Predicting Oscars Through Semantic Web Mining

Krauss, J. Nann, S. Fuehres, H. Gloor, P.

This paper extends earlier work to predict Academy Award nominations based on online forum analysis. Through applying a bag-of-words approach in combination with network metrics to the Oscar Buzz subforum at the Internet Movie Database (IMDb) we make predictions for Academy Awards winners for the main Oscar categories: best picture, best director, best actor, best actress. A bag-of-words specifically adapted to communication patterns occurring in the Oscar Buzz subforum leads to significantly higher precision and recall ratios. Consideration of network position of the posters reduces the impact of spammers and increases overall prediction quality.

Communication at IMDb takes place in an unstructured manner. Thus, textual processing and analysis are necessary to identify the discussion's subject (which movie/director/actor/actress) and determine the semantic orientation expressed towards the subject. Through calculating a degree of association with Oscar keywords and weighting it by the poster's network importance we derive the likelihood for a movie/director/actor/actress of winning an Academy Award. The results obtained through IMBb analysis are additionally weighted by Web and Blog buzz analysis through degree-of-separation search on the Web and in Blogs for the same keywords.