PREDICTING THE DIFFUSION OF ONLINE VIDEO CONTENT

Increasing expenditures on online video advertisements, product placement and brand video channels has made companies keen to know the future popularity of such online content. Although there is increasing research activity, there is no reliable forecasting method available that can accurately predict the diffusion process of online videos. With the assumption that such content is likely to follow a product lifecycle process, this study focuses on evaluating the general applicability of the original Bass model. It i) highlights the current research status in forecasting and segmenting of online video content; ii) provides an empirical analysis of the Bass model; and iii) emphasises the research gaps towards the development of new online video diffusion models.

Challenge Overview

In a similar fashion to their use of TV ads companies want to be sure that their marketing budget is used efficiently, reaching a high number of viewers in a reasonable time. Such an evaluation is needed because many companies now invest a substantial amount of money to produce their video content or agreed on costly product placements. With the assumption that such content is likely to follow a product lifecycle process, marketers are keen to know the current life cycle phase, which then specifically allows actions that prolong the video's popularity or suggests to replace such content. This idea lends itself to the use of diffusion models, such as the Bass model which has been used widely to predict the long term life cycle of new product innovations.

Although, there is an increasing research activity in the field of the online video popularity prediction, there is no standard method established yet. One of the main issues is that the social media content tends to have bursts and spikes depending on the viral widespread. Moreover, studies have reported that several viewing pattern exists. Thus, there is no guarantee that one model can fit all patterns equally well. In addition, segmentation of such patterns implies a secondary challenge. Moreover, many of the proposed models have not been assessed very extensively and, therefore, empirical evidence for plausible forecasts is needed.

Solving the problem

The empirical evaluation was performed on two independent YouTube datasets containing 49 and 149 time series. The former dataset was divided into five clusters according to the K-Spectral Centroid (K-SC) algorithm, which is suggested in the literature as suitable for clustering online viewing patterns. The well-established Bass model is compared alongside three benchmark univariate time series models, namely ARIMA, ETS and Naïve using R software.

Results and Achievements

Overall, the Naïve model performs best! However, on two out of five clusters, accounting to about 50% of the sample videos, the Bass model shows superior performance. This result leads to the conclusion that the Bass model is well suited well for forecasting some type of online content. In addition, clustering showed its effectiveness and the findings suggest that it is likely that more than one forecasting model is necessary due to the heterogeneity of online behaviour. Nevertheless, additional research is needed to improve the Bass model and adapt it to the purpose at hand to become more broadly applicable. Future research directions include improving the parameter estimation, incorporating explanatory variables and consider further extensions to increase model flexibility and overcome the assumed unimodality in the diffusion of a product that limits the model's applicability to online video.