CS310: Project Specification

Matthew Smith
ID: 1003843

October 27, 2012

1 Project Description

Giving Up Smoking: Modelling how Social Networks Impact the Breaking of Habits.

From existing research, it is known that social networks, where a network is defined as "a pattern of interconnections among a set of things" (Easley & Kleinberg 2010), affect the behaviour of the individuals involved in them – for example the concept of 'peer influence', where one group member has their behaviour affected by how the rest of the group act. A natural extension of this is that groups can cause the creation of habits within a person, such as smoking (Harakeh & Vollebergh 2012), but another aspect to consider is whether these networks can aid in breaking habits. This project will investigate individuals giving up smoking with respect to the social networks that they are involved in, with particular attention being paid to how their role and involvement in these networks changes over time.
In addition to the research-based aim, there is an second, technical aim – as a whole, the project will be an assessment of whether social network impact modelling is effective and manageable as a part of a larger, commercial level environment. This will be assessed through working with the company SandTable, who work on the modelling of many aspects of human behaviour. To judge this aspect of the problem, analysis will be performed on the resulting model to determine if it would be of any value to SandTable.

2 Objectives

To help with the planning the project, it is split into the following stages. Since there is a large research element to it, at the most detailed objective level, there is still a lot of space for definition - this will allow for said research to drive how the project proceeds.

2.1 Main Objectives

Overall, the project has two main aims:

1. To investigate the effects of social networks on attempts to give up smoking, particularly how the individual's networks change over the attempt period.

2. To consider whether the model would be appropriate to fit in with the commercial level model that SandTable (referred to as the company) have produced, and furthermore, whether it would be feasible to manage.

2.2 Sub-objectives For Objective 1

Evidently, these objectives are wide ranging and quite general. Objective 1 can be further split into the following objectives:

1. To research AI modelling techniques and the agent-based system approach to use as a base for the rest of the project.
2. To investigate the methods through which social networks can be represented in a modelling environment, focussing on a graph based representation, where nodes represent individuals with certain characteristics and edges are the connections between these individuals.
3. Work the psychological aspect of influence on smokers in groups into the model, as the basis for design decisions when it comes to recreating human behaviour.
4. To model, as accurately as feasible, a group of individuals that contains smokers, non-smokers and those giving up in terms of interactions and behaviour.
5. To model the change in social networks over time as the habit changes, and how this affects the individual giving up.
6. Investigate methods of generating networks and analyse their suitability in representing social networks - examples of these types of network are *small-world*, *scale-free* and *preferential attachment*. On top of this, build tools to parse and output formatted graph data so that real-world datasets can be interpreted and visualised.
7. Introduce the ability to put in vague statistics, such as "50% of the sample are smokers", and simulate this situation.
8. To gain data and analyse this data, gaining some conclusion about the problem at hand.

2.3 Sub-objectives of Objective 2

From the perspective of objective 2, the task is more focussed to the overview of the system:

1. Provide analysis as to whether the model would be a useful addition to a larger, wide-ranging model.
2. Consider if, when scaled, this implementation would be useful to a large model which would require considerably more computing power.
3. Examine if the design used in this project would be applicable to the company's model, with focus on whether it would be maintainable.
4. To look at whether the defined input methods of situational information for a model run through would be useful to the company.
5. Analyse whether data outputted would provide any useful information to the company.

3 Methods

Since the project is research-based, it has a natural flow to it – for example, a certain amount of research must be completed before the main models are started. With respect to main objectives, objective 1 will largely be completed before objective 2. Within objective 1, sub-objectives 1-6 are partly research based with some experimentation occurring at the same time, giving the potential to extend these tools to be parts of the final model and simulation. Sub-objective 7 largely depends on the research and example programs being built, so can only take place after these steps are done, whereas 8 will be done simultaneously with the second objective. To complete objective 2, it will be considered throughout development, particularly parts 2 and 3, as these influence the implementation approach. Sub-parts 1, 4 and 5 will fit in with 1.8 as they all involve considering the system from a completed perspective, as a final review of the project as a whole. In general, objective 2 can only be started after objective 1 hits the development of the main model, and completed once objective 1 is complete.

4 Resources

The resources for this project are:

- Java, used for development of the models, interaction methods and most elements of the project.
- Perl/Python for writing housekeeping scripts and creating test data to be used within the system.
- Repast Simphony, a reactive agent modelling environment which is the basis for the creation of the model, providing modelling utilities and a time-step based environment.
- GitHub, *git* based source control allows the tracking of versions on both the main project and test projects.
- JUNG, an in-memory graph representation library, used for experimenting with network generation.
- Gephi, graph visualisation software.

5 Timetable

The estimated project timetable can be seen on page 6 - this details the steps taken from doing basic research through to final analysis and presentation. Please note that the documentation for the project will be started in Week 5 of Term 1, and maintained throughout.

6 Legal, Social, Ethical and Professional Issues

This project should raise none of these problems – the only potential ethical issue would be if a dataset relating to giving up smoking was used. To avoid issues with this, it would be ensured that the data was anonymous and only contained the information important to use for the project; through this, the project will have no concerns. Furthermore, this project, whilst researching giving up smoking, is not casting smoking in a bad way, but will remain neutral in this respect throughout.

Term	Week Number	Approximate Dates	Task
1	3-9	15th Oct - 30th Nov	Research approach to the main model, using medical and psychological theory as a basis for research into technical details. Also experiement with basic models, reconfiguration and creation of networks (specifically *small-world, scale-free* and *preferential attachment*, giving focus to modelling human relations and networks. Project tools will also be developed, such as dataset parsers and formatted output generation.
1	10-14	3rd Dec - 4th Jan	Develop begins on a basic model of individuals and their interactions, including some tests on small networks. Some work will also be done with on social networks changing over time and the ability to provide basic statistics for network generation.
2	1-2	7th Jan - 19th Jan	Work will be done on expanding to more realistic social networks by improving the human model and interaction methods with a view of the work already completed. More advanced social network changes over time will be incorporated along with attempts to generate realistic networks for testing. Also consider how this would merge to the SandTable environment.
2	3-5	21st Jan - 8th Feb	Finalise model and interactions, whilst fine-tuning network generation, input methods and social network changes over time. Some testing on larger networks will be carried out.
2	5-7	11th Feb - 22nd Feb	Perform simulations and analyse results, ensuring no errors exist in the model - if they do, examine their extent and try to fix them. Consider whether the system will fit in with SandTable and their models.
2	8	25th Feb - 1st Mar	Prepare the presentation and results for demonstrating the whole project.
2	9-10	4th Mar - 15th Mar	Present the project, whilst continuing to write up the results and analysis.
2	10-15	18th Mar - 19th Apr	Finalise write up before submission.

Table 1: Project Timetable

er

References

Easley, D. & Kleinberg, J. (2010). *Networks, Crowds and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, chapter 1, p. 1.

Harakeh, Z. & Vollebergh, W. (2012). The impact of active and passive peer influence on young adult smoking: an experimental study., *Drug and Alchohol Dependence* **121**: 220–3.