

Exploring the Use of Computational Linguistics for Automated Formative Feedback in the Humanities



Background

- This project is a small, experimental collaboration between the School of Informatics, and the School of Divinity at Edinburgh University.
- We wanted to investigate the use of computational linguistic techniques in providing automated formative feedback to students on their written work.

Methodology

We split the project into two parts:-

- Defining the criteria markers used in assessing student work - this was done using manual and computational methods
- Reviewing possible computational techniques - we also performed a few brief exploratory experiments using these.

Conclusions

- There is good evidence that formative feedback is a valuable component of the student experience.
- This exploration suggested that there is real scope for using computational techniques to produce a practical tool for providing such feedback.

Defining Quality Criteria

We manually analysed comments that markers had made on student essays and wiki entries, to extract three categories of quality criteria:

Referencing: e.g. *peer-reviewed, academic print journals are generally considered better than solely electronic resources.*

Style/Terminology: e.g. *avoid making normative claims, such as referring to the Shia as a "deviant sect".*

Structure: e.g. *the first paragraph of your work should outline the main argument of your paper.*

We performed an automated analysis using the LIWC and Wordsmith tools, to identify criteria which correlated with the quality (as measured by the awarded grade):

LIWC - Grammar: e.g. *more highly graded essays tended to have less use of words relating to emotion (positive and negative).*

LIWC - Punctuation: e.g. *more highly graded wikis tended to have less use of period, semicolon, and punctuation overall.*

Wordsmith: e.g. *more highly graded essays have more distinct terms (non-repetitive).*

Computational Techniques

Analysing Style: a range of comparatively straightforward techniques are available for analysing "surface features". For example:

A "POS tagger" can mark parts-of-speech which can then be used to identify the use of the passive voice, or too many prepositions.

Simple "regular expressions" can be used to check the format of references.

Simplified English dictionaries can be used to check for the use of too many simple words.

Analysing Content: more sophisticated techniques are available for analysing "deep features" (meaning). For example:

TextTiling to measure cohesion: *cohesion can be used as a proxy for detecting a clear question and a thesis.*

Google Scholar Citation Count: e.g. *does this reference seem to be an important one for this topic?*

Techniques such as these deal with different aspects of the text, and are easily "fooled". This makes fully automated marking (for example) very difficult. However, they are well-suited to the generation of automated formative feedback where the student has the option to reflect on feedback and reject it where it is inappropriate.