2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

Ensembling via Prediction Market Mechanisms for Time-series Forecasting

Christian Merkwirth† and Maciej J. Ogorzałek†

†Department of Information Technologies
Faculty of Physics, Astronomy and Applied Computer Science
Jagiellonian University
al. Reymonta 4, 30-059 Krakow, Poland
Email: maciej.ogorzalek@uj.edu.pl,christianmerkwirth@gmail.com

Abstract—In this paper we propose the use of prediction market mechanisms for time series forecasting. Prediction markets give predictions much closer to real outcomes than polls. F. Hayek proposed in 1945 that the market acts as an information aggregation mechanism. While prediction markets were recently applied by economists to predict the outcome of elections or the probability of a single event, our proposed mechanism is intended to improve the aggregation of a series of forecasted values to an ensemble forecast. Prediction market approaches assume that all information about a security is reflected in its price. It is expected that in a similar manner the output of the participating models in an ensemble could be aggregated via prediction market mechanisms instead of simply pooling results. Having to bet a certain amount of credit points for each forecasted value makes the creator of a model more aware of the specific strengthes and weaknesses of the participating model. Market mechanisms can thus be used for building useful ensembles of models which might outperform classical aggregation methods.

1. Introduction

In many disciplines of science and technology we are confronted with processing and interpretation of huge amounts of measurement data. Typically we have :
Given: A time-series of input-output-pairs $(\vec{x}^{\mu}, y^{\mu})$ with $\mu = 1, \ldots, N$ or a functional dependence $y(\vec{x})$ (possibly corrupted by noise)
we would like to **chose a model** (function) $\hat{f}$ out of some hypothesis space $\mathcal{H}$ as close to the true f as possible
Models built on the basis of measured time series can be used for forecasting - estimation of the future values of variables.

Many sophisticated methods have been developed for system modeling based on measurement data. Among those methods model building using statistical learning techniques [5] play very important role. Any type of model constructed has to pass the validation stage - the quality of models can be evaluated based on the generalization error ie. performance of the model on the new (unseen) data. In a logical way we select the model with lowest (estimated) generalization error.

2. Ensemble Methods

Unfortunately there is no known method for optimal model selection. Usually a number of models are constructed using the training data and later their performance is evaluated on the test data sets. The best performing model is selected among those tested.

In recent years some researchers have proposed to combine outputs of several models trained separately. This technique is called ensambling [5] and a number of variants such as boosting, bagging etc. have been proposed as special cases of ensemble building. Building an ensemble of models one can consider:
- Simple average $\bar{f}(\vec{x}) = \frac{1}{K} \sum_{k=1}^{K} f_k(\vec{x})$ or
- Weighted average $\bar{f}(\vec{x}) = \sum_k \omega_k f_k(\vec{x})$ with $\sum_k \omega_k = 1$

The ensemble generalization error is always smaller than the expected error of the individual models. An ensemble should consist of well trained but diverse models.

2.1. The Bias/Variance Decomposition for Ensembles

Let us consider the case where we have a given data set $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ and we want to find a function $f(\mathbf{x})$ that approximates y also for unseen observations of $\mathbf{x}$ or gives estimates (predicts) of the future values of y for x outside the training set. The expected generalization error $Err(\mathbf{x})$ given a particular $\mathbf{x}$ and a training set D is

$$Err(\mathbf{x}) = E[(y - f(\mathbf{x}))^2 | \mathbf{x}, D] \quad (1)$$

where the expectation $E[\cdot]$ is taken with respect to the probability distribution P. The Bias/Variance Decomposition of $Err(\mathbf{x})$ is

$$\begin{aligned} Err(\mathbf{x}) &= \sigma^2 + (E_D[f(\mathbf{x})] - E[y|\mathbf{x}])^2 \\ &\quad + E_D[(f(\mathbf{x}) - E_D[f(\mathbf{x})])^2] \quad (2) \\ &= \sigma^2 + (\mathrm{Bias}(f(x)))^2 + \mathrm{Var}(f(x)) \quad (3) \end{aligned}$$

where the expectation $E_D[\cdot]$ is taken with respect to all possible realizations of training sets D with fixed sample size N and $E[y|\mathbf{x}]$ is the deterministic part of the data and σ^2 is the variance of y given $\mathbf{x}$. Balancing between the bias and the variance term is a crucial problem in model building. Let us consider the case of an ensemble average $\hat{f}(\mathbf{x})$

consisting of K individual models

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{K} \omega_i f_i(\mathbf{x}) \qquad w_i \geq 0, \tag{4}$$

where the weights may sum to one $\sum_{i=1}^{K} \omega_i = 1$. If we put this into eqn. (2) we get

$$Err(\mathbf{x}) = \sigma^2 + \text{Bias}(\hat{f}(x))^2 + \text{Var}(\hat{f}(x)), \tag{5}$$

As the bias term in eqn. (5) is the average of the biases of the individual models one should not expect a reduction in the bias term compared to single models.
The variance term of the ensemble could be further decomposed [10] and contains the variances of the ensemble members and the cross terms which disappears if the models are completely uncorrelated [7]. The reduction of the variance of the ensemble is related to the degree of independence of the single models. This is a key feature of the ensemble approach.

Krogh et al. [8] derive the equation $E = \bar{E} - \bar{A}$ which relates the ensemble generalization error E with the average generalization error $\bar{E}$ of the individual models and the variance $\bar{A}$ of the model outputs with respect to the average output. When keeping the average generalization error $\bar{E}$ of the individual models constant, the ensemble generalization error E should decrease with increasing diversity of the models $\bar{A}$. Hence we try to increase A by using two strategies:

1. Resampling: We train each model on a randomly drawn subset of 80% of all training samples. The number of models trained for one ensemble is chosen so that usually all samples of the training set are covered at least once by the different subsets.

2. Variation of model type: We employ two different model types, which are linear models trained by ridge regression and k-nearest-neighbor (k-NN) models with adaptive metric.

Ensembling methods have found numerous applications for approximation and classification purposes [5, 9]. Recently successful approaches to prediction have been also reported [17, 18].

From the above description it becomes clear that there is no method known for selection of a (quasi-)optimal ensemble - optimal selection of participating models. In this paper we propose one possible approach to ensemble optimization based on so-called prediction market approaches. Prediction markets [6] despite the lack of theoretical background [16] have been used to predict results of elections , distribution of prices of goods [14, **?**] or proliferation of diseases [1]. We propose to construct a special type of prediction market mechanism adopted to time-series prediction.

3. Description of the method

Without loss of generality, we describe the method in context of forecasting a daily time-series. The method can be applied to any other sampling sampling schema such as weekly, hourly, tick-based in a straightforward fashion.

3.1. Terminology

Forecast organizer The forecast organizer is the entity that organized the forecast. The organizer declares what has be to forecaste over which forecasting period and specifies the height of the cash reward. The organizer also supplies all participating models with the same historical and accompanying data. The forecast organizer can be a person, group, company, institution, computer program or web service.

Forecast value The forecast value $\hat{v}_i(m_j)$ is the value forecasted by a participating model m_j for the i-th day of the forecast period.

Forecast period In the setting considered in this study, we assume that a valid forecast is constituted of daily forecast values that cover the whole forecasting period of one or more days, e.g. all days of the upcoming month.

Participating model Each model m_j of the ensemble of models that is able to generate a series of forecast values that covers the whole forecast period is called a participating model. A participating model can thus be a human applying a certain forecasting technique or an algorithm or computer program that automatically delivers a forecast.

Betting points The betting points $p_i(m_j)$ is the amount of points bet by model m_j on the forecast value $\hat{v}_i(m_j)$ for the i-th day. The general idea of betting points is to allow models to express their relative confidence into the respective predicted value.

Reward points The reward points r_i is the amount of credits received for winning the i-th day competition.

Cash reward The forecast organizer declares the total amount of real money C (e.g. 1000$) that will be available for rewarding the participating models according to their collected reward points.

3.2. Proposed market mechanism

We have several participating models $\mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}, \mathbf{m_4}, \ldots$ generating a set of predictions for each day for the current forecast period (typically one week or one month). The market mechanism works as follows:

$$
\begin{array}{lcccccccccc}
\text{model/day}(i) & 1 & 2 & 3 & 4 & \dots & \dots & 28 & 29 & 30 \\
\mathbf{m_1} & 1.5(30) & 1.6(30) & 1.7(30) & \dots & \dots & \dots & 4.2(0) & 4.3(0) & 4.4(0) \\
\mathbf{m_2} & 2.2(10) & 2.15(10) & 2.1(10) & \dots & \dots & \dots & 0.85(10) & 0.8(10) & 0.75(10) \\
\mathbf{m_3} & 1.2(30) & 1.2(0) & 1.2(0) & \dots & \dots & \dots & 1.2(3.0) & 1.2(0) & 1.2(0) \\
\mathbf{a} & 1.47 & 1.74 & 1.8 & \dots & \dots & \dots & 1.11 & 0.8 & 0.75
\end{array} \quad (6)
$$

Figure 1: Matrix showing possible different strategies for betting. The number denote the forecast values for each day with associated betting points (in parenthesis). The last row shows the aggregated (ensemble) forecast a_i. The aggregation was performed by computing the weighted average.

- Each participating model m_j obtains for each day of the forecast period a pre-assigned number of betting points (say 10 points for each of the 30 days = total of 300 points). These betting points are added to the total balance P of betting points for model m_j. From these total amount betting points can be freely distributed on the days of the current forecast period.
- Each participating model m_j has publish for the each day of the forecast period two quantities, the forecast value $\hat{v}_i(m_j)$ i and the amount of betting points $p_i(m_j)$ it associated to this forecast value.
- The total balance of betting points for a model can not become negative. Thus the betting points distributed over the days of the forecast period may not exceed total balance of betting points.
- Unused betting points in the total balance can be saved for future forecast periods.
- The aggregated forecast value a_i for day i are computed based on the forecast values $\hat{v}_i(m_j)$ and betting points $p_i(m_j)$ of all participating models.

For example the assigned betting points would be as shown in the matrix in Fig.1

The respective lines in the table show three typical approaches or strategies chosen by the competitors. The first one has high confidence in his short-term predictions and puts all betting points on the first guesses. The second competitor has apparently little knowledge about the quality of the forecast on different days and chooses to put equal bets for all predictions. The third competitor has high confidence on the bets only for selected days (e.g. a specific weekday) and chooses to bet zero points on other days.

After the true outcome of the time-series for the forecast period is known, according to the accuracy and number of points bet on each day reward points will be distributed. In the long run, competitors making better forecasts will get higher rewards.

3.3. Computing aggregates on the basis of daily bets

The aggregate forecast a_i for the i-th day can be computed in at least two ways:

1. As weighted average

$$a_i = 1/\sum_j p_i(m_j)\sum_j p_i(m_j)\hat{v}_i(m_j) \quad (7)$$

2. As weighted median where the forecast value ranked cumulative sum of points reaches 50% of the total points for this day.

3.4. Calculation of reward points

The winning model m_w of the i-th day's competition is the model for which is forecasted value $\hat{v}_i(m_j)$ is closest to the actual outcome of the time-series $\hat{v}_i$ on this day. In case of a tie, all models that are the same far from the actual outcome are considered to be the winning models. Each winning model receives the following reward points (these are related to so-called "pari mutuel" scheme [12]):

$$r_i(m_j) = p_i(m_j)/\sum_j p_i(m_j) \quad \text{if } |\hat{v}_i(m_j) - v_i| \min \quad (8)$$

$$r_i(m_j) = 0 \quad \text{else} \quad (9)$$

The total reward points for model m_j are calculated as $r(m_j) = \sum_i r_i(m_j)$.

3.5. Reward distribution

All participating models obtain a share c_j of the cash reward C proportional to their total reward points $r(m_j)$:

$$c_j = C\frac{r(m_j)}{\sum_k r(m_k)} \quad (10)$$

3.6. Accumulating betting points

Models that participate repeatedly could be assigned in addition to the default betting points that every model receives prior to delivering its forecast with an extra amount of betting points. The amount of additional betting points would be derived from the amount of reward points received for the previous forecast period. In addition to carrying over unused betting points from previous forecast periods, this would enable above-average performing models to accumulate betting points.

This mechanism would have two beneficial effects:

1. The stimulation of constant participation of models in repeated forecast periods.

2. Models which perform above average would have the possibility to influence the daily aggregate stronger than poorly performing models or models that participate for the first time.

In case of a model that performed well repeatedly, but then degraded in performance for an arbitrary reason, the winner-takes-all mechanism when determining the winner(s) of a daily competition would quickly wash away the accumulated betting points of such a model and thus diminishes the influence of this model on future aggregates.

4. Discussion

The proposed mechanism should be able to improve the aggregation of forecast to ensemble forecast by incorporating information about the quality of each forecasted value from the participating groups. Additionally the reward schema encourages participants to tune their forecast to the best possible accuracy that can be obtained by a given forecasting technique. This action can be related to adjusting weights in the aggregate sum of outputs of models in the ensemble. When applied for repeated forecasting periods, participants showing a constantly above-average performance will accumulate betting points and can thus influence the aggregated prediction stronger in the direction of their forecast values which proved superior in the past.

A possible drawback of the method is that in case all participating models assign zero betting points to a certain day of the forecast, there is no information on how to aggregate the forecasted values to the ensemble forecast. To avoid a gap in the forecast series one could compute the unweighted average of the forecast values. No reward points will be assigned to any of the participating models in such a day.

References

[1] Avian Influenza Prediction Market: Prospectus, http://fluprediction.uiowa.edu/fluhome/AvianFlu prospectus.html

[2] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123140, 1996. [Online]. Available: citeseer.ist.psu.edu/ breiman96bagging.html

[3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees". Wadsworth International Group, Belmont, CA., 1984.

[4] L. Hansen and P. Salamon, "Neural Network Ensembles," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993 1001, 1990.

[5] T. Hastie, R. Tibshirani, J. Friedman, "The elements of ststistical leaqrning: Data Mining, Inference and Prediction", Springer Verlag, 2003.

[6] http://hayekcenter.org/

[7] A. Krogh and P. Sollich, "Statistical mechanics of ensemble learning," Physical Review E, vol. 55, no. 1, pp. 811825, 1997.

[8] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in Advances in Neural Information Processing Systems, G. Tesauro, D. Touretzky, and T. Leen, Eds., vol. 7. The MIT Press, 1995, pp. 231238. [Online]. Available: citeseer.ist.psu. edu/krogh95neural.html

[9] L.I. Kuncheva, "Combining Pattern Classifiers", Wiley-Interscience, 2004.

[10] C. Merkwirth, M. Ogorzalek, and J. Wichard, "Stochastic gradient descent training of ensembles of dt-cnn classifiers for digit recognition", in Proceedings of the European Conference on Circuit Theory and Design ECCTD03, vol. 2, Krakw, Poland, September 2003, pp. 337341.

[11] U. Naftaly, N. Intrator, and D. Horn, "Optimal ensemble averaging of neural networks," Network, Comp. Neural Sys., vol. 8, pp. 283296, 1997.

[12] http://en.wikipedia.org/wiki/Parimutuel betting

[13] M.P. Perrone and L. N. Cooper, "When Networks Disagree: Ensemble Methods for Hybrid Neural Networks," in Neural Networks for Speech and Image Processing, R. J. Mammone, Ed. Chapman-Hall, 1993, pp. 126142.

[14] P.M. Polgreen, F.D. Nelson, G.R. Neumann, "Use of Prediction Markets to forecast infectuous disease activity", *Clin. Infect. Dis*,Jan 2007, vol.44, No.2, pp.272-279

[15] P.M. Polgreen, F.D. Nelson, G.R. Neumann, "Using Prediction Markets to Forecast Trends in Infectuous dideases", *Microbe*, vol.1, No.10, pp.459-465, 2006.

[16] J. Surowiecki, "The wisdom of crowds: why the many are smarter then the few and how collective wisdom shapes business, economies, societies and nation", Random House Inc., New York, 2004

[17] J. Wichard and M. Ogorzaek, "Iterated time series prediction with ensemble models," in Pro- ceedings of the 23rd International Conference on Modelling Identification and Control, 2004.

[18] J.D. Wichard, M. Ogorzalek "Time Series Prediction with Ensemble Models applied to the CATS Benchmark", Neurocomputing, Vol. 70, Nos. 13-15, pp. 2371-2378, Aug. 2007.