Information Sciences 252 (2013) 96–105

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Exponential smoothing with credibility weighted observations

Ronald R. Yager

Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States

A R T I C L E I N F O

Article history:
Received 8 November 2011
Received in revised form 18 June 2013
Accepted 9 July 2013
Available online 19 July 2013

Keywords:
Exponential smoothing
Time series
Aggregation
Forecasting
Importance

A B S T R A C T

Our interest is in time series data smoothing. We view this process as an aggregation of previously observed values. We first discuss the features desired of a good smoothing operator. We particularly note the conflict that exists between our desire for minimal variance and desire to use the freshest data. We describe a number of commonly used smoothing techniques, moving average and exponential smoothing. We then consider the extension of these methods to the case where the observations can have different credibility or importances. Specifically we develop an extension of the exponential smoothing method to the case where the observations can have different importance weights in the smoothing process.

1. Introduction

Time series data smoothing is a task that occurs in many applications and is pervasively used as a tool in for prediction or forecasting [3,6] and learning in evolving systems [1]. Among the most popular methods used to implement this process are moving average and exponential smoothing [4]. In this work we consider a variation of the smoothing problem in which all the observations are not all valued the same, they may have different importances or credibility. Here we associate a weight with each observation indicating its usefulness or importance in the smoothing process. Our objective here then is to extend these smoothing methods, notably the exponential smoothing to handle this type of data.

2. Aspects of time series smoothing

In [9] we discussed some fundamental aspects of time series smoothing and averaging. In time series forecasting or prediction we are interested in using a sequence of observations about some variable, x_t for $t = 1$ to n, to predict a future value for the variable. We are interested in obtaining an estimate of x_{n+1} based upon an aggregation of the earlier values. We denote this aggregation as Agg $(x_1, \ldots, x_n)$. An important factor that determines the form for Agg depends on our assumption about the underlying pattern generating the data. One common assumption, the one which shall initially use here, is that the underlying variable a is almost constant, it maybe slowly varying, and we are observing $x_t = a + e_t$ where e_t is some random error with mean zero and constant variance. More generality may be obtained if we consider $a(t)$, however for our purposes this not required. In this case we are using the observations $x_1, \ldots, x_n$ to obtain some estimate $\bar{a}$ of a. We then use this estimation of a as our predictor of x_{n+1}. Here the estimated value of a, $\bar{a}$, and the estimate of x_{n+1} are the same. One approach to obtaining $\bar{a}$ is to use a mean aggregation. In this case $\bar{a} = F(x_1, \ldots, x_n) = \sum_{j=1}^{n} u_j x_j$ where u_j are a collection of weights such that $u_j \in [0, 1]$ and $\sum_{j=1}^{n} u_j = 1$. This provides an unbiased estimate of a. In the temporal environment this aggregation operation is often called smoothing [4]. The collection of the u_j is referred to as the weighting vector.

E-mail address: yager@panix.com

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.07.008

While on the surface this appears just like an ordinary application of a mean aggregation operator [2] there are a number of features that are special about making these calculations in the framework of time series data. One is the repetitive nature of the task. We are constantly getting new readings for x_t and then using these to update our estimate for $\bar{a}$. In order to formally deal with this sequential updation we shall sometimes find it convenient to use the term a_n to indicate our smoothed value $F(x_1, \ldots, x_n)$. Thus at time n, a_n and $\bar{a}$ are synonyms.

Another special feature of the time series environment is that while we are assuming that the underlying value a is fixed it is often more realistic to allow for the possibility of slow drift. That is a more realistic assumption is that a is quasi-constant.

As we shall subsequently see these two special features of the temporal environment will play an important role in the determination of the weights, the u_j.

The repetitive nature of the calculation of $\bar{a}$ has two immediate implications. One is that it would be advantageous to make the calculation of $F(x_1, \ldots, x_n)$ as simple as possible. In particular, if we could take advantage of prior calculations this would help. We should note that while beneficial, in this age of great computational power and cheap memory this is not as important as in the past. The second implication however is more significant. The repeated updation task requires that we are going to implement many calculation of the form

$$F(x_1, \ldots, x_n), F(a_1, \ldots, a_n, x_{n+1}), F(a_1, \ldots, a_{n+1}, a_{n+2})$$

where each of these is a mean aggregation. Since each of these aggregations will involve a different number of arguments we shall be using weighting vectors of different dimensions. Since it is well known [5] that the mean operator is not generally associative this implies there is no mandated manner for performing the aggregation as we add values. However, it is important that all of these calculations be done in some kind of consistent manner. Here then the issue of consistently calculating $\bar{a}$ for different values of n involves the appropriate determination of weighting vectors of growing dimensions.

The earlier mentioned feature of time series data, the allowance for possible variation in the underlying value a being estimated, has as an implication that not all observations should be treated the same. In particular more weight should be assigned to the most recent observations. Thus we have a preference for weighting vectors such that $u_j \geqslant u_i$ for $j > i$.

One characterizing feature of this smoothing operation is the average age of the data being used in the aggregation. If n is the current time then the age of the piece of data x_t is calculated as AGE(t) = $n - t$. Using this we get as the average of the data being used in the aggregation

$$\overline{AGE} = \frac{\sum_{j=1}^{n} u_j \; AGE(j)}{\sum_{j=1}^{n} u_j}$$

Since $\sum_{j=1}^{n} u_j = 1$ then $\overline{AGE} = \sum_{j=1}^{n} u_j(n - j)$. This can also be expressed as $\overline{AGE} = n - \sum_{j=1}^{n} j u_j$.

As we previously indicated we have some preference for fresh or *youthful* data. We can of course, have the freshest data if we select $u_n = 1$ and have all other u_j equal zero. In this case $\overline{AGE} = 0$. However there is some other conflicting objective that we must consider.

As we noted our observations are of the form $x_t = a + e_t$ where e_t is assumed to be a random noise component with mean zero and variance σ^2. Each piece of data has a variance of σ^2. Since our objective is to find a good estimate for a, we desire to have a small variance in our estimate of $\bar{a}$.

With

$$\bar{a} = \sum_{t=1}^{n} u_j x_t = \sum_{t=1}^{n} u_t (a + e_t)$$

where $\sum_{j=1}^{n} u_j = 1$ we get as our expected value, $E_x[\bar{a}] = a$, thus this is an unbiased estimate. To find the variance of our estimate we calculate

$$\text{Var}(\bar{a}) = E_x\left[\sum_{t=1}^{n} (u_t x_t - a)^2\right]$$

where E_x denote the expected value operator. Under the assumption that the observations are uncorrelated we obtain

$$\text{Var}(\bar{a}) = \sum_{t=1}^{n} u_t^2 \sigma^2 = \sigma^2 \sum_{t=1}^{n} u_t^2$$

One objective is to minimize the value of the variance. Since we have no control over the value of σ^2 this task reduces to the problem of trying to make $\sum_{t=1}^{n} u_t^2$ as small as possible. In the following we shall denote $H(u) = \sum_{t=1}^{n} u_t^2$. Thus we want to get a small value for $H(u)$ where we are constrained by the conditions that $\sum_{j=1}^{n} u_j = 1$ and $u_j \in [0, 1]$.

We now see that our objective in obtaining the weights is to try to find weights satisfying $u_j \in [0, 1]$ and $\sum_{j=1}^{n} u_j = 1$ that make $H(u) = \sum_{t=1}^{n} u_t^2$ as small as possible and while also making $\overline{AGE} = \sum_{j=1}^{n} u_j(n - j)$ small. As we shall see the goals of trying to make H (u) and $\overline{AGE}$ small are essentially conflicting under the conditions that $\sum_{j=1}^{n} u_j = 1$ and $u_j \in [0, 1]$.