

## Popularity prediction in content delivery networks

Nesrine Ben Hassine, Dana Marinca, Pascale Minet, Dominique Barth

► **To cite this version:**

Nesrine Ben Hassine, Dana Marinca, Pascale Minet, Dominique Barth. Popularity prediction in content delivery networks. IEEE 26th Annual International Symposium on Personal, Indoor and Mobile radio Communications (PIMRC), Aug 2015, Hong Kong, China. 10.1109/PIMRC.2015.7343641 . hal-01244908

**HAL Id: hal-01244908**

**<https://hal.inria.fr/hal-01244908>**

Submitted on 16 Dec 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Popularity Prediction in Content Delivery Networks

Nesrine Ben Hassine<sup>\*†</sup>, Dana Marinca<sup>†\*</sup>, Pascale Minet<sup>\*</sup>, Dominique Barth<sup>†</sup>

<sup>\*</sup>Inria, Rocquencourt, 78153 Le Chesnay Cedex, France,

Email: nesrine.ben-hassine@inria.fr, pascale.minet@inria.fr

<sup>†</sup>PRiSM, University of Versailles, Versailles, France,

Email: dana.marinca@prism.uvsq.fr, dominique.barth@prism.uvsq.fr

**Abstract**—Content delivery networks (CDNs) face a large and continuously increasing number of users solicitations for video contents. In this paper, we focus on the prediction of popularity evolution of video contents. Based on the observation of past solicitations of individual video contents, individual future solicitations are predicted. We compare different prediction strategies: SES, DES and Basic. The best tuning of each strategy is determined, depending on the considered phase of the solicitation curve. Since DES and Basic experts outperform the SES expert, our method combines DES and Basic experts to predict the number of solicitations within a phase and automatically detect the phase changes, respectively. This self-learning and prediction method can be applied to optimize resources allocation in service oriented architectures and self-adaptive networks, more precisely for the CDN cache nodes management.

## I. INTRODUCTION

With the emergence of media platforms (e.g. YouTube, Daily Motion), multimedia contents have experienced a significant growth of users solicitations. Consequently, the nature of most of the traffic circulating on internet nowadays is a User-Generated Content (UGC) traffic and Video-on-Demand (VoD) content. YouTube is known to be the most popular UGC. YouTube claims to serve 460 million visits per day, whereas Netflix, a VoD system, serves 22 million visits per day. Content Delivery Networks (CDNs) cope with such enormous and growing number of content solicitations by means of caching and duplicating contents near the end users. Indeed, popularity is defined as the relationship between a content and the users who visualize it. The dynamics of content solicitation can be better understood by studying content popularity. To transform observed information in valuable knowledge, one needs to make a step forward and go beyond observing and explaining the past to build models that will predict future behavior.

When a content becomes popular, it will generate an increase in traffic, which in turn can lead to an important delay, or even a saturation of the network. An accurate prediction of a content popularity can help to manage the content storage in a CDN in order to maximize the number of hits and minimize the number of unsatisfied users because of large delays. An inaccurate prediction, on the other hand, can lead to misallocated resources [1]. The collective user behavior in terms of content can also be used to build dynamic data replication strategies in order to lighten the traffic directed to the same cache and/or to be closer to different groups of users who request the same

content but are geographically distant from each other [2]. Popularity is used to decide on the one hand which content should be cached and if the cache size does not allow storing the new content, which content should be ejected [3]. Moreover, studying content popularity can be useful for transport optimization in some applications thanks to hybrid multicast-unicast transport using multicast for the most popular contents and unicast for the less popular programs [4].

Furthermore, popularity can be used to support enhanced applications that are a source of revenues for the CDN providers. Indeed, popularity is an important measure for online marketing. If the popularity can be predicted accurately, better marketing strategies can be adopted or different advertising rates can be charged. Since popularity indicator is tied directly to ad revenue, revenue might be accurately estimated over time if all parties know how many views a content is likely to attract [5], [6]. Thus, predicting content popularity of various contents is of great importance to manage various services and more particularly new services with added values.

In this paper, we present a self-learning and prediction method able to accurately predict the future number of solicitations of individual video contents. For the simulations, we have used real traces on video contents stored on the YouTube website, concerning the daily number of solicitations since content creation until now. The method is based on the definition of several experts, each one predicting the future solicitation according to its own strategy. Different prediction strategies are compared: Single Exponential Smoothing (SES) [7], Double Exponential Smoothing (DES) [7] and a basic one (Basic) [8] issued from control theory. For each prediction strategy, the best tuning of parameters is determined. We show the interest of distinguishing phases in the popularity evolution of a video content to improve the accuracy of expert prediction. Our method combines the best DES expert per phase and the Basic expert.

## II. RELATED WORK

There have been several research regarding the study of the popularity of multimedia contents. Popularity can be measured in different ways. It is mostly expressed by the number of views over a period of time [6], [9], [10], [11] and this is the definition adopted in this paper. Yet, in [12] feedbacks were the metric used to quantitatively evaluate web content popularity. For instance, comments outline the opinion of users with respect to the content, sharing proves

its fame. However, comment analysis requires a semantic study whereas sharing evaluation must be done considering all social networks. Another evaluation of popularity is based on ratings that describe the quality of the content and can provide an estimation of the popularity thanks to an observed correlation between user rankings and content views [13].

Understanding the popularity of contents and the relationship between requests and contents age is fundamental in the design of efficient communication systems. Indeed, a content cannot be requested before it is uploaded and it is expected that even popular contents will not be requested forever. Once published at a given time, a content popularity vanishes over time. Besides, contents exhibit different profiles of popularity [14]: some contents are able to accumulate hundreds of millions of views, whereas the vast majority can only attract a few. Some contents have the same order of popularity magnitude where the average number of solicitations per day is low but covers a lot of days. Some can be highly but shortly popular; others may have been uploaded and/or removed during the same day.

Since the number of views over a period of time is the simplest way to evaluate popularity, we adopt this definition. The popularity prediction is to estimate the number of views associated with a given content at some time in the future, based on its past evolution [10].

Since the understanding of user intents and preferences is key to personalized services, predicting content popularity based on users behavior has been the subject of several researches. Popularity was proved to follow Zipf distribution when analyzing the viewing rate of contents [14], [15], [16]. In [17], Yu et al enumerate previous works investigating the application of Zipf's law to web access and prove that the distribution of users request is indeed described by Zipf's law [16]. However, Cha et al had found that the number of views is better described by power-law with exponential cut-off. In fact, several studies use real traces to determine the law of the distribution of content popularity. All agree on distribution following a Zipf law for the most popular contents. For less popular contents, popularity is remarkably lower than the predicted popularity using the Zipf distribution [18]. To have more accurate results, we use real traces for this study.

This process of prediction encompasses two approaches: the first is to predict the exact value of the popularity. The second approach assumes that we have an idea on the popularity bounds and predicts a class membership where a class includes all contents whose popularity belongs to a given interval.

Analyzing the popularity characteristics of UGC system proves that a strong correlation exists between the popularity of each content at early and later stages during its lifetime. Hence linear regression [18] is used as a prediction method. Szabo and Huberman [6] present a linear regression model by using logarithmic transformation based on the number of views. The authors of [19] choose to apply regression to predict popularity based on different feature spaces.

In [10], Ahmed et al suggest a predictor based on modeling the transitions that content makes between being a member of different classes of behavior over time. Lee et al. [18] [19] evaluate the probability that a given content will receive more than a given number of hits. Lakkaraju and Ajmera [20] use Support Vector Machines to predict whether a given content will fall in a group that attracts less than 10%, 25%, 50%, 75%, 100% of the attention.

In addition, classification based on the evolution of content popularity shows that there is a small number of temporal patterns [10]. Gursun et al [11] find that there are two types of videos: frequently-accessed and rarely-accessed contents. Frequently-accessed contents are consistently popular over long time periods (e.g: a year). However, rarely-accessed contents show rapid changes in popularity; they are viewed during a small portion of the year.

In [20], Crane and Sornette analyzed the popularity evolution patterns of YouTube contents and identified two main classes: contents experiencing no marked peak in popularity that attract little attention or experience just some popularity fluctuation. This class is called Memoryless. The other class includes contents experiencing bursts in popularity and contains three sub-classes. Viral contents popularity increases slowly, up to a peak, decreasing also slowly afterwards. Quality contents experience a very sudden peak in popularity and a slow decay afterwards. Junk contents know a burst of popularity then their popularity drops quickly afterwards.

### III. PROBLEMATIC AND PROPOSED METHODOLOGY

The problem we are interested in is to predict the popularity of video contents retrieved from YouTube. Given a collection of video contents, we study how the popularity of each content evolves since it was uploaded to the YouTube platform till the current time. Based on this evolution description, we are seeking to estimate the popularity of each content in the future. We take into account information available in the YouTube website, (regardless if the content was created or shared somewhere in the network) to observe the number of solicitations of each considered video content during an observation window and predict the number of solicitations.

In this study we use real traces of daily solicitation evolution extracted from YouTube. Let  $C$  be the collection of  $N$  video contents, and  $c$  any video content from this collection. Let  $E_i$ ,  $i \in [1, M]$  be the set of experts predicting the future solicitations of each content  $c$ . An expert  $E_i$  is an entity able to analyze an observation window  $ow_i$  over the past (several consecutive past days) and to predict the solicitation evolution over a prediction window  $pw_i$ , indeed the next day. We define  $PS_i$  the prediction strategy of the expert  $E_i$ . Let  $y_t$  be the real number of solicitations at time  $t$ .

### IV. PREDICTION STRATEGIES

We consider different prediction strategies to predict the popularity of contents.

### A. Single Exponential Smoothing (SES) prediction

The value  $p_{i,t}$  predicted by any Single Exponential Smoothing (SES) [7] expert  $E_i$  at time  $t$  is given by:

$$p_{i,t-ow_i} = y_{t-ow_i}$$

For each  $k \in (t - ow_i, t]$

$$p_{i,k} = \alpha * y_{k-1} + (1 - \alpha) * p_{i,k-1},$$

where  $\alpha$  is a smoothing factor  $\in (0, 1)$ .

### B. Double Exponential Smoothing (DES) prediction

Let  $S'_{i,t}$  denote the single smoothed value obtained by applying SES to  $y_t$ .

$$S'_{i,t} = \alpha y_t + (1 - \alpha) S'_{i,t-1}.$$

Let  $S''_{i,t}$  denote the double smoothed value obtained by applying SES to  $S'_{i,t}$  using the same  $\alpha$ .

$$S''_{i,t} = \alpha S'_{i,t} + (1 - \alpha) S''_{i,t-1},$$

with  $0 < \alpha < 1$ ,  $S'_{i,0} = S''_{i,0} = y_0$ .

The value predicted by a Double Exponential Smoothing (DES) expert [7] at time  $t$  is given by:

$$p_{i,t+1} = L_{i,t} + T_{i,t},$$

where  $L_{i,t} = 2S'_{i,t} - S''_{i,t}$  denotes the estimated level

and  $T_{i,t} = \frac{\alpha}{1-\alpha}(S'_{i,t} - S''_{i,t})$  the estimated trend.

### C. Basic expert

We also consider a basic expert  $E_i$  that at time  $t > 1$  predicts the value  $p_{i,t}$  which is equal to the previous value observed  $y_{t-1}$  plus an increment equal to the difference between the two last observed values. Hence, we get:

$$p_{i,t} = y_{t-1} + (y_{t-1} - y_{t-2}) = 2y_{t-1} - y_{t-2}.$$

This expert has similarities with a PID controller in control theory [8]. Notice that this expert is very simple and requires no parameter tuning.

### D. Error of prediction

We assess the performance of these prediction strategies using the absolute value error function. The relevance of a prediction is measured in terms of instantaneous loss. In this paper, the instantaneous loss is evaluated by the absolute value of the difference between the predicted value and the real one. Hence, at trial  $t > 0$  any expert  $E_i$  predicting  $p_{i,t}$  has an instantaneous loss given by  $|p_{i,t} - y_t|$ .

### E. Improvements of expert predictions

In this section we show how experts improve their prediction when they have additional knowledge about bounds on the values provided by the environment.

*Property 1:* If  $\forall t, Min \leq y_t \leq Max$ , with  $Min$  and  $Max \in Z$ , then any expert  $E_i$  predicts a value  $p_{i,t} \in [Min, Max]$  and has an instantaneous loss  $|p_{i,t} - y_t| \leq |Max - Min|$ .

*Proof:* Since  $\forall t$ , we have  $Min \leq y_t \leq Max$ , any expert  $E_i$  will predict a value  $p_{i,t}$  that meets  $Min \leq p_{i,t} \leq Max$ . More precisely, if the value computed by the expert is smaller than  $Min$ , the expert predicts  $Min$ . Similarly, if the value computed by the expert is larger than  $Max$ , the expert predicts  $Max$ . Considering these two inequations, we get  $Min - Max \leq p_{i,t} - y_t \leq Max - Min$ . Since both  $Max$  and  $Min$

are integers  $\in Z$ , we finally get  $|p_{i,t} - y_t| \leq |Max - Min|$ . That is the instantaneous loss of expert  $E_i$  is upper bounded by  $|Max - Min|$ . ■

*Property 2:* If  $\forall t, Min \leq y_t \leq Max$ , with  $Min$  and  $Max \in Z$ , then the expert  $E_i$  predicting the value  $p_{i,t} = (Max - Min)/2$  has an instantaneous loss  $|p_{i,t} - y_t| \leq |Max - Min|/2$ .

*Proof:* We notice that  $\forall t, (Max - Min)/2$  is the median value of  $[Min, Max]$ . Hence, whatever the value we have of  $t \in [Min, Max]$ , an expert  $E_i$  predicting a value  $p_{i,t} = (Max - Min)/2$  has an instantaneous loss  $|p_{i,t} - y_t| \leq |Max - Min|/2$ . ■

*Property 3:* If  $\forall t > 0, y_{t-1} \leq y_t \leq y_{t-1} + B$ , with  $B \in N$ , then any expert  $E_i$  predicts a value  $p_{i,t} \leq y_{t-1} + B$  and has an instantaneous loss  $|p_{i,t} - y_t| \leq B$ .

*Proof:* Since  $\forall t > 0$ , we have  $y_{t-1} \leq y_t \leq y_{t-1} + B$ , any expert  $E_i$  will predict a value  $p_{i,t}$  that meets  $y_{t-1} \leq p_{i,t} \leq y_{t-1} + B$ . Hence, we get by considering these two inequations,  $-B \leq p_{i,t} - y_t \leq B$ . Since  $B$  is a natural integer, we finally get  $|p_{i,t} - y_t| \leq B$ . That is the instantaneous loss of expert  $E_i$  is upper bounded by  $B$ . ■

## V. PERFORMANCE EVALUATION

### A. Real traces

For the simulations we use datasheets representing the solicitations of YouTube contents. The trace file associated with each content represents the number of solicitations per day, from the online storage of the video content considered up to the current day. The analyzed traces concern contents belonging to different categories: social, music, sport, news, films, cartoons. The traces present video solicitations during durations varying from 7 months to 5 years. For instance, Figure 1a and Figure 1b depict the total number of solicitations for contents belonging to the category *Music* during 1878 days, and *Society* during 534 days, respectively.

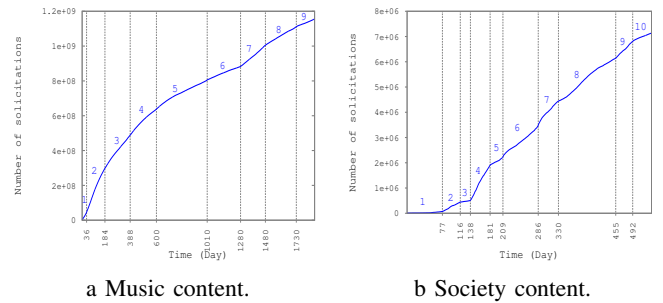


Fig. 1: Cumulated number of solicitations over time.

### B. Use of phases in expert prediction

We analyze different contents with different categories and we notice that contents belonging to the same category have the same appearance of the solicitation curve. Throughout the monitoring duration of each content we observe the existence of periods of consecutive days presenting similar daily increments of the number of solicitations. We denote such period a phase (see for example

Figure 1a, where we can distinguish 9 phases in the popularity evolution of the *Music* video content). The increments of solicitations are very different from one phase to another, determining different slopes on the solicitation curve. We show how to take advantage of these phases to improve the accuracy of expert predictions.

We first define the concepts of best expert on a phase and best global expert. We define  $E_G$  as the best global expert, that is the expert minimizing the sum of its cumulated losses over all phases. For any phase  $\phi$ , we define  $E_{B(\phi)}$  as the best expert on phase  $\phi$ , that is the expert minimizing its cumulated loss over phase  $\phi$ .

*Property 4:* The sum over all phases of the cumulated loss of the best global expert is higher than or equal to the sum over all phases of the cumulated loss of the best expert on each phase.

$\sum_{\phi} L_G^{\phi} \geq \sum_{\phi} L_{B(\phi)}^{\phi}$ , where  $L_G^{\phi}$  denotes the cumulated loss over phase  $\phi$  of  $E_G$  the best global expert and  $L_{B(\phi)}^{\phi}$  is the cumulated loss over phase  $\phi$  of the best expert  $B(\phi)$  on this phase.

*Proof:* For any phase  $\phi$ , we have by definition of the best expert on phase  $\phi$ ,  $L_i^{\phi} \geq L_{B(\phi)}^{\phi}$ . Applying this inequation to the best global expert  $E_G$ , we get  $L_G^{\phi} \geq L_{B(\phi)}^{\phi}$ . Summing over all phases, we get the property. ■

### C. Tuning of experts parameters

Our first goal is to select the best parameters for each given phase for different content solicitation profiles, and for each expert type. For instance, for SES and DES experts, the parameters to determine are the size of the observation window and the value of alpha. For these values of parameters, the experienced loss should be minimal. Notice that for the basic expert, there is no parameter to tune. SES and DES experts use a smoothing parameter  $\alpha$  which varies over time depending on the correlation with the old environment outputs. So, to optimize them, we choose to minimize the cumulated loss between our prediction and the environment output. The optimum value of  $\alpha$  at time  $t$  is used in the prediction of  $t+1$ .

#### 1) SES expert:

Figures 2, 3 and 4 depict the cumulated loss for SES experts ignoring the notion of phase, in phase 5 and in phase 6 of the content depicted in Figure 1b, respectively. A small size of the observation window (i.e.  $ow = 2$ ) is the best, whatever the value of  $\alpha$ . A value of  $\alpha = 0.2$  is recommended when the number of solicitations increases very weakly. Otherwise, a value of  $\alpha$  close to 1 (i.e.  $\alpha = 0.9999$ ) provides the best results, minimizing the cumulated loss.

#### 2) DES expert:

Figures 5, 6 and 7 depict the cumulated loss for DES experts using different sizes of observation window  $ow_i \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ , expressed in days, where  $\alpha \in [0.7, 0.9999]$ , without considering phase and then in phases 5 and 6 of the content depicted in Figure 1b, respectively. Simulations show that values of  $ow_i > 10$  and also values of  $\alpha \in ]0, 0.7]$  are not significant, the cumulated loss on the analyzed phases being too large.

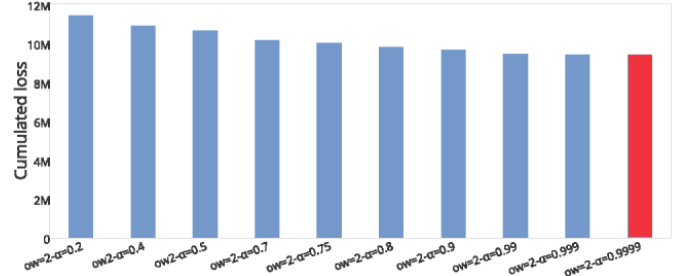


Fig. 2: Cumulated loss for SES experts over all the period

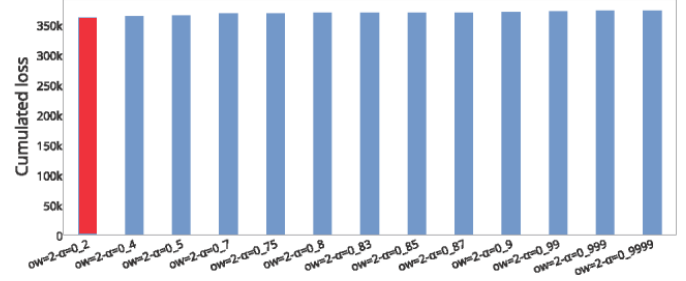


Fig. 3: Cumulated loss for SES experts in phase 5

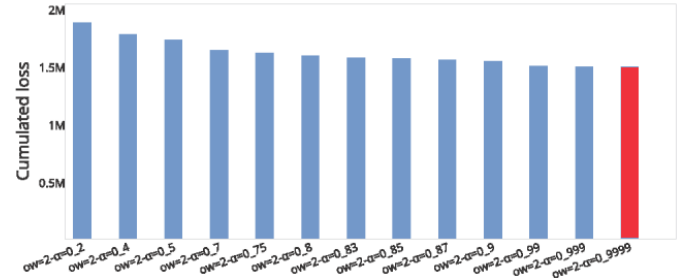


Fig. 4: Cumulated loss for SES experts in phase 6

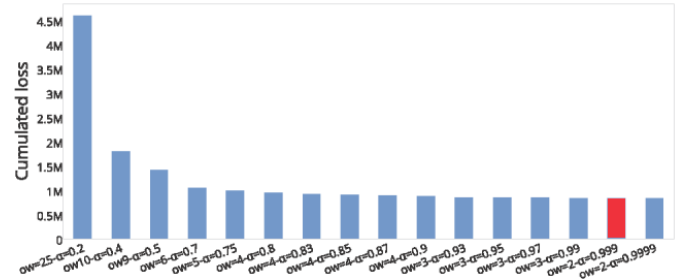


Fig. 5: Cumulated loss for DES experts over all the period

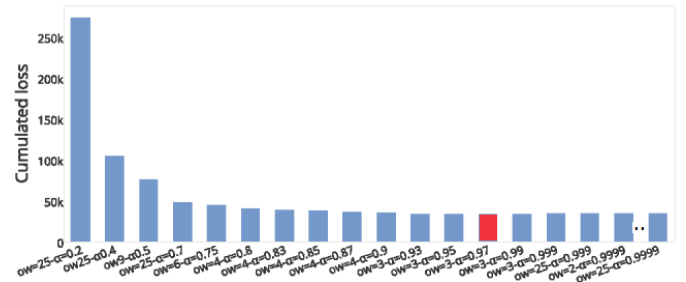


Fig. 6: Cumulated loss for DES experts in phase 5

### D. Choice of the best expert per phase

We recall that we define the best expert as the expert minimizing its cumulated loss. In this section, we show that

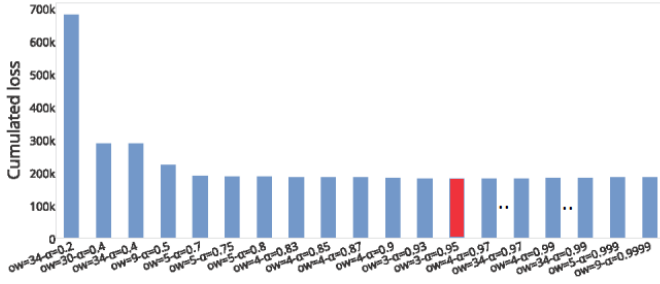


Fig. 7: Cumulated loss for DES experts in phase 6

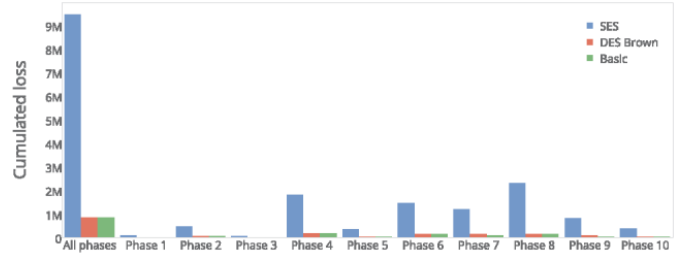


Fig. 9: Cumulated loss for SES, DES and Basic experts.

the best expert over all the period predicts worse than the best expert per phase.

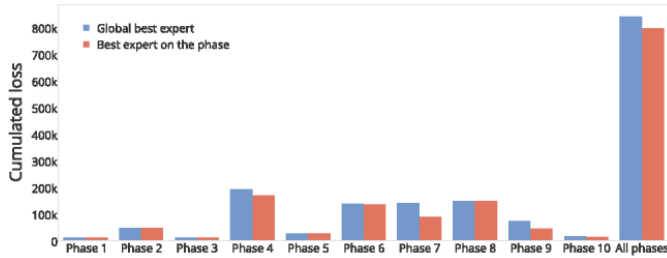


Fig. 8: DES: cumulated loss on total period vs sum of cumulated loss per phase

As expected, Figure 8 shows that real traces confirm Property 4.

#### E. Comparison between SES, DES and Basic experts

For different contents we compare the performances of the best SES experts, the best DES experts and the basic expert in different phases. The best expert is this minimizing the cumulated loss. Figure 9 clearly shows that the best DES expert outperforms the best SES expert. The cumulated loss of the DES expert is about ten times smaller. Such a result is expected, since DES experts have been designed to take into account the trend in the values observed. Surprisingly, the Basic expert and the best DES expert per phase provide close performances. Simulation results show that Basic is better than DES on 7 phases over 10 phases, but globally the cumulated loss over all phases is smaller for DES. The explanation comes from the sensitivity of the DES expert to small deviations, unlike the Basic expert. More precisely, during phase 7 small deviations on the number of solicitations contribute to increase the instantaneous loss of the DES expert. Hence, the DES expert exhibits a larger cumulated loss. Prediction of the best DES expert per phase and the Basic expert are very close as depicted in Figure 10.

#### F. Automatic detection of phase changes

Since the instantaneous loss of the Basic expert is usually low, except when the phase changes Figure 11, we decide to use this Basic expert to detect phase changes. Furthermore, this expert adapts faster to phase changes.

We distinguish four types of phase changes, depending on the slope variation of the number of solicitations:

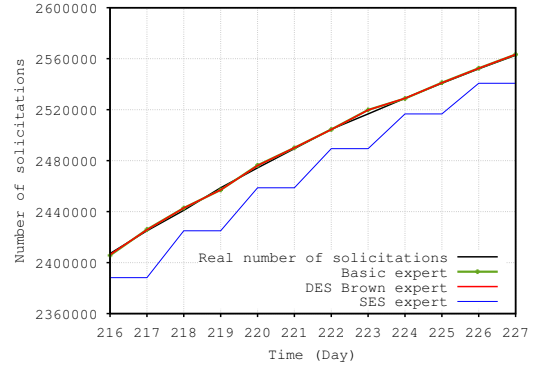


Fig. 10: Prediction of the best SES, DES and Basic experts.

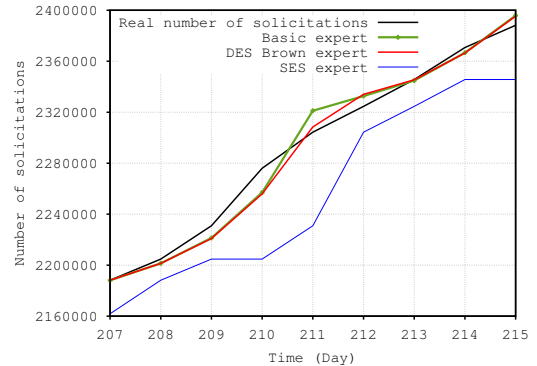


Fig. 11: Reactivity of Basic expert at phase changes.

- Detection of a strong increase in the number of solicitations: see for instance time  $t = 138$  on Figure 12.a; if  $l_{i,t} \geq 2l_{i,t-1}$  and  $l_{i,t-1} < 2l_{i,t-2}$  then there is a phase change at time  $t$  with a strong increase.
- Detection of a soft increase in the number of solicitations: see for instance time  $t = 429$  on Figure 12.b; if  $l_{i,t} < 2l_{i,t-1}$  and  $l_{i,t-1} < 2l_{i,t-2}$  and  $l_{i,t} \geq 1.5l_{i,t-2}$  and  $l_{i,t-1} < 1.5l_{i,t-3}$  then there is a phase change at time  $t$  with a soft increase.
- Detection of an increase in the number of solicitations smaller than expected: see for instance time  $t = 184$  on Figure 12.c; if  $l_{i,t} \leq 0.5l_{i,t-1}$  and  $l_{i,t-1} > 0.5l_{i,t-2}$  then there is a phase change at time  $t$  with an increase smaller than expected.
- Detection of an increase in the number of solicitations slightly smaller than expected: see for instance time  $t =$



496 on Figure 12.d; if  $l_{i,t} \leq 0.7l_{i,t-1}$  and  $l_{i,t} \leq 0.7l_{i,t-2}$  and  $l_{i,t} > 0.6l_{i,t-2}$  then there is a phase change at time  $t$  with an increase slightly smaller than expected.

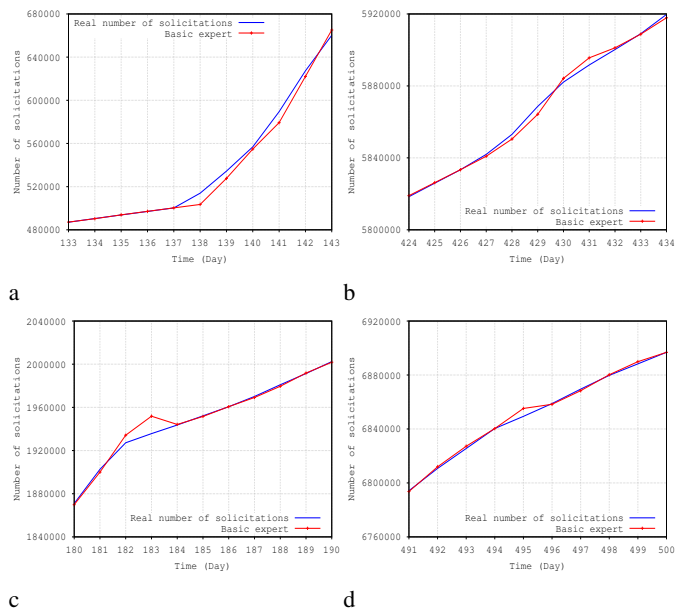


Fig. 12: Phase changes automatically detected.

## VI. CONCLUSION

We presented a self-learning and prediction method able to predict the number of solicitation individually for each video content in a CDN. The model is based on several experts, each expert having its own prediction strategy (e.g. SES, DES and Basic). The best parameters tuning is determined for each prediction strategy and its accuracy is evaluated. The benefit of using phases during popularity evolution was proved, a phase being characterised by a similar slope of the cumulative solicitations evolution curve. Indeed, during one phase one expert can be identified as being more performant to predict than the others. The use of this machine learning method allows the CDN to self-adapt to users solicitations by caching the most popular contents near the end users. More generally, such method can be applied to decide which content should be replicated to improve the performance of video applications and maximize the satisfaction degree of users.

## ACKNOWLEDGMENT

This work has been partially funded by the ANR project Netlearn ANR-13-INFR-004.

## REFERENCES

- [1] H. Abrahamsson and M. Nordmark, "Program popularity and viewer behaviour in a large tv-on-demand system," in *Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM, 2012, pp. 199–210.
- [2] D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, and K. K. Ramakrishnan, "Optimal content placement for a large-scale vod system," in *Proceedings of the 6th International Conference*. ACM, 2010, p. 4.

- [3] J. Famaey, F. Iterbeke, T. Wauters, and F. De Turck, "Towards a predictive cache replacement strategy for multimedia content," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 219–227, 2013.
- [4] D.-E. Meddour, A. Abdallah, T. Ahmed, and R. Boutaba, "A cross layer architecture for multicast and unicast video transmission in mobile broadband networks," *Journal of Network and Computer Applications*, vol. 35, no. 5, pp. 1377–1391, 2012.
- [5] A. Kupavskii, A. Umnov, G. Gusev, and P. Serdyukov, "{Predicting the Audience Size of a Tweet}," 2013.
- [6] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010.
- [7] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006.
- [8] K. H. Ang, G. Chong, and Y. Li, "Pid control system analysis, design, and technology," *Control Systems Technology, IEEE Transactions on*, vol. 13, no. 4, pp. 559–576, 2005.
- [9] H. Pinto, J. M. Almeida, and M. A. Gonçalves, "Using early view patterns to predict the popularity of youtube videos," in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 365–374.
- [10] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, "A peek into the future: Predicting the evolution of popularity in user generated content," in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 607–616.
- [11] G. Gursun, M. Crovella, and I. Matta, "Describing and forecasting video access patterns," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011, pp. 16–20.
- [12] P. Schultes, V. Dörner, and F. Lehner, "Leave a comment! an in-depth analysis of user comments on youtube." in *Wirtschaftsinformatik*, 2013, p. 42.
- [13] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 15–28.
- [14] L. Carlinet, T. Huynh, B. Kauffmann, F. Mathieu, L. Noirie, and S. Tixeuil, "Four months in daily motion: Dissecting user video requests," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2012 8th International*. IEEE, 2012, pp. 613–618.
- [15] S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing web-based video sharing workloads," *ACM Transactions on the Web (TWEB)*, vol. 5, no. 2, p. 8, 2011.
- [16] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1. IEEE, 1999, pp. 126–134.
- [17] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng, "Understanding user behavior in large-scale video-on-demand systems," in *ACM SIGOPS Operating Systems Review*, vol. 40, no. 4. ACM, 2006, pp. 333–344.
- [18] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 5, pp. 1357–1370, 2009.
- [19] M. Tsagkias, W. Weerkamp, and M. De Rijke, "News comments: Exploring, modeling, and online prediction," in *Advances in Information Retrieval*. Springer, 2010, pp. 191–203.
- [20] R. Crane and D. Sornette, "Robust dynamic classes revealed by measuring the response function of a social system," *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15 649–15 653, 2008.