Methods Assignments: M4

Redesigning Youtube Recommendation System

Gaurav Pande
gpande@gatech.edu

Abstract. We all want to design things that touch our lives somewhere so that we can interact with them better. I want to design a better intuitive recommendation system for the Youtube mobile application. Specifically, I want to focus on the task of playing with the recommendations of videos in the mobile application like it's a child's play to remove recommendations, to add recommendations, to mark/block videos from recommendations.

Qualitative Evaluation

I will be using surveys method for the qualitative evaluation of the wireframe prototype which I have described in my previous assignment.

Plan

The survey will be hosted in the Georgia tech peersurvey[1] platform and will be given by my Georgia tech colleagues and family and friends. The result of the survey will be stored through the method of **Software Logging** on the peersurvey platform. The response of each and every participant will be logged **anonymously and asynchronously.** The users will record their input using the wireframe image of the prototype which will be attached in the survey.

Content of evaluation

As of now, I am planning to include the following questions to be asked in the qualitative evaluation:

- Your age group you belong to
- How often do you use the youtube mobile application?

[1] https://peersurvey.gatech.edu

- What did you think when you saw the left/right swipe in a video to remove/add it from your recommendation list?
- What do you think of the "UNDO" feature to undo your actions?
- What did the bubbles tell you after you watch each video?
- What objective do the bubbles fill you think after you watch each video?
- Do you think yourself as an expert user of youtube or novice?
- If you are an expert, then what do you think about the gesture controls like tick right or wrong on the screen to mark a video as "Interested" or "Not Interested"
- Do you think this prototype will be helpful in managing your recommendations on youtube?

Through the above set of question, I should be able to answer the following data inventory questions like:

- Who are the users?
- What is the context of the task?
- What are their goals?
- What are their tasks?
- What are their subtasks?

It would greatly help to answer if the prototype met the **functional and learnability goals** from the requirements. It would be helpful to know from the direct feedback on the swiping feature to manage the recommendation that will help me understand in evaluating the prototype.

Empirical Evaluation

For this evaluation, I will be using my second prototype of voice-based control of the recommendation on the youtube mobile application. The intent or the goal of my evaluation here is to evaluate the new voice-based interface in comparison to the existing interface for managing or controlling the recommendations. Since currently there is no way of **blocking** channels or video from an artist or category in the youtube, the whole idea is to leverage the voice-based commands to **control** your video browsing experience as rich as possible. The following questions will be gathered for an empirical evaluation:

- How long does it take a user to find the "Not interested" button in the existing youtube mobile application?
- How satisfied a user is with the existing functionality of "Not interested" feature?
- How long does it take a user to learn or get used to this functionality of "Not interested" feature?
- How quickly a user is able to find the content he wanted to watch?

Null Hypothesis: The null hypothesis for all the above-mentioned questions will be that both interfaces are equal in terms of time to complete a specific task related to the "Not Interested" feature to manage the recommendation in the application. Let's represent this null hypothesis for each of these parameters being equal as:

H(voice) == H(touch)

Alternate Hypothesis: The alternate hypothesis can be:

H(voice) < > H(touch)

Using the above questions we can do the empirical evaluation to find out the differences if there are any in our prototype. We should be able to find out the differences in terms of using the option of "Not interested" feature that current interface serves with the new voice-based interface. This way we would be able to find which of the interface is better.

I will be using the **within-subject experiment method,** as there is not much of a **significant** difference in the terms of the time execution using touch as well as the voice-based interface. In this **within-subject experiment method,** each subject can use both the interfaces, i.e touch based as well as voice-based to give us a better evaluation. I will be using at least 10 participants to evaluate the interface and provide answers to the above questions. I will also be dividing the 10 people group into 2 groups to avoid the variable where we don't know which interface the user is going to see first.

- Let's call one group as Group1 and another one as Group2.
- Use **random** assignment to assign people to 2 groups.
- Group1 users will be showcased the voice-based interface first and then they will be moved to the touch-based existing interface.

- Group2 users will be showcased the touch-based interface first and then moved to the new voice-based interface.

Data Generation

1. How long does it take a user to find the "Not interested" button in the existing youtube mobile application?

   This question will generate the **nominal quantitative data,** and I will be using the **CHI-squared test** to analyze the data to test our hypothesis to determine:

   H(null)-Distribution is equal for the time taken between the methods.

   H(alternate)-Distribute is unequal for the time taken between the methods.

2. How satisfied a user is with the existing functionality of "Not interested" feature?

   As this is an **ordinal quantitative data,** Kolmogorov-Smirnov test to analyze the above type of data to determine:

   H(null) – Distributions are equally ranked between the methods or

   H(alternate) – Distributions are un-equally ranked between the methods

Similarly, the other questions are also nominal data, So I will be using the **CHI-squared test** for them.

The major **lurking variable** includes here is that we don't know the user will see which interface first so, to tackle with that we have assigned the users into groups and randomly assign the interface to each group.

Predictive Evaluation

I will be doing the predictive evaluation for the textual prototype mentioned in my previous assignment. I would like to use the **GOMS** model to evaluate my textual prototype. GOMS model will be used to address the task of marking a video as Not interested or to say in other words remove a video from the recommendation list based on the sentiment or emotions of the person emitting via

EEG(electroencephalography[2]) and then a software will be used to determine the activity or mood of the person and according to the software we will add or remove the video from the recommendation list. I am attaching the details for this task in the appendix for a better understanding of how this prototype works.

This will then be compared with the current approach of scrolling down for videos and marking a video by touching 3 dots and then clicking on the button of "Not Interested".

- The **goal** of each of the above-mentioned tasks will be to add or remove a video from the recommendation list of a youtube mobile application user.
- The **operators** for the tasks will be like: attach the wearable neural device, open the neural application on the phone, select your favorite mood today like how you are feeling, and then open youtube, and then start browsing the videos from the mobile application.
- The **selection** here is based on the different kinds of emotions or mood that a user is feeling and accordingly the application will show the content and remove and add the recommendation from your application.

I will be evaluating an expert user in accomplishing a single goal of managing his recommendation with zero cognitive loads.

Preparing to execute

I will be using the following 2 evaluation techniques:

Qualitative evaluation

- No extra recruitment needed, I will already get the response from the survey.
- A qualitative approach is necessary for a way to know why the user is interested, and what does he actually wants to do.

Predictive evaluation

- I think having a GOMS model is necessary as it gives you the overall flow of your design, though it certainly not fetch the granular needs and demands, it

[2] https://en.wikipedia.org/wiki/Electroencephalography

definitely helps in performing the goals and subtask across the broad set of the users' group.

- As the textual prototype is based on virtual reality it will be important to have a GOMS model to understand what are the goals and sub-goals to better identify the results.

Reasons for not selecting the empirical evaluation method is :

- Given at this early stage of my work, I think it would be unreasonable to calculate the time it would take the user group to perform the evaluation at this stage.
- Also, the prototype voice-based interface is not yet completely ready for the users to evaluate.

Appendix

Textual prototype mentioned in the M3 assignment:

This prototype will include the virtual reality and I would like to add it as a textual prototype to better describe my intentions while designing the 3rd design alternative:

*"Integration of the **virtual reality** to control the recommendation with **zero cognitive loads**"*

How can I achieve zero cognitive load

Imagine a wearable device like neurable[3] which is made of electrodes to measure your brain activity. The purpose of this device will be to measure your brain neuro-signal to determine a user's intention or mood while watching a video, like for example if while watching a video I am feeling good that the device should be able to capture the brain signals and accordingly mark that video or channel as "Interested" or "Not Interested". The device will use the dry electrodes to record the brain activity via EEG(electroencephalography[4]) and then a software can be used to determine the activity or mood of the person and according to the software will add or remove the video from the recommendation list. The user has to literally do nothing except watching a video thereby reducing the cognitive load to zero as the user will not have to remember his interest or whether he like or not liked the video the device will start improving quickly as it will consider the brain activity.

[3] http://www.neurable.com/
[4] https://en.wikipedia.org/wiki/Electroencephalography