40.016: Analytics Edge

Week 4 Lecture 1

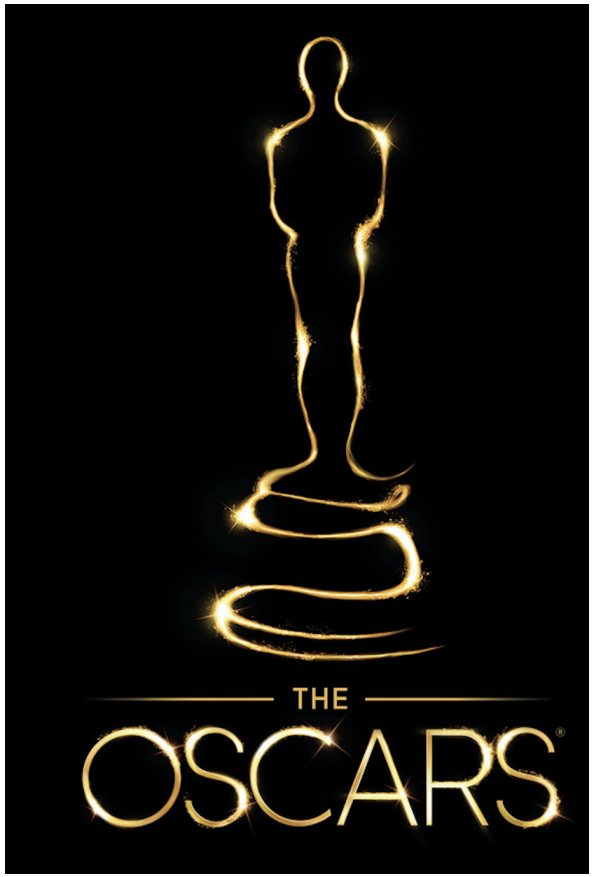
Predicting the Oscars

Term 6, 2020

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

The Academy Awards

- Popularly known as the Oscars.
- Started in 1929.
- The Academy of Motion Picture Arts and Sciences (AMPAS).
- Honors excellence in acting, directing, writing, editing, etc.
- Primarily English language movies
- . . . made by a US production house . . .



Award: nomination and voting

- Eligible voters: Around 6000 members of AMPAS from different disciplines (actors, writers, designers, directors)
- Each branch of AMPAS vote to determine nominees from their own category.
  – In special cases like Best Picture, everyone votes.
- Typically the top 5 voted in each category are chosen to be Oscar nominees.
- Nominees announced in around January.
- All members (in most cases) vote for deciding final winners.
- Oscars ceremony occurs late February , early March.
- `https://www.youtube.com/watch?v=8KeOxeuiZjs`

Effect of the Academy Awards

- Viewership
  - 1998 (70th Academy Awards): 57 million viewers
    Best picture: Titanic
  - 2015 (87th Academy Awards): 37 million viewers
    Best picture: Birdman
  - 2020 (92nd Academy Awards): 23.6 million viewers
    Best picture: Parasite
- Significant impact on ticket sales (see `http://boxofficemojo.com`)
- Winning an Oscar leads to many better deals for the parties involved.
- Of course, an early prediction may lead to having an arbitrage opportunity for some.

How can on predict the Oscars

- Is it possible to predict the Oscars awards with any degree of accuracy?
- Many predictions are there using the buzz of awards, and critics choices.
- Develop a data driven approach.
- Follow: **Applying Discrete Choice models to predict Academy Award winners** by I. Pardoe and D.K. Simonton in *JRSS(B), 2008* .
- Prediction is done after the 5 nominees are announced.
- Prediction for four categories: Best picture, Best director, Best actor and Best actress.

Multinomial Choice Model

Multinomial Choice Model

Multinomial Choice Model

Multinomial Choice Model

Multinomial Choice Model: estimation

Quality of fit

1. Akaike's Information Criterion (AIC):

$$AIC = -2LL + 2p$$

where LL: Log-likelihood, p : no. of parameters. Smaller the AIC, the better.

2. Likelihood ratio index (McFadden's index)

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

where $\hat{\beta}$ is the estimated value of the parameters. LL(0) refers to the log-likelihood when all the parameters are set to 0 (no model).

3. Percent correctly predicted
This identifies the alternative with the highest probability for each individual observation and determining whether or not this was what the actual choice was.