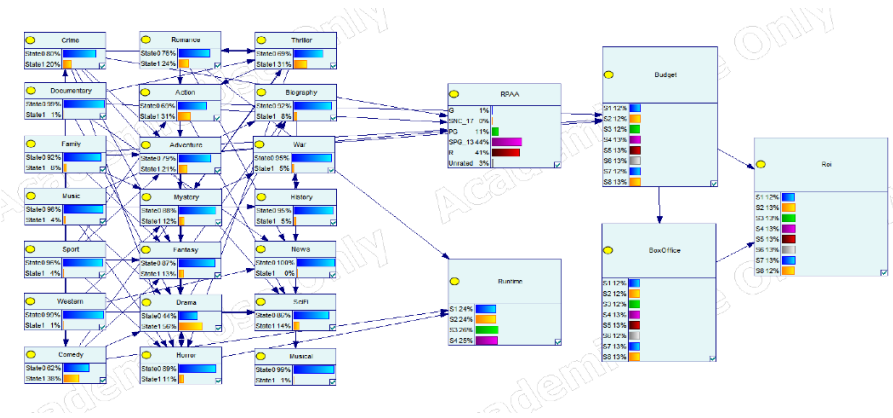
Identifying Temporal Variation in Movie Success Factors with Bayesian Networks

Marchetto U., Bertolotti F., Roman S., Piras A., Mari L.

Causal inference is a set of methods for identifying cause-effect relationships between variables to understand how changes in one variable affect another. Bayesian networks are one of these methods and graphically represent how variables, the nodes of the network, affect each other. These relationships are represented by oriented arcs and imply the existence of non-independence between variables.
Bayesian networks can be applied to different types of management problems, as a tool to identify and understand business risks, in particular to understand which factors affect profit. In this study, Bayesian networks were used to assess the factors that influence the profitability of films in the US market between 2000 and 2019.
The data used for this analysis was obtained from a previous work and classifies 2333 films, and the preliminary analysis was performed using a data visualisation tool to gain insight into their distribution patterns and make informed decisions about the subsequent discretisation. Once sufficient a priori information had been obtained on the dataset, the Bayesian network was constructed with GeNie to assess which elements influence the economic success of a film, including a priori knowledge to ensure that the arcs were properly oriented. Once the network was built, the forces of influence between the nodes and the main sensitivity factors were analyzed using tornado diagrams.



While the initial analysis was conducted on the full dataset, it was later divided into two subsets, the first comprising films released between 2000 and 2009, and the second films released between 2010 and 2019. A number of interesting trends and juxtapositions of variables emerged, including a strong influence of adventure and action films, an index of film classification by age juxtaposed with films labelled 'family', 'thriller' and 'romance'. For the period before 2010, this mode is less pronounced. However, the presence of documentaries, westerns, news and dramas, accompanied by the emerging influence of action films, stands out, especially in terms of their profitability potential in relation to their budget. In terms of the post-2010 period, a greater success for fantasy and science fiction is emerging, with a juxtaposition with the adventure genre. The relationship that appears most frequently in all three networks analysed is the composition of different values in terms of ROI level, specifically Budget-RPAA-Action-Adventure. To the best of our knowledge, this analysis has never been attempted before, and it makes it possible to use causal inference to gain a better understanding of how systems evolve from changes in the models trained on the data they generate.