DSC180A Quarter 2 Project Proposal

Lance Zhu
University of California, San Diego
La Jolla, CA 92093
lazhu@ucsd.edu

Eric Zhang
University of California, San Diego
La Jolla, CA 92093
zezhang@ucsd.edu

Abstract

In the continuous of Quarter 1 data collecting project, Quarter 2 will be focusing on **building predictive models for recommendation system in windows applications, readjustment of foreground windows, and removing pop-up harassment windows and massages.** We will apply model such as LSTM and Markov models.

1 Introduction

Nowadays, technologies have become an inseparable part of our life. Especially, the invention of computers and windows applications provided a great convenience for users so that we can better utilize them and perform routine tasks such as filing, data processing, and even photo editing, at a much more efficient pace. Later days with the development of artificial intelligence, people start to think about whether we can maximizes the capability of computers to help us achieve harder tasks such as opening an application in advance prior to telling the computer to do so. And that's the goal we want to accomplish at the end for this project.

We will make predictions on which applications does the user need at certain time, and then before the user click the application and wait for it to start, we can program a model to send this signal to the computer and open the application earlier to save the user's time and we can make a short list of recommended applications which users may want to open.

Based on the data of the position of a given window of the application, we can predict the position of the window of a specific application so that when a user open the application next time, we can also adjust the size and position of the window automatically according to the user's habit so that the user does not have to spend time on adjusting the window.

In addition, we plan on enhancing a ad-closing feature where it would detect and close when we are browsing on the internet, we often find advertisements and harassment messages pop up in a window. Sometimes, those windows appear again and again which are very annoying. By analyzing the time of an opening window and other information of an opening window, we can determine whether the window is an advertisement or harassment message to the user or not. If it is, we can mark it as "spam" and next time if the window pops up, the system will help cancel the window as soon as possible so the user will not be disturbed by it.

2 Methodology

2.1 Preparation

First, we will collect the data we require using the input library we have already built and tested in our quarter 1 project. We will use the foreground window input library to collect the data for two to three week to make sure we have enough data for the future data analysis and model training.

Then, we will start to clean the data and make analysis on the data we collected. First, we need to convert the data we collected using the C language based input library into a dataset that we can process using python. Such as we need to convert the timestamp, and calculate the duration time for each line of data, and we need to clean the columns and get rid of the columns that we don't need for our prediction model.

2.2 Application prediction using Markov models

Next, with the cleaned dataset, we can start to train our prediction model. Our goal is by predicting the application launch pattern of a user, we can recommend an application or several applications that the user is going to open so the user can open the application he wants faster and save his time as well. We can achieve this by using Markov models. Markov model is a stochastic model used to model a changing systems, which assumes that the future states depend on the current states. The Markov models is a good tool to help us achieve our predictive task. We started to build the Markov model for constructing the transition matrix. Then, we calculated the conditional probability of an opened application and its previous application. We are able to find out the most likely application the user is going to open based on the current opened application.

2.3 Foreground window readjustment

Continuously, based on the name and image of an application as well as the position(top left, top right, bottom left, bottom right) of a window, we can fit the data we collected to the prediction model using Keras. Then, we can predict the size and position of a window for a specific application the user will potentially want to put and the system will resize and reposition the window as soon as the user opens it.

2.4 Removing pop-up harassment windows

Lastly, we are going to determine whether the window is an advertisement or harassment message and help the user to cancel that window. Most of the time, the user will cancel the window as soon as this kind of window shows up, so the time duration of this opening window is very short. In this situation, we will assign a 'spam' score to it. If the window with the same application name, image, and class shows up multiple times and was canceled immediately multiple times, we will assign a higher and higher 'spam' score to the window. If it passes a threshold we set, this window would be classified as 'spam'. When the window shows up next time, it would be canceled by the system as soon as possible.

3 Literature

There aren't many conclusions in this field of study, and usually most of the predictions could be private or personal, and the models can also be targeted. There are some researchers who have studied about predicting user actions, as the researchers discuss in their paper A Blended Deep Learning Approach for Predicting User Intended Actions. They collect the data about user attrition and retention for certain applications and make models to track the critical factors for users and predict the chance a user will decide to stop using their service. But speaking about making predictions on computer users' activities, there are not many results, and that is why we think it will be useful to make such a model to save time for computer users. We have already been working on this topic in the quarter 1 project. We have already successfully built an input library that we will use to collect the data we need for the next step of research.

4 Individual Responsibilities

Langrun Zhu will be responsible for cleaning the dataset we collected using the foreground window input library, applying validation methods on the dataset to create training and testing dataset for the next step, and training different models to predict the user App launch pattern using packages such as Keras and Tensorflow. Langrun will also construct Markov model, build transition matrix,

and output the conditional probability of the Markov model with the data we collected. Also, Langrun is responsible for analyzing the position information of the foreground windows and building predictive models out of it.

Zeming Zhang will be responsible for collecting dataset using the foreground window input library and perform Exploratory data analysis with the data and represent it to the class. Zeming will make graphs to visualize and demonstrate the information of the data in various aspects. Also, Zeming will construct a Markov model in order to make predictive tasks with Langrun. Zeming is also responsible for testing the window position predictive models. Moreover, Zeming should build a function to detect the 'spam' window by setting up 'spam' score threshold, implementing classification function.

5 References

*T*an, F., Wei, Z., He, J., Wu, X., Peng, B., Liu, H., & Yan, Z. (2018). A Blended Deep Learning Approach for Predicting User Intended Actions. doi:10.48550/ARXIV.1810.04824