Learning From The Crowd : An Evolutionary Mutual Reinforcement Model for Analyzing Events

Debanjan Mahata and Nitin Agarwal
Department of Information Science,
University of Arkansas at Little Rock, U.S.A
dxmahata@ualr.edu, nxagarwal@ualr.edu

***Abstract*—Social media is inarguably a powerful medium for mobilizing support for various real-life events be it for social, political, or economic transformation. Further, in contrast to the generic information obtained from the mainstream media, novel and specific information available at social media sites makes them valuable sources for event analysis. However, due to the power law distribution of the Internet, these overwhelmingly large number of sources are buried in the Long Tail making it extremely challenging to identify the quality sources among them. In this research, we propose an evolutionary mutual reinforcement model to confront these challenges. Due to absence of ground truth, a novel evaluation strategy is introduced. The results indicate tremendous potential. 25% to 130% information gain is obtained with the proposed approach when compared against the state-of-the-art baselines, viz. Google blog search and Icerocket blog search. Further, our ranking methodology is capable of identifying the highly informative sources much earlier than the aforementioned baselines. The proposed model affords an apparatus for micro and macro event analysis.**

Keywords: event analysis, social media, mutual reinforcement, specificity, closeness, information gain.

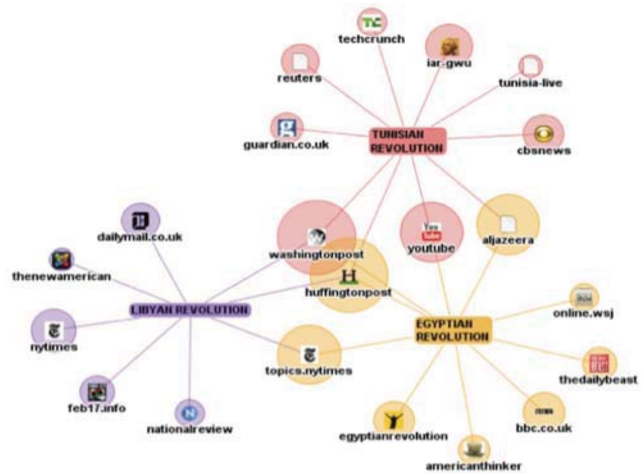

Figure 1: Top 10 Google search results for “Egyptian Revolution”, “Libyan Revolution”, and “Tunisian Revolution”.

I. Introduction

Social media has brought a paradigm shift in the way people communicate and share information. As observed recently, social media played an important role in mobilizing support for various events such as, ‘The Arab Spring’, ‘Occupy Wall Street’, ‘Sandy relief efforts’, ‘London Riots’, ‘The Spanish Revolution’ and others. Social media serves as a parallel, yet distinct source of information about real-life events along with the mainstream media space [21]. Mainstream media sources often gloss over the intricate details while covering a real-life event. They are often biased, regulated by the government, and may not portray the true picture of an event [11]. Whereas, social media sources (e.g. blogs) often contain unbiased, uninhibited, and unedited opinions from people. Thus the sources, which are obtained from social media could potentially provide a rather ‘closer’ or an ‘on-ground’ view of the events with novel information. In fact blogs have been accepted as credible sources of information when compared with mainstream media sources [13].

Due to the power law distribution of the Internet [1], the ‘Short Head’ is generally dominated by the mainstream media sites. As illustrated in Figure 1 the top 10 Google search results for “Egyptian Revolution", “Libyan Revolution", and “Tunisian Revolution" retrieved mainstream media sources. Consequently, the social media sites get buried in the Long Tail [16]. Inspired by Chris Anderson’s quote [3], i.e. “With an estimated 15 million bloggers out there, the odds that a few will have something important and insightful to say are good and getting better”, we explore techniques that would help in identifying the buried Long Tail social media sources containing highly specific information about an event.

Identifying highly informative or ‘specific’ sources related to a real-life event from social media presents various challenges. Overwhelming number of the sparsely linked Long Tail social media sources [2] makes it challenging to identify informative sources. The entities (person, organization, place, etc.) mentioned in the sources act as the atomic units of information. The sources which are ‘specific’ to an event most likely contain entities ‘closer’ or highly relevant to the event. Accordingly, ‘close’ entities can be obtained from ‘specific’ sources, hence presenting a dilemma in assessing the quality of the sources for event related ‘specific’ information content, which makes it a nontrivial task. Furthermore, the challenges with the unstructured content and colloquial usage of language predominant in social media cannot be ignored. Conventional information retrieval based evaluation measures help in identifying the most relevant and authoritative sources, however, these sources may not be the most novel or offer specific information. Therefore, new evaluation measures are required to estimate the performance of any proposed model. In this

ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN

research, we develop a methodology based on the principle of mutual reinforcement that helps in identifying highly 'specific' sources and confront the challenges mentioned above. Next, we present the related work and highlight the contributions of our research to the literature.

II. Related Work

User-generated data from various social media platforms, related to real-life events, have been studied to perform wide range of analysis. Socio-political inferences were drawn by studying sentiments and opinions of people towards public and political events from Twitter [25], as well as blogs [22]. Twitter has been extensively used as a source for analyzing information circulated during natural disasters and crisis situations [24], [6]. Tweets related to events have been extracted, summarized and visualized, in order to have a deeper understanding of the events [18], [19].

Due to huge number of informal sources in social media it is a difficult task to identify high quality sources related to the real-life events. Researchers have built semantic web models for efficient retrieval of event specific media sources [23]. Event related contents have been found leveraging the tagging and location information associated with the photos shared in Flickr [20]. *Becker et al* [4], studied how to identify events and high quality sources related to them from Twitter. In order to identify the genuine sources of information, credibility and trustworthiness of event related information were studied from Twitter [10]. New methods were investigated for filtering and assessing the verity of sources obtained from social media by journalists [7].

Several methods have been developed in the past for identifying and ranking quality sources from the web. PageRank [5] took advantage of the link structure of the web for ranking web pages. It was further improved for making it sensitive to topic based search [12]. Graph based approaches were used for modeling documents and a set of documents as weighted text graphs, and for computing relative importance of textual units for Natural Language Processing [8]. Mutual reinforcement principle was used for identifying Hubs and Authorities from a subset of web pages [14].

The proposed methodology makes a unique contribution to the literature of event analysis by modeling the relationship between sources and entities through an evolving mutual reinforcement system for identifying 'specific' sources and 'close' entities w.r.t an event. Next, we present the formal problem definition.

III. Problem Definition

Given a finite set of events ξ, we take an event $E_j \in \xi$ such that, $1 \leq j \leq \mid \xi \mid$, a set of 'p' sources denoted by ϕ_{E_j}, and a set of 'q' entities denoted by σ_{E_j}, related to the event E_j. We define two functions κ (*specificity*) and τ (*closeness*) such that:

$$\kappa : S_i \rightarrow [0,1] \tag{1}$$

$$\tau : e_i \rightarrow [0,1] \tag{2}$$

where, $S_i(\in \phi_{E_j})$, is the i^{th} source, and $e_i(\in \sigma_{E_j})$ is the i^{th} entity, so that we can get two ordered sets $(\varphi_{E_j}, \varsigma_{E_j})$ for the set of sources in ϕ_{E_j} and entities in σ_{E_j}, such that:

$$\varphi_{E_j} = \{S_1, ..., S_i, S_j, ..., S_p \mid \kappa(S_i) \geq \kappa(S_j), i < j\} \tag{3}$$

$$\varsigma_{E_j} = \{e_1, ..., e_i, e_j, ..., e_q \mid \tau(e_i) \geq \tau(e_j), i < j\} \tag{4}$$

φ_{E_j} is ordered in decreasing order of specificity S_i. ς_{E_j} is ordered in decreasing order of closeness e_i.

IV. Methodology

The relation between specific sources and close entities is modeled following the Mutual Reinforcement Principle, which extends the basic Mutual Reinforcement Principle to consider the evolving knowledge learned about an event. However, the model requires an apriori or seed knowledge about an event, which is provided in terms of an event profile or an event dictionary. Next, we discuss the construction of event dictionaries.

A. *Event Dictionaries*

Each event E_j is profiled by constructing an event dictionary (σ_{E_j}). In order to calculate specificity of a source, we need to start with an initial set of close entities. At the same time, these close entities are better acquired from the specific sources. To solve this dilemma, we construct event dictionaries, from independent sources which are completely separate from the sources (ϕ_{E_j}) that need to be ranked.

Formulation of initial closeness scores: We calculate the 'closeness' score ($\tau(e_i)_{E_j}$) of each entity (e_i) for event E_j in order to construct the event dictionaries, by using equations 5 and 6 based on tf-idf measure. More details on the formulation of the closeness measure is provided in [17].

$$\tau(e_i)_{E_j} = e_i f\_IE_j f = f(e_i, E_j) * IE_j f(e_i) \tag{5}$$

$$IE_j f(e_i) = log(\frac{\mid \xi \mid}{\mid E_j \in \xi : e_i \in E_j \mid}) \tag{6}$$

Since we extract the entities from the sources related to the events, we cannot have an entity that does not belong to any of the events. Therefore, in eq. 6 the denominator is always greater than zero. Following steps are taken to construct the event dictionaries:

1) Entities are extracted from the sources collected from GlobalVoices (http://globalvoicesonline.org), using AlchemyAPI (http://alchemyapi.com) and their corresponding $\tau(e_i)_{E_j}$ values are calculated using equation 5. Due to colloquial nature of the sources entities could take several forms. For example, the entity 'Tahrir Square' occur as 'Tahreer', 'El-Tahrir', etc. Multiple representation of the same entity are disambiguated by applying pattern matching.
2) An entity may occur in multiple events and hence can be present in multiple event dictionaries with different $\tau(e_i)_{E_j}$ scores. Since the range of closeness scores is different for each event, we normalize $\tau(e_i)_{E_j}$ scores w.r.t an event between 0 and 1. The normalization affords a relative closeness assessment of an entity across multiple events. The higher the $\tau(e_i)_{E_j}$ score of an entity the closer it is to the event.
3) The entities are then ranked according to the descending $\tau(e_i)_{E_j}$ scores.

The dictionaries thus obtained from the above mentioned procedure are static and serve as a good source of apriori knowledge about the event. However, as we discover new knowledge from specific sources, it is desirable to update the event dictionaries. Next, we discuss how the dictionaries help in identifying specific sources, which in turn help in improving the dictionary.

B. Mutually Reinforcing Sources and Entities

Given an event $E_j \in \xi$, a set of sources (ϕ_{E_j}) and entities (σ_{E_j}), related to the event, we define two column vectors: ***'Specificity'*** ($\kappa_{\mathbf{E_j}}$) and ***'Closeness'*** ($\tau_{\mathbf{E_j}}$).

$$\kappa_{\mathbf{E_j}} = <\kappa(S_1)_{E_j}, \kappa(S_2)_{E_j}, ..., \kappa(S_p)_{E_j}>^T \tag{7}$$

$$\tau_{\mathbf{E_j}} = <\tau(e_1)_{E_j}, \tau(e_2)_{E_j}, ..., \tau(e_q)_{E_j}>^T \tag{8}$$

where, $\kappa(S_i)_{E_j}$ ($\in range(\kappa)$, from equation 1) represents the *'specificity'* score of i^{th} source $S_i(\in \phi_{E_j})$, for $1 \leq i \leq p$ and $\tau(e_i)_{E_j}$ ($\in range(\tau)$, from equation 2) represents the *'closeness'* score of i^{th} entity $e_i(\in \sigma_{E_j})$, for $1 \leq i \leq q$. Each source S_i may contain related as well as unrelated information about various events. If we consider the set of events ξ, then each $\kappa(S_i)_{E_j}$ is itself a vector of 'specificity' values of the source S_i w.r.t the events as expressed in equation 9.

$$\kappa(S_i)_{E_j} = <\kappa(S_i)_{E_1}, \kappa(S_i)_{E_2}, ..., \kappa(S_i)_{E_{|\xi|}}> \tag{9}$$

Similarly, each entity e_i may be related to various events. If we consider the set of events ξ, then each $\tau(e_i)_{E_j}$ is itself a vector of 'closeness' values of the entity e_i w.r.t the events as expressed in equation 10.

$$\tau(e_i)_{E_j} = <\tau(e_i)_{E_1}, \tau(e_i)_{E_2}, ..., \tau(e_i)_{E_{|\xi|}}> \tag{10}$$

However, while representing $\kappa(S_i)_{E_j}$ and $\tau(e_i)_{E_j}$ as an element of the vectors $\kappa_{\mathbf{E_j}}$ and $\tau_{\mathbf{E_j}}$, respectively, we only choose the entry for the j^{th} event under consideration.

We construct a bipartite graph, $G = (V, U)$ representing the mutual relationship between the sources and the entities, where $V \in \phi_{E_j}, \sigma_{E_j}$, is the set of vertices for G, and U is the set of undirected edges. The sources without entities are discarded during this process.

The presence of an entity in a source is not sufficient to determine its specificity. In order to express the specificity of a source w.r.t an event, we need to consider the closeness value of the entities present in the source. Given the closeness $\tau(e_n)_{E_j}$ of an entity e_n w.r.t an event E_j, obtained from the event dictionary, a weight w_{mn} is assigned to the edges of the graph, which expresses the magnitude by which an entity is related to a source S_m.

The significance of an entity e_n in a source S_m is then,

$$\frac{f(e_n, S_m)}{\sum_{n=0}^{q} f(e_n, S_m)} \tag{11}$$

where, $f(e_n, S_m)$ is the frequency of occurrence of the entity e_n in the source S_m. Therefore mathematically,

$$w_{mn} = \frac{\tau(e_n)_{E_j} * f(e_n, S_m)}{\sum_{n=0}^{q} f(e_n, S_m)} \tag{12}$$

The adjacency matrix of the bipartite graph G is denoted by L, and is defined as follows:

$$L_{mn} = \begin{cases} w_{mn} & if\ (m,n) \in U \\ 0 & otherwise \end{cases}$$

Following the Mutual Reinforcement Principle the relationships between specificity scores of sources and closeness scores of entities for event E_j can be denoted as follows,

$$\kappa_{\mathbf{E_j}} = \mathbf{L}\tau_{\mathbf{E_j}} \tag{13}$$

$$\tau_{\mathbf{E_j}} = \mathbf{L^T}\kappa_{\mathbf{E_j}} \tag{14}$$

Substituting the values for $\kappa_{\mathbf{E_j}}$ and $\tau_{\mathbf{E_j}}$, we derive the following equations,

$$\kappa_{\mathbf{E_j}} = \mathbf{LL^T}\kappa_{\mathbf{E_j}} \tag{15}$$

$$\tau_{\mathbf{E_j}} = \mathbf{L^T L}\tau_{\mathbf{E_j}} \tag{16}$$

Equations 15 and 16 are characteristic equations of an eigensystem, where the solutions to $\kappa_{\mathbf{E_j}}$ and $\tau_{\mathbf{E_j}}$ are the respective eigen vectors with the corresponding eigenvalue of 1.

To emphasize the relationship between the sources and the entities, we make a major contribution by modifying the way the equations 15 and 16 are solved. We make the matrices $\mathbf{LL^T}$ and $\mathbf{L^T L}$ evolutionary while solving the equations. Since each of the equations is a circular definition, the final specificity and closeness scores are computed using the power iteration method [9]. Each iteration improves specificity and closeness scores reflecting their mutual relationship. As we move towards getting the specific sources and close entities in each iteration, we update the weights (w_{mn}) assigned to the edges between the sources and the entities by the newly calculated closeness scores for the entities. This results in renewed reinforcement of the relationship at every iteration by getting closer entities from better sources and vice-versa. This essentially helps the model incorporate the newly discovered knowledge about the events. More precisely, the improved understanding of the relationship between the source and the entities vis-a-vis an event is incorporated into the model.

The updation of the edge weights and the matrices with k^{th} iteration is represented as follows,

$$w_{mn(k)} = \frac{\tau(e_{n(\mathbf{k-1})})_{E_j} * f(e_n, S_m)}{\sum_{n=0}^{q} f(e_n, S_m)} \tag{17}$$

$$L_{mn(k)} = \begin{cases} w_{mn(k)} & if\ (m,n) \in U \\ 0 & otherwise \end{cases}$$

where, $L_{mn(k)}$ represents the adjacency matrix for graph G, and $w_{mn(k)}$ denotes the edge weight for the edge between m^{th} source and n^{th} entity at the k^{th} iteration. $\tau(e_{n(\mathbf{k-1})})_{E_j}$ represents the closeness score of the entity e_n w.r.t the event $E_j(\in \xi)$, obtained from the evolving event dictionary for event E_j at $(k-1)^{th}$ iteration.

If, $\kappa_{\mathbf{E_j}}(\mathbf{k})$ and $\tau_{\mathbf{E_j}}(\mathbf{k})$ be the specificity and the closeness scores, at the k^{th} iteration, the iterative process for generating the final solution are,

$$\kappa_{\mathbf{E_j}}(\mathbf{k}) = \mathbf{L_{k-1}L_{k-1}^T}\kappa_{\mathbf{E_j}}(\mathbf{k-1}) \tag{18}$$

$$\tau_{\mathbf{E_j}}(\mathbf{k}) = \mathbf{L_{k-1}^T L_{k-1}}\tau_{\mathbf{E_j}}(\mathbf{k-1}) \tag{19}$$

In order to get 1 as the largest eigenvalue and, $\kappa_{\mathbf{E_j}}$ and $\tau_{\mathbf{E_j}}$ as the principal eigen vectors, the matrices $\mathbf{L_{k-1}L_{k-1}^T}$

and $\mathbf{L^T_{k-1}L_{k-1}}$ needs to be stochastic and irreducible [15] at every step of our evolutionary process. In the present case, since the graph G is a bipartite graph, matrices $\mathbf{L_{k-1}L^T_{k-1}}$ and $\mathbf{L^T_{k-1}L_{k-1}}$ are already irreducible.

In order to make the matrices $\mathbf{L_{k-1}L^T_{k-1}}$ and $\mathbf{L^T_{k-1}L_{k-1}}$ stochastic, we take the following steps at each iteration,

- Dividing the non-zero entries of the matrices $\mathbf{L_{k-1}L^T_{k-1}}$ and $\mathbf{L^T_{k-1}L_{k-1}}$ by the summation of all the entries in a row.
- Assigning $1/n$ to the zero entries of $\mathbf{L_{k-1}L^T_{k-1}}$ and $1/m$ to the zero entries of $\mathbf{L^T_{k-1}L_{k-1}}$, respectively.

We perform a comparative analysis with a conventional binary static matrices represented as follows,

$$L_{mn} = \begin{cases} 1 & if\ (m,n) \in U \\ 0 & otherwise \end{cases}$$

Table I: Details of Data Collected.

Service Used	Event	Number of Blog Posts
GlobalVoices	Egyptian Revolution	234
	Libyan Revolution	86
	Tunisian Revolution	77
Google Blogger	Egyptian Revolution	579
	Libyan Revolution	600
	Tunisian Revolution	484
Icerocket Blog Search	Egyptian Revolution	5900
	Libyan Revolution	2198
	Tunisian Revolution	1220

V. Data Collection

Blog posts from GlobalVoices, Blogger, and Icerocket are collected for the study. The details of the dataset used is given in Table I. The dataset includes 11,378 blog posts from various blogging platforms like blogspot.com, wordpress.com, livejournal.com, typepad.com, etc. Non-english blogs are removed. The data from GlobalVoices is used for constructing event dictionaries (σ_{E_j}), as explained in Section IV. We collect blog posts related to the three events from Blogger using Google Search, and from other blogging platforms using Icerocket blog search. We extract the sources retrieved by the search engines along with the ranks and consider as our baseline. The collected blog posts are parsed to extract, *URL* of the blog and blogpost), *blog text*, *language*, and *rank* from the respective search engines. Alchemy API is used to extract entities.

Table II: Comparison between the number of iterations taken by the proposed evolutionary mutual reinforcement model and the conventional mutual reinforcement model.

Number of Iterations		
Conventional Model	Our Proposed Model	Event
17	4	Egyptian Revolution
19	5	Libyan Revolution
10	5	Tunisian Revolution

VI. Experiment and Analysis

We first discuss the experimental setup, followed by the comparative analysis between the proposed evolutionary mutual reinforcement model and conventional mutual reinforcement model. We then, introduce a novel evaluation strategy comparing the proposed model with the two baseline models.

A. *Experimental Setup*

The methodology discussed earlier is implemented on the collected datasets. We consider ξ = {“Egyptian Revolution", “Libyan Revolution", “Tunisian Revolution"} as our set of events. We construct the event dictionaries (σ_{E_j}) using GlobalVoices (explained in IV-A). The methodology is implemented on the set of sources (ϕ_{E_j}) from Blogger and Icerocket and the set of entities (σ_{E_j}) from the event dictionary. The threshold value for convergence μ of the power iteration method is decided empirically. On convergence of the power iteration method, we get a ranked set of sources (φ_{E_j}) and entities (ς_{E_j}) in terms of how specific and close they are to the event, respectively. Conventional mutual reinforcement model was also implemented for comparative evaluation.

B. *Comparing Conventional and Evolutionary Mutual Reinforcement Models*

In this experiment, we compared the convergence rate of the two models, viz., the proposed evolutionary mutual reinforcement model with the conventional mutual reinforcement model. As observed from II, the evolutionary model consistently outperforms the conventional model for all the three events by converging significantly faster. Next, we compare the accuracy of the proposed evolutionary model with the conventional model and the baselines in terms of how quickly the models identify specific sources for the events.

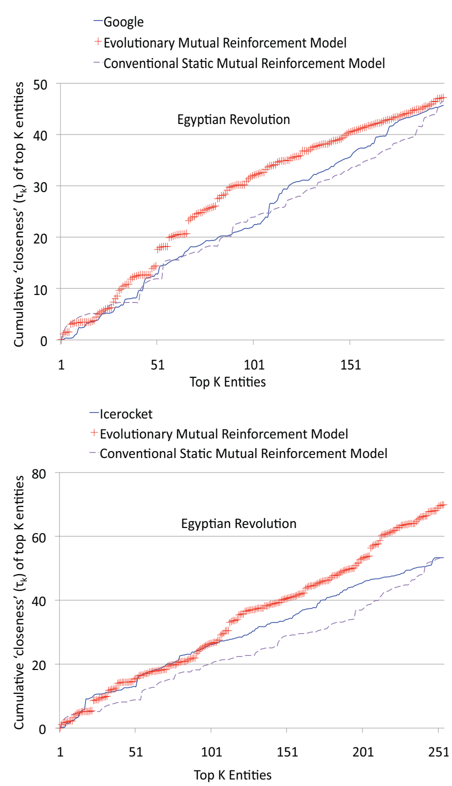

Figure 2: Model Validation.

C. *Baseline Comparisons*

Due to lack of benchmark datasets we use the search results obtained from Google and Icerocket Blog Search as baselines for validation. Standard information retrieval measures for evaluation (DCG, NDCG, MAP) could not be used due to the absence of ground truth. Next, we discuss a novel strategy to evaluate the effectiveness of the proposed evolutionary mutual reinforcement model in identifying highly informative sources.

Search engines are designed to give the most relevant sources for a query. Presumably, the source contain relevant information to the query. This relevant information could be considered as relevant entities for an event, when the search query is an event. Hence, the higher a source is ranked by the search engine, the more informative or specific it should be to the event. We use this notion to evaluate the proposed model.

We compare closeness ($\tau(e_i)_{E_j}$) scores of the entities obtained from the sources ranked according to the search engines, conventional model, and the proposed model. From the three ranked lists (search engine, the proposed evolutionary model, and conventional static model), each source is visited and the entities are extracted. The '$\tau(e_i)_{E_j}$' values are assigned to these entities by referring to the respective event dictionary (ς_{E_j}) and are ranked in descending order. Hence we obtain lists of same entities arranged differently depending upon the ranking of the sources. In order to compare the gain of information from the three lists we take the top 'K' entities from each list and calculate the sum of their '$\tau(e_i)_{E_j}$' values and plot them against the value of 'K' in Figure 2. We start from K=1 and go on increasing its value till the number of entities are exhausted in all the three lists. It is evident from Figure 2 that the information gain is fastest for the proposed model. Figure 2 shows that information about the Egyptian Revolution is gained quicker using our model, which could identify specific sources earlier than the search engines as well as the conventional model. Similar results were also obtained for the other two events (Libyan and Tunisian Revolution). This implies that the sources ranked higher by the proposed model are more specific than the ones ranked by the search engines and the conventional model. These highly specific sources are also very informative. As presented earlier they help in learning useful information about the event due to the presence of close entities. As we have already observed that these sources are often Long Tail sources, we can conclude that Long Tail sources that are ranked lower by the search engines, if identified accurately often provide more specific information than the highly ranked short head sources.

VII. Conclusions

In this paper, we highlighted the need for exploring the social media sources to study an event. We demonstrated that social media sources have the capability to provide very specific and novel information. However, the sheer volume of social media sources and the Long Tail characteristics (e.g., link sparsity, colloquial language, etc.) make it extremely challenging to identify the specific sources. Towards this direction, we developed a methodology that utilizes relevant entities as a mechanism to identify specific sources in a mutual reinforcement framework. To consider the dynamic relationship between the specific sources and close entities, the methodology leverages an evolutionary mutual reinforcement principle. Experiments conducted on real-world datasets demonstrate faster convergence and better accuracy of the evolutionary mutual reinforcement model over the conventional mutual reinforcement model. Furthermore, the evolutionary mutual reinforcement model outperformed one of the most-widely used search engines, i.e., Google Blog Search and IceRocket. It was observed that the search engines ranked the specific sources surprisingly low, thereby reducing the chances of their discovery. The poor hyperlink connectivity of these Long Tail social media sources was contemplated to be a big reason behind their low ranks in traditional search engines.

References

[1] L. Adamic et al. Power-law distribution of the world wide web. *Science*, 287(5461):2115–2115, 2000.

[2] N. Agarwal et al. Identifying the influential bloggers in a community. In *WSDM*, pages 207–218. ACM, 2008.

[3] C. Anderson. *Long Tail, The, Revised and Updated Edition: Why the Future of Business is Selling Less of More*. Hyperion, 2008.

[4] H. Becker et al. Selecting quality twitter content for events. In *ICWSM*, 2011.

[5] S. Brin et al. The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 30(1):107–117, 1998.

[6] F. Cheong et al. Social media data mining: A social network analysis of tweets during the 2010–2011 australian floods. In *PACIS*, 2011.

[7] N. Diakopoulos et al. Finding and assessing social media information sources in the context of journalism. In *Human Factors in Computing Systems*, pages 2451–2460. ACM, 2012.

[8] G. Erkan et al. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res. (JAIR)*, 22:457–479, 2004.

[9] G. Golub et al. *Matrix computations*, volume 3. Johns Hopkins University Press, 1996.

[10] M. Gupta et al. Evaluating event credibility on twitter. In *SDM*. Citeseer, 2012.

[11] N. Hamdy et al. Framing the egyptian uprising in arabic language newspapers and social media. *Journal of Communication*, 2012.

[12] T. Haveliwala. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *TKDE*, 15(4):784–796, 2003.

[13] T. Johnson et al. Wag the blog: How reliance on traditional media and the internet influence credibility perceptions of weblogs among blog users. *Journalism & Mass Communication Quarterly*, 81(3):622–642, 2004.

[14] J. Kleinberg. Authoritative sources in a hyperlinked environment. *JACM*, 46(5):604–632, 1999.

[15] A. Langville et al. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2004.

[16] LOmariba. Is new media posing a serious challenge to traditional media?". Technical report, University of Westminster, 2009.

[17] D. Mahata and N. Agarwal. What does everybody know? identifying event-specific sources from social media. In *CASoN*, pages 63–68. IEEE, 2012.

[18] A. Marcus et al. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Human factors in computing systems*, pages 227–236. ACM, 2011.

[19] A. Popescu et al. Extracting events and event descriptions from twitter. In *World Wide Web*, pages 105–106. ACM, 2011.

[20] T. Rattenbury et al. Towards automatic extraction of event and place semantics from flickr tags. In *SIGIR*, pages 103–110. ACM, 2007.

[21] S. Reese et al. Mapping the blogosphere professional and citizen-based media in the global news arena. *Journalism*, 8(3):235–261, 2007.

[22] V. Singh et al. Mining the blogosphere from a socio-political perspective. In *Computer Information Systems and Industrial Management Applications*, pages 365–370. IEEE, 2010.

[23] R. Troncy et al. Linking events with media. In *International Conference on Semantic Systems*, page 42. ACM, 2010.

[24] S. Vieweg et al. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Human factors in computing systems*, pages 1079–1088. ACM, 2010.

[25] A. Younus et al. What do the average twitterers say: A twitter model for public opinion analysis in the face of major political events. In *ASONAM*, pages 618–623. IEEE, 2011.