Predicting User Behaviour in Web Applications

Summary of the master thesis "Predicting Navigational Patterns in Web Applications using Machine Learning Techniques" (2024)

Patric Balan & Gustav Jönemo

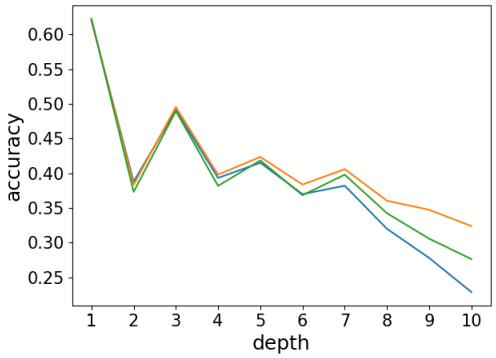
February 2024

We all know the frustration of using a computer program you know nothing about. You wish someone would help you, why not the program itself?

The foundation of the thesis is to make predictions, using machine learning, on what users are about to do when using a web application. This process involves collecting millions of log entries from the application, which are documents that contain information about how the application is being used. These entries are then processed in multiple steps, where data is cleaned, sorted, and interpreted. This process of collecting web usage logs and extracting important information is called "web usage mining". In the thesis, four different sets of data were used, based on a single original data set but processed in different ways depending on the use case. On each of these data sets, machine learning models could then be trained. The three different models used in this project are all based on artificial neural networks, which mimic how the brain functions, with many interconnected neurons that interact in complex ways. When a machine learning model is trained, it looks at one part of a data set and tries to predict another, checking if it got it right as it goes along. Each time a prediction is made, it compares the result to previous ones, adjusting its network so that it always keeps changes that result in better performance. This way, the model can learn complex patterns that would otherwise be difficult for a human to understand.

The models were trained on web logs from a customer support web tool. A customer support employee needs to be able to provide help to a customer without wasting time in the web tool. The need for an efficient workflow in customer support gave rise to the approach of predicting future user inputs, since this could help guide support staff when using the tool so that they can serve customers faster and more effectively.

The models are fed sequences of actions, where an action refers to an interaction with the web tool, such as a click or navigation. Furthermore, they process these sequences in two different ways. They can either look at the next action, always trying to predict one step ahead, or they can look at the last action, which means looking many steps into the future, trying to predict where the user wants to end up. Testing was performed on both of these methods. The best next-action model achieved an accuracy of over 75% in predicting the next user action and the best result for predicting a user's last action reached 62%. Accuracy varies depending on how far into the future the predictions are made, as seen in the figure below.



Accuracy as a function of depth, which is the number of steps into the future predictions are made. The different lines correspond to the performance of different models.

This result is promising, but future work might be needed to assess the techniques used in the thesis on larger web tools, as the tool used proved to be rather small. To build on this, future testing should be performed to measure the efficiency gained from using the model's predictions.

Predicting future inputs could also be used on other web platforms to give the user a better and more personalized experience. For example, the web page could be altered to better match user behavior and even improve the loading time by pre-loading web resources.