Experts versus Novices: Analyzing Behavioral Variability in Complex Process Environments

Nico Elbert
Julius-Maximilian-University Würzburg, Germany, nico.elbert@uni-wuerzburg.de

Christoph M. Flath
Julius-Maximilian-University Würzburg, Germany, christoph.flath@uni-wuerzburg.de

Behavioral Variability and Process Conformance in Complex Process Environments: Analyzing 18 Million Events

Research Paper

Nico Elbert[1] and Christoph M. Flath[1]

Julius-Maximilians-Universität Würzburg, 97070 Würzburg, Germany
{nico.elbert,christoph.flath}@uni-wuerzburg.de

Abstract. This paper introduces a novel approach to studying user behavior through the User Behavior Mining Framework, analyzing a unique log of 948,251 complex user interaction traces across 78,547 individuals featuring logs of low-level activities in a complex process environment. Employing methods such as directly-follows-graph analysis and trace clustering and next action prediction, the research uncovers the impact of varying experience levels on user interaction patterns and enhances predictive modeling for action forecasting in complex scenarios. This work not only addresses a significant gap in the field by leveraging an underutilized data source but also highlights the importance of rich, detailed datasets for a comprehensive understanding of user behavior and system interaction.
Keywords: User behavior mining, UI logs, Process mining, Data sourcing

1 Introduction

Information systems (IS) are increasingly focused on enhancing user interactions through automation and optimization. To accomplish this goal understanding user behavior is key. It allows researchers to better comprehend, predict, or influence how users interact. By analyzing these interactions, researchers can map the underlying processes, reveal behavioral influences, and ultimately enhance these digital environments. As highlighted by Rehse et al. (2024), the study of user behavior has been approached through qualitative methods (Amoako-Gyampah 2007), laboratory experiments (Burton-Jones & Straub Jr 2006), and the solicitation of direct user feedback (Parks 2012). While valuable, these methods have limitations as they tend to primarily capture the behaviors users actively perceive. This may give an incomplete picture as users tend to be unable to track their behavior over extended periods (Hoffmann et al. 2019).

With the User Behavior Mining (UBM) Framework, Rehse et al. (2024) put forward a principled approach to studying user behavior in a data-driven way, through the analysis of user interaction (UI) logs. Such logs offer rich behavioral traces of user interactions, allowing for an in-depth examination of user behaviors at a scale and granularity not previously feasible (Dumais et al. 2014). Rehse et al. (2024) note a significant challenge in this research area: the reluctance of individuals and organizations to share detailed data necessary for such studies, caused by concerns about privacy, regulatory restrictions, and considerations of competitive advantage (van der Aalst et al. 2017, Tenopir et al. 2011). While some datasets capturing user interactions are publicly available (van der Aalst et al. 2012), they tend to represent fairly simple processes. These datasets[1], covering areas like customer interaction (Dees & van Dongen 2016), order management (van Dongen 2019), and incident handling (van Dongen 2014), provide valuable snapshots of user-system engagement and process management, but often fall short in complexity and lack the comprehensive user metadata needed for an in-depth analysis of user behavior. This limitation points to a significant gap in the field, underlining the need for a broader range of intricate and accessible data sources that not only highlight different aspects of complex user interaction but also include rich, detailed user metadata for a more complete understanding of user behavior and system interaction.

Our paper aims to contribute to the methodological advancement of UBM by showcasing how the UBM Framework can be employed to systematically examine behavioral variability between different user groups. We leverage a novel data set featuring a UI log of 948,251 complex traces across 78,547 unique users, demonstrating the framework's capability to uncover meaningful patterns from complex user interaction data. The log contains low level activities, consisting of *Case ID*, *Timestamp*, *Target Element*, *Action Type* and *Context Parameters*.

The log captures competitive interactions in the real-time strategy (RTS) game *Age of Empires II*. The environment features complex time-critical action sequences, so called build orders. These build orders are in principle comparable to chess openings, yet due to uncertainty, hidden information and a much larger game environment significantly richer. Driven by personal and intrinsic motivation, users exhibit highly rational decision-making, engage in long-term skill learning and collaborate within organizational structures for collective success (Reitman et al. 2020). As noted by Wagner (2006) these interactions capture human behavior in a naturalistic environment over extended time periods, addressing the challenges highlighted by Hoffmann et al. (2019). To the best of our knowledge this is the first study applying UBM to such data in a process mining context. However, other studies, such as the works of Clement (2023) and Ching et al. (2021) recently demonstrated, how such data can serve as a large-scale controlled experimental proxy.

We leverage the UBM framework to make several key contributions to the field. Notably, we explore the relationship between user experience and behavior variability. Through directly-follows-graph analysis and trace clustering, we uncover how various levels of experience affect interaction patterns within the environment. Furthermore, we enhance predictive modeling techniques for user actions in complex scenarios. Our work investigates the predictability of actions across different experience levels, aiming to accurately forecast decisions in challenging contexts. Lastly, we address a notable research gap by introducing a novel data source. The use of RTS UI logs presents a unique, underexplored opportunity for UBM research.

[1] https://www.processmining.org/event-data.html

2 UBM Methodology

Our approach is grounded in the UBM Framework as conceptualized by Rehse et al. (2024). This section recaps the framework's component and in turn sets the stage for its application in our study. The UBM Framework presents a systematic method for analyzing user interactions within software applications. This four-part framework encompasses UBM data, UBM technology, UBM objectives, and theoretical contributions, aiming to provide a holistic approach to understanding and improving user-system interactions.

2.1 Data

Central to the UBM Framework is the utilization of UI logs, which are detailed records of user activities during software interaction. These logs, capturing everything from basic clicks to complex task executions, serve as the atomic units for analysis. The framework emphasizes the challenge of categorizing these diverse interactions, acknowledging the need to discern between different user actions for a meaningful analysis of behavior.

2.2 Technology

UBM technology integrates a wide range of data analysis techniques tailored for UBM data, spanning from innovative to adapted existing methods designed to address the unique demands and objectives of UBM. Process mining is recognized as a critical tool, providing both a conceptual and technical framework suitable for exploring and confirming aspects of user behavior through UBM data. Despite its potential, the specific traits of UBM data, such as its variability and granularity, as well as the integration of behavioral theories, may limit the direct applicability of standard process mining methods.

Rehse et al. (2024) categorize the landscape of available techniques into exploratory and confirmatory data analysis techniques. Exploratory techniques aim to reveal new knowledge about user behavior captured in the data, addressing questions related to user interactions, navigation efficiency, usability issues, and identifying abnormal or fraudulent behaviors through process mining. This includes process discovery to visualize user navigation paths, process enhancement for identifying usability bottlenecks, anomaly detection for spotting unusual behaviors, and process prediction to forecast future actions, interaction durations or the ultimate outcome of a process.

Confirmatory techniques focus on testing hypotheses about user behavior, seeking to validate or refute assumptions regarding deviations from intended design, reasons for user drop-offs, efficiency improvements through new features, or changes in behavior following software updates. These techniques often require adapting existing process mining methods, like stochastic conformance checking, to assess user behavior against expected models, and exploring concept drift detection and causal analysis to investigate changes in behavior or causal relationships within user interactions.

2.3 Objectives

The objectives of the UBM Framework are highly depended on the underlying software application and the tasks for which it is used. Rehse et al. (2024) categorize resulting objectives in the areas analysis, assistance, and automation. In context of analysis, UBM aims to provide insights into user behavior, identifying usability problems, system design deviations, and user groups with similar patterns. For assistance, the focus shifts towards supporting users in navigating and utilizing complex or flexible software more efficiently. This involves leveraging UBM to offer real-time, context-specific guidance and recommendations, potentially steering users away from common pitfalls or optimizing their navigation through the system based on their current activities and past user behavior patterns. Lastly, automation through UBM targets the identification and execution of repetitive tasks that can be automated, typically employing Robotic Process Automation techniques. By analyzing user actions, UBM facilitates the discovery of high-potential automation opportunities and the generation of automation scripts, streamlining operations and reducing manual effort.

By delving into UI logs, the framework seeks to uncover usability issues, tailor user interfaces to distinct user groups, and identify opportunities for automating repetitive tasks. This tripartite goal underlines the framework's commitment to enhancing both user experience and system efficiency.

2.4 Theory

According to the UBM framework, data, technology, and objective are sufficient to instantiate a UBM application, especially in technical domains. By leveraging established theories regarding system usage or real-life behavior, UBM applications can enrich their analytical depth and prediction capabilities and gain access to general and causal explanations about the interactions between humans and IT artifacts. The integration of such theories from IS and Human-Computer Interaction provides a structured approach to analyzing user behavior, enriching both practical outcomes and theoretical advancements in the field.

3 Study Design

Using the UBM framework, our study seeks to systematically analyze user behavior, drawing on the rich data available from user interactions in a complex RTS environment. We appply the UBM framework to elucidate the influence of user experience on behavior and to demonstrate the practical application of such UI logs as a novel source for such analysis.

3.1 Data Sourcing

For our study, we collected a log of 948,251 publicly available UI traces, featuring complex competitive interactions in the RTS game *Age of Empires II*. This dataset was compiled by systematically scraping *AoE2 Insights*, a public analytics platform[2]. The log contains low level activities per user and interaction, consisting of *Case ID*, *Timestamp*, *Target Element*, *Action Type* and *Context Parameters*, as depicted in Table 1. The log is annotated with a range of meta data features, specifying the configuration of the environment. Our dataset features 78,547 different users, each with multiple appearances between April 2020 and September 2023. Every user is associated with an Elo rating, assessing a users experience and skill level at the time of the interaction. As we can identify users by their respective User ID, our dataset also features the development of this Elo rating alongside a record of all historic interactions as well as the user's nationality and preferences regarding the configuration of the environment. Further, a subset of 188,068 traces is annotated with an exogenous label, categorizing the featured trace-type.

Table 1. Sample of the Dataset

CaseID	Timestamp	Target Element	Action Type	Parameters
138015780_1	00:00:01	TC-01	Queue 17	Amount:1
138015780_1	00:00:02	TC-01	Queue 17	Amount:1
138015780_1	00:00:03	TC-01	Queue 17	Amount:1
138015780_1	00:00:04	TC-01	Queue 17	Amount:1
138015780_1	00:00:04	S-01	Move	(x=14,y=17)
138015780_1	00:00:04	V-01	Build 70	(x=9,y=12)
138015780_1	00:00:05	V-02	Build 70	(x=12,y=8)
...	...	...	...	...

3.2 Data Preprocessing

To analyze influential factors on user behavior, comparable subgroups in the dataset are crucial. Given the complexity of the underlying processes, especially for process mining related methodologies, we focus on similar processes, using the exogenous labels and narrow configurations for filtering.

The filtering is applied progressively over three stages of configuration adjustments (S1, S2, S3), each further constraining various settings within the environment. In S1 and S2 we apply filters on the general environmental settings, specified by the level type (S1) and the through the user selected set of available actions (S2). To standardize the processes further, we define a critical stop event, which represents a crucial part in the process of the interaction. In S3 we filter for this and other process related features. This step-by-step constriction results in a series of logs with user interaction traces, the quantities of which are depicted in Figure 1 (left).

We partition the refined dataset into three Elo classes —low, medium, and high—to represent user groups of varying skill levels. Our goal is to generate substantial databases for each subgroup to ensure statistical robustness. These subgroups are determined not to be contiguous to one another to maximize intragroup homogeneity and intergroup

[2] `http://www.aoe2insights.com`

heterogeneity. The intervals for these subgroups are selected based on the sample distribution of specific trace types within the overall population, as illustrated in Figure 1 (right).

Given the different numbers of samples in each Elo class, we employ a subsampling strategy aimed at maintaining a target standard deviation (SD) of user Elo ratings within our dataset. To achieve this, we iteratively exchange samples of the larger populations, until we meet the SD of the smallest subgroup within a tolerance of 5%. This method ensures that the heterogeneity of users skill levels is consistent across our groups, allowing for more reliable comparisons. This strategic separation facilitates an examination of user behavior across varied levels of skill with a focus on retaining comparative integrity within each subgroup.

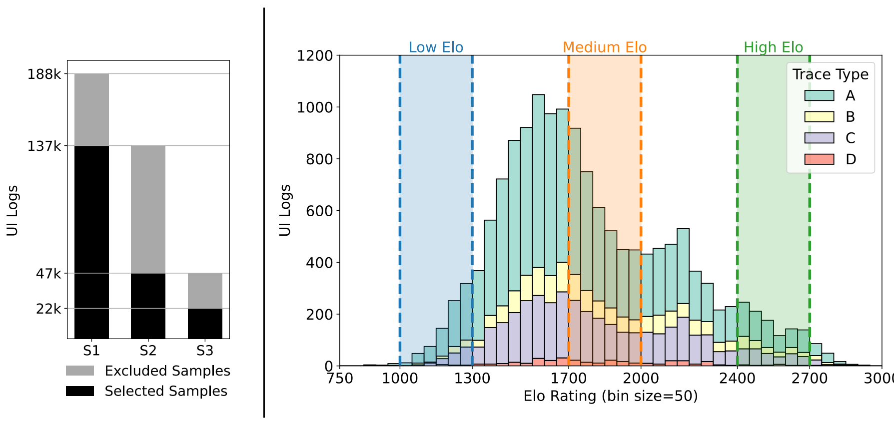

Figure 1. Number of UI traces in Logs of Filter Stages (S1,S2,S3) [left] and Distribution of UI Traces by Trace Type over Elo Values annotated with defined Elo Classes [right]

To maintain consistency among our approaches, we use the same set of configurations for all of our analysis, defined by the selected trace type 'A' and the level of granularity, achieved by including and excluding specific action types from the log. To analyze the behavioral variability and demonstrate the suitability of such UI logs, we conceptualize two UBM applications as proposed by Rehse et al. (2024).

3.3 Usage Pattern Variation

In our first application, summarized in Figure 2, we aim to validate and understand the impact of user skill level on behavioral patterns in user interaction logs. Abb et al. (2022) propose to address the challenge often posed by the complexity of the UI log data (Leno et al. 2021), by applying trace clustering. In line with Song et al. (2009) and Elbert et al. (2023), we utilize the Euclidean distance between aggregated traces to group similar behaviors, allowing us to verify the exogenous labels initially used for categorization.

Subsequent to the clustering process, we generate Directly-Follows Graphs (DFGs) by deploying timeline-based methods (Kaur et al. 2023). Given the nature of the captured interactions, we annotated the actions with a counter, enhancing the structure and interpretability of the timeline-based DFGs. Further we initially limit the horizon of possible

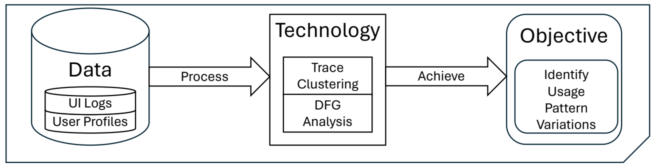

Figure 2. Applying the UBM Framework to Identify Usage Patterns Variations

number of actions or time passed in a process, and increase these limits iteratively. This allows us to compare the processes of getting to a certain state of an interaction, in addition to analyzing the full process. These DFGs serve as a tool to visualize and analyze the sequence and variance in user interactions, effectively capturing the flow of activities and user engagement over time.

To assess the influence of user skill on interaction patterns, we measure and compare several DFG metrics across various user skill levels. Key metrics include DFG variance, the count of unique traces, trace lengths and number of edges within these graphs. For conformance checking we calculate the Earth Mover's Distance (EMD), among the different comparable DFG populations (Leemans et al. 2019). These metrics provide insights into the complexity and diversity of user behavior, revealing the influence of skill level on interaction patterns within the environment.

3.4 Predictive Modeling

Our second application, summarized in Figure 3, concentrates on process prediction, which aims to forecast the future behavior of a running process instance (Evermann et al. 2017). We aim to predict user's subsequent actions and overall process outcomes based on their interaction traces.

We utilize a transformer-based model, GPT-2 (Radford et al. 2019), to forecast the next action within a user's log sequence. This approach leverages the sequential nature of user actions, aiming to predict subsequent steps based on historical data and create recommendations for next process steps as demonstrated by Rehse et al. (2024).

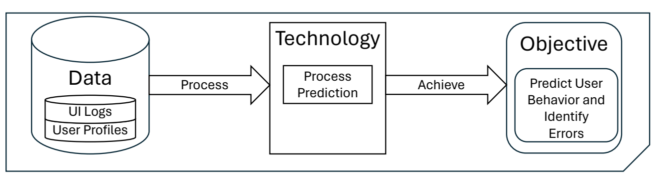

Figure 3. Applying the UBM Framework
to Predict User Behavior and Identify Errors

Further, we employ aggregate trace representations to predict the final outcome of a process (success or not) using gradient boosted trees (Chen & Guestrin 2016). This methods allows us to assess the overall performance of users and identify potential errors in their interaction sequences. By evaluating the quality of predictions across different skill levels, we can determine the extent to which user experience influences the predictability of their actions and the likelihood of successful outcomes.

4 Results

This section presents the tangible outcomes of our applications, illustrating the framework's broad applicability and effectiveness.

4.1 Usage Pattern Variation

In our analysis of usage pattern variations, as outlined by our first UBM application (see Figure 2), we apply trace clustering to our dataset to segregate distinct trace types into multiple clusters. The aggregation of trace representations within the process allows us to verify the exogenous trace type labels with which the log is annotated.

We create timeline-based DFGs as suggested by Kaur et al. (2023), illustrated in an excerpt in Figure 4. The complexity of the interactions and resulting DFGs, even when narrowed down significantly, posed a challenge in comprehension. However, the simplification of the DFGs through adjustments in log horizon and granularity facilitated insights into the early steps and phases of the underlying processes.

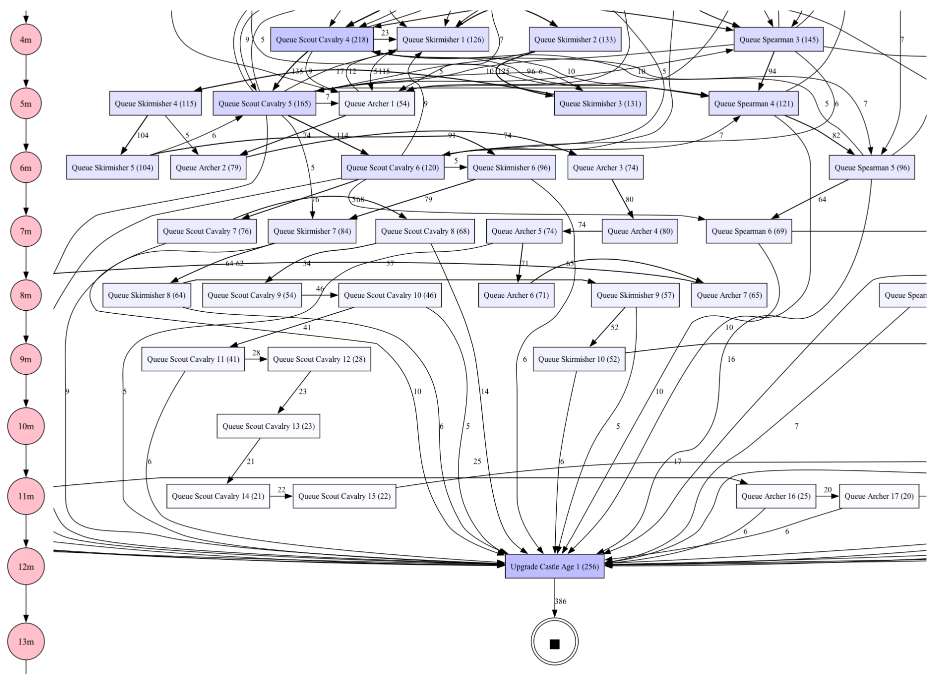

Figure 4. Excerpt of the timeline-based DFG for High Elo Users performing Trace Type A

To quantitatively compare the impact of user skill levels on user behavior, we computed the EMD as well as other key metrics like DFG variance, number of nodes, edges, and unique traces across different Elo classes. The comparative results for the EMD over all created DFGs are depicted in Figure 5. The results highlight a significant difference between the different DFGs, increasing with the size of the skill gap.

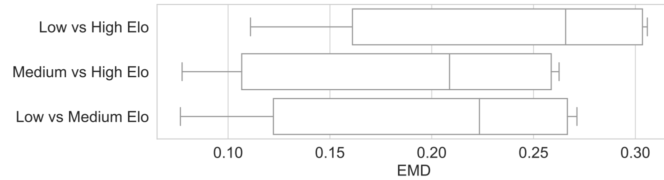

Figure 5. EMD between DFGs of different Elo Classes over all different configurations

Figure 6 illustrates the development of the number of unique traces per DFG over elapsing process time. After 5 minutes into the process, a difference in the number of unique traces becomes evident, underlying a higher uniformity in the behavior of higher skilled users. In a later stage of the game, around 20 minutes, the highest Elo class rapidly increase the number of unique traces, which can be an indicator for a higher adaptability in the later stages of the process.

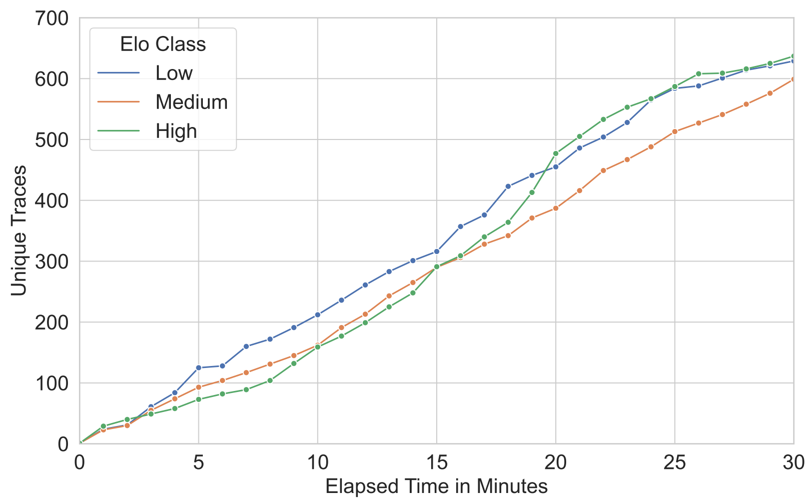

Figure 6. Number of Unique Traces Over the Elapsed Game Time in Minutes

4.2 Predictive Modeling

Our investigation into predictive modeling focused on the predictions of subsequent actions and process outcomes. Employing traces represented as sequences of actions, our transformer based model (Radford et al. 2019) was tasked with forecasting the subsequent action.

To systematically evaluate the model's predictive performance across different user expertise levels, we constructed action sequences for each Elo class. These sequences were formulated by restricting the number of actions from the beginning of the process, referred to as the action depth. Each sequence was then utilized to train and subsequently validate independent models specific to that action depth range. Figure 7 presents the aggregated results, illustrating the model's accuracy across incremental action depth intervals, each spanning five actions. This interval grouping was chosen to mitigate the variability in the data and to present a clearer picture of the model's predictive capabilities at varying stages of action sequences. As depicted in Figure 7 the model demonstrated high accuracy, particularly for traces early in the process. This accuracy decreased with the depth of the action sequence, yet remained commendable across all Elo classes. These findings imply that the predictability of user actions at the onset of interaction is relatively consistent across skill levels, a notable point considering the distinct variations in the number of unique traces observed among different Elo classes.

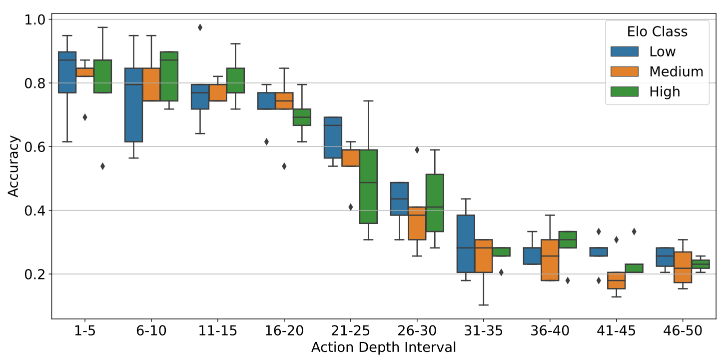

Figure 7. Accuracy for Next Action Prediction for Full Traces with Restricted Action Depth

To further our analysis, we applied gradient boosted trees (Chen & Guestrin 2016) based on aggregated traces to predict the final outcome of a process, evaluating success probability as time progressed. Figure 8 depicts the growing accuracy of our outcome predictions as the elapsed time of interaction increases. The quality of these predictions did not exhibit a marked difference across Elo classes, demonstrating that our model could uniformly assess user performance over time, regardless of skill level.

The proficiency of the outcome prediction model offers a useful tool for identifying which actions may decrease the probability of success, thus pinpointing potential errors in process execution. The consistent quality of predictions across different user experiences suggests that the model is capturing fundamental aspects of user behavior that are common across skill levels, offering insights into process execution that are largely independent of individual user skill or experience.

5 Discussion

The analysis of usage patterns revealed through our UBM applications necessitates a discerning examination of the metrics employed. Particularly in complex environments,

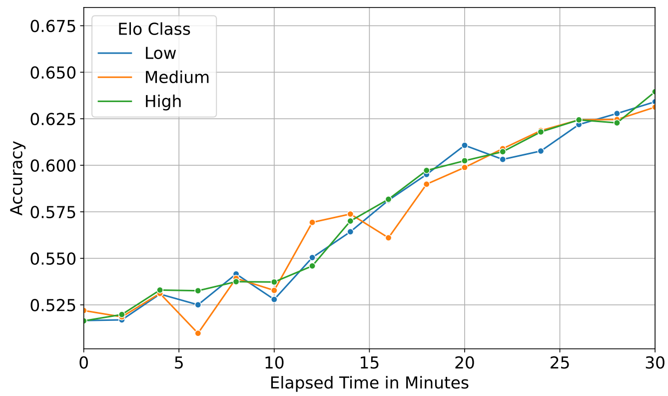

Figure 8. Accuracy of Outcome Prediction for Full Traces with Increasing Elapsed Time

such as the one featured in our UI log, the interpretation of metrics like the number of unique traces demands contextualization. Contrary to typical process environments where high variance may signify inefficiency or deviation from the standard, in such complex environments, this variance can reflect a more nuanced scenario.

The usage of objective labels pertaining to the user's skill level offers a lens through which we can view higher Elo users' actions as more optimal (Anderson et al. 2017). For example, the initial proliferation of unique traces among lower Elo users may be indicative of randomness and a lack of strategic coherence, whereas higher Elo users exhibit more uniform behaviors early in the process. This trend inverts later in the interaction, where an uptick in unique traces among more skilled users signals their ability to adapt and navigate the process flexibly, thereby optimizing their behavior.

The application of the EMD in our study has quantitatively captured the qualitative differences in behaviors among users of varying skill levels. The increase in EMD with the skill gap suggests nuances in behavior that conventional metrics such as the number of edges, unique traces, or trace lengths do not capture.

Our next action prediction models shed light on the predictability of user actions, demonstrating high accuracy across all Elo classes early in the process. Yet, this finding prompts a closer inspection of whether alternative metrics, like top-n accuracy, could provide a more fitting evaluation. Given the frequency and partial interchangeability of actions in such complex environments, a broader accuracy measure could account for the marginal impact of different action orders on the resultant process state.

These findings underline the importance of careful metric selection and interpretation in complex process environments within the domain of UBM. It is clear, that while certain metrics can provide valuable insights, they must be chosen and analyzed with consideration of the unique dynamics present in the domain of study. This acknowledgment paves the way for more sophisticated and contextually aware approaches in future UBM research, particularly as we venture into environments with highly variable and intricate user behaviors.

6 Outlook and Conclusion

Our study demonstrates the ways in which user behavior varies according to user experience, particularly in the context of navigating complex processes within RTS environments. Our research has not only addressed the question posed by Rehse et al. (2024) regarding the behavioral differences between experienced and inexperienced users but has also leveraged large-scale UI log datasets to perform comprehensive UBM.

The successful implementation of our UBM applications underscores the potential of environments such as RTS games for UBM research. The detailed and publicly available dataset used has proved to be an invaluable resource, circumventing common privacy and security concerns, as described by Rehse et al. (2024), while providing the necessary features for a thorough UBM study, including user identification and the complexity of processes involved.

Looking forward, our research opens for several promising directions in the field of UBM, allowing researchers to address the challenges outlined by Rehse et al. (2024):

- **Conceptualization and Standardization:** The variety of log types and levels of granularity offered by this and comparable environments create a unique opportunity to develop sophisticated methods for log homogenization. This includes the creation of standardized exchange formats that could serve as benchmarks for other domains experiencing high log heterogeneity. The presence of publicly accessible logs, in combination with transparent underlying processes, provides an excellent foundation for efforts aimed at achieving a greater level of uniformity and interoperability in UBM datasets.
- **Theoretical Expansion and Grounding:** Our findings highlight the opportunity for theoretical development within UBM. Future studies have the potential to integrate complex user behaviors with existing theories or to devise new frameworks that more accurately reflect the intricate nature of user interactions in a variety of settings.
- **Confirmatory and Causal Analysis Techniques:** The dataset enables the development and benchmarking of techniques aimed at deriving insights from logs that are both highly variable and finely granular. There is ample scope for both developing and applying confirmatory and causal analysis techniques for complex UI data and advancing our general understanding of user behavior.

By delving into these domains, subsequent research can further enrich the discipline of UBM, enhancing our ability to analyze complex user behaviors and contributing to the evolution of more intuitive, secure, and efficacious models of user interaction. In reflecting upon our work, it's crucial to recognize the challenges in applying complex UI logs from RTS environments to UBM. The process of accurately mapping these logs for process-oriented analysis demands rigorous precision to maintain validity. Despite these challenges, when handled carefully, this approach unlocks significant insights, enriching the UBM field. Therefore, we advocate for the strategic use of such data, asserting that, with thoughtful analysis, it broadens our understanding of user behavior and opens new pathways for research.

References

Abb, L., Bormann, C., van der Aa, H. & Rehse, J.-R. (2022), Trace clustering for user behavior mining., *in* 'ECIS'.

Amoako-Gyampah, K. (2007), 'Perceived usefulness, user involvement and behavioral intention: an empirical study of erp implementation', *Computers in human behavior* **23**(3), 1232–1248.

Anderson, A., Kleinberg, J. & Mullainathan, S. (2017), 'Assessing human error against a benchmark of perfection', *ACM Transactions on Knowledge Discovery from Data (TKDD)* **11**(4), 1–25.

Burton-Jones, A. & Straub Jr, D. W. (2006), 'Reconceptualizing system usage: An approach and empirical test', *Information systems research* **17**(3), 228–246.

Chen, T. & Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '16, Association for Computing Machinery, New York, NY, USA, p. 785–794.

Ching, K., Forti, E. & Rawley, E. (2021), 'Extemporaneous coordination in specialist teams: The familiarity complementarity', *Organization Science* **32**(1), 1–17.

Clement, J. (2023), 'Missing the forest for the trees: Modular search and systemic inertia as a response to environmental change', *Administrative Science Quarterly* **68**(1), 186–227.

Dees, M. & van Dongen, B. (2016), 'Bpi challenge 2016: Questions'.
URL: *https://data.4tu.nl/articles/dataset/BPI_Challenge_2016_Questions/12687320/1*

Dumais, S., Jeffries, R., Russell, D. M., Tang, D. & Teevan, J. (2014), 'Understanding user behavior through log data and analysis', *Ways of Knowing in HCI* pp. 349–372.

Elbert, N., Stein, N. & Flath, C. M. (2023), Process and strategy mining in real-time strategy games, *in* 'Wirtschaftsinformatik 2023 Proceedings', p. Article 98.
URL: *https://aisel.aisnet.org/wi2023/98*

Evermann, J., Rehse, J.-R. & Fettke, P. (2017), 'Predicting process behaviour using deep learning', *Decision Support Systems* **100**, 129–140.

Hoffmann, P., Mateja, D., Spohrer, K. & Heinzl, A. (2019), Bridging the vendor-user gap in enterprise cloud software development through data-driven requirements engineering., *in* 'ICIS'.

Kaur, H., Mendling, J., Rubensson, C. & Kampik, T. (2023), 'Timeline-based process discovery', *arXiv preprint arXiv:2401.04114* .

Leemans, S. J., Syring, A. F. & van der Aalst, W. M. (2019), Earth movers' stochastic conformance checking, *in* 'Business Process Management Forum: BPM Forum 2019, Vienna, Austria, September 1–6, 2019, Proceedings 17', Springer, pp. 127–143.

Leno, V., Polyvyanyy, A., Dumas, M., La Rosa, M. & Maggi, F. M. (2021), 'Robotic process mining: vision and challenges', *Business & Information Systems Engineering* **63**, 301–314.

Parks, N. E. (2012), Testing & quantifying erp usability, *in* 'Proceedings of the 1st Annual conference on Research in information technology', pp. 31–36.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. & Sutskever, I. (2019), 'Language models are unsupervised multitask learners'.

Rehse, J.-R., Abb, L., Berg, G., Bormann, C., Kampik, T. & Warmuth, C. (2024), 'User behavior mining', *Business & Information Systems Engineering* pp. 1–18.

Reitman, J. G., Anderson-Coto, M. J., Wu, M., Lee, J. S. & Steinkuehler, C. (2020), 'Esports research: A literature review', *Games and Culture* **15**(1), 32–50.

Song, M., Günther, C. W. & van der Aalst, W. M. (2009), Trace clustering in process mining, *in* 'Business Process Management Workshops: BPM 2008 International Workshops, Milano, Italy, September 1-4, 2008. Revised Papers 6', Springer, pp. 109–120.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011), 'Data sharing by scientists: practices and perceptions', *PloS one* **6**(6), e21101.

van der Aalst, W. M., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., Van Den Brand, P., Brandtjen, R., Buijs, J. et al. (2012), Process mining manifesto, *in* 'Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9', Springer, pp. 169–194.

van der Aalst, W. M., Bichler, M. & Heinzl, A. (2017), 'Responsible data science'.

van Dongen, B. (2014), 'Bpi challenge 2014: Activity log for incidents'.
URL: *https://data.4tu.nl/articles/dataset/BPI_Challenge_2014_Activity_log_for_incidents/12706424/1*

van Dongen, B. (2019), 'Bpi challenge 2019'.
URL: *https://data.4tu.nl/articles_12715853/1*

Wagner, M. (2006), On the scientific relevance of esports., *in* 'International conference on internet computing', pp. 437–442.