International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:6, No:12, 2012

Parameter Estimation for Viewing Rank Distribution of Video-on-Demand

Hyoup-Sang Yoon

Abstract—Video-on-demand (VOD) is designed by using content delivery networks (CDN) to minimize the overall operational cost and to maximize scalability. Estimation of the viewing pattern (i.e., the relationship between the number of viewings and the ranking of VOD contents) plays an important role in minimizing the total operational cost and maximizing the performance of the VOD systems. In this paper, we have analyzed a large body of commercial VOD viewing data and found that the viewing rank distribution fits well with the parabolic fractal distribution. The weighted linear model fitting function is used to estimate the parameters (coefficients) of the parabolic fractal distribution. This paper presents an analytical basis for designing an optimal hierarchical VOD contents distribution system in terms of its cost and performance.

Keywords—VOD, CDN, parabolic fractal distribution, viewing rank, weighted linear model fitting

I. Introduction

THE video-on-demand (VOD) service is enabled by the development of broadband communication technology and large-capacity storage units as well as advances in contents delivery network (CDN). CDN was developed to distribute multimedia content to the multiple users with minimizing network delay and bandwidth usage [1].

Gelman et al. [2] and Sincoskie [3] proposed a VOD service architecture using CDN technology. Since the VOD service requires investment in infrastructure to support the large number of subscribers and contents, a lot of research on VOD service delivery architectures have been conducted to reduce investment and service operation costs [4], [5]. In particular, various research have been conducted on the replica replacement of CDN servers [6], [7] and the contents allocation in each CDN server [5], [8], respectively. The main goals of such problems are to increase service quality and to reduce delivery cost. In order to achieve the goals, several factors need to be considered such as the network structure, contents storage server cost, server installation cost, transmission cost, and contents popularity [1]. Among these factors, content popularity (viewing rank) is one of the most important to design the CDN architecture for VOD service.

In order to manage the COD contents effectively, Günther et al. [9] deployed Zipf's law in approximating the viewing rank probability distribution, while Adamin and Huberman [10] proved that various Internet phenomena activities (e.g., the number of site access) can be modeled using Zipf's law. Guo et al. [11] investigated the viewing ranks of various multimedia services used in colleges and small organizations as well as YouTube to estimate the viewing rank distribution function.

H. S. Yoon is with the Department of Management Information Systems, Catholic University of Daegu, Rep. of Korea (e-mail: hsyoon@cu.ac.kr).

According to Guo et al., [11], the function of the viewing rank of VOD are more approximated by the stretched exponential function or parabolic fractal function than Zipf's law.

To the best of author's knowledge, research on the viewing rank distribution of the large-scale commercial VOD services are hard to find. The objective of this paper is to collect and analyze the VOD viewing history data of IPTV operators and to calculate the viewing rank function using the parabolic fractal function proposed by Laherrëre and Deheuvels [12]. In addition, our purpose is to confirm that the viewing rank function of large-scale commercial VOD services does not conform to Zipf's rule and to calculate the accurate viewing rank distribution function in order to help effectively distribute and manage VOD contents.

Section II deal with the distribution functions like Zipf's law, stretched exponential, and parabolic fractal. In Section III, the viewing rank distribution function is calculated using the parabolic fractal function, and its statistical significance is discussed. Finally, Section IV presents conclusions with the contributions of this paper and further research.

II. Viewing Rank Function Distribution

Previous research works of CDN for managing Internet multimedia files generally assumed that the popularity of a file conformed to Zipf's law [13]. Zipf's law has similar characteristics to the power law (or Pareto) distribution proposed to describe the phenomenon that the probability of appearing a certain value is approximated by inverse proportionality to the power of the value [14]. Such distributions are used to describe the appearance of many objects having small value and a small number of objects having large values. In other words, it can describe the phenomenon whereby most contents are rarely viewed and only a few contents are widely viewed.

In VOD systems, new contents are regularly updated, but once received and serviced, the content is not updated (fetch-at-most-once), so a viewer is not likely to repeatedly view the same content. On the other hand, web pages are repeatedly updated after the launch of the service (fetch-repeatedly), therefore the users generally keep visiting [15]. Such a difference makes that the relationship between the web page visiting rank and the viewing count conforms to Zipf's law, while the relationship between the VOD viewing rank and the viewing count does not.

Guo et al. [11] shows that Zipf's law cannot resemble the relationship between the actual multimedia content viewing rank and the number of viewing. Its evidence is the curvature in the Zipf of Zipf-like distribution function of the viewing rank to the number of viewing in log-log scale. Consequently, it is proposed to use the stretched exponential function or parabolic fractal function as the distribution function. The stretched exponential function was proposed to explain the 'fat tails' phenomenon and meant that the power of the viewing count and log value of the viewing can be expressed by a linear relation ($x_i^c \sim \log(i)$) [16].

The parabolic fractal function adds a secondary term to Zipf's function. Equation (1) shows the parabolic fractal funtion, where i is the viewing rank of a content, x_i is the number of viewing, and the parameters c_0, c_1, and c_2 are constants.

$$\log(x_i) = c_0 + c_1 \log(i) + c_2 \log(i)^2 \tag{1}$$

In this paper, we analyze the viewing log of the commercial VOD service and to estimate the relationship between the viewing count and the viewing rank by using the parabolic fractal function. As a result, we estimate the parameters c_0, c_1, and c_2.

III. Parameter Estimation and Analysis

A. Data Collection

The viewing data used in this paper were collected from one month's viewing history of all subscribers to the commercial IPTV VOD service. The collected data were segmented into four weekly data, and the viewing count of each content was summed and its rank calculated from each segmented data.

A one-week segmentation interval was used since a change in the viewing count was observed weekly because of viewer's life cycle and because terrestrial broadcasting contents, which were the most popular, were updated weekly. The measured viewing data including the title of the contents, genre, viewing count, and viewing data.

B. Weighted Linear Regression Model

In this paper, we used the LinearModel.fit function built into the Statistics Toolbox of MATLAB to calculate the parameters (c_1, c_2, and c_3) of the parabolic fractal function. The linear regression model used in this paper is as follows.

$$\bar{y} = \beta_2 r^2 + \beta_1 r + \beta_0 \tag{2}$$

In (2), y and r denote $\log(x_i)$ and $\log(i)$, respectively, and β_0, β_1, and β_2 are estimates of parameters c_0, c_1, and c_2, respectively. Since more precise estimation to the more popular content is expected to be better, this paper deployed the weighted linear model fit instead of the normal linear model fit. To assign a higher weight to the higher viewing rank when calculating the sum of square error, the viewing count of i-th content was used as the weight (w_i) of the sum of square error of i-th content as follows.

$$SS = \sum_{i=1}^{N} w_i \left(\hat{y}_i - y_i\right)^2 \tag{3}$$

Using the weighted linear model fit, the weekly data showing the relationship between the VOD viewing count and rank were analyzed. Fig. 1 shows the result of the Week 1 data. The data expressed in dots represent the viewing rank (horizontal axis) and the viewing count (vertical axis) of the contents, while the solid line is the parabolic fractal function fitted using the weighted linear model.

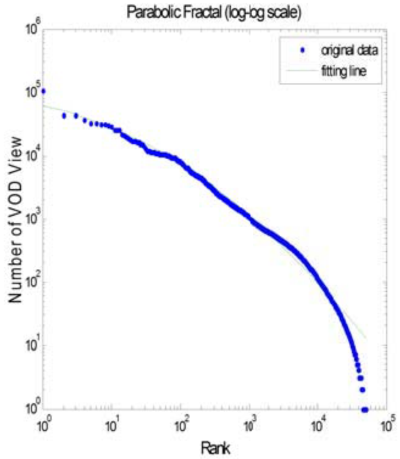

Fig. 1. Weighted Linear Model Fitting of Week 1 Data

TABLE I
Coefficient Prediction of Week 1 Data

	Estimate	SE	tStat	p
β_0	11.0070	0.0054	2045.00	0
β_1	-0.2108	0.0019	-111.94	0
β_2	-0.0525	0.0002	-336.15	0

C. Result of Analysis

As an analysis result, we can see the solid line (see Fig. 1) is well the the value of R^2 is calculated as 0.986. Table 1 shows the statistical significance of the predicted value of each coefficient of the linear model. Since the values are very small, they can be considered as statistically significant.

Figs. 2–4 show the results of the analysis of the VOD viewing data for weeks 2–4, respectively.

The same analysis as shown in Table 1 was performed on each data for weeks 2–4. Therefore, we obtained the results that each data is statistically significant. Table 2 shows the parameters of parabolic fractal distribution function calculated by the weighted linear model fitting. It indicates that the parameters of each week were similar to that of each other week. The average of each parameter is presented in the last row of Table 2.

The parabolic fractal distribution function using the average parameters shown in Table 2 is expressed as follows.

$$\log\left(x_i\right) = 11.186 - 0.247 \log(i) - 0.047 \log(i)^2 \tag{4}$$

To calculate the viewing rank and viewing probability ratio (i.e., the ratio of viewing count of a content to total viewing counts), (4) is converted as follows.

$$x_i = \exp\left[11.186 - 0.247 \log(i) - 0.047 \log(i)^2\right] \tag{5}$$

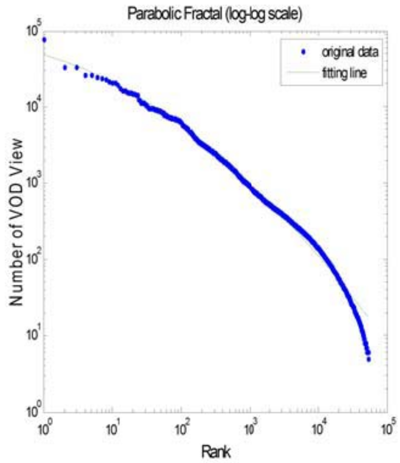

Fig. 2. Weighted Linear Model Fitting of Week 2 Data

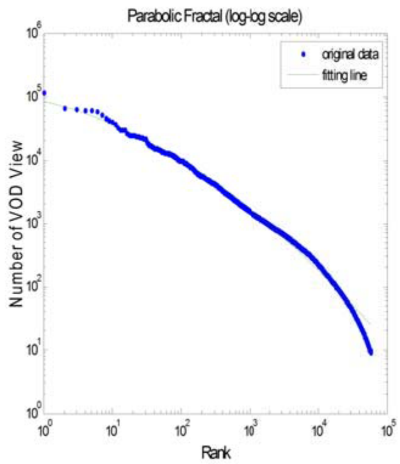

Fig. 3. Weighted Linear Model Fitting of Week 3 Data

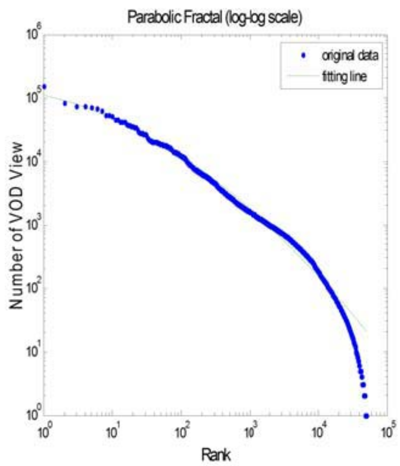

Fig. 4. Weighted Linear Model Fitting of Week 4 Data

Since the viewing count of the content whose viewing rank is i is x_i and the total viewing count is $\sum x_i$, the viewing probability can be calculated as follows.

TABLE II
LINEAR MODEL COEFFICIENTS AND R^2

Week	c_0	c_1	c_2	R^2
1	11.007	−0.211	−0.052	0.986
2	10.787	−0.267	−0.042	0.992
3	11.331	−0.257	−0.044	0.992
4	11.618	−0.253	−0.050	0.987
AVG	11.186	−0.247	−0.047	–

$$p_i = \frac{\exp\left[11.186 - 0.247\log(i) - 0.047\log(i)^2\right]}{\sum_{i=1}^{N}\exp\left[11.186 - 0.247\log(i) - 0.047\log(i)^2\right]} \quad (6)$$

IV. CONCLUSION

This paper analyzed the relation between the viewing rank and the viewing count (or probability) of large commercial VOD services and used the parabolic fractal function to calculate the viewing rank distribution function. As the number of subscribers using VOD services grows rapidly and the demand for large-capacity storage systems to support high-resolution contents services increases, accurate calculation of the viewing rank distribution function has become very important to reduce the operational costs and to improve performance of VOD service systems.

As a result, this research can be used in the allocation of VOD contents to reduce costs and to improve the performance of service systems and user services. Although Thouin and Coates [15] proposed the P2P (peer-to-peer) VOD service structure to resolve the cost and scalability problem of hierarchical VOD services, the conventional VOD service structure is expected to be used until various copyright and service quality issues are cleared.

As a further research, we will develop a new cost model of VOD service operation and solve the contents allocation problem using the contents viewing distribution function proposed in this paper.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012-0003476)

REFERENCES

[1] F. Thouin and M. Coates, "A review of the content delivery network," McGill University, Tech. Rep., 2005.
[2] A. D. Gelman, H. Kobrinski, L. S. Smoot, S. B. Weinstein, M. Fortier, and D. Lemay, "A store-and-forward architecture for video-on-demand service," in *Communications, IEEE International Conference on*, vol. 2, 1991, pp. 842–846. [Online]. Available: http://dx.doi.org/10.1109/ICC.1991.162477
[3] W. D. Sincoskie, "System architecture for a large scale video on demand service," *Computer Networks and ISDN Systems*, vol. 22, no. 2, pp. 155–162, Sep. 1991. [Online]. Available: http://dx.doi.org/10.1016/0169-7552(91)90007-Y

[4] D. Deloddere, W. Verbiest, and H. Verhille, "Interactive video on demand," *Communications Magazine, IEEE*, vol. 32, no. 5, pp. 82–88, May 1994. [Online]. Available: http://dx.doi.org/10.1109/35.281582
[5] Y.-D. Lin, H.-Z. Lai, and Y.-C. Lai, "A hierarchical network storage architecture for video-on-demand services," in *Local Computer Networks, 1996., Proceedings 21st IEEE Conference on*, 1996, pp. 355–364. [Online]. Available: http://dx.doi.org/10.1109/LCN.1996.558164
[6] M. Yang and Z. Fei, "A model for replica placement in content distribution networks for multimedia applications," in *Communications, 2003. ICC '03. IEEE International Conference on*, vol. 1. IEEE, May 2003, pp. 557–561 vol.1. [Online]. Available: http://dx.doi.org/10.1109/ICC.2003.1204238
[7] S. Buchholz and T. Buchholz, "Replica placement in adaptive content distribution networks," in *Proceedings of the 2004 ACM symposium on Applied computing*, ser. SAC '04. New York, NY, USA: ACM, 2004, pp. 1705–1710. [Online]. Available: http://doi.acm.org/10.1145/967900.968238
[8] F. Schaffa and J. P. Nussbaumer, "On bandwidth and storage tradeoffs in multimedia distribution networks," in *INFOCOM '95. Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Bringing Information to People. Proceedings. IEEE*. IEEE, Apr. 1995, pp. 1020–1026 vol.3. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.1995.515978
[9] R. Günther, L. Levitin, B. Schapiro, and P. Wagner, "Zipf 's law and the effect of ranking on probability distributions," *International Journal of Theoretical Physics*, vol. 35, no. 2, pp. 395–417, Feb. 1996. [Online]. Available: http://dx.doi.org/10.1007/BF02083823
[10] L. A. Adamin and B. A. Huberman, "Zipf's law and the internet," *Glottometrics*, vol. 3, pp. 143–150, 2002.
[11] L. Guo, E. Tan, S. Chen, Z. Xiao, and X. Zhang, "The stretched exponential distribution of internet media access patterns," in *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing*, ser. PODC '08. New York, NY, USA: ACM, 2008, pp. 283–294. [Online]. Available: http://dx.doi.org/10.1145/1400751.1400789
[12] J. Laherrére and P. Deheuvels, "Parabolic fractal distributions in nature," *Comptes rendus de l'Académie des sciences, Série II a: Sciences de la Terre et des Planétes*, vol. 322, no. 7, pp. 535–541, 1996.
[13] J. Ni, D. H. K. Tsang, I. S. H. Yeung, and X. Hei, "Hierarchical content routing in large-scale multimedia content delivery network," in *Communications, 2003. ICC '03. IEEE International Conference on*, vol. 2. IEEE, May 2003, pp. 854–859 vol.2. [Online]. Available: http://dx.doi.org/10.1109/ICC.2003.1204454
[14] M. E. J. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, Sep. 2005. [Online]. Available: http://dx.doi.org/10.1080/00107510500052444
[15] F. Thouin and M. Coates, "Video-on-Demand networks: Design approaches and future challenges," *IEEE Network*, vol. 21, no. 2, pp. 42–48, Mar. 2007. [Online]. Available: http://dx.doi.org/10.1109/MNET.2007.334311
[16] J. Laherrère and D. Sornette, "Stretched exponential distributions in nature and economy: fat tails with characteristic scales," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 2, no. 4, pp. 525–539, Apr. 1998. [Online]. Available: http://dx.doi.org/10.1007/s100510050276

Hyoup-Sang Yoon received the BS, MS, and PhD degrees in Industrial Engineering from Korea Advanced Institute of Science and Technology, Daejeon, Korea in 1997, 2000 and 2007, respectively. He is currently an assistant professor in the Department of Management Information Systems, Catholic University of Daegu. His research interests include QoS of IPTV, design of experiment, and performance optimization of communication systems.