

Grey theory based BP-NN co-training for dense sequence long-term tendency prediction

Journal:	<i>Grey Systems: Theory and Application</i>
Manuscript ID	GS-02-2020-0024.R3
Manuscript Type:	Full Length Paper
Keywords:	Grey prediction, Neural network, Co-training, Topic tendency prediction, Markov chain state transition

Title: Grey theory based BP-NN co-training for dense sequence long-term tendency prediction

Author(s) information:

Yuling Hong^{1,2}, Qishan Zhan^{1*} and Yingjie Yang³

¹School of Economics and Management, Fuzhou University, Fuzhou, People's Republic of China,

²Computer Engineering College, Jimei University, Xiamen, People's Republic of China,

³Institute of Artificial Intelligence, De Montfort University, Leicester, United Kingdom

Structure abstract:

Purpose:

The purpose of this paper is to solve the problems existing in topic popularity prediction in online social networks and advance a fine-grained and long-term prediction model for lack of sufficient data.

Design/methodology/approach:

Based on GM(1,1) and neural networks, a co-training model for topic tendency prediction is proposed in this paper. The interpolation based on GM(1,1) is employed to generate fine-grained prediction values of topic popularity time series and two neural network models are considered to achieve convergence by transmitting training parameters via their loss functions.

Findings:

The experiment results indicate that the integrated model can effectively predict dense sequence with higher performance than other algorithms, such as NN and RBF_LSSVM. Furthermore, the Markov chain state transition probability matrix model is used to improve the prediction results.

Practical implications:

Fine-grained and long-term topic popularity prediction, further improvement could be made by predicting any interpolation in the time interval of popularity data points.

Originality/value :

The paper succeeds in constructing a co-training model with GM(1,1) and neural networks. Markov chain state transition probability matrix is deployed for further improvement of popularity tendency prediction.

Funding information:

This work is supported by the Royal Society and NSFC International Exchanges project (IEC\NSFC\170391), the Education Commission Project of Fujian Province, China (JAT170327) and the National Science Foundation of Fujian Province, China (2018J01539).

Grey theory based BP-NN co-training for dense sequence long-term tendency prediction

Abstract

Purpose - The purpose of this paper is to solve the problems existing in topic popularity prediction in online social networks and advance a fine-grained and long-term prediction model for lack of sufficient data.

Design/methodology/approach - Based on GM(1,1) and neural networks, a co-training model for topic tendency prediction is proposed in this paper. The interpolation based on GM(1,1) is employed to generate fine-grained prediction values of topic popularity time series and two neural network models are considered to achieve convergence by transmitting training parameters via their loss functions.

Findings - The experiment results indicate that the integrated model can effectively predict dense sequence with higher performance than other algorithms, such as NN and RBF_LSSVM. Furthermore, the Markov chain state transition probability matrix model is used to improve the prediction results.

Practical implications - Fine-grained and long-term topic popularity prediction, further improvement could be made by predicting any interpolation in the time interval of popularity data points.

Originality/value - The paper succeeds in constructing a co-training model with GM(1,1) and neural networks. Markov chain state transition probability matrix is deployed for further improvement of popularity tendency prediction.

Keywords Grey prediction, Neural network, Co-training, Topic popularity prediction, Markov chain state transition

1. Introduction

Dense prediction refers to predicting fine-grained values of more than two values within a time interval based on information obtained from the past and current data points. In real social network platforms, prediction of topic popularity is usually provided for the next moments, maybe just up or down. The classic historical topic popularity prediction models employed the popularities at a specific time or the popularities in a series of time points at the early stage, and establish a model (e.g. regression) to reveal the developing trends of the popularity. Two of the existing methods are the SVM (Martens et al., 2008) and SH model (Szabo and Huberman, 2010). It is hard to precisely and timely forecast the popularity of some topics on the online social networks over a relatively long period of time (Wang et al., 2015). These models either have insufficient prediction accuracy or the required parameters are too complex to predict the future values of long-term and fine-grained time series for the complex social network. Furthermore, natural language processing, content extraction and other analysis approaches have been deployed (Kallus et al., 2010) to analyze the type of events, the characters involved and the emotional tone. Linear regression and a trained decision tree classifier have also been combined for the early time popularity prediction (Liu et al., 2015). To improve the forecasting accuracy, K-harmonic means and Particle Swarm Optimization were recently employed (Nie and Zhang, 2017).

As the popularity of the topic becomes increasingly random and non-linear, many scholars used grey models which appear as one of the most effective prediction models for very small samples (Deng, 1982). Combining the grey theory with the limit state equation, Zhu *et al.* (2013) established structural reliability model based on small-sample of training data and uncertainty of variable distribution. Wang *et al.* (2014) proposed an algorithm based on grey system theory to predict topic trends which are in a short life circle. Beyond that, Li *et al.* (2019) and Chen *et al.* (2019) improved the grey model for time series prediction. GM(1,1) model, a neural network and grey correlation analysis method were combined to construct a prediction model on short-term data sequence (Duan *et al.*, 2019, Hong *et al.*, 2020).

In this paper, the combination of grey theory and neural network approach makes use of their characteristics in dealing with fine-grained prediction and long-term series prediction. A new concept of co-training algorithm is introduced and then a novel topic popularity forecasting model is built based on GM(1,1) and neural networks. With the advantage of small data size and simple algorithm of grey prediction models, good performance of BP neural networks working for non-linear systems, and an integrated collaborative training process were finally integrated to relieve the problem of dense long-term prediction and improve the accuracy of topic popularity prediction. The performance of this integrated model is compared with the performance of GM(1,1), neural networks and RBF_LSSVM method (Wang *et al.*, 2016) independently. Our experiment results show the superiority of the proposed algorithm.

The remainder of this paper is organized as follows. Section 2 shows an overview of the related work, the motivations and objectives of the proposed model. Section 3 presents hybrid co-training strategies, including neural networks and grey models. Section 4 shows the comparison of RBF_LSSVM, GM(1,1), NN and GNNC. Markov chain state transition is shown in Section 5 to improve forecast results. Section 6 concludes this work and briefly discusses the possible future work.

2. Preliminaries

2.1 Dense Long-term Topic Tendency Nowcasting Problem

Denote $(c_1, c_2, \dots, c_m) \in R^m$ as m characteristics of a topic and define its popularity as x . And then if the observations about a topic are periodically recorded, an original popularity time sequence $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_t)$ is obtained. The dense long-term topic tendency nowcasting problem is to get a fine-grained length- K popularity sequence $(\hat{x}_{t+0.5}, \hat{x}_{t+1}, \hat{x}_{t+1.5}, \hat{x}_{t+2}, \hat{x}_{t+2.5}, \dots, \hat{x}_{t+K})$ in the near future.

2.2 GM(1,1) model

The processes of GM(1,1) model construction and solutions are as follows (Liu and Lin, 2010).

Step1. Assuming a given initial time sequence is as $\hat{X}^{(0)} = (\hat{x}_{(1)}^{(0)}, \hat{x}_{(2)}^{(0)}, \dots, \hat{x}_{(n)}^{(0)})$, where $\hat{x}_{(k)}^{(0)} \geq 0$ and n is the number of the original data. $\hat{X}^{(1)} = (\hat{x}_{(1)}^{(1)}, \hat{x}_{(2)}^{(1)}, \dots, \hat{x}_{(n)}^{(1)})$ is first-order accumulated generating operation (1-AGO) sequence of $\hat{X}^{(0)}$, in which

$$\hat{x}_{(k)}^{(1)} = \sum_{i=1}^k \hat{x}_{(i)}^{(0)}, k = 1, 2, \dots, n. \quad (1)$$

Step 2. Due to the fact that the time series is composed of discrete individuals, there are boundaries instead of intervals, thus there is no value between each pair of adjacent individuals. Therefore, the interior is calculated from the average of the left and right boundaries. $Z^{(1)} = (z_{(2)}^{(1)}, z_{(3)}^{(1)}, \dots, z_{(n)}^{(1)})$, in which

$$z_{(k)}^{(1)} = 0.5(\hat{x}_{(k)}^{(1)} + \hat{x}_{(k-1)}^{(1)}), (k = 2, 3, \dots, n). \quad (2)$$

Then the grey differential equation is

$$\frac{d\hat{x}^{(1)}}{dt} + a\hat{x}^{(1)} = w. \quad (3)$$

And the general equation of GM(1,1) model is

$$\hat{x}_{(k)}^{(0)} + az_{(k)}^{(1)} = w, \quad (4)$$

where $z_{(k)}^{(1)}$ is called the background value of GM(1,1) model.

$$\text{Let } Y = \begin{bmatrix} \hat{x}_{(2)}^{(0)} \\ \hat{x}_{(3)}^{(0)} \\ \vdots \\ \hat{x}_{(n)}^{(0)} \end{bmatrix}, B = \begin{bmatrix} -z_{(2)}^{(1)} & 1 \\ -z_{(3)}^{(1)} & 1 \\ \vdots & \vdots \\ -z_{(n)}^{(1)} & 1 \end{bmatrix}, \hat{\gamma} = [a, w]^T = (B^T B)^{-1} B^T Y, \text{ the solution of equation(3) is}$$

as equation (5).

$$\hat{x}_{(t)}^{(1)} = e^{-at} \left(\hat{x}_{(1)}^{(1)} - \frac{w}{a} \right) + \frac{w}{a} \quad (5)$$

Step 3. According to equation(1), $\hat{x}_{(1)}^{(1)} = \hat{x}_{(1)}^{(0)}$, then put it, a and w into equation (5), $\hat{x}_{(k+1)}^{(1)}$ = $e^{-a(k)} \left(\hat{x}_{(1)}^{(0)} - \frac{w}{a} \right) + \frac{w}{a}$, ($k = 1, 2, \dots, n-1$) is obtained.

In this way, the forecasting value $\hat{x}_{(k+1)}^{(0)}$ can be obtained from accumulative equation (6).

$$\hat{x}_{(k+1)}^{(0)} = \hat{x}_{(k+1)}^{(1)} - \hat{x}_{(k)}^{(1)} = e^{-a(k)} (1 - e^a) \left(\hat{x}_{(1)}^{(0)} - \frac{w}{a} \right) \quad (k = 1, 2, \dots, n-1). \quad (6)$$

2.3 Basic principle of BP-Neural network (BPNN)

BPNN is essentially a multilayer feedforward neural network which uses an optimized gradient descent algorithm to continuously regulate the network weights and thresholds (Gao et al., 2013). It is widely employed in many applications such as classification, function approximation, pattern recognition and data compression (Zhang et al. 2016).

2.4 Co-training Model

Co-training model is currently a popular semi-supervised machine learning method constructing two different classifiers and using a small-scale labeled operation to label large-scale unlabeled operations (Blum and Chawla, 2001). Two different learners are used in their independent learnings on the feature set, and the learning results of the two learners are combined to obtain solutions on a new sample, to reduce the error between the network output and the real output (Nigam and Ghani, 2000; Slivaka, Kovacevic and Konjovic, 2010).

Table 1 presents original co-training algorithm (Blum and Mitchell, 1998). Here the examples are labeled 'positive' or 'negative' based on their membership in one of the two classes.

Table 1 The co-training algorithm

Given:

- A set L of labeled training examples
- A set U of unlabeled examples
- Two feature perspectives positive x_1 and negative x_2 describing the examples

Create a pool U' of examples by choosing u examples at random from U .

Loop for k iterations:

- Use L to train a classifier l_A that considers only the portion x_1 of x .
 - Use L to train a classifier l_B that considers only the portion x_2 of x .
 - Allow l_A to label the most confidently predicted p positive and n negative examples from U' .
 - Allow l_B to label the most confidently predicted p positive and n negative examples from U' .
 - Add these self-labeled examples to L .
 - Randomly choose $2p+2n$ examples from U to replenish U' .
-
-

3. Grey theory based NN co-training algorithm (GNNC)

3.1 Background and interpolation method

Due to factors such as application scenarios and storage, historical data of topics is usually sparsely collected and data with small time intervals are sometimes unavailable, and the long-term topic tendency prediction depends more on the data which are recently predicted but not on the existing historical data. It is necessary to predict fine-grained and long-term data based on time-sparse historical popularity.

The grey prediction model can provide data predictions for nonlinear and uncertain systems with small dataset. However, there is often too much deviation in the real-world application. For instance, in a switch, a failure or an interference situation, the prediction of the series will be abnormal, which will destroy the stability of the grey model. In contrast, the artificial neural networks can predict the occurrence of some special situations by learning predictable mutation data. Nevertheless, it still has its weakness when the experimental data are not natural sufficient for redundant training. In the other hand, too many training models need larger network structure, which not only reduces the learning efficiency, but also consumes a lot of resources. In addition, a small number of meaningful data points may be overwhelmed by a large amount of normal data, making it unable to express the truth.

In this paper, a grey model is employed to interpolate the historical data on the time scale and two neural networks are used to form a mirror network iterative training system based on mutually provided predictive data, learning from the information provided by the original data and the information provided by the grey model. During the iterations, it constrains parameters of the two mirrored neural networks to converge them and make the results unambiguous. It is an attempt to strengthen the training set and improve accuracy of both weak initial neural networks, while using only an insufficient amount of available data.

As a consequence, based on the GM(1,1) model and the neural networks, a machine learning method for long-term topic tendency nowcasting is proposed as followed.

Assume that $P_i = (p_{i1}, p_{i2}, \dots, p_{im}) \in R^m$ and $W_i = (w_{i1}, w_{i2}, \dots, w_{im}) \in R^m$ means m characteristics of the i th topic and their corresponding weights. The time sequence of i th topic popularity is defined as:

$$\text{Topic} = (X_1, X_2, \dots, X_n), X_i = (x_i^1, x_i^2, \dots, x_i^m) = \frac{W_i \odot P_i}{(W_i, P_i)}, i = 1, 2, \dots, n. \quad (7)$$

For each k , we use GM(1,1) model to interpolate values between $x_{(k)}^{(0)}$ and $x_{(k+1)}^{(0)}$, consequently yield a sequence $(x_{(1)}^{(0)}, \hat{x}_{(1)}^{(0)}, x_{(2)}^{(0)}, \hat{x}_{(2)}^{(0)}, \dots, x_{(k)}^{(0)}, \hat{x}_{(k)}^{(0)}, \dots, x_{(t)}^{(0)}, \hat{x}_{(t)}^{(0)})$, in which $\hat{x}_{(k)}^{(0)}, (k = 1, 2, \dots, t)$ are interpolated values, and each of which is generated by GM(1,1) from the 4-dimension sub-series $(x_{(k-3)}^{(0)}, x_{(k-2)}^{(0)}, x_{(k-1)}^{(0)}, x_{(k)}^{(0)})$ (Deng, 1982).

3.2 Modelling Densely-Forecasting Structure

To obtain a forecasting value $\hat{x}_{(k+1)}^{(0)}$, we deploy a sequence $(x_{(k-m+1)}^{(0)}, x_{(k-m+2)}^{(0)}, \dots, x_{(k)}^{(0)})$, $(m = 2, 3, \dots, k)$ as a neural network's input.

Therefore, we use the sequence $\hat{X}^{(0)} = (\hat{x}_{(1)}^{(0)}, \hat{x}_{(2)}^{(0)}, \dots, \hat{x}_{(t)}^{(0)})$ as input and $X^{(0)} = (x_{(t+1)}^{(0)}, x_{(t+2)}^{(0)}, \dots, x_{(t+p)}^{(0)})$ as the training label to train NN_A . In the same time, the labelled series $\hat{X}^{(1)} = (\hat{x}_{(1)}^{(1)}, \hat{x}_{(2)}^{(1)}, \dots, \hat{x}_{(t)}^{(1)})$ can be obtained on the basis of the input series $\hat{X}^{(0)}$. On the other hand, we use the sequence $X^{(0)} = (x_{(1)}^{(0)}, x_{(2)}^{(0)}, \dots, x_{(t)}^{(0)})$ as input and $\hat{X}^{(0)} = (\hat{x}_{(1)}^{(0)}, \hat{x}_{(2)}^{(0)}, \dots, \hat{x}_{(t)}^{(0)})$ as the training label to train NN_B . Meanwhile, we get the labelled series $\hat{X}^{(2)}$ from NN_B based on the input series $\hat{X}^{(1)}$. After that, this series $\hat{X}^{(2)}$ is used as input and $X^{(0)}$ as the training label to obtain NN_A and the labelled series $\hat{X}^{(3)}$. And then, $\hat{X}^{(3)}$ is employed as input and $X^{(0)}$ as the training label and the labelled series $\hat{X}^{(4)}$ is obtained. In this way, after many iterations of training, when it reaches the number of iterations or the threshold, it stops training.

Schematic diagram of GNN based co-training model structure and model of grey theory based neural networks are shown in Fig. 1.

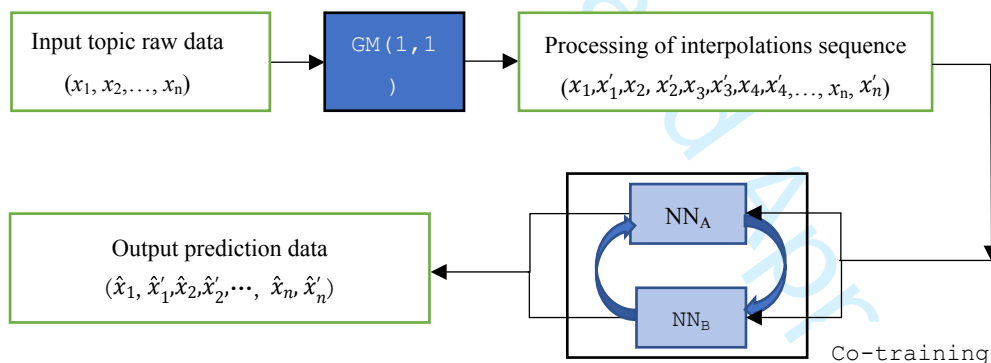


Fig. 1 Schematic diagram of GNN based co-training model structure

The specific algorithm of grey theory based BP-NN co-training for densely popularity tendency prediction is shown in Table 2.

Table 2 The integrated grey theory based NN co-training algorithm

Given:

- Time series $\tilde{X}^{(0)} = \left(x_{(1)}^{(0)}, \hat{x}_{(1)}^{(0)}, x_{(2)}^{(0)}, \hat{x}_{(2)}^{(0)}, \dots, x_{(t)}^{(0)}, \hat{x}_{(t)}^{(0)}, x_{(t+1)}^{(0)}, \hat{x}_{(t+1)}^{(0)}, x_{(t+2)}^{(0)}, \dots, x_{(n)}^{(0)}, \hat{x}_{(n)}^{(0)} \right)$, where $\hat{x}_{(1)}^{(0)} = 0.5(x_{(1)}^{(0)} + x_{(2)}^{(0)})$, $\hat{x}_{(2)}^{(0)} = 0.5(x_{(2)}^{(0)} + x_{(3)}^{(0)})$, $\hat{x}_{(3)}^{(0)} = 0.5(x_{(3)}^{(0)} + x_{(4)}^{(0)})$, $\hat{x}_{(4)}^{(0)} = 0.5(x_{(4)}^{(0)} + x_{(5)}^{(0)})$.
- Time series $\hat{X}_t = (\hat{x}_{(t-3)}^{(0)}, \hat{x}_{(t-2)}^{(0)}, \hat{x}_{(t-1)}^{(0)}, \hat{x}_{(t)}^{(0)})$ of labeled popularity examples generated by GM(1,1)
- 4-time sub-series $\tilde{X}_t = (x_{(t-3)}^{(0)}, x_{(t-2)}^{(0)}, x_{(t-1)}^{(0)}, x_{(t)}^{(0)})$ of raw popularity examples
- Two future views X_{t+1} and \hat{X}_{t+1} describing the examples.

Loop for k iterations:

- Use X_{t+1} to train a learning machine NN_A that considers only the \hat{X}_t portion of $\tilde{X}^{(0)}$.
 - Use \hat{X}_{t+1} to train a learning machine NN_B that considers only the \tilde{X}_t portion of $\tilde{X}^{(0)}$.
 - Allow NN_A to use GM(1,1) to analogy $(n-4)$ -dimension examples $(\hat{x}_{(5)}^{(1)}, \hat{x}_{(6)}^{(1)}, \dots, \hat{x}_{(n)}^{(1)})$ for $\tilde{X}^{(0)}$.
 - Allow NN_B to use GM(1,1) to analogy $(n-4)$ -dimension examples $(x_{(5)}^{(1)}, x_{(6)}^{(1)}, \dots, x_{(n)}^{(1)})$ for $\tilde{X}^{(0)}$.
 - Calculate $e(W_A^{k+1})$ and $e(W_B^{k+1})$, and e_{co} .
 - If $e_{co} > e_{ter}$, then add these re-predicted examples $\tilde{X}^{(0)}$ to update $\tilde{X}^{(1)}$ and renew W_A^{k+1} and W_B^{k+1} .
-
-

3.3 Evaluation of the integrated prediction system

3.3.1 Cost function of GNNC model

Let $x_{(i)}^{(0)}$ be the input sequence and $\hat{x}_{(i)}^{(k)}$ be the prediction result of each iteration. The objective function is to calculate the difference between the output of the GNNC model and the input, which is shown as $E(x_{(i)}, \hat{x}_{(i)}) = \frac{1}{t} \sum_{i=1}^t w_i (x_{(i)}^{(0)} - \hat{x}_{(i)}^{(k)})^2$. The cost function of the integrated model for dense long-term topic popularity tendency prediction is presented as formulas (8) and (9). The first term of the cost function is the objective function. The second one is used for regularization. The third one imposes the punishment on discrepancy between NN_A and NN_B , so it forces the training of NN_A and NN_B converges to an identical one as long as we gradually enlarge λ_A and λ_B while k increases. This guarantees the divergence and ambiguity of training results are eliminated.

$$L(W_A^{k+1}) = \sum_{i=1}^t E(x_{(i)}^{(0)}, \hat{x}_{(i)}^{(k)}) + \lambda_A^{(0)} \|W_A^k\|_2^2 + \lambda_B^{(k+1)} \|W_A^{k+1} - W_B^k\|_2^2 \quad (8)$$

$$L(W_B^{k+1}) = \sum_{i=1}^t E(x_{(i)}^{(0)}, \hat{x}_{(i)}^{(k)}) + \lambda_B^{(0)} \|W_B^k\|_2^2 + \lambda_A^{(k+1)} \|W_B^{k+1} - W_A^k\|_2^2 \quad (9)$$

Here, λ_A and λ_B are two parameters for loss function. The optimal setting of λ_A is determined by trials of tuning it in the first iteration of co-training, and then it is fixed in successive iterations. λ_B is set to zero at the beginning of co-training, and then linearly increases against k th iteration.

$$W_A^{k+1} = \arg \min_{W_A^{k+1}} \left[\sum_{i=1}^t E(x_{(i)}^{(0)}, \hat{x}_{(i)}^{(k)}) + \lambda_A^{(0)} \|W_A^k\|_2^2 + \lambda_B^{(k+1)} \|W_A^{k+1} - W_B^k\|_2^2 \right] \quad (10)$$

$$W_B^{k+1} = \arg \min_{W_B^{k+1}} \left[\sum_{i=1}^t E(x_{(i)}^{(0)}, \hat{x}_{(i)}^{(k)}) + \lambda_B^{(0)} \|W_B^k\|_2^2 + \lambda_A^{(k+1)} \|W_B^{k+1} - W_A^k\|_2^2 \right] \quad (11)$$

3.3.2 Prediction accuracy rate

In order to demonstrate the accuracy of the proposed prediction model, the actual value $x_{(k)}^{(0)}$ and the predicted value $\hat{x}_{(k)}^{(0)}$ can be compared. Eqs. (12) is the accuracy evaluation criteria used in this work.

$$\text{Mean Relative Error: } MRE = \frac{1}{n} \sum_{i=1}^n \frac{|x_{(k)}^{(0)} - \hat{x}_{(k)}^{(0)}|}{x_{(k)}^{(0)}} \times 100\% \quad (12)$$

4. Comparison of RBF_LSSVM, GM(1,1), NN and GNNC

The dataset of the Weibo is used to verify the computing performance of the GNNC algorithm on the real social network. The topic popularity is extracted from Weibo for the period from 22:00 14th Dec 2019 to 22:30 15th Dec 2019 as shown in Table 3, where the data are highly nonlinear and unstable. According to the GNNC model established above, data from 22:00 14th to 19: 00 15th are used for training, and the rest of the data is used for testing. This period of time is the time when most online users are active in a day.

Table 3 Topic popularity data (Data Source: Weibo 14.12.2019 to 15.12.2019)

time	popularity	time	popularity
22:00	0.604	13:00	44.747
23:00	2.604	14:00	20.019
0:00	4.392	15:00	11.830
1:00	3.653	16:00	10.659
2:00	1.732	17:00	20.123
3:00	1.947	18:00	43.284
4:00	4.928	19:00	61.662
5:00	4.586	19:30	88.112
6:00	16.015	20:00	95.167
7:00	19.293	20:30	98.668
8:00	44.447	21:00	100.000
9:00	79.287	21:30	99.000
10:00	80.578	22:00	101.896
11:00	100.000	22:30	102.987
12:00	100.000		

Each prediction result and the corresponding relative error (RE) can be found in Table 4. Comparison of the popularity prediction of RBF_LSSVM model^[26], grey theory based NN co-training (GNNC) model, the traditional GM(1,1) and Neural Network (NN) model is shown. In the experiment of RBF_LSSVM, the interpolation is also generated via its previous four data points.

Table 4 Comparison of popularities prediction and RE from 19: 30 to 22: 30

Time	GM(1,1) prediction	RE of GM	NN prediction	RE of NN	RBF_LSSVM prediction	RE of RBF_LSSVM	GNNC prediction	RE of GNNC
19:30	79.6845	9.6%	71.478	18.9%	82.1011	6.8%	85.159	3.4%
20:00	82.721	13.1%	87.167	8.4%	97.159	2.1%	96.167	1.1%
20:30	85.998	12.8%	91.4395	7.3%	99.9905	1.3%	96.816	1.9%

21:00	89.275	10.7%	95.712	4.3%	102.822	2.8%	99.798	0.2%
21:30	92.812	6.3%	96.81	2.2%	104.3045	5.4%	99.285	0.3%
22:00	96.348	5.4%	95.899	5.9%	105.787	3.8%	100.998	0.9%
22:30	99.548	3.3%	97.018	5.8%	110.801	7.6%	103.892	0.9%
MRE		8.74%		7.54%		4.26%		1.23%

Table 4 shows that the GM(1,1) interpolation method works reasonably well in topic popularity prediction, but the prediction accuracy of the GNNC model is far greater than that of the traditional GM(1,1) model and that it is also superior to prediction accuracy of the NN model. The MRE of GNNC is the smallest among them, which means that on a list of 100 topics, the prediction accuracy is nearly 97%. Especially, when it comes to the densely forecasting time interval, the prediction accuracy of GNNC is much better than the others. Furthermore, the RE of the intermediate points calculated with the RBF_LSSVM algorithm is larger than others, which may be related to the lack of sensitivity of the data to the sparsity of the data. Probably because RBF_LSSVM is not sensitive enough to the data sparsity. Therefore, GM(1,1) interpolation approach is effective for various methods, such as NN and RBF_LSSVM, and it works better with NN.

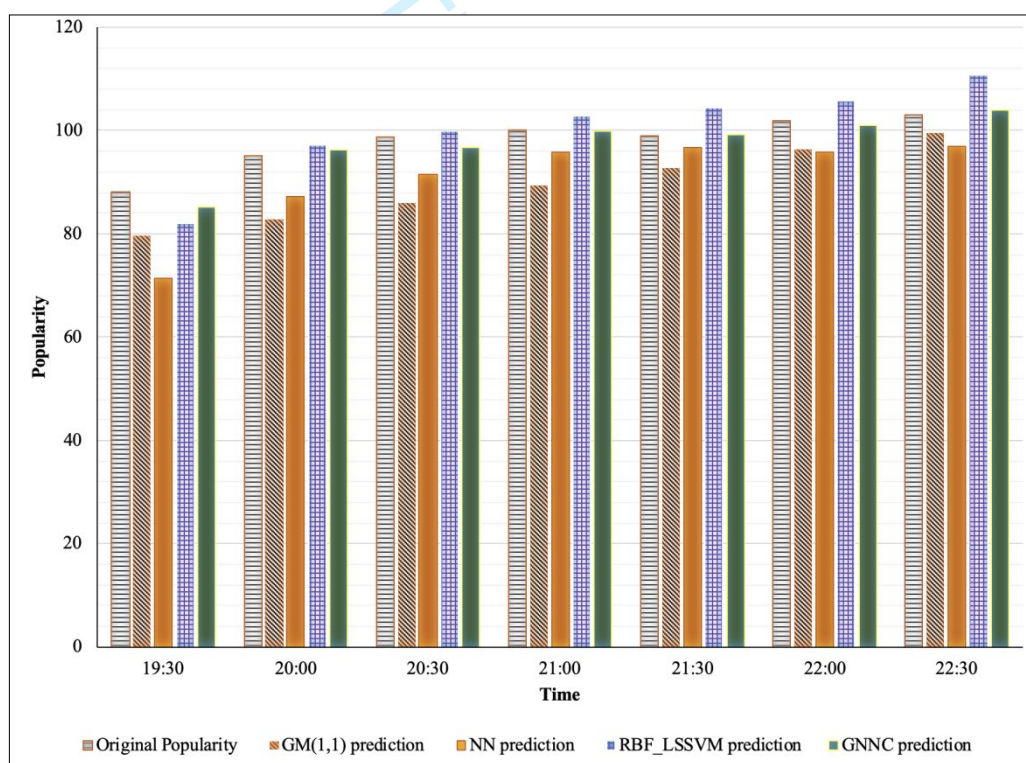


Fig.2 Comparison chart of popularities prediction

Fig.2 shows after 19:30, more and more people involved in network topics after their dinner. Among these comparative models, GNNC model proposed in this paper and RBF_LSSVM have better prediction performance than GM(1,1) and NN models. Moreover, GNNC performed more stable than RBF_LSSVM.

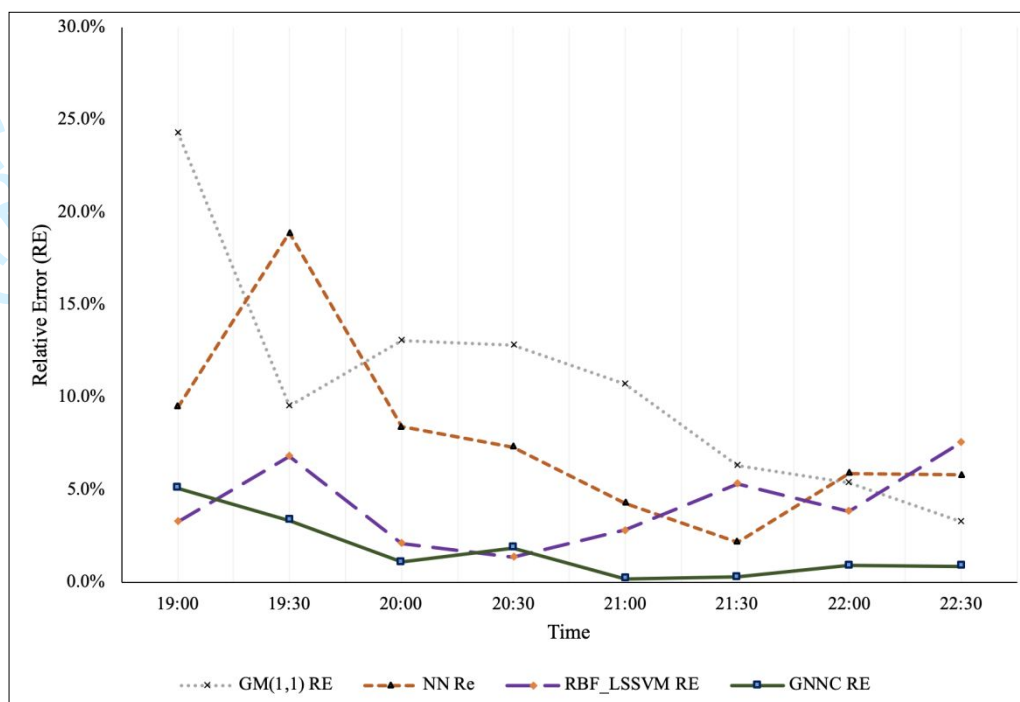


Fig.3 Comparison chart of relative errors

As can be seen from Fig.3, the relative error (RE) of GNNC model preforms the best among these algorithms. At the moment 19:30 when more users were online, NN had a relatively obvious turning point. Although the LSSVM model can provide better performance on S-type data, the GM (1,1) model is good at mapping the monotonically increasing data. According to the GNNC algorithm, its RE at half-time are less than 5% which means that this model has the ability to fine-grained prediction for a long-term sequence.

5. Markov chain state transition applied in topic tendency prediction

The Markov prediction method estimates the future tendency of the system based on the transition probability between states. The transition probability reflects the influence of various random factors and some inherent regularity between states, which can well describe the data sequence. It is very suitable for data prediction problems with random fluctuations. Because that the non-stationary random sequence of topic popularity changes in accordance with the characteristics of Markov chain, it can be divided into 5 states. State01 means “the most popular” in the list, while State02, State03, State04 and State05 refer to “popular”, “ordinary popular”, “little popular” and “unpopular”.

According to the above analysis, co-training improvements are made to the topic popularity prediction based on the GM(1,1) model and neural networks from 2019-12-13 22:00 to 2019-12-14 22:30. Data at the time period of 2019-12-13 22:00 to 2019-12-14 19:00 is used as training dataset while data from 2019-12-14 19:30 to 22:30 is used as testing dataset. According to the Markov chain prediction approach (Liao, 2015), the predicted state vectors from 19:30 to 22:30 are obtained, as shown in Table 5.

Table 5 State probability of grey theory based NN co-training prediction results

State	Time
-------	------

Vectors	19:30	20:00	20:30	21:00	21:30	22:00	22:30
State01	0.333	0.349	0.342	0.342	0.344	0.343	0.341
State02	0.333	0.206	0.192	0.195	0.197	0.196	0.194
State03	0	0.095	0.134	0.137	0.129	0.121	0.115
State04	0.333	0.302	0.245	0.212	0.197	0.190	0.188
State05	0	0.048	0.086	0.114	0.134	0.146	0.163

As is shown in Table 6, after the improvement of the Markov chain probability method, the actual state will be in the most popular range. And their maximum possible probabilities are 33.3%, 34.9%, 34.2%, 34.2%, 34.4%, 34.3%. It means that it is a time when more users have leisure time on online social network and the popularity tendency of the topic may face three kinds of states at the next moment. Later, the topic became more and more popular with online users, which is in accordance with the real situation in Table 3.

Table 6 Forecast result of topic popularity tendency from 19:30 to 22:30

Time	States	State probability	Prediction interval value	Time	States	State probability	Prediction interval value
2019-12-14 19:30	01	0.333	[81-100]	2019-12-14 21:30	01	0.344	[81-100]
	02	0.333	[61-80]		02	0.197	[61-80]
	03	0	[41-60]		03	0.129	[41-60]
	04	0.333	[21-40]		04	0.197	[21-40]
	05	0	[0-20]		05	0.134	[0-20]
2019-12-14 20:00	01	0.349	[81-100]	2019-12-14 22:00	01	0.343	[81-100]
	02	0.206	[61-80]		02	0.196	[61-80]
	03	0.095	[41-60]		03	0.121	[41-60]
	04	0.302	[21-40]		04	0.190	[21-40]
	05	0.048	[0-20]		05	0.146	[0-20]
2019-12-14 20:30	01	0.342	[81-100]	2019-12-14 22:30	01	0.337	[81-100]
	02	0.192	[61-80]		02	0.191	[61-80]
	03	0.134	[41-60]		03	0.111	[41-60]
	04	0.245	[21-40]		04	0.187	[21-40]
	05	0.086	[0-20]		05	0.175	[0-20]
2019-12-14 21:00	01	0.342	[81-100]				
	02	0.195	[61-80]				
	03	0.137	[41-60]				
	04	0.212	[21-40]				
	05	0.114	[0-20]				

Fig.4 shows the state probability of prediction result by GNNC from 19:30 to 22:30. The future development tendency of these five states is around 0.5 or even lower. The future topic popularity in State01 can still maintain a probability of nearly 35% which means that users on the network pay more attention after their dinner.

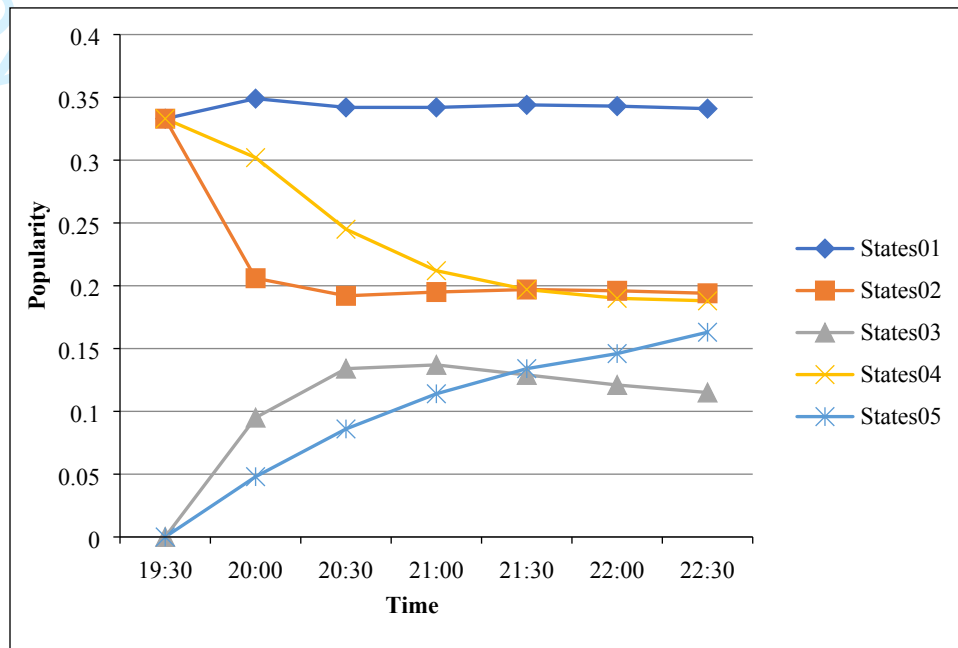


Fig.4 Prediction state probability from 19:30 to 22:30

6. Conclusions

Long-term and fine-grained prediction of topic popularity provides a quantitative basis for the analysis of topic tendency on social networks. The long-term, fine-grained popularity prediction of new topics is highly unpredictable, and the management and services of social networks are affected to a large extent especially at emergency management decision-making.

This work proposes a new model to combine grey models with co-training based neural networks for prediction of topic popularity tendency of online social network. Firstly, the GM(1,1) prediction is introduced to predict the internal interpolation to improve the intensive representation of the sequence of topic popularity. Secondly, based on the prediction of co-training frame, GNNC model is proposed to achieve highly accurate prediction. Thirdly, the method of Markov chain state transition probability matrix is deployed for further improvement of popularity tendency prediction. In the future research, it is important and meaningful to analyze the non-equidistant data of topic popularity streams and utility measure of Markov prediction method, and involve more than two neural networks in the co-training.

References

- Blum A. and Mitchell T. (1998), "Combining labeled and unlabeled data with co-training", *Proceeding of the 11th Annual Conference on Computational Learning Theory in USA*, pp.92–100.
- Blum A. and Chawla S. (2001), "Learning from labeled and unlabeled data using graph mincuts", *Proceeding of the 18th International conference on machine learning in USA*, pp.19-26.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Chen Y., Liu H. and Hsieh H. (2019), "Time series interval forecast using GM(1,1) and NGBM(1,1) models", *Soft computing*, Vol.23, No.3, pp.1541-1555.
- Deng J.L. (1989), "Introduction to grey system theory", *The Journal of Grey System*, Vol.1, No. 1, pp.1-24.
- Duan H., Liu Y. and Wang D. (2019), Prediction of a multi-mode coupling model based on traffic flow tensor data[J]. *Journal of Intelligent and Fuzzy Systems*, Vol.36, No.2, pp.1691-1703.
- Gao X., Mo L., Wen Q. and Katayama S. (2013), "Neural network model for recognizing joint offset during fiber laser welding", *Weld Journal*, Vol. 92, No.9, pp.251-257.
- Hong Y. L., Yang Y. J., Zhang Q. S. and Wu L. (2020), "An integrated model combining grey methods and neural networks and its application to burst topic tendency prediction", *Journal of Grey System*, accepted, article in press.
- Li C., Yang Y.J. and Liu S.F. (2019), "A new method to mitigate data fluctuations for time series prediction", *Appl Math Model*, Vol.65, pp.390-407.
- Liao P. M. (2015), "Commodity market state prediction based on markov chain state transition probability matrix", *Statistics and Decision*, Vol.422, No.2, pp.97-99.
- Liu S. and Lin Y. (2010), *Grey Systems: Theory and applications*. Understanding Complex Systems. Springer: Berlin, Germany, pp.92-111.
- Szabo G. and Huberman A.B. (2010), "Predicting the popularity of online content", *Communications of the ACM*, Vol.53, No.8, pp.80-88.
- Kwak H., Lee C., Park H. and Moon S. (2010), "What is Twitter, a social network or a news media?", *Proceedings of the 19th International Conference on World Wide Web in USA*, pp. 591-600.
- Martens D., Bruynseels L., Baesens B., Willekens M., and Vanthienen J. (2008), "Predicting going concern opinion with data mining". *Decision Support Systems*, Vol.45, No.4, pp.765-777.
- Nie F. and Zhang P. (2017), "Study on prediction and early warning model of public opinion basing on K-harmonic means and Particle Swarm Optimization", *Information Research*, Vol. 235, No.5, pp.6-9.
- Wang P., Wu C., Gao X. (2016), "Research on subway passenger flow combination prediction model based on RBF neural networks and LSSVM". *Proceeding of 2016 Chinese control and decision conference in China*, pp.6283-6287.
- Wang X., Qi L., Chen C., Tang J., Jiang M. (2014), "Grey system theory based prediction for topic trend on internet", *Engineering Applications of Artificial Intelligence*, Vol.29, No.3, pp.191-200.
- Wang Z., Shou L., Chen K., Chen G., & Mehrotra S. (2015), "On summarization and timeline generation for evolutionary tweet streams", *Knowledge and Data Engineering*, Vol.27, No.5, pp.1301-1315.
- Zhang S., Wang B., Li X., Chen H. (2016), "Research and application of improved gas concentration prediction model based on grey theory and BP neural network in digital mine", *Proceeding of the 9th International Conference on Digital Enterprise Technology-Intelligent Manufacturing in the Knowledge Economy Era in China*, pp.471-475.
- Zhu X. and Fang Z. (2013), "'DGM-AMSAA' Model of reliability growth based on the small sample"[C]. *Proceedings of the 2013 IEEE International Conference on Grey Systems and Intelligent Services in China*, pp.123-126.