Maggie Kidd, Katherine Steiner

Introduction

For our machine learning final project, we were interested in understanding what combinations of movie attributes make movies more or less likely to pass the Bechdel Test. A film can pass the Bechdel Test on three different levels. A movie passes the test if it:

1. has at least two named female characters,
2. who speak to each other,
3. about something other than a man.

The test is important because it highlights different trends in the movie industry and brings attention to the fact that not every movie depicts women as multifaceted people. As women in a male-dominated industry, we appreciate visibility and representation in movies and we think it would be interesting to look at how our chosen attributes contribute to the likelihood that a movie will pass the test. Our model might also be a good choice for someone who is trying to make a movie, because we think our test will link a lot of different successful movie attributes in potentially helpful ways.

Data and Attributes

We used Bechdel Test Movie List's API to get information on which movies pass and which movies fail the Bechdel Test. These API calls returned the movie's score, a integer in the range 0 to 3, which represents the number of criteria the movie meets (so a movie with a score of 0 completely fails the Bechdel Test, a movie with a score of 3 completely passes). Generally speaking, we expect "similar" movies to have the same score. Therefore, in order to get a more divisive data set, we only considered the movies with a score of 0 or 3. This {0,3} value we used as our nominal class. The API calls also returned the movie's title, year, and Internet Movie Database (IMDb) id.

To get more information about these movies, we used the Open Movie Database's (OMDb) API. By using the IMDb id obtained from the Bechdel Movie Test List's API call, we were able to get each movie's country of origin, short plot description, production company, rating (e.g. PG-13), genre, writers, directors, and IMDb rating (e.g. 8.1/10).

In order to get all this data into a format that we could use, we did some preprocessing. In particular, some movies had multiple countries of origin, so we only used the first country listed as an attribute. Similarly, some movies had multiple genres, so again we just used one. We found a list of 1000 popular boy and girl names in the US. Then, we extracted the director's first name and compared it to these lists to decide the likely gender of the director. We did this also for the movie's writers. Finally, we came up with a short list of “feminine-coded” words (she, her, hers, woman, women, girl, morther, daughter, wife) and counted the number of times these words appeared in the short plot description, using this number as an attribute.

Building the Model

We split the data we had into 70% (~900 instances) to use for a training set and 30% (~380 instances) to use for a test set. Our original goal in building a model was to get more accuracy on our test set from our training set than using a ZeroR model; ZeroR was chosen here because it simply predicts the mode of the training instances for each of the test instances, giving us a minimum accuracy for what we should consider a good model. Trying ZeroR, we got 34% correctly classified instances.

In WEKA, we tried a couple of different models on our data before we settled on one that gave the highest results. KStar is an instance-based classifier that is similar to nearest neighbor, determining the class of a test instance based on the class of similar instances in training, using an entropy-based distance function. KStar, when run on the test set without any fine-tuning of the parameters, gave a result of 73% correctly classified instances. However, when we increased the parameter "globalblend" from 20 to 30, the accuracy went up 2%, resulting in 75% correctly classified instances.

Analyzing Data

After we ran the model on our data, we used WEKA's "Visualize Data" Tool to check out how much of different attributes comprise passing or failing movies. We saw that one convincing attribute was how many female and male writers helped write the movie. In each of the following graphs, red means that the instance passed the Bechdel Test, while blue means that the instance fails.

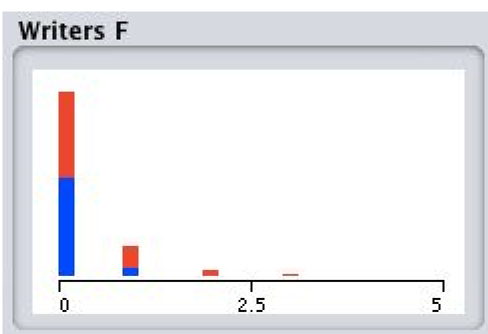

Fig. 1: Number of female writers

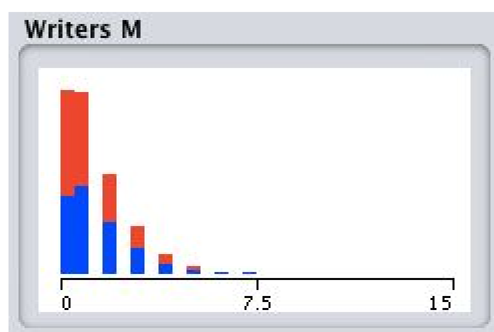

Fig. 2: Number of male writers

While there are not as many instances with female writers, we can see that red, meaning that the instance passed the test, dominates the female writers' graph. Movies with no female writers tend to be more evenly split among passing and failing movies. However, as the number of female writers increases, we see that more of these movies passed the test. Movies with more male writers are shown to not pass the Bechdel test as often.

Similarly, we can take a look at how the gender of the director affects the likelihood of a movie to pass or fail.

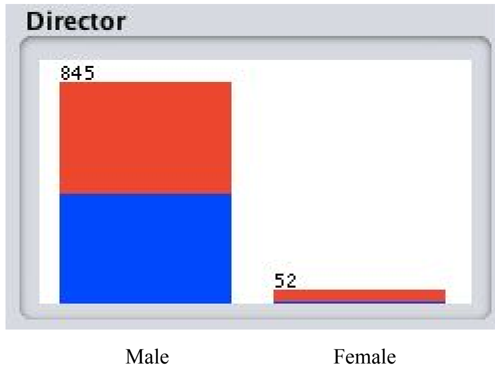

Fig. 3: Movies with male/female directors

It is clear that when a movie has a male director, it has about an even chance of passing or failing the Bechdel test. However, when a movie has a female director, it appears to have over 80% chance of passing the test.

One more attribute that is not necessarily extensible without the use of prediction analysis, but is interesting because it proves our intuition, is the year the movie was produced. As expected, older movies tend to fail the Bechdel Test more often.

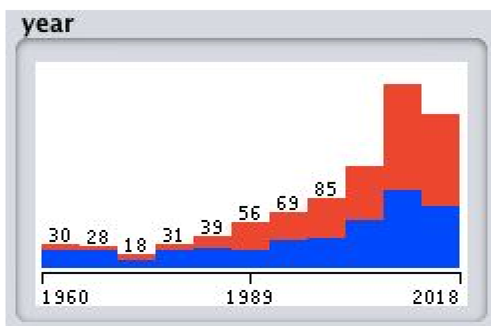

Fig. 4: Year movie was produced

It appears as if movies that were produced before the year 1985 predominantly did not pass the Bechdel Test, and from then on movies were evenly likely to pass or not pass.

Who did What

Katherine - Data Processing and Modeling, Website
Maggie - Data Analysis, Write-Up