

Evolution and Influence Measurement Association Information Network

Wenchao Li
Shandong University of
Science and Technology,
Qingdao 266590, China

Jijuan Jiang
Shandong University of
Science and Technology,
Qingdao 266590, China

Cong Liu*
Shandong University of
Science and Technology,
Qingdao 266590, China

Abstract: This paper mainly using SIR model, based on least squares polynomial fitting GM (1,1) model, Information flow reasonable quantitative analysis and evaluation of the impact of information and the public interest and public opinion to make a qualitative analysis. First, the study of journalism and communication problems model. Using SIR model divided population into three categories: Communicators, unknown person and rem. Then contra the news spread found the SIR model. Through qualitative analysis get rationality of SIR model. Then get specific data unknown persons in accordance with the SIR model was built using the least squares polynomial fitting based on those data, the unknown polynomial fitting. If the data is in line with the model it can be considered as news data, otherwise, the data is not news. Besides, to validate the reliability of model and to forecast the trend of the information flows today. Getting a group of reliable news media data through the local statistical offices personnel is consistent with the model. Since there is no question as to the data to obtain information dissemination five times, so this article by setting the initial value of SIR model, Since there is no question as to get information dissemination five periods of data, so this article by setting the initial value of SIR model to obtain information dissemination five periods of data, and then use the GM (1,1) model predict unknown persons data for each day, obtained dissemination of information specific data of each day. Whereby the trend today of unknown people the number of information dissemination, through testing the GM (1,1) model, it showed that the use of GM (1,1) model to predict today's information communication situation is entirely possible.

Keywords: SIR model; least squares polynomial fitting; GM (1,1) model

1. DESCRIPTION

Information is spread quickly in today's tech-connected communications network; sometimes it is due to the inherent value of the information itself and other times it is due to the information finding, its way to influential or central network nodes that accelerate its spread through social media. While content has varied -- in the 1800s news was more about local events (e.g., weddings, storms, deaths) rather than viral videos of cats or social lives of entertainers -- the prevailing premise is that this cultural characteristic to share information (both serious and trivial) has always been there. However, the flow of information has never been as easy or wide-ranging as it is today, allowing news of various levels of importance to spread quickly across the globe in our tech connected world. By taking a historical perspective of flow of information relative to inherent value of information, the Institute of Communication Media (ICM) [1], [2] seeks to understand the evolution of the methodology, purpose, and functionality of society's networks. Specifically, your team as part of ICM's Information Analytics Division has been assigned to analyze the relationship between speed/flow of information vs inherent value of information based on consideration of 5 periods: in the 1870s, when newspapers were delivered by trains and stories were passed by telegraph; in the 1920s, when radios became a more common household item; in the 1970s, when televisions were in most homes; in the 1990s, when households began connecting to the early internet; in the 2010s, when we can carry a connection to the world on our phones. Your supervisor reminds you to be sure to report the assumptions you make and the data you use to build your models.

The problems to be solved are:

- 1) Develop one or more model(s) that allow(s) you to explore the flow of information and filter or find what qualifies as news.
- 2) Validate your model's reliability by using data from the past and the prediction capability of your model to predict the information communication situation for today and compare that with today's reality.

2. SYMBOL CONVENTIONS AND ASSUMPTIONS

2.1 Symbol Conventions

$s(t)$: This news did not know who (unknown persons) ratio in the population;

$i(t)$: People know this and to spread the news (communicators) the proportion in the population;

$r(t)$: People do not know but this news spread (out of those) the proportion in the population;

λ : Day transmission rate communicators;

μ : Spread spread their number;

σ : Day stop transmission rates.

2.2 Conditions Model assumptions

Due to constraints time and our limited ability level, in order to ensure the reasonableness of rationality of the model and the model of the flow of information, the following assumptions are made.

1) In order to simplify the problem, assume that people were divided into three categories: unknown, disseminators and were removed.

2) Takes into account the short time the news spread, so the news media during the investigation of the total number of regions N unconsidered a birth or death, do not consider migration, and in days of measurement.

3. ANALYSIS OF THE PROBLEM

For question 1, develop one or more models, allowing you to explore the flow of information, filtering, or can find content that can be the news. Due to the spread of infectious diseases is similar to flow of information, so this paper intends to model the flow of information use the SIR Epidemic Model, by setting the parameters of the first test to obtain a model of the flow of information, through qualitative analysis, to test model is reliable. Then depending on the model obtain the data, this paper using the least squares polynomial fitting to fit the flow of information polynomial model. If the data obtained are consistent with the model, it can be considered the given data as news data, otherwise, the data is not news.

For question 2, using past data verify the reliability of the model; forecast the situation of today's information communication, verify the ability of prediction, for comparison with today's reality. This paper by looking for other data get reliable information, then obtained the resulting data about the time of trend graph, and compared with in the model image of Question 1, if reliable data is consistent with the model, can be obtained the model with a certain reliability. By looking for information to give the title to the information flow in the five periods, and this paper intends to use GM (1,1) model based on specific data of five periods' information flows, obtained communication skills today, and compare with today's reality.

4. MODELS

4.1 Establishment of the News' Model

The problem to be solved is: Develop one or more models allowing you to explore the flow of information, filtering, or can find content that can be the news. Due to the spread of infectious diseases is similar to flow of information, so this paper intends to model the flow of information use the SIR Epidemic Model, by setting the parameters of the first test to obtain a model of the flow of information, through qualitative analysis, to test model is reliable. Then depending on the model obtain the data, this paper using the least squares polynomial fitting to fit the flow of information polynomial model. If the data obtained are consistent with the model, it can be considered the given data as news data, otherwise, the data is not news.

4.1.1 SIR model on the news

Summary of population represented by N, Model assumptions are as follows:

1) This article will be divided into three groups,

a. Person does not know this news (unknown persons), $s(t)$ represents an unknown proportion in the population;

b. People who know and spread the news (communicators), $i(t)$ indicates the proportion of communicators in the crowd;

c. People who know the news, but does not spread (out of person), $r(t)$ shows proportion of out of person in the population.

2) Day spread rate of communicators is λ , day stop transmission rate is μ . Their dissemination number of Disseminate is $\sigma = \lambda / \mu$.

3) Taking into account that the news spread in a short time, therefore during area the total number of news media in the study is the constant N, that is not considered either birth or death, nor migrating, and measured in days.

Under the above scenario, crowd flow chart is as follows:



Figure. 1 Crowd flow chart

The assumption is clearly:

$$\begin{aligned} s(t) + i(t) + r(t) &= 1 \\ N \frac{ds}{dt} &= -\lambda N s i \\ N \frac{di}{dt} &= \lambda N s i - \mu N i \\ N \frac{dr}{dt} &= \mu N i \end{aligned}$$

The proportion of unknown person and communicators in the initial time are $s_0(s_0 > 0)$ and $i_0(i_0 > 0)$ (here assume the initial value of the removed persons is $r_0 = 0$), So get SIR model initial value problem are as follows

$$\begin{cases} \frac{ds}{dt} = -\lambda s i, & s(0) = s_0 \\ \frac{di}{dt} = \lambda s i - \mu i, & i(0) = i_0 \\ \frac{dr}{dt} = \mu i, & r(0) = 0 \end{cases}$$

And then have $s + i + r = 1$ get $dr/dt = -di/dt - ds/dt$, thus the third formula equation becomes identity, thereby simplifying the model as follows

$$\begin{cases} \frac{ds}{dt} = -\lambda s i, & s(0) = s_0 \\ \frac{di}{dt} = \lambda s i - \mu i, & i(0) = i_0 \end{cases}$$

Above initial value problem can not get analytical solution, only be obtained by numerical solution to get numerical solution.

For example, take $\lambda = 1, \mu = 0.3, i(0) = 0.02, s(0) = 0.98$, then get numerical solution is obtained in the following table, $i(t), s(t)$ and i - s curve as shown below correspondingly.

Table 1. $i(t)$ and $s(t)$ numerical solution of SIR model

t	0	1	2	3	4	5	6	7	8
i(t)	0.02	0.039	0.0732	0.1285	0.2033	0.2795	0.3312	0.3444	0.3247
s(t)	0.98	0.9525	0.9019	0.8169	0.6927	0.5438	0.3995	0.2839	0.2027
t	9	10	15	20	25	30	35	40	45
i(t)	0.2863	0.2418	0.0787	0.0223	0.0061	0.0017	0.0005	0.0001	0
s(t)	0.1493	0.1145	0.0543	0.0434	0.0408	0.0401	0.0399	0.0399	0.0398

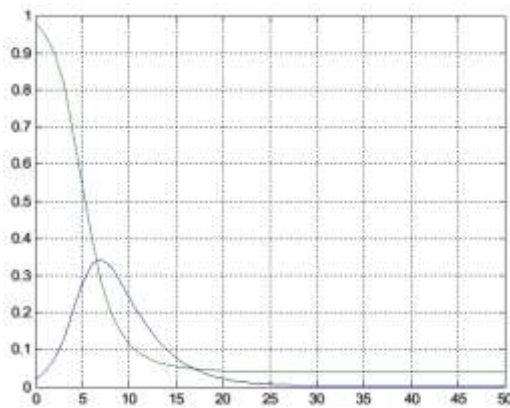


Figure. 2 $i(t)$ and $s(t)$ curve of SIR model

From the perspective of qualitative analysis, the results of this model has been broadly in line with the trend of the time in real life people of a news attention. In real life, when a news just happened, people's attention will increase sharply this news, then quickly reached a peak, and then the news will concern a sharp decline, then forgot this news slowly. The unknown number of this news will be a sharp reduction in a very short period of time, but there is always a small group of people learn the news because of this reason or another. From the perspective of qualitative analysis point of view, SIR model established in this paper according to the propagation of news.

4.1.2 Polynomial fitting based on least squares

As mentioned above, from the perspective of qualitative analysis, it can be obtained: the dissemination of information relating to the establishment of the above-SIR propagation model in line with real-life news. Therefore, this article will based on the data obtained by above example, to be a polynomial fitting based on least squares information data for the news data.

4.1.2.1 polynomial fitting model based on least squares

Curve fitting problem formulation is, given a set of (two-dimensional) data, namely n point on the plane (x_i, y_i) , $i = 1, 2, \dots, n$, x_i different from each other, seek a function (curve) $y = f(x)$, $f(x)$ under certain guidelines so that all the data points closest to that is the best fit curve.

Linear least squares is the most common method to solve curve fitting, the basic idea is

$$f(x) = a_1 r_1(x) + a_2 r_2(x) + \dots + a_m r_m(x)$$

Where $r_k(x)$ is a pre-selected set of linearly independent functions, a_k is undetermined coefficients ($k = 1, 2, \dots, m, m < n$).

Guideline is make square of distance δ_i that $y_i, i = 1, 2, \dots, n$, and $f(x_i)$ and with a minimum, known as a A least squares criterion.

1) Determine the coefficient a_k

$J(a_1, \dots, a_m) = \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n [f(x_i) - y_i]^2$ note for the sake of the a_1, \dots, a_m to a minimum. Simply use Extreme necessary conditions

$$\frac{\partial J}{\partial a_k} = 0 (k = 1, 2, \dots, m)$$

obtained the linear equations on a_1, \dots, a_m

$$\sum_{i=1}^n r_j(x_i) \left[\sum_{k=1}^m a_k r_k(x_i) - y_i \right] = 0, (j = 1, 2, \dots, m)$$

which is

$$\sum_{k=1}^m a_k \left[\sum_{i=1}^n r_j(x_i) r_k(x_i) \right] = \sum_{i=1}^n r_j(x_i) y_i, (j = 1, 2, \dots, m)$$

$$R = \begin{bmatrix} r_1(x_1) & \dots & r_m(x_1) \\ \vdots & \vdots & \vdots \\ r_1(x_n) & \dots & r_m(x_n) \end{bmatrix}_{n \times m},$$

$$A = [a_1, \dots, a_m]^T, Y = [y_1, \dots, y_n]^T$$

From above, the equation can be expressed as $R^T R A = R^T Y$, When $\{r_1(x), \dots, r_m(x)\}$ linearly independent, R full column rank, $R^T R$ reversible. So

$$A = (R^T R)^{-1} R^T Y$$

2) selection of $r_k(x)$ function

Faced with set of data $(x_i, y_i), i = 1, 2, \dots, n$, use linear least squares curve fitting, first and foremost, but also the crucial step is to appropriately select $r_1(x) \dots r_m(x)$. If through mechanism analysis, should be able to know what kind of function between y and x , it is easy to determine $r_1(x) \dots r_m(x)$. Unable to know the relationship between x and y . Typically can plot the data $(x_i, y_i), i = 1, 2, \dots, n$, visually determine what kind of curve fitting them should use. People often use a curve

(i) a straight line

$$y = a_1x + a_2$$

(ii) the polynomial

$$y = a_1x^m + \dots + a_mx + a_{m+1}$$

(generally, $m=2,3$, not too high)

(iii) hyperbola (only one)

$$y = \frac{a_1}{x} + a_2$$

(iv) an exponential curve

$$y = a_1e^{a_2x}$$

For the exponential curve, the need for variable substitution before fitting into the linear function a_1, a_2 .

Given a set of data, what kind of curve fitting the best, possible on the basis of intuitive judgments to elect several curves, then compare, behold Article *J* of the smallest least squares curve index.

3) Least Squares

In unconstrained optimization problems, some important special cases, for example, the objective function by the square of the number of functions and composition. Such functions can generally be written as:

$$F(x) = \sum_{i=1}^m f_i^2(x), x \in R^n$$

Where $x = (x_1, \dots, x_n)^T$ general assumptions $m \geq n$. We minimization problem of the kind of function:

$$\min F(x) = \sum_{i=1}^m f_i^2(x)$$

called least-squares optimization problem.

4.1.2.2 The least squares polynomial fitting of News
Based on data obtained using the Matlab program of above SIR model. The unknown proportion of the number of time-varying polynomial equation:

First Order: $s(t) = -0.0188t + 0.6284$

Second Order: $s(t) = 0.0013t^2 - 0.0716t + 0.8900$

Third Order: $s(t) = -0.0001t^3 + 0.0052t^2 - 0.1355t + 1.0733$

Fourth Order: $s(t) = 0.0000t^4 - 0.0002t^3 + 0.0080t^2 - 0.1593t + 1.1134$

As can be seen, in this fourth order polynomial coefficient of the highest degree term $s(t)$ is close to zero, so in this pair $s(t)$ only polynomial fitting to Fourth - order polynomial. And the

minimum square error form one to four order shown in Table 2.

Table 2. least squares polynomial fitting difference of $s(t)$

Order Number	First Order	Second Order	Third Order	Fourth Order
The minimum square error	0.9956	0.3253	0.0811	0.0673

As can be seen from Table 2, $s(t)$ least squares polynomial fitting difference from the Third Order Fourth Order are just 0.0138, and the fourth-order polynomial coefficients of the highest order term $s(t)$ is close to zero. Therefore, in this select third-order polynomial fitting of $s(t)$, so the fitting function of $s(t)$ is

$$s(t) = -0.0001t^3 + 0.0052t^2 - 0.1355t + 1.0733$$

The image of fitting function and the image of specific data are shown in Figure 3.

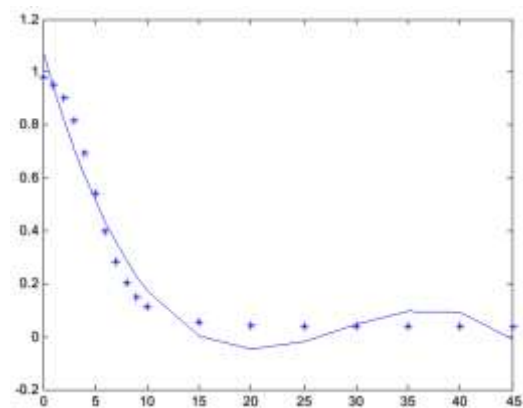


Figure. 3 The image of fitting function and the image of specific data

4.1.3 Data determined in the news

As mentioned above, from the perspective of qualitative analysis, it can be obtained, the dissemination of information relating to the establishment of the above-SIR propagation model in line with real-life news. And the data obtained in the SIR model of news spread performed function fitting. Therefore, the data was subjected to qualitative analysis(its scatter plot, unknown number with respect to time trends, etc.),and performed polynomial fitting to this set of data .If this set of data consistent with the established model can be considered that it is news, otherwise the message is not news.

4.2 Validation of model reliability and prediction information propagation today

4.2.1 Verify the reliability of the model

Linked to a relative working in local statistical offices by their families, Since people get a local news dissemination of data, and use this data to verify the reliability of the model herein, Although only we got a set of reliable data, data is little, but there are certain confirmatory. The proportion of the number of unknown data obtained as shown in Table 3.

Table 3. a local news unknown proportion of data

day	0	1	2	3	4	5	6	7	8	9
unknown people	0.98	0.95	0.88	0.76	0.58	0.39	0.25	0.15	0.10	0.07
day	10	11	12	13	14	15	16	17	18	19
unknown people	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01
day	20	21	22	23	24	25	26	27	28	29
unknown people	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
day	30	31	32	33	34	35	36	37	38	39
unknown people	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

A scatter plot local news by unknown proportion of the data shown in Figure 4.

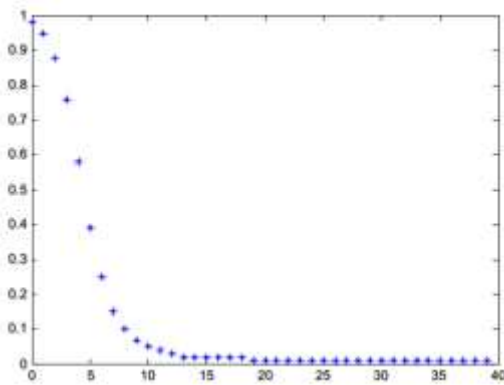


Figure. 4 Scatter plot in a local news unknown proportion of data

As can be seen from the figure, the number of unknown proportion of data that reliable data trends over time is almost the same with the image built SIR model above, to some extent, this article can explain the model has a certain reliability.

4.2.2 Based on GM (1,1) model to predict the spread of today's information

4.2.2.1 GM (1,1) basic ideas and principles

Gray theory think that every random amount is in a certain range, a certain period of time changes the amount of gray and gray process, incomplete of system information, uncertainty is the greyness. Do not look for their data processing and statistical probability distribution law, but processing the raw data for certain. It becomes a regular time-series data, found mathematical model based on this. Using GM (1,1) model based on the cumulative number of generated columns, model works as follows:

1) a cumulative for raw data

Let the original gray data $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)$,

denoted $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$

Make a cumulative yield, get

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), (k = 1, 2, \dots, n)$$

among them

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$$

Cumulative number of columns overcome volatility and randomness of raw data, change into strong regularity increasing the number of columns, be ready to build predictive models in the form of differential equations.

2) establish GM (1,1) model

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

Differential the equation is the gray forecast model GM (1,1), a, u which is a constant, it can be obtained by least squares fitting

$$\begin{pmatrix} a \\ u \end{pmatrix} = (B^T B)^{-1} B^T Y_n$$

among them

$$B = \begin{pmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{pmatrix}$$

$$Y_n = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))$$

Differential $\frac{dx^{(1)}}{dt} + ax^{(1)} = u$ solutions (referred to as time response function) is

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a}$$

Formula prediction the formula is the number of columns, Because the formula is predicted value once accumulated generating sequence can be original series value determined by the following formula

$$\hat{x}^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$$

4.2.2.2 Based applications GM (1,1) Model

As the group members did not find the specific data of five period about the situation of communicate news and information, therefore we decided to use the SIR model established above, by changing its initial value, to get information communication situation of five times specific data. Although this article does not use reliable data, but presents a solution to the problem. The initial value of the group on the basis of relevant experts on the use of the five periods as shown in Table 4

Table 4. The initial value of the SIR model in Table 3 of the five periods

period	1870s	1920s	1970s	1990s	2000s
λ	0.5	0.7	0.9	1.1	1.3
μ	0.2	0.25	0.3	0.35	0.4

According to SIR model, using Matlab software program to obtain five periods of time the number of unknown persons $s(t)$ and communicators $i(t)$ with the proportion of time trends as shown in figure 5.

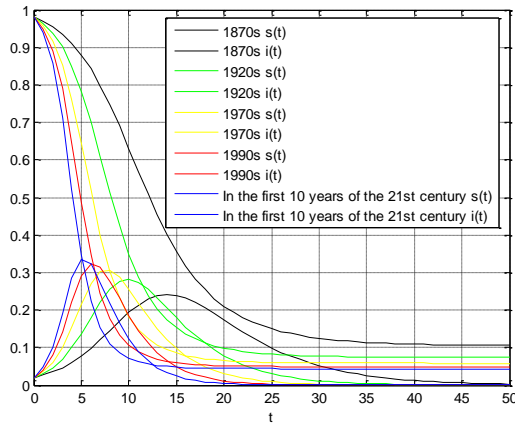


Figure. 5 the proportion of unknown persons $s(t)$ and disseminators $i(t)$ of five times trends over time

Due to a proportion of unknown person $s(t)$ can fully describe the situation with the information communication trends over time, so this article only to proportion of unknown persons $s(t)$ use the gray GM (1,1) prediction model for prediction. Use five times SIR data model of every day as the original data, to five times as the time line, use the gray GM (1,1) prediction model for prediction. Since the data is simulated to use its initial value of five times by changing the SIR model and actual data by a certain gap, so this article only gives a way to solve this problem. Using the gray GM (1,1) forecast model incorporates data five times per day, used on GM (1,1) model to forecast every day respectively, since the first day is the initial value of the data, five times is the same data, so to the next day, for example, obtained by today's information communication situation as shown in Table 5.

Table 5. The next data dissemination of information of the five periods

number	1	2	3	4	5
Unknown person	0.9687	0.963	0.9564	0.949	0.9404

Step 1 Ratio Test level

Establish the number of unknown proportion of five periods of time-series data as follows,

$$x_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(5)) \\ = (0.9687, 0.9630, 0.9564, 0.9490, 0.9404)$$

Level ratio

$$\sigma^{(0)}(k) \\ \sigma^{(0)}(k) = x^{(0)}(k-1) / x^{(0)}(k) \\ \sigma^{(0)} = (\sigma^{(0)}(2), \sigma^{(0)}(3), \dots, \sigma^{(0)}(5)) \\ = (1.0059 \quad 1.0069 \quad 1.0078 \quad 1.0091)$$

judge Level ratio, since all

$$\sigma^{(0)}(k) \in [1.0059 \quad 1.0091], k = 2, 3, \dots, 5$$

Therefore, you can use $x^{(0)}$ as satisfactory GM (1,1) Model

Step2 $x_1^{(0)}$ accumulate the original data, namely

$$x_1^{(1)} = (0.9687 \quad 1.9317 \quad 2.8881 \quad 3.8371 \quad 4.7775)$$

Step3 Construction of a data series B and Vector Data Y .

$$B = \begin{pmatrix} -\frac{1}{2}(x_1^{(1)}(1) + x_1^{(1)}(2)) & 1 \\ -\frac{1}{2}(x_1^{(1)}(2) + x_1^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x_1^{(1)}(4) + x_1^{(1)}(5)) & 1 \end{pmatrix}$$

$$Y_n = (x_1^{(0)}(2), x_1^{(0)}(3), \dots, x_1^{(0)}(5))$$

Step4 Compute \hat{a}

$$\hat{a} = (B^T B)^{-1} B^T Y = \begin{pmatrix} 0.0079 \\ 0.9750 \end{pmatrix} = (a, u)^T$$

Step5 Found Model

$$\frac{dx_1^{(1)}}{dt} + 0.0079 x_1^{(1)} = 0.9750$$

take $x_1^{(1)}(0) = x_1^{(0)}(1) = 0.9687$ obtaining the solution (referred to as time response function) to

$$\hat{x}_1^{(1)}(k+1) = (x_1^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a} = 123.504 - 122.536e^{-0.789409-2t}$$

Step6 Seeking to generate the number of column values

$\hat{x}_1^{(1)}(k+1)$ and restore the value of the model.

Make $k = 1, 2, \dots, 4$, from the function of the response time above can be considered $\hat{x}^{(1)}$, and take

$$\hat{x}^{(1)}(1) = \hat{x}^{(0)}(1) = x^{(0)}(1) = 0.9687$$

by $\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)$, take $k = 2, 3, \dots, 5$, get

$$\hat{x}^{(0)} = (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(5)) \\ = (0.9687 \quad 0.9635 \quad 0.9559 \quad 0.9484 \quad 0.9409)$$

4.2.2.3 Result analysis

Based on information dissemination of the model the original value and the model value of the five times, as well as residuals of the model, relative error and the average of accuracy in Table 6.

Table 6 Prediction and forecasting accuracy of the model

No.	Original value	Model values	Residuals	Relative error	Residual level
1	0.9687	0.9687			
2	0.9630	0.9635	-0.0004997	0.0519%	0.0020
3	0.9564	0.9559	0.0004763	0.0498%	0.0010
4	0.9490	0.9484	0.0005927	0.0625%	0.0001

Table 6 we can get, the accuracy of the model are higher, so it can be used to predict today's information communication situation

$$\hat{x}^{(0)}(6) = \hat{x}^{(1)}(6) - \hat{x}^{(1)}(5) = 0.9389$$

4.2.2.4 Inspection of GM (1,1) Model

Since our analysis of 2050 later in the relationship between communication and the ability to use the network in the case of today's information communication prediction data GM (1,1) Model. To illustrate the effectiveness of the prediction and assessment of the results, we will test to GM (1,1) model,

testing accuracy of GM(1,1) model , we will use the method of residual test .

The definition k point (or time) of the original time series $x^{(0)}$ is $x^{(0)}(k)$.form $x^{(0)}$ obtained the calculation of gray model is $\hat{x}^{(0)}(k)$ called $q(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$ is residuals of k point (or time) ; Relative error $\varepsilon(k)$ and average relative error $\varepsilon(avg)$ are defined as follows:

$$\varepsilon(k) = q(k) / x^{(0)}(k) * 100 \%$$

$$\varepsilon(avg) = 1/(n-1) \sum_{k=2}^n |\varepsilon(k)|$$

For $\varepsilon(k)$, the general requirements $\varepsilon(k) < 20\%$, preferably $\varepsilon(k) < 10\%$. Solving the test results of GM (1,1) model using MATLAB , we get $\varepsilon(avg) = 1.89\% < 10\%$.It showed that use GM (1,1) model to predict today's information communication situation is entirely possible.

Similarly, to predict on the second and third information communication skills, can get communication of information available the trend over time today , as shown in Figure 6.

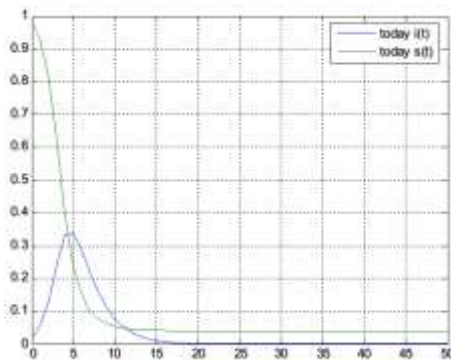


Figure 6 Map information communication trends today forecast

5. MODEL EVALUATION

5.1 Analysis of the advantages of the model

In this paper, conduct a comprehensive study to the flow of information ,the information flow has similarities with the spread of infectious diseases. So use the SIR model to model of the information flow, according to data obtained SIR model based on least squares polynomial fitting, then qualitative analysis to the SIR model, and the use of reliable data be verified to ensure that the flow of information on the SIR model established with a certain reliability.

Since there is no problem to get the information in the data flow to five times, using a modification of SIR model the initial value data obtained as the data flow of information five times, and then use the gray GM (1,1) model to predict for today and the flow of information in 2050.Although the initial data is simulated, there is no certain practical meaning, this model gives an effective way to solve such problems.

Prediction model network environment can be a good characterization of the news media. On this basis, this paper qualitative analyzes how in today's Internet world, so that the public interest and views are changed by information networks.

Topology or intensity form of messages or sources , information networks and how they are used to disseminate information and influence public opinion, studies on the links between social network information and to make a reference.

5.2 Analysis disadvantage of the model

Since news events related to network news, forums, blog, microblogging and other Internet information resources, information capacity, high redundancy and susceptible to noise interference. To collect comprehensive information on these and filtering, on the one hand a huge workload and high labor costs, On the other hand due to the passage of time, some of the information on the Internet will disappear, to be covered by the new information, Therefore, quantitative study of this article only collect and use information representative corpus as samples. Although, in the foregoing chapter we have made under the existing study based on the results of the theory and use of the public emergency network news also from a public opinion which reflects the evolution of the development and changes of the network news coverage. However, Insufficient sample of the present study was to analyze the qualitative part brings some limitations.

Since public opinion guidance and management is a very practical work, specific to the implementation level. We need to deal with a large number of complex issues and deal with complicated things, due to the lack of implementation experience, guidance and management research in the public emergency network public opinion aspects, This article does not find significant innovations, suggestions and measures proposed are mainly to further enrich the existing research-based. . In the future, we prefer to integrate the optimization with some traditional data mining approaches [3], [4], [5] to find more effective ways for optimization.

6. REFERENCES

- [1] Changjie Chen,YiMing Wei: Based Information Sharing for Sustainable Development Evaluation and Optimization [J](Systems Engineering:2011)
- [2] Marshall T. Poe, A History of Communications: Media and Society from the Evolution of Speech to the Internet (New York: Cambridge, 2011)
- [3] Ting Lu, Qi Gao, Xudong Wang, Cong Liu, "Modified Page Rank Model in Investigation of Criminal Gang", International Journal of Science and Engineering Applications, 4(3): 100-104, 2015.
- [4] Baohua Liu, Xudong Wang, Qi Gao, Cong Liu, "A K-Means based Model towards Ebola Virus Prorogation Prediction", International Journal of Science and Engineering Applications, 4(3): 95-99, 2015.
- [5] Libao Zhang, Faming Lu, An Liu, Pingping Guo and Cong Liu*, "Application of K-Means Clustering Algorithm for Classification of NBA Guards", International Journal of Science and Engineering Applications, 5(1), pp. 1-6, 2016.