Stock Price Prediction Through the Mixture of Gaussian Processes via the Precise Hard-cut EM Algorithm

Shuanglong Liu and Jinwen Ma(✉)

Department of Information Science, School of Mathematical Sciences and LMAM, Peking University, Beijing 100871, China
jwma@math.pku.edu.cn

Abstract. In this paper, the mixture of Gaussian processes (MGP) is applied to model and predict the time series of stock prices. Methodically, the precise hard-cut expectation maximization (EM) algorithm for MGPs is utilized to learn the parameters of the MGP model from stock prices data. It is demonstrated by the experiments that the MGP model with the precise hard-cut EM algorithm can be successfully applied to the prediction of stock prices, and outperforms the typical regression models and algorithms.

Keywords: Mixture of Gaussian processes · EM algorithm · Parameter learning · Stock price · Times series prediction

1 Introduction

The stock market has the characteristics of high return and high risk [1], which has always been concerned on the analysis and forecast of stock prices. Actually, the complexity of the internal structure in stock price system and the diversity of the external factors (the national policy, the bank rate, price index, the performance of quoted companies and the psychological factors of the investors) determine the complexity of the stock market, uncertainty and difficulty of stock price forecasting task [2]. Because the stock price is collected according to the order of time, it actually forms a complex nonlinear time series [3]. Some traditional stock market analysis methods, such as stock price graph analysis (k line graph [4]), cannot profoundly reveal the stock intrinsic relationship, so that the prediction results are not so ideal on stock price. Stock price prediction methodologies fall into three broad categories which are fundamental analysis, technical analysis (charting) and technological methods.

From the view of mathematics, the key to effective stock price prediction is to discover the intrinsic mapping or function, and to fit and approximate the mapping or the function. As it has been quickly developed, the mixture of Gaussian processes (MGP) model [5] is a powerful tool for solving this problem. But most of the MGP models are very complex and involve a large number of parameters and hyper-parameters, which makes the application of the MGP models very difficult [6]. Thus, we adopt the MGP model which proposed in [7] with excluding unnecessary priors and carefully selecting the model structure and gating function. This MGP model

D.-S. Huang et al. (Eds.): ICIC 2016, Part III, LNAI 9773, pp. 282–293, 2016.
DOI: 10.1007/978-3-319-42297-8_27

remains the main structure, features and advantages of the original MGP model. Moreover, it can be effectively applied to the modeling and prediction of nonlinear time series via the precise hard-cut EM algorithm. In fact, the precise hard-cut EM algorithm is more efficient than the soft EM algorithm since we could get the hyper-parameters of each GP independently in the M-step. It was demonstrated by the experimental results that this precise hard-cut EM algorithm for the MGP model really gives more precise prediction than some typical regression models and algorithms.

Along this direction, we apply the MGP model to the short-term stock price forecasting via the precise hard-cut EM algorithm. The experimental results show that this MGP based method can find potential rules from historical datasets, and their forecasting results are more stable and accurate.

The rest of this paper is organized as follows. In Sect. 2, we give a brief review of the MGP model and introduce the precise hard-cut EM algorithm. Section 3 presents the framework of stock price forecasting and the experimental results of the MGP based method as well as the comparisons of the regression models and algorithms. Finally, we give a brief conclusion in Sect. 4.

2 The Precise Hard-cut EM Algorithm for MGPs

2.1 The MGP Model

We consider the MGP model as described in [7]. In fact, it can be viewed as a special mixture model where each component is a GP. The whole set of indicators $Z = [z_1, z_2, \ldots, z_N]^T$, inputs X and outputs Y are sequentially generated and the MGP model is mathematically defined as follows:

$$p(z_t = c) = \pi_c, t = 1, 2, \ldots, N, \tag{1}$$

$$p(x_t | z_t = c, \theta_c) = N(x_t | \mu_c, S_c), t = 1, 2, \ldots, N, \tag{2}$$

$$p(y|X, \theta) = \prod\nolimits_{c=1}^{C} N(y_c | 0, K(X_c, X_c | \theta_c) + \sigma_c^2 I_{N_c}) \tag{3}$$

where $K(x_i, x_j) = g^2 \exp\{-\frac{1}{2}(x_i - x_j)^T B (x_i - x_j)\}$, $B = \mathrm{diag}\{b_1^2, b_2^2, \ldots, b_d^2\}$, and Eq. (2) adopts Gaussian inputs in most generative MGP models [8–10]. $\theta_c = \{\pi_c, \mu_c, S_c, g_c, b_{c,1}, b_{c,2}, \ldots, b_{c,d}, \sigma_c\}$ are the parameters in the c-th GP component and $\theta = \{\theta_c\}_{c=1}^{C}$ denotes all the parameters in the mixture model.

The generative structure is prominent and clear for the MGP model, and the model avoids the complicated parameters setting. In various GP components, Gaussian means μ_c are different so that each component concentrates on the different region and this mixture model can fit multimodal dataset.

2.2 The Precise Hard-cut EM Algorithm

To avoid the computational complexity of Q function, it is reasonable to use the hard-cut version of the EM algorithm and we then can efficiently learn the parameters for the MGP model. In fact, the precise hard-cut EM algorithm [7] is a good choice and we summarize its procedures as follows:

Algorithm 1. The Precise Hard-cut EM Algorithm

1: Initialization of indicators:
Cluster $\{(x_t, y_t)\}_{t=1}^{N}$ into C classes by the k-means clustering, and set z_t as the indicator of the *t*-th sample to the cluster.
2: **repeat**
3: M-step:
Calculate π_c, μ_c and S_c in the way of the Gaussian mixture model:

$$\pi_c = \frac{1}{N}\sum_{t=1}^{N} I(z_t = c) \tag{4}$$

$$\mu_c = \frac{\sum_{t=1}^{N} I(z_t=c)x_t}{\sum_{t=1}^{N} I(z_t=c)} \tag{5}$$

$$S_c = \frac{\sum_{t=1}^{N} I(z_t=c)(x_t-\mu_c)(x_t-\mu_c)^T}{\sum_{t=1}^{N} I(z_t=c)} \tag{6}$$

and obtain the GP parameters by maximizing the likelihood.
4: E-step:
Classify each sample into the corresponding component according to the MAP criterion:

$$z_t = \mathrm{argmax}_c p(z_t = c|x_t, y_t) = \mathrm{argmax}_c \pi_c N(x_t|\mu_c, S_c) N(y_t|0, g_c^2 + \sigma_c^2) \tag{7}$$

5: **until** Either the component remains the same in the previous iteration, or the iteration number reaches certain threshold.
6: Output the estimated parameters of MGP.

After the convergence of the precise hard-cut EM algorithm, we have obtained the estimates of all the parameters for the MGP. For a test input x^*, we can classify it into the z-th component of the MGP by the MAP criterion as follows:

$$z = \mathrm{argmax}_c p(z^* = c|x^*) = \mathrm{argmax}_c \pi_c N(x^*|\mu_c, S_c) \tag{8}$$

Based on such a classification, we can predict the output of the test input via the corresponding GP using

$$\hat{y}^* = K(x^*, X)\left[K(X,X) + \sigma^2 I\right]^{-1} y \tag{9}$$

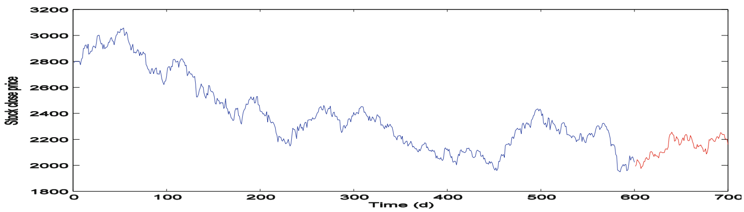
In the next section, the precise hard-cut EM algorithm for the MGP model will be used for the stock closing price prediction, and the obtained results will be compared with the classical regression models and algorithms.

3 Stock Price Prediction

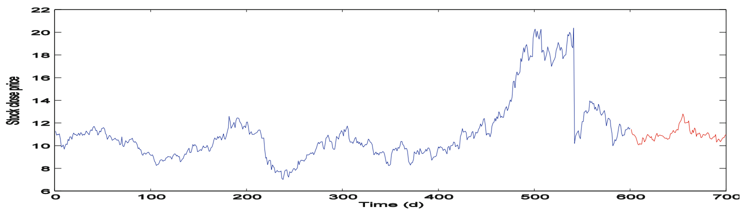
3.1 The General Prediction Model

The time series can be denoted as $\{s(t)\}_{t=1}^{\infty}$. For time series prediction task under certain conditions, Taken's Theorem [11] ensures that for some embedding dimension $d \in N^+$ and almost all time delay $\tau \in N^+$, there is a smooth function $f : R^d \rightarrow R$ so that $s(t) = f[s(t-d\tau), \ldots, s(t-2\tau), s(t-\tau)]$. Thus, a natural choice of the training dataset can be $\{x_t, y_t\}_{t=1}^{N}$ where $x_t = [s(t-d\tau), \ldots, s(t-2\tau), s(t-\tau)]$ and $y_t = s(t)$, and the test dataset $\{x_t^*, y_t^*\}_{t=1}^{L}$ can be set in the same way. In this way, time series prediction task can be transformed into the regression problem which aims at estimating and approximating the unknown function f.

We utilize Shanghai Composite Index (stock code: 000001) and Donghua energy (stock code: 002221) stock closing prices datasets from 2011 to 2013 which are downloaded from the Dazhihui software, and generate training datasets and test datasets which are respectively shown in the blue curve and red curve in Fig. 1.



(a)



(b)

Fig. 1. Shanghai and Donghua stock closing price curves from 2011 to 2013, blue curve represents 600 training data and red curve represents 100 test data. (a). Shanghai stock closing price curve. (b). Donghua stock closing price curve. (Color figure online)

For $\mathrm{d} = 1, 2, 3, 4$ and $\tau = 1, 2, 3, 4$, firstly we generate 700 samples, and every sample is a $\mathrm{d} + 1$ dimensions vector. The first d data are the input sample of our model and the last data is the output. Secondly, we normalize all the training and test outputs by $\mathrm{y} \rightarrow (\mathrm{y} - \mathrm{m})/\sigma$, where m and σ denote the mean and the standard variance of the training outputs, respectively. Again, 700 samples are divided two parts, including 600 training samples and 100 test samples.

3.2 Prediction Results and Comparisons

We implement the precise hard-cut EM algorithm for MGPs (referred to as PreHard-cut) on the training dataset, and verify its performance on the test dataset. Actually, we implement it on each of the 16 normalized training datasets, get the trained MGP model and make the prediction. We finally de-normalize the prediction by $\hat{y} \rightarrow \hat{y}\sigma + \mathrm{m}$. In order to compare its prediction performance, we run the MGP model with the other EM algorithms and some typical regression models and algorithms as follows:

(1) The LOOCV hard-cut EM algorithm (referred to as LOOCV) proposed in [12] for MGPs, which approximates the posteriors and the Q function via the leave-one-out cross validation mechanism;
(2) The variational hard-cut EM algorithm (referred to as VarHard-cut) proposed in [13] for MGPs, which approximates the posteriors via the variational inference;
(3) The Radial Basis Function neural network with Gaussian kernel function (referred to as RBF), the classical regression algorithm which makes prediction by linear combinations of radial basis functions.

The prediction accuracy is evaluated by the root mean squared error (RMSE) on each experiment, which is mathematically defined as follow

$$\mathrm{RMSE} = \sqrt{\frac{1}{L}\sum\nolimits_{t=1}^{L} (\hat{y}_t - \mathrm{y}_\mathrm{t})^2}, \tag{10}$$

where y_t and $\hat{y}_t$ denote the output true value of the t-th test sample and its predictive value, respectively. Meanwhile, we compare the efficiency of these algorithms by the total time consumed for both the parameter learning and the prediction, with an Intel(R) Core(TM) i5 CPU and 16.00 GB of RAM running Matlab R2014a source codes for all the experiments.

Before the parameter learning, some prior parameters have to be specified, including the number C of GP components for the MGP model, the number of pseudo inputs (PI) for the variational hard-cut EM algorithm and the number of neurons in the hidden layer (HL) for the RBF model. Without additional explanation, some typical values of these parameters are tested and these ones are selected and presented with the least prediction RMSEs.

The RMSEs as well as the best values of the predetermined parameters for each algorithm on each dataset are listed in Table 1. We find that in terms of prediction accuracy, the precise hard-cut EM algorithm rank the first in the dataset with

$d = 3, \tau = 1$, which demonstrates the advantage in Shanghai and Donghua stock closing price prediction. And the predictive results are better than the results using the generalized RBF neural network in paper [14, 15]. The variational hard-cut EM algorithm for MGP model is comparable with the precise hard-cut algorithm on accuracy. But the last one is more stable and uniformly optimal with $d = 3, \tau = 1$ on Shanghai and Donghua stock price prediction. The LOOCV hard-cut EM algorithm for MGP model and the RBF model are not qualified for stock price prediction. Besides, Table 1 also shows a general decrease trend on prediction RMSEs with the embedding dimension d, since a large d means more information in the inputs.

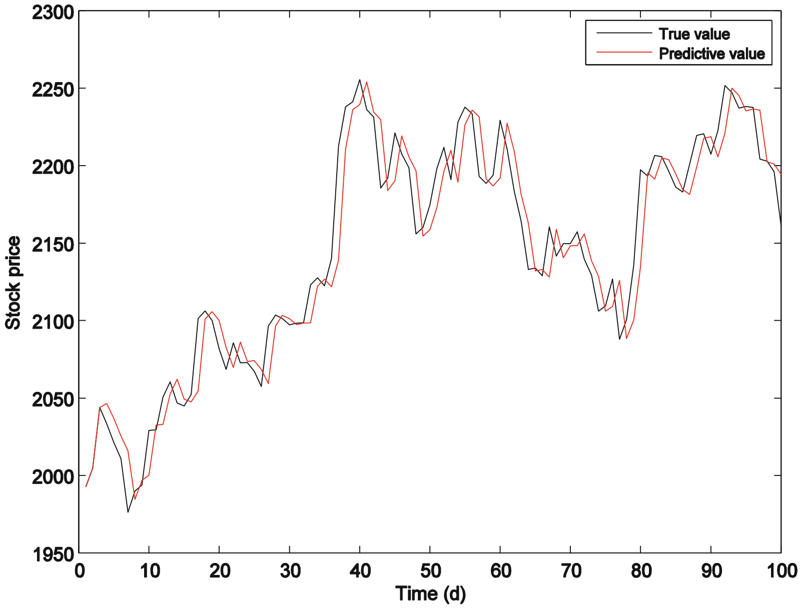
Table 1. The RMSEs for Shanghai and Donghua stock closing price prediction.

d	τ	PreHard-cut		LOOCV		VarHard-cut		RBF	
		Shanghai	Donghua	Shanghai	Donghua	Shanghai	Donghua	Shanghai	Donghua
1	1	21.1933	0.2224	**21.2151**	0.2218	21.2106	**0.2194**	**21.5791**	**0.2204**
1	2	31.8931	0.3174	32.5150	0.3258	31.5769	0.3134	33.4016	0.3188
1	3	38.8945	0.3713	40.4267	0.3864	39.1431	0.3654	41.3143	0.3784
1	4	43.9104	0.4144	45.9821	0.4207	43.8947	0.4030	47.9299	0.4159
2	1	21.2464	0.2192	21.4326	0.2209	21.2879	0.2195	21.8939	0.2216
2	2	32.0199	0.3133	32.8077	0.3428	31.8553	0.3146	33.5479	0.3232
2	3	39.3107	0.3737	41.6617	0.3918	38.0136	0.3556	47.1278	0.3621
2	4	43.7612	0.3940	45.8229	0.4418	43.7895	0.4016	53.1479	0.3975
3	1	**21.0782**	**0.2183**	21.4797	**0.2203**	**21.0879**	0.2206	22.9824	0.2272
3	2	32.0317	0.3192	33.1330	0.3547	31.5318	0.3159	36.3263	0.3126
3	3	39.1218	0.3448	42.0259	0.4089	38.5492	0.3545	43.5631	0.3579
3	4	44.9140	0.3901	52.6917	0.4585	43.7351	0.3726	57.8089	0.4096
4	1	21.1804	0.2225	21.8933	0.2212	21.1461	0.2219	22.9345	0.2341
4	2	33.3603	0.3125	33.3099	0.4030	32.1945	0.3104	38.2823	0.3131
4	3	39.5789	0.3695	43.0564	0.4681	39.1205	0.3586	54.2466	0.3705
4	4	45.9405	0.4067	58.1039	0.5645	45.3620	0.3960	72.0162	0.4536

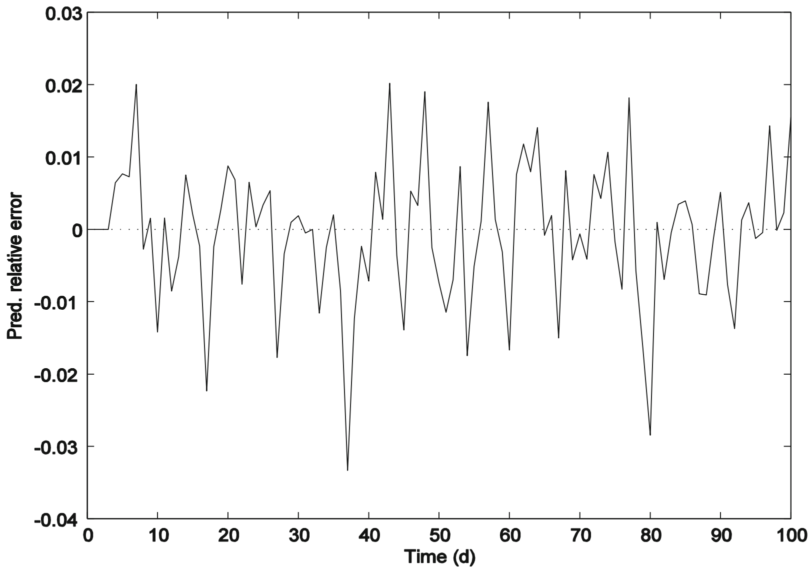
Moreover, the proposed technique has good scalability, but for stock price prediction 600 days stock closing price data are enough on the grounds that time span is up to two years!

Figures 2 and 3 show the best forecasting results with the parameters $d = 3$ and $\tau = 1$, which intuitively show the validity of the predictions. In Shanghai and Donghua test samples, the real and predicted values of the next 100 days are anastomotic and the best prediction RMSEs are 21.0782 and 0.2183 represented in bold in Table 1 respectively. The true values of the test samples are in good agreement with the predicted values, and the corresponding prediction errors are in actual allowable range which are mainly in ± 0.2 and ± 0.4 respectively as shown in Figs. 2 and 3.

The total time consumptions are shown in Table 2. We see that the precise hard-cut EM algorithm takes slightly longer. Nevertheless, no algorithms take longer than 6 min such that the remaining time is adequate for engineers to adjust the output power.

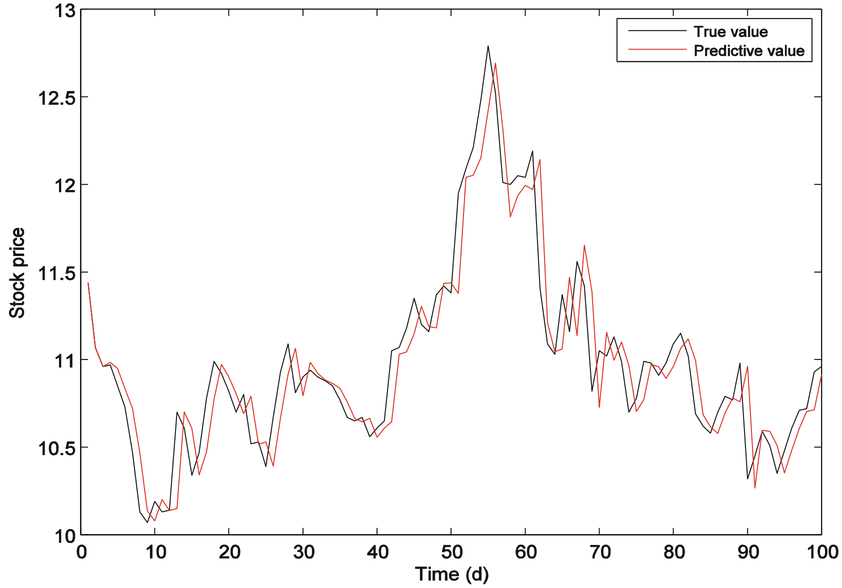


(a)

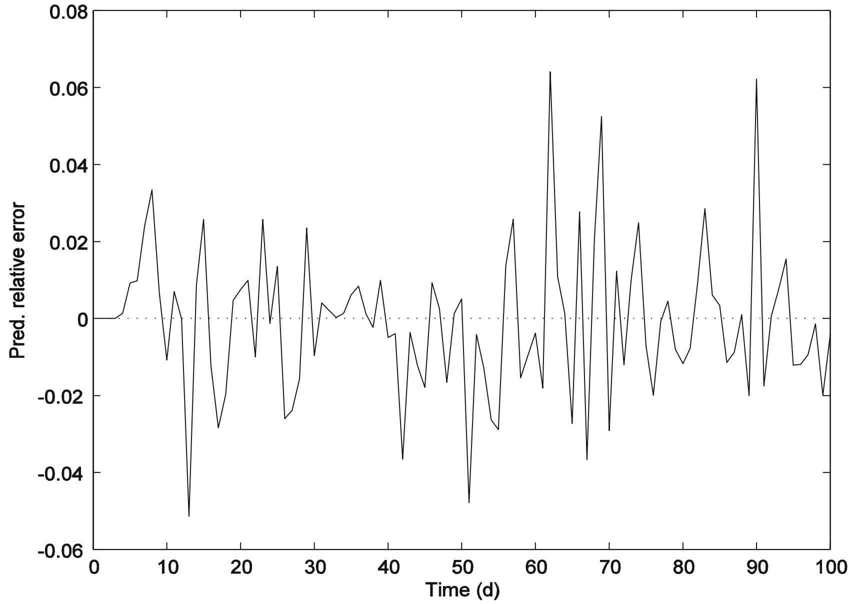


(b)

Fig. 2. (a). The prediction results of Shanghai stock closing price data; (b). The corresponding errors of Shanghai stock closing price data. (Color figure online)



(a)



(b)

Fig. 3. (a). The prediction results of Donghua stock closing price data; (b). The corresponding errors of Donghua stock closing price data. (Color figure online)

Table 2. The time consumptions for Shanghai and Donghua stock closing price prediction.

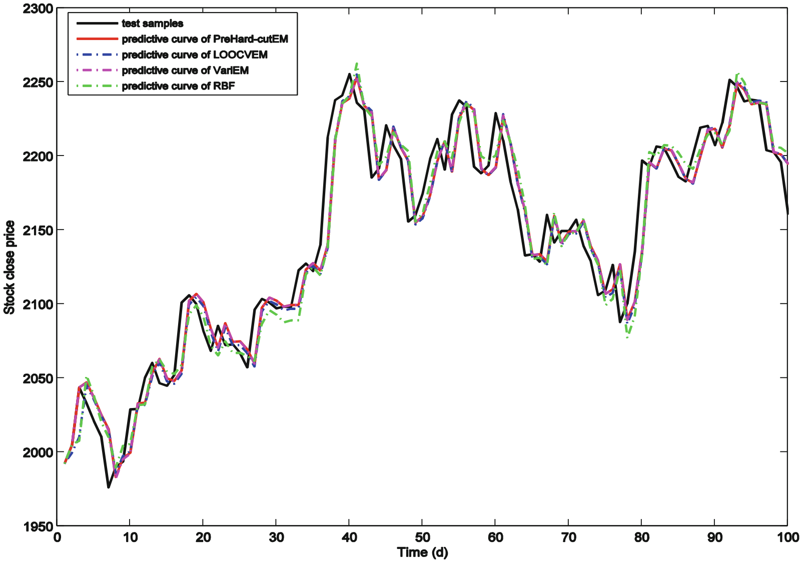
d	τ	PreHard-cut		LOOCV		VarHard-cut		RBF	
		Shanghai	Donghua	Shanghai	Donghua	Shanghai	Donghua	Shanghai	Donghua
1	1	59.5766	120.2869	77.7581	103.5980	84.8291	72.0887	6.6121	0.8491
1	2	77.6747	110.2947	66.0087	86.9751	60.2808	86.6801	1.9530	0.5538
1	3	89.4715	105.6588	56.4041	90.7678	71.2045	59.5331	1.3176	1.2327
1	4	110.5455	101.7635	54.8922	87.8987	82.5639	65.4218	0.3662	0.7490
2	1	72.1279	224.2375	88.2539	104.4790	97.8030	105.5201	0.7642	0.5208
2	2	93.3649	140.5822	65.2119	78.5423	122.9084	114.4137	0.7450	0.5049
2	3	112.4934	126.7071	54.1397	89.6782	101.0682	147.6291	1.2788	0.4943
2	4	152.2296	131.7796	55.6939	75.1938	101.6412	134.8176	0.3793	0.5290
3	1	176.9017	211.0905	70.8960	101.1628	129.4165	120.8225	0.7593	0.7829
3	2	115.6750	198.6486	59.2711	76.5386	127.8695	130.1543	1.2261	0.8615
3	3	115.2516	198.8352	64.5232	79.8847	126.0775	130.8770	0.7630	0.6358
3	4	130.5847	180.8102	53.0291	73.4977	117.0582	138.7150	0.3607	0.7125
4	1	247.9665	266.5358	75.6300	86.7739	153.3917	160.6503	1.2926	1.0299
4	2	146.0758	215.2569	57.9361	78.0230	155.2156	169.2637	0.7964	0.7965
4	3	81.0702	255.1925	53.7123	74.4203	155.8308	180.2085	1.2557	0.7503
4	4	129.1127	225.2805	51.6860	76.9758	149.6366	159.0454	1.8375	0.7804

Therefore, accuracy is the key factor in selecting the appropriate model and algorithm for stock price forecasting, so the precise hard-cut EM algorithm for the MGP model is a wonderful choice.

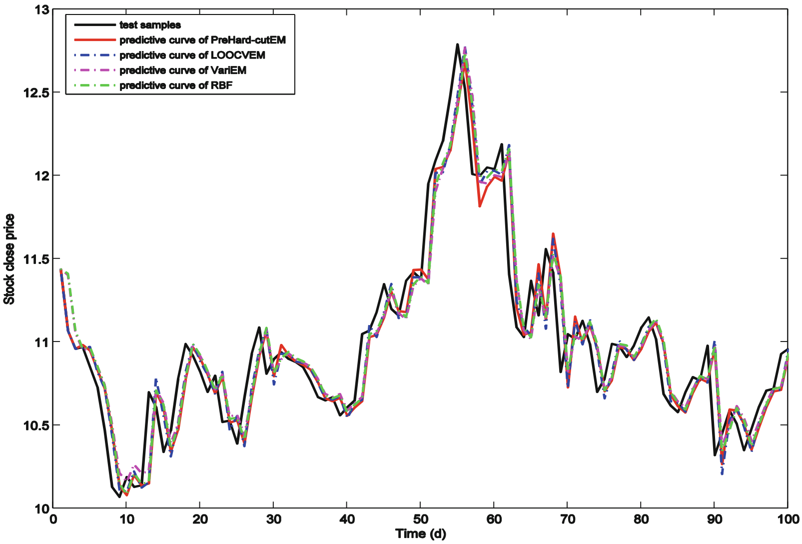
The best predictive curve for each algorithm is shown in Fig. 4. It can be found that the precise hard-cut EM algorithm and the variational hard-cut EM algorithm fit the true stock price extremely well except when the stock price reaches a peak or a trough, where there is a dramatic turn of the stock price. However, the two predictive curves are still within small and acceptable range around the true stock price even during the period of the peak and the trough. Besides, at some moments, the prediction of the precise hard-cut EM algorithm is closer to the true stock price than the variational hard-cut EM algorithm. The LOOCV hard-cut EM algorithm and the RBF model are not suitable for stock price forecasting.

Some remarkable results from Figs. 2, 3 and 4 is that the predicted prices seem to be displaced some constant time. Because the predicted price $s(t)$ is based on before d stock price: $s(t-d\tau), \ldots, s(t-2\tau), s(t-\tau)$.

In order to further explore how to improve the performance of the precise hard-cut EM algorithm, we plot the prediction RMSEs for $d = 1, 2, 3, 4, 5$ and $\tau = 1, 2, 3, 4$ respectively in Fig. 5. It can be observed from Fig. 5 that the RMSE generally decreases with the increasing of d and the decreasing of τ. When $d \geq 3$, the RMSE is considerably low and its variation with d and τ is very tiny. Therefore, an appropriate large embedding dimension d ensures a precise forecasting in stock price.

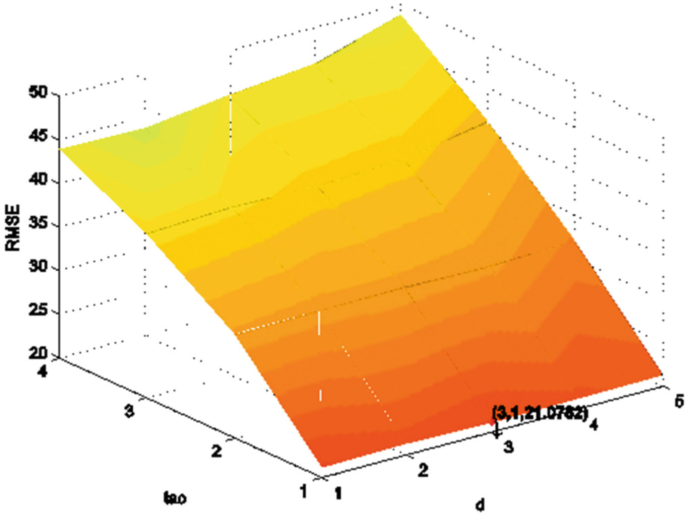


(a)

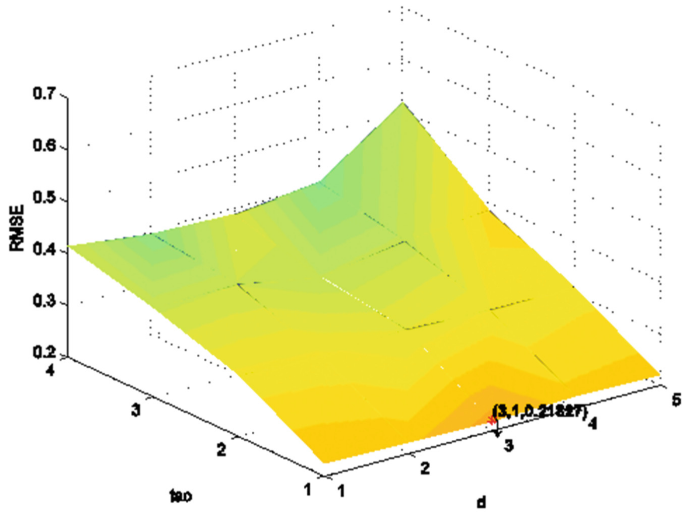


(b)

Fig. 4. (a). Comparisons of each algorithm for the predictive curves of Shanghai stock closing price in 100d test data. (b). Comparisons of each algorithm for the predictive curves of Donghua stock closing price in 100d test data. (Color figure online)



(a)



(b)

Fig. 5. (a). The predictive RMSEs for Shanghai stock closing price in 100d test data in the precise hard-cut EM algorithm with various values of d and τ. (b). The predictive RMSEs for Donghua stock closing price in 100d test data in the precise hard-cut EM algorithm with various values of d and τ.

4 Conclusion

We have successfully applied the MGP model via the precise hard-cut EM algorithm to modeling and predicting the time series of stock prices. The experiment results demonstrate that this MGP based method via the precise hard-cut EM algorithm turns out to be valid, feasible and highly competitive on prediction accuracy with acceptable time consumption, and outperforms some typical regression models and algorithms.

Acknowledgement. This work was supported by the Natural Science Foundation of China for Grant 61171138.

References

1. Sun, W., Guo, J., Xia, B.: Discussion about stock prediction theory based on RBF neural network. Heilongjiang Sci. Technol. Inf. (22), 130 (2010)
2. Fu, C., Fu, M., Que, J.: Prediction of stock price base on radial basic function neural networks. Technol. Dev. Enterp. 23(4), 14–15, 38 (2004)
3. Liu, H., Bai, Y.: Analysis of AR model and neural network for forecasting stock price. Math. Pract. Theory **41**(4), 14–19 (2011)
4. Zhu, Y.: Research on Stock Prediction Methods. Xi'an: Master Thesis of Northwestern Polytechnical University (2006)
5. Tresp, V.: Mixtures of Gaussian processes. In: Advances in Neural Information Processing Systems, vol. 13, pp. 654–660 (2000)
6. Meeds, E., Osindero, S.: An alternative infinite mixture of Gaussian process experts. In: Advances in Neural Information Processing Systems, 18, pp. 883–890 (2006)
7. Chen, Z., Ma, J., Zhou, Y.: A precise hard-cut EM algorithm for mixtures of Gaussian processes. In: Huang, D.-S., Jo, K.-H., Wang, L. (eds.) ICIC 2014. LNCS, vol. 8589, pp. 68–75. Springer, Heidelberg (2014)
8. Joseph, J., Doshi-Velez, F., Huang, A., Roy, N.: A Bayesian nonparametric approach to modeling motion patterns. Auton. Robots **31**(4), 383–400 (2011)
9. Sun, S.: Infinite mixtures of multivariate Gaussian processes. In: International Conference on IEEE Machine Learning and Cybernetics (ICMLC), vol. 3, pp. 1011–1016 (2013)
10. Tayal, A., Poupart, P., Li, Y.: Hierarchical double dirichlet process mixture of Gaussian processes. In: AAAI, pp. 1126–1133 (2012)
11. Zhou, Y., Zhang, T., Sun, J.: Multi-scale gaussian processes: a novel model for chaotic time series prediction. Chin. Phys. Lett. **24**(1), 42–45 (2007)
12. Yang, Y., Ma, J.: An efficient EM approach to parameter learning of the mixture of Gaussian processes. In: Liu, D., Zhang, H., Polycarpou, M., Alippi, C., He, H. (eds.) ISNN 2011, Part II. LNCS, vol. 6676, pp. 165–174. Springer, Heidelberg (2011)
13. Nguyen, T., Bonilla, E.: Fast allocation of Gaussian process experts. In: Proceedings of the 31st International Conference on Machine Learning, pp. 145–153 (2014)
14. Liu, S., Ma, J.: The application of diagonal generalized RBF neural network to stock price prediction. Highlights Sciencepaper Online **7**(13), 1296–1306 (2014)
15. Zheng, W., Ma, J.: Diagonal log-normal generalized RBF neural network for stock price prediction. In: Zeng, Z., Li, Y., King, I. (eds.) ISNN 2014. LNCS, vol. 8866, pp. 576–583. Springer, Heidelberg (2014)