DATA-BASED IDENTIFICATION OF SHORT TERM PREDICTORS FOR STOCK MARKET TRENDS USING HETEROGENEOUS MODEL ENSEMBLES

Stephan Winkler[a], Bonifacio Castaño [b], Sergio Luengo[b], Susanne Schaller[a], Gabriel Kronberger[a], Michael Affenzeller[a,c]

[a] Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria
[b] Department of Physics and Mathematics, University of Alcalá, Madrid (Spain)
[c] Institute for Formal Models and Verification, Johannes Kepler University Linz
Altenberger Straße 69, 4040 Linz, Austria

[a] stephan.winkler@fh-hagenberg.at, michael.affenzeller@fh-hagenberg.at, susanne.schaller@fh-hagenberg.at, gabriel.kronberger@fh-hagenberg.at
[b] bonifacio.castano@uah.es, sergio.luengo@edu.uah.es

ABSTRACT

We here show the application of heterogeneous ensemble modeling for training short term predictors of trends in stock markets. A sliding window approach is used; model ensembles are iteratively learned and tested on subsequent data points. The goal is to predict trends (positive, neutral, or negative stock changes) for the next day, the next week, and the next month.

Several machine learning approaches implemented in HeuristicLab and WEKA have been applied; the models produced using these methods have been combined to heterogeneous model ensembles. We calculate the final estimation for each sample via majority voting, and the relative ratio of a sample's majority vote is used for calculating the confidence in the final estimation; we use a confidence threshold that specifies the minimum confidence level that has to be reached.

We show results of empirical tests performed using data of the Spanish stock market recorded from 2003 to 2013.

Keywords: financial data analysis, ensemble modeling, trend classification, machine learning

1. INTRODUCTION AND OVERVIEW

A lot of research on using machine learning for predicting stock market trends has already been discussed in the literature, see for example (Kaboudan et al. 2002), (Potvina et al. 2004), (Summers et al. 2004), (Mallick et al. 2008), and (Bodas-Sagi et al. 2012).

The aim of the research discussed in this paper is to learn classifiers for short-term trends in stock markets. We use several machine learning techniques for training sets of heterogeneous models that are able to estimate trends in stock market data.

We particularly focus on short term prognosis, i.e., models are trained that are designed to predict trends for a short future period: The goal is to predict the trend for the next trading day, the next 5 days, and the next 20 days. Trends are classified as positive, neutral, or negative, so that the goal is to learn classification models that are able to correctly predict the classification of future trends.

For evaluating this approach we apply a sliding window strategy: The given data are split into partitions, and each partition is used for training models that shall explain the following test data points.

In the following section we describe the sliding window short term training and testing approach used here, and in Section 3 we give an overview of the modeling methods applied. In Section 4 we summarize the results achieved using this approach on data of the Spanish stock market (recorded 2003 – 2013), and Section 5 concludes the paper.

2. SHORT TERM PROGNOSIS AND SLIDING WINDOW TREND CLASSIFICATION

We simulate real world applications for stock market trend estimations by applying training and test on the data using a sliding window approach. As depicted in Figure 1, we split the given time series data in partitions; the training of models (as explained in the following section) is done for each partition separately, and all models are tested only on data samples that are directly subsequent to the respective training samples. This simulates short term prediction for stock market data that are (in real life) updated periodically.

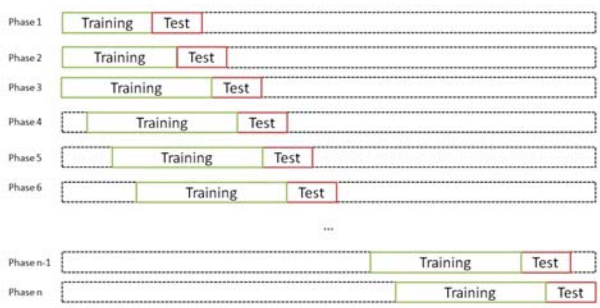

Figure 1: Sliding window short term prognosis training and testing of models.

For the analyzed target variables we are in this research work interested in the classification of their trends. I.e., the goal is to train classification models that estimate the trend of a variable *V* for the next n samples (days) as positive (rising by at least *e*%), negative (falling by at least *e*%), or neutral.

3. HETEROGENEOUS ENSEMBLE MODELING AND CONFIDENCE ESTIMATION

The following techniques for training classifiers have been used in this research study:

- Neural networks (Nelles 2001),
- k-nearest-neighbor classification (Duda et al. 2000),
- support vector machines (Vapnik 1998; Chang & Lin 2001),
- genetic programming (Koza 1992; Affenzeller et al. 2009; Winkler 2009),
- decision trees (Kotsiantis 2007), and
- random forests (Breiman 2001).

Most of these machine learning methods have been implemented using the HeuristicLab framework (Wagner et al. 2014), a framework for prototyping and analyzing optimization techniques for which both generic concepts of evolutionary algorithms and many functions to evaluate and analyze them are available; for learning decision trees the WEKA framework (Hall et al. 2009) was used. Details of these methods are summarized in the appendix of this paper.

For all methods we have used parameter settings that were identified as the best settings by the approach described in (Winkler et al. 2011). There we have described the use of evolutionary optimization for identifying optimal modeling parameters for the here presented data collections as well as the modeling methods applied here.

Each modeling method was executed five times using 10-fold cross validation. All so created models were then applied on their respective test partitions; the final classification for each sample was calculated in the following way (as described in detail in (Winkler et al. 2014(a)) and (Winkler et al. 2014(b))):

For each sample *s* various models are trained; models are here referred to as m_i where *i* represents the current model index. Subsequently, a voting over all models for each sample *s* is performed:

$$vote(c,s) = |\{ m_i : m_i(s) = c \}| \quad (1)$$

where *c* represents an arbitrary, but fix class *c*.

The final classification *fc* for a sample *s* is calculated by using a majority voting of votes *vote(c,s)*:

$$fc(s) = argmaxc\ (vote(c,s)) \quad (2)$$

Finally, we define the confidence of classification *c* for a sample *s* (which represents the confidence towards this classification) on the basis of a set of models *m*:

$$conf(c,s,m) = |\{ m_i : m_i(s) = c \}| \,/\, |m| \quad (3)$$

As we are here facing ternary classification tasks, the classification confidence for any final classification *fc* (that is a majority vote winner and thus must have more than 1/3 of the votes) will always be in the interval]1/3, 1].

The overall work flow followed here is graphically shown in Figure 2:

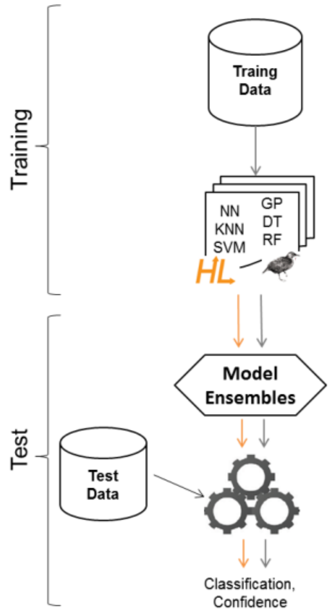

Figure 2: Hybrid ensemble modeling with confidence estimation.

In the empirical tests documented in the following sections we use a confidence threshold $\theta \in [0 \dots 1]$. If the confidence for a sample's classification is smaller than this threshold, then there is no estimation statement for this sample. As a consequence, the ratio of samples for which a classification statement is given will be below 100%; we expect this samples coverage ratio to decrease for increasing values of θ.

4. EMPIRICAL STUDY

For this testing this approach we have used quotes from 10 of the most representative shares in the Spanish market for the years 2003 - 2013. In that period we have identified three types of different market developments: bull market (a long upward price movement where prices in the end of the period are higher than in the beginning), bear market (a long downward price movement where prices in the end are lower than in the beginning), and sideways market (prices in the end and the beginning are almost the same) (Kirkpatrick & Dahlquist 2006).

In the following we describe the executed data preprocessing steps, the target variables for which we trained estimators, and show statistics on the results achieved using the sliding window short term approach with heterogeneous model ensembles.

4.1. Raw Data

We have obtained data from http://www.Yahoo.com/finances using the *quantmod* R library (Ryan 2008). This data source contains most of the Spanish stock market data recorded on a daily basis since 2000. For our experiments we have selected a set of relevant stocks listed on the IBEX35 index.

For all stocks we use the following daily available information:

- The ***adjusted close price*** represents the price of a share at the end of a trading day and takes into account the dividends and splits of that specific share.
- The ***volume*** represents the number of trading actions that were executed for a share on a trading day.

The companies we have chosen are 10 of the most representative ones in the Spanish stock market for which reliable data can be obtained in a period of 10 years from January 2003 to January 2013. These are:

- Telecommunication: **TEF** (Telefónica)
- Electricity: **ELE** (Endesa)
- Banks: **POP** (Popular), **SAN** (Santander) and **BBV** (Bilbao Vizcaya)
- Oil industry: **REP** (Repsol)
- Construction: **FCC** (Fomento de Construcciones y Contratas) and **SYV** (Sacyr Vallehermoso)
- Insurance: **MAP** (Maphre)
- Engineering: **IDR** (Indra)
- Index: **IBEX35** (summarizes the most capitalized companies in the Spanish stock market; shown in Figure 3)

For these variables the data from January 02, 2003 until January 02, 2013 have been collected; in total, 2548 samples (i.e., data of 2548 days) are available.

Figure 3: The IBEX35 index from 2003.01 to 2013.01

4.2. Data Preprocessing, Definition of Features

For each share *S* at time *t* we calculate the following features that are used as input for training classifiers:

- *VolStd10(S,t)*: Standard deviation of the volume of *S* in time period [t-10, …, t]
- *VolStd25(S,t)*: Standard deviation of the volume of *S* in time period [t-25, …, t]
- *PriceAvg10(S,t)*: Average price of *S* in time period [t-10, …, t]
- *PriceAvg25(S,t)*: Average price of *S* in time period [t-25, …, t]
- *PriceStaq(S,t)*: Short term averages quotient, calculated as *PriceAvg25(S,t) / PriceAvg10(S,t)*
- PriceStd10(S,t): Standard deviation of the price of *S* in time period [t-10, …, t]
- PriceStd25(S,t): Standard deviation of the price of *S* in time period [t-25, …, t]

4.3. Target Variables

For a share *S* the goal in this research work is to train classifiers for the following targets:

- *PriceFutureTrend1(S,t)*: Trend (positive, neutral, or negative) for the price of *S* at time *t+1* relative to the price of *S* at time *t*
- *PriceFutureTrend5(S,t)*: Trend (positive, neutral, or negative) for the price of *S* at time *t+5* relative to the price of *S* at time *t*
- *PriceFutureTrend20(S,t)*: Trend (positive, neutral, or negative) for the price of *S* at time *t+20* relative to the price of *S* at time *t*

A trend is classified positive if the difference is at least +1% and negative if the difference is at least -1%.

4.4. Results

For five selected stocks (BBV, ELE, TEF, MAP, and POP) as well as the IBEX35 index we have executed the heterogeneous short term learning and prognosis approach described in Sections 2 and 3; we trained models for the classification of the trend for the next day, the next week (i.e. in five trading days), and the next month (i.e. in 20 trading days). The number of training and test samples in each training / test cycle was set to 100 and 10, respectively. The modeling methods listed in Section 3 were used 20 times independently with varying parameter settings leading to a set of test classification for each sample.

The results for these test series are summarized in Table 1. For varying confidence thresholds θ we report classification accuracies and sample coverage ratios.

We see that for all target variables increasing the confidence threshold θ leads to increased classification accuracies, but also to decreased sample coverage ratios.

Furthermore, we also see that for all target variables the classification accuracy for the close future's trend ranges from approx. 0.5 (for θ=0) to approx. 0.7 (for θ=1). As expected, the classification accuracies for the trend for the next day are higher than those for the next week. Surprisingly, the accuracies for the next month's trend are for all target variables higher than those for the next week.

We exemplarily show classification rates vs. samples coverage ratios for IBEX in Figure 4. In Figures 5 and 6 we show the classification accuracies and confidences for BBV for the trends for the next day and the next week, respectively. In Figure 7 we show the 30 day moving average of classification confidence, accuracy, and samples coverage for IBEXA prognosis for the next day and θ=0.8.

Target variable	θ	classification accuracy	samples coverage
BBV 1d	0.0	0.490	1.000
	0.5	0.493	0.906
	0.8	0.543	0.553
	1.0	0.603	0.136
BBV 5d	0.0	0.484	1.000
	0.5	0.491	0.936
	0.8	0.509	0.694
	1.0	0.517	0.227
BBV 20d	0.0	0.574	1.000
	0.5	0.582	0.939
	0.8	0.605	0.766
	1.0	0.629	0.413
ELE 1d	0.0	0.494	1.000
	0.5	0.514	0.931
	0.8	0.561	0.676
	1.0	0.668	0.311
ELE 5d	0.0	0.447	1.000
	0.5	0.454	0.944
	0.8	0.479	0.672
	1.0	0.494	0.242
ELE 20d	0.0	0.504	1.000
	0.5	0.508	0.942
	0.8	0.523	0.756
	1.0	0.568	0.394
TEF 1d	0.0	0.507	1.000
	0.5	0.521	0.929
	0.8	0.562	0.713
	1.0	0.633	0.306
TEF 5d	0.0	0.449	1.000
	0.5	0.447	0.931
	0.8	0.480	0.660
	1.0	0.512	0.188
TEF 20d	0.0	0.542	1.000
	0.5	0.549	0.940
	0.8	0.568	0.759
	1.0	0.587	0.416
IBEXX 1d	0.0	0.545	1.000
	0.5	0.551	0.928
	0.8	0.601	0.707
	1.0	0.690	0.388
IBEX 5d	0.0	0.458	1.000
	0.5	0.458	0.934
	0.8	0.481	0.653
	1.0	0.506	0.190
IBEX 20d	0.0	0.564	1.000
	0.5	0.576	0.937
	0.8	0.595	0.753
	1.0	0.613	0.396
MAP 1d	0.0	0.456	1.000
	0.5	0.474	0.919
	0.8	0.521	0.625
	1.0	0.628	0.237
MAP 5d	0.0	0.484	1.000
	0.5	0.489	0.938
	0.8	0.509	0.685
	1.0	0.575	0.245
MAP 20d	0.0	0.550	1.000
	0.5	0.561	0.943
	0.8	0.579	0.793
	1.0	0.594	0.461
POP 1d	0.0	0.460	1.000
	0.5	0.472	0.909
	0.8	0.540	0.627
	1.0	0.672	0.283
POP 5d	0.0	0.467	1.000
	0.5	0.474	0.938
	0.8	0.509	0.690
	1.0	0.531	0.267
POP 20d	0.0	0.593	1.000
	0.5	0.605	0.942
	0.8	0.622	0.776
	1.0	0.649	0.445

Table 1: Heterogeneous short-term prognosis modeling test results. The confidence threshold θ is varied leading to varying classification accuracies and sample coverage ratios.

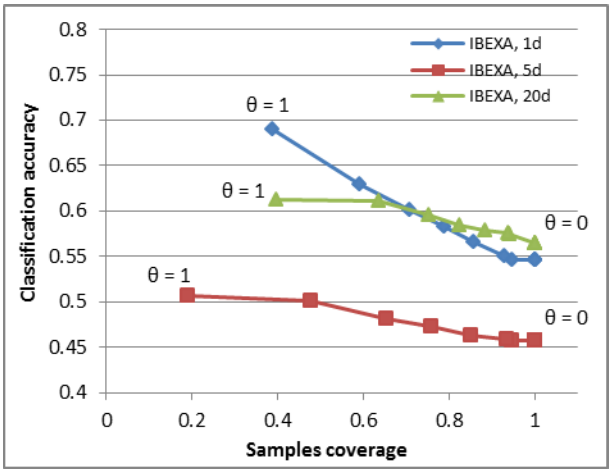

Figure 4: Classification accuracy vs. samples coverage for IBEX trend for the next day, next week, and next month for θ varied from 0.0 to 1.0.

5. CONCLUSION, OUTLOOK

In this paper we have described the application of heterogeneous short time modeling and prognosis to stock market data. In the test series results we have shown that the use of heterogeneous model ensembles leads to increased classification accuracy for the near future's trend of stocks and indices; this can be increased even more by increasing the threshold for the classification confidence, bearing in mind that this decreases the ratio of samples for which a classification statement can be given.

Future work shall focus on the application of this approach to data of other stock markets as well as the identification of variable impacts.

REFERENCES

Affenzeller, M., Winkler, S., Wagner, S., & Beham, A. (2009). *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall/CRC. ISBN 978-1584886297.

Bodas-Sagi, D.J., Fernández-Blanco, P., Hidalgo, J. I., & Soltero-Domingo F.J. (2012) A parallel evolutionary algorithm for technical market indicators optimization. *Natural Computing*, Vol. 12 (2), pp 195-207. Springer.

Breiman, L. (2001). Random Forests. *Machine Learning* 45 (1): 5-32.

Chang, C.-C., & Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*. Wiley Interscience, 2nd edition.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten I.H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11(1).

Kaboudan, M. (2002) GP forecasts of stock prices for profitable trading. *Evolutionary computation in economics and finance*, pp. 359-382. Springer.

Kirkpatrick, C.D. & Dahlquist, J. (2006) *Technical Analysis: The Complete Resource for Financial Market Technicians*. Financial Times Press.
Kotsiantis, S. (2007) Supervised machine learning: A review of classification techniques. *Informatica*, 31:249–268.
Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. University of Michigan Press.
Ljung, L. (1999). *System Identification – Theory For the User*, 2nd edition. PTR Prentice Hall, Upper Saddle River, N.J.
Mallick, D., Lee, V.C., Yew Soon Ong (2008). An empirical study of Genetic Programming generated trading rules in computerized stock trading service system. *Proceedings of the 2008 International Conference on Service Systems and Service Management*, pp. 1-6.
Nelles, O. (2001). *Nonlinear System Identification*. Springer Verlag, Berlin Heidelberg New York.
Potvina, J-Y., Sorianoa, P., Vallee, M. (2004) Generating trading rules on the stock markets with genetic programming. *Computers & Operations Research* 31(7), 1033-1047. Elsevier.
Ryan, J. A. (2008) Quantmod: Quantitative Financial Modeling Framework. *Google academy*.
Summers, B., Griffithsz, E., & Hudson, R. (2004) Back to the future: an empirical investigation into the validity of stock index models over time. *Appl. Fina. Econ.* 14, 209-214.
Vapnik, V. (1998). *Statistical Learning Theory*. Wiley, New York.
Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., & Affenzeller, M. (2014) Architecture and Design of the HeuristicLab Optimization Environment. In *Advanced Methods and Applications in Computational Intelligence, Topics in Intelligent Engineering and Informatics Series*, pp. 197-261. Springer
Winkler, S. (2009). *Evolutionary System Identification - Modern Concepts and Practical Applications*. Schriften der Johannes Kepler Universität Linz, Reihe C: Technik und Naturwissenschaften. Universitätsverlag Rudolf Trauner. ISBN 978-3-85499-569-2.
Winkler, S.M., Schaller, S., Dorfer, V., Affenzeller, M., Petz, G., & Karpowicz, M. (2014) Data based prediction of sentiments using heterogeneous model ensembles. Submitted to *Soft Computing*.
Winkler, S.M., Affenzeller, M., Schaller, S., & Stekel, H. (2014) Data Based Prediction of Cancer Diagnoses Using Heterogeneous Model Ensembles - A Case Study for Breast Cancer, Melanoma, and Cancer in the Respiratory System. In *Proceedings of the GECCO 2014 Workshop on Medical Applications of Genetic and Evolutionary Computation (MedGEC 2014)*, 2014.

AUTHORS BIOGRAPHIES

STEPHAN M. WINKLER received his PhD in engineering sciences in 2008 from Johannes Kepler University (JKU) Linz, Austria. His research interests include genetic programming, nonlinear model identification and machine learning. Since 2009, Dr. Winkler is professor at the Department for Medical and Bioinformatics at the University of Applied Sciences (UAS) Upper Austria at Hagenberg Campus; since 2010, Dr. Winkler is head of the Bioinformatics Research Group at UAS, Hagenberg.

BONIFACIO CASTAÑO received his PhD in Physics from Complutense University of Madrid (UCM) in 1985. He has collaborated with multiple international research organizations, for example with CERN in Geneva, where he worked as a research fellow in High Energy Physics from 1985 to 1986. Since 1989 he is staff professor at the Department of Physics and Mathematics at the Alcalá University of Madrid (UAH). His reseach interests include computer science, wireless communications, robotics, and genetic programming applied to stocks markets analysis.

SERGIO LUENGO's received a degree in Computer Science from the University of Alcalá de Henares (UAH), Spain, in 2013. Currently he is M.Sc. student in Artificial Intelligence at the Technical University of Madrid (UPM) and is a researcher at the Computational Intelligence Group in UPM.

SUSANNE SCHALLER received her Masters of Science in Biomedical Informatics in 2009 and 2010 from the University of Skoevde and Upper Austrian University of Applied Sciences. She is a research and teaching associate at the Research & Development GmbH, in Hagenberg. Her research interests include system biology, medical image analysis, tissue engineering, data mining and machine learning.

GABRIEL KRONBERGER received his PhD in engineering sciences in 2010 from JKU Linz, Austria, and is a research associate at the UAS Research Center Hagenberg. His research interests include genetic programming, machine learning, and data mining and knowledge discovery.

MICHAEL AFFENZELLER has published several papers, journal articles and books dealing with theoretical and practical aspects of evolutionary computation, genetic algorithms, and meta-heuristics in general. In 2001 he received his PhD in engineering sciences and in 2004 he received his habilitation in applied systems engineering, both from the Johannes Kepler University of Linz, Austria. Michael Affenzeller is professor at UAS, Campus Hagenberg.

Figure 5: Classification accuracies for the next day's and next week's trend for BBV over the analyzed time period. The confidence threshold θ set set to 0.8. Classification accuracies for the next day's trend decrease after 2007.

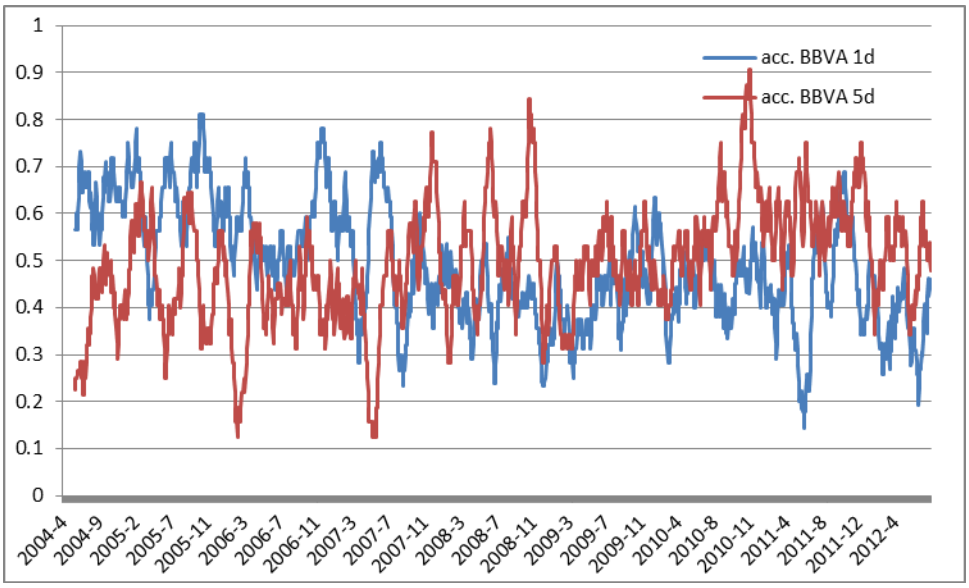

Figure 6: Classification confidence for the next day's and next week's trend for BBV over the analyzed time period. We see that the confidences for both the next day's and the next weeks trends are rather high until 2007; after 2007 the confidence for the next day's classification decreases which can be interpreted as an indicator for changes in the analyzed system.

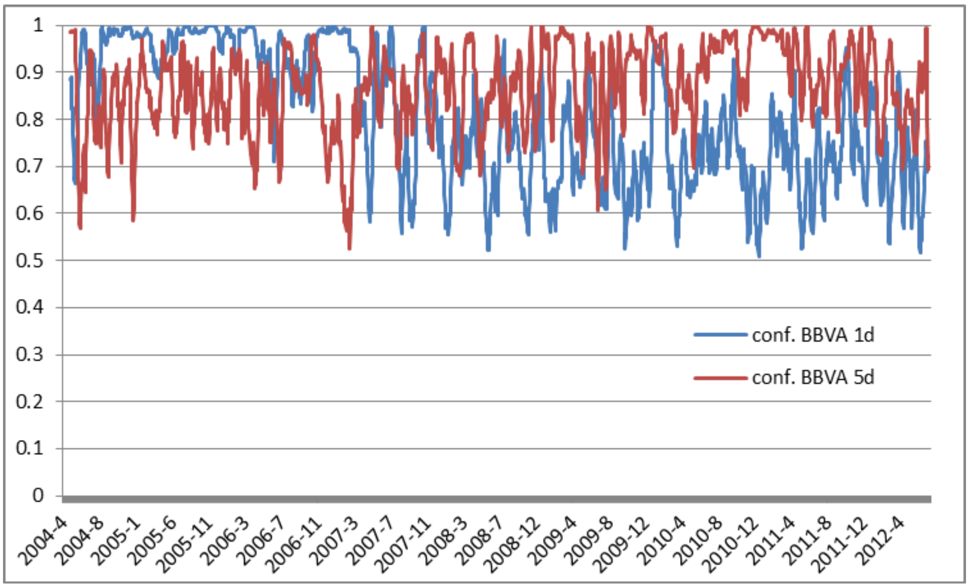

Figure 7: Classification confidence, accuracy, and samples coverage for IBEXA prognosis for the next day (θ=0.8). We here again see that classification confidence and samples coverage decrease after 2007 which can be interpreted as an indicator for changes in the analyzed system.