Working together
www.rcis.ro

Revista de Cercetare si Interventie sociala

ISSN: 1583-3410 (print), ISSN: 1584-5397 (electronic)

A COMBINATION OF HIDDEN MARKOV MODEL AND ASSOCIATION ANALYSIS FOR STOCK MARKET SECTOR ROTATION

Jiao CHEN, Chi XIE, Zhijian ZENG

Revista de cercetare și intervenție socială, 2018, vol. 63, pp. 149-165

The online version of this article can be found at:
www.rcis.ro, *www.doaj.org* and *www.scopus.com*

Published by:
Expert Projects Publishing House

expert projects
publishing

On behalf of:
„Alexandru Ioan Cuza" University,
Department of Sociology and Social Work
and
HoltIS Association

REVISTA DE CERCETARE SI INTERVENTIE SOCIALA
is indexed by Clarivate Analytics (Web of Science) -
Social Sciences Citation Index
(Sociology and Social Work Domains)

A Combination of Hidden Markov Model and Association Analysis for Stock Market Sector Rotation

Jiao CHEN[1], Chi XIE[2], Zhijian ZENG[3]

Abstract

The use of Hidden Markov Model in stock market sector rotation is not investigated in the past. In this research, we consider an industry sector index portfolio based on the Shenwan first-class classification and propose state transition matrix for investment. In particular, we design an correlation analysis strategy that initialized state probability transition matrix Additionally, we design the observation state sequence which consisting of a series of stocks. Using Pearson's Correlation Coefficient to screen out the 10 stocks with the highest correlation in each industry sector. We put these parameters into the HMM and use the Baum-Welch algorithm to obtain the iterative solution results. Using the solved matrix into the back test program, the results show that the strategy returns well.

Keywords: HMM, association analysis, Pearson correlation coefficient, Baum-Welch algorithm, apriori.

[1] College of Business Administration, Hunan University, CHINA. E-mail: cjasd@163.com

[2] College of Business Administration, Hunan University, CHINA; Center of Finance and Investment Management, Hunan University, CHINA. E-mail: xiechi_hnu@163.com *(Corresponding author)*

[3] College of Business Administration, Hunan University, CHINA; Center of Finance and Investment Management, Hunan University, CHINA. E-mail: celia116@126.com

Introduction

There are extensive studies in many domains that deal with Hidden Markov Model (HMM). Early study included the use of HMM for pattern recognition. A speech recognition algorithm based on HMM is proposed by Zhang and Zhang (2011). Mohamed and Ramachandran Nair (2012) describe the development of a context independent, small vocabulary, connectionist-statistical continuous Malayalam speech recognition system. Liu and Chua (2010) present a new test to distinguish between meaningful and non-meaningful HMM-modeled activity patterns in human activity recognition systems. Park and Lee (2011) present a real-time 3D pointing gesture recognition algorithm for mobile robots, based on a cascade hidden Markov model (HMM) and a particle filter. Recently, a number of studies have applied HMM in biological gene engineering (Krogh *et al.,* 2001; Malekpour, Pezeshk, & Sadeghi, 2016; Nikdelfaz & Jalili, 2018). In addition to research on model applications, some studies have focused on model improvements. Zhu, Ye and Gao (2009) used POS algorithm to optimize the initial values of model parameters and proposed the APHMM model. Xu *et al.,* (2017) improved the model with the K-means algorithm to make the prediction results more accurate. Caccia and Remillard (2017) proposed a multivariate autoregressive HMM and found that the improved model has obvious advantages.

Due to the time series nature of the financial industry, it is required to be able to model in the time dimension, which leads to the widespread application of HMM. Srivastava *et al.* (2008) applied the model to credit fraud detection and results show that using HMM for information fraud detection is effective. Xu, Chen and Fu (2015) applied the model to the fuel futures market and conducted VAR risk measurement on the fuel futures market, portraying the volatility of China's fuel futures market. Thomas, Allen and Kingsbury (2002) linked the interest rate process to the credit risk process through HMM to study the relationship between them. What's more, Boyle and Draviam (2007) and Liew and Siu (2010) use the model to solve the option pricing problem by studying the relationship between the volatility of the customer's potential assets and the interest rate.

With the global economic crisis in 2008, especially since 2009, the US stock market has entered a bull market for nearly 10 years, and more and more institutions or individual investors use models to analyze the stock market. Researchers use the open, close, high, and low price of the daily stock to predict the stock price rise and fall the next day (Hassan & Nath, 2005; Park *et al.,* 2009; Gupta & Dhingra, 2012, Huang, 2015), applied HMM to stock price forecasting and gave a new way to predict stock prices. In the mainland stock market, some scholars use models to conduct research. In addition to using HMM alone, Hassan (2009) tried combine it with fuzzy models for better results.

HMM general form

The HMM is usually composed of the following five-tuple (S, O, Π, A, B) form:

1. $S=\{s_1,\ldots,s_N\}$ representing a state set, the elements in the state set have subscripts t representing moments.

2. $O=\{o_1,\ldots,o_M\}$ is the output observation state set. In the case of discrete observation densities, M is the number of observation state.

3. Initial state distribution $\Pi=\{\pi_i\}$, $i\in S$, where the π_i represents the probability value of the state i :

$$\pi_i = P(s_1 = i) \quad (2.1)$$

State transition distribution matrix $A = \{a_{ij}\}, i,j \in S$ $A = \{a_{ij}\}, i,j \in S$

$$a_{ij} = P(s_{t+1}|s_t), 1 \leq i,j \leq N \quad (2.2)$$

Observation state probability distribution matrix $B = b_j(o_t)$ $B = b_j(o_t)$ .where the probability function for each of these states j is:

$$b_j(o_t) = P(o_t|s_t = j) \quad (2.3)$$

After modeling the problem as HMM, assuming the model generates a set of data, we can calculate the probability of the observed sequence and the possible initial state sequence. We can also train model parameters based on observed data and obtain more accurate models, then use the trained models to predict the data.

Model building process and data acquisition

We choose 18 industries in the first-level industry except building decoration, electrical equipment, household electrical and mechanical equipment, and obtain index data from 2011 to September 2018, and we also obtain top 10 stock trading data of each industry within the corresponding time period.

Once we have an HMM, there are three problems of interest

1) The Evaluation Problem

Given an HMM and a sequence of observations $O = (o_1, o_2, \ldots, o_T)$, what is the probability that the observations are generated by the model, $P(O|\lambda)$?

2) Then Decoding Problem

Given a model λ and a sequence of observations $O = (o_1, o_2, \dots, o_T)$, what is the most likely state sequence in the model that produced the observations?

3) The Learning Problem

Given a model λ and a sequence of observations $O = (o_1, o_2, \dots, o_T)$, how should we adjust the model parameters $\{A, B, \pi\}$ in order to maximize $P(O|\lambda)$.

According discussed in the previous section, the modeling of this section belongs to the Learning Problem, that is, learning model parameters in the case of the observation state sequence determination. The main parameter we focused is the hidden state probability transfer matrix and use this probability transfer matrix to guide the direction of industry investment. Therefore, the appropriate set of observation states and the definition of the implicit state set are very important.

Statistical description and correlation analysis

The definition of hidden state

The industries we selected are 18 industries in the first-level industry except Building Decoration, Electrical Equipment, Textile and Garment, Household Electrical and Mechanical Equipment, namely Agriculture Animal Husbandry and Fishery(AAHF), Mining, Chemical Industry(CI), Non-Ferrous Metal(NFM), Food and Beverage(FB), Light Industry(LI), Medical Biology(MB), Transportation(Trans), Commercial Trade(CT), Leisure Services(LS), Comprehensive Industry(CPI), National Defense and Military Industry(NDMI), Computer, Media, Communication (Comm), Banking, Non-bank Finance(NF), Automobile (Auto). From the data in Figure 1, the industry's ups and downs synchronism are obvious, but the trend of trade-offs between industries is also obvious.

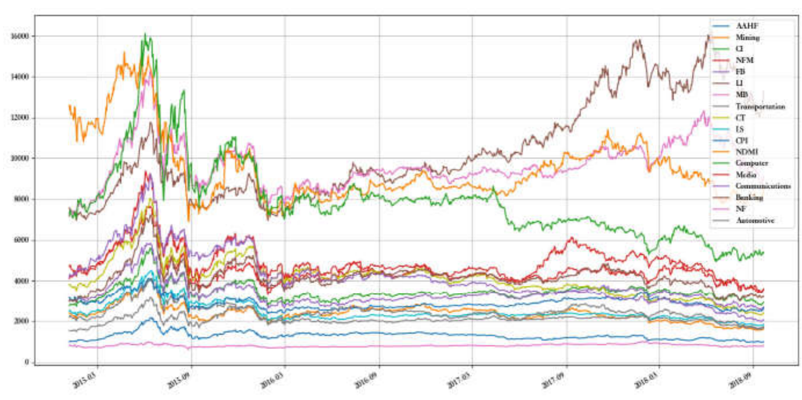

Figure 1. Industry index has risen and fallen in recent years

We define the hidden state set as

$$\mathbf{S} = \left\{ \begin{matrix} AAHF, Mining, CI, NFM, FB, LI, MB, Trans, CT, LS, CPI, \\ NDMI, Computer, Media, Comm, Banking, NF, Auto \end{matrix} \right\} \quad \mathbf{(4.1)}$$

The hidden state set contains 18 industry sectors such as Defense Military and Computer. We define the set of observation states to be the 10 most relevant stocks under the corresponding section. Pearson Correlation Coefficient (Pearson Correlation Coefficient) is used to compute the daily closing data of stocks on the 20 days, 60 days, 90 days, 180 days, 270 days, 365 days, 2 years, and 3 years with the corresponding interval index data of the selected industry, to get the industry sector with the highest average Pearson correlation expectation. The final sector and stock correlation distribution is shown in the *Table 1*.

Table 1. Stock and industry index Pearson correlation coefficient

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
AAHF	0	0	2	0	2	5	22	48	76	1
Mining	1	0	0	0	1	0	5	26	105	9
CI	0	0	0	0	0	0	1	7	43	3
NFM	0	0	0	0	0	5	5	41	89	1
FB	0	0	0	0	0	0	0	5	62	9
LI	0	0	0	0	0	2	11	36	48	1
MB	0	0	0	0	0	0	4	15	25	0
Trans	0	0	0	0	0	2	12	20	30	1
CT	0	0	0	0	0	0	1	19	106	17
LS	0	0	0	0	0	0	2	16	25	1
CPI	0	0	0	0	0	0	2	9	27	0
NDMI	0	0	2	0	0	2	6	11	16	7
Computer	0	0	0	0	1	0	3	21	28	3

Media	0	0	0	0	0	0	3	11	52	10
Comm	0	0	2	0	0	1	1	15	52	4
Banking	0	0	0	0	0	3	10	18	93	0
NF	0	0	0	0	0	1	3	6	5	1
Auto	0	4	3	6	10	24	56	76	66	1

From the distribution map, we find that the correlation coefficient between the listed company's stock and the corresponding industry sector index is generally high. It also shows that the stocks of Chinese listed companies are greatly affected by the industry's rotation effect. In addition, it is worth noting that among the ten stocks with the highest correlation coefficient among the industry sectors, not all of them are the stock classifications specified in the ShenWan first-level industry classification, as shown in the *Table 2*.

Table 2. Industry sector correlation top ten stocks main business areas

Banking	Comm	CI	CI	Media	Auto
NDMI	NF	NF	NF	NF	NF
Banking	CI	CT	Media	NFM	Banking
Comm	Media	MB	Trans	Comm	Comm
Media	Media	Banking	Media	Comm	Media
Computer	Auto	Computer	Computer	Computer	Computer
NDMI	NDMI	NDMI	NDMI	NDMI	NDMI
Mining	LI	Mining	Mining	NF	CPI
MB	Trans	MB	Computer	Auto	LS
Banking	NFM	LS	Auto	CI	CT
CI	Trans	Trans	Trans	Trans	Trans
Media	FB	MB	Trans	MB	MB
FB	CI	FB	LI	Trans	LI
CT	Trans	MB	Auto	CI	FB
NFM	AAHF	CI	NFM	NFM	NFM
LI	LI	Trans	CPI	MB	CI
CI	NFM	NFM	Auto	NFM	Mining
MB	CI	AAHF	Computer	AAHF	AAHF
5	4	3	2	1	

6	Media	NF	MB	Media	Banking	MB	MB	Mining	AAHF	MB	FB	MB	Banking	CT	NFM	LI	LS	NFM
7	NFM	Mining	CT	Banking	Auto	Computer	NDMI	Trans	CT	CI	Trans	MB	MB	CI	NFM	CT	NFM	CT
8	CI	NF	CI	Media	Media	Trans	NDMI	CI	CI	MB	Trans	CI	MB	AAHF	CI	FB	Mining	AAHF
9	Auto	NF	CI	MB	CI	Computer	NDMI	Trans	Mining	Banking	Mining	MB	FB	CI	NFM	CT	Computer	MB
10	Media	NF	FB	CPI	Media	Computer	NDMI	Mining	AAHF	Comm	Mining	LS	MB	AAHF	LI	Auto	CI	CT

It is easy to see that only the top ten stocks in the Financial industry (including Non-bank Financial industry) are consistent with the ShenWan first-class classification. In addition, the Defense Military industry, Trans, Computer, Real Estate industry and ShenWan's first-level industry classifications are more consistent, and other industries have different levels of mixed stocks. The reason for this phenomenon, we believe that on the one hand there are stocks that have excess returns relative to the industry sector, and there are also "market-making" behaviors of more institutional users in the secondary market. On the other hand, for a listed stock, its main business will adjust its capital supply direction according to different time, different market background and profitability. Relatively speaking, because of its industry specificity, the financial industry has very low possibilities for adjusting its main business direction. Secondly, with the "cross-border effect" of the Internet industry in recent years, various industries have cases of Internet transformation. This explains why the financial industry's strong correlation stocks are consistent with ShenWan's first-class classification, and there are more or less one or two computer-based listed company stocks in various industries.

In summary, we take the performance of the stock prices of the top 10 listed companies extracted by various industry sectors as the observation state set, and the observation sequence will use the performance of the weekly corresponding stocks as the elements of this observation sequence.

$$O = \{O_1, O_2, \dots O_{18}\} \quad (4.2)$$

The prior distribution of hidden state

Our goal is to estimate the probability transfer matrix between hidden states by using the performance of stock sets corresponding to the industry sector, that is, the probability distribution of rotation between industry sectors.

For such problems that cannot guarantee the properties of convex functions, in the process of using an algorithm like EM and gradient descent, the problem of local optimal solutions is inevitable. In view of the special problems of the industry sector rotation this paper, we propose that the initial distribution of the parameters of the algorithm startup timing can be supplemented according to some existing prior knowledge. In short, it is possible to analyze the possible rotation relationship between industry sectors, artificially adjust the search starting point of the parameter space. Therefore, we put aside the performance of stocks corresponding to the industry sector and use the correlation analysis technology to find the possible rotation probability distribution between industries in the process of the rise and fall of the industry sector index.

The association rules analyze the database through a specific rule algorithm, and mine the hidden and valuable associations between different data item sets in the database, so as to give the associated feature description of the data set. Its purpose is to help decision makers analyze historical data and the characteristics and laws of current data in order to build predictive models.

We define the relationship of association analysis rules

$$\mathbf{R: X \rightarrow Y \ (4.3)}$$

Where $\mathbf{X \subset I, Y \subset I}$ and $\mathbf{X \cap Y = \emptyset}$, representing the itemset X appearing in a trading period $\mathbf{T_i}$ $\mathbf{T_i}$, it is possible that the itemset Y also appears with probability P. The association analysis rules we care about are measured using the following two metrics: Degree of Support and Degree of Confidence.

For itemset X, we set $\mathbf{count(X \subseteq T)}$ to the total number of transaction periods containing X in transaction set D, and the support of itemset X

$$\mathbf{support(X) = \frac{count(X \subseteq T)}{|D|} \ (4.4)}$$

The support degree of the association analysis rule R represents the ratio of the transaction number of the itemset X and the itemset Y to the $|D|$ $|D|$ in the transaction set, that is

$$\mathbf{support(X \rightarrow Y) = \frac{count(X \cup Y)}{|D|} \ (4.5)}$$

The degree of support reflects the probability that itemset X and itemset Y appear simultaneously.

The confidence of the association analysis rule R is the ratio of the number of transactions including itemset X and itemset Y to the number of transactions containing itemset X, i.e.

$$\mathbf{confidence(X \to Y) = \frac{support(X \to Y)}{support(X)}} \quad (4.6)$$

Confidence reflects the probability that the itemset Y is included in the same period in the transaction cycle. In general, the higher the support and confidence, the stronger the correlation.

Combining the concept of transition probability matrices in the HMM above, we define the rotation probability transfer matrix between industry sectors as

$$\mathbf{A_{n\times n}} = f\left(\begin{bmatrix} a_{00} & \cdots & a_{0n} \\ \vdots & \ddots & \vdots \\ a_{n0} & \cdots & a_{nn} \end{bmatrix}\right) \quad (4.7)$$

$$a_{ij} = \epsilon \times confidence(i \to j) + (1-\epsilon) \times support(i \to j), \epsilon \in [0,1] \quad (4.8)$$

$$f = Softmax \quad (4.9)$$

Where n is equal to the total number of industry sectors, and $\mathbf{a_{ij}}$ represents the association rule score calculated by the i-th industry sector to the j-th industry sector through support degree and confidence degree, and the Softmax function will guarantee that the column vector of matrix A is equal in weight, and the sum is 1. The specific score is the summation weighted by the confidence degree and the support degree, because the support degree reflects the linkage property of the related item set in more cases, and the confidence degree more reflects the rotation property of the related item set, setting the parameter ε =0.8, which increases the weight of the confidence score.

The industry sector rotation model discussed belongs to the third type of HMM which is the known observation state sequence, which estimates the model parameters. The process includes constructing the hidden state and observation state and designing the state sequence. From the perspective of machine learning, the estimation process of model parameters is the training process of the model. Then we need to define the parameters that participate in the model training.

The set of observation state sequences is mentioned as

$$O = \{O_1, O_2, \dots, O_n\} \quad (4.10)$$

Where O_i represents the i-th observation state, we define the observation state as the sum of the cycle price performance of the top 10 stocks with the highest average correlation coefficient among various industry sectors. The average correlation coefficient is composed of the average of the Pearson Correlation Coefficients of the 20days, 60days, quarter, half year, three quarters, full year and two years corresponding to the daily closing price of the stock and industry sectors.

$$Period = \{20, 60, 90, 180, 270, 365, 730\} \quad (4.11)$$

$$Coef(i,j) = \frac{1}{n}\sum_{p}^{n} Pearson(i,j,p), p \in Period \quad (4.12)$$

$$O_i = Top(\{X \mid X \in Coef(i,j)\}, 10) \quad (4.13)$$

$Pearson(i,j,p)$ represents the Pearson Correlation Coefficient of the *i*-th industry index and stock j in p day, and $Coef(i,j)$ represents the average coefficient value after summing up the all cycles.

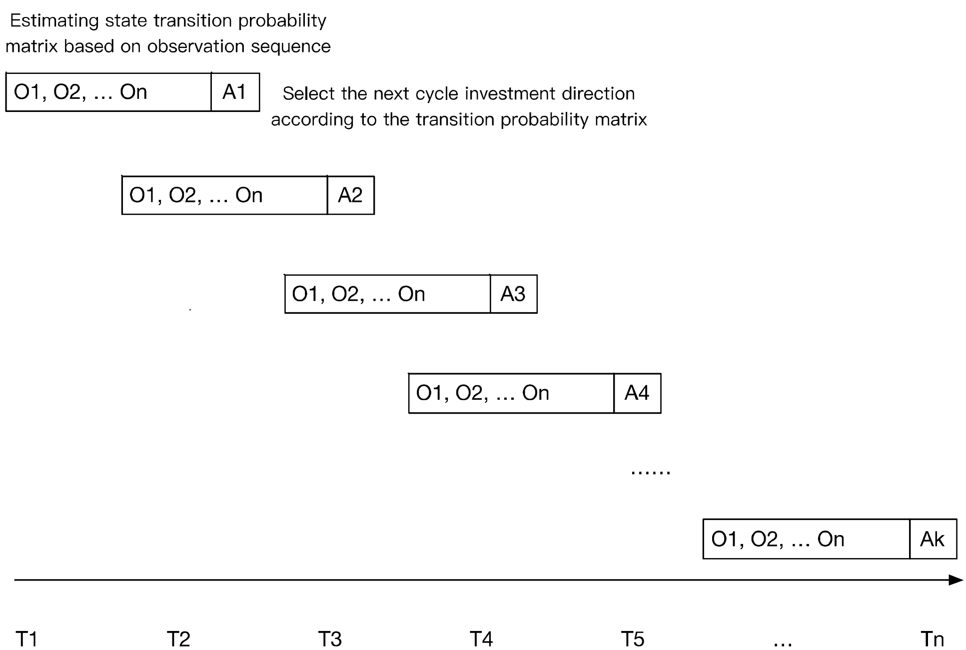

Figure 2. Sliding window iteration parameter flow

After obtaining the definition of the observation state sequence set, we will use the sliding time window form shown in Figure 2, periodically iteratively obtain the parameters of the model, and use the obtained state probability transfer matrix to guide our industry investment decision, and finally verify the validity of the model in the back test.

Market data experiments and results

The data comes from the China A-share daytime data from Wind Database from 2011 to 2018. We have extracted 1,696 stocks of listed companies that have traded since 2011, of which 860 are from the Shanghai Stock Exchange and 836 from the Shenzhen Stock Exchange (including 800 small and medium-sized board stocks and 36 GEM stocks). Industry sector index data includes AAHF, Mining, CI, NFM, FB, LI, MB, Trans, CT, LS, CPI, NDMI, Computer, Media, Comm, Banking, NF, Auto, totally 18 industries, from 2011 until now, close price daily data.

The observation state definition is the sum of the cycle price performance of the top 10 stocks with the highest average correlation coefficient among various industry sectors. The *Table 3* shows the average of the top 10 stocks of various industry sectors and their Pearson Correlation Coefficients.

Except for the banking industry sector, the correlation coefficient of representative stocks of observational states selected by other industries exceeds 0.9, which shows that they are basically representative of the industry. In addition, we use O_1 O_1 -O_{18} O_{18} to replace the observation state from top to bottom. We aggregate the daytime data of the stock into 5-day cycle data by accumulating the daily closing price, and obtain the best-performing observation state sequence every week according to the cycle data.

In order to calculate the a priori hidden state probability transfer matrix, we need to first generate the support degree and confidence degree between all industry sectors through the frequent itemset mining algorithm, namely Apriori.

Table 3. Industry sector correlation top 10 stocks and correlation coefficient mean

0.92	002054	002206	000988	002222	002087	300036	002042	300014	600360	600884	Auto
0.87	601988	601998	601088	000402	601328	601318	601166	601398	601939	600015	NF
0.94	600527	000936	600738	600796	600498	000408	600854	600865	600728	002160	Banking
0.95	600601	600825	300007	600845	002073	000727	002184	002022	600180	600037	Comm
0.96	600635	001696	600289	002177	600571	002017	002232	600580	600088	002154	Media
0.95	600789	600343	600268	600686	002293	600835	002163	600118	600862	600893	Computer
0.95	600030	000750	000686	000783	000623	601788	601099	600999	600837	000776	NDMI
0.92	000898	601186	600031	600546	000552	000039	000157	600808	600089	000001	CPI
0.94	000767	601005	600317	600572	601989	600307	600780	600675	600653	600642	LS
0.96	002274	600488	600243	600776	600206	000428	002328	600299	002144	000682	CT
0.93	600020	600674	600863	000027	601872	600396	600022	600284	600798	600221	Trans
0.92	000915	002230	600612	000042	600215	002262	600848	600350	600064	600993	MB
0.92	000661	600260	002050	600308	600426	002008	002078	600183	600887	600004	LI
0.96	002221	600415	002251	600881	002225	600033	002275	002129	000572	002064	FB
0.94	000789	600550	000537	600456	601168	000735	600596	600362	002218	601958	NFM
0.95	002083	600239	002135	002060	600616	600270	600704	002327	600329	600586	CI
0.96	002297	600397	000758	601007	000521	000589	600311	000060	600327	000751	Mining
0.94	002099	600310	002326	600257	600236	002055	600261	600353	002023	002100	AAHF
Mean of Coefficient	10	9	8	7	6	5	4	3	2	1	

We use the ups and downs of the daily index of the industry sector as a sign of whether the itemset, and also aggregated into a 5-day cycle up and down according to the cumulative form of the 5 days. The industry rotation probability matrix calculated is shown in Figure 3. This probability transfer matrix will be put into the Baum-Welch algorithm as a priori probability transfer matrix parameter.

The parameter estimation of the hidden state transition probability matrix is estimated using the Baum-Welch algorithm, and the derivation process of the algorithm is not described here. Briefly describe the pseudo code as Figure 4.

As shown in the sliding window iterative parameter estimation graph in Figure 2, we accumulate the 1881 trading days since 2011 and aggregate them into 5 trading days, totally 564 trading cycles, and finally divide them into 12 algorithm iteration cycles, and 47 observation states of the cycle are iterated. The estimation of the industry rotation probability transfer matrix is performed every iteration cycle, and the estimated value is used to adjust the sector position. Each adjustment is based on the industry with the highest probability of the next rotation in the previous state $\max_{s_{t+1}} P(s_{t+1}|s_i)$. The back test effect is shown in *Figure 5*.

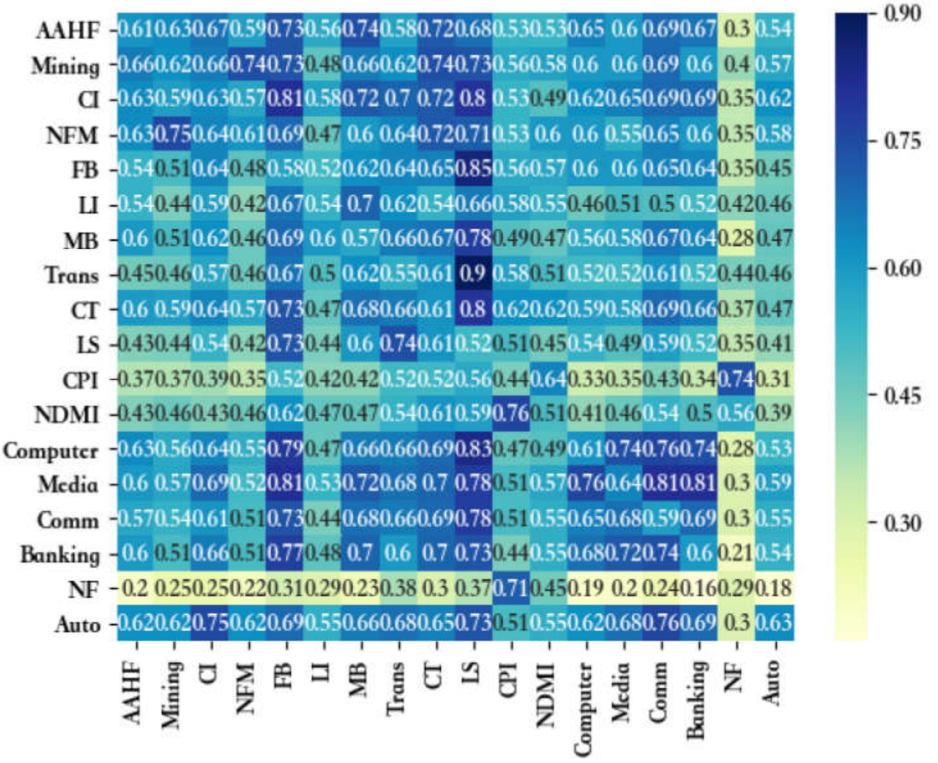

Figure 3. Industry rotation probability matrix

Input: Observation State Sequence $D = \{O_1, O_2, \dots, O_D\}$

Output: HMM $a_{ij}, b_j(k), \pi$

1). Initalize all $\pi_i, a_{ij}, b_j(k)$

2). loop Observation State Sequence D

3).Forward-Backward algorithm update $\gamma_t^{(d)}(i), \xi_t^{(d)}(i,j)$

4). Update $\pi_i, a_{ij}, b_j(k)$

$$\pi_i = \frac{\sum_{d=1}^{D} \gamma_1^{(d)}(i)}{D}$$

$$a_{ij} = \frac{\sum_{d=1}^{D} \sum_{t=1}^{T-1} \xi_t^d(i,j)}{\sum_{d=1}^{D} \sum_{t=1}^{T-1} \gamma_t^d(i)}$$

$$b_j(k) = \frac{\sum_{t=1}^{T} 1_{y_t=v_k} \gamma_t(t)}{\sum_{t=1}^{T} \gamma_t(t)}$$

5). If $\pi_i, a_{ij}, b_j(k)$ convergence, Terminate, else goto step 2)

Figure 4. Baum-Welch algorithm pseudo-code process

Figure 5. Back test results

Table 4. Back test statistics

Statistical indicators	Value
Annualized rate of return	11.96%
Alpha	0.05
Beta	0.394
Shape rate	0.267
Maximum withdrawal	53.759%

From the back test results in *Table 4* , the rotation combination is better to capture the bull market in the middle of 2015 years, but it does not avoid the bear market afterwards. In the relatively stable market environment afterwards, the rotation combination yield is better.

Conclusion

We use the correlation analysis to calculate the prior probability distribution of the HMM state transition matrix, and propose the use of HMM to analyze the industry sector rotation. We define the industry sector as a hidden state node and a strong correlation stock as an observation state node and designed a moving time window algorithm to iteratively generate a state probability transfer matrix. The industries we selected are 18 industries in the first-level industry and use the data from the 2011 to 2018 Chinese stock market to test. In summary, from the beginning of 2011, about 12% of annualized income is slightly insufficient, and it is inevitably affected by the 2018 Sino-US trade war.

Acknowledgements

(1) National Natural Science Foundation of China; Evolution Behaviors of the complex financial network dynamic and the study on crisis contagion control; NO.71373072

(2) National Natural Science Foundation of China; On risk overflow in the stock market based on the time-frequency analysis; NO.71340014

References

Boyle, P. & Draviam, T. (2007). Pricing exotic options under regime switching. *Insurance: Mathematics and Economics*, *40*(2), 267-282.

Caccia, M. & Remillard, B. (2017). Option Pricing and Hedging for Discrete Time Autoregressive Hidden Markov Model. *Les Cahiers du GERAD*, *60,* 1-33.

Gupta, A. & Dhingra, B. (2012). Stock market prediction using Hidden Markov models. Paper presented at *Proceedings of the Student Conference on Engg and Systems (SCES)*, Allahabad, Uttar Pradesh, India, 1-4.

Hassan, M.R. & Nath, B. (2005). Stock Market Forecasting Using Hidden Markov Models: A New Approach. *IEEE fifth International Conference on Intelligent Systems Design and Applications*, 192-196.

Hassan, M.R. (2009). A Combination of Hidden Markov Model and Fuzzy Model for Stock Market Forecasting. *Neurocomputing*, *72*(16-18), 3439-3446.

Huang, R. (2015). Stock Price Forecasting Analysis Based on Hidden Markov Model (HMM). Qing Dao: Qing Dao University.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomesll. *Journal of Molecular Biology*, *305*(3), 567-580.

Liew, C. & Siu, T.K. (2010). Martingale Representation and Admissible Portfolio Process with Regime Switching. *Stochastic Analysis and Applications*, *29*(1), 106-120.

Liu, X.H. & Chua, C.S. (2010). Rejection of non-meaningful activities for HMM-based activity recognition system. *Image and Vision Computing*, *28*(6), 865-871.

Malekpour, S.A., Pezeshk, H., & Sadeghi, M. (2016). MGP-HMM: Detecting genome-wide CNVs using an HMM for modeling mate pair insertion sizes and read counts. *Mathematical Biosciences*, *279*, 53-62.

Mohamed, A., & Ramachandran Nair, K. N. (2012). HMM/ANN hybrid model for continuous Malayalam speech recognition. *Procedia Engineering*, *30*, 616-622.

Nikdelfaz, O. & Jalili, S. (2018). Disease genes prediction by HMM based PU-learning using gene expression profiles. *Journal of Biomedical Informatics*, *81*, 102-111.

Park, B.S., Song, D.H., Kim, H.M., Choi, B.S., Lee, H., & Lee, J.O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature, 458*(7242): 1191-1195

Park, C.B. & Lee, S.W. (2011). Real-time 3D pointing gesture recognition for mobile robots with cascade HMM and particle filter. *Image and Vision Computing*, *29*(1), 51-63.

Srivastava, A., Kundu, A., Sural, S., & Majumdar, A. (2008). Credit Card Fraud Detection Using Hidden Markov Model. *IEEE Transactions on Dependable & Secure Computing*, *5*(1), 37-48.

Thomas, L.C., Allen, D.E., & Markel-Kingsbury, N. (2002). A hidden Markov chain model for the term structure of bond credit risk spreads. *International Review of Financial Analysis*, *11*(3), 311-329.
Xu, K., Chen, N., & Fu, Q.W. (2015). VAR Risk Measurement of Chinese Fuel Futures Market Based on HMM-GJR. *Friends of Accounting*, *16*, 27-30.
Xu, Z.J., Xie, R., Liu, J., & Mei, Y. (2017). Improved Hidden Markov Model and Its Application in Financial Forecasting. *Chinese Journal of Engineering Mathematics*, *34*(5), 470-480.
Zhang, J., & Zhang, M.A. (2011). Speech Recognition System Based Improved Algorithm of Dual-template HMM. *Procedia Engineering*, *15*, 2286-2290.
Zhu, J.Y., Ye, H.Y., & Gao, Y. (2009). Stock Price Forecast Combination Model Based on Hidden Markov Model. *Computer Engineering and Design*, *30*(21), 1945-4948.