Proceedings of 2012 4th International Conference on Machine Learning and Computing
IPCSIT vol. 25 (2012) © (2012) IACSIT Press, Singapore

Trends Predicting of Topics on Twitter based on MACD

Rong Lu+, Zhiheng Xu, Yang Zhang and Qing Yang

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing 100190, China

Abstract. This paper presents a novel method to predict the trends of topics on Twitter based on MACD (Moving Average Convergence-Divergence), which is one of the simplest and most effective momentum indicator in technique analysis of stocks. The MACD turns two trend-following indicators, moving averages, into a momentum oscillator by subtracting the longer moving average from the shorter moving average. As a result, we monitor the key words of topics on Twitter, and use the longer moving average and the shorter moving average to track their longer and shorter trends, respectively. Then, we redefine the trends momentum with two moving averages according to the developmental characteristics of topics on Twitter, and use it to predict the trends. Experimental results show that the proposed method is very simple and effective.

Keywords: Trends predicting, Topics on Twitter, Social media, MACD

1. Introduction

Twitter is a very popular micro-blogging and social-networking service. Today, more than 160 million users around the world are using it to remain socially connected to their friends, family members and co-workers[1]. It allows users to use a short text within a limit of 140 characters as their posts (also called ***tweets***). It also employs a social-networking model called "***following***"[2], in which the user can choose any other users that she wants to follow without any permission or reciprocating by following her back. The one she follows is her ***friend***, and she is the ***follower***. Being a follower on Twitter means that she receives all the updates from her friends[3]. The short text and social-network functionality make Twitter very easy for topics diffusion. Some topics on Twitter may get a lot attention. Many users will ***retweet***(share to their followers) the tweets about the topics to their followers. The reasons why the topics become hot are very complicated. However, we still want to know what topic will become a hot one on Twitter in the future.

Traditionally, researches about trends of topics mainly focus on emerging trends detection(ETD). An emerging trend is a topic area that is growing in interest and utility over time. For example, Extensible Markup Language(XML) emerged as a trend in 1990s[4]. Porter et. al. proposed an emerging trends detection system called Technology Opportunities Analysis System(TOA) in 1995, which illustrates the range of information profiles possible by examining research and development publications and patents. There are also other emerging trends detection system, like Timeminner[5], Patent Minner[6], HDDI[7]. These methods were try to find the emerging trends in the text database and visualize the evolution of topics.

Recently, LDA-style Topic Model becomes one of the hottest research spot in machine learning and information retrieval. It was also introduced in emerging trends detection. However, the original LDA [8] is independent with time. So, several topic models with temporal information were proposed, such as DDTM (Discrete-time Dynamic Topic Model)[9], CDTM(Continuous Time Dynamic Topic Model)[10], TOT(Topics over Time)[11] and TAM(Trend Analysis Model)[12]. DDTM requires that time be discredited, and the complexity of variational inference for the DDTM grows quickly as time granularity increases. The CDTM improve DDTM by using Brownian motion to model the latent topics through a sequential collection

+ Corresponding author. Tel.: + 18611643446.
E-mail address: rlu@nlpr.ia.ac.cn.

of documents. TOT considers the mixture distribution over topics is influenced by both word co-occurrences and the document's timestamp. TAM focuses on the difference between temporal words and other words in each document to detect topic evolution over time.

All the researches about ETD mainly pay attention to the evolution of topics or long-term changes of trends. However, our work focus on the short-term changes of trends. We want to predict wheather a topic on Twitter will be popular in the next few hours or it will die. The prediction is real-time.

For the purpose of predicting trends of topics on Twitter, we borrow a widely used indicator in stock technique analysis: MACD, which was developed by Gerald Appel in the 1960s. The MACD turns two trend-following indicators, moving averages, into a momentum oscillator by subtracting the longer moving average from the shorter moving average. As a result, the MACD offers the best of both worlds: trend following and momentum.

So, we first monitor some key words of topics on twitter, and compute two different time span moving averages in real-time. Then we subtract the longer period moving average from the shorter one. The difference value is defined a **trends momentum** of a topic. Finnally, we use the trend momentum to predict the trends of topics in the future. Experimental results show that, our method is simple and very effective.

2. Trend Momentum

In this section, we give the formal definition of trends momentum. As described above, the trends momentum is related to two moving averages. So, we first give the definition of **moving average**. We divide the continuous time into several successive equal-sized time slice. In time slict t_i, the occurrence of a key word is denoted as $f(t_i)$, and the time-window size of the moving average is k. At the **n***th* time slice, the moving average($MA(n,k)$)is:

$$MA(n,k) = \frac{\sum_{i=n-k+1}^{n} f(t_i)}{k} \tag{1}$$

If $n < k$, the moving average is defined as:

$$MA(n,k) = \frac{\sum_{i=1}^{n} f(t_i)}{n} \tag{2}$$

The moving average can track the trends of topics very well. Moreover, different sized moving average can track different period of trends. As the original MACD, the **trend momentum** is defined as:

$$TM(n) = MA(n,k_s) - MA(n,k_l) \tag{3}$$

k_s is a shorter sized time window, while k_l is a longer sized time window. The trend momentum is the just the value subtracting the longer moving average from the short one. However, the characteristic of topics on Twitter is a little different from the price of stocks. When the longer moving average of a topic represented by a key word is high for a while, the topic often last for a long time period as a very hot topic. But, if the price of a stock is high for a while, it is probably just the beginning of falling down. So we redefine the trend momentum as follows:

$$TM(n) = MA(n,k_s) - (MA(n,k_l))^{\alpha} \tag{4}$$

where $0 < \alpha < 1$.

In Equation 4, we discount the longer moving average by an exponent α. The longer moving average is the reduced item of trend momentum. So, when a topic is very hot, even if its longer moving average is high, its trend momentum will also be bigger than other topics.

However, the value of trend momentum defined in Equation 4, is volatile. So, in practice, we also use moving average to smooth it:

$$Momentum(n) = MA(TM(n), k) \tag{5}$$

In MACD analysis of stock, there are several rules for predicting the future trends of the price. We choose the simplest and most effective rule. When the value of trend momentum of a topic changes from negetive to positive, the trends of the topic will rise. When it changes from positive to negetive, the topic is dying.

3. Data Preparation

For the purpose to validate the proposed method. We crawled two different datasets from Twitter. One is news event dataset，in which we consider every news event as a topic on Twitter. And the other is Twitter trends(hot topics provided by Twitter itself every hour) dataset.

For the news dataset, we first crawled the headlines of top news events from the RSS[1] of The Associated Press website. At the same time, the tweets which match all the words in the headlines are crawled from Twitter. In this dataset, there are 1118 headlines, and more than 450 thousand tweets.

For the Twitter trends dataset, we use the Stream API[2] provided by Twitter to sample about 1% tweets of all all public statuses on Twitter. Meanwhile, we also crawled the Twitter trends in the same period. In this dataset, there are more than 20 million tweets, and 1072 trends topics. The two datasets is used for experements and evaluation in Section 4.

4. Experiments and Evaluation

In this section, we first describe two cases to show how our trends prediction method works. Then a quantitative evaluation is given.

In Figure 1, this case is selected from the news dataset. And the news headline is "PayPal cuts WikiLeaks from money flow", which was released by the Associated Press at 13:15 on December 4th, 2010. At about 80 minutes after the news first appeared on Twitter, the trend momentum of this news event first time changes from negetive to positive. Then the tweets talk about the news event is significant increase after that time point. And at about the 230 minutes, the trend momentum changes from positive to negetive. Then, very few tweets still talk about it.

[1] http://hosted.ap.org/lineups/TOPHEADS-rss_2.0.xml?SITE=KFWB&SECTION=HOME
[2] https://dev.twitter.com/docs/streaming-api/methods

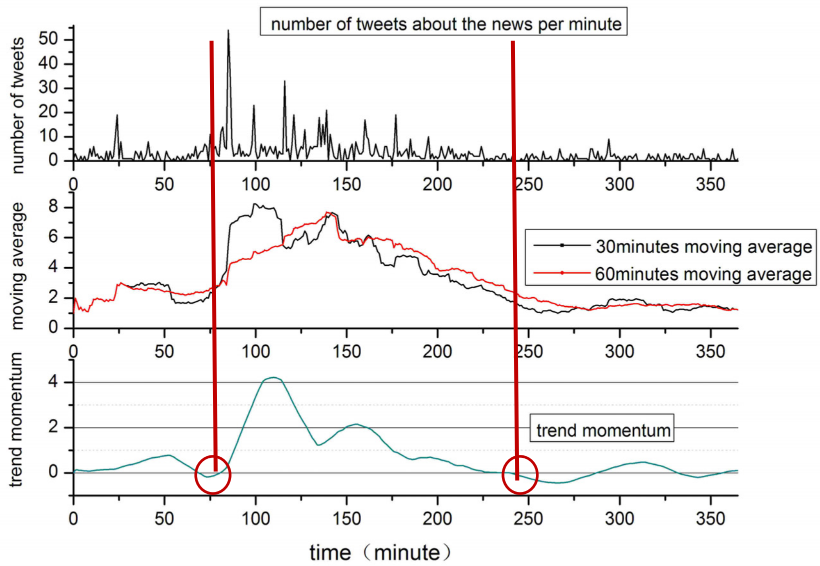

Fig. 1: Trend momentum of a news coverage.

The second case is from the Twitter trends dataset. The key word is “ipad”, which becomes the Twitter trends topic at 22:00, March 24th. Its trend momentum is shown in Figure 2. At the 34th hour(10:00, March 24th), the value of trend momentum changes from negetive to positive. According to our method, we predict the trends will rise at this time point. The moving average then stay high after that as expected. After 12 hours, Twitter chose the word “ipad” as a hot topics, when the topic “ipad” had already begun to cool down. This case shows a strong contrast between the hot topics selection method of Twitter and ours. Our method has a very good forward-looking feature, while Twitter’s method is lagging far behind.

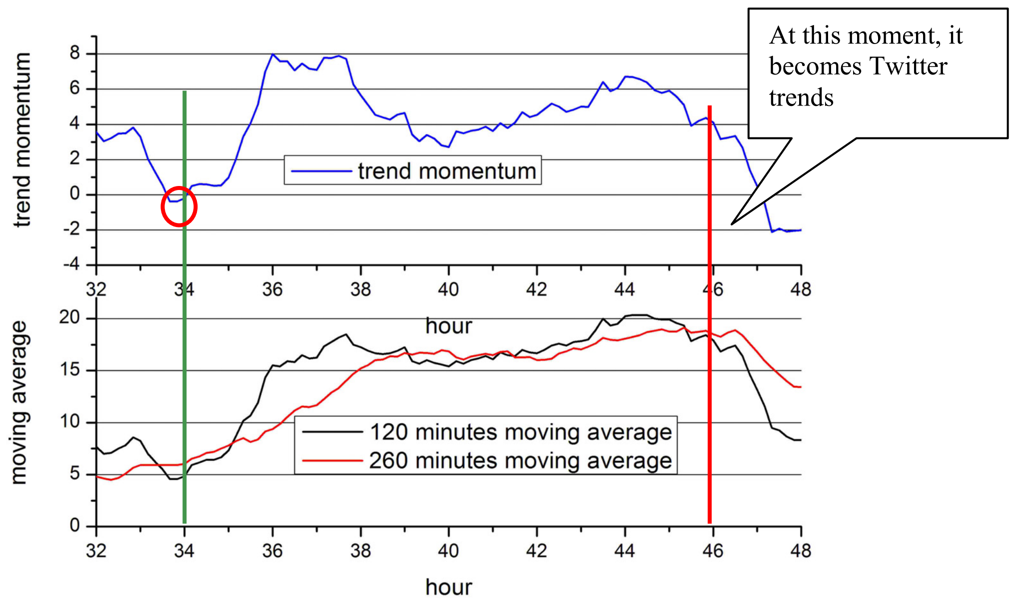

Fig.2: Trend momentum of a Twitter trends(“ipad”)

The two examples above show how to use trend momentum to predict the trends of a topic or a news event on Twitter. Then, we present a quantitative evaluation.

Normally, if a news event or a topic get a lot attention, it will have at least one concentrated discussion period. In this period, the density of tweets talked about the news event or topic is the highest. So, for all the news in the news dataset, we compute the tweets density in the life span of every news, and define the highest point as the main peak. Then, we use our trend momentum method(TM) to predict the main peak. If at a certain time point, our method gives an indication that the main peak is arriving, we call this time point the **prediction point**. The baseline method is a fixed threshold method(FT), in which a fixed threshold θ is given. If the tweets density exceeds the threshold θ, it considers the main peak is arriving. Table 1 shows the results.

Table. 1: Trend Momentum Method vs. Fixed Threshold Model

methods	**Time span from prediction point to the main peak(minutes)**	**omission rate**
TM	24.35	**10.98%**
FT $\theta=10$	65.80	15.04%
FT $\theta=25$	24.49	42.68%
FT $\theta=40$	14.63	77.24%

Our method has the lowest omission rate. Sometimes, the prediction point appears behind the main peak, in this situation, a missing prediction happens. The rate of all missing predictions is called **omission rate**.

In the Twitter trends dataset, we also examine the rate that how many Twitter trend topics have the feature that the trend momentum changes from negetive to positive before it becomes the Twitter trend topics. Results are shown in Table 2.

Table. 2: Time range vs. rate that has the feature.

Time range	0 ~ 4 hours	4 ~ 8 hours	8 ~ 12 hours	12 ~ 16 hours
Rate	10.22%	24.34%	29.69%	10.21%

Before the topic become a Twitter trend topic, about 74.46% trends topics, their trend momentum experienced the process from negetive to positive in the last 16 hours.

5. Summaries

In this paper, we try to predict the topic trends on Twitter based on an indicator (MACD), which is a commonly used in analyze trends of price of stocks. We did a little modification to the original MACD. Two different sized of time-window moving average are given, and trend momentum is difined based on them. Experiments show that the proposed method is simple and effective. In the future, we will integrate more characteristics of Twitter into our method. For example differernt user post tweets at different time, the influences are not the same. Simple frequency count may not be very reasonalbe.

6. References

[1] Huberman, B.A.; Romero, D.M.; Wu, F. Social networks that matter: Twitter under the microscope. First Monday [J], 14(1), (2009)

[2] Weng, J.; Lim, E.P.; Jiang, J.; He, Q.: TwitterRank: finding topic-sensitive influential twitterers. Proc. of the third WSDM, 261-270(2010).

[3] Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media? Proc. of the 19th international conference on World wide web, 591-600(2010).

[4] Kontostathis, A., Galitsky, L., Pottenger, W.M., Roy, S., Phelps, D.J.: Survey of Text Mining: Clustering, Classification, and Retrieval.Survey of Text Mining, 185-224(2003)

[5] Swan, R., Jensen, D.: Timemines: Constructing timelines with statistical models of word usage. KDD-2000 Workshop on Text Mining, 73-80(2000)

[6] Lent, B., Agrawal, R., Srikant, R.: Discovering trends in text databases. Proc. of the 3rd Int'l Conference on Knowledge Discovery in Databases and Data Mining, 227-230(1997)

[7] Pottenger, W.M., Kim, Y.B., Meling, D.D.: HDDI™: Hierarchical Distributed Dynamic Indexing. Journal of Data mining for scientific and engineering applications, (2001)

[8] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. The Journal of Machine Learning Research, volume(3)：993-1022(2003)

[9] Blei, D. M., Lafferty, J. D. Dynamic topic models. Proc. of the 23rd ICML, 113-120, (2006)

[10] Wang, C., Blei, D.: Continuous time dynamic topic models. Proc. of UAI, (2008).

[11] Wang, X., McCallum, A.: Topics over time: a non-Markov continuous-time model of topical trends. Proc. of the 12th ACM SIGKDD. 424-433(2006).

[12] Kawamae, N.: Trend analysis model: Trend consists of temporal words, topics, and timestamps. Proc. of fourth WSDM. 317-326(2011)