SindhUniv. Res. Jour. (Sci. Ser.) Vol.48 (2) 445-448 (2016)

SINDHUNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Modeling and Analysis of Search Interests for Viral Videos in YouTube: A Case Study

A. WAQAS, A. R. GILAL*, M. MEMON, H.A. M. MALIK, M. KARBASI, B. R. MEMON**

Department of Computer Science, Kulliyah of Information and Communication Technology, International Islamic University Malaysia

Received 8thJune2015 and Revised 5th February 2016

Abstract: A video that becomes popular by means of Internet sharing is called viral video. YouTube plays an important role for providing a platform to host and share videos. The search interest of users varies from time to time for such viral videos. The number of views approach to a peak level in short time span after the posting of videos and then continues to decline. In this paper, we have applied three different strategies 1) quadratic function, 2) exponential function, and 3) piecewise function to model the relative search interest for viral videos over time.

Keywords: popularity prediction, view model, viral videos, YouTube, search interests

1. INTRODUCTION

YouTube was launched in 2005 by Chad Hurley, Steve Chen and Jawed Karim and became the most popular medium over Internet for providing a new generation of short video sharing service (Cheng,*et, al* 2008), (Burgess, 2013). YouTube hosted many viral videos in its short lifetime(Jiang, *et, al* 2014). "A viral video is a video that becomes popular through a viral process of Internet sharing, typically through video sharing websites, social media and email" (Wikipedia, 2015). For instance, a viral video in YouTube named "Gangam Style" was posted by Psyon July 15, 2012 and it has greatest sensation becauseit has the record for acquiringhighest views so far. By the January 19th, 2015, this video got 2,213,446,637 views. Moreover, the music video named "Baby" posted by a user Justin Bieber was posted on February 19th, 2010 and had 1,128,987,649 numbers of views till January 19th, 2015. Another sensation, comparatively smaller, for the music video "The Fox" posted by Ylvis on September 3, 2013 and it had the 483,443,674 number of views by the end of January 19, 2015. Despite the varying degrees of popularity, we witnessed that the videos seem to follow a similar trend. Many approaches are used to model and detect the viral videos over internet and also for spreading phenomena. For instance, an agent based model for viral video diffusion in social media is presented by (Kvasnička, 2015), (Kvasni, 2014)Tom Broxton et al. discusses about strategies for catching a viral video (Broxton, *et, al* 2013). Pinto et al. presented the approaches to predict the popularity of YouTube videos by using early view patterns (Pinto, *et, al* 2013).

In this paper, we have selected three viral videos "Gangam Style", "Baby" and "The Fox" on YouTubeas a case study for this study and recorded the relevant quantitative data. Furthermore, we have modeled the popularity of these videos over time according to the search interests.

2. DATA COLLECTION AND ANALYSIS

The data for this study is collected using Google Trends(Choi, 2012),(Carneiro, 2009),(GoogleTrends, 2015)as it computes interest in a search query using an algorithm by identifying the frequency of searched term. Though, the collected data is relative to the total global search volume and it does not express absolute search volume. "Absolute search volume is the number of times a query was searched within a month". Whereas, relative data means that each data point is divided by the highest point and multiplied by 100 (Yagi, 2014). Therefore, the collected date does not represent the number of views but the proportion of views relative to the total number of searches in general. The relative search volume (φ) is found bythe following equation (1)

$$\boldsymbol{\varphi = \frac{\alpha_t}{\gamma} \times 100} \qquad \textbf{(1)}$$

Where, α_t represents the absolute search volume at a given time (t) and γ is the maximum absolute search volume.

Table 1 gives the data recorded from Google Trends for these three movies over the first fourteen months after posting and **(Fig.1**) portrays the trend for relative search interests (RSI).

++Corresponding author:Department of Computer Science, Sukkur IBA Pakistan. Email: ahmad.waqas@iba-suk.edu.pk

* Department of Computer and Information Sciences, University Technology PETRONAS, Malaysia

**Institute of Information and Communication Technology, University of Sindh, Jamshoro, Pakistan

Table 1: Relative Search Interest over the First 14 Months after Posting

Month	Relative Search Interest (RSI)		
	Gang am Style	Baby	The Fox
0	0	0	0
1	18	13	43
2	57	44	100
3	95	85	77
4	100	100	58
5	89	87	33
6	64	75	23
7	47	58	16
8	39	49	13
9	32	42	11
10	25	35	10
11	20	29	8
12	16	29	7
13	13	27	7
14	10	30	7

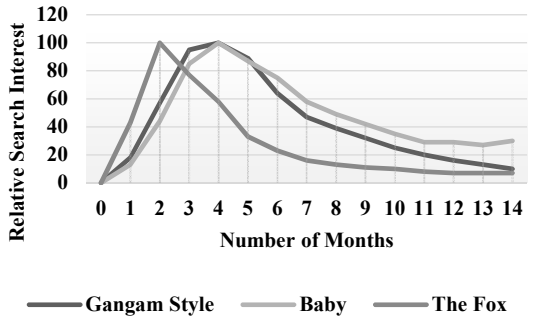


Figure 1: Trend for RSI for the First 14 Months since posting

3. MODELLING THE SEARCH INTERESTS FOR VIRAL VIDEOS

1.1. Modeling with Quadratics

Modeling selected dataset using a quadratic function allows to analyze the first section of data, following the peak. In order to create this quadratic, we used the peak data point (where relative search interest = 100) and the two data points from both side of peak. For instance, to define a quadratic model for the "Gangam Style", we used the peak data point, data pointat 4 months after the video was posted (2 months before and 2 months after the peak). Therefore, the domain of the function is $\{x|2 \leq x \leq 6\}$. With this specified data, we modeled it as quadratic function. **(Fig 2)** represents the quadratic function model of relative search interest for (a) "Gang am Style", (b) "Baby" and (c) "The Fox". The data points used to model the curve is represented by darker points on graph.

Table 2 shows the quadratic function for each video model. We have determined the coefficient of determination (R^2) that gives the percentage of variance. The closer the R^2 value to 1, the stronger the correlation between data and the function.

The quadratic function model for the viral video "Gangam Style" shows the highest correlation since the$R^2$is closer to 1whereas, "The Fox" has the least correlation. One reason for this is the fact that popularity of "The Fox" got a quicker increase in the relative search interest to reach up to 100 only after 2 months of posting. Nevertheless, correlation coefficient with values 0.9186 is also high which means that a quadratic is an appropriate function for the data up to the peak point. Whereas, "Gangam Style" and "Baby" reached 100 after 4 months.

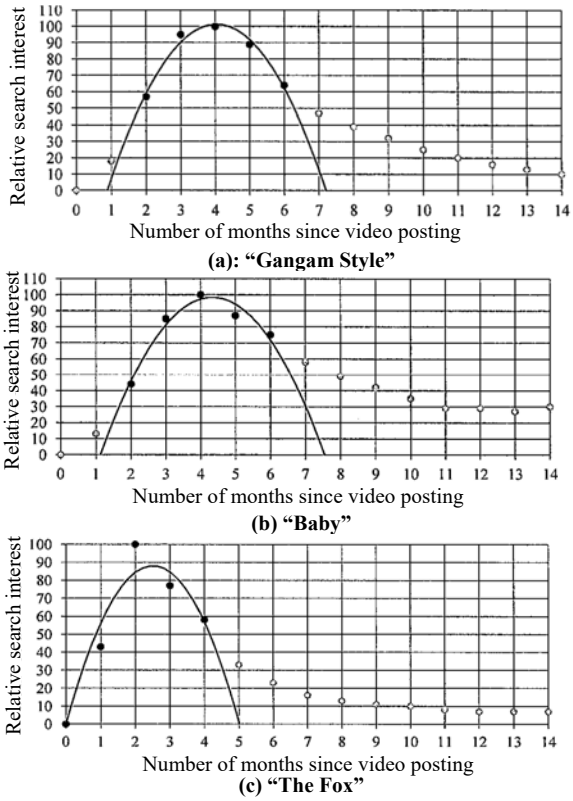


Figure 2: Quadratic Model for Relative Search Interest

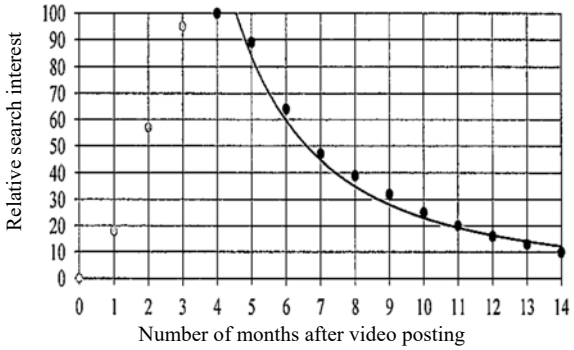
Table 2: Quadratic Function Model and R^2Values

Viral Video	R^2	Quadratic Function Model	
Gangam Style	0.9735	$y = -10.143x^2 + 81.943x - 64.2$	(2)
Baby	0.9513	$y = -9.571x^2 + 82.971x - 81.4$	(3)
The Fox	0.9186	$y = -14x^2 + 70.2x$	(4)

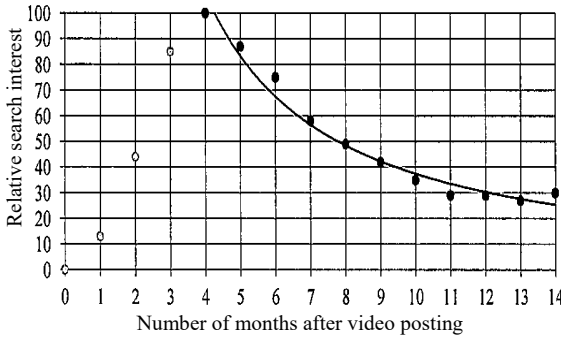
1.2. Exponential Curve

Modeling the above dataset using exponential function allows to analyze the data after peak. For developingexponential curve, we selected data starting from peak value to data point 14 months after. For "Gangam Style", we have utilized peak value that occurred at 4 months later when video was posted, and rest of the data points afterwards to 14 months. Hence, the domain for this function is$\{x|4 \leq x \leq 14\}$.**Fig. 3** represents the exponential function model of RSI for (a) "Gangam Style", (b) "Baby" and (c) "The Fox". The

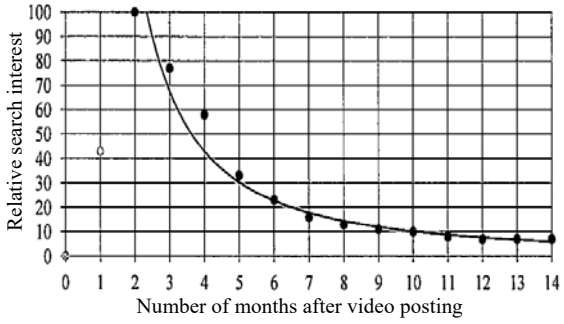
data points used to model the curve arepresentedas darker points.



(a) "Gangam Style"



(b) "Baby"



(c) "The Fox"

Fig.3: Exponential Model for Relative Search Interest

Table 3 shows the exponential function for each video model. We have determined the coefficient of determination (R^2).

Table 3: Exponential Function Model and R^2Values

Viral Video	R^2	Exponential Function Model
Gangam Style	0.9737	$y = 1699.9x^{-1.869}$ (4)
Baby	0.9669	$y = 533.53x^{-1.154}$ (5)
The Fox	0.9773	$y = 377.79x^{-1.573}$ (6)

The exponential models containhigh R^2 values which is the evidence for stronger correlation. Furthermore, it exhibits the decrease in relative search interest after peak that depicts the similar trend.

1.3. Piecewise Function

It is factual that representation of the above dataset is different in two domains, we developed a piecewise function for better representation and modelling. We split the data into two distinct domains, the first represents data before peak and the other represents data after peak. For example, for the video "Gangam Style", the piecewise model function is given by following equation (7):

$$g(x) = \begin{cases} -10.143x^2 + 81.943x - 64.2, & and 0 \le x \le 4 \\ 1699.9x^{-1.869}, & and 4 \le x \le 14 \end{cases} \quad (7)$$

This implies that quadratic function is used to represent data until peak arrives (where$x = 4$), and exponential function is used to represent data after peak.

$$\mathbf{b(x)} = \begin{cases} -9.5714x^2 + 82.971x - 81.4, & and 0 \le x \le 4 \\ 533.53x^{-1.154}, & and 4 \le x \le 14 \end{cases} \quad (8)$$

$$\mathbf{f(x)} = \begin{cases} -14x^2 + 70.2x, & and 0 \le x \le 2 \\ 379.79x^{-1.573}, & and 2 \le x \le 14 \end{cases} \quad (9)$$

Equation (7), (8) and (9) are resultant functions for representation of trends of these three viral videos. However, in order to find a final universal function, we need to observe more movies and dataset.

4. A GENERALIZED MODEL FOR SEARCH INTERESTS FOR VIRAL VIDEOS

The above described function in (7), (8) and (9) models the dataset accurately for their respective video. But, they hold trivial difference for x-values at their peaks. Therefore, to be more universal, we have collected data for 8 viral videos limiting with posting date between 2007 and 2014 as shown in table 4.We, then calculated the average for every month after posting with the help of following equation (10):

$$\textbf{Average} = \frac{V_1+V_2+V_3+V_4+V_5+V_6+V_7+V_8}{8} \quad (10)$$

We then computed the relative average search interest in the video with respect to time by using equation (1). Figure 4 depicts the relative average search interests using a quadratic and exponential function and can be represented by the piecewise function.

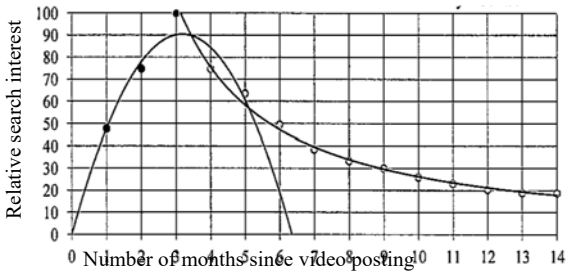


Fig.4: Relative Average Search Interest for Eight Viral Videos

Table 4: Relative Search Interest over the First 14 Months after Posting

Month	Relative Search Interest								Relative Average Search Interest
	Gangam Style	Baby	The Fox	Sneezing Panda	Chocolate Rain	Harlem Shake	How Animals Eat their Food	We Can't Stop	
0	0	0	0	0	0	0	0	0	0
1	18	13	43	27	27	68	100	2	48
2	57	44	100	28	32	100	55	51	75
3	95	85	77	100	100	33	33	100	100
4	100	100	58	48	45	14	20	81	75
5	89	87	33	44	49	8	17	70	64
6	64	75	23	36	38	6	12	53	50
7	47	58	16	38	30	4	10	39	38
8	39	49	13	34	26	4	9	34	33
9	32	42	11	31	29	3	11	28	30
10	25	35	10	32	25	3	9	23	26
11	20	29	8	37	22	3	6	19	23
12	16	29	7	33	18	2	6	16	20
13	13	27	7	31	18	2	6	14	19
14	10	30	7	33	17	2	5	13	19

The model functions used to figure out the relative average search interest as shown in Fig. 4 are $\mathbf{y} = -\mathbf{8.9724x^2} + \mathbf{56.956x}$ and$\mathbf{y} = \mathbf{372.9x^{-1.153}}$. Therefore, the generalized model function for the relative average search interest ($h(x)$) is given as (11):

$$\boldsymbol{h(x)} = \begin{cases} \mathbf{-8.9724x^2 + 56.956x}, \ and 0 \leq \boldsymbol{x} \leq \mathbf{3} \\ \mathbf{372.9x^{-1.153}}, \ and 3 \leq \boldsymbol{x} \leq \mathbf{14} \end{cases} \quad \textbf{(11)}$$

5. CONCLUSION

In this study, we analyzed various viral videoswith different popularities posted during different time spans. Every single viral video got a peak view for popularity, no matters the peak endured fora month or more, but the peak in views is always present and very apparent. We analyzed three specific videos and modeled the relative search interest of these viral videos using quadratic, exponential and piecewise functions. Furthermore, we analyzed eight different videos to generalize this model given in equation (11). It is concluded that exponential model function is more appropriate than the quadratic for this type of dataset.

REFERENCES:

Broxton, T., Y. Interian, J. Vaver, and M. Wattenhofer, (2013) "Catching a viral video," J. Intell. Inf. Syst., vol. 40, No. 2, 241–259.

Burgess, J., G. Joshua, (2013) "YouTube, Online Video and Participatory Culture," John Wiley and Sons. 38–57.

Carneiro,H. A., E. Mylonakis, (2009) "Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks," Clin. Infect. Dis., Vol. 49, No. 10, 1557–1564.

Cheng, X., Dale, C., and Liu, J. (2008, June). Statistics and social network of youtube videos. In Quality of Service, 2008. IWQoS 2008. 16th International Workshop on 229-238. IEEE.

Choi, H., H. Varian, (2012) "Predicting the Present with Google Trends,"Econ. Rec., Vol.88, No. SUPPL.1, 2–9.

Google, "Google Trends," (2015). [Online]. Available: https://www.google.com/trends/.(Date Accessed: Jan, 18, 2015)

Jiang, L., Miao, Y., Yang, Y., Lan, Z., and Hauptmann, A. G. (2014, April). Viral video style: a closer look at viral videos on YouTube. In Proceedings of International Conference on Multimedia Retrieval 193Pp. ACM.

Kvasnič, M., (2014) "Search for the optimal strategy to spread a viral video : An agent-based model optimized with genetic algorithms," in 32nd International Conference on Mathematical Methods in Economics 2, Vol. 548553.

Kvasnička, M., (2014) "Viral Video Diffusion in a Fixed Social Network: An Agent-based Model," Procedia Econ. Financ., Vol. 12, No. 334–342.

Pinto, H., J. M. Almeida, M. A. Gonçalves, (2013) "Using early view patterns to predict the popularity of youtube videos," Proc. sixth ACM Int. Conf. Web search data Min.- WSDM '13, No. 365.

Wikipedia, (2015) "Viral video." [Online]. Available: https://en.wikipedia.org/wiki/Viral_video. (Date Accessed: Dec 28, 2014)

Yagi, M., (2014) "How do you interpret Google Trends's search volume index?,". [Online]. Available: https://www.quora.com/How-do-you-interpret-Google-Trendss-search-volume-index.(Date Accessed: Jan, 18, 2015)