Interactive Event Sequence Prediction for Marketing Analysts

Fan Du
Adobe Research
San Jose, CA
fdu@adobe.com

Eunyee Koh
Adobe Research
San Jose, CA
eunyee@adobe.com

Shunan Guo
Adobe Research
San Jose, CA
sguo@adobe.com

Sungchul Kim
Adobe Research
San Jose, CA
sukim@adobe.com

Sana Malik
Adobe Research
San Jose, CA
sana.malik@adobe.com

Zhicheng Liu
Adobe Research
Seattle, WA
leoli@adobe.com

CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3382971

Abstract
Timestamped event sequences are analyzed to tackle varied problems but have unique challenges in interpretation and analysis. Especially in event sequence prediction, it is difficult to convey the results due to the added uncertainty and complexity introduced by predictive models. In this work, we design and develop ProFlow, a visual analytics system for supporting analysts' workflow of exploring and predicting event sequences. Through an evaluation conducted with four data analysts in a real-world marketing scenario, we discuss the applicability and usefulness of ProFlow as well as its limitations and future directions.

Author Keywords
Predictive analytics; event sequence analysis; visualization.

CCS Concepts
•Human-centered computing $\rightarrow$ Visualization;

Introduction
Sequences of timestamped events, such as patients' treatment records, users' website visit logs, and students' learning activities, have been widely collected and analyzed to provide valuable insights and novel applications. With recent advances in statistics and machine learning, predictive analytics has gained traction in digital marketing businesses and inspired emerging use cases such as predicting cus-

tomers' intents or behaviors to provide personalized experiences or proactive supports [20].

Visualization techniques have been developed to help analysts explore event sequence data and conduct a variety of tasks, such as inspecting the timeline of individual records [17, 18], discovering aggregated patterns in a large database [5, 15, 23, 24], and identifying the correlations between different event patterns and outcomes [4, 8, 22]. However, most existing solutions focus on analyzing historical events and outcomes of archived records. These tools often have difficulty in dealing with the added uncertainty and complexity introduced by predictive models, which often result in an exponential number of probabilistic sequences. Even understanding the predictions becomes difficult, let alone finding insights or making decisions.

Our previous study evaluated different design options for visualizing event sequence predictions [9]. Feedback from the study participants suggested an important next step of evaluating the visualizations in real-world applications with domain experts. Motivated by this need, we implemented the previously proposed visualization designs and guidelines into an interactive system and conducted iteratively refinements through interviews with 12 domain experts. Our direct contributions are:

- The design and implementation of ProFlow, a visual analytics system for supporting analysts' workflow of exploring and predicting event sequences.
- An evaluation conducted with four data analysts in a real-world marketing scenario and a discussion on the applicability and usefulness of ProFlow as well as its limitations and future directions.

Related Work

Extensive research has developed visualization techniques for exploring event sequence data and conducting various analytical tasks, such as inspecting the timeline of individual records [17, 18], discovering aggregated patterns in a large database [5, 15, 23, 24], or identifying the connections between different event patterns and their outcomes [4, 8, 22]. However, most existing solutions only focus on analyzing historical events and outcomes of archived records and often have difficulty in dealing with the added uncertainty and complexity introduced by predictive models

On the other hand, visualization tools have been designed for evaluating the performance of predictive models [1, 2, 10, 12, 16, 21] or interpreting the prediction results to support decision making [11, 13, 14, 25]. Yet, very few were dedicated to supporting the exploration of event sequence predictions. In a closely related work, Gilch et al. [7] use Monte-Carlo method to simulate sports teams' progression in a game and show each team's probability of reaching different stages in the Sankey diagram [19]. The Sankey diagram can produce a clear overview of such progression sequences only in simplified situations, for example, a team can only progress to one designated stage after its current stage. In general applications such as digital marketing, a customer may have a probability of opening multiple emails or visiting multiple websites, which will create massive links and cause significant visual clutters.

Our work explores designs for dealing with the added uncertainty and complexity introduced by predictive models and develops a visual analytics system for supporting analysts' workflow of exploring and predicting event sequences.

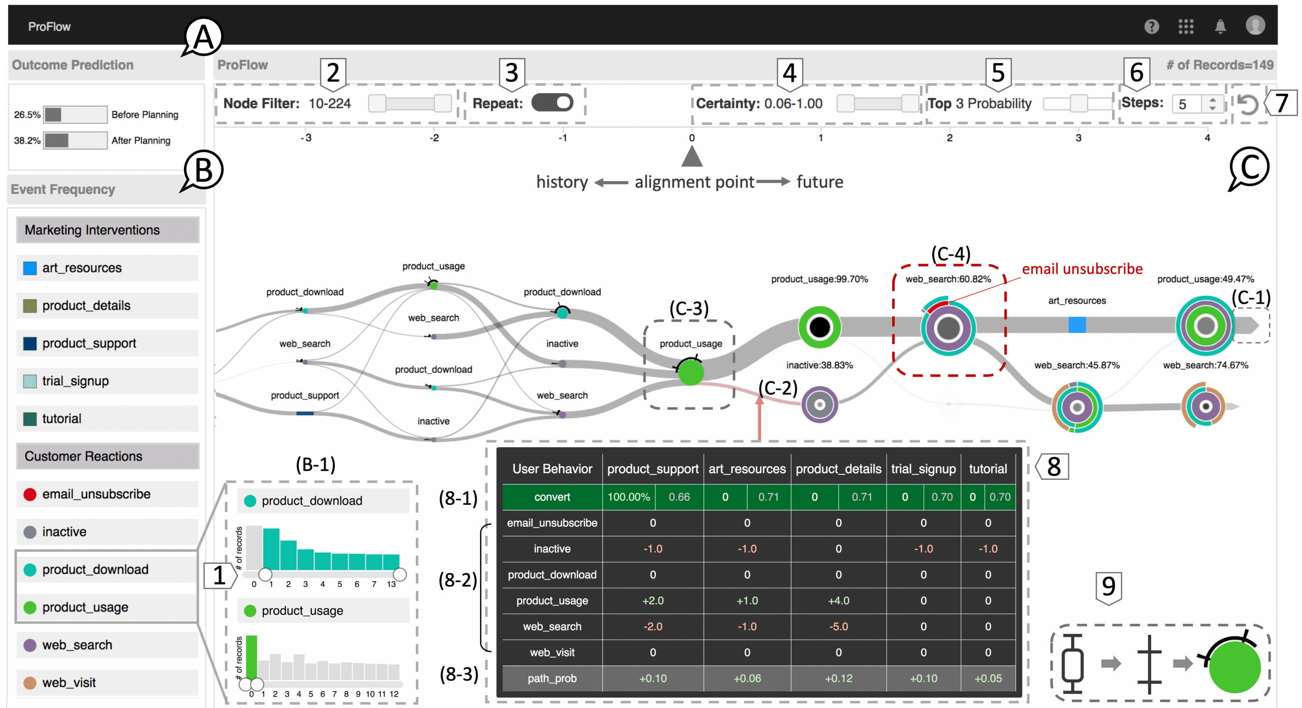

Figure 1: The user interface of ProFlow, a visual analytics system for exploring prediction results of event sequences. Continued on the left.

Figure 1 (cont'd). The ProFlow interface includes (A) an outcome prediction view showing the percentage of records that are predicted to achieve the desired outcome with and without the planned interventions, (B) event frequency filters showing the overall distribution of events in each category, and (C) event sequence view showing the history and future prediction of event sequences. It also provides various controls for filtering records (1), simplifying event sequences (2-3), and facilitating exploration of event sequence prediction results (4-6).

In this example, the analyst found that for customers in a major group (C-4), although "Web Search" was their common top prediction, about 10% of them have an alternative prediction "Email Unsubscribe" indicating a potential risk that needed to be managed with marketing interventions.

Description of ProFlow

ProFlow implements the design considerations and guidelines derived in our previous study [9] for visualizing event sequence predictions. During the development process, we regularly presented our prototypes and collected feedback from 12 expert partners, included 8 marketing practitioners (4 analysts, 3 marketers, and 1 product manager) and 4 machine learning practitioners (2 researchers and 2 data scientists). In this section, we first give an overview of ProFlow's workflow and then introduce each system component for supporting the workflow.

Workflow

ProFlow provides three coordinated interface components (Fig. 1) to support a workflow of interactively predicting and exploring event sequences. Marketing analysts may get started by reviewing customers' history in the event sequence view and estimated purchase rates in the outcome prediction view to identify customer groups of interest. They can use the event frequency filter to simplify the display by keeping only customers meeting specific criteria (e.g., more than three product usages). Then, analysts can generate and explore predictions for the identified group and

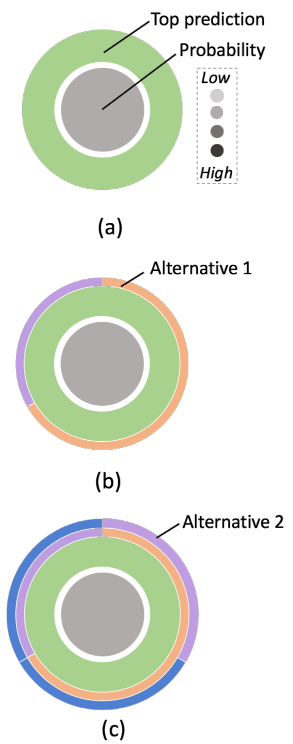

Figure 2: (a) Design for showing the most probable prediction at each step and its probability. (b-c) Alternative predictions with a lower probability can be added by users.

the event sequence view will be expanded to display the predicted future paths. Analysts can make action plans by inserting marketing interventions into the future paths and review how the interventions affect the predictions.

Exploring Historical Activities

ProFlow summarizes event sequences using the Sankey diagram [19] as the primary representation, since it is widely used in real-world commercial tools and familiar to analysts. In the Sankey (Fig. 1(C)), sequences are aligned by their most recent events, with historical events on the left and predicted future events on the right. At each step, records with the same event are aggregated into a node and the links show the transitions of the records between two steps. The size of nodes and the width of links are proportional to the number of records. Nodes' colors encode event categories. Users can hover on a node or link to highlight the complete path of the associated records.

ProFlow provides several simplification strategies [6] to cope with the volume and variety of event sequences. Users can extract records using the event frequency filters (Fig. 1(1)), for example, only keeping customers who have at least one product download but zero product usage (Fig. 1(B-1)) to find those who might need help with installation. Users can also choose to consolidate repeated events of the same category (Fig. 1(3)) and ProFlow will show a circular boxplot (Fig. 1(9)) around the upper half of each node to provide summary statistics of the number of events being consolidated. Finally, to support exploring common activity patterns, users can use a node filter (Fig. 1(2)) to hide nodes of small populations.

Exploring Future Activities and Outcomes

Analysts can generate predictions of a group of records by clicking on the tail link after their last event (Fig. 1(C-1)). With a sequence of events as the input, ProFlow implements a Time-Aware Recurrent Neural Network [3] to predict each record's future activities and outcomes (e.g., purchase or dropout). While each historical event belongs to a certain category, future events generated by predictive models are probabilistic, each represented by a list of event categories ranked by uncertainties. The added complexity results in an exponential number of possible future sequences as we predict multiple steps into the future for multiple records. Even understanding the predictions becomes difficult, let alone finding insights or making decisions.

To simplify the prediction results of multiple steps and multiple records, ProFlow constructs the future portion of the Sankey diagram by only showing the most probable predictions at each step as the outer rings of the nodes (Fig. 2(a)). Uncertainties of the most probable predictions are encoded by the opacity of the dots at the center of the nodes. This design provides an overview of the most probable future paths. Analysts can easily identify which event categories are more popular and certain in the predictions, so as to find records of interest for inspection.

Alternative predictions with a lower probability can be added for fine-grained inspections (Fig. 2(b-c)). Here, each node represents a group of records with the same most probable prediction, shown as the first outer ring. Each additional ring represents a level of alternative predictions of the group, where the uncertainty increases as the level grows. The colored arcs show the popularity of the event categories of the alternative predictions.

Making Action Plans

ProFlow allows analysts to dynamically add intervention events to the future paths to quickly test action plans and receive immediate feedback. As illustrated in Fig. 1(C-2), analysts can click on a link to plan an intervention for a group between two specific predicted actions. A preview table (Fig. 1(8)) will pop up, where the columns represent different types of interventions and for each intervention, the table also provides a summary of its expected effects on the predictions. After the intervention has been added or modified, ProFlow will rerun the prediction by taking the intervention as part of the historical activities and update the predictions of subsequent activities and outcomes.

Evaluation

To understand the applicability and usefulness of ProFlow in real-world scenarios, we conducted a case study with 4 data analysts who worked at the marketing department of a large software company. Two of them managed and analyzed customer data in their daily jobs (E1, E2). The other two built statistical models for marketing applications (E3, E4). The case study lasted about a month consisting of interviews, data preparation, and data exploration. E1-3 also participated in our system prototyping iterations. During the meetings, we provided training of the final ProFlow prototype and answered questions.

Dataset

The analysts provided a real-world marketing dataset of 38,155 customer records. The data consisted of 30-day behavior logs of the customers and provided ground-truth labels of the customers' outcomes (i.e., made a purchase or not). We used 80% of the data for model training and 20% for testing. The models achieved an 85.75% accuracy for predicting the most probable next event and a 74.9% accuracy for predicting outcomes (precision=65.1%, recall=52.9%). Among customers who had a trial, 525 who had not made a purchase were extracted and analyzed in this use case. We tweaked their logs for privacy protection.

Process

The analysts were interested in predicting customers' future behaviors and outcomes to make email sending plans for the following days. After loading the data, they checked the outcome prediction view and found that only 19.7% of the customers were predicted to make purchases. The analysts needed to improve the outcomes to meet their goals. They started by reviewing the overview of the customers' recent activities. By tuning the node filter, high-level activity patterns were revealed and showed that the customers split into two major groups. Customers in the first group were very engaged and had frequent download and product usage events. In contrast, the other group received a lot of marketing emails but continuously being inactive. By drilling down to each group, The analysts found that 32.5% of the active customers had a positive outcome prediction compared to 17.9% in the inactive group, indicating a correlation between trial usage and making purchases.

The analysts decided to focus on the group with active product usage and sent them emails to promote purchases. They double clicked on this group to exclude other records and clicked on the tail link to trigger the prediction for the next 7-step of future activities. The analysts estimated that this group of customers would continue to be active since their future paths contained consistent product usage events with high probabilities. They did not raise any concern from the top predictions so they added the alternative predictions. A red arc in a major node (as illustrated in Fig. 1(C-4)) immediately caught their eyes—although the top prediction of this node was "web search" about 10% of the customers had an "unsubscribe" event as the alternative prediction, indicating a potential risk.

The analysts further drilled down by clicking on the red arc and realized that no customers in this subset had a positive outcome prediction. By reviewing their historical activities, the analysts noticed that these customers had received several emails about "art resource" and "product support" in the past but only used the product once. They thought the customers might need help with using the product. They clicked on the link before their next action to review the intervention table. By comparing different email content (see an example in Fig. 1(8)), they decided that sending a "product support" email was the most appropriate to provide help. ProFlow immediately predicted that it would increase these customers' product usage and had a good chance of converting them to purchase. The analysts added this email in the plan and zoomed out to the whole group. With the plan, ProFlow estimated that the purchase rate of this group would increase by 12.7%, from 32.5% to 45.2%.

Feedback
Overall, all domain experts agreed ProFlow is a useful predictive analytics tool and were willing to adopt it in their daily workflow. In this section, we summarize the experts' feedback and discuss future directions.

Event Sequence Prediction: All the analysts liked ProFlow's support for predicting future events, as E1 said *"behavior predictions can tell me what customers are likely to do and how I can serve them"* and that *"most existing tools only predict aggregated scores."* E4 who built marketing applications was impressed by ProFlow's capability of aggregating and analyzing multiple records and commented: *"Most predictive models are designed to handle a sequence at a time. My clients always wished to have a user interface to explore the prediction results of multiple sequences."*

Action Planning: Both E1 and E2 were excited about ProFlow's support for "what-if" analysis because *"it enables a data-driven way to make decisions."* E2 explained that *"it can help me better personalize customers' experience and adjust my team's A/B testing strategies"* and E1 added that *"data and statistics are especially helpful for identifying the real problems in hard situations when nothing looks obvious."* E3 and E4 also commented that *"combining predictions and intervention planning is a long-desired feature for marketing applications."*

Ease of Learning and Use: All the analysts agreed ProFlow is *"fairly easy to use given some training,"* and E1 emphasized that *"the visualization explains so much in such a compact format."* Several design considerations were applauded: *"it is helpful to show uncertainty explicitly since people may forget (E1)," "right align the flow by now makes a lot of sense (E2),"* and *"showing customer activity with circles and action plan with squares looks very intuitive (E4)."*

Limitations and Future Directions: E1 and E2 suggested adding visual encodings for categorizing customer activities of different marketing channels (e.g., website, product, email, offline), which would allow analysts to exclude or prioritize certain types of activities. They also wished to customize ProFlow's predictive models and include scores such as engagement and satisfaction along with the behavior predictions. E3 and E4 suggested further testing ProFlow in real commercial products, which have access to more customer records and fine-grained activity attributes.

Conclusion
This paper presents ProFlow, a visual analytics system for supporting analysts' workflow of exploring and predicting event sequences. We report on a case study evaluating ProFlow's applicability, usefulness, and limitations in a real-world marketing scenario. In future work, we will extend ProFlow's capabilities to support cross-channel analytics and customizable predictions. We will continue to evaluate ProFlow in real products using larger and richer datasets.

REFERENCES

[1] Bilal Alsallakh, Allan Hanbury, Helwig Hauser, Silvia Miksch, and Andreas Rauber. 2014. Visual methods for analyzing probabilistic classification data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1703–1712.

[2] Markus Bögl, Wolfgang Aigner, Peter Filzmoser, Tim Lammarsch, Silvia Miksch, and Alexander Rind. 2013. Visual analytics for model selection in time series analysis. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2237–2246.

[3] Charles Chen, Sungchul Kim, Hung Bui, Ryan Rossi, Eunyee Koh, Branislav Kveton, and Razvan Bunescu. 2018. Predictive analysis by leveraging temporal user behavior and user embeddings. In *ACM International Conference on Information and Knowledge Management*. 2175–2182.

[4] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2016. EventAction: Visual analytics for temporal event sequence recommendation. In *Proceedings of the IEEE Visual Analytics Science and Technology*. 61–70.

[5] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2017a. Finding similar people to guide life choices: Challenge, design, and evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 5498–5544.

[6] Fan Du, Ben Shneiderman, Catherine Plaisant, Sana Malik, and Adam Perer. 2017b. Coping with volume and variety in temporal event sequences: strategies for sharpening analytic focus. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1636–1649.

[7] Lorenz A Gilch and Sebastian Müller. 2018. On Elo based prediction models for the FIFA Worldcup 2018. *arXiv preprint arXiv:1806.01930* (2018).

[8] David Gotz and Harry Stavropoulos. 2014. DecisionFlow: Visual analytics for high-dimensional temporal event sequence data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1783–1792.

[9] Shunan Guo, Fan Du, Sana Malik, Eunyee Koh, Sungchul Kim, Zhicheng Liu, Donghyun Kim, Hongyuan Zha, and Nan Cao. 2019. Visualizing Uncertainty and Alternatives in Event Sequence Predictions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 573.

[10] Josua Krause, Adam Perer, and Enrico Bertini. 2014. INFUSE: Interactive feature selection for predictive modeling of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1614–1623.

[11] Josua Krause, Adam Perer, and Kenney Ng. 2016a. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 5686–5697.

[12] Josua Krause, Adam Perer, and Harry Stavropoulos. 2016b. Supporting iterative cohort construction with visual temporal queries. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 91–100.

[13] Bum Chul Kwon, Ben Eysenbach, Janu Verma, Kenney Ng, Christopher De Filippi, Walter F Stewart, and Adam Perer. 2018. Clustervision: Visual supervision of unsupervised clustering. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 142–151.

[14] Yafeng Lu, Rolando Garcia, Brett Hansen, Michael Gleicher, and Ross Maciejewski. 2017. The state-of-the-art in predictive visual analytics. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 539–562.

[15] Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. 2013. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2227–2236.

[16] Sugeerth Murugesan, Sana Malik, Fan Du, Eunyee Koh, and Tuan Manh Lai. 2018. DeepCompare: Visual and interactive comparison of deep learning model performance. In *Symposium on Visualization in Data Science*.

[17] Catherine Plaisant, Brett Milash, Anne Rose, Seth Widoff, and Ben Shneiderman. 1996. LifeLines: Visualizing personal histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 221–227.

[18] Catherine Plaisant, Rich Mushlin, Aaron Snyder, Jia Li, Daniel Heller, and Ben Shneiderman. 1998. LifeLines: Using visualization to enhance navigation and analysis of patient records. In *Proceedings of the AMIA Symposium*. 76–80.

[19] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. 2005. Interactive Sankey diagrams. In *IEEE Symposium on Information Visualization*. 233–240.

[20] Eric Siegel. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Wiley Hoboken.

[21] Hendrik Strobelt, Sebastian Gehrmann, Michael Behrisch, Adam Perer, Hanspeter Pfister, and Alexander M Rush. 2018. Seq2Seq-Vis: A visual debugging tool for sequence-to-sequence models. *arXiv preprint arXiv:1804.09299* (2018).

[22] Krist Wongsuphasawat and David Gotz. 2012. Exploring flow, factors, and outcomes of temporal event sequences with the Outflow visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2659–2668.

[23] Krist Wongsuphasawat, John Alexis Guerra Gómez, Catherine Plaisant, Taowei David Wang, Meirav Taieb-Maimon, and Ben Shneiderman. 2011. LifeFlow: Visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1747–1756.

[24] Emanuel Zgraggen, Steven M. Drucker, Danyel Fisher, and Robert DeLine. 2015. (s|qu)eries: Visual regular expressions for querying and exploring event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2683–2692.

[25] Xun Zhao, Yanhong Wu, Dik Lun Lee, and Weiwei Cui. 2018. iForest: Interpreting random forests via visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 407–416.