Arpan Kumar Kar · P. Vigneswara Ilavarasan
M.P. Gupta · Yogesh K. Dwivedi
Matti Mäntymäki · Marijn Janssen
Antonis Simintiras · Salah Al-Sharhan (Eds.)

LNCS 10595

Digital Nations – Smart Cities, Innovation, and Sustainability

16th IFIP WG 6.11 Conference on
e-Business, e-Services, and e-Society, I3E 2017
Delhi, India, November 21–23, 2017, Proceedings

ifip

Springer

Arpan Kumar Kar · P. Vigneswara Ilavarasan
M.P. Gupta · Yogesh K. Dwivedi
Matti Mäntymäki · Marijn Janssen
Antonis Simintiras · Salah Al-Sharhan (Eds.)

Digital Nations – Smart Cities, Innovation, and Sustainability

16th IFIP WG 6.11 Conference on
e-Business, e-Services, and e-Society, I3E 2017
Delhi, India, November 21–23, 2017
Proceedings

Springer

Editors
Arpan Kumar Kar
Indian Institute of Technology Delhi
New Delhi
India

P. Vigneswara Ilavarasan
Indian Institute of Technology Delhi
New Delhi
India

M.P. Gupta
Indian Institute of Technology Delhi
New Delhi
India

Yogesh K. Dwivedi
Swansea University
Swansea
UK

Matti Mäntymäki
University of Turku
Turku
Finland

Marijn Janssen
Delft University of Technology
Delft
The Netherlands

Antonis Simintiras
Gulf University for Science and Technology
West Mishref
Kuwait

Salah Al-Sharhan
Gulf University for Science and Technology
West Mishref
Kuwait

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-68556-4 ISBN 978-3-319-68557-1 (eBook)
DOI 10.1007/978-3-319-68557-1

Library of Congress Control Number: 2017955726

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume presents the proceedings of the 16th International Federation of Information Processing (IFIP) Conference on e-Business, e-Services and e-Society (I3E) held at the Indian Institute of Technology, Delhi, during November 21–23, 2017. The IFIP-I3E conference is highly interdisciplinary in nature and focuses on academic contributions in varied domains of electronic business, services, and society. It had great participation from academia, industry, and practitioners who are either working directly in the domain or are in the process of exploring it. The central theme of the 16th edition of the conference was "Digital Nations – Smart Cities, Innovation, and Sustainability." The idea of digital nations perfectly aligns with the three Es of the conference series. As you get through the following pages, you will discover interesting ideas in the domains of governance, social media, and analytics that are shaping our lives today or most likely will alter the digital nations of tomorrow (Zuiderwijk and Janssen 2014; Rana et al. 2017). From any benchmark, the trajectory of technology developments in the post-Internet period is extraordinary and never seen before (Khatwani et al. 2014). It is complex but fascinating; it is bringing people and communities together, enterprises to collaborate on strengths to grab opportunities as competition becomes intense and nations to break boundaries allowing societies to interoperate across different domains and communities. It is producing an amalgamation of various cultures and the emergence of a new way of living – a society that is truly global with global citizenship (Carter et al. 2016).

In the digital society, the issues emerging are plenty and complicated due to the networkedness and possibilities of understanding the complexities through big data analytics (Janssen et al. 2012; Chauhan et al. 2016). The governance is not so much about digital spread as access is fast bridged by expanding mobile networks and broadband as it is about ability in enabling digital-based services to everyone to their expectations. The literacy gap around the globe will remain a cause of concern in the use and adoption of new media, even though the younger generation is fast in adoption. Digital literacy programs will need continuous monitoring and revision to keep pace with the technology change. The underpinning factor in leveraging the new media is having clarity about the culture and community where one is operating. This has not been explored much. Social media has fast become a widespread fascination for everyone, much more than normal social life, raising new challenges and opportunities related to community, culture, and business (Rathore et al. 2016; Lakhiwal and Kar 2016). Understanding variability in cross-cultural boundaries, particularly East vs. West, is necessary in developing a social media strategy for different contextual settings. It hardly needs any emphasis how trust plays a major role in allowing people to overcome the perception of risk and insecurity. Gender effect is also found to play a key role in social media use.

The use of information and communication technologies (ICTs) is greatly impacting the economy (Chew et al. 2010) and governance (Gupta and Jana 2003). Citizens

Contents

Social Media and Web 3.0 for Smartness

Adoption of Smart Services

Outlier Detection Among Influencer Blogs Based on off-Site Web Analytics Data

Reema Aswani[1], S.P. Ghrera[2], Satish Chandra[3(✉)], and Arpan Kumar Kar[1]

[1] Indian Institute of Technology, Delhi, India
reemaswani@gmail.com, arpan_kar@yahoo.co.in
[2] Jaypee University of Information Technology, Waknaghat, Solan, India
sp.ghrera@juit.ac.in
[3] Jaypee Institute of Information Technology, Noida, India
satish.chandra@jiit.ac.in

Abstract. In the current scenario, with the exponential increase in the use of internet, organizations are continuously thriving for visibility on the web. This has opened new avenues in influencer marketing. Several portals encourage these marketers to build content for the purpose of digital marketing. However, the content building process produces a lot of spam within these websites when done in bulk. This is often done in order to establish their presence by using techniques including article spinning and keyword stuffing. This study thus attempts to identify these spam websites using a dataset comprising 2751 websites using bio inspired outlier detection approaches. We use publically available key performance indicators (KPIs) through which websites that create spam content to boost the amount of text in the domain are identified. A hybrid wolf search algorithm (WSA) and bat algorithm (BA) integrated with K-means are used to classify these websites into spam. Findings indicate that metrics including Domain Authority, Page Authority, Moz Rank, Links In, External Equity Links, Spam Score, Alexa Rank, Citation Flow, Trust Flow, External Back Links, Referred Domains, SemRush URL Links and SemRush Hostname Links play an important role in identifying spam. The proposed approach may prove beneficial in segregating spam influencer websites for effective influencer marketing.

Keywords: Outlier detection · Bio inspired computing · Wolf search algorithm · Bat algorithm · Web analytics · Spam detection · Machine learning

1 Background

The exponential increase in the use of internet in this era of digitization across the world has become an important source of competitive edge for the marketing of products and services [1]. This explosion of digital marketing has completely revamped the way business is done and also affects the brand positioning strategy of the

A.K. Kar et al. (Eds.): I3E 2017, LNCS 10595, pp. 251–260, 2017.
DOI: 10.1007/978-3-319-68557-1_23

organizations [2]. Organizations have realized the importance of web visibility for better customer engagement [3]. These organizations have thus started adopting ways to artificially boost their presence on the web using digital marketing specifically opening new avenues for influencer marketing. Influencer marketing is an approach to marketing that focuses on individuals that advise the decision-making consumers. Such people are referred to as influencers and often play a critical role in the customer engagement process [4]. These influencers often need to build huge amount of content in order to maximize web visibility.

The use of web analytics for enhancing digital marketing has been in practice for the last few decades. However, organizations are still not able to fully utilize the core potential of these techniques for improvising their web visibility. Studies highlight opportunities and practices in web analytics that organizations may adopt for better online marketing [5]. The optimization comprises of two primary categories of on-site (a measure of actual visitors on the website) and off-site web analytics (comprising of tools measuring website audience) [6]. One primary reason for failing to achieve the desired promotion from web analytics in online marketing is inexperienced and unskilled influencers. These influencers in order to expedite the process use unethical practices like artificially generating keywords and links to build low quality content. This not only results in that result ineffective off-site analytics but may even prove to be detrimental to the customer if detected by search engines [6, 7]. After the Google's Panda and subsequent updates such malpractices for artificially boosting the web site rank on search engines results page have resulted penalization and website delisting from search engines [8]. This study thus primarily focuses on identifying outlier influencer websites for the purpose of effective off-site web analytics.

There are several freelancing platforms including Blogmint, Influencer, Upwork and Craiglist that offer freelancers to build content on topics that may be utilized for generating back links and keywords for the customer website [9, 10]. These techniques attract traffic to the customer website and artificially boost the website rank. However, the influencers in the process to expedite the process generate low quality content that is often not original and use techniques like article spinning, keyword stuffing, link building and link farming [6] making the website quality a key driver for successful e-business [11]. The customer is often not aware of the adverse effects of such techniques and thus in the long run these may even lead to penalization by search engines. Studies in literature also discuss about website selection for advertising campaigns [12]. To avoid such spam within the website, our study proposes an outlier detection approach that uses website KPIs to identify spam influencer websites that indulge in low quality content building. Metrics like page rank, page authority, domain authority, alexa rank, google index, social shares, trust flow, citation flow, links, external equity link; external back links, referred domains and domain age are used as indicators for identifying spam influencer websites. A spam score is further associated with each of the 2751 websites considered for the analysis. A bio inspired wolf search and bat algorithm integrated with K-Means is used for subsequently segregating the outlier websites.

2 Research Methodology

This study uses a mixed research methodology where in the data collected surrounding the website KPIs for 2751 influencer blogs on unique domains. A statistical t-test is conducted on the normalized data for the two sets of influencer web domains, with low and high spam score. Further, the significant metrics are used as KPIs for analyzing whether the influencer is spam or not using bio inspired optimization approaches integrated with K-Means for mining outliers. The subsequent sub-sections highlight detailed discussions surrounding the analysis.

2.1 Data Collection and Metric Identification

The data is extracted through an API from the SEO Rank website (https://seo-rank.my-addr.com/) that provides a holistic list of selected metrics provided by various data providers like Majestic [13], Ahref [14], Moz [15], SemRush and Webmaster tools. These data providers have developed ranking mechanisms that are used worldwide for identifying the position of a page in organic search. A list of metrics considered for the analysis is demonstrated in Table 1.

Table 1. Description of website metrics for off-site analytics and digital marketing.

	Data provider	Metric	Description
1.	Moz	Domain Authority	Prediction of the ranking of domain on search engines. Depends on links, Moz Rank and other metrics
2.		Page Authority	Prediction of how a given URL may be ranked on search engines, associated with number of links, Moz Rank, and others
3.		Moz Rank	Link popularity score indicative of importance of the page on the web
4.		Moz Trust	Link trust checks for links from trustworthy sources
5.		Links In	Links to the web page, includes equity, or non-equity both internal and external links
6.		External Equity Links	Number of external equity links to the URL
7.		Spam Score	Based on number of sites penalized (de-listed) containing links to the web page
8.	Alexa	Alexa Rank	Global Alexa rank of webpage
9.		Alexa Links number	Number of links to the web page
10.		Country Rank	Alexa Rank in the popular country

(*continued*)

Table 1. (*continued*)

	Data provider	Metric	Description
11.	Majestic SEO	Citation Flow	Uses site link counts to the web page to see how influential the page is
12.		Trust Flow	Trustworthiness of the page based on link to trustworthy neighbours
13.		External Back Links	Total external back links to the web page
14.		Referred Domains	Total unique domains having links to the website
15.	SemRush	SemRush Rank	Domain rank by SemRush
16.		URL Links	Links to the mentioned web page
17.		Hostname Links	Links to the domain
18.	Ahrefs	URL Rating	Strength of web pages' back link profile and its chances of being ranked high in Google
19.		Domain Rating	Strength of website's domain back link profile
20.		Ahrefs Rank	Ranking based on size and quality back link profile
21.		Live & Fresh Index	List of live and dead links for the website

A total of 21 metrics are considered for the study, the data providers are mentioned along the metrics. This study uses a collective list of the metrics as KPIs for detecting spam influencer websites. The spam score is used as the criteria for dividing the data set into two for identifying the statistical significance of the metrics for subsequent analysis.

2.2 Statistical Analysis

The dataset is divided into two equal sets and 500 influencer websites each having a spam score less than 5 and greater than 5 are taken as sample for conducting a statistical t-test to identify metrics that are significantly different in the two sets. Since, the range of values of each of the metrics is considerably varied; min-max normalization is used to standardize the data to a 0–1 range. Subsequently t-test is conducted and the metrics having a p-value less than 0.05 are considered insignificant for further analysis. A list of remaining 12 significant metrics is highlighted in Table 2.

Table 2. Statistically significant metrics having a p-value greater than 0.05

Metric	P-value	Metric	P-value
Domain authority	0.038	Citation Flow	7.23E-28
Page authority	0.030	Trust Flow	0.0003
Moz rank	1.03E-15	External Back Links	0.002
Links in	0.048	Referred Domains	3.94E-32
External equity links	1.25E-05	SemRush URL Links	7.12E-28
Alexa rank	9.23E-15	SemRush Hostname Links	0.045

The final dataset for analysis thus comprises of the 13 significant attributes namely Domain Authority (DA), Page Authority (PA), Moz Rank (MR), Links In (LI), External Equity Links (ELL), Alexa Rank (AR), Citation Flow (CF), Trust Flow (TF), External Back Links (EBL), Referred Domains (RD), SemRush URL Links (UL) and SemRush Hostname Links (HL) for 2751 influencer websites. Subsequent sub-sections model the identified metrics to segregate outliers using bio inspired computing algorithms [16].

2.3 Outlier Detection

After the statistical t-test that identifies significant metrics, a hybrid bio inspired approach is used for detecting outlier influencer websites. Outlier detection is a popular approach when identifying data points that do not comply with majority of the data set based on selective metrics. There are several studies in literature that demonstrate various outlier detection approaches. An exhaustive list of outlier detection approaches with a comparison of motivation, comparison and disadvantages is highlighted with a categorization into statistical models, neural networks, machine learning and hybrid systems [17]. Chandola et al. [18] further provide an exhaustive review of the techniques by grouping the existing studies into six main categories based on classification, clustering, nearest neighbor, statistical, information theoretic and spectral. They further highlight the widespread applications of these approaches across domains encompassing cyber-intrusion detection [19], fraud detection [20], medical anomaly detection [21], image data [22], textual anomaly detection and sensor networks [23].

With the huge data influx, there are studies for outlier detection in high dimensional data [24–26]. However, these approaches are computationally intensive often NP hard and may even lead to a locally optimum solution [27]. Since the data under consideration is huge and may also be unstructured textual data. This is creates need of integrating approaches that do not converge to a local optima. The meta-heuristic approaches are known to help in reaching to a globally optimum system [28]. Further, bio inspired algorithms have been one of the most popular optimization techniques and mimic swarm behavior for optimization problems [16, 29, 30]. Tang et al. [31] thus integrate a few popular bio inspired algorithms with K-means to avoid the local convergence. This study thus utilizes the integrated bio inspired wolf search algorithm for outlier detection. We thus use the 2751 influencer websites comprising of 14 attributes including KPIs for each website for identifying these outliers.

The wolf search algorithm (WSA) is one such optimization approach that is said to overcome local optima by imitating the wolf preying behavior [31, 32]. Another similar wolf hunting approach for grey wolves is used in literature for detecting outliers integrated with k-nearest neighbor [33]. In the current study, the number of clusters is identified as 2 for normal and outlier data points. The wolf population is initialized with visual distance and escape probability. The initial centroids are assigned for the two clusters. The fitness for the centroid in each wolf is calculated and the best solution is identified. The random preying behavior of the wolf is done by selecting a companion having the best solution within the visual distance. If the fitness of the companion is better than the self fitness of the wolf the companion is selected and is thus approached. After the prey is hunted the wolf randomly selects a position beyond the visual range and the process is repeated from the new location. The centroids with the best fitness are considered as the final solution.

Further, the results are compared with the integrated bat algorithm (BA) which uses the echolocation behavior of bats to find the prey and differentiate between different insects even in the dark [31, 34]. The bat algorithm is one of the most popular algorithms used for several engineering, multi-objective and constrained optimization problems [35–37]. For the integrated bat approach along with the two clusters, the bat population, frequency factor and loudness are initialized. The initial clusters are randomly assigned or the bat population. For each bat, the initial centroids are similarly identified. The fitness of the centroids is computed and the best solutions are identified. Further, the new solution is generated by adjusting the frequency and velocity. If the randomly generated solution is greater than the defined pulse rate, a new best solution is selected from the best solutions from each of the bats. The new solutions are accepted by adjusting pulse rate and loudness for subsequent iterations. The pulse rate is increased and the loudness is decreased for the next iteration.

Thus the bio-inspired algorithms help in identifying the best cluster centroids over iterations. The formulation of centroids is mainly iteratively guided by the search agents in the mentioned approaches. Since the dataset considered for this study requires only two clusters and has a total of 14 attributes for which the centroid values need to be computed.

The $Centroid_{ij}$ is value of the centroid for i^{th} cluster and j^{th} attribute. Thus, the $Centroid_{ij} = \sum_{k=1}^{SolSpace} weight_{ki} datapoint_{kj} / \sum_{k=1}^{SolSpace} weight_{ki}$. The centroids largely depend on the weight that tells whether the data point belongs to the cluster or not. $weight_{ki} = 1$, $if\ datapoint_k \in cluster_i; else\ weight_{ki} = 0$. Once the best cluster centroids are identified for the two clusters of outliers and normal data points, a distance measure is subsequently used segregate the outliers. The subsequent section demonstrates the findings.

3 Findings

The K-means integrating WSA and BA algorithms have been used in this study for detecting outliers. The use of bio-inspired algorithms avoids locally optimum solutions. The study demonstrates the segregation of outlier influencer websites based on certain KPIs that have been extracted for a set of 2751 influencer websites using APIs. A total of 13 attributes are considered for detecting the outlier influencers for off-site web analytics. The spam score is excluded for the classification and is used for the validation. Table 3 highlights the cluster centers for the remaining 12 metrics. The table lists the cluster centroids for the authentic blogs (A) and outlier blogs (O) for both WSA and BA.

Table 3. Cluster centers for K-Means Integrated WSA and BA

		DA	PA	MR	LI	ELL	AR	CF	TF	EBL	RD	UL	HL
WSA	A	0.12	0.19	0.18	0.69	0.32	0.74	0.55	0.76	0.84	0.58	0.46	0.75
	O	0.08	0.11	0.09	0.45	0.17	0.31	0.48	0.36	0.61	0.24	0.13	0.31
BA	A	0.20	0.27	0.24	0.81	0.39	0.96	0.59	0.97	0.98	0.40	0.31	0.50
	O	0.11	0.12	0.09	0.33	0.17	0.01	0.51	0.06	0.49	0.35	0.21	0.44

The results for the two approaches used for the purpose show that the bat algorithm shows higher accuracy. Out of 2751 influencer websites, 1254 websites were identified as outliers based on their spam score and manual examination. The bat algorithm correctly identified 1218 giving an accuracy of 97.12% while the wolf search algorithm correctly identified 1203 with an accuracy of 95.93%. However, time taken to converge to the optimum solution is 22.61 s for BA while it is just 16.18 s for WSA. The Fig. 1. demonstrates the outlier plots for WSA and BA.

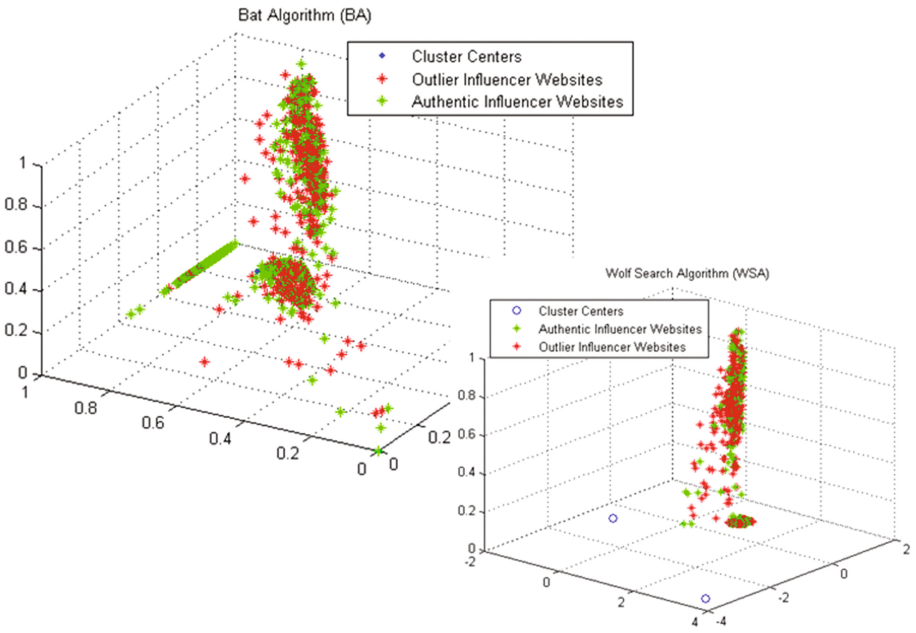

Fig. 1. Outlier plots for WSA and BA

Thus, the findings indicate that a large number (45.58%) of influencer websites are actually outliers. The reason behind this is that majority of influencer websites being categorized as outliers is because these blogs are heavily dependent on techniques like article spinning, link farming and keyword stuffing for content building and subsequent promotion. They often pick up original content and spin/manipulate the content by paraphrasing and including keywords related to the consumer domain to gain traction. These practices are often deemed unfit when it comes to digital marketing. However, the customers adopting these services are often not aware of such malpractices adopted by the websites. This has adverse effects on the consumer website in the long run and may even result in penalization. The use of KPIs in identifying such outlier influencers thus segregates these websites on the basis of publically available metrics from several service providers.

4 Conclusion

With the increased internet use and online marketing opportunities, organizations have realized the importance of web visibility and have started leveraging the power of internet to reach a larger audience for their products and services. This has opened new avenues for digital marketing especially influencer marketing where on several portals have emerged to encourage these influencers to build content for customer businesses. However, this process of content building generates a lot of spam content within these websites when done in bulk for a large consumer base and often involves techniques like article spinning and keyword stuffing for user traction. Such practices are not considered ethical as per the search engine guidelines and affect the consumers adversely. This study thus attempts to use publically available influencer website KPIs, a total of 13 attributes including Domain Authority, Page Authority, Moz Rank, Links In, External Equity Links, Spam Score, Alexa Rank, Citation Flow, Trust Flow, External Back Links, Referred Domains, SemRush URL Links and SemRush Hostname Links for 2751 influencer websites. Further, K-means integrated bio-inspired computing techniques are used for detecting and segregating outliers from the extracted data. Findings indicate that such approaches overcome local optima problems and give globally optimum solutions for such NP hard and computationally extensive data. Further, it is seen that the integrated bat algorithm gives better accuracy than wolf search algorithm as demonstrated in existing literature when the approach is used for clustering [31]. Our study re-establishes the same for the web analytics data set under consideration for outlier detection by extending the proposed approach.

5 Implications and Future Scope

This study uses KPIs and segregates outlier influencer websites that is beneficial for off-site web analytics. This may be useful for preventing consumer investments to such spam influencers that may adversely affect the websites position on search engines in the long run. Apart from the KPIs, content based analytics including keyword density, lexical diversity, meta information and topic modeling may also be incorporated in the analysis.

Future studies can be extended to using social media analytics for further validation of the results since social media platforms are utilized by consumers for raising concerns regarding the services used by them. These platforms specially Twitter and Facebook profiles of such influencer websites provide a lot of information in the form of user generated content that may be integrated with the existing metrics to reinforce the findings. An empirical validation of the results can also be done using a structured questionnaire for the consumers opting for such influencer marketing services and the short term and long term impact of the same on their visitors and web visibility. An existing study surrounding an analysis of results suggested by search engines for market share establishment can also be extended for influencer marketing [38].

References

1. Leeflang, P.S., Verhoef, P.C., Dahlström, P., Freundt, T.: Challenges and solutions for marketing in a digital era. Eur. Manag. J. **32**(1), 1–12 (2014)
2. Dou, W., Lim, K.H., Su, C., Zhou, N., Cui, N.: Brand positioning strategy using search engine marketing. MIS Q. **34**(2), 261–279 (2010)
3. Sawhney, M., Verona, G., Prandelli, E.: Collaborating to create: The Internet as a platform for customer engagement in product innovation. J. Interact. Mark. **19**(4), 4–17 (2005)
4. Brown, D., Hayes, N.: Influencer Marketing: Who Really Influences Your Customers?. Routledge, London (2008)
5. Chaffey, D., Patron, M.: From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. J. Direct Data Digital Mark. Pract. **14**(1), 30–45 (2012)
6. Malaga, R.A.: Worst practices in search engine optimization. Commun. ACM **51**(12), 147–150 (2008)
7. Moreno, L., Martinez, P.: Overlapping factors in search engine optimization and web accessibility. Online Inf. Rev. **37**(4), 564–580 (2013)
8. A Complete Guide to Panda, Penguin, and Hummingbird. Search Engine Journal. http://www.searchenginejournal.com/seo-guide/google-penguin-panda-hummingbird. Last accessed 15 Feb 2017
9. Jain, A., Dave, M.: The role of backlinks in search engine ranking. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 3(4) (2013)
10. Zuze, H., Weideman, M.: Keyword stuffing and the big three search engines. Online Inf. Rev. **37**(2), 268–286 (2013)
11. Lee, Y., Kozar, K.A.: Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach. Decis. Support Syst. **42**(3), 1383–1401 (2006)
12. Kar, A.K.: A decision support system for website selection for internet based advertising and promotions. In: Sengupta, S., Das, K., Khan, G. (eds.) Emerging Trends in Computing and Communication. LNEE, vol. 298, pp. 453–457. Springer, New Delhi (2014). doi:10.1007/978-81-322-1817-3_48
13. Positive link building using Majestic tools and metrics. Majestic Blog. https://blog.majestic.com/training/positive-link-building-with-majestic-tools/. Last accessed 10 Feb 2017
14. Ahrefs' SEO Metrics Explained (Finally). Ahrefs Blog. https://ahrefs.com/blog/seo-metrics/. Last accessed 10 Feb 2017
15. A Practical Guide to Content and Its Metrics. Moz Blog. https://moz.com/blog/practical-guide-content-metrics. Last accessed 15 Feb 2017
16. Kar, A.K.: Bio inspired computing–A review of algorithms and scope of applications. Expert Syst. Appl. **59**, 20–32 (2016)
17. Hodge, V., Austin, J.: A survey of outlier detection methodologies. Artif. Intell. Rev. **22**(2), 85–126 (2004)
18. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comput. Surv. (CSUR) **41**(3), 15 (2009)
19. Patcha, A., Park, J.M.: An overview of anomaly detection techniques: Existing solutions and latest technological trends. Comput. Netw. **51**(12), 3448–3470 (2007)
20. Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y., Sun, X.: The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decis. Support Syst. **50**(3), 559–569 (2011)

21. Laurikkala, J., Juhola, M., Kentala, E., Lavrac, N., Miksch, S., Kavsek, B.: Informal identification of outliers in medical data. In: Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology, pp. 20–24 (2000)
22. Stein, D.W., Beaven, S.G., Hoff, L.E., Winter, E.M., Schaum, A.P., Stocker, A.D.: Anomaly detection from hyperspectral imagery. IEEE Signal Process. Mag. **19**(1), 58–69 (2002)
23. Basu, S., Meckesheimer, M.: Automatic outlier detection for time series: an application to sensor data. Knowl. Inf. Syst. **11**(2), 137–154 (2007)
24. Aggarwal, C.C., Yu, P.S.: Outlier detection for high dimensional data. In: ACM SIGMOD Record, pp. 37–46. ACM (2001)
25. Kriegel, H.P., Zimek, A.: Angle-based outlier detection in high-dimensional data. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 444–452. ACM (2008)
26. Zhou, X.Y., Sun, Z.H., Zhang, B.L., Yang, Y.D.: Fast outlier detection algorithm for high dimensional categorical data streams. Ruan Jian Xue Bao(Journal of Software) 18(4), 933–942 (2007)
27. Chawla, S., Gionis, A.: k-means–: A unified approach to clustering and outlier detection. In: Proceedings of the 2013 SIAM International Conference on Data Mining, pp. 189–197. Society for Industrial and Applied Mathematics (2013)
28. Blum, C., Roli, A.: Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM Comput. Surv. (CSUR) **35**(3), 268–308 (2003)
29. Binitha, S., Sathya, S.S.: A survey of bio inspired optimization algorithms. Int. J. Soft Comput. Eng. **2**(2), 137–151 (2012)
30. Chakraborty, A., Kar, A.K.: Swarm intelligence: a review of algorithms. In: Patnaik, S., Yang, X.-S., Nakamatsu, K. (eds.) Nature-Inspired Computing and Optimization. MOST, vol. 10, pp. 475–494. Springer, Cham (2017). doi:10.1007/978-3-319-50920-4_19
31. Tang, R., Fong, S., Yang, X.S., Deb, S.: Integrating nature-inspired optimization algorithms to K-means clustering. In: Seventh International Conference on Digital Information Management (ICDIM), pp. 116–123. IEEE, Macao (2012)
32. Tang, R., Fong, S., Yang, X.S., Deb, S.: Wolf search algorithm with ephemeral memory. In: Seventh International Conference on Digital Information Management (ICDIM), pp. 165–172. IEEE, Macao (2012)
33. Aswani, R., Ghrera, S.P., Chandra, S.: A novel approach to outlier detection using modified grey wolf optimization and k-nearest neighbors algorithm. Indian J. Sci. Technol. 9(44) (2016)
34. Yang, X.S.: A new metaheuristic bat-inspired algorithm. In: González, J.R., Pelta, D.A., Cruz, C., Terrazas, G., Krasnogor, N. (eds.) Nature inspired cooperative strategies for optimization (NICSO 2010), Studies in Computational Intelligence, vol. 284, pp. 65–74. Springer Springer, Heidelberg (2010)
35. Yang, X.S., Gandomi, A.H.: Bat algorithm: a novel approach for global engineering optimization. Eng. Comput. **29**(5), 464–483 (2012)
36. Yang, X.S.: Bat algorithm for multi-objective optimisation. Int. J. Bio-Inspired Comput. **3**(5), 267–274 (2011)
37. Gandomi, A.H., Yang, X.S., Alavi, A.H., Talatahari, S.: Bat algorithm for constrained optimization tasks. Neural Comput. Appl. **22**(6), 1239–1255 (2013)
38. Utsuro, T., Zhao, C., Xu, L., Li, J., Kawada, Y.: An empirical analysis on comparing market share with concerns on companies measured through search engine suggests. Global J. Flex. Syst. Manage. 1–17 (2017)