Technical Disclosure Commons

Defensive Publications Series

December 2020

Predicting Content Views Using Finite Integrals

Brian Mulford

Michael de Ridder

Colby Ranger

T. J. Gaffney

Jay Mithani

Predicting Content Views Using Finite Integrals

ABSTRACT

Video hosting and sharing services enable creators and advertisers to create campaigns that engage viewers. To price the advertisements, and to give advertisers on the campaign an idea of the popularity of the content, the viewership is predicted. Both under- and over-prediction of views are associated with penalties, respectively of wasted inventory and capacity crunches. View estimations based on channel average suffer from sample bias and invisible trends. This disclosure describes techniques of in-flight view prediction, e.g., predictions of views done after the launch of a campaign for the remaining days of a campaign. The predictions of the total views on a line-up of in-flight videos are based on the distributions of prior view history. The described predictor delivers continuously improving predictions for live videos, and enables determination of whether a campaign is meeting view goals. It thereby enables real-time fine-tuning of inventory and capacity for the remaining days of the campaign.

KEYWORDS

- View prediction
- Video sharing
- Video hosting
- Advertising campaign
- Log-likelihood estimation
- Advertising inventory
- Riemann sum
- Finite integral

BACKGROUND

Video hosting and sharing services enable creators to create a set of videos that serve as a campaign to engage with viewers. Such services can facilitate agreements between creators and brands that want to advertise on the campaign. A campaign can last a certain duration, e.g., thirty days. To price the advertisements, and to give potential advertisers on the campaign an estimate of the likely popularity of the videos, viewership prediction techniques are employed.

Under-delivery of viewership (less viewers than predicted) can leave advertisers feeling shortchanged. To compensate, additional creators may have to be activated at a late stage, which is expensive and drains capacity. Over prediction (more viewers than predicted) results in wasted capacity and non-realized proceeds from excess view efficiency. View estimations based on channel average (e.g., for a creator's channel) can suffer from sample bias and invisible trends. Overall, inaccurate view predictions result in lost revenue or revenue potential.

Viewership can be predicted prior to or after the launch of a campaign. View prediction done after campaign launch is known as in-flight prediction, and can be used to fine-tune predicted viewership for the remaining days of the campaign.

Without tools to understand how an in-flight campaign is performing relative to its expected prediction, it is difficult for account managers at video hosting/sharing services to accurately judge when there is a need to involve or hire more creators. Since there is a contractual requirement for campaigns to meet their views goals, missing a view goal results in longer campaigns as more creators need to be involved, reducing throughput for other campaigns. On the other hand, exceeding a view goal by a large margin results in wasted inventory, which can also negatively affect other campaigns.

Existing techniques of view prediction are based on fixed time intervals (e.g., 30 days) from the moment a video is made public to a selected interval are based on historical view distributions. While useful, such view predictions don't factor live videos that accumulate views to condition the predictions. The nature of views being difficult to predict, prior distributions that are inaccurate are difficult to detect until close to the end of the time interval.

DESCRIPTION

This disclosure describes using parametric simulations to estimate, or predict during flight, the distribution of sums of content views, which is used to estimate, for example, the probability of exceeding a given view-threshold. This process of adding simulations amounts to approximating the Riemann (finite sum) integral of the k-fold convolution of probability distribution functions. The techniques deliver continuously improving predictions for live videos, enabling determination of whether a video (or campaign) is meeting view goals or not. The described conditioned model results in predictions of high accuracy (with an error of less than 5%), enabling near real-time adjustments to campaigns.

Per the techniques, a distribution is created of the ratio of the number of views v_d on the dth day of the campaign to the number of views ($v_{d+1}+v_{d+2}+ \ldots + v_N$) on the remaining days of the campaign for historical videos that are similar to the videos of the present campaign. Here, N is the total number of days in the campaign.

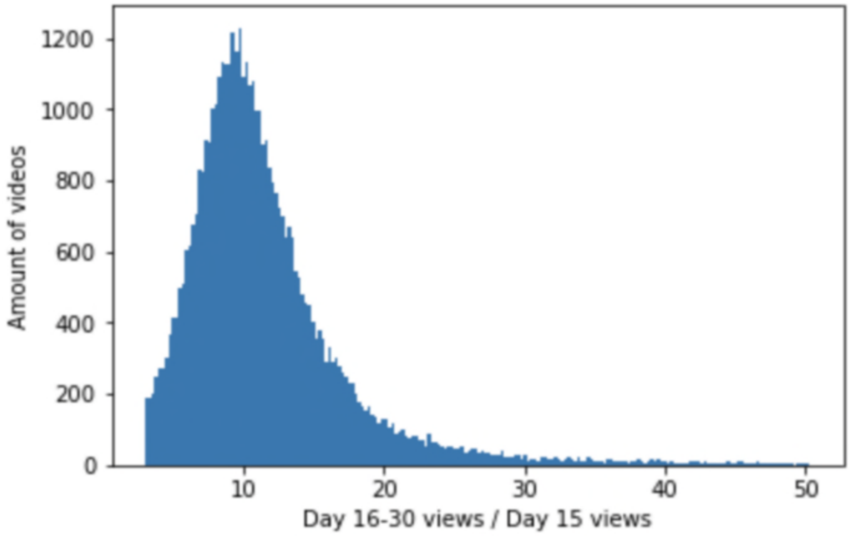

Fig. 1: Example histogram of raw ratios of day 16-30 views over day-15 views

Fig. 1 illustrates an example of such a distribution, of the ratios of historical views between days $d+1$ and N to the number of views of day d. The y-axis is the number of videos of the sample set corresponding to the particular number of day 16-30 views. In this example, $d=15$ and the campaign duration N is 30 days.

From the distribution of Fig. 1, across the set of videos, the most likely number of views between days 16 and 30 is ten times the number of views on day 15. Thus, if a video received 500,000 views in the first 15 days, of which 30,000 took place on the fifteenth day, then the most probable number of views between days 16 and 30 would be 30,000×10=300,000. The most probable total number of views over the campaign duration (from day 1 to day 30) would be 300,000+500,000 = 800,000.

The known sums of the view-numbers of the videos of the present campaign at day d form another distribution, known as the line-up distribution. To estimate the number of views in the remaining days of the campaign (e.g., over the days $d+1$ through N), the current rate of change of views (e.g., the views gained the last day less the views on day d) is multiplied by scaling factors from the distribution of the ratio of historical views between days d+1 and N to the number of views of day d (e.g., the distribution of Fig. 1). Effectively, a sample of the predicted distribution of views between days d+1 and N is obtained by multiplying a sample of the line-up distribution at day d by a sample of the distribution of the ratio of historical views between days d+1 and N to the number of views of day d (e.g., the distribution of Fig. 1). The predicted distribution of views between days d+1 and N is obtained by repeating the above multiplication procedure a large number (e.g., >10,000) of times. In this manner, the techniques use prior view history, and the distributions thereof, to estimate the total views on the entire line-up of in-flight videos.

To get the probability that a target view goal is exceeded, the output of the lineup prediction simulations and the in-flight prediction simulations are pairwise summed. The ratio of the number of simulations that exceed a target view goal to the total number of simulations gives the probability of exceeding the target view goal.

Because the final output is derived from a combination of distributions, it is possible to assign a variance, or confidence level, to the prediction. For example, a combination of the lineup distribution with a *single* number from the in-flight view predictor (the raw ratio of Fig. 1) would result in decreasing variance as the number of published videos in a lineup increases. Eventually, a point would be reached (erroneously) with no variance, and all predictions at either 0% or 100%.

Measuring the accuracy of view prediction

To measure the accuracy of view prediction, the predictor is run against n past campaigns. The accuracy of prediction can be measured using measures such as the win fraction, the likelihood metric, and the fraction of caught failures, defined as follows.

- *Win fraction*: The win fraction is a measure of how often a given model wins over another. For each day of a campaign, a predictor model provides a probability that a campaign will meet a target view goal. If the campaign actually meets its target view goal, the model that returned the higher probability receives a win. If the campaign failed to meet its target view goal, then the predictor model that returned the lower probability receives a win. Performing this evaluation each day over all test campaigns gives a win fraction for each model. The win fraction of the in-flight predictor is higher than that of other predictors and improves with the number of days in the campaign.
- *Likelihood metric*: Likelihood is the probability of the observed outcome given a model. A higher likelihood metric is better. This metric is calculated as follows.

$$L(d) = \prod_{i=1}^{n} [p_{i,d} \cdot e_i + (1 - p_{i,d}) \cdot (1 - e_i))]$$

  where $L(d)$ is the likelihood metric on day d, $p_{i,d}$ is the predicted probability as of day d of hitting the view goal of campaign i, and e_i is 1 if campaign i hit its view goal at campaign's end (e.g., 30 days), and 0 otherwise. The described in-flight predictor outperforms other predictors on the measure of log-likelihood, with performance improving with the number of days a campaign has been in flight.
- *Fraction of caught failures*: The percent of future failures caught is a measure of how early a predictor determines that a campaign will fail before it actually reaches its end (e.g., the 30-day mark). Each day of a campaign, a predictor determines whether or not to warn that a campaign is behind pace. If this campaign is one that actually does fail, it is considered a future failure that the in-flight predictor caught. If this campaign ends up meeting its target view goal, it is considered a false positive. Performing this evaluation over all campaigns gives the percentage of future failures that are caught and the false-

positive rate. As the number of in-flight days increases, the percentage of failing campaigns that a predictor is able to catch tends to 100%. Additionally, the false-positive rate is found to be 0% or nearly so. The predictor can thus be utilized to reliably provide a dashboard feature that displays a failing-campaign warning.

The described view-prediction techniques can be combined with other short-term (e.g., over the length of a few hours) or long-term (over days or weeks) predictors to arrive at a consolidated view-prediction.

CONCLUSION

This disclosure describes techniques of in-flight view prediction, e.g., predictions of views done after the launch of a campaign for the remaining days of a campaign. The predictions of the total views on a line-up of in-flight videos are based on the distributions of prior view history. The described predictor delivers continuously improving predictions for live videos, and enables determination of whether a campaign is meeting view goals. It thereby enables real-time fine-tuning of inventory and capacity for the remaining days of the campaign.