A New Parameter Estimation Method for a Zipf-like Distribution for Geospatial Data Access

Rui Li, Wei Feng, Hao Wang, and Huayi Wu

Many reports have shown that the access pattern for geospatial tiles follows Zipf's law and that its parameter α represents the access characteristics. However, visits to geospatial tiles have temporal and spatial popularities, and the α-value changes as they change. We construct a mathematical model to simulate the user's access behavior by studying the attributes of frequently visited tile objects to determine parameter estimation algorithms. Because the least squares (LS) method in common use cannot obtain an exact α-value and does not provide a suitable fit to data for frequently visited tiles, we present a new approach, which uses a moment method of estimation to obtain the value of α when α is close to 1. When α is further away from 1, the method uses the associated cache hit ratio for tile access and uses an LS method based on a critical cache size to estimate the value of α. The decrease in the estimation error is presented and discussed in the section on experiment results. This new method, which provides a more accurate estimate of α than earlier methods, promises more effective prediction of requests for frequently accessed tiles for better caching and load balancing.

Keywords: Cache, least squares method of estimation, moment method of estimation, modeling, geographic information system.

Manuscript received Mar. 29, 2013; revised June 19, 2013; accepted June 25, 2013.

This work was supported by the National Natural Science Foundation of China (Grant No.41071248 and No. 41371370), the National High Technology Research and Development Program of China (Grant No. 2012AA12Z401) and the LIESMARS Special Research Funding. The mathematical model in section 2 of this paper was presented at the International Conference on Computer Science and Education (ICCSE 2012). We acknowledge the recommendations from the International Conference.

Rui Li (phone: +86 13607168286, ruili@whu.edu.cn), Wei Feng (subrainbow@qq.com), Hao Wang (alexwhu@163.com), and Huayi Wu (wuhuayi@whu.deu.cn) are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, P.R., China.

I. Introduction

The relationship between humankind and Earth is changing greatly through the growing popularity of virtual geographic environments, such as Digital Earth, 3D Digital City, and virtual communities [1]. Human beings now interact with these systems on a regular basis [2], [3]. Many researchers believe that studying the behavior of users of these systems can provide critical cues for system maintainers and designers [4]-[7]. Using historical log files to gather statistics and analyzing system running behavior and the user's interaction modes to design better methods for system maintenance can achieve an excellent match with running instances of actual systems and improve system performance and features. This approach has already been demonstrated by some studies, such as those of Fang and others [8] and Krumm and Horvitz [9].

However, it has been shown that tile access is characterized by bias and repeatability and that the access pattern satisfies the "20/80 rule" from sociology. Fisher developed the first method, Hotmap, to visualize the user's behavior using log files [5]. Hotmap analyzed a number of tiles with the number of hits for each and found that the hits follow a power-law distribution. Talagala and others observed that tile popularity follows a Zipf distribution [10]. However, these studies did not go into further detail. The authors of [10] analyzed the number of hits on each tile in Digital Earth using the GlobeSIGht software, which they developed, studied the characteristics of tile access patterns, and confirmed that the geospatial tile access pattern satisfies a Zipf-like power-law distribution very well [11]. The distribution parameters were specified by rules and lines [12], [13].

Networked geographic information systems (GISs) are complex. The studies mentioned above all determined the

http://dx.doi.org/10.4218/etrij.14.0113.0293

user's access behavior from sample log files and obtained the distribution parameter α by fitting a distribution curve to this data, most of these curves being based on Zipf's law with one or two parameters and fixed α [14]. Sometimes, however, geospatial tile access has both temporal and spatial characteristics [15], [16]. If the popularity of some tiles changes, the value of the parameter α will change accordingly.

This paper studies the characteristics of the distribution parameter α mentioned above. In tile access behavior research, one of the key requirements is to determine the value α over time. However, it takes a long time to extract sample data from log files. Therefore, it is essential to investigate ways of constructing a mathematical model to simulate the user's access behavior and to estimate the value of α accurately. It is also important to study the attributes of frequently accessed tile objects to build parameter estimation algorithms. These factors motivated the problems that are solved through the research presented herein.

The rest of this paper is organized as follows. Section II introduces a Zipf-like distribution of tile access behavior and presents a mathematical model to simulate tile access, which the authors of this paper presented at the 7th International Conference on Computer Science and Education [17], and it is the basis for studying the parameter estimation methods for a Zipf-like distribution. Section III presents the moment method of estimation (MME), which is used to estimate the value of α under the condition that α is close to 1, as well as the least squares (LS) method based on a critical value of the number of tiles in the cache, which we refer to as CCLSM and which is used to estimate the value of α under the condition that α is further away from 1. Section IV contains a summary and conclusions.

II. Mathematical Simulation Model of ZIPF-like Distribution

1. Tile Access Follows Zipf-like Distribution

Log files at various points in time were obtained from the Digital Earth GlobeSIGht server. Curves were fitted to these data to obtain a value for α, revealing that the tile access pattern from a real-world GIS follows a Zipf-like distribution, as stated in (1) [13]. P_i is the frequency of the i-th most frequent tile, and C is a normalized constant:

$$P_i = C/i^{\alpha}. \quad (1)$$

To analyze online interactive systems, accurate historical log files are required, which can support visualization of the changes in the characteristics of tile visiting patterns. Unfortunately, it takes a long time to collect such highly accurate log file data. For example, all the data samples used in Hotmap were obtained by sampling tile access logs from 2006 and early 2007 [5]. Therefore, it is essential to simulate tile access behavior to study the variation of distribution characteristics and to obtain an effective estimate of the distribution parameter α.

2. Mathematical Simulation Model of Zipf-like Distribution

Existing research studies [5], [10], [13] are all based on historical log files, fitting a distribution curve to the user behavior they represent and using LS estimation to obtain an approximate value of α. This method is easy to implement but encounters limitations due to the insufficient size of the samples and historical logs, which are inadequate for present-day needs. To study tile access behavior more accurately, in the research reported herein, a mathematical simulation model (MSM) of a Zipf-like distribution (the Zipf-like MSM) is constructed, using the mathematical model to simulate requests for tile access and studying the attributes of frequently visited tile objects. In this model, supposing α is the Zipf-like distribution parameter, N the total number of tiles, and M the total number of user requests, partition the range of [0, 1] into N intervals and create a random value lying in the range (0, 1) for M times. Each random value is bound to lie in an interval. Thus, these random values follow the Zipf-like distribution function:

$$F(k) = P\{X \le k\} = \sum_{x_i \le k} P\{X = x_i\} = \sum_{i=1}^{k} (C/i^{\alpha}),$$

$$i = 1, 2, \cdots, k,$$

where $F(k)$ is the function to get the access frequency of the k-th tile. According to the value of $F(0), \ldots, F(k), \ldots, F(N)$, the size of interval k can be represented as $F(k+1) - F(k)$. The model can generate sequences of visits to each tile automatically in a short time. The process is as follows.

Algorithm 1. Generating sequence of visits.

Step 1. For N tiles, *Visits* is an array containing the number of visits to each tile, where *Visits* = $\{V_i \mid 1 \le i \le N\}$. Initialize *Visits*: For $i = 1$ to N $(V_i = 0)$.

Step 2. Because C in the Zipf-like law ($P_i = C/i^{\alpha}$) is a constant, it can be calculated as

$$C = \left(\sum_{i=1}^{N} 1/i^{\alpha}\right)^{-1}.$$

Step 3. Assuming that tile access requests follow a Zipf-like distribution, let F be an array of length N+1 which fits a Zipf-like distribution, $F = \{F[k] \mid 0 \le k \le N\}$, and is determined as follows:

If $k = 0$: $F[0] = 0$;

If $k = 1,2,\ldots, N-1$: $F[k] = P\{X \le k\} = \sum_{x_i \le k} P\{X = x_i\} = \sum_{i=1}^{k} (C/i^{\alpha})$.

If $k = N$: $F[N] = 1$.

Step 4. Create a random value, R, based on the system clock and lying in the range (0, 1) for M time periods. For each value of R, if $F[k] \le R < F[k+1]$, $Visits[k] = Visits[k] + 1$.

Each R can be regarded as a visit to a tile, and interval k can be regarded as tile k, so R lying in interval k can be regarded as a visit to tile k. For R lying in the range (0, 1), a value will be determined in $F[k]$, which fits the Zipf-like distribution. Thus, by adding up the random values in $Visits[k]$, it is possible to obtain a sequence of the number of visits to each tile. Therefore, the sequences of visits generated by the Zipf-like MSM follow a Zipf-like distribution.

3. Simulation and Comparison

Based on the results of the authors' previous work on log files [13], the input parameter values for the Zipf-like MSM can be set according to (2). Figure 1 shows the results of the model-based simulation, which are consistent with the curve presented in Wang and others' work [13]. The curve in Fig. 1 describes the phenomenon clearly: the tile access curve is made up of several subsections. When $i = 100$, there is an obvious inflexion point; for each subsection, the logarithm of tile access frequency and the logarithm of tile access rank are related by a line whose slope is negative. This satisfies the Zipf-like law very well, with the tile access exhibiting the heavy-tailed characteristic inherent in the Zipf-like law. This means that many tiles are located in low access frequency areas, whereas only a few tiles are located in high access frequency areas.

$$P_i = \begin{cases} 0.05843, & i = 1, \\ 0.00819/i^{0.32931}, & 2 \le i \le 100, \\ 0.18357/i^{1.00074}, & 101 \le i \le 6571. \end{cases} \quad (2)$$

Figure 2 presents the simulation data ordered by the number of hits on each tile, as shown in Fig. 9 of Fisher's paper on Hotmap [5]. In contrast to Fisher's work, this research is based on more than one million data points, and there is enough information to define a data distribution. In Fig. 2, the number of data points represented is less than one hundred thousand, and the data distribution is therefore sparser. However, taken together, the two figures illustrate the same phenomenon. It can therefore be concluded that the sequences of tile visits produced by the Zipf-like MSM satisfy a Zipf-like distribution and form the same graphical shape as those from sample historical logs in the reports of Fisher [5] and Wang and others [13]. Therefore, this approach to using a mathematical method

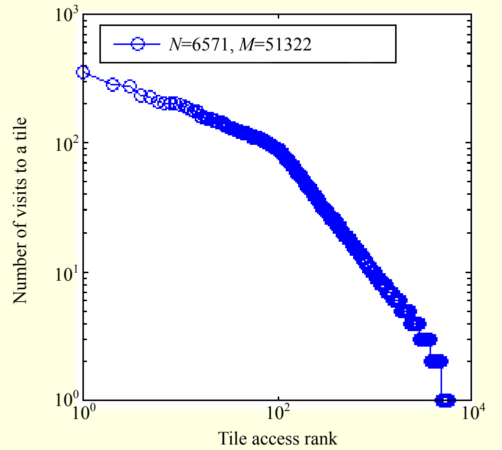


Fig. 1. Sequence for number of visits to each tile generated by Zipf-like MSM.

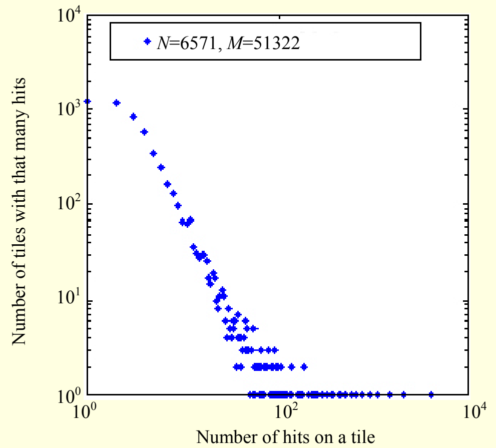


Fig. 2. Frequency charts of tile use generated by Zipf-like MSM.

to simulate the user's access behavior is credible and can be used to verify the estimation methods for the distribution parameter α in section III, since it can generate customized access patterns that can be applied to the experiments for the estimation methods in a short time rather than depending upon the time-consuming collection of data logs, which was the technique used in previous studies [5], [10], [13].

III. Estimation of Distribution Parameter α

1. Moment Method of Estimation

When using log files to study the user's behavior, determining the value of the distribution parameter α remains an intractable problem. One common method is to construct a linear regression equation using LS estimation or equivalently maximum entropy to obtain the parameter value [18], [19].

http://dx.doi.org/10.4218/etrij.14.0113.0293

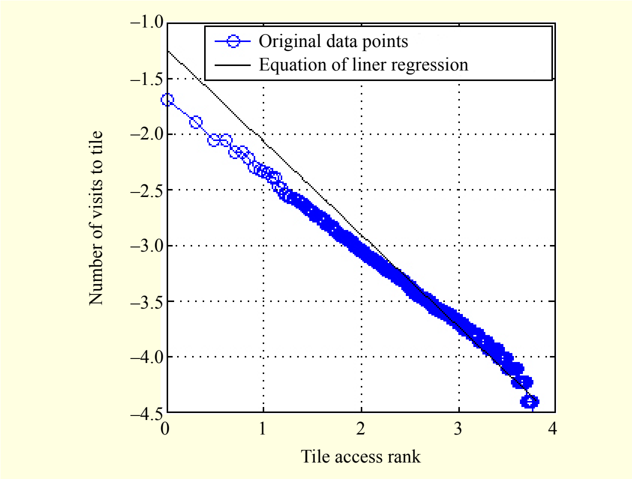


Fig. 3. LS method of estimation for distribution parameter α.

Then, the R^2 value of the fit can be obtained from (3) to determine whether the quality of the estimate is satisfactory. The closer the R^2 value is to 1, the better the estimate is.

$$R^2 = 1-(SSE\,/\,SST),$$

$$SSE = \sum_{i=s}^{t}(y_i-\overline{\overline{y}}_i)^2, \qquad (3)$$

$$SST = \sum_{i=s}^{t}(y_i-\overline{y}_i)^2 .$$

Various research has shown that the LS method of estimation cannot obtain an accurate α-value, as shown in Fig. 3. Fang and Jeong [20] demonstrated that the LS estimation is very sensitive to anomalies and proposed a novel robust probabilistic model to address the issue. If the value of α estimated using the LS method is 0.80372, the corresponding R^2 is 0.90657. The LS method concentrates on fitting all the data, but most of the data is in the low frequency access areas, so data from the high frequency access areas is not of noted interest here. As a result, the fitted line shows a sharp decrease.

Therefore, this paper proposes a new method using the MME to obtain a value for α. Assume that the sequence of tile visits produced by the Zipf-like MSM is a random variable, X, with N being the total number of tiles. From (1), the following is obtained:

$$i=(C/P_i)^{-\alpha}. \qquad (4)$$

For X, its moment of origin is

$$E(X)=\sum_{i=1}^{N}(i\times P_i). \qquad (5)$$

From (4) and (5), the following is obtained:

$$E_{\text{changed}}(X)=\sum_{i=1}^{N}P_i\times(C/P_i)^{-\alpha}. \qquad (6)$$

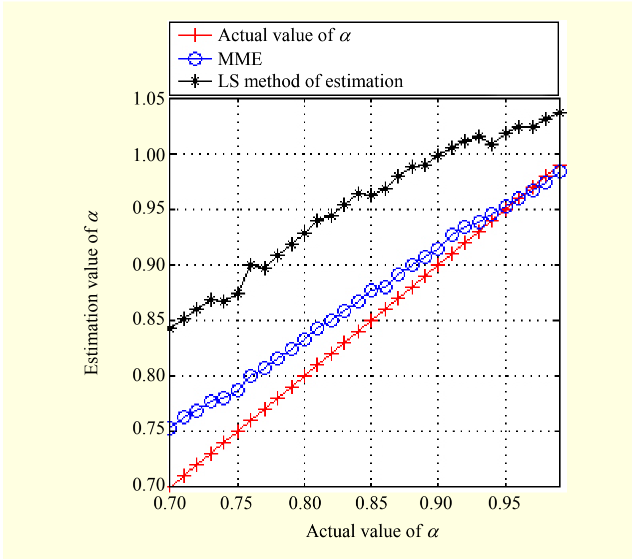


Fig. 4. Comparison of MME and LS method of estimation.

Table 1. Comparison of estimation errors.

Estimation errors (%)				
Actual value of α	0.60 - 0.69	0.70 - 0.79	0.80 - 0.89	0.90 - 0.99
LS method of estimation	19.5588 - 24.6833	16.3165 - 20.4714	11.2697 - 16.0494	4.737 - 10.8889
MME	7.5735 - 12.5833	4.4304 - 7.6143	1.9438 - 4.1625	0.0417 - 1.8462
CCLSM	2.8657 - 4.8333	2.7722 - 4.8961	2.4773 - 3.3256	
CCLSM with curve translation (offset value of α is 0.03)	0.1014 - 1.6119	0.0001 - 2.0263	0.1628 - 1.0000	

Algorithm 2. Moment method of estimation.

Step 1. Obtain $E(X)$ from (5). And sort P_i to obtain set P.

Step 2. Use the difference value minimum method to find the optimal value of α.

Step 2.1. Let S be the set of estimation errors. Compute the elements of S as follows:

for α = 0.0001: 0.9999 step 0.0001

$$E_{\text{changed}}(X)=\sum_{i=1}^{N}P_i\times(C/P_i)^{-\alpha},$$

$$j=j+1,$$

$$S_j = abs(E-E_{\text{changed}}),$$

end.

Step 2.2. Find the minimum value in set S, which will have index number j:

$$[S_{\min}, j]=\min(S).$$

Step 2.3. $\alpha(j)$ is the desired estimate.

For an actual parameter α, it is certain that there is a set S, $S=\{S_j \mid 1\le j\le 9999\}$, containing a minimum error value,

S_{min}. The smaller the step value for α, the smaller S_{min} will be. Moreover, S_{min} is unique, so the estimated value of α is also unique.

Next, the MME is compared with the LS method of estimation. The input parameter to the Zipf-like MDM is varied from 0.70 to 0.99. From Figure 4, it is obvious that the MME is more accurate than the LS method of estimation. When the value of α is closer to 1, the estimation of α by the MME is more satisfactory. However, as α moves further away from 1, the effectiveness of the fitting procedure steadily decreases. Table 1 shows a comparison of estimation errors between the MME and the LS method.

2. CCLSM

Subsection II.3 mentioned that the total number of visits to each tile exhibits high access frequency areas and low access frequency areas. For every tile in the low access frequency areas, the number of visits is generally less than one hundred, and some tiles are hardly accessed at all. Caching all tile visit requests is therefore a waste of cache resources. For tiles in high access frequency areas, requests to the tile server from the network involve substantial repetition, and this repetition can be used by the server to manage the tile access cache efficiently. The question becomes how best to use this repetitive characteristic of tile access.

On the other hand, using the LS method of estimation and a value of α for the Zipf-like distribution, the fit in the high access frequency areas is not suitable, and there is a heavy-tailed mass in the low access frequency areas. This factor has a negative effect on the study of the user's behavior when visiting highly popular tiles. One possible solution might be to give up some data in the low access frequency areas in an effort to improve the accuracy of the estimate.

In view of the situation described above, this research determines that the key approach is to find the critical value of the number of the most popular tiles, abandon those tiles with a number of visits less than the critical value, and make the curve fitted to the high access frequency areas correspond more closely with the actual values. To do this, we study the relationships among the size of the tile cache, the hit ratio, and the distribution parameter α.

Shi and others [14] proposed the following equation:

$$\begin{cases} k = N \times h^{1/(1-\alpha)}, \ \alpha < 1, \\ k \approx e^{h \times N}, \ \alpha = 1, \end{cases} \tag{7}$$

where k is the number of most popular tiles that need to be cached, namely, the critical value, h is the steady-state cache hit ratio, and N is the total number of tiles.

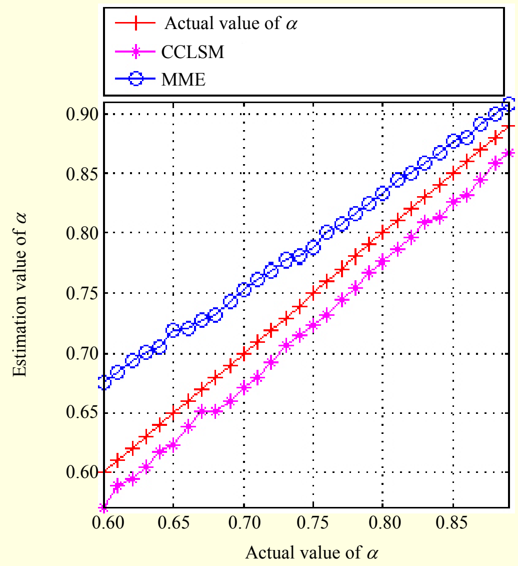


Fig. 5. Comparison of LS method of estimation based on critical value of number of cached tiles to MME.

From (7), under the assumption of a given number of cached tiles, k, and a given hit ratio, h, α can be calculated. Using an approach based on the MME described above, this paper proposes an algorithm that uses the critical value of the number of cached tiles and the LS estimation method to estimate the value of α. The algorithm can be stated as follows.

Algorithm 3: LS estimation method based on critical value of number of cached tiles.

Step 1. Input the tile access data from the Zipf-like MSM. Calculate the access probabilities P_i and sort them in ascending order to obtain the array P. Variable r is the length of P, and k is in $(1, r)$.

Step 2. From $k = N \times h^{1/(1-\alpha)}$, set bounds on α and h and determine the critical cache size, k.

Step 2.1. Boundary of α. Using the MME to obtain a primary value of α, α_{pri}, for $0 < \alpha \le 1$, set $\alpha_{min} = \max(0, \alpha_{pri} - 0.5)$ and $\alpha_{max} = \min(\alpha_{pri} + 0.5, 1)$. Iterate the α-value from α_{min} to α_{max} by an increment of 0.01. Variable u represents the number of iterations required.

Step 2.2. Boundary of h. In each loop of Step 2.1, obtain a value for α, $1 \le k \le r$, and determine h_{min} and h_{max}. For practical purposes, to ensure that the hit ratio h is not too low, set $h_{min} \ge 0.5$. For each loop in Step 2.1, h is varied from h_{min} to h_{max} by an increment of 0.01. Variable v represents the number of iterations required.

Step 2.3. For each outer loop and each inner loop, determine a value of α to calculate a unique k value.

Step 3. Using the top k tiles' P_i ($i = 1, 2, \ldots, k$) as input objects to the LS estimation algorithm, obtain an estimate of α and the coefficient of determination R^2. Calculate the matrix $R^2_{u \times v}$ and find the largest element and its index. The value of α corresponding to this index is the desired estimate.

http://dx.doi.org/10.4218/etrij.14.0113.0293

The CCLSM is compared with the MME. The input parameter to the Zipf-like MDM is varied from 0.60 to 0.89 in increments of 0.01. From Fig. 5, it is apparent that as α moves further away from 1, the estimation error of the α-value obtained using CCLSM becomes smaller than that obtained using the MME. In such cases, CCLSM therefore improves the estimation error. The fitted curve obtained with CCLSM is parallel with the curve representing the actual values. In practice, certain correction methods, such as translating the fitting curve, can be used to obtain an accurate estimate of α. Table 1 shows the comparison of estimation errors between the MME, CCLSM, and CCLSM with curve translation.

IV. Conclusion

It has been proven that a Zipf-like distribution can be used to describe the popularity of tile objects. The distribution parameter α represents the tile access characteristics. In geoscientific applications, if the popularity of some tiles changes, then the value of parameter α changes also. Research into the estimation of α-values is the foundation of managing the tile cache resource efficiently and also plays an important role in the design of networked GISs.

Using a Zipf-like law, we analyzed the effect of the distribution parameter α from two vantage points: 1) Based on log files of tile visits and mathematical statistical methods, we constructed a mathematical simulation model of a Zipf-like distribution and studied the attributes of highly popular tile objects for parameter estimation algorithms; 2) Because the LS method in common use cannot obtain an accurate value for α and cannot achieve a suitable fit for data from highly popular tiles, which are the main concerns of this research and are found in certain high frequency access areas, this paper presented a new approach, which uses the MME to obtain values for α when α is close to 1. When α is further away from 1, using the cache hit ratio for tile access requests, the proposed method uses an LS method of estimation based on a critical value of the number of cached tiles to estimate α.

Zipf's law indicates the interactive pattern and modes between users and networked GISs. Accurate estimation for distribution parameter α promises more effective prediction of access distribution and highly popular tiles. It can provide a new theoretical method to design strategies to improve the system performance of networked geographic information applications and the utilization efficiency for networking resources. For example, it can be used to establish an effective caching mechanism in a distributed caching system, such as by caching those tiles that have higher probabilities and responding to requests for these tiles by distributed cache servers to shorten response time for user requests, thus achieving high hit ratios and more effective sharing of spatial data. In addition, it can be used in load balancing strategies in networked GISs, such as by allocating more networking resources (for example, bandwidth resource and computing resource) for the requests to tiles that have higher probabilities, to prevent overconcentration of loads caused by hotspot data access and achieve efficient task distribution and data extraction. In addition, the new parameter estimation method presented in this paper can be used in other applications on the Internet, whose access pattern of web objects also flows in a Zipf-like distribution (for example, the application of webpages, an information search, and video on demand).

References

[1] J.H. Gong, "Man-Earth Relationships Based on Virtual Geographic Environments," *6th Nat. Conf. Cartography GIS Conf.*, Wuhan, Hubei, China, Oct. 30, 2006.

[2] D. Butler, "Virtual Globes: The Web-Wide World," *Nature*, vol. 439, no. 7078, Feb. 16, 2006, pp. 776-778.

[3] D.G. Bell et al., "NASA World Wind: Opensource GIS for Mission Operations," *Proc. IEEE Aerospace Conf.*, Mar. 3-10, 2007, pp. 1-9.

[4] C. Yang et al., "Performance-Improving Techniques in Web-Based GIS," *Proc. Int. J. Geograph.Inf. Sci.*, vol. 19, no. 3, 2005, pp. 319-342.

[5] D Fisher, "Hotmap: Looking at Geographic Attention," *IEEE Trans. Proc. Vis. Comput. Graph.*, vol. 13, no. 6, Nov.-Dec. 2007, pp. 1184-1191.

[6] D. Fisher, "How We Watch the City: Popularity and Online Maps," *Workshop Imaging City, ACM Comput.-Human Interaction*, San Jose, CA, USA, May 2007.

[7] Q. Li et al., "Mining User Similarity Based on Location History," *16th ACM SIGSPATIAL Int. Conf. Geograph. Inf. Syst.*, Irvine, CA, USA, Nov. 5-7, 2008.

[8] Y. Fang, O.A. Omitaomu, and A.R. Ganguly, "Incremental Anomaly Detection Approach for Characterizing Unusual Profiles," *LNCS 5840: Knowledge Discovery from Sensor Data*, M.M. Gaber et al., Eds., Heidelberg: Springer, 2010, pp. 190-202.

[9] J. Krumm and E. Horvitz, "Predestination: Where Do You Want to Go Today?" *IEEE Comput. Archive*, vol. 40, no. 4, Apr. 2007, pp. 105-107.

[10] N. Talagala et al., "The Art of Massive Storage: A Web Image Archive," *IEEE Comput. Society*, vol. 33, no. 11, Nov. 2000, pp. 22-28.

[11] L.A. Adamic and B.A. Huberman, "Zipf's Law and the Internet," *Glottometrics*, vol. 3, no. 1, 2002, pp. 143-150.

[12] H. Wang, *Research on Distributed Load Balancing and Cache Technologies for Multimedia Networked GIS*, doctoral dissertation, Wuhan University, 2009.

[13] H. Wang et al., "Zipf-like Distribution and Its Application Analysis for Image Data Tile Request in Digital Earth," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 35, no. 3, Mar. 2010, pp. 356-359.

[14] L. Shi et al., "Quantitative Analysis of Zipf's Law on Web Cache," *LNCS*, vol. 3758, 2005, pp. 845-852.

[15] A.R. Ganguly et al., "Knowledge Discovery from Sensor Data for Scientific Applications," *Learning from Data Streams: Processing Techniques in Sensor Networks*, J. Gama and M.M. Gaber, Eds., Heidelberg: Springer, 2007, pp. 205-229.

[16] R. Li et al., "A Prefetching Model Based on Access Popularity for Geospatial Data in a Cluster-Based Caching System," *Int. J. Geograph. Inf. Sci.*, vol. 26, no. 10, Oct. 2012, pp. 1831-1844.

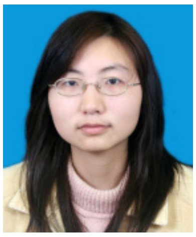
[17] R. Li et al., "A Mathematical Simulation Model for Access Traffic of Geospatial Data," *7th Int. Conf. Comput. Sci. Education*, Melbourne, Australia, July 14-17, 2012, pp. 1127-1129.

[18] S. Chatterjee and A.S. Hadi, *Regression Analysis by Example*, New York: Wiley-Interscience, 1977.

[19] Y. Fang, L. Si, and A.P. Mathur, "Discriminative Graphical Models for Faculty Homepage Discovery," *Inf. Retrieval*, vol. 13, no. 6, Dec. 1, 2010, pp. 618-635.

[20] Y. Fang and M.K. Jeong, "Robust Probabilistic Multivariate Calibration Model," *Technometrics*, vol. 50, no. 3, July 2008, pp. 305-316.

Rui Li received her BE degree and MS degree in computer science from Hubei University and Wuhan University, China, in 1996 and 1999, respectively, and her PhD degree in communication and information systems from Wuhan University in 2006. She is an associate professor in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. She specializes in multimedia networking communication, parallel and distributed computing, distributed real-time systems, and GIS networks.



Wei Feng is a postgraduate in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. She specializes in multimedia networking communication and parallel and distributed computing.

Hao Wang received his BS from East China Jiao Tong University in 2003 and his PhD from Wuhan University in 2009, both in electronics engineering. He specializes in multimedia networking communication and high-performance computing.

Huayi Wu received his BS and MS in mathematical statistics from Fudan University and Wuhan University, China, in 1988 and 1991, respectively. He received his PhD in photogrammetry and remote sensing from Wuhan University in 1999. He is now a full professor in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His major research areas are high-performance geospatial computing and intelligent geospatial web services.

http://dx.doi.org/10.4218/etrij.14.0113.0293