

# STH-Bass: A Spatial-Temporal Heterogeneous Bass Model to Predict Single-Tweet Popularity

Yan Yan<sup>1</sup>, Zhaowei Tan<sup>1</sup>, Xiaofeng Gao<sup>1(✉)</sup>, Shaojie Tang<sup>2</sup>, and Guihai Chen<sup>1</sup>

<sup>1</sup> Shanghai Key Laboratory of Data Science, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China  
gao-xf@cs.sjtu.edu.cn

<sup>2</sup> Department of Information Systems,  
The University of Texas at Dallas, Richardson, USA

**Abstract.** Prediction in social networks attracts more and more attentions since social networks have become an important part of people's lives. Although a few topic or event prediction models have been proposed in the past few years, researches that focus on the single tweet prediction just emerge recently. In this paper, we propose STH-Bass, a Spatial and Temporal Heterogeneous Bass model derived from economic field, to predict the popularity of a single tweet. Leveraging only the first day's information after a tweet is posted, STH-Bass can not only predict the trend of a tweet with favorite count and retweet count, but also classify whether the tweet will be popular in the future. We perform extensive experiments to evaluate the efficiency and accuracy of STH-Bass based on real-world Twitter data. The evaluation results show that STH-Bass obtains much less APE than the baselines when predicting the trend of a single tweet, and an average of 24% higher *precision* when classifying the tweets popularity.

**Keywords:** The Bass model · Predicting popularity · Social network

## 1 Introduction

Twitter, which is centered by users and communications, is one of the best-known social networks all over the world. What differs Twitter from other popular social networks is the asymmetric friend relationship, where a user is able to see tweets posted by his followings, while his tweets can be seen by his followers. Because of this characteristic, Twitter is more suitable for users who want to know celebrities' lives compared to symmetric social network such as Facebook,

---

This work has been supported in part by the China 973 project (2012CB316200), National Natural Science Foundation of China (Grant number 61202024, 61472252, 61133006, 61272443, 61473109), the Opening Project of Key Lab of Information Network Security of Ministry of Public Security (The Third Research Institute of Ministry of Public Security) Grant number C15602, and the Opening Project of Baidu (Grant number 181515P005267).

WeChat, etc. In Twitter, users tweet not only to express their emotions, but also to share their lives with followers. In other words, users tweet for what are worth tweeting, usually things that can lead to people's interest, attention or discussion. Once users have posted these tweets, they will have an expectation that these tweets will become popular, gaining a large number of retweets and favorites.

In recent years, prediction in social networks attracts more and more attention from both academia and industry since social networks have become parts of people's lives. Foresight is always users' favorite, and prediction satisfies the demand. Plenty of contents in social networks are worth predicting, such as user's personality [1], top10 news [2], popular stories [3], and controversial events [4]. Even a film's box office which seems to have nothing to do with social networks can be predicted through contents posted by users [5]. Most of these predictions use methods of machine learning such as Support Vector Machine, Naive Bayesian, Neural Network, etc., plus statistical model to do classification or regression.

Although researchers have done plenty of works on prediction, there are few works about predicting the popularity of a single tweet. Single tweets are components of topics and events. For users, from the perspective of psychology, posting popular tweets will bring satisfaction, and drive them to post more tweets. For Twitter and other official accounts, if they are capable of predicting which single tweet will become popular in advance, then they get an opportunity to lead the trend, and even come up with new hot topics or hashtags generating from that popular tweet. For third-party companies other than Twitter, predicting whether a single tweet will become popular and learning what makes a single tweet popular are useful as well, in that they are likely to create a popular tweet artificially, helping propagate their products and so on. Last but not least, abnormal popular tweets can set alarm for disaster, criminal, or catastrophe.

However, adopting existing topic or event prediction models cannot obtain satisfactory results. Topics or events usually consist of multiple tweets, but the popularity of one topic or event cannot well represent a single tweet's popularity. Actually, predicting the popularity of a single tweet requires different models. Compared with topics or events, which have more information generated from all tweets that they are made up of for prediction, predicting a single tweet's popularity can only use its own textual information and user's information with time line. Besides, the lifespan of a single tweet is much less than a topic or event. Most of tweets are out of sight after posted for one week, which leaves us less time to compare the prediction with real trend, let alone correct the prediction. In addition, traditional machine learning methods with large training set does not work for single tweets prediction, since each tweet is a unique new target to predict. Information of those historical popular tweets helps little, in that we cannot learn a new model with unique parameters for each new tweet.

When it comes to the prediction of online contents popularity, most of related works focus on predicting the popularity of topics, events, or news. To the best of our knowledge, few researches discuss the prediction of a single tweet [6,7].

Zhang et al. [6] only predicts whether a tweet will be retweeted, which is a 2-class classification problem, leaving aside the prediction of the tweet’s trend, which is a regression problem. The accuracy of regression is harder to ensure, in that the trend of a tweet changes greatly everyday after it is posted, and an efficient regression model have to predict as much as possible random changes of the trend. On the other hand, the model in [7] is impractical, in that it requires too many features that are difficult to obtain for predicting the trend of a tweet, such as the post time of all retweets relative to the original tweet, and the number of followers of each retweet user. Currently, there exists no reported study of predicting contents in social network using models modified from economics fields.

In this paper, we design STH-Bass, a *Spatial and Temporal Heterogeneous Bass Model* to predict the future of a single tweet. The Bass model [8] is one of the most widely applied models in management science, and the “Bass Model” paper is one of the Top 10 Most Influential Papers published in the 50-year history of *Management Science* [9]. The model is originally used in economic field to model the sales of a newly put-on-market product. To make up for its deficiency, we work more on spatial and temporal heterogeneity. We would like to predict the trend of a single tweet. In specific, according to the information observed after a single tweet posted for one day, we can use STH-Bass model to predict its favorite count and retweet count later of its whole life cycle. We further predict whether a single tweet will be popular using the results of trend prediction. Our model does not need large training set like traditional machine learning methods. In addition, there is no need of the topology of the social network such as following and follower relationships. Our experiments using real-world Twitter data validate the efficiency and accuracy of the trend prediction, with less absolute percent error and better classification detection.

We summarize our contributions as follow:

- We predict the trend of a single tweet after it is posted for a day, and whether the tweet will be hot in the end. Our model only needs attributes about the tweet to predict and its poster. Large training set and the topology of the network are not needed.
- We are the first to use the Bass model which is a famous statistical model from management science, in social network content prediction. In addition, we combine spatial and temporal heterogeneity into the Bass model, proposing a practical model which can be used in predicting a single tweet’s popularity.
- Our heterogeneous Bass model is not only suitable for predicting a single tweet’s popularity in Twitter, but also suitable for other social networks which have an asymmetric following-follower relationship, such as Weibo and Digg.
- We use real-world Twitter data to examine the efficiency of STH-Bass model, and the simulation results well exhibit the efficiency and accuracy of the trend prediction, with less absolute percent error and better classification detection.

The rest of the paper is organized as follows. In Sect. 2, we introduce the related work in the field of social network prediction. In Sect. 3, we give some preliminaries about our problem statement and data analysis, which is the basis

for parameters of our proposed model in Sect. 4. In Sect. 5, we illustrate our experiments on Twitter data set with discussions. Finally, we conclude our work in Sect. 6.

## 2 Related Work

As soon as social network was brought to people’s eyesight, researchers began to explore it from different perspectives. They first analyzed social network, discovered and found popular things [10,11] and influenced users [12]. Then, they classified things [13,14], and finally they recommended things to users [15], and predicted the future [5].

Prediction in social network goes to different directions as well. Social network is consist of users and contents posted by users. Therefore, one direction of prediction goes to people [1,16]. Another direction goes to contents, which can be classified specifically into predicting events [4,17], topics [18], news [2,19,20] or activities [21]. Most of contents prediction were at collective level. They paid attention to predicting things that a group of people took part in, not things that created by an individual.

Collective prediction such as the prediction of topic and event has grown relatively mature recently. These predictions are mostly classification problems, predicting whether a topic or event will be popular in future. Deng et al. [18] used a probability method of Bayesian combined with generative learning method. Furthermore, they divided time into continuous time intervals and predicted in which interval a topic will be popular. Zhang et al. [17] firstly detected events from burst word clustering, then used linear spread model to predict event popularity. All these topic or event models need extra tools to generate topics or events at first step, then use self-designed model with machine learning methods to reach their goals.

In each social network, there are special contents to predict, such as images on Flickr [22], stories on Digg [3], or hashtags on Twitter [23–25]. Kong et al. [25] is an representative example of predicting bursts and popularity of hashtags. It is smart to classify the life cycle of a hashtag into different statuses. Hashtag is a great innovation to help annotate topics in tweets. It helps predict popularity of online contents. Chang [26] brought in *Diffusion of Innovation* (DOI) theory, which was first applied in the researches in social network. It regarded hashtag as a kind of innovation, and showed that the Bass model [8] is feasible to hashtag prediction. Cui et al. [27] and Yang et al. [28] mentioned the DOI theory to discover popular tweets, but did not apply any model of DOI to predict the popularity of online contents.

For works related to ours most, predicting the popularity of a single story on Digg [3] is the first relevant work we know that paid attention to contents of individual level. It took into account individual behaviors to generate social dynamic model to obtain good performance on Digg. When it comes to Twitter, retweet behavior prediction [6] used a hierarchical Dirichlet process to predict whether a tweet would gain retweets. It was a classification problem which cannot

provide a quantitative outcome of how many retweets a tweet will finally gain. Hong et al. [29] predicted popular messages in Twitter using methods of binary classification and multi-class classification. The newly published [7] is one of the most relevant work that predicts the popularity of a single tweet. However, besides the shortages mentioned in Sect. 1, their model had to obtain the specific information of each user who retweeted, which is hard to track when generating the data set.

### 3 Preliminary

#### 3.1 Problem Statement

After a tweet  $w$  is posted by user  $u$ , the followers of  $u$  can favorite the tweet, which increases its favorite count  $f(t)$ , or retweet, which increases its retweet count  $r(t)$ . Once a follower retweets  $u$ , the tweet can be seen by the follower's followers. Therefore,  $f(t)$  and  $r(t)$  keep increasing until everyone who have seen the tweet stopped favoriting it or retweeting it. The final count of favorite count is denoted as  $fc$ , and the final count of retweet count is denoted as  $rc$ . In order to state our task clearly, we give some definitions first.

**Definition 1** (*Stable Time*). The time  $T$  that satisfies  $f(T) \geq \nu fc$  and  $r(T) \geq \nu rc$  is the **stable time** of a tweet. We set  $\nu = 0.95$  in our experiments.

We set parameter  $\nu$  here because not all tweets can reach their final counts given a fixed period of time. Some of the most popular tweets will continue to be retweeted and favorited for over a month. Therefore, we use  $\nu$  to ensure that the favorite counts and retweet counts of tweets in our data set can finally reach a stable state. This will be explained later in Sect. 3.2.

**Definition 2** (*Popularity Count*). Both favorite count and retweet count represent a tweet's popularity. We define the **popularity count**  $Y(t)$  as:  $Y(t) = \mu f(t) + (1 - \mu)r(t)$ , where  $0 < t \leq T$ , and  $\mu$  is a coefficient to maintain the balance of favorite count and retweet count.

Therefore, predicting the trend of tweet  $w$  turns into predicting  $Y(t)$ , given  $0 < t \leq T$ . At the same time,  $Y(T)$  is the approximate final popularity count of a tweet. If  $Y(T) > \gamma$ , where  $\gamma$  is a threshold, then we regard the tweet as **popular**.

#### 3.2 Data Analysis

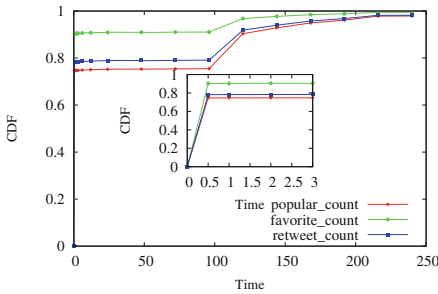
Our data set is collected through Twitter API<sup>1</sup>. In order to efficiently crawl as many tweets as possible, and select the features we need, we first randomly crawl a set of tweets (time from 16th July, 2015 to 23rd July, 2015; with a quantity of 102,756 tweets), and do some data analysis (including Tables 1, 2, and Figs. 1, 2, 3, 4, 5 and 6). We find several characteristics of single tweet:

<sup>1</sup> <https://dev.twitter.com/>.

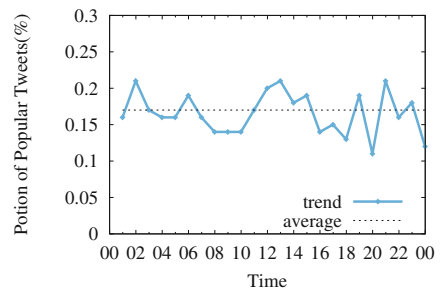
**Table 1.** Tweets’ popular count distribution

| Popular count | 0       | 1–9     | 10–99  | 100–999 | 1000+  |
|---------------|---------|---------|--------|---------|--------|
| Potion (%)    | 74.60 % | 24.06 % | 1.20 % | 0.12 %  | 0.02 % |

- From Table 1 we can see that 74.6 % of original tweets receive no favorite count or retweet count. This is because the users do not have enough followers, or the contents of these tweets are not the type which can be favorited or retweeted, e.g., something sad or meaningless. The rest 25 % non-zero tweets mostly gain 1–10 popular counts. Only 1.34 % tweets gain over 10 popular counts.
- The popular count of a single tweet increases rapidly during the first several hours after it is posted. As we can see from Fig. 1, 75 % tweets’ favorite counts and retweet counts remain unchanged since they were posted. Referring to the characteristic above we know that nearly all of these tweets have no favorite count or retweet count. It is also possible that 0.4 % of these tweets gain several popular counts, and soon reach their stationary. There is a huge increase at 100 h (4–5 days after posted). At this time, 90 % tweets reach their final counts, and nearly 98 % tweets become stable after 240 h (10 days after posted). The rest 1 % tweets are with high probability to become popular tweets, which still receive favorites and retweets, although at a very slow rate.
- We can also see from Fig. 1 that favorite count reaches its stationary much faster than retweet count, which makes sense. It takes little time to click your mouse to favorite a tweet you see. But retweet action includes something like chatting. For example, user *A* retweets user *B*’s tweet, and *B* retweets back with something *A* is interested in. Then *A* will retweet again, and *B* replies back. Finally, the retweet count will keep increasing until their conversation ends.



**Fig. 1.** The CDF of tweets’ stable time (Color figure online)



**Fig. 2.** When are popular tweets created (Color figure online)

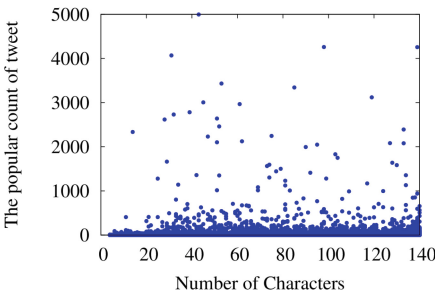
Based on the above characteristics, we use Twitter stream to stochastically crawl tweets which has just been posted on Twitter, and record their favorite

counts and retweet counts at  $t = 0, 0.5\text{h}, 1\text{h}, 2\text{h}, 3\text{h}, 6\text{h}$ . For tweets with  $Y(t = 6\text{h}) > 10$ , we track down their favorite counts and retweet counts at the following  $t = 12\text{h}, 1\text{d}, 2\text{d}, \dots, 10\text{d}$ . For the rest tweets, we only record their popularity count at  $t = 10\text{d}$  as the final popularity count. Finally, our second data set is from 24th July, 2015 to 24th August, 2015, with a quantity of 569,149 tweets. We just crawl information about original tweets (instead of retweets), since the original tweets must be popular if some of its retweets are popular. Besides, we crawl the features about tweets' posters, including their number of followings, followers, tweets they have posted and the favorites they have obtained, etc.

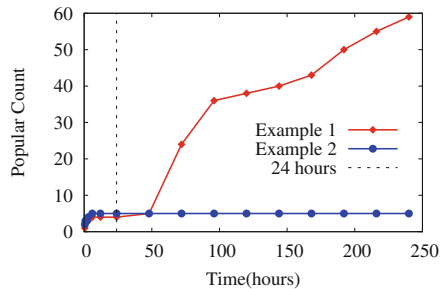
According to our second data set, we dig deeper about the correlation between features of tweets and popularity of tweets. Apart from semantic and emotion features which need extra tools to generate, we study the effect of creating time and length of a tweet, which is simple but worthwhile.

Figure 2 shows the creating time of tweets and the number of popular tweets during 24 hours in a day. The truth is, at what time a tweet is posted does have an impact on its popularity. The dash line gives the average proportion of popular tweets per hour. It is obvious that 1 a.m. – 2 a.m., 11 a.m. – 1 p.m. and 8 p.m. – 9 p.m. are best time periods during which there is a large proportion of popular tweets. However, 7 a.m. – 10 a.m. in the morning is not a good time for posting a tweet, so as 4 p.m. – 6 p.m. in the afternoon, because people are busy going to work and going back home. This is a general analysis, because the creating time of a tweet is not strongly related to its popularity: a popular tweet is born to be popular according to its inner property. Nevertheless, without the right posting time, a popular tweet may submerge under large quantities of tweet streams. In a word, choosing an appropriate time to post helps accelerate the speed of popularity.

Figure 3 shows the relationship between the length of a tweet and its final popular count. Intuitively the length of a tweet is a useful feature for prediction, for that popular tweets always come up with clearly declarative or logical



**Fig. 3.** Number of characters and popular counts of tweets



**Fig. 4.** An example of two tweets which had the same trend at the beginning, but ended differently (Color figure online)

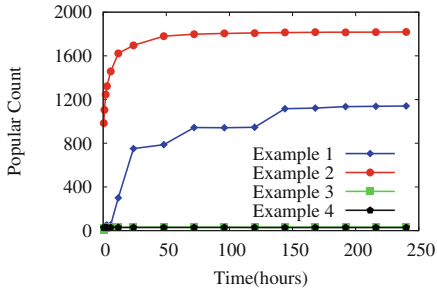
**Table 2.** Popular tweet examples with different length

| Content   | Length | Favorite count | Retweet count |
|---|--------|----------------|---------------|
| I can not deal with people commenting on my shit. I'm a human being, stop leaving comments on my pictures like I don't have eyes to read it | 138    | 814            | 3442          |
| I can remember song lyrics from 2006 but not whatever maths formula we were learning yesterday  | 94     | 2112           | 3019          |
| Her smile puts stars in my eyes   | 31     | 857            | 3212          |
| GOODMORNING   | 11     | 73             | 334           |

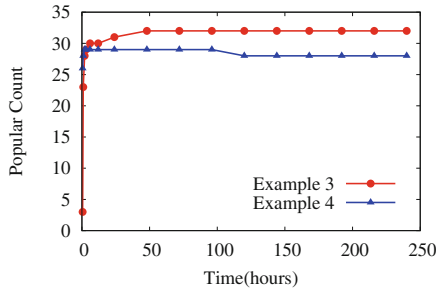
statements, which are all in need of quite a few words. However, it turns out that the contribution of the length to a popular tweet is not significant. Table 2 gives some examples of popular tweets with different length. However, whether the length can be regarded as one of predicting features remains to be examined in experiments later.

Tweets in Fig. 4 have similar trend in the observation period during the first 24 h after posted. However, one gets popular and another stays unpopular.

We also classify the trends of tweets mainly into four (Fig. 5): (1) gets high popular count immediately after posted, and stays popular (Example 2); (2) slowly increases and finally gets high popular count (Example 1); increases at a high rate at beginning but are not popular in the end (Example 3); slowly increases and stays slow in the end (Example 4). The threshold of being popular is 100, meaning that a tweet with final popular count greater than 100 will be classified as popular. Since Example 3 and Example 4 get small number of popular count, their trends have been zoomed in in Fig. 6.



**Fig. 5.** Different trends of tweets (Color figure online)



**Fig. 6.** A clearer illustration of Examples 3 and 4 in Fig. 5 (Color figure online)



## 4 The Heterogeneous Bass Model

The Bass model [8] is one of the most widely applied models in management science. The model was proposed to predict the sales of a new product when it is launched on the market. The model is different from traditional machine learning method in that it does not need large numbers of training set. Given the first several days or months of sales of a new product, we can easily predict the performance of the product later via only two parameters.

The standard formulation of a diffusion process is:

$$h(t) = (p + qH(t))(1 - H(t)) \quad (1)$$

Where  $p$  is the coefficient of innovators (or the impact of factors outside the population),  $q$  is the coefficient of imitators (or the impact of contacts within the population).  $h(t)$  is the hazard rate of adoption, which is also known as the likelihood of purchase at time  $t$ .  $1 - H(t)$  is the probability that one has not yet adopted at time  $t$ . Obviously,  $H(t) = \int_0^t h(x)dx$ . Here  $t$  is represented in *hours*.

Assuming the size of potential buyers is fixed as  $m$ , the number of purchase at  $t$  is:

$$S(t) = mh(t) = pm + (q - p)Y(t) - \frac{q}{m}[Y(t)]^2 \quad (2)$$

Where  $Y(t) = \int_0^t S(x)dx$  is the accumulative number of sales. Then Eq. (2) is the formulation of standard Bass model.

Although the Bass model is an excellent model in economic fields, it has many drawbacks as well. It only has two parameters, which gives little chance to add feature of social network into the model. In addition, the standard Bass model assumes spatial and temporal homogeneity, leading to no distinction of individuals. Therefore, we can combine features of Twitter with the original model, and relax it to individual-level heterogeneity. Due to the limitation of Twitter features, we focus on incorporating spatial heterogeneity, which allows everyone to have different possibilities to favorite or retweet a tweet. According to [30], there are 3 kinds of terms to reflect spatial heterogeneity: the intrinsic probability of adoption, the susceptibility to intra population linkages and the infectiousness of adopters.

Endowing the features we can get from Twitter, we derive the model as below:

$$S(t) = pm + (q - p)Y(t) - \frac{q}{m}[Y(t)]^2 + \alpha\mathbf{x} + \beta\mathbf{y} \quad (3)$$

Here  $\mathbf{x}$  is a vector of variables representing user features, such as the number of followings and followers of a user, the number of tweets a user posted since the creation of his account, the number of favorites the users collected.  $\mathbf{y}$  is a vector of single tweet features including the creating time of a tweet, the number of URLs appearing in a tweet, and the number of characters in the text. Although the number of characters is not the main component contributing to popular

count based on our data analysis in Sect. 3.2, we still take it into consideration here. It should be noted that not all of these features contribute to the final result of our experiments. Those nonsignificant features will be ruled out by **PCA** (Principal Component Analysis).

$$\mathbf{x} = \begin{bmatrix} \log \text{ num Of Followings} \\ \log \text{ num Of Followers} \\ \log \text{ num Of Tweets} \\ \log \text{ num Of Favorites} \end{bmatrix} \quad (4)$$

$$\mathbf{y} = \begin{bmatrix} \log \text{ num Of Creating Time} \\ \log \text{ num Of URLs} \\ \log \text{ num Of Characters} \end{bmatrix} \quad (5)$$

In order to find  $Y(t)$  we must solve the non-linear differential equation:

$$\frac{dY}{dt} = pm + (q - p)Y(t) - \frac{q}{m}[Y(t)]^2 + \alpha\mathbf{x} + \beta\mathbf{y} \quad (6)$$

For simplicity, let  $V = pm + \alpha\mathbf{x} + \beta\mathbf{y}$ . Then we have:

$$\frac{mdY}{q[Y(t)]^2 + m(p - q)Y(t) - mV} = -dt \quad (7)$$

Factoring the denominator on the left of Eq. (7), we have:

$$\frac{mdY}{(Y - y_1)(Y - y_2)} = -dt \quad (8)$$

with  $y_1 = \frac{m(q-p)+\sqrt{\Delta}}{2q}$ ,  $y_2 = \frac{m(q-p)-\sqrt{\Delta}}{2q}$ , and  $\Delta = m^2(p - q)^2 + 4mqV$ .

Change Eq. (8) into:

$$\left(\frac{1}{Y - y_1} - \frac{1}{Y - y_2}\right) \frac{mdY}{y_1 - y_2} = -dt \quad (9)$$

Then we can do integration on both sides of the equation:

$$\int_0^T \left(\frac{1}{Y - y_1} - \frac{1}{Y - y_2}\right) \frac{mdY}{y_1 - y_2} = \int_0^T -dt \quad (10)$$

The solution is:

$$Y(t) = \frac{y_2 e^{-\frac{\sqrt{\Delta}}{mq}t+C} + y_1}{1 + e^{-\frac{\sqrt{\Delta}}{mq}t+C}} \quad (11)$$

Because  $Y(0) = 0$ , constant  $C$  generated by the integration can be solved:

$$C = \ln\left(-\frac{y_1}{y_2}\right) \quad (12)$$

And:

$$Y(t) = \frac{y_2 e^{-\frac{\sqrt{\Delta}}{mq} t + \ln(-\frac{y_1}{y_2})} + y_1}{1 + e^{-\frac{\sqrt{\Delta}}{mq} t + \ln(-\frac{y_1}{y_2})}} \quad (13)$$

Hence we get the spatial and temporal heterogeneous Bass model. Using **Least Square Method**, which is one of the most recommending mathematical methods to fit Eq. (13). Then we can predict the popular count of a tweet at any time  $t$ , where  $1d < t \leq T$ .

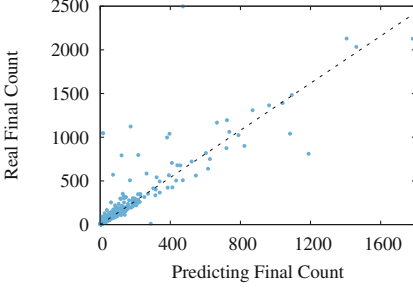
## 5 Experiments

Our experiments are divided into two parts. The first part is the trend predicting, and the second part is the popularity predicting. Due to the fact that the favorite count and retweet count of a tweet varies a lot, some tweets gain large amount of favorite count but much fewer retweet count, while some tweets are on the contrary. Therefore, we set  $\mu = 0.5$  to treat the number of favorite count and retweet count equally. In addition, we set  $\nu = 10d$  and  $\gamma = 100$  according to our data analysis. Furthermore, we treat follower count and friend count as invariant during the period of our experiments. As a matter of fact, these features do not change much for a mature user who has already set up his relationship network.

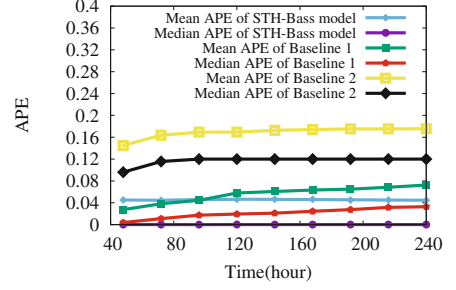
### 5.1 Predicting Trends

Different from traditional machine learning methods, which usually have a training set to train parameters and a test set to evaluate trained models, our heterogeneous Bass model divides each tweet into training part and predicting part. The training part includes the first 7 sample points at  $t = 0.5 \text{ h}, 1 \text{ h}, 2 \text{ h}, 3 \text{ h}, 6 \text{ h}, 12 \text{ h}, 24 \text{ h}$  after a tweet posted. We use **Least Square Method** on these sample points of a single tweet, and the parameters in Eq. (13) are solved. Each tweet has a set of unique parameters to predict its trend in the nearly future. To test the efficiency of our model, we sample 9 time points at  $t = 48 \text{ h}(2 \text{ d}), 72 \text{ h}(3 \text{ d}), \dots, 240 \text{ h}(10 \text{ d})$ . According to our data analysis in Sect. 3, most of popular counts of tweets tend to be stable after posted for 240 h. Therefore, our sample points can well test the whole life cycle of a tweet.

We give the comparison of the predicting final popular count and the real final popular count, as illustrated in Fig. 7. The dash line shows the correlation between the predicted value and the real value, with slope 1.3. Besides several outliers which are far away from the dash line, most of our predicting values are close to real values. When the final popular count is less than 200, our model performs best. As the real popular count becomes larger, our prediction emerges its conservative. There is a high probability that our predicting value is smaller than the real ones. However, it does not affect the precision of predicting popularity.



**Fig. 7.** The predicting final count VS the real final count



**Fig. 8.** The APE of our model and baselines (Color figure online)

Our STH-Bass model only need each tweet’s text information (favorite count, retweet count, created time, length of the tweet, etc.) and it’s poster’s information (friends count, followers count, etc.), which can be easily obtained from Twitter API. For baselines, we initially intend to use the SEISMIC Model [7] as one of our baselines, since it is the most relevant study compared to ours. However, the SEISMIC Model needs extra information (such as time and follower counts) about user  $i$  who contributed to the  $ith$  retweet of a certain tweet. It is difficult to crawl these extra features from Twitter. Without these extra information, the SEISMIC model will lose its accuracy. Hence, we give up on the SEISMIC model, and propose another two baselines, which need a few features about each tweet and its author like our STH-Bass model:

- **Naive Model (Baseline1).** Since there are not many literature about predicting the trend of a single tweet, we come up with a naive model based on the characteristic (analyzed in Sect. 3.2) of our data set. We only use the popular count of  $t = 24$ h to predict the trend of next 9 days. The Naive Model performs rather competitive because 75 % of tweets in our data set reach their final count after posted for one day.
- **Log Linear Regression Model (Baseline2).** Szabo and Huberman [31] analyzed there are strong correlations between early and later times of the logarithmical transformed popularity:

$$\ln Y(t_r) = \ln r(t_i, t_r) + \ln N(t_i) + \xi(t_i, t_r) \quad (14)$$

Therefore, Log Linear Regression Model can also predict well.

We calculate the **Absolute Percentage Error (APE)** for our evaluation. For each tweet  $w$  at time  $t$ , the formulation of APE is as follows:

$$APE(t) = \frac{|Y(t) - popularCount(t)|}{popularCount(t)} \quad (15)$$

Figure 8 shows the Mean APE and Median APE of our model and two baselines. We test 9 time points from 48 h (2 days after posted) to 240 h (10 days after

posted). Both Mean and Median APE of our STH-Bass model change little during the whole period. The Mean APE is around 4 %, and the Median APE is around 1 %. Baseline 2 is far less accurate than STH-Bass in terms of Mean APE and Media APE. Baseline 1 is competitive at first several time points with STH-Bass, but tends to be less accurate later, indicating that there are still plenty of increases of popular counts from 48 h to 240 h.

**Table 3.** The performances of different methods

| Model     | Precision | Recall | F1-score |
|-----------|-----------|--------|----------|
| Our model | 0.997     | 0.796  | 0.886    |
| Baseline1 | 0.949     | 0.792  | 0.863    |
| Baseline2 | 0.757     | 0.950  | 0.843    |

## 5.2 Predicting Popularity

According to Table 1, tweets which finally gain 0–9 popular count has a proportion of over 98 %. It is easy for our model and two baselines to get accuracy rates of over 95 %, because tweets receiving 0–9 popular count do not change much in their trends, and most of them already have reached stable after posted for one day. Therefore, we focus on the rest 2 % tweets, whose trends change relatively a lot, and are hard to reach their final counts after posted for one day. Therefore, we set  $\gamma = 100$ , which indicates that a tweet with  $Y(T) > 100$  is popular.

Once we have predicted the final count of a tweet, we can immediately decide whether a tweet is popular. Table 3 shows the performance of predicting whether a tweet becomes popular by different methods. STH-Bass model gets the highest *Precision* and *F1-score*, while Baseline 2 gets the highest *Recall*. Comprehensively, STH-Bass model is 3 % better than Baseline 1 and 5 % better than Baseline 2 in terms of *F1-score*. In addition, STH-Bass model is 5 % better than Baseline 1 and 24 % better than Baseline 2 in terms of *Precision*, meaning that unpopular tweets are seldom classified as popular ones. As a matter of fact, classification is easier than regression. Although the advantage of STH-Bass is not as obvious as predicting trend in Sect. 5.1, STH-Bass ensures that those popular tweets which the model has classified, will be popular in the end with a high probability.

## 6 Conclusion

In this paper, we propose STH-Bass, a spatial and temporal heterogeneous Bass model to predict the popularity of a single tweet. STH-Bass uses the data features of a single tweet from the first day it has been posted, and can successfully predict the whether this tweet can be popular in the future. More specifically, STH-Bass

can well depict the trend of a single tweet during its life cycle. Our model can even distinguish the tweets which have similar beginning popular count in first 24h, but gain extremely different popular count in the end. We also use real-world Twitter data set to examine the performance of STH-Bass and compare the results with several baseline strategies. The simulation results validate the efficiency and accuracy of STH-Bass model with much less **APE** than baselines when predicting trend of a single tweet, and higher *Precision* and *F1-score* than one of our baselines when classifying the popularity.

## References

1. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. In: ACM CHI Conference on Human Computer Interaction, pp. 253–262 (2011)
2. Kong, L., Jiang, S., Yan, R., Xu, S., Zhang, Y.: Ranking news events by influence decay and information fusion for media and users. In: ACM International Conference on Information and Knowledge Management (CIKM), pp. 1849–1853 (2012)
3. Lerman, K., Hogg, T.: Using a model of social dynamics to predict popularity of news. In: International Conference on World Wide Web (WWW), pp. 621–630 (2010)
4. Popescu, A.-M., Pennacchiotti, M.: Detecting controversial events from twitter. In: ACM International Conference on Information and Knowledge Management (CIKM), pp. 1873–1876 (2010)
5. Asur, S., Huberman, B.: Predicting the future with social media. In: IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, pp. 492–499 (2010)
6. Zhang, Q., Gong, Y., Guo, Y., Huang, X.: Retweet behavior prediction using hierarchical dirichlet process. In: AAAI Conference on Artificial Intelligence (AAAI), pp. 403–409 (2015)
7. Zhao, Q., Erdogdu, M.A., He, H.Y., Rajaraman, A., Leskovec, J.: Seismic: a self-exciting point process model for predicting tweet popularity. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pp. 1513–1522 (2015)
8. Bass, F.M.: A new product growth for model consumer durables. *Manag. Sci.* **15**(5), 215–227 (1969)
9. Bass, F.M.: Comments on “a new product growth for model consumer durables the Bass model”. *Manag. Sci.* **50**(12), 1833–1840 (2004)
10. Lerman, K., Ghosh, R.: Information contagion: an empirical study of the spread of news on digg and twitter social networks. In: The International AAAI Conference on Web and Social Media (ICWSM), vol. 10, pp. 90–97 (2010)
11. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: ACM International Conference on World Wide Web (WWW), pp. 591–600 (2010)
12. Weng, J., Lim, E.-P., Jiang, J., He, Q.: Twiterrank: finding topicsensitive influential twitterers. In: ACM International Conference on Web Search and Data Mining (WSDM), pp. 261–270 (2010)
13. Yang, S.-H., Kolcz, A., Schlaikjer, A., Gupta, P.: Large-scale high-precision topic modeling on Twitter. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pp. 1907–1916 (2014)

14. Zhou, D., Chen, L., He, Y.: An unsupervised framework of exploring events on Twitter: filtering, extraction and categorization. In: AAAI Conference on Artificial Intelligence (AAAI), pp. 2468–2474 (2015)
15. de Macedo, A.Q., Marinho, L.B.: Event recommendation in eventbased social networks. In: Hypertext, Social Personalization Workshop, pp. 3130–3131 (2014)
16. Akram, H.A.A.L., Mahmood, A.: Predicting personality traits, gender and psychopath behavior of Twitter users. *Int. J. Technol. Diffus. (IJTD)* **5**(2), 1–14 (2014)
17. Zhang, X., Chen, X., Chen, Y., Wang, S., Li, Z., Xia, J.: Event detection and popularity prediction in microblogging. *Neurocomputing* **149**, 1469–1480 (2015)
18. Deng, Z.-H., Gong, X., Jiang, F., Tsang, I.W.: Effectively predicting whether and when a topic will become prevalent in a social network. In: AAAI Conference on Artificial Intelligence (AAAI), pp. 210–216 (2015)
19. Castillo, C., El-Haddad, M., Pfeffer, J., Stempeck, M.: Characterizing the life cycle of online news stories using social media reactions. In: ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW), pp. 211–223 (2014)
20. Bandari, R., Asur, S., Huberman, B.A.: The pulse of news in social media: forecasting popularity. In: International AAAI Conference on Web and Social Media (ICWSM), pp. 26–33 (2012)
21. Huang, S., Chen, M., Luo, B., Lee, D.: Predicting aggregate social activities using continuous-time stochastic process. In: ACM international Conference on Information and Knowledge Management (CIKM), pp. 982–991 (2012)
22. McParlane, P.J., Moshfeghi, Y., Jose, J.M.: Nobody comes here anymore, it's too crowded; predicting image popularity on flickr. In: ACM International Conference on Multimedia Retrieval (ICMR), pp. 385–391 (2014)
23. Kamath, K.Y., Caverlee, J.: Spatio-temporal meme prediction: learning what hashtags will be popular where. In: ACM International Conference on Information Knowledge Management (CIKM), pp. 1341–1350 (2013)
24. Zaman, T.R., Herbrich, R., Van Gael, J., Stern, D.: Predicting information spreading in Twitter. In: Workshop on Computational Social Science and the Wisdom of Crowds, Annual Conference on Neural Information Processing Systems (NIPS), pp. 599–601. Citeseer (2010)
25. Kong, S., Mei, Q., Feng, L., Ye, F., Zhao, Z.: Predicting bursts and popularity of hashtags in real-time. In: International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 927–930 (2014)
26. Chang, H.-C.: A new perspective on Twitter hashtag use: diffusion of innovation theory. *J. Am. Soc. Inf. Sci. Technol.* **47**(1), 1–4 (2010)
27. Cui, A., Zhang, M., Liu, Y., Ma, S., Zhang, K.: Discover breaking events with popular hashtags in Twitter. In: ACM International Conference on Information and Knowledge Management (CIKM), pp. 1794–1798 (2012)
28. Yang, L., Sun, T., Zhang, M., Mei, Q.: We know what@ you# tag: does the dual role affect hashtag adoption? In: ACM International Conference on World Wide Web (WWW), pp. 261–270 (2012)
29. Hong, L., Dan, O., Davison, B.D.: Predicting popular messages in Twitter. In: ACM International Conference Companion on World Wide Web (WWW), pp. 57–58 (2011)
30. Strang, D., Tuma, N.B.: Spatial and temporal heterogeneity in diffusion. *Am. J. Soc.* **99**(3), 614–639 (1993)
31. Szaba, G., Huberman, B.A.: Predicting the popularity of online content. *Commun. ACM* **53**(8), 80–88 (2010)