Week 5 Project Proposal – Group 46

Title: To explore gender biases in film scenes using machine learning.

Students: Cliona O'Doherty, 15315679
Hannah Craddock,

Motivation: There is a definite gender bias in film media that can be quantified with ML measures such as the GD-IQ[1], and image analyses that reveal significant differences in the visual representation of male and female characters[2]. In the 100 highest grossing (US domestic) films of 2014-2016, female on-screen and speaking time was at 35%, and men were found to be seen and heard twice as often as women. Biases appear in where female characters appear, with them being more likely to appear in static scenes such as indoors. Here, we delve deeper into the prevalence of gender representation biases in film. By exploring the pattern of co-occurrence of visual objects as they appear in movie scenes we will test whether presence of a male or female can be predicted by a classification model. The best predictive features for each gender may shed light on stereotypical gender associations that are perpetuated in film and media and, by extension in everyday life.

Dataset: We will explore our hypothesis in a dataset of labels generated from 158.4 hr of feature length films. Labels were automatically generated using Amazon Rekognition, returning a set of labels for movie frames at a 200 ms sampling interval. The dataset comprises 2,851,272 timepoints, with a total of 41,330,953 labels of which 2,466 are unique. The dataset was collected by a member of the Cusack Lab at Trinity.

ML techniques being applied: The main portion of this study will train a classification model to predict either male or female from the labels that appear within frame, and in preceding frames. We will use methods for working with text and time series features. We will also perform an association analysis to investigate whether labels that are more strongly associated are more stereotypically associated with the gender to which they are coupled.

Experiments for evaluation: The classification performance of the model will be evaluated using ROC curves and F1 measures, comparing to appropriate baselines. In the case that good classification performance will not be achieved from the movie label data, the association analysis will still reveal interesting insight into the gender links that are present within the movies.

---

[1] https://about.google/intl/en_us/main/gender-equality-films

[2] Jang, J.Y., Lee, S. and Lee, B., 2019. Quantification of Gender Representation Bias in Commercial Films based on Image Analysis. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), pp.1-29.