Absolute Percent Error Based Fitness Functions for Evolving Forecast Models

Andy Novobilski, Ph.D.
College of Engineering and Computer Science
The University of Tennessee at Chattanooga
615 McCallie Ave., Dept. 2302
Chattanooga, TN 37403
(423) 755-4202
Andy-Novobilski@utc.edu

Farhad A. Kamangar, Ph.D.
Department of Computer Science and Engineering
The University of Texas at Arlington
Arlington, TX 76019
(817) 272-3617
Kamangar@cse.uta.edu

Abstract

One aspect of evolutionary computing as a method of data mining, is its intrinsic ability to drive model selection according to a mixed set of criteria. Based on natural selection, evolutionary computing utilizes evaluation of candidate solutions according to a fitness criteria that might or might not share the exact same implementation as the metric used to measure the performance of the selected solution. This paper presents the results of using four different fitness functions to evolve naïve Bayesian networks based on a combination of Mean Absolute Percent Error and Worst Absolute Percent Error values for individual population members. In addition to the error measurements from both the training and forecast evaluations, data is presented that shows APE for individual members during the forecast generation and evaluation phase.

Introduction

One aspect of evolutionary computing as a method of data mining, is its intrinsic ability to drive model selection according to a mixed set of criteria. Based on natural selection, evolutionary computing utilizes evaluation of candidate solutions according to a fitness criteria that might or might not share the exact same implementation as the metric used to measure the performance of the selected solution [1]. For example, the final measurement of a solution might be one of "does it work" while the fitness criteria might be implemented in such a way to give weight to not only does the solution work, but to also consider particular characteristics of the solution such as its overall complexity and ability to accommodate change.

In previous work, a framework for inferencing Bayesian Networks from time series data was presented that used genetic programming to evolve predictive models [2]. The framework relied upon a fitness function that was computed using a normalized version of mean square error to direct the natural selection process. Although additional work shown in [3] validates the use of this fitness metric, it became apparent that in order to compare the results of the natural selection framework with results from other forecasting methods in forecasting competitions such as M-3 [4], a more standard measurement would be needed, with Mean Absolute Percent Error [5], or MAPE, being the best initial candidate for exploration.

One interesting outcome of the work done to utilize MAPE as a performance measure of the generated predictive model, was the exploration of the use of MAPE as part of the fitness function. During this activity, it was noted that reliance upon MAPE as the sole contributor to fitness would cause the framework to produce predictive models that would do consistently well in the Absolute Percent Error (APE) sense for most forecast values, but noticeably worse for others. Given this observation, the question arises as to what impact, if any, would making fitness a function of MAPE and the Worst Absolute Percent Error (WAPE) observed over the training set have on the overall performance of the evolved predictive model.

This paper presents the results of using four different fitness functions to evolve naïve Bayesian networks based on a combination of MAPE and WAPE values for individual population members. The natural selection framework is used to generate predictive models for both a stationary and non-stationary time series from synthesized data. In addition to the error measurements from both the training and forecast evaluations, data is presented that shows the change in MAPE and WAPE by generation during the natural selection process, and the APE for individual members during the forecast generation and evaluation phase.

Fitness and Model Evolution

Simply put, utilizing natural selection to identify predictive models based on time series data involves four basic steps that are common to both genetic algorithms [6] and genetic programming [7]. First, an initial population of candidate models is created randomly from the space of all possible models. Second, the individual members of the population (the predictive models) are trained and then measured as to their relative fitness to other members in the same population. Third, a new population (or generation) is created by selecting members from the existing population in proportion to their fitness to be either carried forward as-is, or combined using the genetic operations of cross-over and/or mutation. Fourth, the second and third steps are repeated until a maximum number of generations have been evolved, or a predetermined criteria for success has been met.

The key step in ranking the members of the population relative to each other is the application of a fitness function to each individual member of the population. The higher the fitness, the more likely the member (an the predictive model it represents) is copied directly forward into the next generation, or is used during the operation of reproduction. For many applications, a simple form of fitness is to measure how well a solution solves the problem. For example, a genetic programming approach might be used to identify a design of a bridge to support a certain amount of weight. In order to rank population members with respect to each other, the major consideration would be whether the bridge can support the required weight. Additional characteristics such as cost to build, number of supports, etc. might also be incorporated in determining fitness to separate acceptable solutions from preferred solutions.

In the case of naturally selecting naïve Bayesian networks for use as predictive models for time series data, there are no absolute acceptable or unacceptable solutions. Therefore, the fitness of a particular individual in the population needs to be related to the individual predictive model's ability to forecast future values of the time series. For the purposes of the work described here, the final measurement of performance will be the Mean Absolute Percent Error, or MAPE. MAPE is defined as:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|\hat{Y}_t - Y_t\right|}{Y_t}$$

By using MAPE as the sole component of the fitness measurement, the natural selection process will evolve a solution that minimizes the average error across the entire training set. However, in the case of several of the experimental datasets used in the framework for past research, this tended to produce a predictive model that was very good for the majority of forecasts and very poor for the others. This lack of consistency in the predictive model's performance causes difficulty in its application to decision-theoretic planning [8], where the ability to identify the utility of the predictive model in terms of how well the model represents the future for all forecasts is an important aspect of solving the problem.

To address this problem, a second form of error measurement could be employed for the fitness function that would seek to minimize the Worst Absolute Percent Error, or WAPE, of all forecasts made by the predictive model being evaluated. WAPE is defined as:

$$WAPE = \max_{t \in \{1,\ldots,N\}}\left[\frac{\left|\hat{Y}_t - Y_t\right|}{Y_t}\right]$$

The next step in the experimental process was to study the impact of using a third fitness function based on the average MAPE and WAPE for a given predictive model. It was hoped that by combining both types of information, the fitness function would drive the natural selection process towards evolving a predictive model that would outperform MAPE alone, while increasing the consistency of the forecasts generated across the entire set of evaluation data. Experimental data supported this hypothesis, but not as strongly as was hoped. Although the consistency of the solution improved, the overall performance did not and in most cases, was out performed by MAPE alone.

Building on the concept of combining MAPE and WAPE, a fourth fitness function was developed that would itself change as the overall performance of the individual predictive model changed. The MAPE and WAPE errors were combined according to:

$$\alpha\, MAPE + \beta\, WAPE$$

where:

$$\alpha = 1 - \beta$$

and:

$$\beta = 1 - \frac{MAPE}{WAPE}$$

Using this weighted combination of MAPE and WAPE causes the fitness function to rank predictive models with the lowest MAPE that is closest to the WAPE as most preferred, while encouraging the natural selection process to separate members that are close in overall performance by how consistent they are in forecasting values across the entire training dataset. This affects the desired behavior of identifying predictive models from the data that are consistent and perform reasonably with respect to the measurement of the Mean Absolute Percent Error across the forecast of both training and evaluation data.

Experimental Results

In order to test the hypothesis that a fitness function based on a variance weighted combination of Mean Absolute Percent Error and Worst Absolute Percent Error, a set of experiments were performed on synthesized data for both a stationary and non-stationary process. The experiments used the framework for naturally selecting Bayesian networks to evolve forecast models based on naïve Bayesian networks for the following four fitness functions:

$F_{MAPE} = 1. - MAPE$

$F_{WAPE} = 1. - WAPE$

$F_{AVG} = 1. - \frac{MAPE + WAPE}{2}$

$F_{\alpha\beta} = 1. - (\alpha MAPE + \beta WAPE)$

The first series, representing the stationary process:

$$y(t) = y(t\text{-}27) + a_t$$

where a_t is a normally distributed random variable with mean 0 is shown in Figure 2. The second series, representing the non-stationary process:

$$y(t) = 1.01\, y(t\text{-}3) - .001\, y(t\text{-}5) + a_t$$

is shown in Figure 3. Both series consisted of 200 data points or which were used to create a training case set of 132 records and an evaluation case set of 56 records. Each training record consisted of the target value combined with the 10 previous values of the time series. The evolutionary search process was further restricted to 10 attributes per variable selected as part of the predictive model. Finally, the population size was selected to be 50 members, with 70 generations evolved from the initial population. The evolutionary process was allowed to continue for the full 70 generations regardless of the convergence of the fitness function, worse case error or mean absolute percent error of the population members. For this experiment, only the fittest member of the population was used in the forecasting part of the experiment. Finally, the synthesized data was chosen to minimize the possibility of APE growing beyond a value of 1.0.

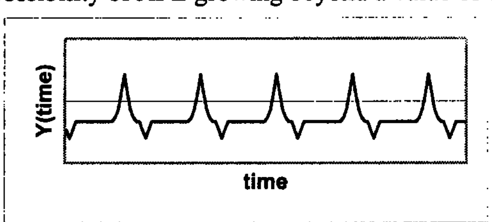


Figure 2 – A synthesized stationary process.

The experimental results, shown in table 1, describe the impact on forecast accuracy using each of the fitness functions has on both the stationary and non stationary process in terms of training error, forecast error, and the difference in performance between training and forecasting. The predictive model delivering the best performance during training evolved during natural selection for both the stationary and non-stationary processes was the one in which F_{MAPE} was used as the fitness function. However, even though the same predictive model evolved using F_{MAPE} performed best for forecasting the stationary process over the evaluation data, it was the $F_{\alpha\beta}$ that did best with the non-stationary process. One interesting point is that in both the stationary and non-stationary process forecasts, the predictive model based on the $F_{\alpha\beta}$ fitness function showed the minimum amount of variance between its performance over both the training data and the forecast evaluation data.

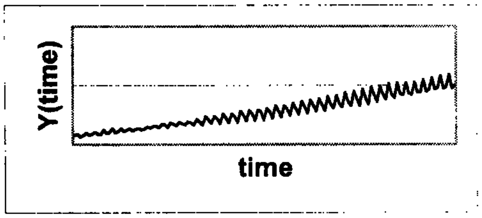


Figure 3 – A synthesized non-stationary process.

Figure 4 presents the results of the forecasts made using predictive models evolved under each of the four fitness functions for the stationary process. It is interesting to note that the combination of both MAPE and WAPE with either the F_{AVG} or $F_{\alpha\beta}$ results in a more consistent Absolute Percent Error during the forecasting process. Also, the best early forecasts are made by F_{MAPE} and $F_{\alpha\beta}$, where the fitness function prefers the minimization of MAPE over WAPE.

The ability of the predictive models produced by fitness functions F_{MAPE} and $F_{\alpha\beta}$ to perform well in the early part of the forecast process is also seen in the models evolved for the non-stationary process. Shown in Figure 5, the results of forecast made using the predictive models evolved under each of the four fitness functions for the non-stationary process show those made by F_{WAPE} and F_{AVG} to have larger APE values than the others. It is also interesting to note, that there is a noticeable difference in the performance of the predictive models for the first 12 forecasts made, with the predictive model generated by the $F_{\alpha\beta}$ fitness model performing more evenly that that of the model generated by the F_{MAPE} fitness function.

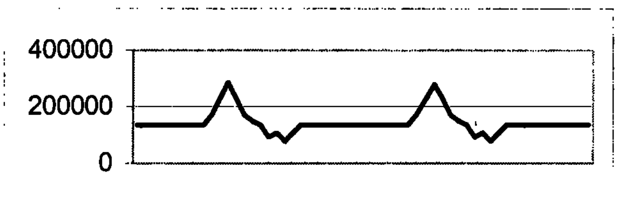
Based on these results, several areas for future study present themselves. First, the type of time series could be expanded to include actual data from multiple interest areas such as economic impact, financial market movement, manufacturing processes, etc. Second, it would be interesting to find out if what, if any, affect the type of Bayesian network (naïve, modified naïve, and classical) would have on the relative quality of predictive models produced by the evolutionary process as driven by the four fitness functions under consideration. Finally, a second set of experiments could be performed to measure the performance of the various predictive models under the constraint of making forecasts that rely upon previous forecast values, versus the current implementation of using known data to affect the forecast value.

Conclusions

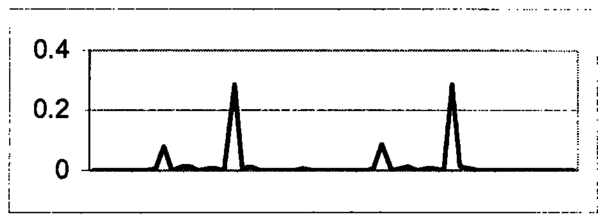
Based on the experimental data presented, it is concluded that the fitness function based on the αβ weighted combination of MAPE and WAPE had a positive impact on the consistency of the naïve Bayesian predictive model produced by the natural selection process. Although not producing the best performing model over the training data, it did produce the model that performed the most consistently of all fitness functions across both the stationary and non-stationary datasets. Finally, the predictive model evolved by the αβ weighted fitness function performed better than the one evolved using MAPE alone for the non-stationary sample data and performed .006 worse than that of MAPE alone for the stationary sample. This indicates that the αβ weighted fitness function would be the best choice in fitness functions when the dataset being mined had equal probability of being produced by a stationary or non-stationary process.

	STATIONARY			NON-STATIONARY		
FITNESS	TRAINING ERROR	FORECAST ERROR	% DIFF	TRAINING ERROR	FORECAST ERROR	% DIFF
F_{MAPE}	0.006403	0.016444	157%	0.006112	0.021415	250%
F_{WAPE}	0.042368	0.055986	32%	0.007074	0.051604	629%
F_{AVG}	0.025857	0.027557	7%	0.007209	0.037668	423%
$F_{\alpha\beta}$	0.021284	0.021843	3%	0.009496	0.018246	92%

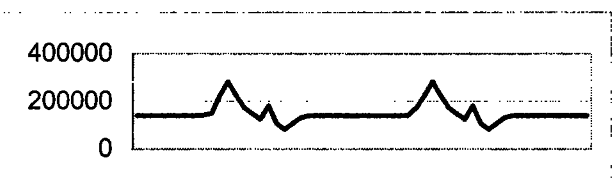
Table 1 – Composite performance data for evolution of naïve Bayesian forecast models.



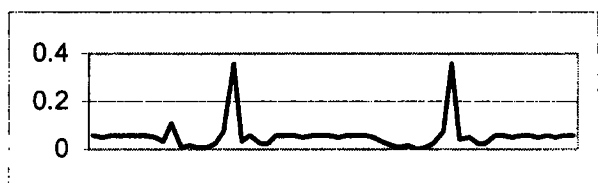
(a) $Y(t)$, $\hat{Y}(t)$ F_{MAPE} for Stationary Process



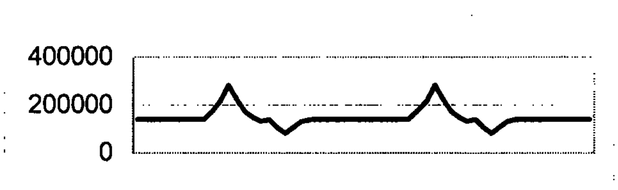
(b) APE(t) F_{MAPE} for Stationary Process



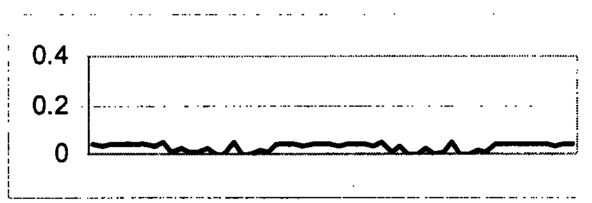
(c) $Y(t)$, $\hat{Y}(t)$ F_{WAPE} for Stationary Process



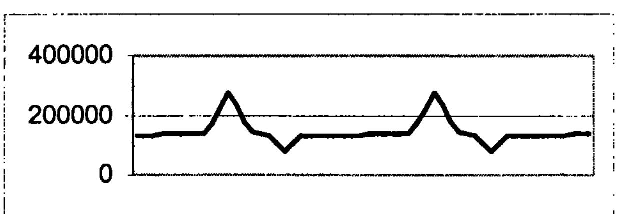
(d) APE(t) F_{WAPE} for Stationary Process



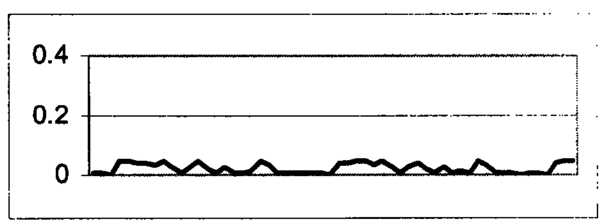
(e) $Y(t)$, $\hat{Y}(t)$ F_{AVG} for Stationary Process



(f) APE(t) F_{AVG} for Stationary Process

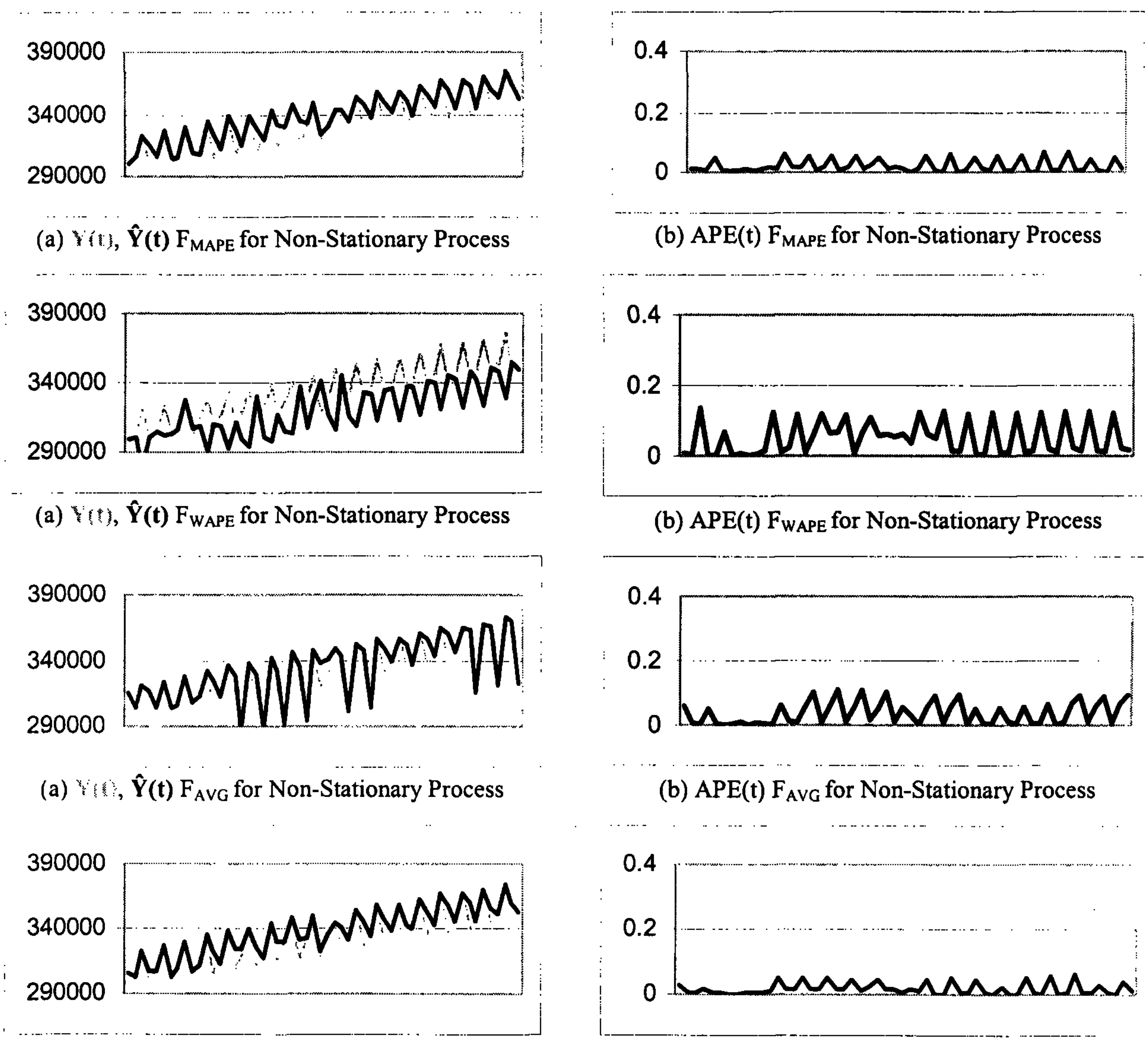


(g) $Y(t)$, $\hat{Y}(t)$ $F_{\alpha\beta}$ for Stationary Process



(h) APE(t) $F_{\alpha\beta}$ for Stationary Process

Figure 4 – Forecast results and Absolute Percent Error for the Stationary Process



(a) $Y(t)$, $\hat{Y}(t)$ F_{MAPE} for Non-Stationary Process

(b) APE(t) F_{MAPE} for Non-Stationary Process

(a) $Y(t)$, $\hat{Y}(t)$ F_{WAPE} for Non-Stationary Process

(b) APE(t) F_{WAPE} for Non-Stationary Process

(a) $Y(t)$, $\hat{Y}(t)$ F_{AVG} for Non-Stationary Process

(b) APE(t) F_{AVG} for Non-Stationary Process

(a) $Y(t)$, $\hat{Y}(t)$ $F_{\alpha\beta}$ for Non-Stationary Process

(b) APE(t) $F_{\alpha\beta}$ for Non-Stationary Process

Figure 5 – Forecast results and Absolute Percent Error for the Non-Stationary Process

[1] Holland, John. 1992. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Cambridge: MIT Press.

[2] Novobilski, Andrew, F. Kamangar. 2000. "Inferencing Bayesian Networks from Time Series Data Using Natural Selection". *The 13th International FLAIRS Conference.*

[3] Novobilski, Andrew. 2000. *Forecasting of Time Series Data Using Naturally Selected Bayesian Networks.* (Ann Arbor, Mich.: University Microfilms).

[4] Hibon, Michele. 2000. "The M-3 Competition Results." *To be published in the International Journal of Forecasting.*

[5] Wun, L.-M. and W. L. Pean. 1991, "Assessing the Statistical Characteristics of the Mean Absolute Error for Forecasting". *International Journal of Forecasting* 7-3,(335-337).

[6] Goldberg, David. 1989. *Genetic Algorithms in Search of Optimization and Machine Learning.* Reading: Addison-Wesley.

[7] Koza, John. 1996. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge: MIT Press.

[8] Haddawy, Peter, S. Hanks. 1998. "Utility Models for Goal-Directed, Decision-Theoretic Planners." *Computation Intelligence* (14)3, August.