

A Momentum Theory for Hot Topic Life-cycle: A Case Study of Hot Hashtag Emerging in Twitter

L. Wang, X. Li, L.J. Liao, L. Liu

Liu Wang, Xin Li, Le-Jian Liao*, Li Liu

School of Computer Science

Beijing Institute of Technology

No. 5 South Zhongguancun Street

Beijing, 100081, China

wangliu2000@163.com, xinli@bit.edu.cn,

liaolj@bit.edu.cn, liuli0407@hotmail.com

*Corresponding author: liaolj@bit.edu.cn

Abstract:

The existing work on mining of hot topics is mainly based on topic multiplicity and attention from users in unit time. With the advent of social networking, the weight has been put on the hot topics which can effectively describe the importance and hotness of a topic. However, the researches on the influence exerted by the accumulation of attention towards hot topics and the alternation between hot topics and outdated ones are still relatively weak. In this paper, a novel algorithm for calculating the hotness of topics is proposed based on momentum. The number of the participants, but also the long tail effect of the historical accumulation on the topic is taken into consideration. Through this algorithm, we can accurately build a model for the hot topics on their emerging growing period and effectively describe the whole life circle of the topic. Additionally, the change between hot topics and old ones can be distinguished efficiently. Our experiments show that the process of a topic growing into a hot topic can be detected explicitly. Potential hot topics can be explored and the overdue ones can be rejected respectively.

Keywords: hashtag, hot topic, aging theory.

1 Introduction

In modern times social networking has become an important resource of real-time news updates. As smart phones and other mobile devices spread, there is a growing tendency that people use social networks such as Twitter and Weibo to obtain the hot issues happening across the world. Detecting the hot topics out from the large-scale of information posted online simultaneously is of significant interest for many reasons. For one, it shortens the time for users to obtain the hot topic, which may play an important role in decision-making. Also from the hot topics detected, one can easily have an understanding of current social dynamics.

Most of the social network systems may provide a ranking list of hot topics through the search or post count of keywords. While this approach fails to take the temporal relations of hot topics into account. Current hot topic detection solutions are mostly based on the topic multiplicity and attention from users in unit time, whereas they neglect the digestion of the hot topics. Also they aim to establish the topics from a clustering of keywords from the content layer, which overlooked the functions of a social network. From the point of view of research methods analysis, physics methods are applied in many fields, such as: Newton's theorem was applied to the [6, 18], the gravity was applied to manufacturing modeling [21], the application of the theorem of momentum was used in [26, 34], [15] combined with momentum method to launch the stock prediction. They carried out the research of modeling by physics methods. From the social networking features, in addition to the security and network mining research [11, 28, 31, 35], also, many scholars are

still studying hashtags' other functions [27, 40], in [1, 37], Xiao and Aldhelaan were using social network to do the research on hot topic discovery and recommendation, Chen [13] combined with evolution model to achieve the topic prediction. The accumulation of topics in social networks and the integration of the topic have strong physics characteristics, so we try to find a way to use physics methods to model the topic in social networks and combine the momentum theorem with hashtags.

In this paper, we aim to detect those emerging hot topics according to the momentum theory with the use of hashtags. The momentum of hashtags can reveal the mechanism of dynamics from the temporal characteristics and quantity characteristics and discover the ideal life cycle of topics. This will result in using momentum theory to dig out hot topics more accurately and effectively. In this paper, the proposed algorithm does not need to collect and store a large number of historical data. Therefore, it is suitable for data streams real-time calculation. Moreover, the nested loops in the algorithm are few and the time complexity is approximate to $O(n)$. Due to high computational efficiency, this algorithm is suitable for large data analysis as well.

The rest of the paper is organized as follows: In Section 2, we give a review of related works. In Section 3, we propose the definition of hot topics. Section 4 describes the theoretical model of topic hotness. In Section 5, we discuss the results of the experiments run. Finally, in Section 6, we present our conclusions and some future research directions.

2 Related Work

Topic Detection and Tracking (TDT) [2] has long been a foundation of the research related to hot topic detection. However, different from the traditional sources of information such as Web pages, texts, etc., information in social networks is often very short and sparse, also spreading rapidly. Under the background of the new era of social media, a series of work have been completed towards those characters. You et al. [7] utilized frequent item set mining algorithm SaM [8] to find out hot topics denoted by combinations of keywords. Thus, detecting the hot topics is comparable to mining frequent patterns from news streams. Similar to that, a recent work by Kim et al. [22] also took geographic elements into consideration, which provides a simple but useful approach to analyze real-time streaming data and finds geographic communities. According to FP-growth [20], Giannella et al. [19] proposed the FP-stream time window and the digestion concept of the support parameters. This new algorithm can let the outdated items expired and accumulate the importance of items based on the timeline as well. However, the algorithm is more complicated. Guo et al. [24] improved the tree structure in the Frequent Pattern stream mining algorithm (FP-stream) [19] and used the new algorithm to detect hot topics from twitter streams, which can lead to time-sensitive results. Lee et al. [17] developed an algorithm for ranking topics, using topic energy to represent the significance of a given topic at each time point within a time period. The strategy in determining topic energy value considers factors such as popularity, burstiness and informativeness, which suits the character of information on social network better. In [29], the key entity significance is computed through traditional "tf-idf" evaluation method [33] in the information retrieval literature. Then, clustering entities are used to generate significant events. Bun et al. [25] computed the value of a term by using tf*idf and clustered terms into a sentence. Yang et al. [39] applied the VSM to the task of news TDT and used a time window with a decaying function (TW-DF) to model the temporal relations between documents and events. Unfortunately however, the strategies above did not consider the temporal relations thoroughly. The newly generated hot topics cannot be distinguished from the outdated ones accurately.

Chen et al. [12, 16] proposed an aging theory to model life cycles of news events. For all user

messages generated from the topic time interval, the aging theory calculates their nutrition and convert the nutrition value to the energy value, and then add into the cumulative energy value of the topic. It also applies energy decaying strategy on the topic. As time goes by, the energy decays gradually. If the increase of the energy is less than the reduction of the energy, the topic will show a trend of attenuation. If the energy value is less than a given threshold, this topic is set to "death" state and is removed from the hot topics list. The aging algorithm defines α as the nutrition transferred factor and β as the nutrition decayed factor, $0 < \alpha < 1$, $0 < \beta < 1$, α decides the increase of nutrition from an input news document and β decides the nutrition loss in a period. Zheng et al. [41] utilized aging theory on candidate topics discovered by a clustering method in each time slot in BBS. With energy values updated at the end of each time slot according to the three functions of aging theory, hot topics in each time slot can be easily found. Chen et al. [14] combined aging theory with a term weighting scheme to extract genuine hot terms. Based on the extracted hot terms, key sentences are then identified and grouped into clusters that represent hot topics by using multidimensional sentence vectors. Cataldi et al. [10] improved the nutrition formula by using "idf" and formalized the keyword life cycle leveraging a novel aging theory intended to mine terms that frequently occur in the specified time interval which are relatively rare in the past. Wang et al. [36] took media focus and user attention into the classical aging theory and proved the importance of these two factors. In comparison with these approaches, our solution extracts hot topics based on hashtags according to the functions of a social network instead of the content aspect. The hot topic emerging theory based on momentum can explicitly distinguish new hot topics from outdated ones.

The methods mentioned emphasized topic extraction from a cluster of sentences or the frequency pattern of topics, but ignore the topic life cycle. The aging theory [12, 16, 41] discovers the life cycle of topic but does not deal with the older topic elimination mechanism very well. Chen et al. [23] proposed a similar idea to our model, defining hot velocity and hot acceleration to recognize hot topics but did not reveal the dynamic characteristics. Different from that model, we define hot topic emerging theory based on momentum. The utilization of momentum theory can successfully describe the whole life circle of a topic, thus sift out the newly emerged hot topics from the old ones more effectively.

3 Definition of Hot Topic

Hot topics in real microblogging systems have three characteristics: 1) the number of posts related to this topic would exceed a threshold; 2) the amount of users concerned should be large, especially those key persons; 3) a hot topic would occur at a short time [23]. In [14], the characteristics of hot topics are concluded based on [25] as those appear on many news channels and go through a life cycle of birth, growth, maturity, and death. In this paper, hot topics are defined as the topics which are both influential and latest. Four characteristics are summarized as follows:

- i Timeliness. The hot topic refers to the people and things that have happened recently. Those topics which have been under discussion for more than seven days or one month cannot be regarded as hot topics.
- ii Development. Due to the dynamics while topics are propagated, the influence of public opinion will spread as time goes by. The hot topics diffract backwards.
- iii Accumulation. The more people are involved into the discussion of the topic, the hotter the topic is. However with more and more attention drawn on a certain topic, it becomes less possible to be hot topic again.

- iv Digestion. As time goes on, the influence of a topic will unavoidably be watered-down. Thus hot topics have the character of obliviousness.

4 Kinetic Model of Topic Hotness

The topic under a certain time point can be regarded as a set of large number of synonymous terms. In other words, to describe this concept from the perspective of physics, we can assume each term as a particle which owns weight and velocity. Besides, the activeness of the particle represents its energy and the activeness of a topic is the reflection of the particles inside the set. As time goes on, the set is dynamically changing with terms increasing, thus the energy of the particles are changing too. In time $t - 1$, the topic set is represented as $Topic_{n(t-1)}$. In time t , the additional subset $Topic_{n\Delta t}$ will join into $Topic_{n(t-1)}$ as the new set $Topic_{nt}$. This process can be regarded as two kinetic physical objects collided and fused into one kinetic physical object with the exchange of energy. In this paper, we adopt conservation of momentum to illustrate this process.

In this paper, we propose to use momentum to represent the physical characters in thermodynamics and dynamics of hot topic. Negative acceleration is imported to represent the digestion of hotness according to the attenuation of the hot topics while propagated. We use momentum equation to present the active momentum of topics which is aroused by discussion.

A topic consists of many synonymous terms. In social networks, a post can be seen as a term, and similar terms constitute a topic. The topic set is varying all the time and the variation composes a time sequence. For a topic n , the topic set can be expressed as:

$$Topic_n := \{Topic_{n1}, Topic_{n2}, \dots, Topic_{nt}\}$$

All kinds of topics are formed as the whole topic set:

$$Topic := \{Topic_1, Topic_2, \dots, Topic_n\}$$

The current topic set is the combination of the former set and the subset increased during the interval, as:

$$Topic_{nt} := Topic_{n(t-1)} \cup Topic_{n\Delta t}$$

$$Topic_{n\Delta t} := \{term_{n1}, term_{n2}, \dots, term_{nm}\}$$

We can also say that the topic under a specific time spot is the set of all the terms emerged before, as:

$$Topic_{nt} := \{term_{n1}, term_{n2}, term_{n3}, \dots, term_{nt}\}$$

4.1 Momentum Modelling

In this paper, we use the equations below to represent the variation when a topic becomes hotter. The topic in $t - 1$ can be seen as a physical entity m_{t-1} with weight and velocity. The increased terms in under that topic can be regarded as another entity $m_{term_{\Delta t}}$. Thus the topic in t as an entity m_t is the combination of $m_{term_{\Delta t}}$ and m_{t-1} after a collision with the velocity of $v_{term_{\Delta t}}$.

The increased momentum of topic entity $m_{term_{\Delta t}}$ is the sum of the momentum of each term added. We can calculate the momentum of each topic as follows:

$$\vec{M}_{term_{\Delta t}} = \sum_{\Delta t=1}^m (m_{term_{\Delta t}} \times \vec{v}_{term_{\Delta t}}) \quad (1)$$

The momentum of the M_t equals to the sum of the momentum of m_{t-1} and $M_{term_{\Delta t}}$. The total momentum is invariable before and after the collision. $\vec{v}_{t-1} + \vec{g}\Delta t$ represents the final velocity after Δt . The development and digestion of topic can be expressed clearly by this equation:

$$\vec{M}_t = m_{t-1} \times (\vec{v}_{t-1} + \vec{g}\Delta t) + \vec{M}_{term_{\Delta t}} \quad (2)$$

The weight of the topic in time t is the sum of the weight of itself before collision and the weight of the topic added, which shows the accumulation of topic. The more the topic is discussed, the heavier the topic is.

$$m_t = m_{t-1} + \sum_{\Delta t=1}^m m_{term_{\Delta t}} \quad (3)$$

In time t , the velocity of the topic is the quotient of total momentum and total weight.

$$\vec{v}_t = \vec{M}_t / m_t \quad (4)$$

The topic reaches a height in a certain velocity after being represented in physics. In this paper we define this height as the hotness of the topic in that time spot.

$$\vec{H}_t = \vec{v}_{t-1} \times t \quad (5)$$

The height is not able to keep growing according to the digestion of the topic. Under this circumstance we use acceleration of gravity g to lower the height (hotness). The equation is modified as follows:

$$\vec{H}_t = \vec{v}_{t-1} \times t_0 + \frac{1}{2} \times \vec{g}\Delta t^2 \quad (6)$$

4.2 Modelling Solution

In social networks, Hashtag (#) is used extensively to determine the topic, which is convenient to aggregate and classify vast amounts of information and let people who follow a certain topic get the relevant information more easily. Twitter showed its significant value in information propagation almost before every emergency and important activity due to the aggregation of using hashtags. We can draw a safe conclusion that hashtag reflects the true tendency of topics and has great potential in the reconstruction and compilation of information. In this paper we use hashtag for identification modeling of topics. A topic under a specific time spot is defined as a set with synonymous hashtags.

$$Topic_{nt} := \{hashtag_{n1}, hashtag_{n2}, hashtag_{n3}, \dots, hashtag_{nt}\}$$

At a specified time, more similar hashtags means the topic is more active. In a specific period of time, topic is a sequence of sets. In this Section we use m for the weight of hashtag, v for the velocity of hashtag, and M for the momentum of hashtag. The topic hotness modeling algorithm is as follows:

5 Experiment Analysis And Result

5.1 Data Corpus and Parameter Settings

In this paper we adopt the historical(2009/6/11-2009/12/31) data from Twitter as corpus. We extracted 460,496 twitter terms and 22,063 hashtags in total, neglecting those topics which lasted

Algorithm 1 Topic Hotness Modeling**Input:** The set of hashtags which emerged during a certain period of time**Output:** The topics list ranked by hotness in each time spot t

```

1: for  $t = t_0$  do
2:   Process the data we collected and aggregate into topic set  $Topic$ .
3:   for each topic  $Topic_n$  from  $Topic$  do
4:     Obtain the set  $\{term_{n1}, \dots, term_{nt}\}$  of the  $Topic_{nt}$ 
5:     Calculate the hotness of  $Topic_{nt}$  by  $\vec{H}_t = \vec{v}_{t-1} \times t_0 + \frac{1}{2} \times \vec{g} \Delta t^2$ 
6:   end for
7:   output the topic list ordered by  $H_t$ 
8:    $t = t + 1$ 
9: end for

```

Table 1: Parameter Setting

Parameter	Value
Hashtag Initial Velocity v	10
Hashtag Weight m	1
Acceleration g	-3
Interval t (day)	1

less than 3 days or contained less than 2 posts, the coverage can meet the needs of simulation experiments. The main parameters of the computer for this modelling are as follows: CPU is Intel(R) Core(TM)2 Duo, Main Frequency is 2.0GHz, Memory Frequency is 777MHz, Memory Capacity is 1.96GB, Running Environment is WINDOWS XP. The rest parameters are set as the Table 1.

5.2 Comparison of Hot Topic detection

In this experiment, our momentum theory (M) is compared to two proposed methods. The baseline method (A) [12, 16] is a basic aging method. Cataldi et al. [10] improved a method (A-TF) which enhanced the aging method by using an augmented normalized term frequency. As a result, top 3 and top 10 generated hot topics from each method are evaluated using official TDT measures including: precision (p), recall (r) and F1-measure ($F1$).

In Table 2, the best score for each item is represented in bold. In the top 3 comparison, our momentum theory achieves both highest precision and recall which results in the best $F1$ score, while the aging method achieves both reasonable precision and recall. In the top 10 comparison, the precision of our momentum theory is still the highest in the first 10 hot hashtags but loses

Table 2: Comparison of Three Methods

	Method	p	r	$F1$
Top 3	A	0.67	0.67	0.67
	M	1.00	0.67	0.80
	A-TF	0.50	0.50	0.50
Top 10	A	0.63	0.71	0.67
	M	0.88	0.78	0.82
	A-TF	0.56	0.83	0.67

Table 3: Top 3 Hashtags Changing Circumstances in A Week in Three Methods

	2009/7/6	2009/7/7	2009/7/8	2009/7/9	2009/7/10	2009/7/11
M	mileycyrus 140army 140mafia	mileycyrus 140army 140mafia	mileycyrus 140army gorillapenis	140army mileycyrus gorillapenis	140army zyngapirates urumqi	zyngapirates hottest100 ashes
A	livestrong iranelection 140mafia	moonfruit nothingpersonal iranelection	turnon turnoff iranelection	notagoodlook iranelection 140mafia	ff followfriday iranelection	unacceptable iranelection ff
A-TF	livestrong iranelection musicmonday	moonfruit iranelection livestrong	turnon turnoff iranelection	iranelection 140mafia notagoodlook	ff followfriday iranelection	unacceptable iranelection ff

recall. We believe that is due to the fact that the momentum mechanism likes to remove the old topics. The recall of A-TF is the best, because it improves the aging algorithm which will be sensitive of hot topics and enlarge its coverage, but there is a reduction in the precision. In the *F1* item, the momentum theory has maintained the highest standards. In conclusion, the comprehensive performance of momentum theory is better than the other two algorithms.

5.3 Comparison of Hot Topic Trending

Due to space limitations, Table 3 represents the top 3 hashtags from 2009/7/6 to 2009/7/11 in the momentum theory (M), the aging algorithm (A) and the improved aging algorithm (A-TF). These results show that the topic trending in M is relatively stable. Hashtag #mileycyrus, #140army and #zyngapirates are stable up to first place in turn. In the last two algorithms, #iranelection is ranked in the top 3 places. However, according to the historical data checking, this hashtag appeared over 20,000 times a day in June and was already ranked as hottest topic at that time in all three algorithms. Clearly, in July #iranelection should not be a hot topic again. Hashtag #140mafia's overall performance is good, the repeating number is higher and the potential of impacting as hottest topic is great, so it is ranked as top 3 in the three algorithms. Although it is stable in M, it is unstable in A and A-TF.

Table 4 presents the top 10 hashtags in two days (2009/7/6-2009/7/7) in three algorithms. Hashtag #mileycyrus remains first place in M in two days, ranks 7th in A, falls out of the top 10 in A-TF and only ranks 17th as 0.635 energy value. It is because there are a lot of old hot topics which interfere with these two days of ranking, while these old hashtags already were the hot topics in an earlier time. #140army is in second place in M, but ranks 13th and 11th in A and A-TF separately, falling out of top 10. #flip2009 is captured only by M, while it gets its highest times in 5th July and 6th July. But due to relatively lower times, it is abandoned by two aging algorithms. As we mentioned before, #140mafia is captured in all three methods. #tcot, #spymaster and #tweetmyjobs had more than thousands of times per day from 2009/6/12 and were listed as hot topics in momentum theory at that time. These are not hot news yet, but these hashtags are still ranked as top 10 hashtags in A and A-TF. #mj is a standard hot hashtag which arise sharply in 7th July as 1,077 times occur, however, it is replaced by #tcot in A and A-TF due to the shortcoming of lower discrimination degree of the aging theory.

Because of the large amount of data and limited space, we take hashtag #mileysycyrus as an example to analyze the trending of different algorithms. Figure 1 Shows the times, hotness and energy value six months profiles of hashtag #mileysycyrus, which is the hottest topic of 6th July in momentum theory.

In Figure 1 A, this hashtag began to appear in 12th June. After that, it did not reproduce for about a week. Later, it appeared sometime in a lower rate. On 5th July, it shot up suddenly

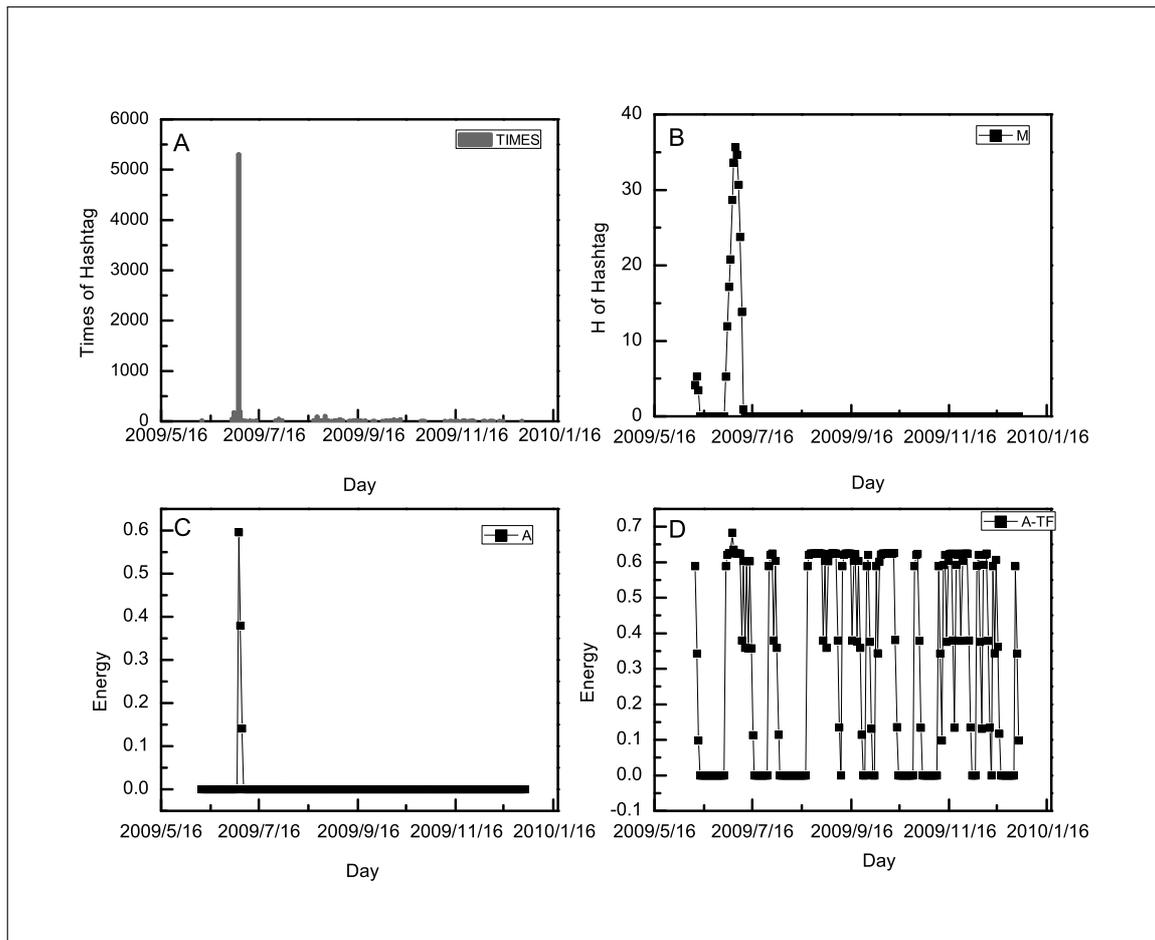


Figure 1: Comparison of #mileysycyrus (A:occurrence number of hashtag per day, B:hotness of hashtag per day in momentum theory, C:energy value of hashtag per day in aging method, D:energy value of hashtag per day in A-TF method)

Table 4: Top 10 Hashtags in Two Days in Three Methods

	Hashtag	M	Times	Hashtag	A	Times	Hashtag	A-TF	Times
7-6	mileycyrus	33.63	187	livestrong	0.56	4026	livestrong	0.67	4026
	140army	27.93	781	iranelection	0.55	2640	iranelection	0.66	2640
	140mafia	27.68	2316	140mafia	0.53	2316	musicmonday	0.66	2966
	moonfruit	26.23	2124	moonfruit	0.51	2124	140mafia	0.66	2316
	gorillapenis	24.98	984	musicmonday	0.50	2966	moonfruit	0.66	2124
	flip2009	21.19	12	spymaster	0.47	1616	spymaster	0.65	1616
	mj	20.86	288	tcot	0.40	1206	tcot	0.65	1206
	rw09	20.64	16	gorillapenis	0.39	984	gorillapenis	0.64	984
	katemcrae	19.36	10	mileycyrus	0.38	187	militarymon	0.64	933
	cmonbrazil	19.11	140	honduras	0.30	676	jobs	0.64	827
7-7	mileycyrus	35.64	18	moonfruit	0.62	5147	moonfruit	0.64	5147
	140army	30.88	461	nothingpersonal	0.54	3670	iranelection	0.63	1902
	140mafia	29.23	1644	iranelection	0.51	1902	livestrong	0.63	1228
	gorillapenis	29.09	559	140mafia	0.49	1644	140mafia	0.63	1644
	moonfruit	25.65	5147	livestrong	0.47	1228	musicmonday	0.63	1423
	cmonbrazil	20.13	47	musicmonday	0.46	1423	spymaster	0.63	936
	urumqi	19.61	153	spymaster	0.40	936	tcot	0.63	839
	mj	19.27	1075	tcot	0.36	839	tweetmyjobs	0.63	1013
	crocmint	19.27	6	tweetmyjobs	0.34	1013	gorillapenis	0.63	559
	xinjiang	19.05	110	threadless	0.30	1192	jobs	0.63	642

and reached 5,302 times, then declined rapidly. Although it repeated a few dozens or hundreds times later, but never rose again sharply. In Figure 1 B, momentum theory captures the hotness rising rapidly on 5th, 6th, 7th July and declining dramatically quickly. This hotness is higher than other hashtags, so it is ranked in first place. The aging theory also finds this change but does not list this hashtag in a higher position (only the 9th place) due to a relative lower energy value compared with other hashtags. The A-TF method almost gets every higher repeat rate moments, but it cannot distinguish the highest point from all higher points clearly. So, it is not able to choose this hashtag from the data corpus efficiently. As a result, it falls out of the top 10 hashtags. In Figure 1, momentum theory presents a higher sensitivity and better ability to discriminate the hottest hashtags from others than the other two algorithms.

Extrapolating the results from the experiment, we can obtain the hottest topics during the period of time we captured by the algorithm we proposed, and compared with the hot hashtags that are extracted by the rank of occurrence.

From the comparison of Table 5, we can explicitly observe that the occurrence of the hashtag may mislead to the generation of a hot topic, such as #followfriday, #ff and #1. Those topics own little valid information, however are regarded as hot according to the high frequency of occurrence. By the algorithm we proposed, the topics which lack realistic meaning will be eliminated by the momentum equation. The topics which survive will be the active ones with realistic meaning.

6 Conclusion and Future Work

In this paper we propose a novel algorithm for hot topic detection based on momentum theory using hashtags for defining a topic. The main contributions of the paper are as follows.

Table 5: Hottest Hashtags in Half A Year

	Momentum Theory		Occurrence		
	Hashtag	Hotness	Hashtag	Days	Times
1	tehran	44.73	ff	113	417802
2	nomaschavez	44.19	iranelection	116	368519
3	happybirthdaymikey	43.13	tcot	120	288091
4	nem	43	jobs	115	243586
5	iranelection	41.14	mobsterworld	85	203445
6	zain	38.93	followfriday	113	188664
7	happybirthdaypink	38.2	1	114	166260
8	teenisland	38.15	musicmonday	109	164399
9	blackbery	37.94	140mafia	105	149078

First we analyze the characteristics of hot topics and conclude into four points: 1) Timeliness. 2) Development. 3) Accumulation. 4) Digestion. Then we build a hot topic detection model with momentum theory. The experiments show that our model can identify those emerging hot topics effectively and accurately.

In the future, we hope to filter the posts under each topic to provide the users with a purer source of information without the distraction from irrelevant posts. Our algorithm standardizes the topic life cycle into a very stable curve which makes the topic prediction possible. Some artificial intelligence techniques have been applied in mathematical modeling [4, 32, 38], and some achievements have been obtained in the prediction of data modeling [3, 5, 9, 30]. We will try to use these techniques to predict the hot topics in the next step.

Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful comments. The authors also thank Yanmei Zhai and Xu Han for their help with data preparation and experiments. This work has been partially supported by National Program on Key Basic Research Project under Grant No. 2013CB329605, NSFC under Grant No. 61300178 and Natural Science Foundation of Beijing under Grant No. 4092037.

Bibliography

- [1] Aldhelaan, M.; Alhawasi, H. (2015); Graph Summarization for Hashtag Recommendation, *2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud)*, 698-702.
- [2] Allan, J.; Carbonell, J.; Doddington, G., et al. (1998); Topic Detection and Tracking Pilot Study Final Report, *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, 194-218.
- [3] Asadi, S.; Hadavandi, E.; Mehmanpazir, F., et al. (2012); Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction, *Knowledge-Based Systems*, 35(15): 245-258.

- [4] Aydemir, E.; Korusu, H.I. (2015); A New Production Scheduling Module Using Priority-Rule Based Genetic Algorithm, *International Journal of Simulation Modelling*, ISBN 1726-4529, 14(3): 450-462.
- [5] Bas, E.; Egrioglu, E.; Aladag, C.H., et al. (2015); Fuzzy-time-series network used to forecast linear and nonlinear time series, *Applied Intelligence*, 43(2): 1-13.
- [6] Blekas, K.; Lagaris, I.E. (2013); A Spectral Clustering Approach Based on Newton's Equations of Motion, *International Journal of Intelligent Systems*, 28(4): 394-410.
- [7] Bo, Y.; Ming, L.; Bing-Quan, L., et al. (2012); Detecting hot topics in technology news streams, *Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*, ISBN 2160-133X, 5:1968-1974.
- [8] Borgelt, C. (2010); Simple Algorithms for Frequent Item Set Mining, *Advances in Machine Learning II*, ISBN 978-3-642-05178-4, 263(16):351-369.
- [9] Caldeira, J.F.; Moura, G.V.; Santos, A.A.P. (2016); Predicting the yield curve using forecast combinations, *Computational Statistics & Data Analysis*, 100: 79-98.
- [10] Cataldi, M.; Caro, L.D.; Schifanella, C. (2010); Emerging topic detection on Twitter based on temporal and social terms evaluation, *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, 1-10.
- [11] Chang, M.K.; Cheung, W.; Tang, M. (2013); Building trust online: Interactions among trust building mechanisms, *Information & Management*, 50(7): 439-445.
- [12] Chen, C.; Chen, Y.-T.; Sun, Y., et al. (2003); Life Cycle Modeling of News Events Using Aging Theory, *Machine Learning: ECML 2003*, ISBN 978-3-540-20121-2, 2837(7):47-59.
- [13] Chen, J.; Yu, J.; Shen, Y. (2012); Towards Topic Trend Prediction on a Topic Evolution Model with Social Connection, *The Ieee/wic/acm International Joint Conferences on Web Intelligence & Intelligent Agent Technology*, 153-157.
- [14] Chen, K.Y.; Luesukprasert, L.; Chou, S.c.T. (2007); Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling, *IEEE Transactions on Knowledge and Data Engineering*, ISBN 1041-4347, 19(8): 1016-1025.
- [15] Chen, T.L. (2012); Forecasting the Taiwan Stock Market with a Novel Momentum-based Fuzzy Time-series, *Review of Economics & Finance*, 2:38-50.
- [16] Chien Chin, C.; Yao-Tsung, C.; Meng Chang, C. (2007); An Aging Theory for Event Life-Cycle Modeling, *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, ISBN 1083-4427, 37(2): 237-248.
- [17] Chung-Hong, L.; Tzan-Feng, C.; Hsin-Chang, Y. (2011); An automatic topic ranking approach for event detection on microblogging messages, *Systems, Man and Cybernetics (SMC), 2011 IEEE International Conference on*, ISBN 1062-922X, 1358-1363
- [18] Galperin, E.A. (2011); Information transmittal, Newton's law of gravitation, and tensor approach to general relativity, *Computers & Mathematics with Applications*, 62(2): 709-724.
- [19] Giannella, C.; Han, J.; Pei, J., et al. (2003); Mining Frequent Patterns in Data Streams at Multiple Time Granularities, *Data Mining Next Generation Challenges & Future Directions*.

- [20] Han, J.; Pei, J.; Yin, Y. (2000); Mining frequent patterns without candidate generation, *SIGMOD Rec.*, ISBN 0163-5808, 29(2): 1-12.
- [21] Hrelja, M.; Klancnik, S.; Balic, J., et al. (2014); Modelling of a turning process using the gravitational search algorithm, *International Journal of Simulation Modelling*, ISBN 1726-4529, 13(1): 30-41.
- [22] Hwi-Gang, K.; Seongjoo, L.; Sunghyon, K. (2013); Discovering hot topics using Twitter streaming data social topic detection and geographic clustering, *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, 1215-1220.
- [23] Jiangfeng, C.; Jianjun, Y.; Yi, S. (2012); Towards Topic Trend Prediction on a Topic Evolution Model with Social Connection, *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*, 1:153-157.
- [24] Jing, G.; Peng, Z.; Tanb, J., et al. (2012); Mining Hot Topics from Twitter Streams, *Procedia Computer Science*, 9(11): 2008-2011.
- [25] Khoo Khyou, B.; Ishizuka, M. (2002); Topic extraction from news archive using TF*PDF algorithm, *Web Information Systems Engineering, 2002. WISE 2002. Proceedings of the Third International Conference on*, 73-82.
- [26] Khulief, Y.A. (2010); Numerical Modelling of Impulsive Events in Mechanical Systems, *International Journal of Modelling & Simulation*, 30: 80-86.
- [27] Kotsakos, D.; Sakkos, P.; Katakis, I., et al. (2015); Language agnostic meme-filtering for hashtag-based social network analysis, *Social Network Analysis & Mining*, 5(1): 1-14.
- [28] Li, M.; Tang, M. (2013); Information Security Engineering: a Framework for Research and Practices, *International Journal of Computers Communications & Control*, 8(4): 578-587.
- [29] Liu, M.; Liu, Y.; Xiang, L., et al. (2008); Extracting Key Entities and Significant Events from Online Daily News, *Intelligent Data Engineering and Automated Learning-IDEAL 2008*, ISBN 978-3-540-88905-2, 5326(26):201-209.
- [30] Liu, C.H.; Xiong, W. (2015); Modelling and Simulation of Quality Risk Forecasting in a Supply Chain, *International Journal of Simulation Modelling*, ISBN 1726-4529, 14(2): 359-370.
- [31] Ma, H.; Lu, Z.; Li, D., et al. (2014); Mining hidden links in social networks to achieve equilibrium Aî, *Theoretical Computer Science*, 556:13-24.
- [32] Ramesh Kumar, L.; Padmanaban, K.; Balamurugan, C. (2016); Optimal Tolerance Allocation in a Complex Assembly Using Evolutionary Algorithms, *International Journal of Simulation Modelling*, ISBN 1726-4529, 15(1): 121-132.
- [33] Salton, G. (1989); *Automatic text processing: the transformation, analysis, and retrieval of information by computer*, ISBN 0-201-12227-8.
- [34] Ternik, P.; Rudolf, R. (2013); Laminar Natural Convection of Non-Newtonian Nanofluids in a Square Enclosure with Differentially Heated Side Walls, *International Journal of Simulation Modelling*, 12(1): 5-16.
- [35] Tran, D.H.; Nguyen, H.L.; Zhao, W., et al. (2011); Towards security in sharing data on cloud-based social networks, *Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on*, 1-5.

- [36] Wang, C.; Zhang, M.; Ru, L., et al. (2008); Automatic online news topic ranking using media focus and user attention based on aging theory, *Proceedings of the 17th ACM conference on Information and knowledge management*, 1033-1042.
- [37] Xiao, Z. (2014); A SOCIAL NETWORK-ORIENTED MINING ALGORITHM FOR HOT TOPIC DATA, *Computer Applications & Software*
- [38] Yang, K.W.; Zhang, P.L.; Ge, B.F., et al. (2015); A Variables Clustering Based Differential Evolution Algorithm to Solve Production Planning Problem, *International Journal of Simulation Modelling*, ISBN 1726-4529, 14: 525-538.
- [39] Yang, Y.; Pierce, T.; Carbonell, J. (1998); A study of retrospective and on-line event detection, *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 28-36.
- [40] Zangerle, E.; Gassler, W.; Specht, G. (2013); On the impact of text similarity functions on hashtag recommendations in microblogging environments, *Social Network Analysis & Mining*, 3(4): 889-898.
- [41] Zheng, D.; Li, F. (2009); Hot Topic Detection on BBS Using Aging Theory, *Web Information Systems and Mining*, ISBN 978-3-642-05249-1, 5854(14): 129-138.