Exponential smoothing methods for interval time series

5 authors, including:

Antonio Muñoz
Universidad Pontificia Comillas
41 PUBLICATIONS **1,311** CITATIONS

Carlos Maté
Universidad Pontificia Comillas
27 PUBLICATIONS **673** CITATIONS

Exponential smoothing methods for interval time series

Javier Arroyo[1], Antonio Muñoz San Roque[2], Carlos Maté[2], and Ángel Sarabia[2] *

1- Universidad Complutense - Facultad de Informática
Profesor José García-Santesmases s/n, 28040 Madrid - Spain

2- Universidad Pontificia Comillas - Instituto de Investigación Tecnológica
Alberto Aguilera 25, 28015 Madrid - Spain

Abstract. An interval time series is a sequence of intervals observed sequentially in time. It allows to describe the behavior of phenomena where variability must be taken into account. In this paper, exponential smoothing methods are adapted to this kind of series with the help of interval arithmetic. A comparison of the forecast performance of the interval exponential smoothing and other methods is carried out. These methods include interval multilayer perceptron and modelling, in a separate way, the series of the interval attributes (minimum and maximum, or center and radius) using classical forecasting methods.

1 Introduction

An interval time series (ITS) is a sequence of intervals observed sequentially in time. ITS has been previously proposed in [1] and [2] with the aim to extend symbolic data analysis [3] to the field of time series. In symbolic data analysis, items are described by symbolic variables: lists of values, intervals, frequency distributions, etc. These variables allow the characterization of complex real-life situations and the summarization of large data sets into symbolic ones retaining the key information but offering a more manageable size. In an ITS, the variable observed through time is an interval variable.

ITS represent phenomena that classical time series (i.e. series where observations are single values) cannot accurately describe, such as when variability must be taken into account. For example, an ITS is suitable to describe the lower and upper monthly water levels of a river at a given location; or the range of daily values of a stock index; or the intervals enclosing the levels of an air-pollutant recorded in several meteorological stations distributed along a city.

ITS can be obtained in sampling or summarization contexts. In a sampling context, an ITS arises recording the lower and upper values in each time interval. In a summarization context, an ITS is obtained summing up a set of values by means of an interval for each considered instant. In these contexts, intervals can arise from the minimum and maximum observed values, but also from the interquartile range or from the middle 90% of the scores (in order to avoid outliers); it depends on the aims of the analysis.

This paper tackle ITS forecasting from different approaches. Section 2 defines interval variable and ITS. Section 3 summarizes the main ideas of interval arithmetic, which will be used as the basis in some ITS forecasting methods. Section 4 briefly shows how to measure errors in ITS. In section 5, exponential smoothing methods for ITS are proposed and some ideas to deal with trend and seasonality are introduced. Section 6 shows other approaches to forecast ITS and special attention is given to the Interval Multilayer Perceptron. In section 7 the forecasting performance of the proposed approaches is analyzed by an example. Finally, section 8 concludes.

2 Definitions

An interval variable, $[X]$, is a variable defined for all the elements i of a set E, where $[X]_i = \{[X_{i,L}, X_{i,U}], -\infty < X_{i,L} \leq X_{i,U} < \infty\}, \forall i \in E$. The value of $[X]$ for the ith element can be denoted by the interval lower and upper bounds $[X]_i = [X_{i,L}, X_{i,U}]$ or, equivalently, by the center and radius $[X]_i = \langle X_{i,C}, X_{i,R} \rangle$, where $X_{i,C} = (X_{i,L} + X_{i,U})/2$ and $X_{i,R} = (X_{i,U} - X_{i,L})/2$, respectively.

An ITS can be denoted by $\{[X]_t\}$ and the value of the variable in t can be expressed as $[X]_t = [X_{t,L}, X_{t,U}] = \langle X_{t,C}, X_{t,R} \rangle$. In order to denote a forecasted value, a hat will be placed above the variable, $[\hat{X}]_t$.

3 Interval arithmetic

Apart from symbolic data analysis, other field related with intervals is interval analysis [4]. This field assumes that, in the real world, observations and estimations are usually incomplete or uncertain. Thus, it considers that, if precision is needed, data must be represented as intervals enclosing the real quantities. The theory of interval analysis is build around this idea.

Interval computations are based in interval arithmetic [4], which can be summarized as follows: Let A and B be two intervals and $\square$ be an arithmetic operator, then $A \square B$ is the smallest interval which contains $a \square b\ \forall a \in A$ and $\forall b \in B$. According to this definition, interval addition, subtraction, multiplication and quotient are respectively, defined by:

$$[A] + [B] = [A_L + B_L, A_U + B_U]$$

$$[A] - [B] = [A_L - B_U, A_U - B_L]$$

$$\begin{aligned}[A] \cdot [B] \quad = \quad &[\min\{A_L \cdot B_L, A_L \cdot B_U, A_U \cdot B_L, A_U \cdot B_U\}, \\ &\max\{A_L \cdot B_L, A_L \cdot B_U, A_U \cdot B_L, A_U \cdot B_U\}]\end{aligned}$$

$$[A]/[B] = [A] \cdot (1/[B]), \text{with } 1/[B] = [1/B_U, 1/B_L]$$

It is worth noting that interval arithmetic subsumes the classical one, in the sense that, if the operands of interval arithmetic are intervals with width zero

(i.e. $[a, a], a \in \Re$), the result of interval operations will be equal to the result obtained by the classical operations.

In interval arithmetic, addition and multiplication are associative and commutative. The distributive law does now always hold, but the subdistributive property is satisfied; it is defined as:

$$[A]([B] + [C]) \subseteq [A][B] + [A][C]$$

4 Error Measures for Interval Time Series

According to [2], error measures for ITS cannot be based in the difference between the observed and the actual interval, because interval subtraction does not faithfully represent the concept of deviation, as $[A] - [A] = [0, 0]$ if and only if $[A] = [a, a]$ with $a \in \Re$. Thus, they propose error measures based on distances for interval data, such as the Hausdorff and the Ichino-Yaguchi distance.

Let $\{[X]_t\}$ be the observed ITS, and $\{[\hat{X}]_t\}$ be the forecast of this ITS with $t = 1, ..., n$, the Mean Distance Error based on Hausdorff distance is defined as

$$MDE_H = \frac{1}{n} \sum_{t=1}^{n} [|X_{t,C} - \hat{X}_{t,C}| + |X_{t,R} - \hat{X}_{t,R}|],$$

and the Mean Distance Error based on the Ichino-Yaguchi distance is defined as

$$MDE_{IY} = \frac{1}{n} \sum_{t=1}^{n} 0.5[|X_{t,L} - \hat{X}_{t,L}| + |X_{t,U} - \hat{X}_{t,U}|].$$

5 Exponential Smoothing methods for ITS

Exponential smoothing methods in classical time series obtain forecasts as the weighted moving average of all past observations where the assigned weights decrease exponentially (see [5] for an up-to-date review). In this section, exponential smoothing methods are adapted to ITS[1]. In order to adapt the methods to ITS, a procedure to average intervals is required. We propose to average intervals using interval arithmetic.

5.1 Average Interval

The interval that averages a set E of n intervals $[X]_i$, $i = 1, ..., n$ is defined as

$$[\bar{X}] = \frac{[X]_1 + [X]_2 + ... + [X]_n}{n},$$

where the arithmetic operations are interval arithmetic operations.

The average interval holds the following properties:

$$\bar{X}_L = \frac{X_{L,1} + X_{L,2} + ... + X_{L,n}}{n}, \quad \bar{X}_U = \frac{X_{U,1} + X_{U,2} + ... + X_{U,n}}{n},$$

---

[1]The notation of the proposed methods will be similar to that in [5]

$$\bar{X}_C = \frac{X_{C,1} + X_{C,2} + ... + X_{C,n}}{n}, \ \bar{X}_R = \frac{X_{R,1} + X_{R,2} + ... + X_{R,n}}{n}.$$

These properties allow us to consider that the average interval is the barycenter of a system of particles (the set of intervals E) where each particle is defined by two coordinates: the lower and the upper bounds, or, equivalently, the center and the radius. Obviously, as the average interval is equivalent to the interval barycenter, it can also be seen as the interval that minimizes the addition of the euclidean distances between itself and each interval of the set E.

The definition of moving averages based in the interval average is straightforward, and will not be tackled in this article.

5.2 Simple Exponential Smoothing

The formula of simple exponential smoothing (SES) in classical time series is

$$\hat{X}_{t+1} = \hat{X}_t + \alpha(X_t - \hat{X}_t),$$

where $\alpha \in [0, 1]$. This equation is written in an error-correction form, while its equivalent recurrence form is given by:

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t.$$

If both equations are adapted to ITS using interval arithmetic, they are not equivalent and

$$\alpha[X_t] + (1 - \alpha)[\hat{X}]_t \subseteq [\hat{X}]_t + \alpha([X_t] - [\hat{X}_t])$$

due to the subdistributive property. Thus, the SES method for ITS will adapt equation in recurrence form as it produces tighter intervals. It is defined as:

$$[\hat{X}]_{t+1} = \alpha[X_t] + (1 - \alpha)[\hat{X}]_t,$$

where $\alpha \in [0, 1]$. The initializing phase requires the value of $[\hat{X}]_1$ which can be the first observed value, $[X]_1$, or the average interval of the first three or four observed values. It is clear, that the SES forecast of an ITS is the weighted moving average of all past observations, where the weights decrease exponentially.

5.3 Exponential Smoothing with Trend

The Holt exponential smoothing method allows forecasting classical time series with trend. This method smoothes both the level and the trend of the series. The adaptation of the Holt method for ITS requires to smooth both components.

In our approach, the level of the HTS in t will be represented by an interval, $[S]_t$; and the trend of the HTS in t will be represented by a single value, T_t, that represents the location of the interval by its center. Both, level and trend, are separately smoothed and later added in order to obtain the forecast.

The exponential smoothing method with additive trend (EST) is defined as:

$$[S]_t = \alpha[X]_t + (1 - \alpha)([S]_{t-1} + T_{t-1}),$$

$$T_t = \gamma(S_{C,t} - S_{C,t-1}) + (1 - \gamma)T_{t-1},$$

$$[\hat{X}]_{t+m} = [S]_t + mT_t,$$

where $\alpha, \gamma \in [0, 1]$, $S_{C,t}$ is the center of the interval $[S]_t$, and m is a factor that multiplies the trend in order to produce forecasts for m periods ahead. The initialization values can be $T_1 = X_{C,2} - X_{C,1}$ and $[S]_1 = [X]_1$, but more sophisticated initialization values can be given.

5.4 Exponential Smoothing with Seasonality

In an ITS, it can be considered that there are two different types of seasonality: first, considering that the seasonal variation only concerns to the location of the intervals; and second, considering that the seasonality affects the whole interval. We will propose two different methods in order to deal with both alternatives.

In the first approach, the seasonal component in t is represented as a single value, I_t, representing the changes in the interval location due to the seasonal effect and where interval location is represented as the interval center. The level in t, $[S]_t$, is a deseasonalized interval, i.e. an interval without the seasonal effect. Forecasts are the addition of the level interval and the seasonal component. The interval exponential smoothing method with additive crisp seasonality (IEScS) is given next:

$$[S_t] = \alpha([X]_t - I_{t-s}) + (1 - \alpha)[S]_{t-1},$$

$$I_t = \delta(X_{c,t} - S_{c,t}) + (1 - \delta)I_{t-s},$$

$$[\hat{X}]_{t+1} = [S]_t + I_{t-s+1},$$

where $\alpha, \delta \in [0, 1]$ and s is the length of the seasonality. The initialization of the model requires a whole season (i.e. the first s periods) and it can be done as follows: $[S]_s = \frac{[X]_1 + ... + [X]_s}{s}$ and $I_1 = X_{C,1} - S_{C,s}$, $I_2 = X_{C,2} - S_{C,s},...,$ $I_s = X_{C,s} - S_{C,s}$.

In the second approach, the level, $[S]_t$, and the seasonality, $[I]_t$, are represented as intervals. In this case, the level is no deseasonalized and both components are independently smoothed in separate equations. The forecast is obtained as the weighted addition of the level and the seasonality interval, where the weight controls the importance of each component. The interval exponential smoothing method with additive interval seasonality (IESiS) is shown next:

$$[S]_t = \alpha[X]_t + (1 - \alpha)[S]_{t-1},$$

$$[I]_t = \delta[I]_t + (1 - \delta)[I]_{t-s},$$

$$[\hat{X}]_{t+1} = \xi[S]_t + (1 - \xi)[I]_{t-s+1},$$

where $\alpha, \delta, \xi \in [0, 1]$, and s is the length of the seasonality. As in the previous approach, the first s periods of the HTS are needed to initialize the model. The initialization of the seasonal component can be done as follows $[I]_1 = [X]_1$, $[I]_2 = [X]_2$, ..., $[I]_s = [X]_s$; while the initial value of the level component can be obtained as $[S]_1 = [X]_1$ for $t = 1$ and applying the smoothing equation to obtain the level for $t = 2, .., s$.

6 Other approaches to forecast ITS

The simplest way to forecast an ITS is the naive model, although its forecasting ability is limited. Letting aside this model, the most straightforward approach consists of transforming the ITS in a pair of classical time series (center and radius, or minimum and maximum) and modelling each of them with an univariate forecasting model or both of them with a multivariate model. This approach has the objection that it can produce wrong intervals (e.g. intervals where $X_L > X_U$ or where $X_C < 0$) as it does not deal with intervals as a whole. On the other hand, interval attributes allow to focus in the features that characterizes the intervals. We consider that the center-radius approach is especially interesting as the center shows the interval location and the radius shows the interval span.

Classical multilayer perceptrons are commonly applied in time series forecasting prediction [6]. In a similar way, the Interval Multilayer Perceptron (iMLP) [7] can be applied to forecast ITS. The iMLP adapts the classical Multilayer Perceptron structure [8] in order to operate on interval-valued input and output data. Thus, it allows to deal with variability (in interval form) in the data set.

Other perceptrons dealing, in some manner, with intervals have been proposed, for example, in [9] and [10]. The one proposed by Beheshti et al. [10] is the more similar to the iMLP, as it has inputs and outputs in interval form. However, in the Behesti perceptron the weights and biases are intervals, whereas, in the iMLP, they are crisp values. Consequently, the estimation of the optimal weights and biases in the Behesti perceptron is done by means of interval computational algorithms and is substantially more complex than the calibration of the iMLP. Thus, the iMLP will be considered to forecast ITS and will be described below.

6.1 The Interval Multilayer Perceptron

An iMLP with n inputs and m outputs is comprised of an input layer with n input buffer units, one or more hidden layers with a non-fixed number of nonlinear hidden units and one output layer with m linear or nonlinear output units. Henceforth, we will consider just one hidden layer with h hidden units and one output ($m = 1$). The operations in the iMLP follow the rules of interval arithmetic (see section 3).

Considering n interval-valued inputs $[X]_i = \langle X_{i,C}, X_{i,R} \rangle = [X_{i,C} - X_{i,R}, X_{i,C} + X_{i,R}]$, with $i = 1, .., n$, the output of the j-th hidden unit is the weighted linear combination of the n interval inputs and the bias. It is worth noting that

the weights of the proposed structure are crisp and not intervals. The linear combination results in a new interval given by:

$$[S]_j = w_{j0} + \sum_{i=1}^{n} w_{ji}[X]_i = \langle w_{j0} + \sum_{i=1}^{n} w_{ji}X_{i,C}, \sum_{i=1}^{n} |w_{ji}|X_{i,R} \rangle.$$

The activation of the j-th hidden unit is obtained by transforming the interval $[S]_j$ using a nonlinear activation function, more precisely, the tanh function. This function is monotonic, then the interval output is given by $f([A]) = [f(A_L), f(A_U)]$. Thus, the resulting interval can be calculated as:

$$
\begin{aligned}
[A]_j &= \tanh([S]_j) = [\tanh(S_{j,C} - S_{j,R}), \tanh(S_{j,C} + S_{j,R})] = \\
&= \langle \frac{\tanh(S_{j,C} - S_{j,R}) + \tanh(S_{j,C} + S_{j,R})}{2}, \\
&\quad \frac{\tanh(S_{j,C} + S_{j,R}) - \tanh(S_{j,C} - S_{j,R})}{2} \rangle.
\end{aligned}
$$

The output of the network, $[\hat{Z}]$, is obtained by transforming the activations of the hidden units using a second layer of processing units. In the case of a single output and a linear activation function, the estimated output interval is a linear combination of the activations of the hidden layer and the bias:

$$[\hat{Z}] = \sum_{j=1}^{h} \alpha_j [A]_j + \alpha_0 = \langle \sum_{j=1}^{h} \alpha_j A_{j,C} + \alpha_0, \sum_{j=1}^{h} |\alpha_j| A_{j,R} \rangle.$$

The iMLP can be used to approximate an interval-valued function. The iMLP crisp weights can be adjusted with a supervised learning procedure by minimizing an error function of the form:

$$E = \frac{1}{p} \sum_{t=1}^{p} d([Z]_t, [\hat{Z}]_t) + \lambda \Phi(\hat{f}),$$

where $d([Z]_t, [\hat{Z}]_t)$ is a measure of the discrepancy between the actual and the estimated output intervals for the t-th training sample with $t = 1, ..., p$; and $\lambda \Phi(\hat{f})$ is a regularization term [11] of the estimated function $\hat{f}([X]_i) : [X]_i \rightarrow [Z]$ with $i = 1, ..., n$. A weighted Euclidean distance function for a pair of intervals $[A]$ and $[B]$ can be used as discrepancy measure:

$$d([A], [B]) = \beta(A_C - B_C)^2 + (1 - \beta)(A_R - B_R)^2.$$

The parameter $\beta \in [0, 1]$ allows to assign more weight to the error in the centers or in the radii. This discrepancy function can be minimized applying a low-memory Quasi Newton method [12] with random initial weights. Second order methods require the calculation of the gradient of the cost function with respect to the adaptive weights ($w's$ and $\alpha's$). These derivatives can be calculated in an

effective way by applying a backpropagation procedure, similar to this proposed in [8] for the standard MLP. More details are given in [7].

Due to the capability of input-output mapping of the iMLP, it can be used for causal forecasting of ITS or for extrapolative ITS forecasting. In the second case, the functional relationship to be estimated by an iMLP can be written as $[X]_{t+1} = f([X]_t, [\hat{X}]_{t-1}, [X]_{t-l})$, where $[X]_t$ is the interval observed at time t.

7 Analysis of the forecasting performance

The original data set consists of records of the monthly mean temperature in 60 weather stations in China from January 1952 to December 1988 (i.e. 444 months). These stations make up a network with a relatively uniform spatial distribution and each one is representative of a particular climate region of China. Data can be obtained in the archive of the Computational and Information Systems Laboratory (http://dss.ucar.edu/datasets/ds578.5/data/). The 60 temperature time series has been aggregated leading to an ITS of 444 monthly periods, where each period represents the interval of the monthly mean temperature throughout China. Figure 1 shows that the ITS has a seasonal pattern with $s = 12$ that concerns not only the interval centers, but also the ranges; it is clear that summer months have less range than winter months.
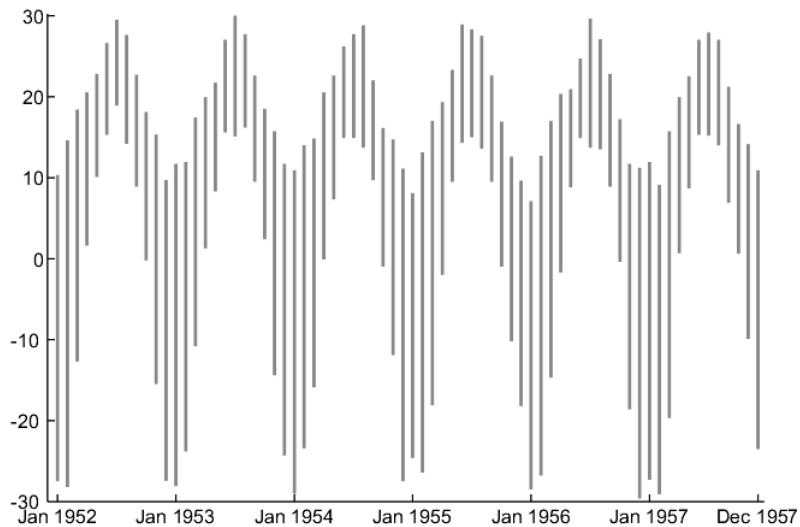

Fig. 1: ITS of the monthly temperature in China (1952-1957).

The training set consists of the first 296 periods, while the test set contains the last 148 periods. The ITS has been forecasted using different approaches:

1. naive model with seasonality: $[\hat{X}]_{t+1} = [X]_{t-s+1}$

2. IES with crisp seasonality: $\alpha = 1$, $\delta = 0.93$

3. IES with interval seasonality: $\alpha = 0.17$, $\delta = 0.19$ and $\xi = 0.03$

4. iMLP with 3 layers ($n = 6, h = 6, m = 1$) and using as inputs $[X]_{t-l+1}$, where $l = \{1, 2, 12, 13, 24, 25\}$,

5. modelling separately the minimum and maximum series:

    (a) exponential smoothing models with additive level and seasonality
       - minimum: $\alpha = 0.101, \delta = 0.1075$
       - maximum: $\alpha = 0.028, \delta = 0.16$
    (b) ARIMA models
       - minimum: ARIMA $(1, 0, 1)(2, 1, 1)_{12}$ without constant
       - maximum: ARIMA $(1, 0, 0)(0, 1, 1)_{12}$ with constant

6. modelling separately the centers and radii series with:

    (a) exponential smoothing models with additive level and seasonality
       - centers: $\alpha = 0.0865, \delta = 0.1274$
       - radii: $\alpha = 0.0602, \delta = 0.1254$
    (b) ARIMA models
       - centers: ARIMA $(1, 0, 1)(2, 1, 1)_{12}$ without constant
       - radii: ARIMA $(1, 0, 0)(0, 1, 1)_{12}$ with constant

Table 1 summarizes the forecasting performance of the considered approaches. The IEScS obtains a forecasting performance worst than the seasonal naive; this is due to the fact that the ITS seasonality affects both, interval range and interval center, and not only centers as the IEScS assumes. The iMLP outperforms the seasonal naive model and the IEScS, but it is less accurate than the rest of the methods. Modelling the univariate series with ARIMA models is the best method in this case. The IESiS obtains a good result, especially, if we consider that it only requires 3 parameters instead of the 8 needed by the ARIMA based approaches. The performance of the IESiS is quite similar to the performance of the models that forecast the univariate series with exponential smoothing methods, but our method is slightly simpler and deals with intervals as a whole.

8   Conclusions

ITS provide a way of modelling the range variation of an observed phenomenon through time. The proposal of methods to forecast and to analyze ITS is an interesting challenge. In this paper, an extension of exponential smoothing methods to ITS has been proposed and the iMLP has been applied to ITS forecasting. The forecasting performance of these methods is promising but it must be improved in the future. We believe that forecasting methods for ITS must deal with intervals as a whole. Therefore, more sophisticated ITS forecasting methods should be proposed. Teles and Brito [1] adapted ARMA models to ITS, a comparison of the accuracy of exponential smoothing and ARMA models in ITS must be done in the future.

model	training	test
1- seasonal naive	2.348	2.432
2- IEScS	3.675	3.45
3- IESiS	1.88	1.729
4- iMLP	2.08	2.07
5a- min-max (exp. smooth.)	1.856	1.703
5b- min-max (ARIMA)	1.508	1.554
6a- cen-rad (exp. smooth.)	1.841	1.669
6b- cen-rad (ARIMA)	1.553	1.539

Table 1: Forecasting performance in terms of the MDE_H.

References

[1] P. Teles and M. P. Brito. Modelling interval time series data. In *3rd IASC world conference on Computational Statistics & Data Analysis*, Limassol, Cyprus, 2005.

[2] J. Arroyo and C. Maté. Introducing interval time series: Accuracy measures. In *COMPSTAT 2006, Proceedings in Computational Statistics*, pages 1139–1146, Heidelberg. Physica-Verlag.

[3] H.-H. Bock and E. Diday, editors. *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin Heidelberg, first edition, 2000.

[4] R.. E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.

[5] E. S. Gardner. Exponential smoothing: The state of the art. part 2. *International Journal of Forecasting*, 22(4):637–666, 2006.

[6] G. Zhang, B. E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks: The state of art. *International Journal of Forecasting*, 14(1):35–62, 1998.

[7] A. Muñoz San Roque, C. Maté, J. Arroyo, and A. Sarabia. imlp: Applying multi-layer perceptrons to interval-valued data. Technical report (submitted to neural processing letters, july 2006), Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, 2006.

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning internal representations by error propagation*, pages 318–362. MIT Press, Cambridge, MA, USA, 1986.

[9] H. Ishibuchi, H. Tanaka, and H. Okada. An architecture of neural networks with interval weights and its application to fuzzy regression analysis. *Fuzzy Sets and Systems*, 57:27–39, 1993.

[10] M. Beheshti, A. Berrached, A. de Korvin, C. Hu, and O. Sirisaengtaksin. On interval weighted three-layer neural networks. In *Annual Simulation Symposium*, pages 188–195. IEEE Computer Society, 1998.

[11] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, 1995.

[12] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, MA, 1984.