

Sajid Sadi

Project Exercises

Overview

The general idea of the project is to springboard off a project I am working on partially for my own edification and partially for the group. It involves the UBERBadges used during the last TTT meeting (also slated for used in the coming meeting in May). The basic idea is to be able to give the sponsors direction support to find projects they are actually interested in, while at the same time allow a greater level of personalization that has been previously possible. The work breaks down into 2 phases from the viewpoint of MAS 965:

- ***Give the user a clearer ability to personalize his or her choices:*** The basic plan here is to use a [Dasher](#)-style interface to allow the user to choose topics with a gradation of interest levels. This sort of interest allows mining of the user's "attraction level" towards a concept (in theory, of course), and it will be a challenge to develop the sort of topic tree that allows this form of interaction.
 - ***Create a transparent interface to the recommender system:*** The personalization is all but useless if the recommendation system cannot convince the user to trust it. The problem here is the short time interval within which the system must work. Essentially, the goals of the system require perfect (or at least relevant) results with zero training. However, if the system tried too hard to be accurate, it would fail to produce a long enough list of meaningful results to last the day. One of the central issues I am exploring here is how to make the system transparent enough so that the user can trust it when it is correct, and understand it when it is wrong.
-

Week 2 Exercise: Relationship Basics

- ***Enticement:*** The sponsors get to see the things they are interested in, instead of following around the group trying to find something that looks interesting. Moreover, it deals with the issue of sponsors forming into large groups, where no one gets any individual attention, and the presenters are forced to be generic in their talks.
- ***Relationship:*** There already exists a system for showing interest in particular projects. By voting on a project, the sponsor is essentially voting on the subtree of interests the project subscribes to. This can then be used to emphasize those interests for future recommendations.
- ***Fulfillment:*** The sponsors get a better experience, and the presenters get to get their message across to people who are actually interested in the same domain and issues.

- **Main goal:** please see above (essentially there are two).
 - **Why should it be relational?** If the sponsors are unable to understand the system, it is likely they would fall back on the pattern of simply following the crowd or relying on the "routes" system that's used in the current TTT handouts. By creating a system that the sponsors can understand and relate with (at least in the context of relating to an expert system that they feel understands their needs), the overall experience is greatly improved.
 - **What about it is relational?** The personalization-interaction-feedback cycle can be seen as a relationship with the system.
 - **Target population:** TTT sponsors (though the DL sponsors have also expressed an interest in this).
 - **How would you evaluate it?** If the system understand the users' wishes, properly deals with their actions, and provides clear and understandable (and hopefully meaningful) feedback, then the system could be considered a success in terms of being able to conduct a sort of relationship with the users. On the other hand, if the users end up rejecting the system, then it has in some sense failed. To put it shortly, if the users use and trust the system, and it improves their experience, then it will be a relatively complete success.
-

Week 3 Exercise: Defining the Relationship

Relationship

I have spent some time thinking about the interactions that the users will have with the program (machine from now on... just for consistency's sake with what I normally use for the paper comments), and I have seriously moved away from portraying the machine as any sort of assistant. There are several issues with doing so:

1. An assistant is normally expected to know a lot about you. However, the personalization process is naturally flawed in that it gathers information not about the holistic user, but a subset of the user's interests as they pertain to this lab. It thus seems unnatural to call a machine an assistant when it is so limited in what it can actually do for you.
2. The assistant is, in general, a helper devoted to you (or to you and a small number of other individuals). In this case, the assistant is devoted to hundreds of sponsors. While the computer has no problem handling this (especially in the limited areas that it will be interacting), it does not seem to resonate mentally with the image of an assistant.
3. The assistant is a long term helping hand, and is constantly by one's side. On the other hand, the machine is only here in the most concrete sense while the sponsor

is at the lab. While I will talk more about the long term in a later section, overall the machine's limitations suggest a more generic title.

So, with that in mind, I propose that the system portray itself as the stereotypical English butler. This sort of role encompasses the knowledge that the system is in possession of in terms of the overall structure and content of the lab (what one might consider it's "butler training"), but at the same time it is somewhat removed and is not specific in its behavior to the extent that a personal assistant would be. This projection also embodies the idea of service, in that the sponsors cannot really thank or talk to the system, but the system is always there to help (which is, in simple terms, the duty that it is meant to carry out). This in some sense goes with the idea of the butler, who is trained and devoted to this sort of service.

Costs and Benefits

The cost to the sponsors of using the system is really minimal (they only have to pay attention, and complete an initial personalization), while the benefits of having a guide through the chaos of projects that is the Media Lab is of course quite high. Even though the sponsors now are accompanied by the hard-working folks of the CASR office, the personnel there have only limited knowledge of what is going on around the lab, and what thing may be of interest to the sponsors. In fact, I generally find that the information on sponsor interests from the office is so minimal that it is best to ignore it and simply quiz the sponsors on their specific interests as an icebreaker for visitors. In fact, it appears that the sponsors are left to do the lion's share of investigation into what groups they should meet. While this is a bad time investment (to help choose groups) for the CASR office, it is a perfect job for a computer. Moreover, the disclosure problems are somewhat lessened by interaction with a system that is perceived as inanimate, which might help with the guidance.

Because the system interacts in very fail-soft ways, even if it is not perfect, little is lost. In the worst case, it is no worse than the sponsor choosing the wrong group to visit him/herself, which happens frequently enough currently. Additionally, if the sponsor maintains a relationship with the system over time, the system becomes more and more fluent at understanding their needs, which helps the sponsors capitalize on their investments of time and money, and helps the Media Lab keep sponsors interested (and hopefully happy).

Establishment

As I mentioned earlier, the cost of the relationship in terms of real investment is the initial personalization. It is clear that sponsors do want and need more guidance in this area, and thus there is a large incentive (enticement...) to complete the personalization. After this, the information is generally handed out in very small packets, minimizing the marginal time costs of continuing the relationship. Additionally, I believe that once the system is accepted by a number of users, the "peer pressure" factor that exists in sponsor meetings (are you using that cool gadget yet?) will cause further adoption of the system.

In some sense, by billing the system as some sort of a search engine for projects, I hope to bypass some of the initial skepticism and “hook” the users.

Maintenance

Of course, the coup de grace of the system is the sponsor event itself. If it can function well within that environment, it will promote further use. The system currently has a voting interface that allows sponsors to choose to receive more information on projects. I hope to mine this data for my own purposes. Firstly, this data allows the machine to further customize its profile of the user’s interest, which over time causes the accuracy of the system to increase. Moreover, this also gives a correlation between the sets of projects that are liked by similar sponsors, which basically gives the machine overall experience on which to draw upon when recommending projects that are not directly within the sponsor’s preference tree, but are liked by sponsors like the user. This essentially boils down to an added collaborative filter overlaid on the personal preferences. Finally, when the requested information is sent to the user, by adding additional information from the system via the web, it is possible to continue the relationship beyond the limits of the event itself. Of course, there are real limitations to the process, in that there are a finite number of projects within the lab. Additionally, the PLDB entries are generally not updated regularly like they are supposed to be (for reasons that we all subscribe to religiously), which adds an extra latency to the data. Over time, though, I expect that sponsors will find the groups and projects they are interested in, and hopefully interact with the project for the duration of the RA’s stay (or even longer), which should cause the interaction to slope off gracefully during the nulls in new information.

Week 4 Exercise: Defining the Relationship

A short summary of how things stand now (as presented to UBERBadge group): the system essentially uses the UBERBadges as input devices to a just-in-time recommender system that guides the sponsors to projects, groups, and even persons (time permitting) relevant to them. Part of the relevancy comes from the personalization system, which will be using a non-list-based method of input to attempt to maximize the “gut-response” veracity of the input. The other part comes from the preferences of those with similar likes and dislikes in respect to the personalization system. The project-voting interface will also be mined in real time to adjust the preferences. The output of the system will be displayed on any available computer running the client software (which will be distributed to all groups before the TTT meet). The software will try to spear out the visitors across all groups as well, to prevent the prevalent flocking behavior, which generally impedes a good visiting experience.

I plan to approach the evaluation of my project in two parts: before and during full deployment during the TTT meet. I will talk more about the pre-deployment evaluations because I feel that those will have more of an effect on the project in the design sense. I realize that my project is somewhat reflexive, which will no doubt introduce a good bit of bias, and this makes end user input all the more important. The small-scale initial deployment will serve both to iron out bugs and allow observation of the users using the actual devices, as well as open-ended interview input from the users about the design.

To carry out the test deployment, I need several things in place:

- Personalization input system complete (at least sufficiently to use)
- Finish data mining from PLDB and old TTT documentation – I am fairly certain I will not be able to get all the advisors to go through the update process by the time I want to test
- Client/server/recommender modules fully complete and running
- Badge system up and running (available now in limited scales)
- Instructions complete – both written and oral
- Pre-testing of systems complete

The design of the system essentially requires almost all of it to be up at a small scale for it to work at all. While the client/server design is nearing completion already, the personalization, recommender, and data miner subsystems are still very preliminary.

Most of these are technical hurdles, but the instructions deserve a bit more discussion. This year the badge description will be limited to 5 minutes, and there will be a video describing its use. I have considered handing out a simple “cheat-sheet” to the badge when the sponsors pick it up. However, the direct user interaction is via a single button press, and it might be more effective to have it covered in the intro video/Bartos session. In some sense, I wish to have the system act “intelligently” insofar as it tries to anticipate the user’s needs (the butler metaphor, essentially), without the user losing control. To that end, it would probably be best to keep it simple as far as interaction goes.

As to the subjects of this initial study, I want to try it out on a few interested sponsors, as well as a number of people (possibly UROPs) who are not familiar to the Media Lab in general. This is because I want to have a system that is general enough for everyone to understand at a fairly low level, but also tailor it to the needs and wants of the sponsors towards whom it is targeted in this case. I will probably test it only with one or two individuals at a time. Though this limit will obscure the effects of the crowd mentality present at the actual event, it should give me a better idea of interaction under the ideal conditions (ie, no external motivators). While the system will have an “anti-flocking” feature, that will be the weakest motivator of recommendations, and act as a simple weighting on the recommendations.

I will both be observing the usage of the system first with minimal instructions, and then after a few minutes, with the full instruction, to get an idea of the intuitiveness of the

interface. I will also have a set of questions (closed response) ready for the interview in addition to the open ended request for suggestions and critique, to attempt to evaluate the overall ease of use of different parts of the system, the extent to which the users felt the system knew them, and the extent to which they trusted the recommendations of the system. This information is relevant to tuning the system to a middle ground between showing the users new things, and showing them what they want to see, that maximizes overall confidence in recommendations.

During the actual meet, I will collect the voting information from the badges (as well as what projects the sponsors choose to receive more information on). In addition to further customization, this can then be correlated with the recommendations to get a positive measure of the success of the system. Not voting is not necessarily an indication of dislike, and thus cannot be considered, though that would in fact be more interesting. Additionally, I hope to give the sponsors a short questionnaire at the exit booth where they return the badges, to get a self-descriptive measure of the same factors I will attempt to gauge during the initial test run. However, the effectiveness of questionnaires is somewhat questionable given the conditions of the administration, and I do not expect to use that information as a data source, but more as a corroboration of the voting data. There is also the additional danger of the sponsors reverting to their “normal” behavior and ignoring the badges altogether. However, I expect that the novelty of the situations should allow for at least partial data of some relevance (of course, with respect to novel situations).

Week 6: Initial Interviews

Who you are going to talk to (users and others involved)

I plan initially to talk to the C*D office people, to get a basic idea of the sort of questions they face most often, and about their methodology for guiding people around the Media Lab. After I have a baseline (and do a first pass modification on the questions), I will talk to some of the sponsors. Concurrently, I will also interview some undergrads to get a sense of how they view the work here and what they would really be interested in seeing. This last part is to get an idea of a middle ground between someone who works here, and someone who comes from the outside, so that I have a good range of thoughts and specs on what people want out of a guidance system like this. Additionally, I will be working closely with the badge teams (both hardware and software) in order to keep the logistics of the deployment in line.

What kinds of information gathering strategies you plan to use

I will be relying on interviews for the most part, since getting the people I am interested in to commit time to focus groups is somewhat difficult. I will also do some focus-group style (if not outright focus-oriented) meetings with the badge team. These will focus more

on the technical and behavioral aspects of the interaction, as well as the interaction between the many projects that will be running on the badges during TTT. I have already done one that focused on the low level technical details, but as things proceed, these will move towards more interaction and less technicalities.

Purpose of the interview and an agenda -- what you want to find out

I have a pretty good idea of the exact need that needs to be addressed. What I will be focusing on in these interviews will be the things that people want out of the presentation itself. Things I am interested in are:

- How the system should justify itself, and to what extent
- How it should direct the user
- How much time the user would want to spend in reading these recommendations
- How often they expect to use the system (technical detail needed to calculate expected loading of system)
- Whether the fact that the system would know their location (and there is talk of a “locate-the-user” facility as well) would trouble them
- How they envision the map, and what map projection can they read most easily
- How hands-free the interaction should be
- What sort of balance between popularity and relevance is good
- Functionality they would like to see in the system in addition to what we are proposing (open ended)

Interview questions

I sort of cover the questions in (3). I will doubtlessly edit these after meeting with the C*D people, but I don't want to add too much more. The questions are all fairly open ended at this phase, because I want to foster ideas, and I think the above will already take more time than I really want people to spend on it.

What props you're using

Maps with different projections, a badge

What will be recorded

I might record audio of the interviews with the sponsors. I will probably run things more informally with the other subjects, and notes should be sufficient in those cases. These recordings will be for internal (my own) use only, so I am not sure what sort of waivers are needed yet.
