Arthur CHANSEL
SCIPER : 324265
08/10/2023

ADA Milestone P1
Dataset: CMU Movie Summary Corpus

Idea 1: Overproductive actors

Is it worth being overproductive in cinema? Not all actors played in so many movies and, conversely, some don't even count their screen appearances anymore. So, who are them? Could we draw similarities between them and contrast them between epochs? On the other hand, very iconic personalities only made a few movies. Probably that the most successful actors spend time on a role to make it perfect and can't afford to play in so many movies. It would be interesting to investigate the overproductive-actors, their revenue, their success, their roles, which actors they play with, and more. What benefits did this strategy bring them?

Idea 2: Pivotal movies

Which movies were a game changer in the history of cinema, and why? When we think about Star Wars, we think of a great success, of an incredible story, but also of a turning point with groundbreaking technology, revolutionizing special effects in cinema and science fiction. We are seeking these pivotal movies out by filtering movies' distribution over time, with specific themes, genres, and more. Then we could understand such dynamics, but also if prior similar movies exist but didn't generate such engagement before. Pivotal movies might be related to technology improvements, such as arrival of sound or colors in movies but not only: other worldwide events also impacted the industry. We saw many movies about war after big conflicts, or cartoons correlated to a new audience. Here, we'd like to find out which movies played that role of milestone, and for what reason!

Idea 3: Movies, the mirror of society

What does the society think? We assume that movies have better success when we can identify with something in, and we could consider them as a mirror of society somehow. Thus, it would be interesting to observe the variation of societal values through movies over years. We'll investigate characters attitudes like violence or love, and world representations like nature or authority, and more. We might even find out themes that we haven't thought of. We'll search if a movie summary contains keywords that we would attribute to a behavior (i.e. "torture" or "kill" for violence). We could also determine how a film highlights a behavior, positively or negatively. Over a period, we would see if a behavior was promoted at the beginning and discredited then, and we might see the switch happen at a specific time. Text mining, NLP methods and supplement data from Stanford CoreNLP pipeline could help us to identify key behaviors. Then, crossing our results with other parameters, such as box office or number of movies with similar attributes could tell us about the engagement or importance of a trend. Finally, we would have a better idea of what society sees most on screen and what it might think at different times.

Feedback

Idea 1 Very novel idea! It is interesting to see how the productivity of actors affect their revenue and success. It may be hard to isolate the effect of productivity from other factors, such as plots and genre. So be careful when you do the analysis and interpretation. But I think this will be a very interesting project. **Idea 2** Very interesting idea. It would be very cool if you could rediscover some of the pivotal movies in the history of cinema via data analysis. You will need to define a clear criteria for pivotal movies. And show how your method of discovering pivotal movies is consistent with the history of cinema. Also avoid overfitting the data if it involves machine learning. **Idea 3** Meaningful idea. NLP is a good way to analyze the text data. You gave an example of "torture" and "kill" for violence. You can try to give a more principled way of identifying the key behaviors. Your project mentor throughout the semester will be Saibo Geng : saibo.geng@epfl.ch. For future discussions specific to your P2 and P3 deliverables, you are encouraged to be in touch with your mentor.