The Lifecycle of a Youtube Video: Phases, Content and Popularity

Honglin Yu, Lexing Xie & Scott Sanner, Australian National University, NICTA

1 The Problem
• How to describe and measure popularity over time?
• How to better predict popularity?

2 Main Contributions
• New representation: popularity phases.
• New method: phase extraction algorithm from popularity history.
• A large-scale, longitudinal measurement study of popularity.
• Better prediction of future popularity using phase representations.

3 Phase Detection
\[ p(t) = at^c + e \]

Figure 1: The complexity of viewcount dynamics: the lifecycles of four example videos. Blue dots: daily viewcounts; red curves: phase segments found by our algorithm. (a) A video with one power-law growth trend. (b) A video with one power-law decay. (c) A video with many phases, including both convex and concave shapes - this video contains a gymnastic performance. (d) A video with seemingly annual growth and decay - this video demonstrates how to vent a air-conditioner, and reaches peaks during each summer. Viewcount shapes such as (a) and (b) are explained by Crane and Sornette’s model (PNAS 2008), but (c) and (d), and many more like them, are not.

4 Dataset

Figure 3: Left: Boxplots of video viewcounts at \( T = 735 \) days, for popularity percentiles quantized at 5% each. Viewpoints of the 5% most- and least-popular videos span more than three orders of magnitude, while videos in the middle bins are within 30% views of each other. Right: The change of popularity percentile from 1.5 years (y-axis, from 0% to 100%) to 2 years (x-axis, in 5% bins). While most videos retain a similar rank, videos from almost any popularity at 18 months of age could jump to the top 5% popularity bin before it is 24 months old (left most boxplot).

5 Observations

Figure 4: Four types of phase shapes and their basic statistics. Right: Red curves are the probability of a video having a new phases in 15-day intervals over time, broken down by phase types. Blue curve is the average daily viewcount.

Phase, Video Type and Popularity

6 Viewcount Prediction

Code/Dataset: https://github.com/yuhonglin/ytphasewithdata

Figure 5: Left: Percentage of videos broken down by the number of phases they have, over (a) popularity percentile and (b) content categories. Middle: Percentage of the four phase types, broken down by (c) popularity percentile and (d) content categories. A general trend is that popular videos and entertainment content (e.g. music videos) have more phases overtime, and more than half of news videos and the least popular videos have one dominant decreasing phase.

Concave phases

Figure 6: Popular and entertainment videos have more concave shapes. Such phase shape cannot be generated from Crane-Sornette model, our ongoing work focus on a generative model that can explain all phase shapes.

Phase types of the most popular videos

Figure 7: Phase type and popularity evolution for the top 5% videos over time. We can see that videos that have jumped by more than 30% in popularity either have new phases or have been in a continuously increasing phase.

6 Observations

Figure 8: Mean normalized MSE for the baseline and phase-informed prediction over different pivot dates (x-axis) for videos with less than 5 phases, broken down by the shape of the last phase of \( x_{t+r} \), \( \Delta t=15 \) days.

Baseline: Multi-linear regression
Phase-aware: Use phase feature to group videos and train separate models for each group.
Phase-informed prediction consistently out-perform baseline approach across all phase types and task settings.