2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)

Next Likely Behavior: Predicting Individual Actions from Aggregate User Behaviors

Bernard J. Jansen
Qatar Computing Research Institute
HBKU
Doha, Qatar
jjansen@acm.org

Soon-Gyo Jung
Qatar Computing Research Institute
HBKU
sjung@hbku.edu.qa

Dianne Ramirez Robillos
School of Statistics
University of the Philippines Quezon City
Diliman, Philippines
diannerobillos@yahoo.com

Joni Salminen
Qatar Computing Research Institute, HBKU; Turku School of Economics
Doha, Qatar: Turku, Finland
jsalminen@hbku.edu.qa

***Abstract*— We report results using n-grams to model user actions with only aggregated data and knowing little about the user. Employing a data set of 33,860 flight bookings from 4,221 passengers, we evaluate the n-gram model for the precision of predicting next likely actions. Results show that our approach can achieve a precision of 21% overall and 88% for some user behavior patterns, which is well above a baseline of 11.8% of recommending the most popular destinations. We achieve this performance with minimal complexity by using a first-order model of two states. However, the coverage is limited to about 21% of the destinations. Implications are that the n-gram approach can predict short customer behavior patterns when individual customer information is not available for specific sequences of actions. The findings have consequences for understanding customer populations for tasks that can be state modeled within the airline domain and similar contexts. Findings may also apply to other user-facing tasks, such as predicting website transitions or next likely clicks.**

Keywords—user segmentation; customer segmentation; n-grams

I. Introduction

Much prior user modeling research has focused on predicting subsequent actions [1] of individuals [2]. However, for many user- and consumer tasks, focusing on the individual level may not be necessary or possible because of the individual data being unavailable. An example of the latter is aggregated data from social media platforms [12]; i.e., the platforms share the data in a format that is not personally identifiable but, at the same time, this makes it difficult for researchers and organizers to model such data.

In this research, to address this issue, we evaluate an n-gram approach to predict the *next likely behavior* (NLB) of a customer using solely aggregated behavioral data from the overall customer population. Knowing very little about the individual customers, we base predictions exclusively on aggregated behaviors of the entire set of customers. Our research is motivated by the premise that aggregate behaviors are, for some tasks, reliably accurate predictors of the NLB action of an individual customer. The NLB is the action that a customer is most likely to implement, even in situations where there is a multitude of next possible actions. These are the assumptions that motivate our research.

We test our assumptions on a dataset of flight bookings using a first-order model [9](i.e., two states, one transition). In the airline industry, there are several benefits in knowing a particular customer's next likely flight destination, including personalized promotions and prediction of load factors for flights. Our research findings show that the methodological approach of n-grams is highly predictive and computationally straightforward for some state patterns. However, we also find that the n-gram method is not applicable for the complete dataset, indicating the need for ensemble approaches for broader coverage of all NLB user patterns from a customer population [18].

In the following sections, we briefly review prior work and then present our research questions. We explain the use of a dataset to investigate behavior prediction patterns using n-grams and then present our results. We end the manuscript with implications for system design and future research.

II. Prior Works

In examining the user behavioral processes, various researchers have used the concept of a *state* to model the sequence of user interactions, including online searching [10], browsing [16], check-ins [8], conversational agents [17], or related activities [5]. In this research, we define a state as a particular action by a customer (i.e., an airline passenger). A chain of such states is a *pattern*. We are interested in predicting the next state in the pattern, which is the customer's NLB.

Prior research has identified user actions on a system typically and then classified these actions into states. With this information, one can build a map or matrix of possible transitions from one state to the next. Each pattern is a sequence of state changes. This use of states and transitions is a stochastic (i.e., unpredictable) process from which one can compare patterns of various lengths to test the significance (i.e., to determine what length of pattern predicts arrival at a specific state). Such stochastic processes [3, 13] are established as effective methods for analyzing patterns, including text [19], and authorship [15]. The n-gram model is an example of such a methodological approach.

However, the n-gram model has typically been employed in a constrained space in which the number of states is limited or where common patterns, such as the use of language, limits or constraints future individual user actions. Additional the focus has typically been on overall aggregate prediction of user action. There has been little investigation, as we do in this research, in (a) evaluating the effectiveness of the n-gram approach for predictive user actions in more unconstrained contexts such as the domain of flight destinations [11] (i.e., a customer can fly to any destination in the network) and (b) using overall aggregate NLB patterns to predict individual actions in a complex and challenging [18] customer context.

Therefore, it is unclear if n-grams would be beneficial in such environments. If it can be shown that n-grams are effective in these unconstrained environments and using aggregate data for individual customer prediction, there are

978-1-6654-2180-5/21/$31.00 ©2021 IEEE

significant benefits in employing such an approach. N-grams require little individual customer data, making it valuable in contexts with little customer data or when privacy concerns [6] limit the collection of individual customer data. These are the motivation for our research.

III. Research Questions

The following research questions (RQs) are addressed in this exploratory research study:

RQ1: *Can an n-gram model based on aggregate data accurately predict the next likely behavior (NLB) for individual customers?*

RQ2: *What percentage of customer action patterns can an n-gram based model effectively predict?*

For RQ1, we develop a probability transition matrix that provides the percentage of transitions among possible flight destinations. Then, we use an n-gram approach to determine what order of states provides the best prediction of future states.

For RQ2, we compare the state patterns with predictability above the baseline to determine what percentage of all state patterns have high predictability.

We discuss our research design in more detail in the following sections.

IV. Data Collection and Methodology

A. Data Collection

Our dataset is a collection of 33,860 flight bookings from 4,221 customers of a major international airline who are members of the airline's frequent flyer program. We ignore all other demographical information of the customer in order to determine what prediction precision and coverage of the data set we can achieve using the aggregated flight patterns solely. The data set consisted simply of unique customer id and the set origin and destination for each booking (e.g., CUS01, ORIGIN01-DESTINATION01, ORIGIN02-DESTINATION02, etc.). There were 166 unique destinations in the dataset. The average number of booking per customer was 9.16 (SD = 6.6, max = 50, min =1).

B. Methodology

We first segment the flight bookings by place of origin (i.e., all flights originating from New York, Chicago, Washington, etc.). We then generate state patterns for each customer. For example, New York City (NYC) flights to some DESTINATION (D) for a user (U) results in a flight pattern of (NYC -D01, NYC -D02, NYC -D03, etc.).

We then remove the place of origin, resulting in a state destination sequence (D01-D02-D03, etc.) for each customer. We then divide each into n-grams; for example, a two-state n-gram for customer U would be (D01-D02, D02-D03, etc.).

We develop a program to automatically generate the transition probabilities. Once we execute the segmentation algorithm against the flight dataset, we then develop a probability transition matrix using a stochastic approach.

C. Stochastic Model

A stochastic model mathematically describes the sequence of states through which users progress via transition probabilities. The value in each cell of a transition probability matrix is the probability of going from the row state to the corresponding column state. Therefore, a transition probability matrix describes a pattern of movements through state space. A stochastic model is both descriptive and predictive. The number of states that one uses to predict future states is referred to as the order of the stochastic model. A zero-order stochastic process refers to the probability of being at a single state (i.e., no prediction). A first-order stochastic process refers to the probability of arriving at a specific state given a prior state.

Conceptually, for this research, the probability matrix is a table of all customer flight bookings. This table is a matrix of transactions from one state (i.e., D_i) to another state (i.e., D_j). An analysis of customer behavior in this manner describes a flight state (i.e., the destination) at a given point in the sequence and suggests which states are most likely to follow one another (i.e., what destination state will come next).

We use n-grams to develop probabilistic patterns. N-grams are a probabilistic modeling approach used for predicting the next item in a sequence; they are $(n-1)$ order Markov models, where n is the gram (i.e., subsequence or pattern) from the complete sequence or pattern. An n-gram model predicts state x_i using states $x_{i-1}, x_{i-2}, x_{i-3}, \ldots x_{i-n}$. The probabilistic model is then presented as: $P(x_i|x_{i-1}, x_{i-2}, x_{i-3}, \ldots x_{i-n})$, assuming that the next state depends only on the last $n-1$ states, which is, again, an $(n-1)$ order Markov model.

The probability matrix for the entire dataset allowed us to address RQ1 and RQ2. With the probability matrix of all occurrences of a given pattern, we could calculate the state transitions during each session at any order. Mathematically, for some model M_S, let S represent a number of states; then the set of states is $S = (0, 1 \ldots n)$, where n is the maximum pattern length for the longest session state transitions in the dataset. We used the maximum requested next state (NP_{max}) at each S as the predicted outcome for that state transition probability.

Using this approach, we can then make predictions on a user's next state transition. Additionally, this straightforward approach has the advantage of being easily implementable. In addition, it allows for evaluating which order of the n-gram is the best predictor for query reformulations within the dataset.

As an example of our algorithm's output, assume the current flight pattern log (L_{fp}) consists of the pattern "ABCDE," and the order of the pattern is three. In this case, the prediction algorithm checks L_{fp} for the pattern "ABC", and the next possible states are "D" and "E," respectively. If the order of the pattern is two, then the algorithm would check for the pattern "AB," "BC," and "CD," returning "C," "D," and "E," respectively. For an order of four, "ABCD" would predict "E."

D. Evaluation of Methodology

For evaluation of the NLB prediction algorithm, we were interested in the metric of (a) the *precision* of the predictions, along with (b) the *coverage* of the user population's state patterns.

As an illustration, consider a booking log of flight patterns consisting of six individual records (see Table 1), each of which represents a flight booking sequence containing state modification patterns: Each state (A, B, C, D, or E) in represents a flight destination booked in the NLB pattern. Using these six sessions, a second-order NLB model (i.e., two

states to predict the third) would generate the n-gram patterns in Table 2.

TABLE I. FLIGHT LOG OF DESTINATIONS (A, B, C, D, E, F).

Log of Flight Bookings (Destinations A though F inclusive)
ABCF
ABCDE
ABCDE
A
AB
AC

TABLE II. N-GRAM NLB PATTERN TABLE OF STATE = 3 (I.E., TWO STATES TO PREDICT STATE THREE).

Predictive pattern	NP_{max}	P@1
AB	C	100%
BC	D	66%
CD	E	100%

Tables 1 and 2 show that our precision of prediction varies from 66 to 100%. However, there also are three sessions that our model does not represent (i.e., "A," "AB," and "AC"), since the length of these patterns is less than $S + 1$. Consequently, the *coverage* (i.e., the percentage of the collection that the path length addresses) is less.

Therefore, we evaluate our prediction algorithm in terms of two metrics: *precision and coverage*. Concerning precision, we calculate precision at 1 (P@1) for each 2-state n-gram by recommending the state D_n from the highest occurrence n-gram with state D_{n-1} (i.e., *how accurately can we predict the single next destination for the customer based on the prior destination?*).

We also calculated P@1 for 3-state n-grams, 4-state n-grams, etc. From our initial evaluation, n-grams of 3 or more states provide prediction precision below the baseline for many patterns, so for the remainder of this research, we focus exclusively on the two-state NLB n-grams model.

Coverage is a measure of the model's applicability to the entire dataset (i.e., how much of the dataset a model of a particular order can accurately address). This is a metric of how much of the entire set of existing patterns that our approach addresses. The idea being that the approach may be highly accurate for a set of patterns; however, if the approach does not address a reasonable percentage of the dataset, it may not be a worthwhile pursuit in terms of practicality.

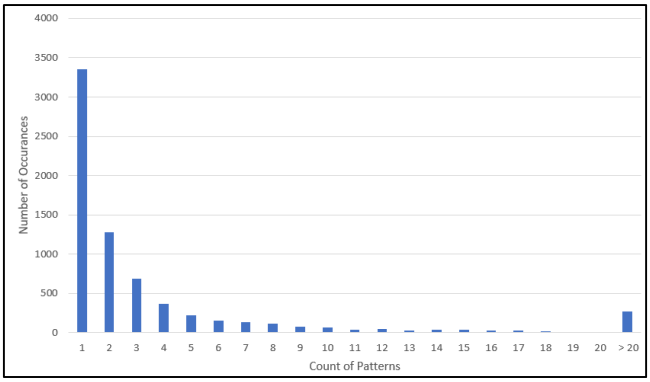

Fig. 1. Distribution of occurrence of two-state patterns. For example, 1 means the pattern occurs one time, and there are 3,357 such patterns. 20 means the pattern occurs 20 times in the dataset, and there are 10 such patterns.

TABLE III. COUNT OF THE OCCURRENCE OF TWO-STATE PATTERNS. FOR EXAMPLE, 1 MEANS THE PATTERN OCCURS ONE TIME, AND THERE ARE 3,357 SUCH PATTERNS. 20 MEANS THE PATTERN OCCURS 20 TIMES IN THE DATASET, AND THERE ARE 10 SUCH PATTERNS.

Count	No.	%	Cumulative
1	3357	47.54%	47.54%
2	1276	18.07%	65.61%
3	685	9.70%	75.31%
4	367	5.20%	80.51%
5	223	3.16%	83.67%
6	158	2.24%	85.91%
7	137	1.94%	87.85%
8	118	1.67%	89.52%
9	78	1.10%	90.62%
10	72	1.02%	91.64%
11	39	0.55%	92.19%
12	53	0.75%	92.94%
13	33	0.47%	93.41%
14	39	0.55%	93.96%
15	36	0.51%	94.47%
16	29	0.41%	94.88%
17	29	0.41%	95.29%
18	19	0.27%	95.56%
19	14	0.20%	95.76%
20	10	0.14%	95.90%
> 20	269	3.81%	100.00%
	7061	100.00%	

V. RESULTS

Evaluation of the data set at State 0 (i.e., the 1st flight) provides us a baseline for evaluation by ranking all destinations. Since we are examining P@1, we would get a P@1 of 11.8% across the entire dataset if we recommend the most popular destination. Therefore, we use this as the baseline for comparison of our n-gram effectiveness. Overall, we have 7,061 two-state transitions representing 38,669 flights, as shown in Table 3 and Figure 1.

A. Two State N-grams

Our approach identifies the 1st state and then recommends the topmost occurring 2nd state, although the 1st state could have more than one 2nd state. Taking this approach, we have 166 predictive patterns for evaluation (2.35% of all two-state patterns), indicating that the span of user action in this context is quite varied.

As shown in Table 4, we have several two-state NLB n-grams with 100% precision (due to single occurrences); however, even two-state patterns that are quite common have high occurrences, well above baseline.

In order to focus on the most impactful patterns, Table 5 shows the topmost frequently occurring two-state patterns and the prediction precision. Overall, the average P@1 was 21.1% (max = 100%, min = 7.2%, SD = 17.0%, median = 15.7%). We have 132 (79.5%) two-state transitions that exceed the baseline and 34 at or below the baseline (20.5%). We can accurately predict 21.6% (7,976) of NLB flights bookings

using a two-state approach by assuming the individual customer will adhere to the overall aggregate flight patterns.

B. Two State N-grams Coverage

Overall, we have 7,061 two-state transitions, and we have 166 predictive patterns. Therefore, our model covers 2.04% of all possible patterns and 7,976 (21%) of all flights. However, again, this exceeds our baseline of recommending the most popular destination, covering only 0.6% of destinations and only 11.8% of flights, which is reasonable given the state complexity of the dataset (i.e., there are lot of states and each state is connected to every other state),

TABLE IV. THE TOP TWO STATE NLB PATTERNS IN TERMS OF PRECISION. ACTUAL DESTINATIONS MASKED FOR COMPANY CONFIDENTIALITY.

State01	State02	Count	Percentage
Dest16	Dest13	1	100.0%
Dest17	Dest19	1	100.0%
Dest21	Dest25	1	100.0%
Dest22	Dest14	1	100.0%
Dest24	Dest3	1	100.0%
Dest10	Dest9	3	66.7%
Dest6	Dest6	51	51.0%
Dest1	Dest14	2	50.0%
Dest11	Dest2	2	50.0%
Dest20	Dest7	2	50.0%
Dest8	Dest18	61	47.5%
Dest26	Dest18	9	44.4%
Dest23	Dest23	32	43.8%
Dest4	Dest4	1888	41.4%
Dest15	Dest15	11	36.4%
Dest12	Dest14	347	34.9%

TABLE V. THE TOP TEN TWO-STATE NLB PATTERNS IN TERMS OF OCCURRENCES.

	State01	State02	Count	Percentage
1	Dest18	Dest18	4580	25.6%
2	Dest26	Dest18	3812	25.6%
3	Dest24	Dest24	2664	23.7%
4	Dest28	Dest28	2154	24.4%
5	Dest6	Dest6	1888	41.4%
6	Dest12	Dest26	1375	22.9%
7	Dest21	Dest18	1064	14.1%
8	Dest4	Dest18	1041	19.9%
9	Dest9	Dest18	943	20.6%
10	Dest2	Dest2	775	19.4%

C. Three State N-grams

We also explored tri-grams. In this approach, we identify the 1st state and 2nd state, and we then recommend the topmost occurring 3rd state, although the 1st state and 2nd state pattern could have more than one 3rd state. Taking this approach, we have 144 predictive patterns for the evaluation of 21,033 unique patterns (0.07% of all three-state patterns). Given the rather low coverage of the dataset, we did not explore these longer patterns further.

VI. DISCUSSION AND IMPLICATIONS

In this research, we have demonstrated that using a straightforward and state-based NLB approach with no individual customer information, and we can achieve quite high precision for predicting the NLB for P@1 for some NLB user patterns. Overall, our NLB n-gram model achieves approximately 21.1% precision for all flight bookings and covers 21.6% of all flights, with minimal complexity and no personally identifying information concerning any customer. Findings show that in non-constrained spaces, n-grams most likely need to be combined with other predictive approaches [4, 7, 14] in an ensemble strategy. The results suggest that the more popular specific transitions collectively are, the easier it is to predict next likely action of an individual customer.

VII. CONCLUSION

This research implements an n-gram approach to model individual customer behavior leveling only aggregate level data. Findings show that a second-order model performs reasonably well in terms of prediction precision while still providing good dataset coverage. The implications are that one can integrate the detection of customer behavior patterns for which NLB n-gram modeling would be effective.

For future research, we aim to expand the number of states and exploring a combination of predictive states for different origins to determine a set of n-gram orders that provides the best combination of precision, coverage, and complexity, along with the attributes for selecting the set of n-grams.

REFERENCES

[1] L. Boratto, G. Fenu, and M. Marras, "Connecting user and item perspectives in popularity debiasing for collaborative recommendation," *Information Processing & Management,* vol. 58, p. 102387, 2021/01/01/ 2021.

[2] T. Chovanak, O. Kassak, and M. Bielikova, "Behavioral Patterns Mining for Online Time Personalization," presented at the Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 2017.

[3] E. Cinlar, *Introduction to Stochastic Processes*. New York: Dover Publications, 2013.

[4] B. D. Davison and H. Hirsh, "Predicting sequences of user actions," 1998.

[5] İ. Dönmez, "Human Activity Analysis and Prediction Using Google n-Grams," *International Journal of Future Computer and Communication,* vol. 7, pp. 32-36, 2018.

[6] J. Golbeck, "I'll be Watching You: Policing the Line between Personalization and Privacy," presented at the Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 2017.

[7] T. Gurbanov, F. Ricci, and M. Ploner, "Modeling and Predicting User Actions in Recommender Systems," presented at the Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, Halifax, Nova Scotia, Canada, 2016.

[8] H.-P. Hsieh, C.-T. Li, and X. Gao, "T-Gram: A Time-Aware Language Model to Predict Human Mobility," 2015.

[9] J. J. Hull and S. N. Srihari, "Experiments in text recognition with binary n-gram and viterbi algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 5, pp. 520-530, 1982.

[10] B. J. Jansen, D. L. Booth, and A. Spink, "Patterns of query modification during Web searching," *Journal of the American Society for Information Science and Technology,* vol. 60, pp. 1358-1371, 2009.

[11] B. J. Jansen, S. G. Jung, D. Ramirez Robillos, and J. Salminen, "Too Few, Too Many, Just Right: Creating the Necessary Number of Segments for Large Online Customer Populations," *Electronic*

Commerce Research and Applications, vol. 49, p. Article 101083, 2021.

[12] B. J. Jansen, J. Salminen, and S. G. Jung, "Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics," *Data and Information Management,* vol. 4, pp. 1-17, 2020.

[13] H. Lin, Y. Zuo, G. Liu, H. Li, J. Wu, and Z. Wu, "A Pseudo-document-based Topical N-grams model for short texts," *World Wide Web,* vol. 23, pp. 3001-3023, 2020/11/01 2020.

[14] D. Noy, "Predicting user intentions in graphical user interfaces using implicit disambiguation," presented at the CHI '01 Extended Abstracts on Human Factors in Computing Systems, Seattle, Washington, 2001.

[15] Y. Seroussi, I. Zukerman, and F. Bohnert, "Collaborative inference of sentiments from texts," in *18th International Conference on User Modeling, Adaptation, and Personalization, UMAP 2010*, Big Island, HI, USA, 2010, pp. 195–206.

[16] Z. Su, Q. Yang, Y. Lu, and H. Zhang, "WhatNext: a prediction system for Web requests using n-gram sequence models," in *Proceedings of the First International Conference on Web Information Systems Engineering*, Hong Kong, China, 2000, pp. 214-222.

[17] H. Sun, D. Cheng, J. Wang, Q. Qi, and J. Liao, "Pattern and content controlled response generation," *Information Processing & Management,* vol. 58, p. 102605, 2021/09/01/ 2021.

[18] S. Thirumuruganathan, J. Salminen, S. G. Jung, D. Ramirez Robillos, and B. J. Jansen, "Forecasting the Nearly Unforecastable: Why Aren't Airline Bookings Adhering to the Prediction Algorithm?," *Electronic Commerce Research,* vol. 21, pp. 73–100, 2021.

[19] W. R. Wright and D. N. Chin, "Personality Profiling from Text: Introducing Part-of-Speech N-Grams," in *User Modeling, Adaptation, and Personalization. UMAP 2014*. vol. 8538, V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, and G. Houben, Eds., ed Cham: Springer, 2014.