Predicting Popularity of Facebook Videos Through Visual Features Using Support Vector Machine Classifier

Bruna M. Dalmoro , Soraia R. Musse

[1] School of Technology, Graduate Program of Computer Science
Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Brazil

bruna.dalmoro@edu.pucrs.br, soraia.musse@pucrs.br

Abstract. *With the popularization of social networks, the sharing and consumption of content in video format becomes easier. Understanding what makes a video popular and being able to predict its popularity in number of views is useful for both content creators and advertising. In this work, we explore visual features extracted from 1,820 Facebook videos in order to predict whether they will reach more than a certain number of views on the seven days after publication. For this purpose, we used Support Vector Machine with Gaussian Radial Basis Function classification model. Using only visual features as predictors, the model with Video Characteristics and Rigidity features combined reached Kappa of 0.7324, sensitivity of 0.8930, and positive predictive value of 0.8930.*

1. Introduction

Currently, there is a great facility to publish videos on the web, mainly through social networks, such as YouTube and Facebook, for example. On YouTube, about 400 hours of videos per minute are published, accessed by 2 billion monthly users who generate billions of views daily [Youtube About 2020]. Regarding content creators, the number of channels with more than one million subscribers grew by more than 65% per year. When it comes to revenues, the number of channels that had six digits annual revenue on YouTube, grew by over 40% per year [Youtube About 2020]. Part of these revenues can come from ads [YouTube Support 2019], as it also happens on Facebook [Facebook for Business 2020]. It is advantageous to advertise on social networks: on mobile devices alone, YouTube reaches more people between 18 and 34 years old in the USA than any other TV channel [Youtube About 2020]. So, understanding what makes a video popular and being able to predict its popularity is a problem that companies like Facebook and Netflix have invested in solving. This predictive power is useful both for advertisements, since they can be directed to videos of greater reach, and for content creators, with regard to the management and production of content based on characteristics that generate more views.

In this paper, we use visual features extracted from 1,820 videos published on Facebook [Trzciński and Rokita 2017], to predict the popularity of videos. We use a Support Vector Machine (SVM) classifier, developed with the *caret* package [Kuhn et al. 2019] on R [R Core Team 2020] software and compare the results of our classification model with the regression model provided from previous competitive work [Trzciński and Rokita 2017]. Our results show that the selected visual features, presented in the paper, are related to the number of views of a video. We also show that

Support Vector Machine with Gaussian Radial Basis Function classification is effective for these kinds of tasks.

2. Related Work

The sharing of content on social networks has been the reality of a large part of the population. Yet, the opportunity to use content sharing platforms as digital advertising channels was identified. So, one challenge is to identify which contents are more relevant and provide more visibility to the advertisement running with it. It has already been explored in literature [Kong et al. 2018], [Trzciński and Rokita 2017], [Khosla et al. 2014].

However, the best scenario is to be able to predict the popularity of content before it is published, when the content are pictures. In [Khosla et al. 2014], the authors used visual features to make popularity prediction. In [Trzciński and Rokita 2017], the authors produced a work aiming the same goal, but with video content. In both work, Support Vector Regression with Radial Basis Function using Gaussian kernel was used. Trzciński and Rokita using data collected from Facebook pages, propose a method called Popularity-SVR, that predicts popularity of an online video using Support Vector Regression (SVR) [Trzciński and Rokita 2017]. The Facebook video data included visual features and temporal features, that is, features captured soon after the content was published, such as number of views over time. To assess the performance of the proposed predictive model, they used Spearman's correlation [Spearman 1904], a non-parametric measure of statistical dependence between two variables. This measure ranges from -1 to 1, where -1 indicates a perfect inverse relationship between these variables, 1 indicates a perfect positive relationship and when the relationship between them is closer to 0, the relationship between them is smaller. When it comes to visual features, the Popularity-SVR shows that, individually, deep features provide the highest Spearman correlation value with video popularity (0.13), followed by the feature groups Clutter (0.12) and Scene Dynamics (0.08). Overall correlation value using all visual features reached over 0.23. However, the best results were obtained when visual features were combined with temporal features, where the Spearman correlation reached over 0.94. The aim of this work is to use only visual features extracted from Facebook videos to predict their popularity. We use Support Vector Machine with Gaussian Radial Basis Function to classify these videos into two groups: the most and the least popular, according to the number of views.

3. Dataset and Features

The dataset[1] used in this work was available by the authors [Trzciński and Rokita 2017]. The available file contains features extracted from 1,820 videos published on Facebook between August 1st and October 15th 2015 from pages such as AJ+[2] and BuzzFeedVideo[3]. Two types of features were considered in the data extract: temporal and visual features. The next sections describe some details about both features data.

3.1. Temporal Features

After the video is published, the temporal features show the number of views, likes, comments, and shares every hour, for seven days after posting, collected by the URL scraper

[1] http://ii.pw.edu.pl/~ttrzcins/facebook_dataset_2015.csv

[2] https://www.facebook.com/ajplusenglish

[3] https://www.facebook.com/BuzzFeedVideo

on the posting page [Trzciński and Rokita 2017]. In this work, we used only visual resources as predictor variables, while temporal resources, such as number of likes, comments, and views were excluded from the analysis. However, we use the number of views at the end of the period as a response variable. Therefore, the main reason for discarding temporal data in this work is that we want to analyze and investigate only visual data, so that this analysis can be produced before publication.

3.2. Visual Features

In this work, we hypothesize that visual features can be used to predict popularity of video content. The visual features were collected directly from the video, using various computer vision algorithms, as described in [Trzciński and Rokita 2017] (please refer to this paper for further details). The list of available visual data that was used in this work is:

1. **Video characteristics:** This class regards general video information, such as duration, frames per second, number of frames, and frame dimensions of the analyzed video.
2. **Dominant color:** The color space of the video was divided into 10 classes (black, white, blue, cyan, green, yellow, orange, red, magenta, and other) and each frame of each video was assigned to one of these classes. In addition, the data set contains information about which class of colors is dominant and what proportion of each color is present for each video.
3. **Face detection:** Presents information about the presence of faces in the video, such as the average number of faces per frame, the proportion of frames with faces, and the average proportion of the face size in relation to the size of the frame.
4. **Text detection:** Similar to face detection, it concerns information about the presence of text in the video, such as the proportion of frames with text and the average proportion of the text size in relation to the size of the frame.
5. **Scene dynamics:** It regards information about the number of shots in the video and classification of the shots as hard and soft cuts.
6. **Rigidity:** They provide information about the average video speed, a clutter metric, and a metric that specifies the video rigidity.

While Trzciński and Rokita propose the Popularity-SVR using Support Vector Regression [Trzciński and Rokita 2017], in our work we use Support Vector Classifier, as described in the next section. Indeed, when we consider this question as a classification problem, we understand that some milestones in terms of video visualizations are more interesting, and probably relevant. For example, 100,000 visualizations or 1 million are maybe good milestones.

4. Method

In this paper, our goal is to predict whether a video will be popular or not based on number of views, given its visual features computed using computer vision algorithms. This section presents the pre-processing phase executed on available data [Trzciński and Rokita 2017], the model tuning to configure SVM hyperparameters, and details about the used SVM classifier model. We performed all analysis and modeling using R software version 3.6.3 [R Core Team 2020], through caret package version

6.0-86 [Kuhn et al. 2008], on a computer with operational system Windows 10x64, processor i7-7500U CPU @ 2.7GHz and RAM 8 GB. Different from the work proposed by Trzcinski et al. [Trzciński and Rokita 2017], in our method, the popularity prediction is treated as a classification problem. In this case, we do not want to provide the exact number of views for a given video, but to identify whether the video will have more views than a certain pre-established milestone, 7 days after its publication. In this work, we tested 5 different milestones according to the number of views: **10,000**, **100,000**, **500,000**, **750,000** and **1 million** views. We named videos that have reached the milestone as *successful-videos*. Further details are presented in Section 5.

4.1. Pre-processing

Pre-processing of data is essential in building a statistical model. In this phase, it is possible to identify missing values or the relationship between the variables that can harm the modeling process. It is during pre-processing that the addition, deletion, or transformation of the dataset is done. According to Kuhn and Johnson [Kuhn and Johnson 2013], data preparation can create or break a model's predictive ability. In the pre-processing of our method, the missing values, zero- and near zero-variance feature predictors were analyzed, identifying correlated feature predictors and linear dependencies.

We identified four correlated feature predictors: number of frames is highly correlated with video duration, frame width is highly correlated with frame height, average proportion of frames with faces is highly correlated with average number of faces per frame, and two features about soft and hard cuts are complementary, so they have a -1 correlation. Consequently, the features about the number of frames, frame width, average proportion of frames with faces, and one of two features about shot cuts have been removed from the features list. Regarding the linear dependencies, QR decomposition [Goodall 1993] is used to determine whether features are linearly independent and then identify the sets of features involved in the dependencies if any. There was no need to treat missing values since the dataset has complete information for all videos. It was tested for feature predictors with zero- and near zero-variance, but none were identified, so there was no need to remove resources in this case. The train and test sets are splitted in proportion of 70% and 30%, respectively, preserving the overall class distribution of the response variable. Finally, we centered and scaled the data, to improve the numerical stability of some calculations [Kuhn and Johnson 2013].

4.2. Support Vector Machine Classifier

The technique used for predictive modeling of popularity of videos is a Support Vector Machine with Gaussian Radial Basis Function classification model. The Support Vector Machines classifier is a binary classifier algorithm that looks for an optimal hyperplane as a decision function in a high-dimensional space [Boser et al. 1992]. Given a set of labeled training patterns $(\mathbf{x}_i, y_i)$, $i = 1, ..., l$ where $\mathbf{x}_i \in \mathbb{R}^n$ e $y \in \{1, -1\}^l$, the algorithm finds the parameters of the decision functions $D(\mathbf{x})$ during a learning phase. The decision function has the following form: $D(\mathbf{x_i}) = \sum_{k=1}^{p} \alpha_k K(\mathbf{x}_k, \mathbf{x_i}) + b$, where $\mathbf{x}_k$ are the support vectors returned by algorithm, α_k are the coefficients, x is a feature vector for a video, the function K is a predefined kernel and b is the intercept. For non-linearly separable problems, Support Vector Machines can not find a separation hyperplane that provides a good generalization. For that, a kernel can be used to transform the data to a higher-dimensional space and thus a linear hyperplane can be obtained to proper separate the different classes. We used the Gaussian radial basis function kernel as follows:

$$K(\mathbf{x}_i, y_i) = exp(-\frac{\|\mathbf{x}_i - y_i\|^2}{\sigma^2}), \tag{1}$$

where $\sigma > 0$ is a parameter from Gaussian kernel. The model hyperparameters are the cost C, from Support Vector Machine, and σ from kernel.

4.3. Model Tuning

For each visualization milestone, we tested 21 different combinations of visual features. To search for the best hyperparameters for the model, we create a grid of values for the hyperparameter of the model. For σ hyperparameter, we define 12 possible values between 0 and 0.5, and for C hyperparameter, 7 values between 0.25 and 8. This setup results in 8,820 different models, one for each combination of visual features, according to Table 1, milestones and pair of hyperparameters.

Table 1. Combinations of group of features, as defined in Section 3.2, tested in the models and abbreviations that we use to refer to the feature setup.

Abbreviation	Combination of Visual Features
V	Video Characteristics
C	Dominant Color
F	Face Detection
T	Text Detection
D	Scene Dynamics
R	Rigidity
VC	V. Char. + Color
VF	V. Char. + Faces
VT	V. Char. + Text
VD	V. Char. + S. Dyn.
VR	V. Char. + Rigidity
VDC	V. Char. + S. Dyn. + Color
VDF	V. Char. + S. Dyn. + Faces
VDT	V. Char. + S. Dyn. + Text
VDR	V. Char. + S. Dyn. + Rigidity
VDRC	V. Char. + S. Dyn. + Rigidity + Color
VDRF	V. Char. + S. Dyn. + Rigidity + Faces
VDRT	V. Char. + S. Dyn. + Rigidity + Text
VDRTC	V. Char. + S. Dyn. + Rigidity + Text + Color
VDRTF	V. Char. + S. Dyn. + Rigidity + Text + Faces
Complete Model	V. Char. + S. Dyn. + Rigidity + Text + Color + Faces

Then, in the model tuning process, we use repeated 10-fold cross-validation, where three separate 10-fold cross-validations were used as the resampling scheme. For each combination of visual features and milestones, the pair of hyperparameters that generated the model with the largest Kappa [Landis and Koch 1977] was selected, thus leaving 105 models.

5. Results

We trained 8,820 different models with different configurations to predict whether a video will be a *successful-video* and which milestone it has achieved. After selecting hyperparameters, 105 models remained, combining different features to predict the number of views according to 5 different milestones: 10,000, 100,000, 500,000, 750,000, and 1 million views. To select the best models among the 105, three different metrics were used: Kappa, Sensitivity, and Positive Predictive Value.

Cohen's Kappa Coefficient is a statistical measure of agreement between classifications, which compares the model's classification with the response variable more robustly than accuracy since it takes into account the chance of the result being the result of chance [Vieira et al. 2010]. Sensitivity is the ability of a model to identify positive cases, that is, the percentage of successful-videos correctly classified in the model among all successful-videos in the dataset. While the Positive Predictive Value measures how many true positives are actually positive, that is, how many of the videos are classified as successful-videos. Landis and Koch proposed a classification of strength of agreement for certain metric value ranges [Landis and Koch 1977], as shown in Table 2. We consider models that have resulted in Kappa with moderate strength as agreement or more, that is $Kappa >= 0.41$, and sensitivity and positive predictive value of at least 0.5. Based on these metrics, 21 models were selected, shown in Table 3.

Table 2. Agreement strength classification for Cohen's Kappa Coefficient proposed by Landis and Koch in [Landis and Koch 1977].

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

As we can see in the results presented in Table 3, the milestone that obtained the best results was 100,000 views. Only three of the 21 selected models generated positive results for other milestones 750,000 and 1 million. One of the possible causes of this better performance of the 100,000 view milestone is related to the balance of the sample since the other milestones are more unbalanced in the amount of successful-videos. The model performed better in the three metrics, highlighted in bold in Table 3, uses the 100,000 views framework and combines the visual features **Video Characteristics + Rigidity**. Video features include duration information, frames per second, and frame dimensions, and the rigidity features provide information about the average video speed, clutter metric, and video rigidity, as defined in Section 3.2. The hyperparameters that generated the best model are $Sigma = 0.04$ and $C = 2$, obtaining Kappa of 0.7324, sensitivity of 0.8930, and positive predictive value of 0.8930.

In [Trzciński and Rokita 2017], the authors describe that when they use only visual features, they obtained better results in the complete model. The features that most

Table 3. Results and configurations of the 21 best models among the 105 models described in Section 4.3, selected based on the Kappa, Sensitivity, and Positive Predictive metrics. The model with the best overall rating is highlighted in bold.

Features	Views	Kappa	Sensitivity	Pos Pred Value	Sigma	C
V	*100k*	0.7288	0.8899	0.8926	0.08	0.5
D	*100k*	0.4317	0.9174	0.7282	0.5	8
R	*100k*	0.4266	0.8746	0.7371	0.5	2
VC	*100k*	0.6892	0.8838	0.8705	0.03	8
VF	*750k*	0.5345	0.5000	0.7576	0.25	5
	100k	0.7144	0.8777	0.8913	0.25	5
VT	*1m*	0.5165	0.5422	0.6338	0.5	8
	750k	0.5386	0.5100	0.7500	0.5	5
	100k	0.7292	0.8869	0.8951	0.5	5
VD	*100k*	0.6856	0.8807	0.8701	0.5	8
VR	***100k***	**0.7324**	**0.8930**	**0.8930**	**0.04**	**2**
VDC	*100k*	0.6503	0.8716	0.8533	0.03	8
VDF	*100k*	0.7126	0.8899	0.8818	0.25	5
VDT	*100k*	0.6978	0.8807	0.8780	0.25	8
VDR	*100k*	0.7009	0.8869	0.8761	0.08	8
VDRC	*100k*	0.6508	0.8685	0.8554	0.02	8
VDRF	*100k*	0.7140	0.8807	0.8889	0.06	8
VDRT	*100k*	0.6987	0.8746	0.8827	0.25	5
VDRTC	*100k*	0.6877	0.8930	0.8639	0.01	8
VDRTF	*100k*	0.7131	0.8869	0.8841	0.04	8
Complete model	*100k*	0.7009	0.8869	0.8761	0.02	5

contributed to the performance of their model were Deep features, Clutter and Scene Dynamics but Deep features were not present in the available dataset, used also in the present work. Therefore, this is the reason why we could not re-implement their method from scratch because we do not have all available data. Even though, our technique and their method [Trzciński and Rokita 2017] are different approaches of the same technique (regression and classification), which makes it difficult to compare the results. Anyway, we believe that we obtained better results once our complete model reached Kappa of 0.7324, sensitivity of 0.8930, and positive predictive value of 0.8930, while their complete model reached 0.23 in Spearman correlation.

6. Final considerations

In this work we proposed a new way of approaching the problem of prediction of popularity of online videos proposed in [Trzciński and Rokita 2017], using the same technique and part of the same dataset, but treating the problem as a classification problem and using only visual features as predictors. Using Support Vector Machine with Gaussian Radial Basis Function we predict which of the 1,800 videos published on Facebook had more than a certain number of views seven days after their publication based solely on visual features so that such an analysis can be produced before publication. Our predictive model performed better when using the features video characteristics and rigidity features, obtaining Kappa of 0.7324, sensitivity of 0.8930, and positive predictive value of 0.8930. As future work, we suggest testing resources that bring other information about the image, such as brightness and saturation. It would also be interesting to compare the results with videos of other types, such as advertisements and video lessons.

References

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press.

Facebook for Business (2020). In-stream ads. Online; accessed 06-Apr-2020.

Goodall, C. R. (1993). 13 computation using the qr decomposition. In *Computational Statistics*, volume 9 of *Handbook of Statistics*, pages 467–508. Elsevier.

Khosla, A., Das Sarma, A., and Hamid, R. (2014). What makes an image popular? In *Proceedings of the 23rd international conference on World wide web*, pages 867–876. ACM.

Kong, Q., Rizoiu, M.-A., Wu, S., and Xie, L. (2018). Will this video go viral: Explaining and predicting the popularity of youtube videos. In *Companion Proceedings of the The Web Conference 2018*, pages 175–178. International World Wide Web Conferences Steering Committee.

Kuhn, M. et al. (2008). Building predictive models in r using the caret package. *Journal of statistical software*, 28(5):1–26.

Kuhn, M. et al. (2019). *caret: Classification and Regression Training*. R package version 6.0-84.

Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling*, volume 26. Springer.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Trzciński, T. and Rokita, P. (2017). Predicting popularity of online videos using support vector regression. *IEEE Transactions on Multimedia*, 19(11):2561–2570.

Vieira, S. M., Kaymak, U., and Sousa, J. M. C. (2010). Cohen's kappa coefficient as a performance measure for feature selection. In *International Conference on Fuzzy Systems*, pages 1–8.

Youtube About (2020). Press. Online; accessed 06-Apr-2020.

YouTube Support (2019). How to earn money on youtube. Online; accessed 06-Apr-2020.