From Crowd Ratings to Predictive Models of Newsworthiness to Support Science Journalism

Paper Link

Authors:

- Sachita Nishal
- Nicholas Diakopoulos

Notes

- Their work comes at the problem from the scientific paper - essentially they are trying to predict whether or not a scientific article might make an interesting news article (as opposed to our work which analyses the news article to attempt to understand which part of the scientific paper is worth mentioning)
- Their definition of newsworthiness is broad and incorporates scientific and social impact but also factors in stuff that we would probably consider less helpful for understanding impact like how controversial the work is.
- They use computer science pre-prints gathered from ArXiv via their API (filtered by category/subject area).
- Their data set is 50 papers in 'train' set, 55 in validation set
- Scientific journalists were paid as experts. Non-specialists were recruited for crowd-sourcing at larger scale (MTurk)
- Crowd-source workers asked to score papers based on the abstract (full text was optionally available but only for gathering additional context/if necessary).
- Workers asked to score along 4 dimensions: actuality, surprise, impact magnitude, impact valence
- They use Likert scales for crowd-rating of newsworthiness
- Crowd source workers achieve moderate associations with the "expert" annotators.

- They train an extra trees random forest classifier with Sentence-BERT feature input (single 768 dim vector used for each article abstract). Additional features include binary indicator if article makes code or data available (hypothesis is that this could increase impact)
- They rank outputs based on likelihood score from the classifier (Extra Trees classifier) and achieve reasonable results

---

Revision #5
Created 2022-11-16 09:52:28 UTC by James
Updated 2024-02-11 16:27:03 UTC by James