Contents

1   2019-03-06

1.1   Preparation

Question: How should this dissertation relate to Cassidy's project outline?
Thoughts on data structures:

- dataframe seems ok for tidy text style data

- nested dataframe may be best for survey response data

- look at other styles (incl. non-tidy), maybe trees, as per parse trees etc

- Document-term matrix for tf-idf (PCA would be interesting here due to massive dimension)

Reading: Text Mining with R: notes

☒ Read Text Mining with R

☒ Assess Twitter api

☒ Play with iNZight/lite, Cassidy's Project

☒ Consider UI

☒ Consider Survey Responses

☒ Draft UI Depictions

1.2   Minutes / Summary

UI is something that will have to be organically developed as we go.

Meeting was mainly demonstration from Jason to Chris about the preparation done, as well as Chris demonstrating to Jason the current iNZight UX. Research so far is good, but too broad to demonstrate in the confines of a weekly meeting - Chris suggested a solution; to set notes in a neat summary format and publish to a github repository

1.3   Actions

☒ Set notes in neat summary form (organise file structure to match)

☒ Push to a private github repository

☒ Give access to Chris

☐ ~~Create twitter developer account~~

☐ ~~Get twitter api access token~~

2   2019-03-13

2.1   Minutes / Summary

This meeting took place while seeing what packages exist already to complete various tasks. We looked though the various packages I have found. Determined that we needed to reel in and pick specific features that we want, according to the heuristic; if it isn't obvious immediately, get rid of it. The primary question we seek to answer with this text analytics program is, what are people talking about, and how do they feel about it. With this in mind, analysis for e.g. writing styles are not to be considered (who wrote shakespeare etc.).

2.2   Actions

- ☐ ~~interface (w. respect to above)~~

- ☐ ~~lit review of twitter analysis (esp. discourse)~~

- ☒ Determine what text analysis really is at base, and how it relates to the primary question

- ☒ find if subtitles can be attained easily enough

- ☒ formal critique of Cassidy's project, integration w/ iNZight

- ☒ scope reduction

- ☒ See similar papers in stats printing room

3   2019-03-20

3.1   Summary

Close assessment of textrank package and background algorithm - on the right track for scope. Chris noted to check for survey response data etc. from those involved with it.

3.2   Actions

- ☒ Create an example corpus for testing

- ☒ Assess textrank performance on the example corpus

4   2019-03-28

4.1   Summary

Presentation of further package discoveries, for development purposes, as well as end-user. Demonstrated the benefits and drawbacks of using textRank, discussed the possibility of lexRank. Presented draft of feature space - including the usefulness of categorisation of text analytics as within- and between- texts. Much discussion dedicated to summarisation.

4.2   Actions

☒ Development of a test-corpus; Chris has emailed some free-form patient responses

☒ Further testing of textRank (try tidy-style) on corpus

☒ Testing lexRank on on test corpus

☒ Formalise feature-space as list

5   2019-04-03

5.1   Summary

Demonstration of summarisation systems, confirmation of use of lexRank, discussion of case of free-form response data - much interest in these capabilities

5.2   Actions

☐ Summarisation of Free responses using the two example datasets; indiv. response as token?

☐ Testing of standard text analysis tools on free-response data - see what is useful

☐ Test grouping variables such as survey group in the analysis

☐ LexRank for keywords

6 2019-04-03

6.1 Summary

6.2 Actions

6.3 TODO