Predicting popularity of online videos using Support Vector Regression

Tomasz Trzcinski and Przemyslaw Rokita

***Abstract*—In this work, we propose a regression method to predict the popularity of an online video based on temporal and visual cues. Our method uses Support Vector Regression with Gaussian Radial Basis Functions. We show that modelling popularity patterns with this approach provides higher and more stable prediction results, mainly thanks to the non-linearity character of the proposed method as well as its resistance against overfitting. We compare our method with the state of the art on datasets containing over 14,000 videos from YouTube and Facebook. Furthermore, we show that results obtained relying only on the early distribution patterns, can be improved by adding social and visual metadata.**

***Index Terms*—Computer Vision, Popularity Prediction, Support Vector Regression, Video Analysis.**

I. Introduction

RECENT years have brought an enormous increase in the popularity of online platforms, such as YouTube, Facebook, Twitter or Instagram, where users can easily share various content with other people. YouTube is the biggest video sharing website with over 1 billion users that watch hundreds of millions of hours and generate billions of views [1]. The most popular social network with almost 1.5 billion registered users is Facebook [2], followed by Instagram with over 400 million users [3] and Twitter with over 300 million active users sending 500 million tweets (short messages) per day [4]. Although not every social network user is equally active in creating and publishing content, it is estimated that 85% of Facebook users actually do engage in content creation process [5]. Among different types of content generated by the users, photos and videos become more and more popular, mainly thanks to the proliferation of mobile devices with embedded high-quality cameras, but also as a result of studies indicating that visual content leads to higher user engagement [6]. Since the amount of visual content accessible online is so high, one should expect that only small portion of this data gains significant popularity, while the rest remains seen only by a small audience [7]. This phenomenon has led to an inception of the term *viral video* which describes a movie uploaded online that is gaining audience in an exponential manner, often reaching millions of views within a few days after publishing.

In this context, the ability of predicting the number of views of a given video can serve multiple causes, from load balancing the throughput of the data centers and servers to adjusting marketing efforts of the media houses that publish advertisements online. The latter application becomes increasingly significant, as marketing agencies spend 13% more money on digital marketing each year, with an estimated $52.8 billion dollars spent in 2015 [8]. Moreover, social networks such as Facebook, allow the marketing agencies to promote their content by increasing the reach of their videos. In this context, estimating the future popularity of the video can improve the allocation of the promotional funds. For instance, if the video of a given publisher is expected to reach 1 million organic views and its predicted view count exceeds this number, the promotional funds can be spent on other less popular videos instead.

Predicting the popularity of videos published online is a challenging problem. First of all, the external context of the content plays an important role in the distribution patterns of the video, *i.e.* if the subject of a video is trending in other media (television, radio, newspapers), its popularity online is also expected to be high. Secondly, the structure of the network built around the publisher such as the number of its friends and followers, and their respective friends and followers, has a substantial impact on the distribution of the content and therefore its future popularity. Last but not least, factors such as the relevance of the video to the final viewer and the relationship between real world events and the content are complex and difficult to capture, increasing the difficulty of popularity prediction.

Nevertheless, in the recent years several attempts have been made to address the problem of online content popularity prediction [9], [10], [11], [12], [13], [14]. Researchers analysed several types of online content, including news articles [11], Twitter messages [15], [16], images [13] and videos [10], [12]. Proposed prediction methods rely either on intrinsic features of the content, such as visual or textual cues [11], [13], or on social features describing the structure of the social network [14] or on early distribution patterns [9], [12]. To our knowledge, not too much attention was paid to the problem of combining different cues to predict the popularity of the online content in the context of online videos.

In this work, we propose a regression method based on Support Vector Regression with Gaussian Radial Basis Functions to predict the popularity of online videos. As video features we use visual cues computed before the video is published as well as early popularity pattern of the video once it is released online. We evaluate our method on datasets containing over 14,000 online videos uploaded to YouTube and Facebook. The contributions of this paper are the following:

- We introduce a new popularity prediction method for

T. Trzcinski and P. Rokita are with the Institute of Computer Science, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland e-mail: t.trzcinski@ii.pw.edu.pl.

Manuscript submitted.

online video content that relies on SVR with Gaussian RBF kernel and show that it outperforms the state of the art.
- We show that results obtained relying only on the early distribution patterns as done in [9], [12], can be improved by adding visual and social metadata, such as number of faces shown throughout the video or the number of comments recorded for a video.
- We collect and present a new dataset of online videos uploaded to the largest social network along with the corresponding social interactions metadata.

The remainder of this paper is organized in the following manner. In Section II we give an overview of the state of the art. In Section III we discuss the features used to predict the popularity of online videos using methods described in Section IV. Section V presents the results and we conclude this work in Section VI.

II. Related work

Due to the enormous growth of the number of Internet users and online data available, popularity prediction of online content has received a lot of attention from the research community. Early works have focused on user web-access patterns [17] and more specifically on the distribution of the video content [18], as it accounted for a significant portion of the Internet traffic and the findings could be used to determine the benefits of caching. Once the general access patterns were understood, the attention of the research community was attracted to the actual popularity prediction of various content types.

Textual content, such as Twitter messages, Digg stories or online news, is typically distributed very fast and catches users' attention for relatively short period of time [19]. Its popularity, measured in number of user actions such as comments, re-tweets or likes, is therefore highly skewed and can be modelled, e.g. with log-normal distribution [20]. Video content exhibits similar heavy-tailed distribution, while its popularity is typically measured by the number of views [21]. Thanks to the availability of the video content and related popularity data via YouTube platform, where every minute over 100 hours of video is uploaded [21], researchers were able to investigate other aspects related to the video content distribution, such as prediction of the peak popularity time of the video [22] or identifying popularity evolution patterns [23]. However, most if not all methods used to predict popularity of a given video rely on its early evolution pattern [9], [10], [12] or its social context [14]. Contrary to the method proposed in this paper, they do not exploit additional visual cues to improve their prediction accuracy.

In particular, [9] observe a log-linear relationship between the views of the YouTube videos at early stages after the publication of the material and later times. The reported Pearson correlation coefficient between the log-transformed number of views after seven and thirty days after publication exceeds 0.9, which suggests that the more popular submission is at the beginning, the more popular it will remain later.

Building up on the log-linear model of [9], [12] proposed to extend their approach with Multivariate Linear (ML) model that uses multiple inputs from previous stages (values of views received by a video in the early times after publication) to predict the future popularity of the video. On top of the Ordinary Least Squares regressor, they also experimented with the Ridge regressor using Radial Basis Functions (RBF) which reduces the prediction error by 20% on average with respect to the method of [9]. In this paper, we follow this lead and propose to use Gaussian RBF as a Support Vector Regression kernel [24].

To improve the prediction accuracy, Xu et al. [14] propose to add information about the structure of publisher's social network, including the proportion of the users who viewed and shared a video as well as the number of their followers. Their so-called Social-Forecast method aims to maximize the forecast reward defined as a trade-off between prediction accuracy and the timing of the prediction. Although the method shows improved accuracy in terms of forecast reward, it requires fairly detailed data concerning social network structure, which is not always available. For instance, Facebook, the social network with the highest number of registered users, does not allow to browse users' history of viewed videos and its followers' counts by public entities. Therefore, the Social-Forecast method, evaluated on the Chinese RenRen social network database where those metrics are publicly available, has to be adapted to other platforms if needed.

Although it is not the focus of this paper, several approaches have been made to predict the popularity of online content based on several information sources [19], [25]. For instance, [25] use data from Twitter to detect YouTube videos that will receive a significant growth in popularity. The model is based on the extraction of popular and trending topics on Twitter and linking them to the corresponding YouTube videos. This results in 70% higher accuracy of significant popularity growth prediction compared to the single-domain models that use only data from YouTube.

All the above mentioned works propose to predict future popularity of online content after the content is made publicly available, i.e. after the publication date. It is much more interesting, although more challenging as well, to attempt to predict the popularity of a given piece of content *before* it is published. Khosla et al. [13] address this problem in the context of images. More precisely, the proposed method analyses visual and social cues of the images published on Flickr to predict its relative popularity after the publication. Using a dataset of over 2 million images, the authors demonstrate that features such as image color or number of friends of the publisher play significant role in determining the future popularity of a given photo. Moreover, using those cues, they are able to predict the normalized view count of images. Following this methodology, we use computer vision algorithms to calculate visual content metadata and verify if combining it with early evolution data can improve prediction accuracy.

III. Features

In this section we discuss features of the videos used to predict their popularity. We start with the description of visual content features that can be extracted before a video

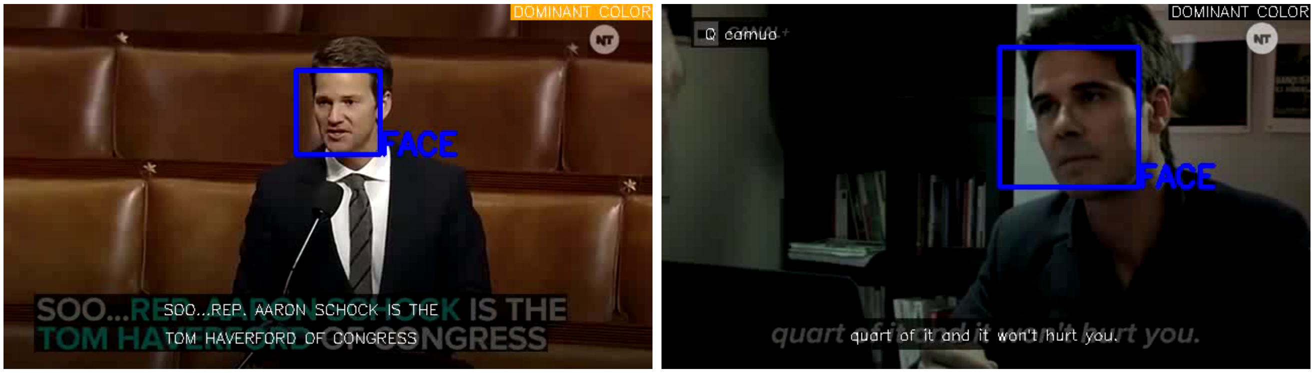

Fig. 1. Results of the visual content analysis of a sample video frame. The dominant color is displayed in the top right corner of the frame. The face is detected using cascade classifier. The text region is detected and faded to enable impainting OCR results. Best to be seen on a screen.

is published online. We then follow with an overview of temporal features recorded after the video was published, such as number of aggregated video views or other interactions available in the social network.

A. Visual features

Features presented here are computed using several computer vision algorithms applied on raw video data. The resulting visual metadata is then used to provide additional cues for the prediction methods.

Video characteristics: We use simple video metadata describing video length, number of frames, video resolution and frame dimensions.

Color: We first cluster the color space into 10 distinct classes depending on their coordinates in the HSV space: *black, white, blue, cyan, green, yellow, orange, red, magenta* and *other*. Then, for each frame of a video, we assign a pixel to a single color and identify the dominant color of every frame. We aggregate the results of the color classification and represent color feature of a video as a histogram of dominant colors across the frames as well as dominant video color.

Face: Using a face detector based on a cascade classifier [26], we detect the region of a frame with a face. We then count number of detected faces per frame, number of frames with faces present and the size of the face regions with respect to the frame size. The results are averaged across all video frames and stored.

Text: With a combination of edge detection and morphological filters, we identify the regions of the image with imprinted subtitles and apply Tesseract-OCR engine[1] to validate the detection. We then report the following textual characteristics of a video: a portion of the frames with imprinted text in the video and an average ratio of the text region size with respect to the frame size.

Scene dynamics: To quantify scene dynamics of a video, we first employ Edge Change Ration algorithm [27] and determine shot boundaries. We then analyse the boundaries distribution and extract the number of shots and an average shot length in seconds. We also classify the shots as hard or soft cuts and save the corresponding histogram of shots.

[1]https://code.google.com/p/tesseract-ocr/

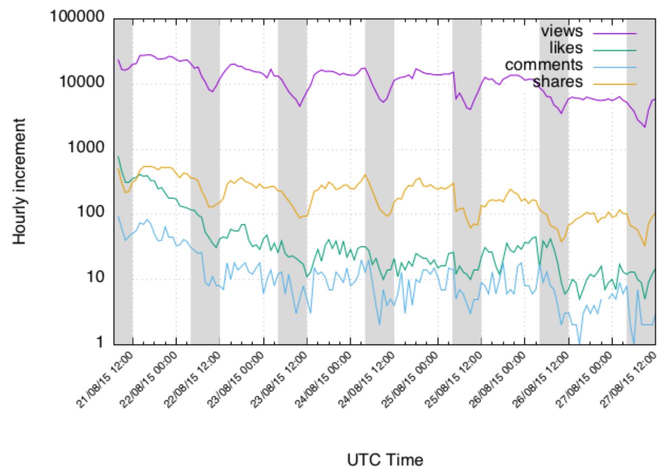

Fig. 2. Plot of hourly increments in number of views, likes, comments and shares for a sample Facebook video. The grey areas indicate night time according to the EDT Time Zone. The evolution patterns of those metrics are used in the paper to predict popularity of a given video.

Clutter: We use a Canny edge detector [28] to quantify the clutter present in the video. We report the ratio of the edge pixels detected and all pixels in a frame, averaged across all frames in a video.

Rigidity: To evaluate the scene rigidity we estimate the homography between two consecutive frames using a combination of FAST feature point detector [29] and BRIEF descriptor [30]. We then save an average number of frames where a valid homography between current and previous frames can be found.

Thumbnail: Building upon the work of [13], we also computed popularity score using Popularity API[2] of the video thumbnail and saved the result.

Fig. 1 shows a sample result of the computer vision analysis of two video frames.

B. Temporal features

Once a video is made available online, we are able to collect data related to its popularity, that is the number of views as

[2]http://popularity.csail.mit.edu/

well as other social interactions aggregates. Figure 2 shows a set of hourly increments in views, likes, comments and shares for a sample Facebook video. The evolution patterns of the video statistics provide an important cue for the popularity prediction methods, as [9] reported high correlation between log-transformed view counts early after the publication and later on. These results are also confirmed by the experiments presented in Section V.

IV. Methods

In this paper, following the works of [9], [12] we cast the problem of popularity prediction as a regression task. More precisely, our goal is to predict the *number of views* of a video v at time t_t, given features available from the first t_r days after publication (where $t_r < t_t$). In this section, we discuss the regression methods used for the prediction in Section V.

A. Univariate Linear (UL) Regression

Based on the high correlation observed between log-transformed early and late popularity counts of online content, [9] proposed to use a simple regressor to predict the future popularity of a given video v. According to this model, the number of views of a video v can be calculated at time t_t as:

$$\hat{N}(v, t_r, t_t) = \exp\left(\alpha(t_r, t_t) \cdot \ln N(v, t_r)\right) \tag{1}$$

where $\exp$ defines natural exponential function, $\hat{N}(v, t_r, t_t)$ defines predicted number of views for video v at time t_t when prediction is made at time t_r. $\alpha(t_r, t_t)$ is a weight learnt from training videos $v_t \in T$ and $N(v, t_r)$ is the number of views at time t_r. Weight $\alpha(t_r, t_r)$ can be computed using the ordinary least squares model.

B. Multivariate Linear (ML) Regression

Pinto et al. [12] propose to extend the UL regression model by including also the views accumulated by the video before t_r. In other words, they increase the dimensionality of the input feature vector. Instead of using a single cumulated view count at time t_r, they sample the timeline between publication time t_0 and reference time t_r and use the number of views received in those sampling intervals (views' *increments* or *deltas*) to form a feature vector. The proposed method called Multivariate Linear (ML) Regression predicts the popularity of the video v at time t_t as a linear combination of the feature values and can be expressed as:

$$\hat{N}(v, t_r, t_t) = \sum_{i=1}^{r} \alpha(t_i, t_t) \cdot \big(N(v, t_i) - N(v, t_{i-1})\big) \tag{2}$$

where $\{\alpha(t_i, t_t)\}_{i=1}^{r}$ are model parameters learned from training data T and the term $\big(N(v, t_i) - N(v, t_{i-1})\big)$ corresponds to the view deltas in the i-th sampling interval.

C. MRBF Regression

The ML Regression model is able to capture more information about the evolution pattern thanks to different weights assigned to time intervals. However, the weights learned from the training data cannot capture the intrinsic variations of the evolution patterns within the training dataset videos. Therefore, [12] propose to extend their ML model by introducing a similarity notion between the videos based on their evolution patterns. The so-called MRBF regression uses Radial Basis Functions (RBF) to calculate the distance between the videos and predicts the number of views based on the views increments as well as distances to a set of pre-selected training videos $v_c \in C$:

$$\begin{aligned} \hat{N}(v, t_r, t_t) &= \underbrace{\sum_{i=1}^{r} \alpha(t_i, t_t) \cdot \big(N(v, t_i) - N(v, t_{i-1})\big)}_{\text{ML regression}} + \\ &+ \underbrace{\sum_{v_c \in C} \omega_{v_c} \cdot \Phi(v, v_c)}_{\text{RBF features}} \end{aligned} \tag{3}$$

where $\Phi(x, y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right)$ is a Gaussian RBF with σ parameter and a set of videos C to be selected during cross-validation. The above problem can be solved with ordinary least squares, similarly to the previously discussed methods. However, the additional set of input features increases the risk of overfitting. Therefore, [12] propose to use Ridge regression [31] instead.

D. SVR

MRBF Regression model encompasses linear and non-linear dependencies within the popularity evolution patterns using a combination of two methods: ML regression and RBF features. Instead of combining those two distinct techniques, we propose a new method that unifies them using Support Vector Regression (SVR) [24] with Gaussian RBF kernel. Using RBF kernel allows us to map the feature vectors into a non-linear space where the relations between early and later view counts are easier to capture. According to the proposed SVR model the popularity of the video v can be predicted as:

$$\hat{N}(v, t_r, t_t) = \sum_{k=1}^{K} \alpha_k \cdot \Phi\Big(X(v, t_r), X(k, t_r)\Big) + b \tag{4}$$

where $\Phi(x, y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right)$ is a Gaussian RBF with σ parameter, $X(v, t_r)$ is a feature vector for video v available at time t_r and $\{X(k, t_r)\}_{k=1}^{K}$ is a set of support vectors returned by the SVR algorithm along with a set of coefficients $\{\alpha_k\}_{k=1}^{K}$ and intercept b. Unless stated otherwise, we use a vector of view deltas as feature vectors, as proposed in [12], that is $X(v, t_r) = \{N(v, t_i) - N(v, t_i - 1)\}_{i=1}^{r}$. We found optimal values for the hyperparameter C of the SVM optimization and σ of the RBF kernel with a grid search in a preliminary set of experiments and in the remainder of this paper the following values are used: $C = 10, \sigma = 0.005$.

TABLE I

YOUTUBE VIDEO DATASETS. RESULTS OF THE PREDICTION FOR UL, ML, MRBF AND SVR METHODS REPORTED AS SPEARMAN RANK CORRELATION $\pm$ 95% CONFIDENCE INTERVAL ($t_r = 6$ DAYS, $t_t = 30$ DAYS). SVR OUTPERFORMS THE COMPETITORS WHILE PROVIDING MORE STABLE PREDICTION ACCURACY (SMALLER CONFIDENCE INTERVAL).

Dataset	UL	ML	MRBF	SVR
Random dataset	0.9093 $\pm$ 0.0038	0.9368 $\pm$ 0.0025	0.9378 $\pm$ 0.003	**0.9558 $\pm$ 0.0017**
Top dataset	0.9494 $\pm$ 0.0032	0.9602 $\pm$ 0.0033	0.9651 $\pm$ 0.0033	**0.9723 $\pm$ 0.0026**

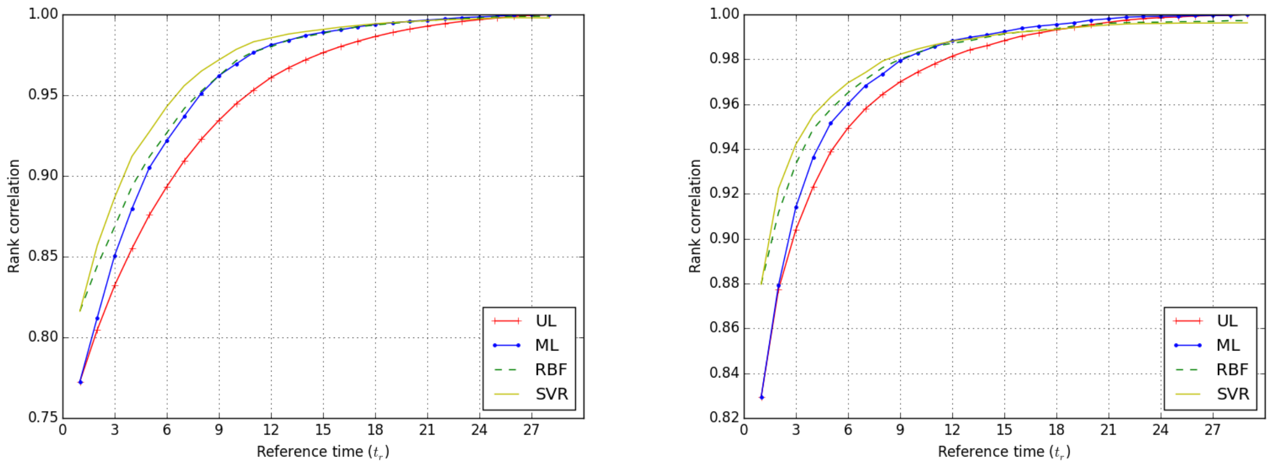

Fig. 3. Prediction results for the YouTube video datasets: Random (left) and Top (right). The reference time t_r indicates number of days since publication and the target time is $t_t = 30$ days. The proposed SVR method outperforms the state-of-the-art methods, among which the MRBF performs the best, for both datasets. The performance improvement is more significant for $t_r < 12$.

V. RESULTS

In this section we compare the state-of-the-art methods described in section IV, namely the UL, ML and MRBF against the proposed SVR method. To that end, we employ 3 datasets containing more than 14,000 videos. For ML and MRBF methods we use implementations obtained from their authors. For UL and SVR we use our own Python implementation based on the Scikit-learn package[3]. We first evaluate the methods using only the data with temporal evolution of the views (without visual or social cues) on two publicly available datasets of YouTube videos: Top and Random [32]. Then, we show how we can improve the prediction precision with additional visual and social cues obtained using a dataset of Facebook videos.

A. Datasets

Top and Random datasets [12] contain data gathered for YouTube videos, such as time evolution of the number of views, comments, favorites and ratings. The Top dataset is a compilation of those results for a total of 27,212 videos taken from the top-100 most popular videos of each country in the world. The Random dataset, contains the same type of data gathered for 24,484 unique randomly selected videos. Similarly to [12], we also preprocess both YouTube datasets and remove the videos with incomplete statistics, with less than 30 days of data or those that receive less than 10 views. The final preprocessed datasets generated this way have 6,829 (Random) and 5,018 (Top) videos.

To evaluate the prediction methods in the context of social media, we also collected data for 1,820 videos uploaded to Facebook between August 1^{st} 2015 until October 15^{th} 2015. The videos were uploaded by several public Facebook publishers, including the AJ+[4] and BuzzFeedVideo[5]. We implemented a crawler that uses Facebook Graph API[6] to browse Facebook publishers' pages and retrieve the publicly available information regarding number of interactions with a given video, that is number of shares, likes and comments. Since the number of views of a video is not publicly available through the Graph API, we retrieve this data using simple URL scraper of a video page. We plan to increase the number of records and release this dataset for the public to enable further research on that topic.

B. Evaluation protocol

To evaluate the performances of prediction methods, we follow the approach of [12] and use 10-fold cross validation. For every dataset used, we randomly split all the samples into 10 equal-sized folds. We then use 9 folds for training and one for testing. We repeat the process 10 times, every time testing the methods on a distinct fold and training them with the remaining 9 folds. We report here average results

[3]http://scikit-learn.org/

[4]www.facebook.com/ajplusenglish

[5]www.facebook.com/BuzzFeedVideo

[6]https://developers.facebook.com/docs/graph-api

across all the 10 test sets along with the corresponding 95% confidence interval. As an evaluation metric we use Spearman rank correlation, as in [13].

C. YouTube datasets

We first evaluate the popularity prediction methods on two sets of YouTube videos: Random and Top datasets. Figure 3 shows the results in terms of rank correlation for reference time $t_r \in (1, 29)$ days and target time $t_t = 30$). The performance of our proposed SVR method is higher than the competitors for both datasets and across the reference time values. The improvement over the state-of-the-art methods is more significant for $t_r < 12$, while for t_r being closer to the target time t_t is saturates. This is no surprise, as the more time passes, the easier the prediction is. Out of the competitors, MRBF performs the best which confirms the results of [12]. For the quantitative analysis, we also show the average results along with the 95% confidence interval for $t_r = 6$ in Table I. Not only does the SVR method perform best, but its 95% confidence interval is also significantly smaller than the other methods, which means that it provides more stable prediction accuracy across different videos.

D. Facebook dataset

Secondly, we evaluate the performances using Facebook dataset. Figure 4 shows the obtained results. They confirm that our proposed SVR method performs better than the competitors. Since the reference time t_r is denoted in hours and is much smaller than the target time $t_t = 7$ days, the improvement is visible across all t_r values. Table II shows the average results for $t_r = 6$ hours along with the corresponding confidence intervals.

In Figure 4 we also plot the results obtained when we use different data sources as the feature vectors for the SVR method. As described in Section IV-D, our proposed SVR method by default uses views increments as feature vectors. Adding time evolution of social interactions data, such as the number of shares, comments and likes, improves the performance of the SVR, especially for the higher values of t_r. We also evaluated the performance of the SVR when relying only on the visual features available before the publication of the video. As shown in Table II, the average rank correlation in this case is equal to 0.197 which is much lower than the results obtained for the SVR based on the temporal evolution of the views. Nevertheless, combining the data from both temporal and visual cues improves the overall prediction accuracy and provides the best performance among all evaluated methods, as can be seen in Figure 4.

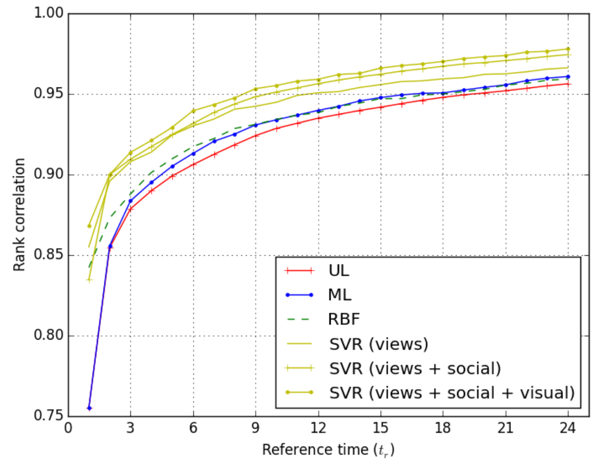

Fig. 4. Prediction results for Facebook dataset. The reference time t_r indicates number of hours since publication and the target time is $t_t = 7$ days. SVR provides better performance than other methods. When adding other types of data to the feature vector, the performance of SVR is improved even more, reaching the peak with features based on the time evolution of the views, social interactions and visual cues.

VI. Conclusions

In this paper, we propose to use Support Vector Regression with Gaussian Radial Basis Functions to predict the popularity of online video content measured as the number of views. We showed that modelling the dependencies between popularity evolution patterns with Gaussian RBF kernel leads to a significant performance improvement. Our method was evaluated on three datasets containing a total of over 14,000 videos. Moreover, our results suggest that using only visual content of the data proves to be insufficient for accurate prediction. Nevertheless, visual cues, combined with social interactions data, can improve the results obtained using only early views evolution data.

Acknowledgments

This work was partially funded by the grant of the Dean of the Faculty of Electronics and Information Technology at Warsaw University of Technology (project II/2015/GD/1).

References

[1] YouTube, "Press statistics," https://www.youtube.com/yt/press/statistics.html, 2015, [Online; accessed 19-October-2015].
[2] Facebook, "Company info," http://newsroom.fb.com/company-info/, 2015, [Online; accessed 06-October-2015].
[3] Instagram, "Press," https://instagram.com/press/, 2015, [Online; accessed 06-October-2015].
[4] Twitter, "Company info," https://about.twitter.com/company, 2015, [Online; accessed 06-October-2015].
[5] Adage.com, "Facebook 85 users creating content," http://adage.com/article/digital/facebook-85-users-creating-content/236358/, 2015, [Online; accessed 06-October-2015].
[6] Twitter, "What fuels a tweet engagement," https://blog.twitter.com/2014/what-fuels-a-tweets-engagement/, 2015, [Online; accessed 16-October-2015].
[7] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007.
[8] TechCrunch, "2015 ad spend rises to $187b, digital inches closer to one third of it," http://techcrunch.com/2015/01/20/2015-ad-spend-rises-to-187b-digital-inches-closer-to-one-third-of-it/, 2015, [Online; accessed 19-October-2015].
[9] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Commun. ACM*, vol. 53, no. 8, pp. 80–88, Aug. 2010.
[10] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. L. Eager, and A. Mahanti, "Characterizing and modelling popularity of user-generated videos." *Perform. Eval.*, vol. 68, no. 11, pp. 1037–1055, 2011.

TABLE II
FACEBOOK VIDEOS DATASET. RESULTS OF THE PREDICTION FOR THE UL, ML, MRBF AND SVR METHODS REPORTED AS SPEARMAN RANK CORRELATION $\pm$ 95% CONFIDENCE INTERVAL ($t_r = 6$ HOURS, $t_t = 7$ DAYS). SVR METHOD BASED ON THE TIME EVOLUTION OF THE NUMBER OF VIEWS OUTPERFORMS THE STATE OF THE ART METHODS. PREDICTION ACCURACY OF THE SVR METHOD WITH ONLY VISUAL CUES IS MUCH LOWER THAN THE ACCURACY OF THE METHODS RELYING ON THE TEMPORAL FEATURES. NEVERTHELESS, COMBINING VISUAL AND TEMPORAL FEATURES LEADS TO HIGHER ACCURACY. THE BEST PERFORMANCE AMONG ALL TESTED CONFIGURATIONS IS ACHIEVED BY THE SVR METHOD WITH TIME EVOLUTION OF THE NUMBER OF VIDEOS, SOCIAL INTERACTIONS AND VISUAL CUES.

UL	ML	MRBF	SVR			
			visual	views	views + social	views + social + visual
0.9061 $\pm$ 0.0366	0.9130 $\pm$ 0.0382	0.9173 $\pm$ 0.0379	0.197 $\pm$ 0.1852	0.9301 $\pm$ 0.0191	0.9316 $\pm$ 0.016	**0.9395 $\pm$ 0.0165**

[11] R. Bandari, S. Asur, and B. A. Huberman, "The Pulse of News in Social Media: Forecasting Popularity," *CoRR*, vol. abs/1202.0332, 2012. [Online]. Available: http://arxiv.org/abs/1202.0332

[12] 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*, ser. WSDM '13, 2013, pp. 365–374.

[13] A. Khosla, A. D. Sarma, and R. Hamid, "What makes an image popular?" in *International World Wide Web Conference (WWW)*, 2014.

[14] J. Xu, M. van der Schaar, J. Liu, and H. Li, "Forecasting popularity of videos using social media," *CoRR*, vol. abs/1403.5603, 2014.

[15] M. Osborne and V. Lavrenko, "V.: Rt to win! predicting message propagation in twitter," in *In: 5th ICWSM*, 2011.

[16] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in twitter," in *Proceedings of the 20th International Conference Companion on World Wide Web*, 2011.

[17] V. Almeida, A. Bestavros, M. Crovella, and A. de Oliveira, "Characterizing Reference Locality in the WWW," in *Proceedings of PDIS'96: The IEEE Conference on Parallel and Distributed Information Systems*, 1996.

[18] M. Chesire, A. Wolman, G. M. Voelker, and H. M. Levy, "Measurement and analysis of a streaming-media workload," in *Proceedings of the 3rd Conference on USENIX Symposium on Internet Technologies and Systems - Volume 3*, 2001.

[19] C. Castillo, M. El-Haddad, J. Pfeffer, and M. Stempeck, "Characterizing the life cycle of online news stories using social media reactions," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work And Social Computing*, 2014.

[20] M. Tsagkias, W. Weerkamp, and M. de Rijke, "News comments: Exploring, modeling, and online prediction." in *ECIR*, ser. Lecture Notes in Computer Science. Springer, 2010.

[21] A. Tatar, M. D. de Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of web content," *Journal of Internet Services and Applications*, vol. 5, 2014.

[22] L. Jiang, Y. Miao, Y. Yang, Z. Lan, and A. G. Hauptmann, "Viral video style: A closer look at viral videos on youtube," in *Proceedings of International Conference on Multimedia Retrieval*, 2014.

[23] 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*, no. 41, pp. 15 649–15 653, 2008.

[24] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems 9*, 1996, pp. 155–161.

[25] S. D. Roy, T. Mei, W. Zeng, and S. Li, "Towards cross-domain learning for social video popularity prediction." *IEEE Transactions on Multimedia*, vol. 15, no. 6, pp. 1255–1267, 2013.

[26] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.

[27] A. Jacobs, A. Miene, G. T. Ioannidis, and O. Herzog, "Automatic shot boundary detection combining color, edge, and motion features of adjacent frames," in *TRECVID 2004 Workshop Notebook Papers*, 2004, pp. 197–206.

[28] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, Jun. 1986.

[29] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.

[30] M. Calonder, V. Lepetit, M. Özuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: computing a local binary descriptor very fast," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281–1298, 2012.

[31] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. Springer New York Inc., 2001.

[32] F. Figueiredo, J. M. Almeida, M. A. Gonçalves, and F. Benevenuto, "On the dynamics of social media popularity: A youtube case study," *ACM Trans. Internet Technol.*, vol. 14, no. 4, pp. 24:1–24:23, Dec. 2014.