Capturing Engagement As A Multidimensional Datapoint

HARISH RAVI, Virginia Tech
SUDIPA SHRESTHA, Virginia Tech
MINH NGUYEN, Virginia Tech

User engagement largely determines website success, making it a crucial asset for any site. Data is considered the oil of today[6], and great products cannot be made without sufficient data to boot. However, requiring users to create an account to capture their engagement can be a missed opportunity for budding companies. In this project, we propose a novel approach for capturing user engagement on sites with minimal direct user interaction, such as likes, comments, and logins. Instead, we utilize site-usage data points like user scroll speed. We hypothesize that these metrics can be just as valuable, if not more so, than direct user engagement data. To test this hypothesis, we will build a movie buff website that uses scroll data on movie-related posts to provide personalized movie suggestions. We hope to demonstrate the power of this framework and provide a valuable tool for companies seeking to enhance their user engagement strategies.

CCS Concepts: • **Human-centered computing → Gestural input**; • **Information systems → Personalization**.

Additional Key Words and Phrases: elasticsearch, user visualization, user engagement, ELK stack

ACM Reference Format:
Harish Ravi, Sudipa Shrestha, and Minh Nguyen. 2018. Capturing Engagement As A Multidimensional Datapoint. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

User data is a valuable resource that can provide insights to companies seeking to improve their existing products or introduce new ones. To succeed in the market, businesses require a significant quantity and quality of user data. Traditional methods of tracking user engagement, such as likes, dislikes, and comments, provide valuable information, but they only scratch the surface. Our project introduces a new dimension of data points by utilizing user scroll logs. By tracking the time users spend on posts or videos, we gain deeper insights into their interests and preferences. We can then use this data to provide targeted advertisements and product recommendations. Our project focuses on providing movie recommendations based on users' scrolling activity. Using the ELK stack, we aggregate, analyze, and visualize user usage data to generate a list of recommended movies. By avoiding the time-consuming sign-up process and leveraging site usage data, we hope to demonstrate the potential value of this approach to businesses looking to improve their user engagement strategies.

2 RELATED WORK

Our project aims to develop a data pipeline that ingests application API data to generate personalized content recommendations for users. One of our unique approaches is incorporating site usage data into our recommendation system,

Manuscript submitted to ACM

which we believe will enhance content relevance. While traditionally collected user-engagement data points like likes and dislikes provide a variety of feedback, the positive unlabelled data, such as clicks and implicit indicators, can also provide valuable insights at scale. Claypool et al. [4] explored this idea by building a recommendation system based on implicit indicators. By collecting anonymized and non-intrusive data, we can gather a wealth of information to improve our recommendation system at scale.

Moreover, Saito et al. (2020) proposed a novel approach to training recommendation systems with implicit feedback data that are missing-not-at-random (MNAR) using an algorithm that jointly models the probabilistic structure in both observed and missing data [7]. By addressing the problem of biased recommendations caused by missing data, the MNAR method achieved higher quality and fairness in comparison to state-of-the-art methods [7].

To inform our pipeline design, we conducted a comprehensive analysis of the technical whitepapers from leading tech firms. By drawing inspiration from their data processing pipelines, we hope to create a scalable and efficient pipeline that can handle large volumes of data while providing accurate recommendations to our users.

Elasticsearch is a popular tool used by companies across various industries to quickly collect data from both logged-in users and site visitors. Its versatility makes it particularly useful for building recommendation systems. For instance, at Facebook, elastic search and deep learning are integral components of the News Feed algorithm [2]. By analyzing data points such as likes and content interactions, the algorithm can indirectly infer users' preferences and tailor future posts to their interests [2]. Using advanced AI models, the system interprets users' intentions and provides personalized recommendations to enhance the user experience. The widespread adoption of Elasticsearch by leading tech firms underscores its importance in the data processing landscape and highlights its potential for powering innovative solutions in various domains.

Meanwhile, Netflix and Youtube are two prime examples of companies that use elastic search to recommend content to millions of users. Netflix collects a variety of data points, including viewing history, rating, scroll activity, and navigation, to help users find relevant content [3]. Similarly, Youtube leverages data on clicks, watch time, survey responses, sharing, likes, and dislikes to enhance users' viewing experiences [5]. Notably, Youtube also uses machine learning to classify "authoritative" videos from its elastic search database to combat misinformation [5]. The use of elastic search by these industry leaders is a testament to its importance in recommendation systems and its potential to revolutionize the way we discover and consume content.

In addition, Uber has successfully integrated elastic search into its software to create a real-time prediction app for estimated time of arrival (ETA) of carpool or UberEATS services [1]. With elastic search's support for multiple search filters, queried data can be quickly and accurately retrieved, allowing for seamless scalability of the product. In addition to elastic search, Uber leverages other components of the ELK stack, including logstash for data indexing and kibana for visualization [1]. By utilizing elastic search and other components of the ELK stack, Uber is able to deliver accurate and timely predictions, ultimately enhancing the overall user experience for its customers.

Robinhood, a user-friendly trading app, utilizes elastic search to support recommended stocks and detect fraud using a variety of data points. These data points include in-app search data, app review analysis, app and ad metrics, company metrics, customer support analysis, and log analysis [8]. By leveraging elastic search, Robinhood can analyze vast amounts of data quickly and efficiently, providing users with personalized recommendations and detecting fraudulent activity in real time. This enables Robinhood to deliver a seamless and secure trading experience to its users, ultimately contributing to the app's overall success.

REFERENCES

[1] 2017. Engineering uber trip distance and duration predictions in real time with Elk. *Uber Blog* (Jul 2017). https://www.uber.com/blog/elk/

[2] 2022. The new AI-powered feature designed to improve feed for everyone. *Meta AI* (Oct 2022). https://ai.facebook.com/blog/facebook-feed-improvements-ai-show-more-less/

[3] Netflix Technology Blog. 2017. What's trending on Netflix? *Medium* (Apr 2017). https://netflixtechblog.com/whats-trending-on-netflix-f00b4b037f61

[4] Mark Claypool, Phong Le, Makoto Wased, and David Brown. 2001. Implicit Interest Indicators. In *Proceedings of the 6th International Conference on Intelligent User Interfaces* (Santa Fe, New Mexico, USA) *(IUI '01)*. Association for Computing Machinery, New York, NY, USA, 33–40. https://doi.org/10.1145/359784.359836

[5] Cristos Goodrow. 2021. On YouTube's recommendation system. *blog.youtube* (Sep 2021). https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/

[6] Yonego Joris Toonders. 2015. Data is the new oil of the Digital Economy. *Wired* (Aug 2015). https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/

[7] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 501–509.

[8] Arpan Shah. 2017. Taming elk. https://robinhood.engineering/taming-elk-4e1349f077c3

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009