived satisfaction in product search. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA (pp. 547–555). New York, NY: ACM.

Tagliacozzo, R. (1977). Estimating the satisfaction of information users. Bulletin of the Medical Library Association, 65(2), 243–249.

Tatler, B.W., & Vincent, B.T. (2009). The prominence of behavioural biases in eye guidance. Visual Cognition, 17(6–7), 1029–1054.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. The Quarterly Journal of Experimental Psychology, 59(11), 1931–1949.

Uno, T., & Grings, W.W. (1965). Autonomic components of orienting behavior. Psychophysiology, 1(4), 311–321.

Veraguth, O. (1907). Das psycho-galvanische Reflex-Phänomen. European Neurology, 21(5), 387–406.

Von Helmholtz, H. (1867). Handbuch der physiologischen optik (Vol. 9). Leipzig, Germany: Voss.

Wagner, H. (1989). The peripheral physiological differentiation of emotions.

Wilcott, R., Darrow, C., & Siegel, A. (1957). Uniphasic and diphasic wave forms of the skin potential response. Journal of Comparative and Physiological Psychology, 50(3), 217–219.

Wittek, P., Liu, Y.-H., Darányi, S., Gedeon, T., & Lim, I.S. (2016). Risk and ambiguity in information seeking: Eye gaze patterns reveal contextual behavior in dealing with uncertainty. Frontiers in Psychology, 7, 1790.

Wu, C.-T., Su, W.-S., Yueh, M.-H., Lin, W.-W., & Yau, S.-T. (2015). 3D surface morphing method based on conformal parameterization (US

Application US20170186208A1). Retrieved from https://patents.google.com/patent/US20170186208A1

Wu, M.-M., & Liu, Y.-H. (2011). On intermediaries' inquiring minds, elicitation styles, and user satisfaction. Journal of the American Society for Information Science and Technology, 62(12), 2396–2403.

Wu, Y. (2012). The application of conformal mapping to stomatology (Master dissertation). Cambridge, MA: Harvard University.

Wu, Y., Liu, Y., Su, N., Ma, S., & Ou, W. (2017). Predicting online shopping search satisfaction and user behaviors with electrodermal activity. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia (pp. 855–856). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Comm-ittee.

Xie, X., Liu, Y., Wang, X., Wang, M., Wu, Z., Wu, Y., … Ma, S. (2017). Investigating examination behavior of image search users. In International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo (275-284). New York, NY: ACM.

Yau, S.-T. (2014). Review of geometry and analysis. In Selected expository works of Shing-Tung Yau with commentary (Vol. I, p. 554). Somerville, MA: International Press.

Yau, S.-T., Gu, X., & Wang, Y. (2002). Analysis of geometric surfaces by comformal structure (US Application US20060013505A1). Retrieved from https://patents.google.com/patent/US20060013505A1

Yokota, T., Takahashi, T., Kondo, M., & Fujimori, B. (1959). Studies on the diphasic wave form of the galvanic skin reflex. Electroencephalography and Clinical Neurophysiology, 11(4), 687–696.

Zhang, F., Zhou, K., Shao, Y., Luo, C., Zhang, M., & Ma, S. (2018). How well do offline and online evaluation metrics measure user satisfaction in web image search? In The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor, MI (pp. 615–624). New York, NY: ACM.

DOI: 10.1002/asi