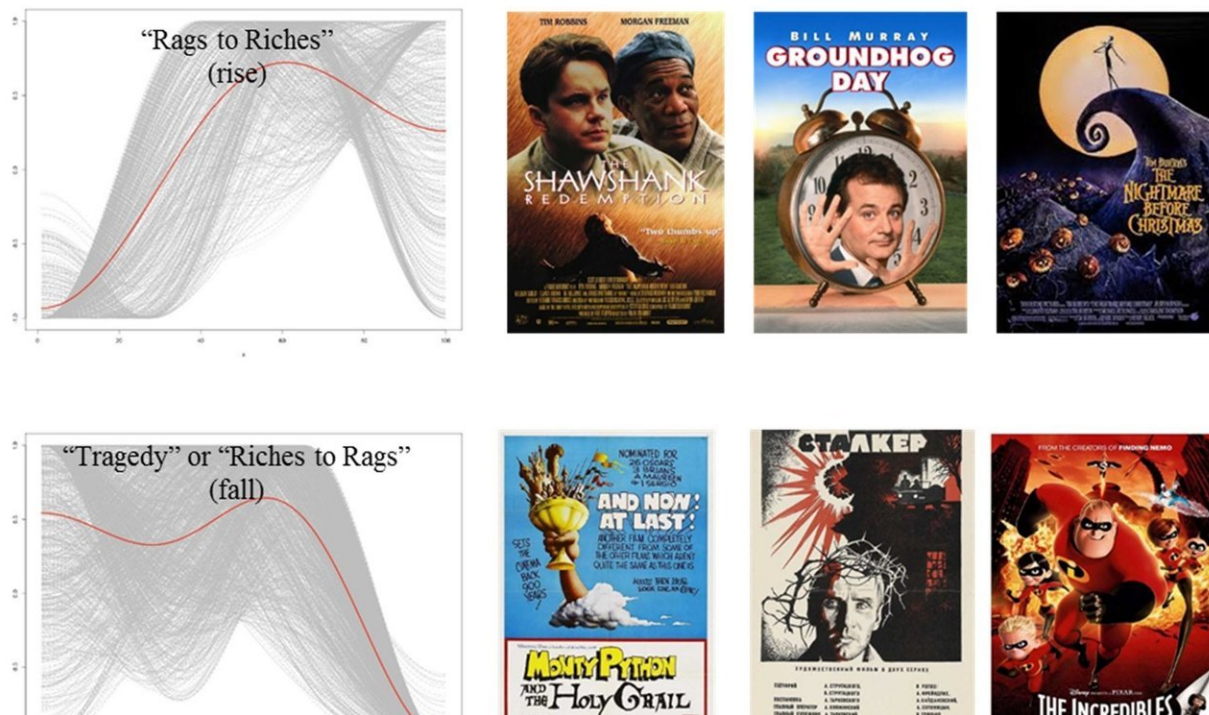# Researchers use machine learning to analyse movie preferences

July 13 2018, by Ingrid Fadelli

Credit: arXiv:1807.02221 [cs.CL]

Could behavioural economics and machine learning help to better understand consumers' movie preferences? A team of researchers from the University of Cambridge, the University of West England, and the Alan Turing Institute dove deeper into this question, in a fascinating study that combines behavioural economics, business and AI.

Marco Del Vecchio, Alexander Kharlamov, Glenn Parry, and Ganna Pogrebna used their diverse skillsets to develop tools that could help the media industry to better understand what content viewers really want to see. Currently, the motion picture, media and entertainment industry selects content offerings based on top-down decisions, typically informed by expertise, experience, surveys and focus groups. "Our main motivation was to understand whether and to what extent we can put viewer perceptions at the heart of the equation," the researchers said.

Their study focused on the emotional journeys of movies, investigating whether these fall in different categories, and whether they are related to a movie's success. The researchers used a dataset of 6,174 movies, each with complete scripts, revenue data, IMDb ratings, and other relevant information.

Using natural language processing (NLP) algorithms, they analysed the movie scripts to determine their emotional journeys and then used these results to explore the relationship between a movie's emotional journey and its success, both in terms of revenue and public reception.

The researchers found that, similarly to novels, stories in movies fit into six main story arcs, or types of emotional journeys that the viewers experience:

- Rags to Riches: "An ongoing emotional rise" (e.g., *The Shawshank Redemption*, *Groundhog Day*, and *The Nightmare Before Christmas*)
- Riches to Rags: "An ongoing emotional fall" (e.g., *Psycho* and *Toy Story 3*)
- Man in a Hole: "A fall followed by a rise" (e.g., *The Godfather*, *The Lord of the Rings: The Fellowship of the Ring*, and *The Departed*)
- Icarus: "A rise followed by a fall" (e.g., *On the Waterfront*, *Mary*

Poppins, and *A Very Long Engagement*)
- Cinderella: "Rise-fall-rise" (e.g., *Rushmore*, *Babe*, and *Spider-Man 2*)
- Oedipus: "Fall-rise-fall" (e.g. *All About My Mother*, *As Good as It Gets* and *The Little Mermaid*)

Movies in the "Man in a Hole" category had the highest box office rankings, as well as the greatest gross worldwide and domestic revenues, irrespective of their genres and production budgets. "'Man in a Hole' succeeds not because it produces the most 'liked' movies, but because it generates the most 'talked about' movies," said the researchers. "The number of IMDb ratings given, as well as the number of user and critics reviews are a lot higher for 'Man in a Hole' movies than for movies in any other emotional arc category."

Despite these movies' better average performance, the researchers note, "It would be an oversimplification to say that the motion picture industry should only produce 'Man in a Hole' movies. A carefully chosen combination of production budget and genre produces a financially successful movie with any emotional shape."

For instance, the Icarus emotional arc was particularly effective for low-budget movies, while the Riches to Rags shape were more likely to be successful with larger budgets of over $100 million.

Science fiction, mystery, and thriller films with happy endings ("Rags to Riches" shape) and comedies with a bad ending ("Riches to Rags" shape) did not perform well at the box office, while "Oedipus"-shaped movies did not do well at award ceremonies and festivals other than the Oscars.

"Our findings and the tool we are working toward may ultimately help writers optimize their scripts during editing or inform producers who have to make an investment decision when faced with a choice between

projects," the researchers said.

Pogrebna and her colleagues are now are seeking out industry partners who can provide them with further data for their studies.

"In the future, we would like to create robust methods to analyse sentiment in all media, including nonfiction such as documentaries and shorter videos such as those on YouTube. Once we have optimised the tool, it would be good to spin out a company that can commercialise the work and get it into the hands of industry colleagues."

Abstract

Much of business literature addresses the issues of consumer-centric design: how can businesses design customized services and products which accurately reflect consumer preferences? This paper uses data science natural language processing methodology to explore whether and to what extent emotions shape consumer preferences for media and entertainment content. Using a unique filtered dataset of 6,174 movie scripts, we generate a mapping of screen content to capture the emotional trajectory of each motion picture. We then combine the obtained mappings into clusters which represent groupings of consumer emotional journeys. These clusters are used to predict overall success parameters of the movies including box office revenues, viewer satisfaction levels (captured by IMDb ratings), awards, as well as the number of viewers' and critics' reviews. We find that like books all movie stories are dominated by 6 basic shapes. The highest box offices are associated with the Man in a Hole shape which is characterized by an emotional fall followed by an emotional rise. This shape results in financially successful movies irrespective of genre and production

budget. Yet, Man in a Hole succeeds not because it produces most "liked" movies but because it generates most "talked about" movies. Interestingly, a carefully chosen combination of production budget and genre may produce a financially successful movie with any emotional shape. Implications of this analysis for generating on-demand content and for driving business model innovation in entertainment industries are discussed.