



IDENTIFYING INFLUENTIAL BLOGGERS ON THE WEB

Jimoh, R. G.

Department of Computer Science
University of Ilorin, Ilorin
jimoh_rasheed@yahoo.com

Awotunde Joseph Bamidele

Department of Computer Science
University of Ilorin, Ilorin
jabonnetbylinks@gmail.com

Enikuomihin, O. A.

Department of Computer Science
Lagos State University, Lagos

ABSTRACT

Blog has taken an important aspect of internet since the introduction of Web 2.0 technology because blog has been a way to influence others who read the blogs. People now have interest in finding materials and friends on the internet. Many users visit blog sites to read the posts and comment on them. Most people do read blog to gather information on things that are important before taking major decisions about them. Because bloggers always share their experience on a topic for others to comment and through this others share their own experience. With the impact that influential bloggers have in a community. The benefits of achieving competitive advantages in a blog community by identifying influential bloggers have created several research gaps and the popularity of these services has made the problem of identifying the most influential bloggers significant, since its solution can lead to major benefits for the users of these services i.e. education, politics, participatory journalism, advertising, searching, commerce etc. The current works in this regard ignore some important aspects of the blogosphere. This paper focuses on using a crossbreed method as an improvement to the existing methodologies. With the introduction of new parameters FBCount and Mining Comments the new approach shows that the score of each blog post reflects quality and goodness of blog posts. A program prototype was designed to calculate the influential bloggers. The results obtained confirm that the current approach could significantly identify influential bloggers on the web and the proposed model has better performance than other approaches. There are still a few avenues for future research. Future work can include full implementation of the program prototype and try to improve on it to directly get the parameters used from the blog post on the web in a blog community, more parameters like twitter shares, G+1s Pin shares etc can be included into the literature and check for the behavior of the influence and future research can investigate more time in deciding weight parameter that is crucial for tuning between different influential factors.

Keyword: *Blog, Blogger, Social networks, Blogosphere, Influential bloggers, Influential, Models.*

INTRODUCTION

The characteristics of blogs, such as permanence, personal responsibility, control over content and writing, enabled a blogger to build an online identity over time [16]. [9] further pointed out that it was not the content published, but the personalities or the identities behind blogs that distinguished blogs from online discussion forums. Unlike an online discussion centering on a topic or a community, a blog was often a personal narrative of feelings, experiences, and thinking [16].

The shift from discussion forum to blogosphere was described as a transition from a shared space governed by group norms to a personal territory where interactions were initiated, centered around, and regulated by individuals. The content of blogs was found to be more varied and individually distinct than that on the discussion forum [16].

[11] argued that the collective structure of an online forum might be more conducive for nurturing substantive online groups, while the individual sovereignty rendered by blogs allowed for more liberty to individual expressions and less conformity to group rules and expectations.

Some theoretical and empirical endeavors have been made to determine the distinct value of blogs. Several researchers have collectively claimed that the unique value of blogs resided in the dual roles they served as both individual and collaborative vehicles. Representative descriptions of such dual roles of blogs include monologues and dialogues [21]; conversations with self and others [17]; individualistic and collaborative [12]; a combination of self-expression and social interactions [19]. [9] described it as a synergy of personal and community learning as blogs provided personal spaces without imposing a community learning agenda or style. At the same time, learners were not alienated but could benefit from feedback from the community, validation, and support. Along the same vein, [5] commented: “blogs represent for authors an opportunity to reach out and connect with an audience never before accessible to them, while maintaining control over their personal expressive spaces.

[12] further explicated the advantages of such a mixture of personal and community space: on one hand, the personal ownership of blogs affords an individual voice and; on the other hand, the interactive and linking mechanisms open the possibility for discussion and



connection. “This clear ownership principle takes advantage of the democratic nature of online communications while preserving the freedom of expression that is valued in a personal diary or journal.” Identifying influential bloggers is a significant problem as the bloggers can take special roles in the society. Commercial companies can use them as their unofficial spokesman to promote their products and services. They can affect individuals voting behavior, so political parties can use their influence to benefit at the vote bank and forge their political agendas. They can be market movers, market promoters. So identifying such influential bloggers is of significant importance as they influence the decisions of others. Identifying the blog sites with greater marketing influence capabilities is crucial in promotion of a product/service.

This paper addresses the problem of identifying influential bloggers in a web blog community. It investigates the problem of identifying influential bloggers by scoring each blog post, posted by bloggers, based on influential factors and ranking bloggers accordingly.

The extensive but still the present model demonstrates that:

- ✓ Influential bloggers are not necessarily active bloggers,
- ✓ Active bloggers are not necessarily influential bloggers,
- ✓ By tuning the weights associated with the parameters of the present model, one can examine how different parameters impact the influence ranking for different needs, and
- ✓ The present model can serve as a baseline in identifying influential bloggers and can be extended by incorporating additional parameters to discover different patterns.

The aim of this paper is to identify influential bloggers at a blog site regardless of the site being influential or not. The issue of identifying influential bloggers looks very similar to finding influential blog sites and authoritative web pages but the techniques used in solving these problems can't be used in this paper. In achieving this aim, the following specific objectives are formulated.

- 1 To study existing models of identifying influential bloggers.
- 2 To propose a better model for identifying influential bloggers.
- 3 To implement the propose model in finding influential bloggers.
- 4 A program prototype was designed to calculate the influential bloggers.

A common technique for identifying influential bloggers is to be study, inspired from the interesting insights present in the literature; we will analyzed the algorithms to find out the areas of improvement. The paper thoroughly investigated blogosphere for additional influential factors that can further improve the influential mining system.

We also measured the novelty of each blog post and utilized the parameter in the new algorithm so that the score of each blog post reflect quality and goodness of blog post. An experiment is carried out and the results obtained. The paper then compares the new results against the previous model.

RELATED WORK

Finding influential blog sites in the blogosphere is an important research problem, which investigates how these blog sites influence the external world and within the blogosphere [10]. The works [2][3], [18], [1][13] applied various network-based parameters and dimensions to examine the influence of commentary information in online social network.

The study of influence in the spread of topics and ideas in the blogosphere has been performed by [8]; [13]; [13][14]. There is a rich foundation in the study of diffusion in networks with decades worth of work from various fields including sociology, biology, economics, epidemiology, physics and marketing [20]. [7] studied the dynamics of information propagation identifying chatter and spikes of topics.

They use the Independent Cascade model of infectious diseases to model the propagation of topics and information. When a blogger writes on a topic there is a probability associated with his neighbors to read it and to further go ahead and write about it.

Moreover, [2] presents a well-defined preliminary influence model to identify influential bloggers and paves the way for building a robust model that allows for finding various types of the influentials. In our model, we extend the concepts and ideas from [3].

[1][13][14] used the models proposed by [8] to maximize the spread of influence in a social network and found influential nodes.

They apply the models to blogosphere to predict influential blogs. Their experiments also show that effective prediction of influential can be done with a simple heuristic based page rank technique.

The issue of identifying influential bloggers looks very similar to finding influential blog sites and authoritative web pages but the techniques used in solving these problems can't be used in the current project.

[2], have introduced the problem of identifying influential bloggers for the first time in 2008. They have proposed "influential flow method" to solve the problem by using an initial set of intuitive properties like recognition, Activity generation, Novelty, and Eloquence.

Later on [15] have introduced productivity and temporal aspects into literature in 2009 which are crucial factors of blogosphere. They have proposed two metrics called MEIBI (Metric for Evaluating and Identifying Blogger's Influence) & MEIBIX (MEIBI extended) index to identify current influential bloggers. In 2011 the same authors has modified their model to capture temporal (time) patterns of blogosphere. They have introduced a model that generates temporal patterns for both Productive bloggers (Active bloggers that blogs very often) and influential bloggers (actual influential bloggers). [6], have proposed a domain specific influential bloggers mining system. In their proposal they introduced a model to mine the top-k influential bloggers according to their interest domains and network proximity. They have proposed a model that can evaluate blogger's influence and develop a domain specific influential blogger mining system.

Inspired from the interesting insights present in the literature, we analyzed the algorithms to find out the areas of improvement. We thoroughly investigated



blogosphere for additional influential factors that can further improve the influential mining system. The modern web world is associated with parameters such as the number of hits, likes, G+ 1s, and sharing on social networking websites. Nevertheless, these parameters can be used as influential factors. The existing model's 'blog scoring' methods have merely depended on numbers (length, number of inlinks and comments) associated with blog posts which makes the model look over simplistic.

Considering all the aspects discussed above, in the new approach we introduced two new influential factors FBCount and LBCount, which is a combination of facebook likes + shares and Living time of blog in network. We combined and improved the ideas and methodologies present in existing models with a better machine learning techniques. While evaluating a blog post instead of just depending on the number we actually mined each comment on the blog post to capture the true tone of the commenter and evaluated the blog post accordingly. We also measured the novelty of each blog post and utilized the parameter in the new algorithm so that the score of each blog post reflect quality and goodness of blog post.

INFLUENTIAL BLOGGER

Beyond any doubt, the number of incoming links to a blog post is strong evidence of its influence. Similarly, the number of comments made to a post is another strong indication that this blog post has received significant attention by the community. Many users express their opinion on the post through comments. Some people agree with the author's idea and some people disagree. Outlinks are one other factor, unlike inlinks they work against the influence. As discussed in Influence Flow

So given facebook count Fb_i^j , a time varying Fb-score for a blog i of blogger j is given by

$$Fb_{i,p}^j(t) = w_{fb} * Fb_i^j * \left(\frac{\theta}{t - t_{i,p}^j + \theta} \right)^\delta \quad (\text{eq. 1})$$

δ, θ are predefined constants whose values are 1, 86400 respectively. $t - t_{i,p}^j$ represents the time difference between the current date and publication date of the post in seconds. w_{fb} the weight parameter of facebook count factor whose value is 5.

When we are considering the activity generated ("Post-Reply relationship" [6]) by a blog post there are important and valuable information embedded with it. First, the influence of each comment may have different impact power, depending on who issues it. The comment might be from an expert on the topic or a general blogger or a normal user.

Their comments on blog post should be treated differently, and it is easy to see an expert's comment

$$C_{i,p}^j(t) = w_c \sum_{v \in C_i^j} \left(\frac{\inf(b_c) * SF(b_c * v)}{TC(b_c)} \right) * \left(\frac{\theta}{t - t_{i,p}^j + \theta} \right)^\delta \quad (\text{eq. 2})$$

w_c is the weight parameter of comment factor whose value is 8.

Method [2], outlinks in a post indicate usage of other's idea to support his/her own idea. i.e. outlinks work against the novelty of the post [2]. Temporal (time) aspects of blogosphere also play a major role in influence, as the blog post that exerted impact on public today may not have any influence after two months or so [15]. The paper hasn't considered the length of the post as influential factor because length doesn't matter as long as the post's content is of good quality.

Finally, we have considered the latest parameters associated with blog posts such as facebook likes, G+ 1's, sharing and length time in network on social networking websites as our new influential factors. These latest parameters are also the indication of direct influence exerted by blogs, same as inlinks.

Before reading random articles, news, web blog or a personal profile, many of us tend to look at the number of hits (visitors), likes (fb-likes, G+ 1s etc.) or sharing (facebook, twitter etc.) on them, so that we can get an intuitive idea of goodness of current reading.

Nevertheless, these hits, likes and sharing can be used as influential factors for the current literature. The act of clicking on fb-like, G +1 or sharing on social web sites most likely indicates the interest or approval of the content read. The number of views, f-likes, G+1's on a web blog post can be treated as the direct influence received by post same as inlinks. People by sharing on facebook and twitter are supporting and voting for the web content, so capturing those shares and votes can help us in evaluating support or acceptance that the web content has received. But keeping in mind over fitting error problem we have considered the combination of facebook likes and shares as "facebook count" parameter and introduced it as a new influential factor into the literature.

would enhance the influence of post more. Secondly, the comments from other bloggers could be positive, negative or neutral, and these sentimental factors also affect the post's influence among commenters. So given set of comments to blog post C_i^j , for each comment x the researcher ought to find the person who issued it and his consent towards the blog post. The researcher uses this information to find the comment score achieved by the blog post. Also each comment is associated with the time stamp which we can use to keep up with the dynamic nature of the blog post. For a received set of comments $C_{i,p}^j(t)$ the comment score for i^{th} post of blogger j is calculated as follows:

Also, when we are scoring each blog post, the model should be able to mine the post content. The score of blog post should also reflect the goodness and quality of



the content. Novelty is one such parameter that is directly related to blog post and can convey quality of content present. So we include novelty in our scoring algorithm. Novelty is inversely proportional to outlinks present in the post [2]. i.e. the more the number of outlinks present in blog post, the more the number of ideas borrowed from others to support the idea present in the post. Also to measure the novelty we should know how novel the content is for which need to build a word space that

represent a general space of ideas. In the new approach the researcher considered each blog post as a document. We first removed all the stop words from the documents and then using TFIDF weighing scheme to extracted all the important keywords from the documents to create word space. After creating word space the researcher project each latest blog post into the word space to estimate the degree of novelty. And the novelty is calculated as follows.

$$N_i^j = W_n \left(1 - \frac{\text{number of words matching}}{\text{total no of words}} \right) - w_o * \text{outlinks} \quad (\text{eq. 3})$$

Where W_n is novelty weight parameter and W_o is outlinks weight parameter whose values are 30 and 1/10, respectively. So finally the influence score for the each blog post is calculated as follows.

The score for post i of blogger j is:

$$PS_{i,p}^j = (Fb_{i,p}^j + C_{i,p}^j + N_i^j) \quad (\text{eq. 4})$$

After calculating the post score $PS_{i,p}^j(t)$ for all blog posts using eq.11 the model assigns Blogger j a Influence- index value at a given instance t , equal to P_t^j , if and only if P_t^j blog posts of blogger j , has got a score $\geq P_t^j$ and remaining $BP(j) - P_t^j$ posts got a score $< P_t^j$.

PROPOSED SYSTEM DESIGN AND IMPLEMENTATION

This research is being carried out for the sole purpose of designing a program for calculating influential blogger in a blog community. It is expected that the user should input blog data and comments on blog into the system after which the program will compute the comments statistics and the influential blogger using the compute formula by extract data from the database.

The Influential Blogger was designed using the client/server architecture. There will be a connection between the blog data, blog summary, comments statistics, comment on post and the results on the database. The reason for choosing Client/Server model

for this application is because it provides adequate security for the resources required for a critical application such as identify of influential blogger. The work is implemented using Java software tool, used to design the user interfaces and calculate influence.

A detail description of the system is shown below using the flowchart and the pseudocode

ComputeInfluence()

1. Theta = 86400
2. Lamda = 1
3. FbWeight = 5
4. AuthorList ← getAuthors()
5. for(AuthorList : Author)
 - 6. Tp ← getTotalPost()
 - 7. Tc ← getTotalComment()
 - 8. TI ← getTotalInlinks ()
 - 9. Fb ← getTotalFBCount()
 - 10. timeDiff = getPostDate () - getCurrDate()
 - 11. Score = FbWeight * Fb * ((Theta / (timeDiff + Theta)) ^ lamda
 - 12. Print Score
 - 13. end-for
13. return

AuthorList = list of Authors in database

Tp = Total number of post by an Author

Fb = Total number of Facebook Count

Fc = Total number of Comment

FI = Total number of Inlinks

timeDiff = time difference in seconds

Score = Influence score

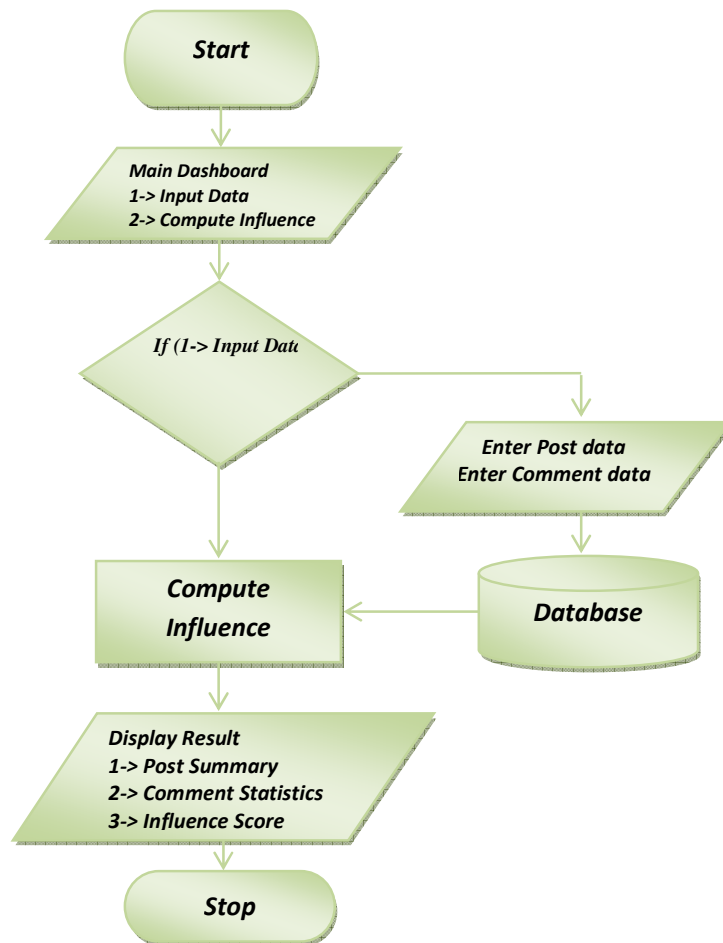


Fig. 1: Flowchart

INFLUENTIAL BLOGGER DESIGN INTERFACE

The first page show is the Menu buttons, which consist of the File menu, Action menu and Help menu fig. 2.

File Menu: This consists of the Blog Data and the Exit buttons. They are used for entering the blog post, comments on post and exit the program respectively fig. 3.

Action Menu: This consists of three buttons: Compute Assessment, Reset and Clear Database. It used to compute the influential blogger using the blog post entered, to undo the entered data and clear the data in the database respectively fig. 4.

Help Menu: It consists of the About button, it is used to know little information about the program fig. 5.

Blog Data: Clicking the Blog data new page will appear where you enter the Blog Post and Comment on Post

where you enter the blog post and comments that the program used in calculating the influential blogger fig. 6.

Blog Post Button: This is where you enter the Author's Name, Title of the post, Date, Total comment, Total Inlinks and FBCount fig. 7.

Comment on Post: This is where you enter the Post ID, Author's Name, Tone and the body of the comment.

Post Summary: This show information that have been input through the blog post button fig. 8.

Comment Statistics: It contains compute comment tone base on the comments entered and the tone of the comments fig. 9.

Result: The final compute influence that compare the new model with the old model fig 10.

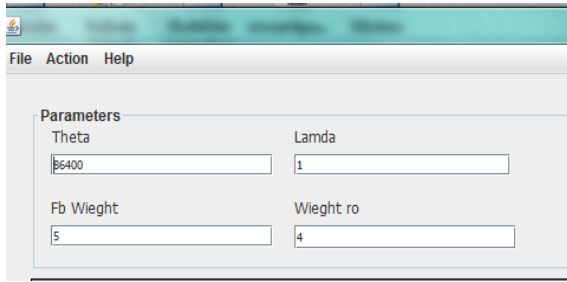


Fig. 2: The Menu Buttons

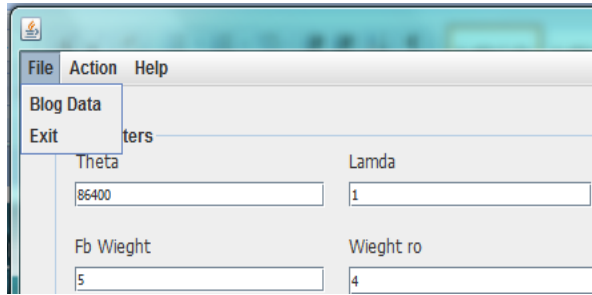


Fig. 3: The File Menu

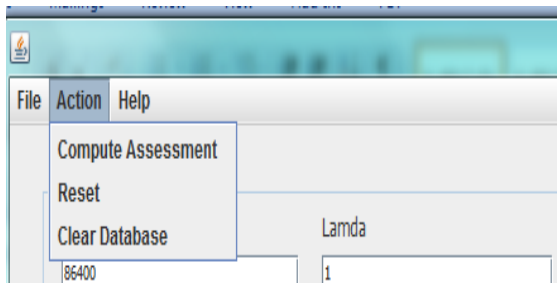


Fig. 4: Action Button

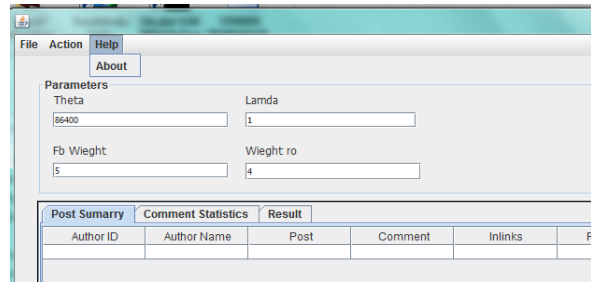


Fig. 5: Help Button

Id	Title	Author	Date	Comment	Inlinks	FBcount
1	Motorola makes il official, melds And...	Darren Murph	22 10 2010 ...	470	59	508
2	Smartphone, content initiatives this ...	Chris Ziegler	22 10 2010 ...	345	21	413
3	Nokia Nuron for T-Mobile review	Vladislav Savov	22 10 2010 ...	382	56	458
4	Engadget give away: win some Beatl...	Laura June	22 10 2010 ...	361	29	402
5	Kong domain, an 'entirely legal' wor...	Nilay Patel	22 10 2010 ...	390	67	359
6	Kindle for iPad and tablets makes th...	Thomas Ricker	22 10 2010 ...	361	29	341
7	For AT & amp, T in the coming month...	Joshua Topolsky	22 10 2010 ...	328	52	311
8	Motorola's Android powered il launc...	paul Miller	22 10 2010 ...	267	7	319
9	BB is very expensive but people are ...	Donald Melanson	22 10 2010 ...	45	15	305
10	Early reports show IE not faring well ...	Tim Steven	22 10 2010 ...	345	519	305
11	Tata Motors' Nano lights up the stre...	Jospeh L. Flatley	22 10 2010 ...	114	13	271
12	Kong domain, an 'entirely legal' wor...	Ross Miller	22 10 2010 ...	273	44	391
13	Spring Design Alex review	Joanna Stern	22 10 2010 ...	298	15	202
14	Engadget give away: win some Beatl...	Ryan Block	22 10 2010 ...	146	15	162
15	George Takei can't show you Sharp'...	Richard Lai	22 10 2010 ...	45	21	165

Fig. 6: Blog Post

Id	PostId	Body	Author	Tone
1	12779	There is some quality engineering right there	Spasewalkr	Positive
2	12779	V for vendetta, i see whats going on here...	TJK	Neutral
3	12779	\$2,500? If I needed a car and these were sold in the US...	Brent1700	Neutral
4	12779	@Brent1700 It won't pass the crash tests. \$2500 can st...	Vanmankline	Negative
5	12779	Lier, lier!Cars on fire	lluvms	Neutral
6	12779	@Abe Of course, if they've only sold 6 of 'em, that's a 5...	Connie	Negative
7	11335	haha wwiwa wow.	Germangabriel	Positive
8	11335	@germangabriel what a legend	dauidmuful	Neutral
9	11335	@germangabriel Now the other manufactures are goin...	Blackstar	Negative
10	11335	You Know, I have the power to blow minds. Literally	Lord Vader	Neutral
11	11335	@germangabriel I haven't enjoyed an advert this much i...	Pdotg	Positive
12	11335	@germangabriel *passes by peter griffin* - "well, helloo...	Saad the other one	Neutral
13	11335	Coming soon to a Jaguar / Land Rover dealer near you.	Wingdo	Neutral
14	12779	Hot as Curry	SmokemeaKipper	Positive
15	12779	This is why airdiners' avionics don't run Vista	Oflife	Positive

Fig. 7: Comments Post



Post Summary		Comment Statistics		Result	
Author ID	Author Name	Post	Comments	Inlinks	FB Count
1	Laura June	1	361	29	402
2	Joshua Topolsky	1	328	52	311
3	Thomas Ricker	1	361	29	341
4	Ross Miller	2	273	44	391
5	Chris Ziegler	1	345	21	413
6	Tim Stevens	1	345	19	305
7	Paul Miller	1	267	7	319
8	Nilay Patel	1	390	67	359
9	Donald Melanson	1	45	15	305
10	Joanna Stern	1	298	15	202
11	Joseph L. Flatley	1	114	13	271
12	Ryan Block	1	146	15	162
13	Vladislav Savov	1	382	56	458
14	Darren Murph	1	470	59	508
15	Richard Lia	1	45	21	165

Fig. 8: Post Summary

Post Summary		Comment Statistics		Result	
Post ID	Positive	Negative	Neutral	Total Comment	Avg Influence
12779	15	17	82	114	79
11335	26	19	37	82	71

Fig. 9: Comment Statistics Button

Post Summary		Comment Statistics		Result			
Author ID	Author Name	BI	BP	I-FBCount	I-Mining Co...	I-(FB + Mining C...	I-New Approach
1	Laura June	56	62	98	76	87	82
2	Joshua Top...	51	56	76	69	73	64
3	Thomas Ric...	56	62	83	76	80	70
4	Ross Miller	17	19	75	24	50	63
5	Chris Ziegler	54	59	101	73	87	84
6	Tim Stevens	54	59	74	73	74	62
7	Paul Miller	41	46	78	56	67	65
8	Nilay Patel	60	67	87	82	85	73
9	Donald Mela...	7	8	74	9	42	62
10	Joanna Stern	46	51	49	63	56	41
11	Joseph L. Fl...	18	19	66	24	45	55
12	Ryan Block	23	25	39	31	35	33
13	Vladislav Sa...	59	65	111	80	96	94
14	Darren Murph	73	80	124	99	112	104
15	Richard Lia	7	8	40	9	25	34

Fig.10: Final Result of the Influential Bloggers



DISCUSSION

When we applied the final model on dataset the ranking has been definite and more accurate than the other models. The column I-New Approach indicates the Influence Indexes of final model. As shown in fig.10 the ranking has been reordered in highlighted sections. The bloggers were ranked without any overlapping among them. As showed in fig.10 V Savov has been ranked 2nd and C Ziegler ranked 3rd. Comparing with I-(FB+Comments) result, T Ricker has taken 5th rank and L June took 6th. And the ranking among J Topolsky, P Miller, D Melanso, and T Stevens is interchanged without any overlapping.

CONCLUSION

Blogs have demonstrated an ability to influence: to shape events, news, educations, politics etc. This makes the technology and those who use it key players in day to day activities of human endeavor. It have been a major communication channel for many internet users where they sharing ideas, opinions, and experience with others through comments.

Blogs have made it possible for people to express their thoughts, voice their opinions, and share their experiences and ideas. Individuals experience a sense of community, a feeling of belonging, a bonding that members matter to one another and their recess needs will be met through online interactions.

Blogs are becoming one of the most popular media of communication and interaction among blog community. Identify influentials blogger will not only allow us to better track and understand the interesting things happening in a virtual world, but also present unique opportunities for academic scholars, political scientist, industry, sales, and advertisements etc.

There exist model that proposed various solutions to identify influential bloggers in web community. Attempts to mathematically determine which blogs are the most influential are currently characterized by the shortcoming of failed to consider important aspects of the blog posts. The existing methodology mainly depended on numbers associated with the blog post i.e. comments, inlinks, outlinks lengths etc. They failed to considered the user tone toward the blog post (i.e. agree or disagree) which is very important while calculating (evaluate) the influence neither process measures influence by considered the novelty of post which indicates the goodness and purity of the content.

Encouraged from the literature review and classy solutions that the current literature lacking, we proposed a method that evaluate each post based on their quality and user consent toward the content present. The competing methods have not taken into account the quality and user consent towards the content present, which we argue are also important ones when evaluate influence of bloggers in a virtual community.

We introduced new parameter which can be used as influential factor (facebook count) into the literature that further evaluated blog posts proficiently. The main motivation for the introduction of facebook count is that they are very important and helped in identify the influence of a blog in web community.

The obtained results verified that the introduction of facebook count (combination of facebook likes and shares) has greatly improved the model. The factors helped the model in finding the right tuning between different blog posts. Also considering number of comments we actually mined each comment to find out the tone of commenter towards the post, which helped the model to capture the true influence of the blog post. Lastly using the novelty parameter the model is able to express the goodness and quality of content.

The experimental results shows that current approach could significantly identify influential of bloggers on the web and the proposed model have better performance than other approaches such Influence Flow Method, MEIBI (Metric for Evaluating and Identifying Blogger's Influence) and MEIBIX (MEIBI extended). This show the current approach has improved over the previous models.

FUTURE WORK

There are still a few of avenues for the future research. First, future work Future work can include full implementation of the program prototype and try to improve on it to directly get the parameters used from the blog post on the web in a blog community. Secondly more parameters like twitter shares, G+1s Pin shares etc into the literature and check for the behavior of the influence. Thirdly, the new approach has used list of positive and negative word phrases to find out the tone of the commenter, but a better mining techniques can further be used in improving accuracy of sentiment factor of the commenter. Lastly, future research can investigate more time in deciding weight parameter that is crucial for tuning between different influential factors.

REFERENCE

- [1] .A. Java, X. Song, T. Finin, and B. Tseng, Why we twitter: understanding microblogging usage and communities," in WebKDD/SNA-KDD, 2007, pp. 56-65.
- [2] Agarwal N. and H. Liu, "Blogosphere: Research issues, tools and applications," ACM SIGKDD Explorations, vol. 10, no. 1, pp. 18–31, 2008.
- [3] Agarwal, N., Liu, H., Tang, L. and Yu, P.S. 2008. Identifying the influential bloggers in a community. Proceedings of the international conference on Web search and web data mining pp. 207–218.
- [4] Agarwal, N., Liu, H.: Modeling and Data Mining in Blogosphere. Moegan & Clay-pool Publishers (2009)
- [5] Bruns, A., & Jacobs, J. (2006). Introduction. In A. Bruns & J. Jacobs (Eds.), Uses of blogs (pp. 1–8). New York: Peter Lang.
- [6] Cai, Y., & Chen, Y., (2010). MASS: a multi--facet domain--specific influential blogger mining system. Data Engineering (ICDE), IEEE 26th International Conference on, vol.,no.,



- pp.1109-1112,1-6 March 2010.
doi:10.1109/ICDE.2010.5447798
- [7] D. Gruhl, David Liben-Nowell, R. Guha, and A. Tomkins. Information diffusion through blogspace. SIGKDD Exploration Newsletter, 6(2):43–52, 2004.
- [8] David Kempe, Jon Kleinberg, and Eva Tardos. Maximizing the spread of influence through a social network. In Proceedings of the KDD, pages 137–146, New York, NY, USA, 2003. ACM Press.
- [9] Efimova, L., & Fiedler, S. (2004). Learning webs: Learning in weblog networks. In Proc. IADIS int. conf. web based communities (pp. 490–494).
- [10] Gill, K., (2004), “How can we measure the influence of the blogosphere?”, Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, WWW 2004.
- [11] Hodkinson, P. (2007). Interactive online journals and individualization. *New Media & Society*, 9(4), 625–650.
- [12] Huffaker, D. (2005). The educated blogger: Using weblogs to promote literacy in the classroom. *AACE Journal*, 13(2), 91–98.
- [13] Java Akshay, Pranam Kolari, Tim Finin, and Tim Oates. Modeling the spread of influence on the blogosphere. In Proceedings of the 15th International World Wide Web Conference, 2006.
- [14] Joachims, T. , Granka, L. , Pan, B., Hembrooke, H, Gay, G., (2005). Accurately interpreting clickthrough data as implicit feedback, Proceedings of the 28th annual international ACM SIGIR conference.
- [15] L. Akritidis,D.Katsaros, and P. Bozanis, “Identifying influential bloggers: Time does matter,” in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., 2009, pp. 76–83.
- [16] Liping Deng & Allan H. K. Yuen (2010) EXPLORING THE ROLE OF ACADEMIC BLOGS IN A BLENDED COMMUNITY: AN INTEGRATIVE APPROACH
- [17] O'Donnell, M. (2006). Blogging as pedagogic practice: Artefact and ecology. *Asia Pacific Media Educator*, 17, 5–19.
- [18] Richardson, M. and Domingos, P. 2001. Mining the network value of customers. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.
- [19] Richardson,W. (2006). Blogs, wikis, podcasts, and other powerful web tools for classrooms. Thousand Oaks, Calif.: Corwin Press.
- [20] Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- [21] Wrede, O. (2003). Weblogs and discourse: Weblogs as a transformational technology for higher education and academic research. In Blogtalk conference. Paper presented at the Blogtalk Conference, Vienna.