Small Data Analyst
1. MSc Project Progress Report

Author: Sebastian Zillessen *
Supervisor: Jeroen Keppens †
Product Owner: Isabel Sassoon ‡

June 22, 2016

1 Introduction

This first project progress report summarises the progress made so far in the MSc Project "Small Data Analyst" and amendments to the original outline and project plan. The project is supervised by Jeroen Keppens and the requirements for the software have been closely evaluated and defined together with Isabel Sassoon, author of the underlying paper [7].

2 Literatur Analysis

To start with the project, a background research has been done. The main focus is Argumentation Theory [3] and different ways of extending Argumentation Frameworks to work with preferences in conflicting argumentation frameworks [5, 2, 1, 6], as this will be an important aspect of the software. These extensions have been quickly reviewed and the chapter on the literature review is almost finished.

However, it has been decided to postpone the further writing of the final report, as I would like to finish the implementation part first. This is reflected in the updated project plan, attached to this report.

3 Use Case Specification and Refinement

To understand and to specify all requirements, a Use Case 2.0 [4] analysis has been performed. These where discussed and approved by Isabel Sassoon to meet the requirements of the theoretical underlying work of statistical model selection. In addition these Use Cases have been grouped, ranked and rated according to the popular MuSCoW (**mu**st, **s**hould, **co**uld, **w**ould) prioritisation scheme. The Use Cases have been grouped in four Release Candidates (RC 1 - 4).

4 RC 1 and Prototype Demo

RC 1 includes the core functionality to perform an analysis on research questions and their possible models by using critical assumptions that are evaluated on the datasets. This involves as well the following Use Cases:

- Entering two default models into the system including definition of critical / non-critical assumptions [UC 1[1]]
- Selection of a research question and performing an analysis of possible models with the selected dataset (AS1) [UC 2,4,5]

In addition to these Use Cases, the setup of the infrastructure has been done in RC 1 as well. The final application will be developed as a web application and be hosted on Heroku. One of the biggest challenges during RC 1 was the ability to execute `R-Scripts` in Ruby on Rails [UC 18]. For this purpose the Gem `rinruby` has been used and slightly modified to ensure compatibility with `ruby 2.2.4`. In addition a `TravisCI` environment has been setup for this fork of the gem to ensure correctness and to improve code quality.

A continuous integration environment has been setup for the further progress of this project including Github, a Travis-CI instance to ensure that all

*sebastian.zillessen@kcl.ac.uk
†jeroen.keppens@kcl.ac.uk
‡isabel.sassoon@kcl.ac.uk

[1]The Use Cases are described as Trello cards. The progress of the

test are always green and an auto-deployment feature to continuously deploy tested code versions to Heroku.

Despite the original schedule, RC 1 includes two additional Use Cases that were planned for later releases ([UC 2-1]: Store completed analyses, [UC 17]: Enter additional research questions into the system, [UC 8]: See assumptions that hold and that do not hold during the analysis).

From these Use Cases various other tasks have been derived and implemented in RC 1. The current deployed version of the application is available online at https://small-data-analyst.herokuapp.com and has been presented in a first demo to Jeroen Keppens and Isabel Sassoon. Overall the current progress is in accordance with the original planned progress and the result has been satisfactory.

5 Current Status of RC2

RC 2 is building up on RC 1 and will include the following new aspects:

- Authentication support. [UC 6, 6-1, 6-2, 6-3]
- Ability to add new assumptions online [UC 12-1, 12-2, 12-5]
- Upload functionality for datasets [UC 7]
- Detailed reports of assumptions that hold and those that do not hold and therefore cause a model not to be possible [UC 8]

In addition the first required code parts for the analysis of the possible models using an argumentation framework will done in RC 2 (although this was initially planned to be part of RC 3, but after receiving the first feedback on RC 1 this change seems reasonable). This will enable us to perform first analyses of the Argumentation Framework describe as AS 2 in [7]. Furthermore access to a big dataset for testing the application has been granted.

6 Project Plan Updates

The original project plan as proposed in the Preliminary Report has been slightly changed, due to various reasons.

First of all, the different project phases have changed: The writing of the final thesis has been postponed and the focus on the development has been increased during the first months. In addition as a learning from RC 1, the duration of the different release candidates have been reduced to iterate faster. A buffer of approximately 4 weeks has been integrated into the planning to be able to react flexible on problems that might occur during the development.

Furthermore the feedback times have been decoupled from the remaining development process, as we decided to have short demo sessions immediately after each release candidate, so the feedback process is reduced to a minimum. The updated project plan can be found at Trello and the most current version is Projectplan_v5.

References

[1] Amgoud, L., and Cayrol, C. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (1998), Morgan Kaufmann Publishers Inc., pp. 1–7.

[2] Amgoud, L., and Cayrol, C. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence 34*, 1 (2002), 197–215.

[3] Dung, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence 77*, 2 (1995), 321 – 357.

[4] Jacobson, I., Spence, I., and Bittner, K. *USE-CASE 2.0 The Guide to Succeeding with Use Cases*. Ivar Jacobson International SA., Dezember 2011.

[5] Modgil, S. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence 173*, 910 (2009), 901 – 934.

[6] Pollock, J. L. Defeasible reasoning. *Cognitive science 11*, 4 (1987), 481–518.

[7] Sassoon, I., Keppens, J., and McBurney, P. Towards argumentation for statistical model selection. In *COMMA* (2014), pp. 67–74.