

Detecting Gender Bias in the Coverage of Politicians in Irish Newspapers Using Automated Text Classification

Susan Leavy
School of Computer Science and Statistics
Trinity College Dublin

Abstract

This paper explores the use of machine learning and natural language processing techniques to identify gender bias in the coverage of politicians in Irish print media. Newspaper content is analysed and systematic differences in the coverage of male and female politicians are highlighted. These differences are examined and evidence of gender bias is identified.

The corpus analysed was comprised of Irish newspaper articles featuring male and female politicians over a 15 year period between 1997 and 2011 and coverage of candidates in the 2011 Irish Presidential Election. Machine learning algorithms were used to identify differences in the coverage of male and female politicians. These differences were analysed for evidence of gender bias and where appropriate, the context in which these differences occurred in the corpus were investigated. A broad range of text classification experiments were explored and the best approaches to using text classification to identify gender bias in text were identified.

This research found that the best approaches to text classification involved using a support vector machine learning algorithm along with a binary representation of the features of the articles. The features that were found to be most useful in identifying gender bias were single words, adjectives, verbs and domain specific lexicons.

Evidence of gender bias was identified in the newspaper coverage of politicians in Ireland. Some of these findings align with findings of previous research on the portrayal of female politicians in the media. Other findings are new and highlight some concerns regarding how female politicians are evaluated by the media and stereotypical portrayals regarding policy and personal issues.

This research presents a corpus-driven approach to analysing gender bias in the coverage of politicians in the media. They show how machine learning algorithms can be used to highlight differences in the coverage of male and female politicians and how these patterns can then be analysed to identify gender bias. This approach facilitates large scale analysis of texts thus addressing some critiques regarding the generalisability of some studies in gender and language that focus on smaller samples of text (Baker, 2014; Neuendorf, 2011). These findings contribute to the current body of literature on the representation of female politicians in the media and present new topics for the analysis of gender bias in media content.

This research presents a broad exploration of patterns of difference in how male and female politicians are represented in the media. These findings could form the basis for further in-depth studies of gender bias. In evaluating a range of approaches to using text classification for identifying gender bias, this research presents a new methodology for researchers interested in examining the representation of female politicians in the media.

References

Neuendorf, K. A. (2011), 'Content analysis a methodological primer for gender research', *Sex Roles* **64**(3-4), 276–289.

Baker, P. (2014), *Using Corpora to Analyse Gender*, Bloomsbury, London.