

Spring 2012

IDENTIFYING INFLUENTIAL BLOGGERS

SivaNaga Prasad Shola
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

Recommended Citation

Shola, SivaNaga Prasad, "IDENTIFYING INFLUENTIAL BLOGGERS" (2012). *Master's Projects*. 322.
DOI: <https://doi.org/10.31979/etd.z25y-57jy>
https://scholarworks.sjsu.edu/etd_projects/322

This Master's Project is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Projects by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.



IDENTIFYING INFLUENTIAL BLOGGERS

A Thesis

Presented to

The Faculty of the Department of Computer Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

SivaNaga Prasad Shola

May 2012

© 2012

SivaNaga Prasad Shola

ALL RIGHTS RESERVED

The Designated Thesis Committee Approves the Thesis Titled

IDENTIFYING INFLUENTIAL BLOGGERS

by

SivaNaga Prasad Shola

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

SAN JOSÉ STATE UNIVERSITY

May 2012

| | | |
|--------------|--------------------------------|------|
| Dr. Teng Moh | Department of Computer Science | Date |
|--------------|--------------------------------|------|

| | | |
|------------------|--------------------------------|------|
| Prof. Frank Butt | Department of Computer Science | Date |
|------------------|--------------------------------|------|

| | | |
|----------------|--------------------------------|------|
| Dr. Sami Khuri | Department of Computer Science | Date |
|----------------|--------------------------------|------|

ABSTRACT

IDENTIFYING INFLUENTIAL BLOGGERS

This project 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. There exists preliminary models that attempted to solve the problem but they lack some of important aspects of the blogosphere. In this project we try to combine and improve the methodologies and ideas present in the previous models. We have introduced a new influence factor, which is a combination of facebook likes and shares, into the literature that can further evaluate blog posts efficiently. To capture the true influence we have mined each comment on a blog post to know the tone (agree or disagree) of the commenter, and evaluated the post accordingly. We also utilized the novelty parameter in our approach which represents the quality or goodness of the blog post. The experiment results proved that our approach is able to capture the true influence of each blog post and rank bloggers efficiently.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor, Dr. Teng Moh for his invaluable support, encouragement, and guidance. I thank my seniors and fellow students for their technical assistance whenever I was stuck. I would like to thank my committee members Dr. Sami Khuri and Frank Butt for their valuable suggestions, inputs and incredible patience in supporting me. I am also indebted to the developers that contributed countless programs and tools to ``Open Source" programming community which I have used to analyze and evaluate my project. Finally the special thanks go to my friends and family for their love, trust and encouragement.

TABLE OF CONTENTS

| | |
|---|----|
| 1. Introduction | 8 |
| 2. Literature Review | 12 |
| 2.1 Influential Flow Method | 12 |
| 2.2 Productivity and temporal aspects | 14 |
| 2.3 Productive and Influential Bloggers | 17 |
| 3. My Approach | 21 |
| 3.1 Influential Factors | 22 |
| 3.2 Influential Bloggers | 23 |
| 4. Experiments & Results | 27 |
| 5. Conclusion & Future work | 53 |
| 6. References | 55 |

List of Tables

| | |
|--|----|
| Table 1: Notion used in MEIBI & MEIBIX model..... | 15 |
| Table 2: Notion used in “Productive and Influential bloggers” model | 18 |
| Table 3: Notion used in my approach..... | 23 |
| Table 4: Authors with respective total number of posts, comments, and inlinks..... | 27 |
| Table 5: Individual blog post scores when facebook count is introduced..... | 31 |
| Table 6: BI vs I-FBCount | 33 |
| Table 7: Individual blog post scores with “mining comments” technique..... | 34 |
| Table 8: Classification of comments of blog post 12779.. .. | 37 |
| Table 9: Classification of comments of blog post 11335... .. | 43 |
| Table 10: Statistics of mined comments..... | 48 |
| Table 11: BI vs. I-Mine comments..... | 49 |
| Table 12: BI vs. I-(FB+Mine comments) | 50 |
| Table 13: BI vs. I-My Approach | 51 |
| Table 14: BI vs. I-indexes..... | 52 |

List of Figures

| | |
|---|----|
| Figure 1: Sample blog post content and parameters..... | 22 |
| Figure 2: Cropped blog post (ID 14890) with influential factors..... | 32 |
| Figure 3: Cropped blog post (ID 11557) with influential factors..... | 32 |
| Figure 4: Cropped blog post (ID 12779) with influential factors and comments..... | 35 |
| Figure 5: Cropped blog post (ID 11335) with influential factors and comments..... | 36 |

1.0 INTRODUCTION

Blogs have been the major communication channel for many internet users for the past couple of years. Bloggers are sharing their ideas, opinions, and experience with others through blog posts. A blog post might comprise of text, image, video, and links other blogs or web pages. Many users visit blog sites to read the posts and comment on them. The people who get attracted or convinced by a post idea will eventually choose their decisions in favor of the ideas borrowed from blog posts. Analysts say that one in ten in America would suggest others where to eat, what to buy, who to vote etc. We call the former as influential, as his/her influence affecting the decision of individuals. Similarly the bloggers who write influential blogs can influence other people, who read or visit their posts. Such bloggers are called influential bloggers.

A blogosphere is a virtual world which comprises of all the blogs and their connections in the universe. We can visualize blogosphere as a network of interconnections between bloggers and their blogs. The behavior of the blogosphere changes dynamically with time. There are two different types of blog sites in a blogosphere, community web blog sites and individual web blog sites. Individual web blog sites are more like personal accounts, journals, and diaries. They are owned and maintained by individual users. They almost have negligible group interaction and negligible collective knowledge, so very less people visit them. The community blog sites are more like forums and discussion boards. These are maintained by group of people and have collective dynamic information. Therefore more number of users visits community blogs.

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.

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. Agarwal, N., Liu, H., Tang, H., & Yu, P. S. have introduced the problem of identifying influential bloggers for the first time in 2008 [1]. 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 Akritidis, L., Katsaros, D., & Bozanis, P. 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 [6]. In 2011 the same authors has modified their model to capture temporal (time) patterns of blogosphere [2]. 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). Cai, Y., and Chen, Y 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 [5].

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. For example, a comment can be positive, negative, or neutral, which means a user can convey his/her consent towards the post in three different ways. A positive comment indicates the reader's agreement and negative comment indicates disagreement and finding such a tone is very important when dealing with the influence exerted by the blog post. Also when scoring a blog post, the score should reflect the quality and goodness of the blog post. The model should learn the quality and goodness of the blog post.

Considering all the aspects discussed above, in our approach we introduced a new influential factor FBCount, which is a combination of facebook likes & shares. 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 our algorithm so that the score of each blog post reflect quality and goodness of blog post.

The rest of paper is organized as the following: we will discuss preliminary models in section 2. Section 3 explains the new factors and approaches used in our approach. Experimental studies and results are described in Section 4 and finally in Section 5 we discuss conclusions and future work.

2.0 LITERATURE REVIEW

In this section we will discuss three different solutions proposed to solve the problem of identifying influential bloggers. First, Influence Flow Method in which authors Agarwal N., Liu H., Tang H., & Yu P. S. have utilized Inlinks, Comments, Outlinks, and the length of the blog post as influential factors to identify influential bloggers. This method has set up ground truth for the problem. Second, Akritidis, L., Katsaros, D., & Bozanis, P. have introduced two metrics MEIBI (Metric for Evaluating and Identifying Blogger's Influence) & MEIBIX (MEIBI extended) indexes to evaluate influence by considering temporal aspects and productivity along with inlinks and comments. They have ignored parameters such as the length of the post and outlinks of posts in the second model. Third, the same authors Akritidis, L., Katsaros, D., & Bozanis, P. have modified the metrics MEIBI & MEIBIX indexes to BP & BI indexes to differentiate between productive and influential bloggers. The following subsections elaborate various aspects of the three models in detail.

2.1 INFLUENCE FLOW METHOD

Agarwal N., Liu H., Tang H., & Yu P. S. have first introduced the problem of identifying influential bloggers to the literature. They proposed "Influence flow method" to find out the influential bloggers in a web community. According to them in the real world an influential person is one s/he is well recognized by others, often does good work that is novel, well received and creates some activity in society. The authors considered these social parameters such as recognition, activity generation, novelty, and eloquence as influential factors for scoring each blog post of a blogger. But how can we relate these

social gestures to blogosphere? Each blog post in blogosphere is associated with parameters like inlinks, outlinks, comments, length of the post etc. We can use these parameters to quantify the influence. The blog post parameters can be related to social gestures as following [1].

Recognition: How well people have received the idea? This can be measured by the inlinks received by the blog post.

Activity Generation: The participatory activity generated i.e. the comments received by the blog posts. More the comments more the activity generated.

Novelty: Conveys a new idea shared in the blog post. Can be measured through the outlinks used in the post. More the outlinks less is the novelty.

Eloquence: The goodness of the blog post can be measured by the length of the post.

Finally the Influence Score for a blog is defined as a function of social gestures. These parameters help the model to trace the intuitive influence flow among the blogs. The influence flow of blog post p is calculated as

$$Influence\ flow(p) = w_{in} \sum_{m=1}^{|i|} I(p_m) - w_{out} \sum_{n=1}^{|o|} I(p_n) \quad (eq. 1)$$

where w_{in} , and w_{out} are the weights that can be used to adjust the contribution of incoming and outgoing influence respectively. And $I(p_m)$, $I(p_n)$ are the influence flow based on inlinks and outlinks, respectively.

For a blog post p , that has i inlinks and θ outlinks, the influence score can be calculated by the following equation:

$$I(p) = w(\lambda) \times (w_{com}\gamma_p + Influence\ flow(p)) \quad (eq. 2)$$

where $w(\lambda)$ is the weight function chose based on the length of a blog post, γ_p denotes the number of comments received by p , w_{com} is weight parameter used to normalize the contribution of the number of comments.

The influence score is calculated for each blog post P that belongs to blogger B . The maximum of all influence scores of Blogger B is represented as the influenceIndex of blogger B .

$$influenceIndex(B) = MAX(I(P_i)) \text{ where } 1 \leq i \leq N.$$

Where N is the total number of posts written by blogger B . The bloggers who have achieved top influenceIndex values are the most influential ones of total bloggers.

2.2 PROCUTIVITY AND TEMPORAL ASPECTS

The influential flow method was able to set some ground truth for influential bloggers problem but model has couple of drawbacks with it. Isolating a single post to identify influential bloggers made this approach over simplistic and it overlooks productivity of the bloggers. The output of the model highly dependent on the user defined weights. Most importantly, this model ignores temporal aspects of blogs which is very crucial for

Blogosphere: Time. A blog that has influenced most today might not be influential after two months. i.e. proposed model should be able to decay the influence score with time. Considering the above issues Akritidis, L., Katsaros, D., & Bozanis, P. have proposed a new model that investigates the problem by two easily computable blogger scoring methods. These methods define two metrics MEIBI & MEIBIX that not only includes temporal aspects but also takes care of productivity of the blogger [6].

The influential factors that have been considered for this model are inlinks, comments, productivity, and time i.e. age of the blog post and inlinks to the blog post. The outlinks factor was ignored in this model. Bloggers use outlinks to support or define their ideas. But it doesn't mean it works against the novelty of the blog post.

Table 1: Notion used in MEIBI & MEIBIX model

| Symbol | Meaning |
|------------------------------------|---|
| BP(j) | Blogger j's blog posts. |
| Bp_j (i) | Blogger j's post i. |
| C_j (i) | Number of Comments on post i of blogger j. |
| R_j (i) | Inlinks referring post i of blogger j. |
| $\Delta TP_j(i)$ | Difference between current date and posted date of post i of blogger j. |
| $\Delta TP(i)$ | Difference between current date and date referral post i submitted. |

MEIBI: (Metric for Evaluating and Identifying Blogger's Influence)

MEIBI metric takes into the account the number of comments, inlinks received by blog post, along with the date of publication. A blog post posted today will become outdated after few months.

The score $S_j^m(i)$ for each blog post i that belong to blogger j is calculated as follows

$$S_j^m(i) = \gamma(|C(i) + 1| \times (\Delta TP_j(i) + 1)^{-\delta} |R_j(i)|) \quad (eq. 3)$$

Where γ and δ are constants assigned values 4 and 1 respectively. After calculating score $S_j^m(i)$ for all blog posts the model assigns Blogger j a MEIBI index value equal to m , if and only if m blog posts of blogger j , has got a score $\geq m$ and remaining $BP(j) - m$ posts got a score $< m$.

This MEIBI takes care of both productivity and influence of a blogger. The top k MEIBI index bloggers are the top k -influential bloggers.

MEIBIX: (MEIBIX extended)

Instead of considering the age of the post we can decay the score for each blog, based on the age of the inlinks received. For each incoming link we assign a weight that depends on the age of the inlink. The idea is formulated and a new score $S_j^x(i)$ is calculated as follows:

$$S_j^x(i) = \gamma(|C(i) + 1| \times \sum_{\forall x \in R_j(i)} (\Delta TP(x) + 1)^{-\delta}) \quad (eq. 4)$$

After calculating score $S_j^x(i)$ for all blog posts the model assigns Blogger j a MEIBIX index value equal to x , if and only if x blog posts of blogger j , has got a score $\geq x$ and remaining $BP(j) - x$ posts got a score $< x$.

The model does not depend on any user defined weights, and evaluates both blog posts and bloggers effectively. A blogger will be influential if and only if he/she blogs influential posts very often.

2.3 PRODUCTIVE AND INFLUENTIAL BLOGGERS

The MEIBI & MEIBIX indexes have considered both temporal aspects and productivity of the blogs, and it was able to show significant improvement over the preliminary model “Influence Flow Method”. Though these metrics were able to define unique index values to each blogger, they fail to differentiate between productivity and influence exactly. Considering this drawback the same authors Akritidis, L., Katsaros, D., & Bozanis, P. have refined their previous model in Feb 2011 to differentiate between productive bloggers and influential bloggers. In their defense productive bloggers are those who posts very often i.e. active bloggers, and influence bloggers are those whose writing’s impact others decisions. Many influential bloggers are also active bloggers. Although productivity and influence do not coincide, the influence of a blogger decays with time. An influential blogger is recognized as such if he/she has written influential posts recently or if the posts have had an impact recently [2]. The impact can occur as described in the next subsection.

Proximal Impact: Indicates the blogger's influence on regular users/members of the community. This is measured through the comments received on a post.

Wide Impact: Indicates the blogger's influence outside the community. This is measured through inlinks used by other bloggers outside the community.

Table 2: Notion used in “Productive and Influential bloggers” model

| Symbol | Meaning |
|-------------|---|
| P_t^j | Blogger j's Productivity at instance t |
| I_t^j | Blogger j's Influence at instance t |
| N^j | Blogger j's blog posts |
| n_i^j | Blogger j's post i. |
| C_i^j | Comments on blog post i of blogger j. |
| R_i^j | Inlinks received by blog post i of blogger j. |
| L_i^j | Length of blog post i of blogger j |
| L | Average length of blog post of blogger. |
| T | Time variable |
| $t_{i,p}^j$ | Time stamp of post i of blogger j |
| $t_{x,l}^j$ | Time stamp of post x referring blogger j |
| $t_{x,c}^j$ | Comment x's time stamp on post of blogger j. |

Productivity Calculation:

For each blog post i of blogger j , a score $U_{i,p}^j(t)$ is measured as

$$U_{i,p}^j(t) = \gamma \frac{L_i^j}{L} \left(\frac{\theta}{t - t_{i,p}^j + \theta} \right)^\delta \quad (eq. 5)$$

where γ , θ , and δ are predefined constants, whose values are 100, 86400, and 1, respectively. The score $U_{i,p}^j$ connects the post with its age and its score decays with time. And the blogger's productivity index is defined next.

After calculating score $U_{i,p}^j(t)$ for all blog posts the model assigns Blogger j a BP- 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 $< P_t^j$.

BP-index defines productive bloggers in terms of recent lengthy blog posts submitted by a blogger.

Influence Calculation:

For each blog post i of blogger j , the dual nature of influence i.e. proximal impact and wide impact can be quantified as follows.

$$V_{i,p}^j(t) = w_l \sum_{\forall x \in R_i^j} \left(\frac{\theta}{t - t_{x,l}^j + \theta} \right)^\delta + w_c \sum_{\forall x \in C_i^j} \left(\frac{\theta}{t - t_{x,c}^j + \theta} \right)^\delta \quad (eq. 6)$$

where $w_l = 100$ and $w_c = 10$. The score $V_{i,p}^j(t)$ connects the comments and inlinks received by post with corresponding age and its score decays with time. And the blogger's influence index is defined as following

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

The BI-Index defines influential bloggers in terms of recent influence. Also it appreciates blog posts that receive many inlinks and comments.

3.0 My Approach


In this section, we will discuss different influential factors considered in our approach and the procedure followed to find influential bloggers in a blog community. Each blogger in a web blog community maintains a blog space, where s/he blogs multiple posts time to time. For each blogger, we calculate the influential score for all the posts and define Influence index (I-index) based on the scores obtained. In our approach we have taken ideas and methodologies in the existing work and modified inefficient areas of the algorithm for better model. Figure 1 depicts one of the blog post from endagnet web blog community. As highlighted in the figure, each blog consists of title, post body, author name, date of post, comments, number of flikes, fshares, twitter shares etc.

NEWS HUBS GALLERIES VIDEOS PODCASTS TOPICS REVIEWS STORE FOLLOW SUBSCRIBE ABOUT / FAQ TIPUS

\$10 USB power outlet leaves no plug behind

By Donald Melanson posted December 5th 2009 10:28AM

Author name and date of post



Don't freak out or anything. But all that time you spent building a **DIY in-wall USB charger** may have been for naught. Of course, you've gained a useful learning experience and potentially gotten a lesson in the dangers of electricity, but you could have simply ordered this TruPower UCS outlet from FastMac, which packs the same two USB charging ports as the DIY outlet, along with two standard power outlets for your other power-hungry devices. Best of all, it only costs the same ten bucks that the DIY option would have set you back, but it looks like it won't start shipping until sometime early next year.

number of comments

SOURCE **FastMac** facebook likes

1,118 people like this. Be the first of your friends.

f-share, twitter share, G+1's

DISCUSS 121

7 1 0 +1 29

TAGS ac, ac outlet, AcOutlet, fastmac, outlet, power, power outlet, PowerOutlet, trupower, trupower ucs, TrupowerUcs, usb, usb charger, usb outlet, UsbCharger, UsbOutlet

FEATURED **BREAKING**

ask Engadget: best sub-\$150 mechanical keyboard? 48 minutes ago

Mobile Miscellany: week of April 16th, 2012 3 hours ago

HTC talks smartphone design: from inception to final product 1 day ago

ASUS Transformer Prime GPS Extension Kit hands-on 1 day ago

Hands-on with HTC's colorful One X cases 1 day ago

COMMENTS

SUBSCRIBE

Comments: on | off

SEE ALL

nuck44

Posted Dec 5th 2009 10:31AM

HIGHEST RANKED

Darn you, engadget!! I don't need more "lessons in the dangers of electricity."

REPORT

+ -

REPLY

This comment has been down-ranked into oblivion. [View comment](#)

eugeneaction

Posted Dec 6th 2009 2:25AM

HIGHEST RANKED

@(Unverified) ACTUALLY it looks like they are playing pong, and they are scared because the puck went wayyyy up there by his ear...

REPORT

+ -

REPLY

bstump

Posted Feb 15th 2010 9:23AM

NEUTRAL

@nuck44 Unfortunately engadget, looks like they are retailing this for twice that, \$20 now. That sucks.

REPORT

+ -

REPLY

dkbraves

Posted Dec 5th 2009 10:32AM

HIGHEST RANKED

if u guys can, please remind us when this ships. U guys should hire samuel l. jackson to say "engadget motha f*cka"

REPORT

+ -

REPLY

decypher44

Posted Dec 5th 2009 10:36AM

NEUTRAL

@dkbraves

I agree. Please post another article when this is released! I'd like to have one (or three!), but want to wait until install video is up and hear back from initial reviews.

REPORT

+ -

REPLY

jawxx

Posted Dec 5th 2009 11:11AM

NEUTRAL

@dkbraves Hal! That'd be rad.

REPORT

+ -

REPLY

Chris Dunninga

Posted Dec 5th 2009 11:20AM

NEUTRAL

REPORT

FACEBOOK ACTIVITY

Hands-on with Smart Monitor's SmartWatch, the seizure sensing wristwatch

618 people recommend this.

HTC Rezound review

307 people recommend this.

Jay Leno's Chevy Volt still has original tank of gas, 11,000 miles later

586 people recommend this.

Man skydives from 13 miles above Earth, isn't satisfied (video)

1,566 people recommend this.

Dragon's Lair landing on XBLA with Kinect support, lets you act out your princess-saving fantasies

204 people recommend this.

Facebook social plugin

Sponsored Links

Orange Savings Account

No fees. No minimums. Nothing standing in your way. [Learn More.](#)

[www.ingdirect.com](#)

Figure 1: Sample blog post content and parameters

3.1 Influential factors:

Without any doubt comments and Inlinks are primary factors that represent the influence exerted or attention received by blog post. 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 Method, 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 [1]. 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 [6]. We haven'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

22

latest parameters associated with blog posts such as facebook likes, G+ 1's, and sharings 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.

3.2 Influential Bloggers

Table 3: Notion used in my approach

| Symbol | Meaning |
|--------------|---|
| I_t^j | Blogger j's Influence at instance t |
| $BP(j)$ | Blogger j's blog posts |
| n_i^j | Blogger j's post i. |
| Fb_i^j | facebook count of ith blog of blogger j. |
| C_i^j | Comments on blog post i of blogger j. |
| R_i^j | Inlinks received by blog post i of blogger j. |
| $Inf(b_c)$ | Influence of the blogger that issued comment c. |
| $SF(b_c, x)$ | Sentiment factor of blogger b_c on comment x. |
| $TC(b_c)$ | Total number of comments of the blogger b_c |
| T | Time variable |
| $t_{i,p}^j$ | Time stamp of post i of blogger j |
| $t_{x,l}^j$ | Time stamp of post x referring blogger j |
| $t_{x,c}^j$ | Comment x's time stamp on post of blogger j. |

Before reading random articles, news, web blog or a personal profile, many of us tend to look at the number of hits (visitors), likes (fblikes, 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 sharings can be used as influential factors for the current literature. The act of clicking on fblike, G +1 or sharing on social web sites most likely indicates the interest or approval of the content read. The number of views, flikes, 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 overfitting 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. So given facebook count Fb_i^j , a time varying Fb-score for a blog i of blogger j is given by

$$Fb_{i,p}^j(t) = w_{fb} * Fb_i^j * \left(\frac{\theta}{t - t_{i,p}^j + \theta} \right)^\delta \quad (\text{eq. 7})$$

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} is the weight parameter of facebook count factor whose value is 5.

When we are considering the activity generated (“Post-Reply relationship” [5]) 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 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 we ought to find the person who issued it and his consent towards the blog post. We use this information to find the comment score achieved by the blogpost. 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

$$C_{i,p}^j(t) = w_c \sum_{\forall x \in C_i^j} \left(\frac{Inf(b_c) * SF(b_c, x)}{TC(b_c)} \right) * \left(\frac{\theta}{t - t_{x,c}^j + \theta} \right)^\delta \quad (eq. 8)$$

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

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

$$N_i^j = w_n \left(1 - \left(\frac{\text{number of words matching}}{\text{total no of words}} \right) - w_o * \text{outlinks} \right) \quad (\text{eq. 9})$$

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:

$$PS_{i,p}^j = (Fb_{i,p}^j + C_{i,p}^j + INLK_{i,p}^j + N_i^j) \quad (\text{eq. 10})$$

where $INLK_{i,p}^j$ is the inlink score received by the blog post i of blogger j given by [6],

$$INLK_{i,p}^j(t) = w_l \sum_{x \in R_i^j} \left(\frac{\theta}{t - t_{x,l}^j + \theta} \right)^\delta \quad (\text{eq. 11})$$

w_l is the inlink weight parameter whose value is 75.

After calculating the post score $PS_{i,p}^j(t)$ for all blog posts using eq.10 the model assigns Blogger j a Influence- index value at a given instance t , equal to I_t^j , if and only if I_t^j blog posts of blogger j , has got a score $\geq I_t^j$ and remaining $BP(j) - I_t^j$ posts got a score $< I_t^j$.

4.0 Experiments & Results

In this section we discuss how the experiments are carried out and the results obtained. And then we compare our results against the previous model. First, we need blog post dataset with necessary fields and parameters. We have requested the authors of “Productive and Influential Bloggers [2]” for dataset and have received a small dataset containing blog posts from engadget web blog community. Table 4 gives the overall picture of the dataset we have received from engadget web blog community.

Table 4: Authors with respective total number of posts, comments, and inlinks.

| Author ID | Author Name | Posts | Total Comments | Total Inlinks |
|-----------|-------------------|-------|----------------|---------------|
| 1 | Joseph L. Flatley | 1352 | 60449 | 7543 |
| 2 | Thomas Ricker | 4798 | 178138 | 36175 |
| 3 | Paul Miller | 5000 | 209425 | 30946 |
| 4 | Darren Murph | 11555 | 630979 | 63028 |
| 5 | Ben Drawbaugh | 218 | 6426 | 1313 |
| 6 | Tim Stevens | 1225 | 61529 | 7132 |
| 7 | Ross Miller | 1385 | 68589 | 9722 |
| 8 | Nilay Patel | 3091 | 181430 | 26381 |
| 9 | Chris Ziegler | 1997 | 130603 | 14560 |
| 10 | Donald Melanson | 4856 | 147420 | 21878 |
| 11 | Laura June | 1060 | 536362 | 6111 |
| 12 | Inhabitat | 8 | 188 | 54 |
| 13 | Vladislav Savov | 944 | 46431 | 8277 |
| 14 | Joshua Topolsky | 2057 | 264627 | 20858 |

| | | | | |
|----|----------------------|------|--------|-------|
| 15 | Chad Mumm | 11 | 598 | 211 |
| 16 | Richard Lawler | 506 | 17279 | 3161 |
| 17 | Ross Rubin | 229 | 4098 | 715 |
| 18 | Trent Wolbe | 104 | 1770 | 269 |
| 19 | Steven Kim | 97 | 1962 | 374 |
| 20 | Michael Gartenberg | 38 | 4932 | 271 |
| 21 | Joanna Stern | 133 | 6075 | 1363 |
| 22 | Christopher Grant | 85 | 1866 | 386 |
| 23 | Jacob Schulman | 102 | 30082 | 618 |
| 24 | Sean Cooper | 232 | 7768 | 1549 |
| 25 | Sean Hollister | 53 | 2243 | 428 |
| 26 | Richard Lai | 134 | 7636 | 1282 |
| 27 | Engadget staff | 38 | 3460 | 359 |
| 28 | Ryan Block | 5643 | 576923 | 25251 |
| 29 | Joshua Fruhlinger | 666 | 18389 | 1825 |
| 30 | Ariel Waldman | 18 | 640 | 100 |
| 31 | Kevin Wong | 4 | 266 | 21 |
| 32 | Jose Andrade | 2 | 33 | 14 |
| 33 | Mark Priestap | 1 | 0 | 0 |
| 34 | Benjamin Heckendorn | 19 | 1224 | 375 |
| 35 | Samuel Axon | 76 | 3297 | 357 |
| 36 | Stephanie Patterson | 34 | 1238 | 204 |
| 37 | Jeremy Toeman | 4 | 157 | 21 |
| 38 | Dante Cesa | 42 | 168714 | 139 |
| 39 | Terrence O'Brien | 2 | 124 | 7 |
| 40 | Evan Blass | 2269 | 62473 | 7951 |
| 41 | Marc Weber Tobias | 11 | 512 | 110 |
| 42 | Erik Hanson | 99 | 1669 | 385 |
| 43 | Barb Dybwad | 1523 | 33634 | 1442 |
| 44 | Sam Sheffer | 2 | 92 | 8 |
| 45 | Conrad Quilty-Harper | 577 | 16994 | 3443 |
| 46 | Peter Rojas | 5897 | 98026 | 4501 |
| 47 | Matt Burns | 15 | 257 | 45 |
| 48 | Brian White | 19 | 273 | 62 |
| 49 | Niall Kennedy | 2 | 65 | 17 |
| 50 | Will O'Brien | 42 | 1537 | 325 |
| 51 | Jeannie Choe | 93 | 1544 | 491 |
| 52 | Adam Nielson | 10 | 2466 | 43 |
| 53 | Josh Fruhlinger | 16 | 277 | 54 |
| 54 | Omar McFarlane | 48 | 476 | 134 |

| | | | | |
|----|--------------------|------|-------|------|
| 55 | Randall Bennett | 49 | 697 | 150 |
| 56 | Cyrus Farivar | 710 | 13190 | 2672 |
| 57 | Stephen Speicher | 12 | 238 | 35 |
| 58 | Jesse D. Lewin | 46 | 326 | 27 |
| 59 | Anand Shimpi | 2 | 206 | 38 |
| 60 | Tom Whitwell | 16 | 197 | 39 |
| 61 | Paul Boutin | 13 | 890 | 265 |
| 62 | Dave Zatz | 5 | 151 | 59 |
| 63 | Stan Horaczek | 65 | 1011 | 291 |
| 64 | Alberto Ballestin | 3 | 82 | 12 |
| 65 | Jeff Wilson | 1 | 10 | 3 |
| 66 | Marc Perton | 2122 | 35721 | 3074 |
| 67 | Andres Palmiter | 6 | 132 | 14 |
| 68 | Fabienne Serriere | 15 | 493 | 35 |
| 69 | Matthew Maier | 1 | 28 | 1 |
| 70 | Kevin C. Tofel | 57 | 545 | 57 |
| 71 | Eliot Phillips | 9 | 417 | 69 |
| 72 | Liam McNulty | 112 | 1152 | 70 |
| 73 | Jonathan Hayter | 1 | 21 | 2 |
| 74 | James Ransom-Wiley | 10 | 57 | 1 |
| 75 | Gareth Edwards | 221 | 691 | 67 |
| 76 | Dan Choi | 1 | 132 | 0 |
| 77 | Alberto Escarlate | 28 | 95 | 16 |
| 78 | Ben Zackheim | 18 | 54 | 0 |
| 79 | Vladimir Cole | 1 | 204 | 1 |
| 80 | Jason Calacanis | 7 | 1071 | 6 |
| 81 | Emily Conrad | 6 | 46 | 4 |
| 82 | Jason Striegel | 3 | 144 | 7 |
| 83 | Dan Wu | 334 | 2604 | 73 |
| 84 | David Thomas | 2 | 17 | 1 |
| 85 | Eric Lin | 81 | 1153 | 18 |
| 86 | Mike Outmesguine | 9 | 111 | 1 |
| 87 | Phillip Torrone | 401 | 5588 | 335 |
| 88 | Susan Mernit | 30 | 273 | 16 |
| 89 | Katie Fehrenbacher | 189 | 573 | 75 |
| 90 | Simon Spagnoletti | 234 | 617 | 46 |
| 91 | Heather Sparks | 15 | 38 | 3 |
| 92 | Mia Kim | 8 | 47 | 2 |
| 93 | Joshua Klein | 74 | 132 | 6 |

The dataset we have received has all the necessary parameters like post title, author, postdate, post content, comments, comment dates, inlinks and inlinks dates etc. We web crawled through each blog post for the additional parameters that are missing for our experiment like fblikes, twitter shares, G +1s etc. We considered the previous model “Productive and Influential bloggers [2]” as our prototype and introduced our parameters one by one. The experiment was carried out in four different stages in order to avoid over fitting error problem i.e. adding too many parameters into model can jeopardize the overall performance of the system. So first, we introduced the new influential factor (facebook count) into prototype model and identified the changes in the result. In the second stage we have replaced “comments score” part in eq. 6 (second half) with our “mining comments” technique (eq. 8) as proposed in section 3.2. In the third stage we applied both facebook count and “mining comments” technique into prototype model. Lastly we incorporated our novelty parameter along with facebook count and “mining comments” technique into previous model which is our final model.

To check how our approach works we need to look at individual blog post influential score and their changes with the each modification. Table 5 shows how the introduction of facebook count has affected the influential score between different blog posts.

Table 5: Individual blog post scores when facebook count is introduced into model

| Post ID | Post Title | Post Author | Post Comments | Post Date | Post Inlinks | Post FBCount | Score_U | Score_V | PS_FB |
|---------|--|-----------------|---------------|----------------|--------------|--------------|---------|---------|-------|
| 11224 | Engadget giveaway: win some Beatles MusicSkins | Laura June | 550 | 3/22/2010 0:00 | 5 | 14 | 18 | 781 | 799 |
| 11113 | Official: Apple now offering iPhones contract free (updated: not unlocked) | Joshua Topolsky | 311 | 3/22/2010 0:00 | 29 | 335 | 16 | 768 | 1186 |
| 13557 | Official: Palm Pre Plus and Pixi Plus for AT&T 'in the coming months' | Joshua Topolsky | 220 | 3/22/2010 0:00 | 23 | 459 | 93 | 694 | 1268 |
| 14890 | Kindle for iPad and tablets makes the scene | Thomas Ricker | 129 | 3/22/2010 0:00 | 32 | 69 | 19 | 569 | 655 |
| 11557 | Google.cn now rerouting to Hong Kong domain, an 'entirely legal' workaround to censorship woes | Ross Miller | 112 | 3/22/2010 0:00 | 26 | 310 | 31 | 519 | 907 |
| 10445 | Sprint to announce 'groundbreaking new device' (HTC Supersonic?) tomorrow | Ross Miller | 161 | 3/22/2010 0:00 | 18 | 81 | 12 | 447 | 548 |
| 11000 | Samsung to announce Galaxy S smartphone, content initiatives this week | Chris Ziegler | 20 | 3/22/2010 0:00 | 31 | 8 | 21 | 437 | 447 |
| 13890 | Early reports show IE not faring well in the post-ballot screen days | Tim Stevens | 182 | 3/22/2010 0:00 | 15 | 45 | 9 | 418 | 475 |

In Table 5 columns Score_U & Score_V represent the productive & influence scores of posts according to “productive and influential bloggers” model, and PS_FB represents the influence score of posts according to resultant model when facebook count is introduced. From left to right the columns of the Table 5 represents post id, title, post author, comments received by post, postdate, inlinks received by post, post FBCount etc. We have taken random blog posts posted on 03/22/2010 to monitor the changes when facebook count is introduced. Since the date for all the blog posts is same the influence scores of blog posts entirely depends on other parameters comments, inlinks, and facebook count. Consider the influence scores (score_V, PS_FB) for blog posts with post_id 14890 and 11557. For both posts there is a subtle difference in number of comments and inlinks but blog post 11557 has greater FBCount than the post 11890,

which has favored the influence score of 11557 to be more in resultant model unlike “productive and influential blogger” model.

Figure 2 & Figure 3 depict cropped blog posts with post_ids 14890 & 11557 respectively. As showed in the figures, we can see the influential factors associated with the blog posts such as author name, date of post, number of comments, fblikes, and facebook shares etc.



Figure 2: Cropped blog post (ID 14890) with influential factors.

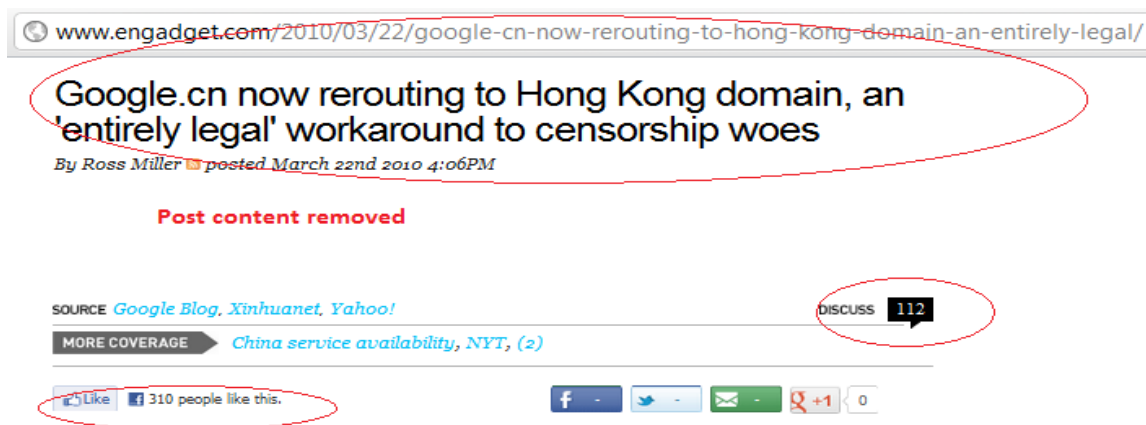


Figure 3: Cropped blog post (ID 11557) with influential factors.

Similarly for each blogger, the resultant model scores each blog post and defines an Influence index from the scores obtained. This Influence index has been used to rank the bloggers among other bloggers in the community. Table 6 depicts how the final ranking of the bloggers has been changed when facebook count is introduced into the model. In Table 6 BP indicates BP-index, BI indicates BI-index of the blogger according to “productive and influential bloggers [2]” and I-FBCount indicates Influence-index of blogger according to resultant model. BP-index is defined out of all score_U(measures how often blogger produces lengthy posts)s, which is not related to influence, so we compare our resultant model with BI-Index. BI-index is defined out of all score_Vs which measures the influence of blog post. Similarly I-FBCount is defined out of all PS_FB of blogger.

Table 6: BI vs I-FBCount

| Author ID | Author Name | BP | BI | I - FBCount |
|-----------|-------------------|----|-----|-------------|
| 4 | Darren Murph | 23 | 114 | 124 |
| 9 | Chris Ziegler | 18 | 94 | 101 |
| 13 | Vladislav Savov | 17 | 91 | 111 |
| 11 | Laura June | 13 | 88 | 98 |
| 8 | Nilay Patel | 13 | 74 | 87 |
| 2 | Thomas Ricker | 13 | 72 | 83 |
| 14 | Joshua Topolsky | 16 | 72 | 76 |
| 3 | Paul Miller | 20 | 67 | 78 |
| 10 | Donald Melanson | 12 | 64 | 74 |
| 6 | Tim Stevens | 14 | 63 | 74 |
| 1 | Joseph L. Flatley | 16 | 56 | 66 |
| 7 | Ross Miller | 15 | 54 | 63 |
| 21 | Joanna Stern | 19 | 42 | 49 |
| 28 | Ryan Block | 3 | 39 | 39 |
| 26 | Richard Lai | 9 | 36 | 40 |

As highlighted in Table 6 the resultant model ranked V Savov second and C Ziegler third. Similarly J Topolsky and P miller ranks are interchanged and R Block and R Lai ranks are also interchanged.

In the second stage of experiment we applied “mining comments” method into the prototype model which has changed the scores of blog posts as following. In Table 7 the last column PS_Comment represent influence score of blog post when tone of the comment is introduced into the prototype model.

Table 7: Individual blog post scores with mining comments technique

| Post ID | Post Title | Post Author | Post Comments | Post Date | Post Inlinks | Score_U | Score_V | PS_Comment |
|---------|--|-------------------|---------------|----------------|--------------|---------|---------|------------|
| 11224 | Engadget giveaway: win some Beatles MusicSkins | Laura June | 550 | 3/22/2010 0:00 | 5 | 18 | 781 | 908 |
| 15112 | Nokia Nuron for T-Mobile review | Chris Ziegler | 112 | 3/22/2010 0:00 | 15 | 119 | 322 | 371 |
| 12779 | Tata Motors' Nano lights up the streets of Mumbai | Joseph L. Flatley | 114 | 3/22/2010 0:00 | 13 | 12 | 308 | 271 |
| 11335 | George Takei can't show you Sharp's fourth pixel, can still blow your mind | Richard Lawler | 114 | 3/22/2010 0:00 | 12 | 9 | 299 | 357 |
| 12890 | Spring Design Alex review | Joanna Stern | 52 | 3/22/2010 0:00 | 15 | 127 | 261 | 252 |
| 12112 | Impossible Project's Polaroid film goes on sale this week | Donald Melanson | 45 | 3/22/2010 0:00 | 15 | 8 | 256 | 242 |
| 12557 | Motorola's Android-powered i1 launching at CTIA | Sean Cooper | 56 | 3/22/2010 0:00 | 14 | 7 | 247 | 236 |
| 11779 | Motorola makes i1 official, melds Android and push-to-talk this summer on Sprint | Ross Miller | 61 | 3/22/2010 0:00 | 13 | 16 | 246 | 229 |

The blog posts with post_id 12779 and 11335 have almost same number of comments but the final influence score has been favored 11335 to be more in the resultant model (PS_Comment) unlike “productive and influential bloggers” model (Score_V).

www.engadget.com/2010/03/22/tata-motors-nano-lights-up-the-streets-of-mumbai/

Tata Motors' Nano lights up the streets of Mumbai

By Joseph L. Flatley posted March 22nd 2010 12:09PM

Post content removed

SOURCE [Indian Autos Blog](#)

MORE COVERAGE [Indian Autos Blog](#)

Like 469 people like this. Be the first of your friends.

DISCUSS 114

TAGS car, compact car, CompactCar, fire, india, Satish Sawant, SatishSawant, tata, tata motors, tata nano, TataMotors, TataNano, vehicle

COMMENTS

SUBSCRIBE Comments: on | off



spasewalkr Posted Mar 22nd 2010 12:12PM **HIGHEST RANKED**

theres some quality engineering right there

REPORT + - REPLY



TJK Posted Mar 22nd 2010 12:27PM **HIGHEST RANKED**

V for vendetta, i see whats going on here...

REPORT + - REPLY



tobsmonster2 Posted Mar 22nd 2010 12:35PM **HIGHEST RANKED**

@spasewalkr
yea, reminds me alot of when i overclocked my athlon x2 on a cheap asrock board =\

REPORT + - REPLY



Cy Starkman Posted Mar 22nd 2010 12:39PM **NEUTRAL**

@tobsmonster2

Imagine if that had been a gigabyte main board.
.. Or a packard bell! Think of the deveststion

REPORT + - REPLY



WilliamH Posted Mar 22nd 2010 12:40PM **NEUTRAL**

@spasewalkr i would like to see this thing with 20" rims. hahahahahaha

REPORT + - REPLY

Figure 4: Cropped blog post (ID 12779) with influential factors and comments

35

www.engadget.com/2010/03/22/george-takei-cant-show-you-sharps-fourth-pixel-can-still-blow/

George Takei can't show you Sharp's fourth pixel, can still blow your mind

By Richard Lawler  posted March 22nd 2010 5:21PM

Post content removed

No. of comments

SOURCE [YouTube](#)

Like  160 people like this.  1  0  0  +1 1

TAGS [advertisement](#), [aquos](#), [cmdr sulu](#), [CmdrSulu](#), [edge lit led](#), [EdgeLitLed](#), [george takei](#), [GeorgeTakei](#), [hdtv](#), [lcd](#), [led](#), [quad pixel](#), [quad-pixel](#), [QuadPixel](#), [quattron](#), [rgb](#), [sharp](#), [star trek](#), [StarTrek](#), [yellow](#)

COMMENTS SUBSCRIBE **Comments** Comments: on | off

**germangabriel** Posted Mar 22nd 2010 5:23PM NEUTRAL
haha wwiwa wow.

**Chris DPSN AggieCEO XBLThe Aggi** Posted Mar 22nd 2010 5:34PM
NEUTRAL
@davidmuful Indeed

**Blackstar** Posted Mar 22nd 2010 5:47PM NEUTRAL
@germangabriel
Now the other manufactures are going to grab the rest of the cast for their ads.
To bad Scotty's dead... he could come in saying "Captain, she can't take any more!"

Figure 5: Cropped blog post (ID 11335) with influential factors and comments

Figure 4 & Figure 5 depict cropped blog posts with post_ids 12779 & 11335, respectively. As the figures show, we can see the influential factors associated with the blog posts like author name, date of post, number of comments, fblikes, and facebook shares etc. We can also see the comments written by different users.

Now let's see how the comments have been classified in each blog post by our "mining comments" method. Table 8 & Table 9 show the classification of comments of posts 12779 & 11335 respectively. The columns Comment_body, comment_author, Tone represents the content of comment, the author that issued the comment and his/her tone towards the comment (evaluated by our model) respectively. For each comment the resultant model considers influence of comment_author along with his/her tone to calculate the comment score. The sentiment factor (SF in eq. 8) for each positive comment is +2, -1 for each negative comment and +1 for a neutral comment. This particular technique unlike prototype model captures the true acceptance of topic by mining comments and finding out commenters consent towards the post. Also the comments are scored based on the users issued the comments and their influence (eq. 8).

Table 8: Classification of comments of blog post 12779.

| Comment post id | Comment Body | Comment Author | Tone |
|-----------------|--|----------------|----------|
| 12779 | theres some quality engineering right there | Spasewalkr | Positive |
| 12779 | V for vendetta, i see whats going on here... | TjK | Neutral |
| 12779 | @spasewalkr yea, reminds me alot of when i overclocked my 37avele x2 on a cheap asrock board =\ | tobsmonster2 | Neutral |
| 12779 | @tobsmonster2 Imagine if that had been a gigabyte main board... Or a 37avelen bell! Think of the deveststion | Cy Starkman | Neutral |
| 12779 | @spasewalkr i would like to see this thing with 20" rims. Hahahahahaha | WilliamH | Neutral |
| 12779 | @spasewalkr To be fair, my car feels similar after I've been for a curry. | TC | Neutral |
| 12779 | @spasewalkr All that cost cutting had to come from somewhere... | AltairDusk | Neutral |
| 12779 | LOL, this doesn't surprise me ☺ | CaptainPlanet | Neutral |
| 12779 | @spasewalkr A 2,500 USD 2010 car. What do you expect? Lol | Shoxite | Neutral |

| | | | |
|-------|--|---------------|----------|
| 12779 | @spasewalkr It wasn't the engineering, it's that Lei in the hood that's causing the problem. | Someguyperson | Neutral |
| 12779 | @spasewalkr Hey! My uncle (how many times removed?) DESIGNED the Nano. | Foxh8er2 | Neutral |
| 12779 | @TjK glorious | BrownSound | Positive |
| 12779 | @spasewalkr I've always said that the Nano looked like it was designed for flamers. | NikAmi | Neutral |
| 12779 | My Bad | Lord Vader | Neutral |
| 12779 | @Lord Vader And let that be a lesson to you. DO NOT park here again! | Rederikus | Negative |
| 12779 | I was just testing out my new Death Star lasers | Lord Vader | Neutral |
| 12779 | Lier, liar!Cars on fire | Iluvms | Neutral |
| 12779 | \$2,500? If I needed a car and these were sold in the US, I'd buy one... even if it did have issues with bursting into flames. | Brent1700 | Neutral |
| 12779 | @Brent1700 You have a life insurance and you're bored, right? | Atkins | Neutral |
| 12779 | @Brent1700 It won't pass the crash tests. \$2500 can still get a decent car here in the US. You just need to know what to look for. | Vanmankline | Negative |
| 12779 | @Brent1700 I don't think any car insurance company would be insane enough to offer comprehensive coverage for this car. So your roadside BBQ payday may never come. | Spamfree | Negative |
| 12779 | @vanmankline I drive a '94 Lincoln Town Car I bought 2 years ago that just turned 100k miles. For \$2500, it is an excellent car, I'm looking for another one. The best seats in the world. | CallDon | Positive |
| 12779 | @Brent1700 Pfft, if only 3 of them have burst into flames, that's no big deal. I'm not saying it shouldn't be fixed, because it should, but wow, can we say media over-blown? | Abe | Neutral |
| 12779 | @Abe EXACTLY I love how ppl blow thing wayyyy out of proportion. And yea, I know you can get a car for \$2500 but, used. And, well, I don't believe in the whole "one man's trash is another man's treasure" thing... | Brent1700 | Positive |
| 12779 | @Abe Of course, if they've only sold 6 of 'em, that's a 50% failure rate. I'm just sayin'..... | Connie | Negative |
| 12779 | @Brent1700 Thats too bad, I was hoping to give you my old lambo, only 3k miles. Since it's used you probably don't want it though. | Scots79 | Negative |
| 12779 | @Abe You must have a lot of faith in large corporations, and very little interest in history. | Glens | Neutral |
| 12779 | @vanmankline What, you mean like this. http://www.theonion.com/video/ford-unveils-new-car-for-cashstrapped-buyers-the-1,14381/ | Mesozoic | Neutral |

| | | | |
|-------|--|---------------|----------|
| 12779 | It has already passed EU crash tests- videos available online- Now for US market- these won't be \$2500 nor a company will not design and produce cars that can't pass the basic tests for which norms are available; by the way did u see the "smart cars" already in US streets? | Jitheshvv | Negative |
| 12779 | @jitheshvv I've seen some. But they're over \$10,000! | Brent1700 | Neutral |
| 12779 | the car's made of plastic. | Tpbftw | Neutral |
| 12779 | Ya, any problems with the Nano EV would be quite shocking. | MarcusMaximus | Neutral |
| 12779 | @MarcusMaximus nice. I see what you did there... | Njeske | Neutral |
| 12779 | What happens to the worst can also happen to the best. http://www.blog.automotiveaddicts.com/thats-one-hot-ferrari-ferrari-is-on-fire | David G | Neutral |
| 12779 | @David G Impressive, yes, but for the price of this one car, you could instead watch about 120 Tatas go up in flames. | Patches66 | Positive |
| 12779 | @David G Not to mention, you have an excuse to say "tata" in mixed company. | Patches66 | Negative |
| 12779 | Oh man and I thought Toyota cars had issues | RLJSlick | Neutral |
| 12779 | Well at least its a small fire. | KAL326 | Neutral |
| 12779 | @KAL326 Well, it's a small car. | Atkins | Neutral |
| 12779 | I guess that's what happens when you try to push a Briggs & Stratton engine to it's limits. | LiqwidZero | Neutral |
| 12779 | @LiqwidZero Don't start the B&S vs Tecumseh flame wars. | Joelaf | Neutral |
| 12779 | That could be a good security system. | Atkins | Positive |
| 12779 | I hope that car has air conditioning. It's probably pretty warm in summer. | Arkweld | Neutral |
| 12779 | I've heard of 'Hot Hatches' but this just takes the cake... | richb93 | Neutral |
| 12779 | @richb93 The car is so cheap that it 39avenen even have a hatch. The back is solid. In order to get to the space in the rear you have to go over the back seats. | Luffy | Neutral |
| 12779 | I suppose they were trying too hard to avoid the reputation that rear-engine cars (like the VW Bug) have poor heaters. | CityZen | Negative |
| 12779 | Their consulting group is outsourcing a portion of the company I work for. Overall, the whole experience is being received about the same. But one difference is that the failure of the consulting group wont result in melted plastic. | Thiels851 | Negative |
| 12779 | @Thiels851 ...and which company is it that you work for? | Bazookafx3 | Neutral |
| 12779 | @bazookafx3 One who's future doesn't bode well if they continue along this path of outsourcing with lesser talent. | Thiels851 | Neutral |
| 12779 | @Thiels851 Wait – where is this Indian company | Valicore | Neutral |

| | | | |
|-------|--|---------------------|----------|
| | outsourcing to? | | |
| 12779 | @Valicore To Engadget commenters! The most populous nation in the world. | SuperGadget | Neutral |
| 12779 | @Valicore: my US based company has hired TCS (Tata Consulting Services) to replace my colleague and myself with 9 people overseas in India. | Thiels851 | Neutral |
| 12779 | @Thiels851 Or, maybe there is something really wrong with how you are teaching them. 9 months! I can teach Japanese to Lil Wayne in that time! | Nhoj281 | Negative |
| 12779 | I especially like the lei hanging out the front. | UnixSystemsEngineer | Neutral |
| 12779 | ZOMGWTFBBQ | vdubdan45 | Neutral |
| 12779 | I saw a 1995-ish Jeep Cherokee pulling a boat up a hill go from overheat to flaming pile of debris in about 15 minutes. Of course those things had the worst cooling system ever created by man. | Glamajamma | Neutral |
| 12779 | @glamajamma But in your example, that sounds like a good reason. The one in the picture looks... parked. | Atkins | Positive |
| 12779 | @Atkins : Read the summary, the car was being driven when it caught fire. After being alerted by a motorcyclist that his car was on fire, I assume the driver, rather wisely, decided to stop. | Wjousts | Neutral |
| 12779 | @Atkins It is now! LOL. | CallDon | Neutral |
| 12779 | Isn't it convenient that he's an insurance agent? | Spamfree | Neutral |
| 12779 | Tata Motors: Cars so HAWT, they spontaneously combust! | Da Black Anarch | Neutral |
| 12779 | tiny wheels are tiny | Fcells | Neutral |
| 12779 | My worst nightmare is to be in traffic with a Tata in front of me and a Toyota behind me. | Edf | Negative |
| 12779 | @edf Your comment, I love | Drey | Positive |
| 12779 | @edf Knock the Tata out of the way then turn off before the Toyota catches up. At worst you'll have a few scratches on the front bumper. | AltairDusk | Negative |
| 12779 | @edf best comment ever | Inertone | Positive |
| 12779 | Ahhh, the Pride of India. | (Unverified) | Positive |
| 12779 | I am torn between not burning to death and my love for tatas, even if they are little tatas. | Glamajamma | Neutral |
| 12779 | Saw the same thing happen to a Fiat 500, circa 1973. Our school's French teacher was driving along, totally engulfed from the door handles back. Folks were waving, and she'd smile and wave back. | Joelaf | Neutral |
| 12779 | @joelaf OT, how do you get the avatar? I have tried to upload several times and can't get mine to show up?? | CallDon | Neutral |
| 12779 | @CallDon Go to your profile. It will give an option to upload pic. BTW, mine's from the Apple vs HTC copyright filing. Mine didn't show up until I signed out | Joelaf | Positive |

| | | | |
|-------|---|-------------------|----------|
| | and signed back in. | | |
| 12779 | @CallDon Happened to me too. I had to settle for GIF format... I think... with the right size too. I think it's 64x64. | Onlymyrailgun | Neutral |
| 12779 | Nano is not designed for US driving speed and habit. It's designed for the very busy street of Mumbai and other city. You'd be hard press to get even 30 MPH just due to congestion. With that in mind, it isn't impossible to get it made and sell for \$2500. | BigFire | Negative |
| 12779 | Reminds me of that scene from Trains, Planes and Automobiles with John Candy....can you smell fire? ☺ | Rod | Neutral |
| 12779 | This car is hot! | Fliguy84 | Positive |
| 12779 | Don't worry folks, it'll be imported as the 2012 Ford Pinto! | Sanskrit | Neutral |
| 12779 | Yoga Flame | XxTheRemedyXx | Neutral |
| 12779 | @XxTheRemedyXx I love street fighter 4 on my itouch | Iluvms | Positive |
| 12779 | @XxTheRemedyXx Congratulations sir, you have just won the Internet. It should arrive in your mail-box in 6-8 weeks. | Krizoitz | Neutral |
| 12779 | Tata Tandori | Joelaf | Neutral |
| 12779 | Is this the Phoenix model? | HikaKao | Neutral |
| 12779 | It looks like Tata...Is on fire!YEEAAAAAAHHHHH!!!!!!!!!!!!!! | TheBennettBrigade | Neutral |
| 12779 | Rattata just got charred by Charizard! | Onlymyrailgun | Neutral |
| 12779 | Scare tactic Propaganda | Echobox | Neutral |
| 12779 | I guess you could say Tata, to the Tata. | Nitesh | Neutral |
| 12779 | their customer service center is in india, like those US companies, right? | Htd | Positive |
| 12779 | "Unexplained fires are a matter for the courts! / Tata Nano! / Tata Nano!" (with sincere apologies to Matt Groening et al) | Vtluu | Neutral |
| 12779 | I wonder if a contributing factor to these problems (of which I'm sure there are many) is the fact that they went with a midengine design. | Beatsandmelody | Neutral |
| 12779 | Here in India tata is facing lots of problem in their cars.....one their car , tata indica ,.....the axle of the car broke down and car lost all its handlings...even the brake was not working..... and the car was only 8 months old..... | XXR | Negative |
| 12779 | Hot as Curry | SmokemeaKipper | Positive |
| 12779 | Coming soon to a Jaguar / Land Rover dealer near you. | Wingdo | Neutral |

| | | | |
|-------|---|---------------|----------|
| 12779 | This is why airliners' avionics don't run Vista. | DigDug | Neutral |
| 12779 | That's one hot ride, bro! | Oflife | Positive |
| 12779 | With that flower garland hanging from the hood, it might actually be a funeral pyre. | Mtkupp | Neutral |
| 12779 | I like Tata's! I like them in all sizes! | Wam9468 | Neutral |
| 12779 | Looks like that 42avele got fried! | Bteeuwen | Neutral |
| 12779 | They're so inexpensive so their disposable. As in, wave "Ta-Ta"! | ChewChewTrain | Neutral |
| 12779 | amazing anti theft system | Gooner14 | Positive |
| 12779 | I'm so hot for Tata. | Kris91 | Neutral |
| 12779 | here comes the jaguar | Nexgenmax | Neutral |
| 12779 | I paid \$2500 for a Ford Temp years ago, it had 60k miles on it and we drove it for 6 years until the transmission went out on it. Never caught on fire! | Bonedog73 | Negative |
| 12779 | My company is in the process of negotiating a contract with Tata to become the main supplier of cabins for the helicopters we build. So yeah, there's that. | Patbergix | Neutral |
| 12779 | Save the Tatas! | Keyser | Neutral |
| 12779 | Great, another exploding Apple product!Whats that? Not an iPod? Really? Someone else makes a Nano? Oh, well....never mind then.... | Krzoitz | Negative |
| 12779 | You get what you pay for. | Purge68 | Neutral |
| 12779 | I live in Chicago. How can I get this feature added to my car? It would be great for winter time. My car would be warm and de-iced before I get in it with this feature. | James T | Positive |
| 12779 | disposable car | Benderfender | Neutral |
| 12779 | Could it be the fancy feathers atop the motor-controller? | Kingu | Neutral |
| 12779 | Tata supplies a lot of the steel which goes into the other cars folks on this blog probably drive. | Naashak | Neutral |
| 12779 | New Toyota ad: Hey, at least our cars won't self-destruct like Tata's. | pika2000 | Neutral |
| 12779 | Only the cheap die young. | BuzzMega | Neutral |
| 12779 | well whats new about burning automobiles,their are plenty of expensive sports cars going up in flame,like the 42avelength42 42aveleng.just do a a google search "42avelength42 42aveleng up in flames". | Aim120 | Neutral |
| 12779 | If you didn't know, there exists American engineering in that Nano. More specifically GM engineering, if you still consider Delphi as part of GM. | Syke | Neutral |

| | | | |
|-------|---|---------------------------------------|----------|
| 11335 | I'm going to wait until they make tvs out of unubtanium | First2Mars | Neutral |
| 11335 | I wonder what he saw there for that brief second... WHOAA!!... :O | TruUnknowN | Positive |
| 11335 | #Nerdgasm | Chris DPSN AggieCEO XBLThe Aggi | Neutral |
| 11335 | @Chris DPSN AggieCEO XBLThe Aggi You look like a guy I've seen on TV! | CarliT | Neutral |
| 11335 | Is he included in the package? | Techno1q | Neutral |
| 11335 | @Techno1q I wish George Takei would include himself in mine | Dystroglycan | Neutral |
| 11335 | @KyleBolton OOOH MY! | Acispades | Neutral |
| 11335 | I had thought he was the Dharma Initiative orientation guy, when I saw this commercial. | Killplay | Neutral |
| 11335 | @killplay as a south 44avel (indian) person, I can confirm that 44avel people do all look alike. Sometimes i confuse my cousins, they all look alike. | Ravissimo | Neutral |
| 11335 | @ravissimo not just that, the lab coat, the dialog, everything | Killplay | Negative |
| 11335 | Edge lit gives you ridiculously thing displays. However, backlit gives your better color reproduction. Oh, the choices... them is good!!!! | Xtasi | Negative |
| 11335 | That was a beautiful thing. | Tohe | Positive |
| 11335 | @KyleBolton So? | Tohe | Neutral |
| 11335 | @KyleBolton The minds not the only thing he can blow | Rhepsfan | Negative |
| 11335 | @Tohe Was just a funny observation, lighten up. | KyleBolton | Neutral |
| 11335 | Don't you internet bastards do it. Don't you dare take that picture and do something dirty to you. You bastards. You internet bastards.George Takei is a nice man. DON'T. YOU. DARE. | Teslanaut | Negative |
| 11335 | @Teslanaut not... starting... photoshop... i.... swear. | UnixSystemsEngi neer | Neutral |
| 11335 | @Teslanaut MUST...RESIST..DAMN..YOU...DARTH...VAADDEEERRR!!! !! | CaptainPlanet | Neutral |
| 11335 | I don't get it: So the fourth pixel requires third input? | Luke | Neutral |
| 11335 | @Luke i'm guessing its a fourth color pixel, to better reproduce colors, its going from RGB to CMYK (google it), it doesn't require a second input it just reallocates different voltages to the pixels to represent a wider range of color | totaldestroyer | Neutral |

| | | | |
|-------|---|---------------|----------|
| 11335 | i'm guessing its a fourth color pixel, to better reproduce colors, its going from RGB to CMYK (google it)"This is not CMYK. | Jason42 | Negative |
| 11335 | @totaldestroyer I'm not entirely sure I get this.RGB: blast green and blue full to get yellow. So what, exactly, is the point of the yellow pixel? | The Madman | Negative |
| 11335 | yeah, I feel like this is just for the marketing, there are two types of primary colors, additive; RGB, and subtractive, CMY | JeremyBenthem | Positive |
| 11335 | @JeremyBenthem Red +Green = yellow in CIE world, Green+Blue = Cyan | Irfan | Neutral |
| 11335 | Yeah i agree, don't see that much point of this. The reason why we have RGB, is that, in most people, cones in our eyes are sensitive to only 3 different frequencies of light (short 45avelength – Blue; | Nmosfet | Positive |
| 11335 | @Luke "Third Input" ?? Really? We're going back to the gay thing again?lulz. | (Unverified) | Neutral |
| 11335 | @jason42 So, the content is being captured in an RGB system, transported in another system, and Sharp is presenting it in LAB system. How is Sharp interpolating the RBG original data, and how is the original LCD technology limiting that LAB space? | (Unverified) | Neutral |
| 11335 | Close...RGB is the additive color spectrum...CMY is the subtractive spectrum. | Gentlefury | Neutral |
| 11335 | To me thin tv's have no real value other than "hey look how thin my tv is" | MurdaFace | Neutral |
| 11335 | @MurdaFace And look how light it is come moving time...or look how the 55" model still works on my stand I had my 42" plasma on...or how it uses 40% less electricity to do the same thing. | Gentlefury | Neutral |
| 11335 | Why hasn't anyone thought of adding yellow before? | Mellerad | Neutral |
| 11335 | @mellerad Because every color of light can be made with red, green, and blue. Yellow, for example, can be made by combining red and green light. | DrTrent | Neutral |
| 11335 | Red green and blue can't reproduce every color. The three define a limited subset of all possible colors. | Garths | Negative |
| 11335 | @garths "Red green and blue can't reproduce every color."I'm sure some physicists would like to have a word with you. | Diamondsw | Neutral |
| 11335 | @diamondsw "A set of primary colors, such as the sRGB primaries, define a color triangle; only colors within this triangle can be reproduced by mixing the primary colors. | Garths | Neutral |
| 11335 | That's cool.I'm sure some physicists would like to have a word with you."Don't be pedantic. The RGB on your monitor can't produce the entire colour gamut. | Jason42 | Positive |

| | | | |
|-------|---|-----------------|----------|
| 11335 | diamondsw You're right. I would like to have a word with him...because he's right. | Jedidove | Positive |
| 11335 | garths Why they chose yellow for the fourth is beyond me. Yellow is well approximated. | Jedidove | Positive |
| 11335 | @jedidove For the same reason Fuji has super saturated yellow film....skin tones. | Gentle fury | Neutral |
| 11335 | Set phasers to "stunned"! :o | ChazClout | Postive |
| 11335 | I think that fish gave him the full monty. | To Hell | Neutral |
| 11335 | how do you pronounce aquos? | Ren021 | Neutral |
| 11335 | @Ren021 Like aqua, but pronounced like aww-kwohs. | Einhanderkiller | Neutral |
| 11335 | @Ren021 Exactly how it is spelled. | Paul34 | Neutral |
| 11335 | @Ren021 I believe its actually awk-wohs. The 'k' sound is part of the first syllable. | Ramifications | Neutral |
| 11335 | @KyleBolton Well, it was an observation. | Jon | Neutral |
| 11335 | Up until 0:09 I thought George Takei had retired from acting and become an engineer for Sharp.Or is it talking about the TV? | PotatoBaron | Neutral |
| 11335 | I don't know how many of yall have taken a art class ans know how many colors can be made with yellow. I have and this is seriously a big deal. I want a 26 inch monitor with this Qauttron technology. | Alexandertron | Neutral |
| 11335 | Alexandertron Yellow is a primary subtractive color, so it's important if you're working with pigments or inks.B69 | Ben F | Negative |
| 11335 | @Ben F True..... BUT this is unarguably a upgrade from a 3 color system. | Alexandertron | Neutral |
| 11335 | @Alexandertron Well, this is definitely an interesting concept, and I hope it succeeds enough where it would benefit us with computers. | Groovedafied | Positive |
| 11335 | William Shatner: Then George, you could give them a karate chop.George Takei: I find that offensive. | Battery199 | Negative |
| 11335 | @battery199 +1 for the Futurama quotes. That's such a great episode.Leonard Nimoy: "I am living in a gefilte fish jar." | CapnShiner | Positive |
| 11335 | I want a side-by-side comparison between a tv with the fourth pixel and a sucky normal tv with only three different colours before I am convinced. | HikaKao | Negative |
| 11335 | @HikaKao That would require you to be physically there, looking at the screens. | Groovedafied | Neutral |
| 11335 | "Actually I'd like to touch" – George Takei | Kaplanfx | Neutral |

| | | | |
|-------|--|-----------------|----------|
| 11335 | @kaplanfx “Young. Wet. Bitches.” | Logicbombde | Neutral |
| 11335 | this is old | chad3670 | Neutral |
| 11335 | If I don’t get him with the TV I don’t want it! >_ | BrownSound | Neutral |
| 11335 | best.advert.ever. | jjrp123 | Positive |
| 11335 | bababooney! Bababooney! | Jonhimself | Neutral |
| 11335 | George Takei, you rock my world! | Kiteless | Positive |
| 11335 | Dry Oatmeal | Longhairbilly | Neutral |
| 11335 | So if video signals are made of red, green, and blue channels, where exactly does this TV get the signal for the yellow? | Franktinsley | Negative |
| 11335 | You are one of the few that is smart enough to grasp the heavy duty marketing involved in this scam. | CrabGrass | Neutral |
| 11335 | I thought that was Sulu... | retro77 | Neutral |
| 11335 | I love that man. | Camperton | Positive |
| 11335 | but.....is it 3D? | iPaul | Neutral |
| 11335 | Raff out Roud | GarrettQ | Neutral |
| 11335 | 1000 Quattloos for the newcomer! It will make an excellent Thrall. | Samurai Jack | Positive |
| 11335 | “oohh wwuuuu!!!”.....”Chocolate rain!!!!” | egy11micro | Neutral |
| 11335 | I’m about 50% sure that is Sun and Non’s doctor from the t.v. show “LOST” | Bhuggs | Neutral |
| 11335 | @bhuggs I can’t tell if you’re kidding or not. George Takei was not on Lost. | (Unverified) | Negative |
| 11335 | George Takei should narrate everything. | Pryomancer | Neutral |
| 11335 | Ooh My! | Zaraki Kenpachi | Neutral |
| 11335 | where is the white paper explaining the actual technology? | Chodaboy19 | Neutral |
| 11335 | lol good stuff. | Tristan88888888 | Positive |
| 11335 | Well seems to me like they ripped off the best of Audi: Quatro + E-Tron = Quatron | Tohe | Positive |
| 11335 | Have you ever noticed on anything but a full red screen most crt’s and lcd’s emit a bluish glow? | Psyopper | Positive |

| | | | |
|-------|--|------------------|----------|
| 11335 | *yawn* cyan, magenta, yellow, black... now where have I seen this before.. | Buoy | Neutral |
| 11335 | saw this commercial on tv for the first time last night and fell off the couch in laughter. Great ad! | Jpstyles21 | Positive |
| 11335 | ..."can still blow your mind". I bet that's not the only thing he blows. Ba-dum-psshhhhhhh.Sorry, George, I couldn't resist. | Johnny Tremaine | Negative |
| 11335 | Sulu. Can you be any Nellyer?!Oh and my year and half old DLP has 6 colors. AND 3D. | (Unverified) | Neutral |
| 11335 | George Takei's laugh is so unique | Prince David | Neutral |
| 11335 | I wonder who will bring out 7 Pixel ROYGBIV screens? | AlthalusTheThief | Neutral |
| 11335 | i hate that guy. After watching that commercial 10 times in a row i want to hit that guy with the stupid tv over the head. | Pretol | Negative |
| 11335 | 3 more colors to gay TVs | NIMBUS | Neutral |
| 11335 | Saw this on BBC America during Top Gear last night. Very funny. | AlwaysThere | Neutral |
| 11335 | Excellent | aolke123 | Positive |
| 11335 | Apparently Panasonic did this back in 1972...it obviously didn't catch on | Howzer | Neutral |
| 11335 | Actually no they didn't forget I spoke. | Howzer | Neutral |
| 11335 | I hardly watch live tv anymore. I Pretty much dvr anything | Sporkinum | Positive |
| 11335 | I like the remix better (via regretsy.com): http://www.youtube.com/watch?v=JfOpgTN2ZqY | Technojunkie | Neutral |

Table 10: Statistics of mined comments

| Post ID | Positive | Negative | Neutral | Total comments | Avg Influence of Authors |
|---------|----------|----------|---------|----------------|--------------------------|
| 12779 | 15 | 17 | 82 | 114 | 83.19 |
| 11335 | 26 | 19 | 69 | 114 | 82.65 |

As shown in Table 10, the classification of comments, the blog post 11335 has received more number of positive comments. Also the average influence of authors commented on

11335 blog post is greater than 12779. So the above example explains the reason why blogpost 11335 is being scored more than 12779.

The Influence-indexes (I-Mine Comments) of the bloggers when we apply the resultant model on dataset are present in Table 11. The column I-Mine comments represent the Influence-index of resultant model when “comment score” part (second half of eq. 6) in “productive and influential bloggers” has been replaced by “mining comments” technique (eq. 8).

Table 11: BI vs. I-Mine comments

| Author ID | Author Name | BP | BI | I - Mine Comments |
|-----------|-------------------|----|-----|-------------------|
| 4 | Darren Murph | 23 | 114 | 104 |
| 9 | Chris Ziegler | 18 | 94 | 92 |
| 13 | Vladislav Savov | 17 | 91 | 91 |
| 11 | Laura June | 13 | 88 | 74 |
| 8 | Nilay Patel | 13 | 74 | 74 |
| 2 | Thomas Ricker | 13 | 72 | 73 |
| 14 | Joshua Topolsky | 16 | 72 | 66 |
| 3 | Paul Miller | 20 | 67 | 65 |
| 10 | Donald Melanson | 12 | 64 | 66 |
| 6 | Tim Stevens | 14 | 63 | 65 |
| 1 | Joseph L. Flatley | 16 | 56 | 56 |
| 7 | Ross Miller | 15 | 54 | 52 |
| 21 | Joanna Stern | 19 | 42 | 45 |
| 28 | Ryan Block | 3 | 39 | 25 |
| 26 | Richard Lai | 9 | 36 | 35 |

As shown in Table 11, when we introduced mining comments technique the ranking among J Topolsky, P Miller, D Mealanson and T Stevens have been interchanged. Also R Block, R Lai’s ranks were interchanged. Even though most of the ranking order doesn’t

change, the technique followed captured the true influence of blog post by capturing the tone of the commenter.

In stage 3 we applied both facebook count and mined comments into the “productive and influential bloggers” model. And we have observed many changes in the ranking order of the resultant model. The I-(FB+Mine comments) indicates Influential-index of the resultant model when both facebook count and mining comments techniques are applied.

Table 12: BI vs. I-(FB+Mine comments)

| Author ID | Author Name | BP | BI | I - (FB+Mine Comments) |
|-----------|-------------------|----|-----|------------------------|
| 4 | Darren Murph | 23 | 114 | 112 |
| 9 | Chris Ziegler | 18 | 94 | 97 |
| 13 | Vladislav Savov | 17 | 91 | 105 |
| 11 | Laura June | 13 | 88 | 77 |
| 8 | Nilay Patel | 13 | 74 | 82 |
| 2 | Thomas Ricker | 13 | 72 | 76 |
| 14 | Joshua Topolsky | 16 | 72 | 67 |
| 3 | Paul Miller | 20 | 67 | 71 |
| 10 | Donald Melanson | 12 | 64 | 69 |
| 6 | Tim Stevens | 14 | 63 | 71 |
| 1 | Joseph L. Flatley | 16 | 56 | 59 |
| 7 | Ross Miller | 15 | 54 | 56 |
| 21 | Joanna Stern | 19 | 42 | 46 |
| 28 | Ryan Block | 3 | 39 | 25 |
| 26 | Richard Lai | 9 | 36 | 38 |

As shown in Table 12 the ranking of bloggers in highlighted sections has been interchanged. Table 12 depicts merged results of stage 1 & stage 2 with better discrimination between authors.

In stage 4 we have applied novelty along with FBCount and mining comments into model which is our final model. To find novelty of a blog we need to build word space which represents a general space of ideas. In our approach we considered each blog post as a document. We first removed all the stop words from the documents and then using TFIDF weighing scheme we extracted all the important keywords from the documents to create word space. After that we projected each latest blog post into the word space to estimate the degree of novelty. The outlinks used in blog post also plays major role in measuring degree of novelty. And novelty is calculated using eq. 9.

Table 13: Bi vs. I-My Approach

| Author ID | Author Name | BP | BI | I - My Approach |
|-----------|-------------------|----|-----|-----------------|
| 4 | Darren Murph | 23 | 114 | 113 |
| 9 | Chris Ziegler | 18 | 94 | 99 |
| 13 | Vladislav Savov | 17 | 91 | 108 |
| 11 | Laura June | 13 | 88 | 81 |
| 8 | Nilay Patel | 13 | 74 | 86 |
| 2 | Thomas Ricker | 13 | 72 | 83 |
| 14 | Joshua Topolsky | 16 | 72 | 73 |
| 3 | Paul Miller | 20 | 67 | 77 |
| 10 | Donald Melanson | 12 | 64 | 74 |
| 6 | Tim Stevens | 14 | 63 | 76 |
| 1 | Joseph L. Flatley | 16 | 56 | 66 |
| 7 | Ross Miller | 15 | 54 | 65 |
| 21 | Joanna Stern | 19 | 42 | 56 |
| 28 | Ryan Block | 3 | 39 | 39 |
| 26 | Richard Lai | 9 | 36 | 49 |

When we applied the final model on dataset the ranking has been definite and more accurate than the other models. The column I-My Approach indicates the Influence

Indexes of final model. As shown in Table 13 the ranking has been reordered in highlighted sections. The bloggers were ranked without any overlapping among them. As depicted in Table 13 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.

Finally all the results of the four experiments were put together side by side in Table 14 for better comparison and understanding.

Table 14: BI vs. I-indexes

| Author ID | Author Name | BP | BI | I - FBCount | I - Mine Comments | I - (FB+Mine Comments) | I - My Approach |
|-----------|-------------------|----|-----|-------------|-------------------|------------------------|-----------------|
| 4 | Darren Murph | 23 | 114 | 124 | 104 | 112 | 113 |
| 9 | Chris Ziegler | 18 | 94 | 101 | 92 | 97 | 99 |
| 13 | Vladislav Savov | 17 | 91 | 111 | 91 | 105 | 108 |
| 11 | Laura June | 13 | 88 | 98 | 74 | 77 | 81 |
| 8 | Nilay Patel | 13 | 74 | 87 | 74 | 82 | 86 |
| 2 | Thomas Ricker | 13 | 72 | 83 | 73 | 76 | 83 |
| 14 | Joshua Topolsky | 16 | 72 | 76 | 66 | 67 | 73 |
| 3 | Paul Miller | 20 | 67 | 78 | 65 | 71 | 77 |
| 10 | Donald Melanson | 12 | 64 | 74 | 66 | 69 | 74 |
| 6 | Tim Stevens | 14 | 63 | 74 | 65 | 71 | 76 |
| 1 | Joseph L. Flatley | 16 | 56 | 66 | 56 | 59 | 66 |
| 7 | Ross Miller | 15 | 54 | 63 | 52 | 56 | 65 |
| 21 | Joanna Stern | 19 | 42 | 49 | 45 | 46 | 56 |
| 28 | Ryan Block | 3 | 39 | 39 | 25 | 25 | 39 |
| 26 | Richard Lai | 9 | 36 | 40 | 35 | 38 | 49 |

5.0 CONCLUSION

Blogs have been the major communication channel for many internet users for past couple of years. Blogosphere is a rapidly growing landscape with time and it is very dynamic in nature. Bloggers are sharing their ideas, opinions, and experience with others through blog posts. The people who get attracted or convinced by these blog posts will eventually act or choose their decisions in favor of the ideas borrowed from blogs. So intuitively bloggers are exerting their influence over internet users with their expertise. Such kinds of bloggers are called influential bloggers.

There exist preliminary models that proposed various solutions to identify influential bloggers in a web community. But they failed to consider important aspects of the blog posts. The existing works mainly depended on numbers associated with the blog post (length, comments, Inlinks, outlinks etc.). They haven't considered the user tone toward (agree or disagree) the blog post which is very important while evaluating the influence. Also none of them have considered the novelty of post which indicates the goodness and purity of the content. Inspired from the literature and sophisticated solutions that the current literature lacking, we proposed a method that evaluates each post based on their quality and user consent towards the content present. We introduced new influential factor (facebook count) into the literature that further evaluated blog posts efficiently.

From results obtained in our experiments we can say that the introduction of facebook count (combination of facebook likes and shares) has definitely improved the model to a great deal. The new factor helped the model in finding right tuning between different blog posts. Also Instead of just considering number of comments we actually mined each comment to find out the tone of commenter towards the post, which helped model to capture the true influence of the blog posts. Finally using the novelty parameter the model is able to express the goodness and quality of content.

Overall our experiment results show that the current approach has improved over the previous models. But in future work we can include more parameters like twitter shares, G +1s etc. into the literature and check for the behavior of the influence. As of now our approach has used list of positive and negative word phrases to find out the tone of the commenter, but better mining techniques can further improve accuracy of sentiment factor of the commenter. Also, we can investigate more time in deciding weight parameters that are crucial for tuning between different influential factors.

6.0 REFERENCES

- [1] Agarwal, N., Liu H., Tang, H., & Yu, P. S. (2008). Identifying the influential bloggers in a community. *WSDM '08: Proceedings of the international conference on Web search and web data mining*.
- [2] Akritidis, L., Katsaros, D., & Bozanis, P. (2011). Identifying the Productive and Influential Bloggers in a Community. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* , vol.41, no.5, pp.759-764, Sept. 2011. doi: 10.1109/WI-IAT.2009.18
- URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5284916&isnumber=5284878>
- [3] Blogosphere. Wikipedia. Retrieved October 4, 2011, from <http://en.wikipedia.org/wiki/Blogosphere>
- [4] Blogs. Wikipedia. Retrieved December 9, 2011, from <http://en.wikipedia.org/wiki/Blog>
- [5] 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
- URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5447798&isnumber=5447611>

[6] Leonidas, A., Dimitrios, K., & Panayiotis, B. (2009). Identifying Influential Bloggers:

Time Does Matter. *Web Intelligence and Intelligent Agent Technologies, 2009.*

WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on , vol.1, no., pp.76-83, 15-18 Sept. 2009. doi: 10.1109/WI-IAT.2009.18

URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5284916&isnumber=5284878>

[7] Mehwish, A., & Muhammad, R. (2010). Sentence based semantic similarity measure

for blog posts. *Digital Content, Multimedia Technology and its Applications (IDC),*

2010 6th International Conference on , vol., no., pp.69-74, 16-18 Aug. 2010. URL:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5568598&isnumber=5568515>

[8] Song, X., Chi, Y., Hino, K., and Tseng, B. (2007). Identifying opinion leaders in the

blogosphere. *CIKM '07: Proceedings of the sixteenth ACM conference on*

Conference on information and knowledge management.