

1 Abstract

User engagement plays an important role in the consumption of news media. Research in this area has focused on creating optimal models to predict this user engagement but has placed less emphasis on the features used to produce these models. In this research the features will be analysed to determine the correlation found between the selected features and the chosen user engagement metrics on an article and the reasons behind these correlations.

Using data scraped from the-journal.ie experiments were set up using machine learning models to predict the user engagement metrics. These models used feature sets containing the data the reader had access to before engaging with each metric. The data used was gathered from the website, and external libraries were used to extract more features for the final feature set. The models produced coefficients relating to each of the features that determined the relationship between those features and the user engagement metrics.

Identifying the importance of these features is the major contribution of this research. The research shows that the title and blurb were more important than the article as predictors of user engagement, likely meaning readers decide on interacting before viewing the article. This has implications for the use of features from the article as predictors of user engagement, as it appears that the decision to interact often occurs before the reader views the article. Another key finding is that figure-based information in the title and blurb is found to be negatively correlated to both the number of views and comments received. The more concise the information provided in these sections, the less need a reader will feel to view them and these articles also tend to be less controversial and therefore receive fewer comments. It is also found that time-related features have a significant impact on the viewing and sharing metrics for an article, while having a substantially smaller effect on commenting behaviour. Overall, this research finds strong correlations between certain features in the articles and the user engagement metrics analysed.