2000

A Data Mining for the Effect of Cognitive Style, Subjective Emotion, and Physiological Phenomena on the Accuracy of Judgmental Time-Series Forecasting

Byoungho Song
Sangmyung University, bhsong@pine.sangmyung.ac.kr

Hung Kook Park
Sangmyung University, parkh@pine.sangmyung.ac.kr

Hyeon Joon Yoo
Sangmyung University, hjyoo@smuc.sangmyung.ac.kr

A Data Mining for the Effect of Cognitive Style, Subjective Emotion, and Physiological Phenomena on the Accuracy of Judgmental Time-Series Forecasting

Byoungho Song
Department of Software Science, Sangmyung University, Korea
bhsong@pine.sangmyung.ac.kr

Hung Kook Park
Division of Information and Telecommunications, Sangmyung University, Korea
parkh@pine.sangmyung.ac.kr

Hyeon-Joong Yoo
Division of Information and Telecommunications, Sangmyung University, Korea
hjyoo@smuc.sangmyung.ac.kr

Abstract

Data mining is finding hidden rules in given dataset using non-traditional methods. The objective is to discover some useful tendency or patterns from the given collection of data. We had mined the rules representing the effect of cognitive style, subjective emotion, and physiological phenomena on the accuracy of subjects' judgmental time-series forecasting. Then we have tried to find out any consistent tendencies in the frequent rules.

Subjects in Analytic-style show more accurate forecasting. Subjects in relaxed mode show more accurate forecasting. And Subjects' left EEG and beta rhythm seem to have a significant effect on their forecasting accuracy. But additional data mining to the other effects should be made.

Category: Technical

Keywords: Data Mining, Emotion Engineering, Judgemental Forecasting

Introduction

The activities of judgmental time-series forecasting can be easily found in our real life, such as estimating the movement of stock exchange indexes, weather forecasting, and so on. Our study team has made so many experiments to analyze the effect of decision-maker's cognitive style, subjective emotion, and physiological measurements (electroencephalograms) on the result of intuitive time-series forecasting. Unfortunately, this experimental measurements itself could not easily reveal the effect or meaningful relationship between them without some new non-traditional approach. And we decided to apply data mining (Agrawal, Imielinski, and Swami, 1993; Barson and Smith, 1997; Chen, Han, and Yu, 1996; Frawley, Shapiro, and Matheus, 1991), a new approach to find something meaningful and hidden from the given collection of data, to our experimental measurements.

Section 2 presents the experimental measurements as the dataset to be mined. Section 3 describes how we prepared the data mining. Section 4 presents our observations, and section 5 discusses the mining results and further research issues.

Experimental Measurements

Subjects

The test was done in two times. At first, 29 students were used to get data about each cognitive style, physiological phenomena, and the accuracy of their forecasting results. And to collect more data for better mining result, additional 48 students were used to get data about each cognitive style, subjective emotion, physiological phenomena, and the accuracy of their forecasting results. Thus, the total number of subjects (and records) is 77. Only the number of records of subjective emotion is 48.

Cognitive Style

According to the popular myers-Briggs Type Indicator (MBTI) test (Marakas, 1999; Myers, 1962), we classify all the subjects' cognitive style into 4 categories: Analytic (A), Behavioral (B), Conceptual (C), and Directive (D). The number of subjects in A was 25 (32.5%), the number of B was 17 (22.0%), the number of C was 23 (29.9%), and the number of D was 12 (15.6%).

Subjective Emotion

The test on 48-students' subjective emotion was based on two characteristics: positive or negative, and stressed or relaxed. There are 4 possible combinations: positive-stressed, positive-relaxed, negative-stressed, and negative-relaxed. We measured each of these 4 index values two times each: before the forecasting begins, and after the forecasting was done. Thus each subject's subjective emotion is represented with 8 values:

- positive-stressed-before-forecasting,
- positive-stressed-after-forecasting,
- positive-relaxed-before-forecasting,
- positive-relaxed-after-forecasting,
- negative-stressed-before-forecasting,
- negative-stressed-after-forecasting,
- negative-relaxed-before-forecasting, and
- negative-relaxed-after-forecasting.

Physiological Phenomena

We measured each subject's electroencephalogram (EEG) values as their physiological phenomena before and during their forecasting. Alpha/beta rhythms in the region of left/right brain lobe were measured. So we could get up 8 values per one subject:

- left-alpha-before-forecasting,
- left-beta-before-forecasting,
- right-alpha-before-forecasting,
- right-beta-before-forecasting,
- left-alpha-during-forecasting,
- left-beta-during-forecasting,
- right-alpha-during-forecasting, and
- right-beta-during-forecasting.

Forecasting Accuracy

We use MAPE (Mean Absolute Percentage Error) to evaluate the accuracy of subjects' forecasting results. The range of possible MAPE values is 0 to 1. The lower the MAPE value is, the better the accuracy should be.

Preparation for Data Mining

For each subject, we have 1 style value (A, B, C, or D), 8 (original) subjective emotion values (in numeric), 8 (original) EEC values (in numeric), and 1 MAPE value (in numeric). In addition, we made some derived (statistical) values (in numeric) from original emotion data and EEG data for more macroscopic analysis:

- The increasing ratio of positive-stressed values between before- and after-forecasting,
- The increasing ratio of positive-relaxed values,
- The increasing ratio of negative-stressed values,
- The increasing ratio of negative-relaxed values,
- The increasing ratio of left-alpha EEG values between before- and during forecasting,
- The increasing ratio of left-beta EEG values,
- The increasing ratio of right-alpha EEG values,
- The increasing ratio of right-beta EEG values,
- Total volume of left EEG before forecasting,
- Total volume of left EEG during forecasting,
- The increasing ratio of left EEG,
- Total volume of right EEG before forecasting,
- Total volume of right EEG during forecasting,
- The increasing ratio of right EEG,
- Total volume of alpha EEG before forecasting,
- Total volume of alpha EEG during forecasting,
- The increasing ratio of alpha EEG,
- Total volume of beta EEG before forecasting,
- Total volume of beta EEG during forecasting,
- The increasing ratio of beta EEG,
- Total volume of EEG before forecasting,
- Total volume of EEG during forecasting, and
- The increasing ratio of EEG.

Data mining requires partitioning every continuous (numeric) value range into several zones. We partitioned all the numeric properties into three levels: *high*, *low*, and *middle*. Highest 30% was assigned to 'high'; Lowest 30% was assigned to 'low'; and the rest 40% was assigned to 'middle'. Thus 23 of 77 MAPE values were 'high', 31 of 77 MAPE values were 'middle', and 23 of 77 MAPE values were said to be 'low'. The same values at any boundary were considered to be 'middle'.

The Mining Results

So far, we have found so many rules (relationships) between arbitrary pair of properties. Infrequent rules were removed and we tried to find out any consistent tendencies in the rest frequent rules. The rest of this section consists of our observations. The portions of high accuracy (low MAPE), middle accuracy (middle MAPE), and low accuracy (high MAPE) will be written in this order in '(' and ')' at the end of any tendencies or rules.

Observations on the Effect of Cognitive Style

Observation 1: The subjects in style A had a tendency to make high accurate (low MAPE) forecasting *(10/25, 9/25, 6/25).*

Observation 2: The subjects in style B had a tendency to make low accurate forecasting *(3/17, 8/17, 6/17)*.

And we can't find any meaningful tendencies in style C *(7/23, 9/23, 7/*23) and style D *(3/12, 5/12, 4/12).*

Observations on the Effect of Subjective Emotion

Observation 3: There was a tendency that regardless of positive or negative emotion, the higher relaxed level the subject shows at before-forecasting, the higher accuracy (s)he achieves, and when the lower relaxed level is shown, the lower accuracy is achieved. The evidence is:

- negative-relaxed-before-forecasting (*low*) *-> (4/11, 1/11, 6/11)*
- positive-relaxed-before-forecasting (*mid*) *-> (5/22, 9/22, 8/22)*
- positive-relaxed-before-forecasting (*high*) *-> (6/12, 3/12, 3/12)*

Observation 4: In contrast, there was a tendency that regardless of positive or negative emotion, the higher stressed level the subject shows at before-forecasting, the lower accuracy (s)he achieves, and when the lower stressed level is shown, the higher accuracy is achieved. The evidence is:

- negative-stressed-before-forecasting (*high*) *-> (3/12, 4/12, 5/12)*
- positive-stressed-before-forecasting (*low*) *-> (5/11, 4/11, 2/11)*
- positive-stressed-before-forecasting (*high*) *-> (2/10, 3/10, 5/10)*

Observation 5: There was a tendency that regardless of stressed or relaxed, the higher positive level the subject shows at after-forecasting, the higher accuracy (s)he achieves, and when the lower positive level is shown, the lower accuracy is achieved. The evidence is:

- positive-stressed-after-forecasting (*low*) *-> (3/15, 7/15, 5/15)*
- positive-stressed-after-forecasting (*mid*) *-> (7/18, 7/18, 4/18)*
- positive-relaxed-after-forecasting (*low*) *-> (3/16, 6/16, 7/16)*
- positive-relaxed-after-forecasting (*mid*) *-> (7/17, 8/17, 2/17)*

Observation 6: And, there was a tendency that regardless of stressed or relaxed, when the subject shows high or low negative level at after-forecasting, (s)he would achieved high or low accuracy, and when the level is middle, the accuracy is middle, too. The evidence is:

- negative-stressed-after-forecasting (*low*) *-> (5/13, 3/13, 5/13)*
- negative-stressed-after-forecasting (*high*) *-> (3/12, 6/12, 3/12)*
- negative-relaxed-after-forecasting (*low*) *-> (5/13, 2/13, 6/13)*
- negative-relaxed-after-forecasting (*high*) *-> (6/16, 4/16, 6/16)*
- negative-relaxed-after-forecasting (*mid*) *-> (4/19, 12/19, 3/19)*

The tendencies in the Observation 5 and 6 may come from the fact that when a subject is confident of his(her) decision just before, (s)he feels happy; and when (s)he is diffident of his(her) decision just before, (s)he feels unhappy.

Observations on the Effect of Physiological Phenomena

Observation 7: There was a slight tendency that the accuracy is low when the total volume of EEG at before-forecasting is low *(7/23, 7/23, 9/23).*

But there was no evidence that the accuracy is high when the total volume of EEG at before-forecasting is high *(6/23, 11/23, 6/23).*

Observation 8: There was a slight tendency that the accuracy is low when the total volume of EEG at during-forecasting is low (6/23, 8/23, 9/23), and the accuracy is low when the total volume of EEG at before-forecasting is low *(7/23, 11/23, 5/23).*

From the Observation 7 and 8, we can't find any distinct tendency between total volume of EEG and the accuracy.

Observation 9: There was a tendency that the accuracy is low when the total beta volume at before-forecasting is low *(7/23, 7/23, 9/23)*, and the accuracy is high when the total beta volume at before-forecasting is middle *(11/31, 12/31, 8/31)*.

But we can't find the effect of 'high' total beta volume at before-forecasting on the accuracy *(6/23, 11/23, 6/23)*.

Observation 10: There was a tendency that the accuracy is high when the increasing ratio of beta is low (i.e., when beta decreases) *(7/23, 11/23, 5/23)* or high (i.e., when beta increase significantly) *(9/23, 7/23, 7/23)*, and the accuracy is low when the increasing ratio of beta is middle (i.e., when beta is relatively invariant) *(7/31, 13/31, 11/31)*.

From the Observation 9 and 10, we can find the fact that the effect of beta rhythm on the accuracy is true.

Observation 11: There was a tendency that the accuracy is low when the left EEG (alpha + beta) volume at during-forecasting is low *(6/23, 8/23, 9/23)*, and the accuracy is high when the left EEG volume at during-forecasting is middle *(11/31, 12/31, 8/31)*.

But we can't find the effect of 'high' total EEG volume from the left-brain at during-forecasting on the accuracy *(6/23, 11/23, 6/23)*.

Observation 12: There was a tendency that the accuracy is high when the increasing ratio of left EEG is low (i.e., when the left EEG decreases) *(9/23, 8/23, 6/23)* or high (i.e., when the left EEG increase significantly) *(8/23, 10/23, 5/23)*, and the accuracy is low when the increasing ratio of left EEG is middle (i.e., when the left EEG is relatively invariant) *(6/31, 13/31, 12/31)*.

From the Observation 11 and 12, we can find the fact that the effect of EEG from the left-brain lobe on the accuracy is true.

Concluding Remarks

Subjects in style A (Analytic) seem to be more accurate, and subjects in style B (Behavioral) seem to be less accurate. It means that if we hire analysts in style A, we would have more opportunity to be happy.

Subjects in relaxed mode seem to be more accurate. It means that if we make our analysts relaxed, we would have more opportunity to be happy.

Subjects' left EEG and beta rhythm seem to have a significant effect on their forecasting accuracy. It means that if we can control our analysts' left EEG and beta signal, we would have more opportunity to be happy.

The effect of subjective emotion and physiological phenomena on the accuracy of style A (or B) subject's forecasting, the effect of cognitive style and physiological phenomena on the accuracy of relaxed (or stressed) subject's forecasting, and the effect of cognitive style and subjective emotion on the control over left EEG and beta signal are much worth analyzing in depth. Moreover, the effect of cognitive style on subjective emotion, the effect of cognitive style on EEG, the effect of subjective emotion on EEG, the effect of EEG on subjective emotion, and accurate subjects' dominant cognitive style, dominant subjective emotion, and dominant EEG signal should be analyzed as well.

Acknowledgements

This research was supported by 1999 research funds from Korea Research Foundation (KRF-99-042-C00141).

References

R. Agrawal, T. Imielinski, and A. Swami, "Database Mining: A Performance Perspective," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, No. 6, Dec. 1993.

A. Berson and S. J. Smith, "Data Warehousing, Data Mining, and OLAP," McGraw-Hill Pub., 1997.

M.-S. Chen, J. Han, and P. S. Yu, "Data Mining: An Overview from a Database Perspective," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883, Dec. 1996.

W. J. Frawley, G. piatetsky-Shapiro, and C. J. Matheus, "Knowledge Discovery in Databases: An Overview," Knowledge Discovery in Databases, AAAI/MIT Press, pp. 1-27, 1991.

D. Heckerman, "Bayesian Networks for Data Mining," Data Mining and Knowledge Discovery, Kluwer Academy Pub., Vol. 1, No. 1, pp 79-119, 1997.

S. J. Hoch, Schkade, "A Physiological Approach to Decision Support Systems," *Management Science*, Vol. 42, No. 1, pp. 51-64, 1996.

Kdnuggets, "Software for Data Mining and Knowledge Discovery," http://www.kdnuggets.com/software/index.html, (Current May 3, 2000).

Knowledge Systems Group, Dept. of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway, "The ROSSETA Homepage," http://www.idt.unit.no/~aleks/rosetta/rosetta.html (Current May 3, 2000).

George M. Marakas, “Decision Support Systems,” Prentice-Hall Pub., 1999.

I. B. Myers, “Manual for the Myers-Biggs Type Indicator,” Princeton, 1962.

M. Mehta, R. Agrawal, and J. Rissanen, "SLIQ: A Fast Scalable Classifier for Data Mining," EDBT, 1996.

R. Ramakrishnan, "Database Management Systems," McGraw-Hill Pub., 1997.

Thomas L. Ruble, Richard A. Cosier, “Effects of Cognitive Styles and Decision Setting on Performance,” *Organizational Behavior and Human Decision Process*, Vol. 46, No. 2, pp. 283-312, 1990.