LMU Munich
Frauenlobstraße 7a
D–80337 München
Institut für Informatik

Master Thesis

Intent Prediction with Vectorized Sequential UI Tree Data

August Oberhauser

august.oberhauser@campus.lmu.de

Course of Study: Medieninformatik

Examiner: Prof. Dr. Sven Mayer

Supervisor: Florian Bemmann, M.Sc.

Commenced: March 1, 2023

Completed: Juli 1, 2023

Kurzfassung

<Short summary of the thesis>

Abstract

The interaction of a user with an end device such as a smartphone or a computer is very diverse and difficult to predict. Nevertheless, user-specific (personalized) as well as global (collaborative) patterns can possibly be worked out with the help of preceding user interactions. These could be used to predict the intention of a user or a group of users. It is interesting to know to what level of detail these predictions can be made reliably. By making use of the continous on-device data an attempt can be made to gain more insights in the user behavior or even forecast their next actions.

It suggests itself to implement this with the help of user interactions in sessions on Android devices. For this purpose, the Sequential UI Tree data of the device could be tracked, filtered and labeled and then trained with a machine learning model to find similar interaction sequences and then make predictions. These can then be very coarse, such as predicting the next app. Or they can be very detailed, e.g., determining the next user action, such as filling out a form field.

A concept will be developed on how a model for predicting user intent could be built and how it could be applied to the user session. To this end, possibilities for collecting and vectorizing sequential UI trees (e.g., from the Android Accessibility Service) will be discussed (e.g., via Recurrent Neural Network (*RNN*) [10] [1] [5], *Seq2Seq* Model [2], *Screen2Vec* Model [8], *Intention2Text* [13], *Html2Vec* [12]), which are designed to predict the user intent. Here, privacy and feature pre-filtering in UI data plays an important role. After that, personalized as well as collaborative data can be used in a hybrid approach. This model should then be made available to the user in an Android app service and, depending on the level of detail, suggest upcoming apps or actions to the user at a suitable time. It should also be considered whether the user can contribute to the learning process and improve suggested actions through feedback (labeling). The performance of the model can be measured, for example, by indicators such as the amount of training data and time spent on the learning process. The effectiveness can be evaluated by accuracy metrics in predicting, for example, app categories [9] or complete test sequences via Rico [3] or ERICA [4].

Furthermore, the machine learning model could provide the following benefits in addition to intent prediction:

- reduction of the complexity and size of the UI tree
- creation of user groups that have similar behavior when using digital UI systems [7]
- elimination of technical expertise on individual features that would be required to manually compare user sessions [6]
- consideration of a user's history over time (sequential)
- comparison of user interactions without providing privacy invasive information
- supporting app developers to improve their app design and usability
- application in psychology and market research
- pre-loading of processes on devices (energy savings) [11]

As listed above, many fields of application can profit by elaborating such a system. It would be exciting to know, how the concrete concept would look like and if it can be implemented successfully e.g. to improve the user experience on end-user devices.

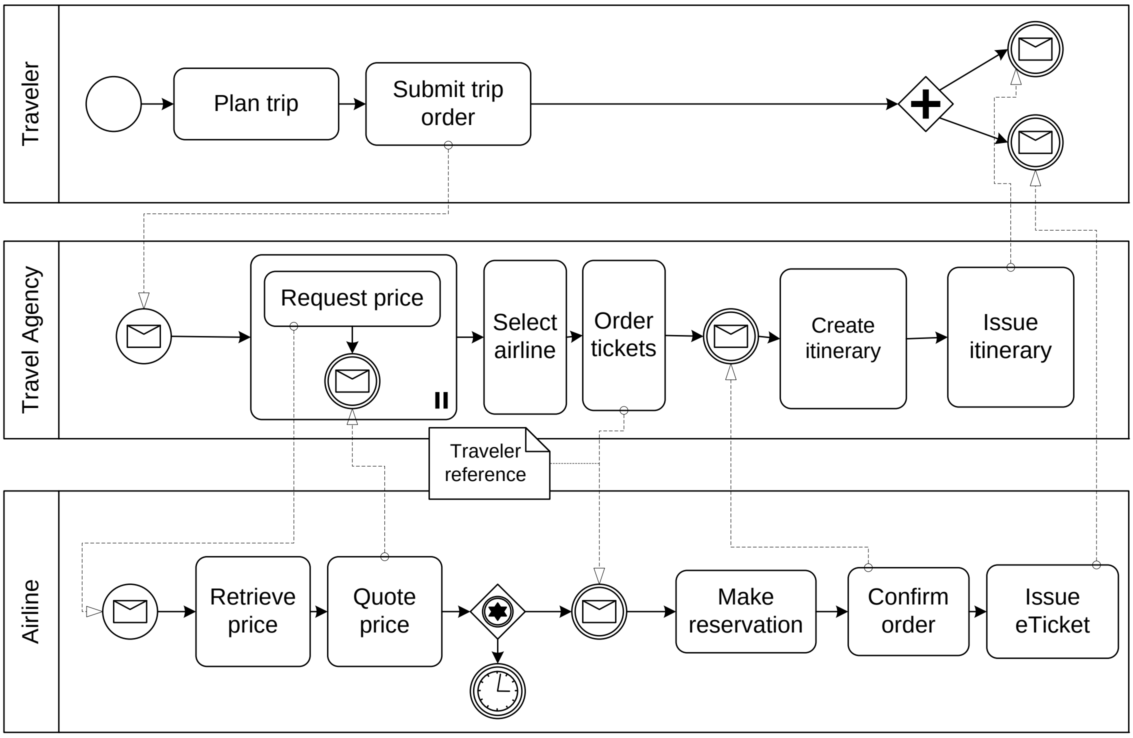

Figure 1: Possible procedure using a Machine-Learning algorithm to predict the next intent from a beginning user session: The input (1) can be a sequence of Android tree data. With help of a Machine-Learning-Model (2) (e.g. RNN) a vector representation can be trained and then predict the most probable action or screen (3) from a given starting sequence, but also can be improvde through the users feedback.

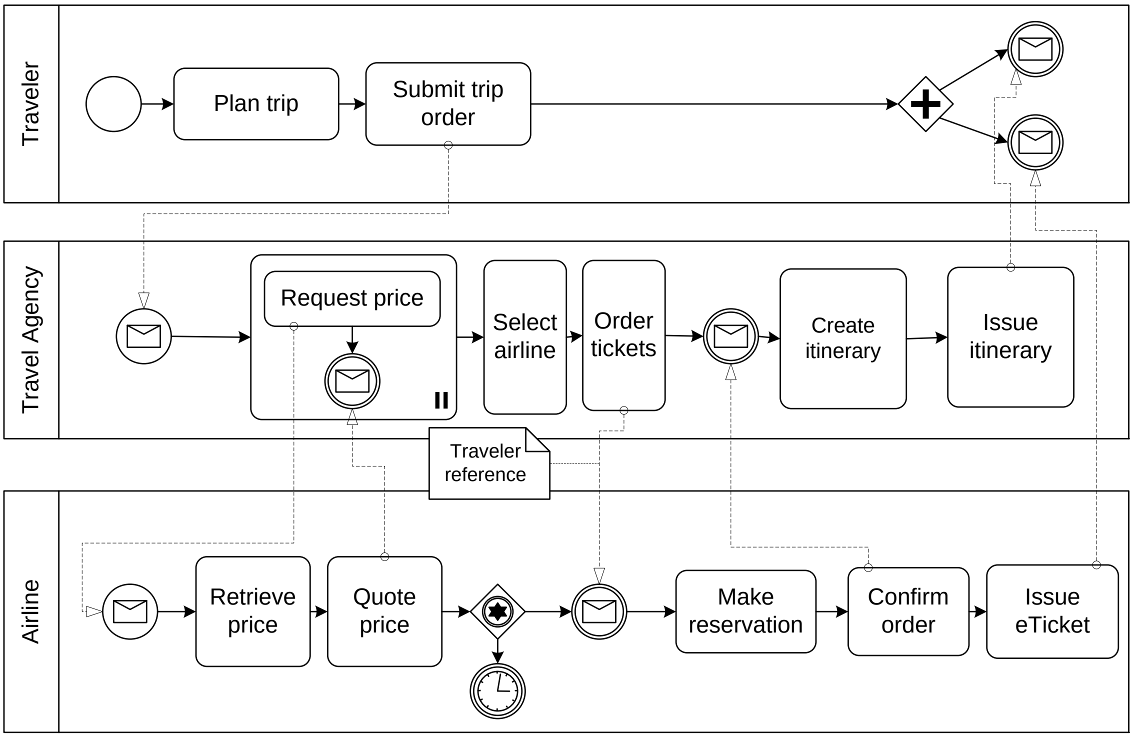

Figure 2: Schedule as a Gantt Chart

Contents

List of Figures

List of Tables

1 Introduction

This is a typical human-computer interaction thesis structure for an introduction which is structured in four paragraphs as follows:

1.1 Motivation for transforming Android UI tree data to vectors

P1.1. What is the large scope of the problem?

P1.2. What is the specific problem?

P2.1. The second paragraph should be about what have others been doing

P2.2. Why is the problem important? Why was this work carried out?

P3.1. What have you done?

P3.2. What is new about your work?

P4.1. What did you find out? What are the concrete results?

P4.2. What are the implications? What does this mean for the bigger picture?

2 Related Work

Describe relevant scientific literature related to your work.

a) Doc2Vec and Word2Vec b) Tree2Vec c) Html2Vec d) Activity2Vec

2.1 Vector models

2.1.1 Seq2Seq Model

2.2 Time Series / Sequence models

2.2.1 RICO / RicoSCA

"Rico is a public UI corpus with 72K Android UI screens mined from 9.7K Android apps. [...] We manually removed screens whose view hierarchies do not match their screenshots by ask- ing annotators to visually verify whether the bound- ing boxes of view hierarchy leaves match each UI object on the corresponding screenshot image.This filtering results in 25K unique screens."

2.2.2 Screen2Vec

2.2.3 Screen2Words

2.2.4 Intention2Text

2.2.5 Html2Vec

3 Methodology

3.1 Android UI Data

3.1.1 Data tree structure

3.1.2 Retrieval of UI data via Android Accessibility Service

3.2 Machine Learning

3.2.1 Preprocessing

3.2.1.1 Normalization, Feature selection

Such as Filtering privacy invasive details

Parameterizing the vectorization process a) Vector length b) Weighting of features c) Manipulating individual parameters of model

3.2.2 Supervised vs Unsupervised vs Semisupervised

3.2.3 Under and Overfitting

3.2.4 Evaluation Metrics

3.2.5 Neuronal Nets

Activation Functions Cost function Gradient - Regression: Continous Values - Classification: Multiple class - One Class

Tensors

LSTM 4 dimensional

Embedding before LSTM

TimeDistributedLayer

3.2.5.1 Deep Neuronal Nets

3.2.5.2 Convolutional Neuronal Nets

3.2.5.3 Recurrent Neuronal Networks and LSTMs / GRU

3.2.6 Layers

- Embedding layer Dimension near the actual average length of features - Dense Layer

Positive Integer to Dense Vectors of fixed size

3.2.6.1 Autoencoders

4 Results

4.1 Datasets

- Problem with sequential data sets

4.1.1 Rico

- Too less frames.
- No transition between apps.

4.2 Preprocessing Android UI tree data

4.2.1 Filtering privacy invasive details

4.2.2 Normalization, Feature selection

Dealing with variable length data tf.io.VarLenFeature()

4.3 Evaluation

Multiple approaches

AutoEncoder:

- Encoder -> Decoder -> LSTM -> Decoder - Encoder -> LSTM -> Decoder - LSTM -> Encoder -> Decoder (AutoEncoder)

Decoder can either only decode to x and y or to whole UI tree.

4.3.1 Mean Squared Error

4.3.2 F1 Score

5 Application of Android UI tree vectors

5.1 Automation and testing of Android apps

5.2 UI design similarities

5.3 Action prediction models, User behavior modeling

5.4 Behavioral analyses for smartphone usage patterns

6 Conclusion and Future Work

Summary

Outlook

Future directions for research in this area

ChatGPT – Image Recognition – Limitation als Ausblick

References

[1] Saumya Bansal, Niyati Baliyan. "Remembering past and predicting future: a hybrid recurrent neural network based recommender system". In: *Journal of Ambient Intelligence and Humanized Computing* (2022), pp. 1–12.

[2] François Chollet. *A ten-minute introduction to sequence-to-sequence learning in Keras*. 2017. URL: https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html (visited on 02/15/2023).

[3] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, Ranjitha Kumar. "Rico: A mobile app dataset for building data-driven design applications". In: *Proceedings of the 30th annual ACM symposium on user interface software and technology*. 2017, pp. 845–854.

[4] Biplab Deka, Zifeng Huang, Ranjitha Kumar. "ERICA: Interaction mining mobile apps". In: *Proceedings of the 29th annual symposium on user interface software and technology*. 2016, pp. 767–776.

[5] Mauro Di Pietro. *Modern Recommendation Systems with Neural Networks*. 2022. URL: https://towardsdatascience.com/modern-recommendation-systems-with-neural-networks-3cc06a6ded2c (visited on 03/08/2023).

[6] Alireza Ghods, Diane J Cook. "Activity2vec: Learning adl embeddings from sensor data with a sequence-to-sequence model". In: *arXiv preprint arXiv:1907.05597* (2019).

[7] Kasthuri Jayarajah, Youngki Lee, Archan Misra, Rajesh Krishna Balan. "Need accurate user behaviour? pay attention to groups!" In: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 2015, pp. 855–866.

[8] Toby Jia-Jun Li, Lindsay Popowski, Tom Mitchell, Brad A Myers. "Screen2vec: Semantic embedding of gui screens and gui components". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–15.

[9] Google LLC. *Choose a category and tags for your app or game*. 2023. URL: https://support.google.com/googleplay/android-developer/answer/9859673 (visited on 02/15/2023).

[10] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, Paolo Cremonesi. "Personalizing session-based recommendations with hierarchical recurrent neural networks". In: *proceedings of the Eleventh ACM Conference on Recommender Systems*. 2017, pp. 130–137.

[11] Zhihao Shen, Kang Yang, Wan Du, Xi Zhao, Jianhua Zou. "Deepapp: a deep reinforcement learning framework for mobile application usage prediction". In: *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 2019, pp. 153–165.

[12] Peihuang Wu, Jiakun Zhao. "Distributed representations of html page". In: *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*. IEEE. 2022, pp. 160–164.

[13] Che-Hsuan Yu, Hung-Yuan Chen, Fang-Yie Leu, Yao-Chung Fan. "Understanding mobile user intent using attentive sequence-to-sequence RNNs". In: *Wireless Internet: 12th EAI International Conference, WiCON 2019, TaiChung, Taiwan, November 26–27, 2019, Proceedings 12*. Springer. 2020, pp. 51–61.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature