

Expert-based on-line learning and 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. Expert-based on-line learning and prediction in Content Delivery Networks. IWCMC 2016 - The 12th International Wireless Communications & Mobile Computing Conference, Sep 2016, Paphos, Cyprus. pp.182 - 187, 2016, <10.1109/IWCMC.2016.7577054>. <hal-01411119>

HAL Id: hal-01411119

<https://hal.inria.fr/hal-01411119>

Submitted on 7 Dec 2016

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.

Expert-based On-line Learning and Prediction in Content Delivery Networks

Nesrine Ben Hassine^{*†}, Dana Marinca^{†*}, Pascale Minet^{*}, Dominique Barth[†]

^{*}Inria Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12,

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

[†]DAVID, University of Versailles, Versailles, France,

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

Abstract—Machine learning techniques can be used to improve the quality of experience for the end users of Content Delivery Networks (CDNs). In a CDN, the most popular video contents are cached near the end-users in order to minimize the contents delivery latency. The idea developed hereafter consists in using prediction techniques to evaluate the future popularity of video contents in order to decide which should be cached. We consider various prediction methods, called experts, coming from different fields (e.g. statistics, control theory). We assess these experts according to three criteria: cumulated loss, maximum instantaneous loss and best ranking. We also show the importance of a decision maker, called forecaster, that predicts the popularity based on the predictions of selection of several experts. The forecaster based on the best K experts outperforms in terms of cumulated loss the individual experts' predictions and those of the forecaster based on only one expert, even if this expert varies over time.

I. INTRODUCTION

Content Delivery Networks (CDNs) have undergone an ever-increasing success, leading to an important increase in multimedia traffic that could result in unacceptable delays for the mobile or fixed end users. The user mobility strongly impacts the solicited caches: their content should be updated dynamically to reflect users' solicitations. To mitigate this effect, caching techniques are used in nodes near the end users. Hence, to decide which video contents should be cached is a challenging task that requires an in-depth analysis of the popularity of video contents over time.

In this paper, we focus on the use of machine learning techniques in CDNs. The remainder of this paper is organized as follows. In this section, we present the context of this study and some related work. In Section II, we define the theoretical framework including the problem statement. In the context of our paper, an expert refers to a software component that implements a prediction computation method. We define all the experts used (e.g. DES, Basic, polynomial or Savitzky-Golay regression, etc) coming from various fields (e.g. statistics, control theory) and explain how they compute their predictions. These predictions are used by a decision maker, called forecaster, to build its own prediction. We define two types of forecasters: Best Expert (BE) and K Best Experts (KBE). In Section III, we report our simulation results obtained from real traces of YouTube. We consider various video contents and evaluate the prediction accuracy of the forecasters and experts considered. We also compare the

experts using multiple criteria (e.g. cumulated loss, maximum instantaneous loss, best ranking). Finally, Section IV gives the conclusion.

A. Context

Our goal is to predict the popularity of video content in a CDN. As in many other studies [1], [2], [3], [4], we evaluate the popularity of a video content by its number of solicitations over a period of time. Hence, we predict the video content popularity by approximating the function representing the evolution of the number of solicitations over time.

For that purpose, we will evaluate different prediction methods applied on the same data extracted from real traces of YouTube CDN. We adopt the following approach where the popularity prediction is made in two steps. In the first step, each expert computes its prediction using its own prediction strategy (e.g. average, regression, exponential smoothing). In a second step, the forecaster uses the predictions given by the experts to build its own prediction. Here again, different strategies are possible. For instance, the forecaster may select the best expert or a group of best experts. Having selected this group of experts, the forecaster may use averaging or voting techniques. The best experts are those that provide the most accurate predictions. The accuracy of predictions is evaluated by the cumulated gap between the prediction and the real value up to current time. We will first justify the choice of this two-step approach based on experts and forecasters.

B. Related work

Let us consider a dynamic system whose parameters often change over time. We consider an online prediction problem where system is observed at different times and there is a challenge of maximizing prediction accuracy. This problem occurs in a multitude of machine learning applications.

Gu and Wang [5] use on-line prediction in a data stream processing application. They combine Markov models and Bayesian classification methods to predict the bottleneck that could appear, and its probable cause, while clustering data streams.

Many studies have used on-line prediction to deal with the limited battery lifetime problem. In [6], the authors employ a statistical approach to predict online the battery lifetime for embedded and mobile devices. They incorporate both recently

observed power dissipation values with pre-computed off-line reference signature. The battery lifetime prediction [8] is based on the usage patterns of individual users. Furthermore, in [7] an online approach is used during runtime to identify whether a failure will occur in the near future based on the monitored current system state.

In a previous study [9], we applied on-line prediction to predict the quality of links in a wireless sensor network. We combined a range of learning techniques and we identified the best parameter tuning of these techniques which provides a better approximation of the real values of the quality of the links.

II. THEORETICAL FRAMEWORK

A. Problem statement

From YouTube platform, we extract the real traces describing the daily solicitation evolution over time of randomly chosen video contents. We consider that the number of solicitations defines the popularity of a video content. We then focus on any video content among the selected ones. Based on its daily solicitations in the time interval $[0, t]$, the goal is to predict its popularity during the day $t + 1$.

The prediction model is based on a set of *experts*, each expert being a logical entity which computes and predicts the future number of solicitations for each video content using its own computation method (Section II-B). A special expert, called *forecaster*, predicts the number of solicitations based on a selection of best expert predictions (Section II-D). We will show that forecaster predictions are more accurate than individual experts predictions (Section III-B).

Let us introduce some notations:

- $\{E_i, i \geq 1\}$ the set of experts.
 - ow_i the observation window of expert E_i .
 - $\{F_j, j \geq 1\}$ the set of forecasters.
- For any given video content:
- y_t the number of solicitations at time t .
 - $p_{i,t}$ the prediction of the expert E_i for time t .
 - $\hat{p}_{j,t}$ the prediction of the forecaster F_j for time t .

B. Experts definition

The prediction methods used by the proposed experts come from statistics or control theory. These prediction methods are based on exponential smoothing [10], average of corrections, regression [11] or filtering [12].

1) Double Exponential Smoothing (DES) Expert:

The DES expert [10], uses exponential smoothing to compute its prediction. A DES expert E_i applies the exponential smoothing twice: first on y_t to compute $S'_{i,t}$ and second on $S'_{i,t}$ to compute $S''_{i,t}$. Both smoothings use the same smoothing factor α with $0 < \alpha < 1$,

$$S'_{i,0} = S''_{i,0} = y_0.$$

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

At time t , the DES expert E_i predicts the value

$$p_{i,t+1} = L_{i,t} + T_{i,t} \quad (1)$$

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.

2) Constant Basic (CB) expert:

The Constant Basic (CB) expert has been defined in [13]. The expert E_i predicts at time t the value for time $t + 1$ given by:

$$p_{i,t+1} = y_t + c_i * (y_t - y_{t-1}) \quad (2)$$

where c_i denotes the correction coefficient. In this prediction method, c_i is fixed.

3) Dynamic Basic (DB) expert:

The Dynamic Basic (DB) expert E_i uses the same prediction method as the previously defined one, except that the correction coefficient is computed dynamically as follows:

$$c_{i,t} = \frac{y_t - y_{t-1}}{y_{t-1} - y_{t-2}}. \quad (3)$$

4) Arithmetical Moving Average (AMA) adjusted expert:

The Arithmetical Moving Average (AMA) adjusted expert, E_i predicts a value that is adjusted using a coefficient equal to the arithmetical average of dynamic correction coefficients over the solicitation values in its observation window ow . $|ow|$ represents the number of values in ow .

$$p_{i,t+1} = y_t + \bar{c}_{i,t} * (y_t - y_{t-1}) \quad (4)$$

with $\bar{c}_{i,t} = \frac{\sum_{k \in ow} c_{i,k}}{|ow|}$, where $c_{i,k}$ is computed according to Equation 3.

5) Geometrical Moving Average (GMA) adjusted expert:

The Geometrical Moving Average (GMA) adjusted expert, E_i predicts a value that is adjusted using a coefficient equal to the geometrical average of dynamic correction coefficients over its observation window ow .

$$p_{i,t+1} = y_t + \bar{c}_{i,t} * (y_t - y_{t-1}) \quad (5)$$

with $\bar{c}_{i,t} = \sqrt[|ow|]{\prod_{k \in ow} c_{i,k}}$, where $c_{i,k}$ is computed according to Equation 3.

6) Polynomial Regression (PR) expert:

The Polynomial Regression (PR) experts smooth the time function to approximate the real profile, by an n -degree polynomial extrapolation. Given n the degree of the polynomial, the coefficients $a_k, k \in [0, n]$ involved in the regression are calculated in such a way that they fit the relationship among data in the observation window ow . At time t , the Polynomial Regression expert E_i predicts the value:

$$p_{i,t+1} = a_n(t+1)^n + a_{n-1}(t+1)^{n-1} + \dots + a_1(t+1) + a_0. \quad (6)$$

7) Savitzky Golay (SG) regression expert:

Savitzky Golay (SG) regression experts smooth the time function of the real profile by a Savitzky Golay filter. The idea of the Savitzky Golay filter is to find the set of coefficients a_0, \dots, a_n which best preserve the shape of the features present in the sampled profile. The approach is to make a polynomial regression of the observations in the window ow around its center k , and then evaluate this polynomial at time $k + |ow/2| + 1$. The coefficients a_l , with $l \in [0..n]$, involved in the regression are computed by a convolution of the input data in ow with the coefficient vector given by Equation 7.

$$a_l = \sum_{j=-|ow|/2}^{|ow|/2} v_j y_j \quad (7)$$

$$p_{i,t+1} = a_0 + a_1(t-k) + \dots + a_n(t-k)^n \quad (8)$$

The observation window size $|ow|$ should be odd and greater than n the degree of the polynomial.

The arithmetical average smoothing corresponds to the case where all coefficients a_l , with $l \in [0, n]$ are equal to $1/|ow|$. To determine the smoothed value for each data point, the local polynomial regression is better than adjacent averaging because it tends to preserve features of the data such as peak height and width, which are usually eliminated by adjacent averaging.

C. Accuracy of an expert

The prediction accuracy of an expert E_i is evaluated by its loss regarding the gap between its prediction and the real solicitation value. We define several types of losses for any expert E_i , computed at each time instant t :

- the instantaneous loss, denoted $l_{i,t}$, meets:

$$l_{i,t} = |y_t - p_{i,t}|;$$

- the cumulated loss, denoted $L_{i,t}$, satisfies:

$$L_{i,t} = \sum_{k=1}^t |y_k - p_{i,k}| = \sum_{k=1}^t l_{i,k};$$

- the maximum instantaneous loss, denoted $lmax_{i,t}$, meets:

$$lmax_{i,t} = \max_{k \in [0,t]} l_{i,k}.$$

D. Forecasters definition

At each time t , a forecaster selects some experts that minimize their cumulated loss and computes a prediction value based on the selected experts predictions. We propose two types of forecaster.

1) Best Expert (BE) forecaster:

At each time t , the BE forecaster F_j selects exactly one expert E_i that minimizes the cumulated loss. The forecaster predicts the value predicted by the selected expert E_i .

2) K Best Experts (KBE) forecaster:

At each time t , a KBE forecaster F_j selects the experts that provide the K smallest cumulated losses. The forecaster predicts the average of the predictions given by those best experts. Notice that since several experts may predict the same value, the forecaster may select more than K best experts.

E. Accuracy of a forecaster

Similarly, the accuracy of a forecaster F_j is evaluated by the gap between its prediction and the real solicitation value at each time instant t . We define several types of losses for any forecaster F_j :

- the instantaneous loss, denoted $\hat{l}_{j,t}$, meets:

$$\hat{l}_{j,t} = |y_t - \hat{p}_{j,t}|;$$

- the cumulated loss, denoted $\hat{L}_{j,t}$, satisfies:

$$\hat{L}_{j,t} = \sum_{k=1}^t |y_k - \hat{p}_{j,t}| = \sum_{k=1}^t \hat{l}_{j,t};$$

- the maximum instantaneous loss, denoted $l\hat{m}ax_{j,t}$, meets:

$$l\hat{m}ax_{j,t} = \max_{k \in [0,t]} \hat{l}_{j,k}.$$

F. Algorithm of a KBE forecaster

The algorithm of the KBE forecaster is given by Algorithm 1. To begin, the forecaster computes its instantaneous loss and its cumulated loss. The forecaster is allowed to see the losses and the predictions of all experts. At each time t , it observes the K^{th} smallest cumulated loss among all expert cumulated losses. It selects the experts having at time t a cumulated loss less than or equal to the observed cumulated loss and constitutes its group of best experts. Subsequently, it computes the average of the predictions of its best experts and predicts this value.

We can notice that the BE forecaster can be considered as a specific case of a KBE forecaster, where exactly one expert is chosen among those providing the minimum cumulated loss.

Algorithm 1 KBE Forecaster's algorithm at time $t > 0$

```

 $\hat{l}_{j,t} = |\hat{p}_{j,t} - y_t|$ 
 $\hat{L}_{j,t} = \hat{L}_{j,t-1} + \hat{l}_{j,t}$ 
Selected_Experts =  $\emptyset$ 
Sort the experts by increasing  $L_{i,t}$ 
Last_cumul = the  $K^{th}$  value for the cumulated loss
 $\hat{p}_{j,t} = 0$ 
for  $i \in [1, n]$  do
  if  $L_{i,t} \leq Last\_cumul$  then
    Selected_Experts = Selected_Experts  $\cup \{E_i\}$ 
     $\hat{p}_{j,t} = \hat{p}_{j,t} + p_{i,t}$ 
  end if
end for
 $\hat{p}_{j,t} = \hat{p}_{j,t} / size(Selected\_Experts)$ 

```

III. SIMULATION RESULTS

A. Real traces

In this study, we extracted from YouTube platform, real traces describing the daily solicitation evolution over time of randomly chosen video contents. The daily solicitations concern the period from the instant of the online availability up to the current time. We then focus on any video content among the selected ones. Knowing its daily number of solicitations in the time interval $[0, t]$, the goal is to predict how many times this content will be solicited during the day $t + 1$.

Figure 1 depicts the evolution of solicitations day by day for different video content profiles used in our tests. We observe that, during the lifetime of these contents, they have totally different profiles of solicitations: order of magnitude and evolution curve of solicitations.

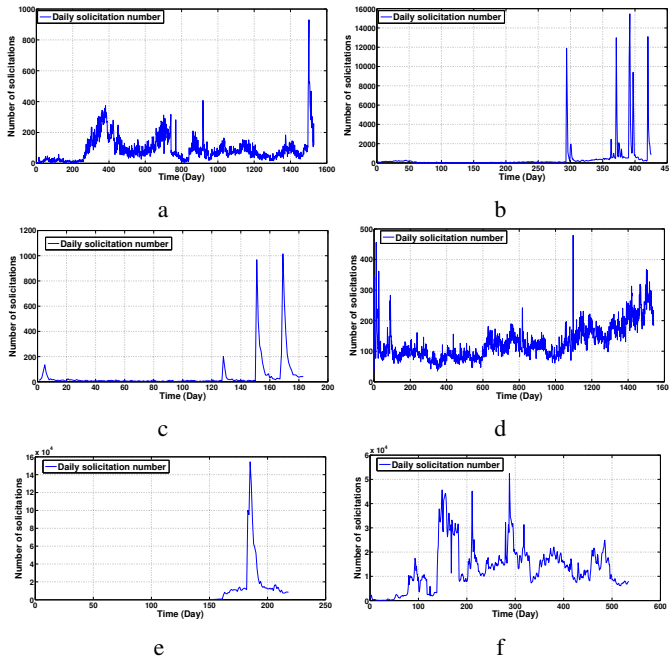


Fig. 1: The evolution of solicitations of video contents.

B. Interest of the KBE forecaster

In the first series of experiments, we compare the BE and KBE forecasters in terms of cumulated loss. Simulation results show that there exists an integer $K > 1$ such that a forecaster's prediction based on K advices is better than the prediction based on only one advice. As we can see in Figures 2a to 2d, the best results in terms of cumulated loss are not obtained by the BE forecaster in any of these video contents. Furthermore, the BE forecaster provides the worst results for video contents 1a, 1e and 1f. The best results are obtained by a KBE forecaster, which uses the predictions of all the experts that have the K smallest cumulated losses.

As depicted in Figures 2a to 2d, the best value of K varies with the video contents. It is $K = 2$ for video contents 1b and 1e, whereas it is $K = 3$ for video contents 1a and 1f. This can be explained by the fact that if B experts among the

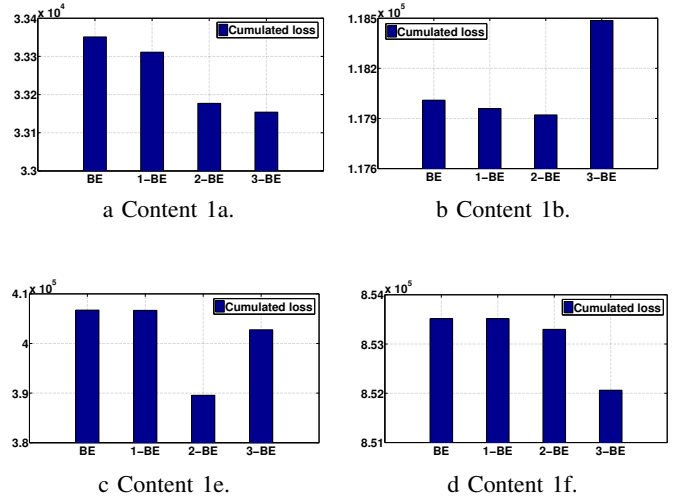


Fig. 2: The cumulated loss of BE forecaster vs the cumulated loss of KBE forecaster.

N experts made bad predictions, the KBE forecaster avoids selecting them as much as possible. Besides, the greater the variation in the number of solicitations for a given content, the greater the interest in increasing the value of K in order not to deviate too much from the real value. However, when the number of solicitations is relatively stable, the BE forecaster, based only on the best expert, predicts accurately and therefore involving other experts in the prediction process could move the prediction away from the real values.

Now we focus on the predictions of different forecasters and compare them with the real value. We distinguish two cases:

- When the demand evolution is quasi-regular, all forecasters give predictions close to real values, as depicted in Figure 3.
- When there is a sudden change in the demand evolution, some forecasters react faster than others, as illustrated in Figure 4. The BE forecaster is the slowest to react.

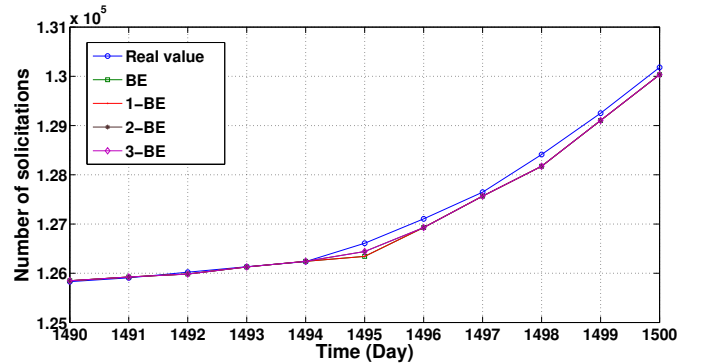


Fig. 3: The prediction of forecasters in stable periods.

C. Accuracy of expert predictions

In this section, we are interested in the accuracy of the prediction of the experts. Table I shows that the best experts

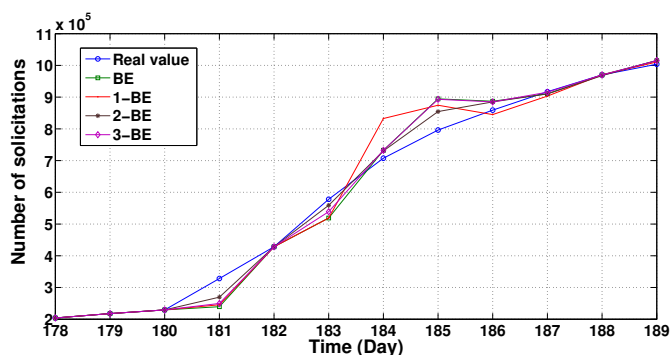


Fig. 4: The reactivity of forecasters in phase change.

selected by the forecaster to build its prediction depend on the video content considered. For instance, the video content 1b selected the DES, Constant Basic, Dynamic basic and Polynomial Regression as the best experts over time, whereas the video content 1e selected AMA and GMA as the best experts. Table I justifies the fact that we need to use different experts. In addition, simulation results show that these best experts vary over time. Hence, a forecaster is needed to select the best experts at any given time.

TABLE I: Best experts for different contents.

	DES	Constant basic	Dynamic basic	AMA	GMA	PR
1a	×					
1b	×	×	×			×
1c	×	×	×			
1d	×	×				
1e				×	×	
1f	×	×		×	×	

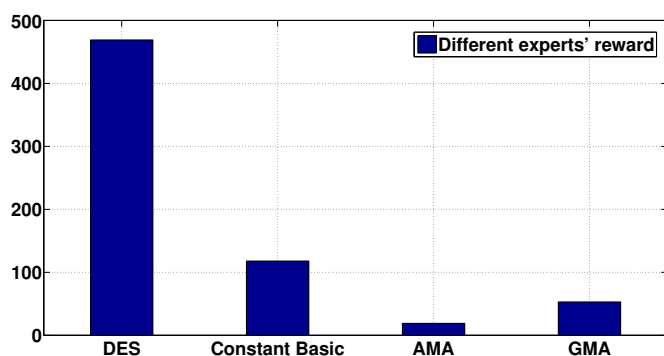


Fig. 5: Rewards obtained by different experts.

Figure 5 depicts the rewards obtained by the experts considered for the video content 1f. An expert is rewarded each time its cumulated loss belongs to the $K > 1$ smallest ones. We observe that DES is considered as a best expert 465 times, whereas AMA is selected only 23 times. This illustrates the importance of the forecaster's role.

We now evaluate the best experts with regard to three objectives we want to minimize simultaneously. These objectives

are 1) the cumulated loss, 2) the maximum instantaneous loss and 3) the inverse of the reward. We consider two video contents: 1a for Figure 6 and 1b for Figure 7. For each content, we depict the accuracy of three experts, each expert minimizing an objective.

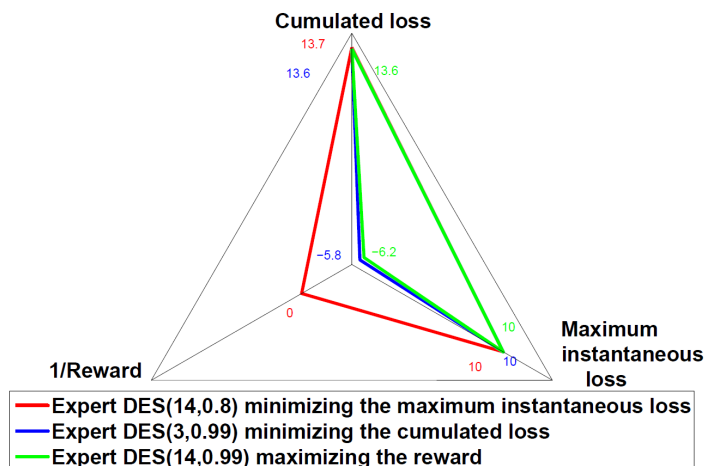


Fig. 6: The objectives minimized for content 1a.

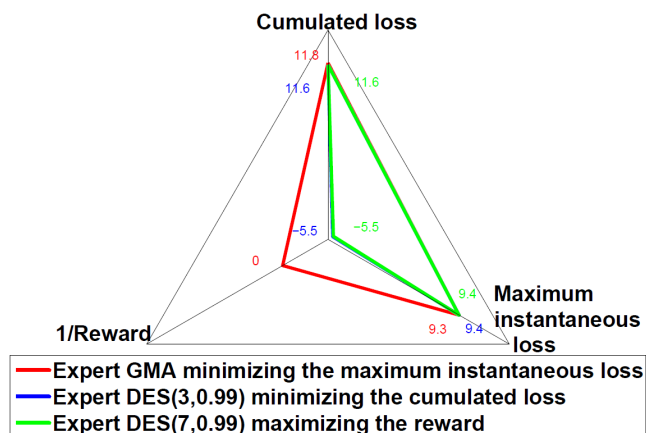


Fig. 7: The objectives minimized for content 1b.

As observed in Figures 6 and 7, no expert simultaneously minimizes the three objectives. However, in Figure 6, we notice that the DES expert with parameters $ow = 14$ and $\alpha = 0.99$ maximizes the reward and minimizes the cumulated loss. The expert minimizing the maximum instantaneous loss has a poor reward. It is rarely selected as a best expert. In Figure 7, we obtain similar conclusions but with different experts. The expert maximizing the reward and minimizing the cumulated loss is DES with the parameter $\alpha = 0.99$. The Geometrical Moving Average adjusted expert is the one minimizing the maximum instantaneous loss.

IV. CONCLUSION

Machine learning techniques can be used in CDNs to predict the popularity of video contents. Each expert uses its own prediction method to build its prediction. In this paper, we evaluated the performances of a wide range of experts using various prediction methods (e.g. exponential smoothing, basic, adjustment with different corrections, polynomial regression, Savitzky Golay filtering). We used real traces of video content solicitations in YouTube CDN. Simulation results show that, on the one hand, the best experts (i.e. those minimizing the cumulated loss) depend on the profile of the video content considered, and on the other hand, the best experts vary over time. This is why we introduced a decision maker, called a forecaster, that builds its own prediction from the predictions given by the best experts. By comparing two types of forecasters we concluded that the KBE forecaster outperforms the BE forecaster. In a further work, we will use the content popularity predictions to decide the contents that will be cached in the nodes near the end user. A certain percentage of a cache size will be reserved for the most popular contents. The remaining storage capacity will be shared among the solicited contents depending on their predicted popularity. Such a caching technique will be compared with classical techniques in terms of hit ratio.

REFERENCES

- [1] G. Szabo and B. A. Huberman, *Predicting the popularity of online content*, Communications of the ACM, vol. 53, no. 8, pp. 8088, 2010.
- [2] H. Pinto, J. M. Almeida, and M. A. Goncalves, *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. 365374.
- [3] 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. 607616.
- [4] G. Gursun, M. Crovella, and I. Matta, *Describing and forecasting video access patterns*, in INFOCOM, 2011 Proceedings IEEE. IEEE, 2011, pp. 1620.
- [5] X. Gu, and H. Wang *Online Anomaly Prediction for Robust Cluster Systems*, ICDE 2009, Shanghai, China.
- [6] Y. Wen, R. Wolski, and C. Krintz *Online Prediction of Battery Lifetime for Embedded and Mobile Devices*, PACS 2005
- [7] F. Salfner, M. Lenk, and M. Malek *A Survey of Online Failure Prediction Methods*, CSUR 2010
- [8] J. Kang, S. Seo, and J.W. Hong *Personalized Battery Lifetime Prediction for Mobile Devices based on Usage Patterns*, JCSE 2011
- [9] D. Marinca, P. Minet, and N. Ben Hassine *An Efficient Learning Technique to Predict Link Quality in WSN*, PIMRC 2014, Washington, USA.
- [10] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*, Cambridge University Press, 2006.
- [11] R. Shumway, D. Stoffer, *Time Series Analysis and Its Applications*, Springer, 2013
- [12] D. Tong, Z. Jia, X. Qin, C. Cao, C. Chang, *An Optimal Savitzky-golay Filtering Based Vertical Handoff Algorithm in Heterogeneous Wireless Networks*, Journal of Computers, November 2014.
- [13] N. Ben Hassine, D. Marinca, P. Minet, D. Barth, *Popularity Prediction in Content Delivery Networks*, PIMRC 2015, Hong-Kong, China, September 2015.