# An Approach to Qualitative Radial Basis Function Networks over Orders of Magnitude

**Xari Rovira** [*] **and Núria Agell** [*]
ESADE, Universitat Ramon Llull
Av. Pedralbes, 62
08034 Barcelona (Spain)
{agell, rovira}@esade.edu

**Mónica Sánchez**[*]**, Francesc Prats**[*] **and Xavier Parra**[*]
Universitat Politècnica de Catalunya
Pau Gargallo, 5
08028 Barcelona (Spain)
{monica.sanchez, francesc.prats, xavier.parra}@upc.es

**Abstract**

This paper lies within the domain of supervised learning algorithms based on neural networks whose architecture corresponds to radial basis functions. A methodology to use RBF when the descriptors of the patterns are given by means of their orders of magnitude is developed. A qualitative distance is constructed over the discrete structure of absolute orders of magnitude spaces. This distance is reliant on a metric structure defined in $R^n$. The aim is to capture the remoteness between the components of the patterns by locating labels with respect to extreme magnitudes. An application to a financial problem of the described learning method is given and it permits to compare results obtained from a qualitative treatment with those from a quantitative treatment.

**Keywords:** Learning Algorithms, Radial Basis Functions, Orders of Magnitude Reasoning.

## Introduction

Qualitative Reasoning has developed techniques that permit to formalize the human capabilities to perceive, analyze, understand and model real problems. In addition in the process of understanding the world around us, measurement is crucial. The measurement of reality needs, at the same time, the use of numbers, orders of magnitude and categories. One of the goals of qualitative reasoning is obtaining models allowing the treatment of problems in which variables described in different scales appear.

On the other hand, learning processes can be considered as one of the main human reasoning capabilities.

The majority of learning algorithms, fundamental subjects for AI, are based uniquely on numerical scales, whereas in human learning also non numerical variables take part, both qualitative variables described by orders of magnitude and attributes.

In this sense, the process followed by the rating agencies analysts to classify bonds issued by a company according to their credit risk is a good example of how reasoning and learning are used in a human qualified assessment.

In this paper it is shown that the learning capability of a neural network, specifically Radial Basis Function Networks (RBF), is improved when using qualitative information.

In the improvement of RBF's performance special attention is given to the qualitativization of the domain of activation functions. Orders of magnitude models are introduced in the development of this non linear type of algorithm and the selection of the stopping criteria is discussed. All these purposes are conducted to be applied to the financial problem of the determination of a credit risk prediction measure. For this problem the orders of magnitude and tendencies of the variables involved are considered more relevant than their exact numerical values [2].

Section 2 is devoted to homogenizing scales via a discretization that takes into account the qualitative categorical variables. To this aim the absolute orders of

[*] GREC- Knowledge Engineering Research Group

magnitude model with granularity n, OM(n), constructed via a symmetric partition of the real line, is considered and a location function and a distance in OM(n) are defined. Section 3 gives the basic concepts of RBF networks, highlights the importance of stepwise activation functions for these kinds of learning algorithms, and describes the forward selection algorithm to determine the centers for the treatment of qualitative information. In Section 4, an example of the proposed method is given involving some companies and their financial ratios, and results obtained from quantitative and qualitative methods are compared. The paper ends with several conclusions and outlines some proposals for future research.

## The discretization process

In classification processes the situation in which the numerical values of some of the data are unknown, and only their qualitative descriptions are available - given by their absolute or relative orders of magnitude [13] - is not unusual. In other situations, the numerical values, even though they might be available, are not relevant for solving the proposed problem. In that sense, in problems in which there are involved either quantitative or categorical variables, the values of a given numerical variable have frequently different meanings depending on the categories of an attribute. Regarding the financial problem of a firm's credit risk measurement, a particular value of, for instance, the variable leverage has different significances depending on the country in which the firm operates.

In this section a discretization of quantitative variables that permits to unify the measurement scales considered is presented.

This process is carried out by considering different landmarks for a fixed variable, in relation with the chosen category of an attribute, i.e., a Bayesian point of view of the discretization process is performed.

To this end, a supervised discretization method based on the CAIM Method (Class-Attribute Interdependence Maximization) is used [6], [9]. The CAIM Method is one of the most recently proposed discretization methods and it provides interesting results in a reduced computational time. The method divides the range of a variable into a very small number of intervals that can be found automatically. Nevertheless, in this work a variant of the method will be used in which the number of intervals will be previously fixed by the user [5].

The input variable F and the output variable C can be considered as random variables and their contingency table (quanta matrix) can be seen in Table 1.

| Class | Interval | | | | | Class Total |
|---|---|---|---|---|---|---|
| | $[d_0, d_1]$ | … | $(d_{r-1}, d_r]$ | … | $(d_{n-1}, d_n]$ | |
| $C_1$ | $q_{11}$ | … | $q_{1r}$ | … | $q_{1n}$ | $M_{1+}$ |
| : | : | … | : | … | : | : |
| $C_i$ | $q_{i1}$ | … | $q_{ir}$ | … | $q_{in}$ | $M_{i+}$ |
| : | : | … | : | … | : | : |
| $C_S$ | $q_{S1}$ | … | $q_{Sr}$ | … | $q_{Sn}$ | $M_{S+}$ |
| Interval Total | $M_{+1}$ | … | $M_{+r}$ | … | $M_{+n}$ | M |

**Table 1:** Quanta Matrix

From the quanta matrix in [5] the following coefficient is considered:

$$CAIM^*(C,D|F)=\sum_{r=1}^{n}\frac{\max_r^2}{M_{+r}}.$$

Where: D stands for a particular discretization scheme on F, given by a set of n landmarks, n is the number of discrete intervals, r iterates through all intervals, $\max_r$ is the maximum value within the $r^{th}$ column of the quanta matrix and $M_{+r}$ is the total number of continuous values of attribute F that are within the $r^{th}$ interval.

The algorithm chosen consists of maximizing the coefficient CAIM* by considering a fixed number n of intervals.

**Unifying scales by discretization**

Discretization is a process of converting the range of a continuous variable into a finite set of intervals. This allows to generate a variable with a smaller number of distinct values or ordered labels by determining a finite number of landmarks to establish the intervals.

Within the frame of Artificial Intelligence, a key factor is extracting information from heterogeneous data. Considering variables described via a common set of orders of magnitude can be helpful to manage them simultaneously. Qualitative orders of magnitude reasoning tackle problems in such a way that the principle of relevance is preserved [7]; that is to say, each variable involved in a real problem is valued with the required level of precision.

Given a categorical variable C, whose range is $\{C_1, \ldots, C_S\}$, and a quantitative variable V, it is considered the discretization of V conditioned by the category $C_j$, $V(C_j)$, by a set of landmarks. The set of landmarks and the corresponding labels to distinguish the ordered classes, can be either defined by experts [2] or automatically generated by a discretization algorithm [6], [9].

After this process, discretized quantitative variables will be defined in a common orders of magnitude space with granularity n, OM(n).

The *absolute orders of magnitude models* [11], [13], work with a finite set of symbols or qualitative labels obtained via a partition of the real line, where any element of the partition is a basic label. These models provide a mathematical structure which unifies sign algebra and interval algebra through a continuum of qualitative structures built from the rougher to the finest partition of the real line. This mathematical structure, the Qualitative Algebras or Q-Algebras, have been studied in depth [1].

In this section the absolute orders of magnitude model is briefly described [1], and a new function to determine the location of labels is defined. The model used is a generalization of the model introduced in [11].

The absolute orders of magnitude model of granularity n, OM(n), is defined from a symmetric partition of the real line in 2n+1 classes:

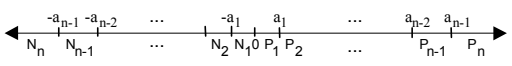


**Fig.1:** Partition of the real line

where $N_i=[-a_i,-a_{i-1})$, $0=\{0\}$ and $P_i=(a_{i-1},a_i]$.

Each class is named *basic description or basic element,* and, using the notation introduced in [1], is represented by a label of the set $S_1$:

$$S_1=\{N_n, N_{n-1}, ..., N_1, 0, P_1, ..., P_{n-1}, P_n\}.$$

*The quantity space S* is the set of labels in the form [X,Y] for all $X,Y\in S_1$, with $X<Y$ (i.e., $x<y$ for all $x\in X$ and $y\in Y$):

$$[X,Y]=\begin{cases} X, & \text{if } Y=0; \\ Y, & \text{if } X=0; \\ \text{the smallest interval with respect} \\ \text{to inclusion containing X and Y,} & \text{if } X\neq 0 \text{ and } Y\neq 0. \end{cases}$$

The relation $\leq_P$, *to be more precise than* (given $X,Y\in S$, X is more precise than Y ($X\leq_P Y$) if $X\subseteq Y$) is an order relation in S.

For all $X\in S-\{0\}$, the *basis of X* is the set:

$$B_X=\{B\in S_1-\{0\}: B\leq_p X\},$$

and, given a basic element $U\in S_1$, the *U-expansion* of X is:

$$X_U=\text{Min}\{Y\in S: X\leq_P Y \text{ and } U\leq_P Y\},$$

the minimum label that is *less precise* than X and U, i.e., the smallest interval with respect to the inclusion containing X and U.

Note that $X_U$ does not depend on the values of the landmarks used to determine the real line partition.

**The location function**

In order to define a qualitative RBF, a distance between qualitative vectors will be considered. To this end, due to the lack of good properties of the qualitative difference in an OM(n) structure, a function that measures the relative position between patterns by means of a distance in $R^n$ is defined.

Each element X in S will be codified by a pair ($l_1(X)$, $l_2(X)$) of integers: $l_1(X)$ is the number of basic elements in $S_1-\{0\}$ that are "between" the basis of X and $N_n$, and $l_2(X)$ is the number of basic elements in $S_1-\{0\}$ that are "between" the basis of X and $P_n$.

These numbers permit each element to be "located" in S, where all different levels of precision are considered. These two numbers are necessary for the location to capture the particular level of precision of each element. The addition of $l_1(X)$ and $l_2(X)$ allows to obtain the number of basic elements contained in X.

In addition, this "location" can be extended to any pattern defined by k orders of magnitude variables.

As an example to illustrate this process, let us consider the absolute orders of magnitude model with granularity 3, OM(3):

$$S_1=\{N_3, N_2, N_1, 0, P_1, P_2, P_3\},$$

and

$S=S_1\cup\{[N_3,N_2],[N_2,N_1],[N_1,P_1],[P_1,P_2],[P_2,P_3],[N_3,N_1],$
$[N_2,P_1],[N_1,P_2],[P_1,P_3],[N_3,P_1],[N_2,P_2],[N_1,P_3],[N_3,P_2],$
$[N_2,P_3],?\},$

where "?" stands for $[N_3,P_3]$.

The location of the label $X=[N_1,P_2]$ is the pair (-2,1), because there are left two basic elements at the left hand side of X and only one at its right hand side:

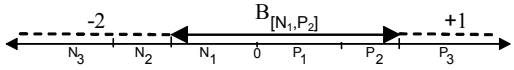


**Fig. 2:** The location function

The formal definition of the *location function* is $l: S\rightarrow Z^2$ such that:

$$l(X)=(l_1(X),l_2(X))=$$
$$(-\text{Card}(B_{X_{N_n}})+\text{Card}(B_X),\text{Card}(B_{X_{P_n}})-\text{Card}(B_X)).$$

This is a way to codify labels by points on a Euclidean plane, in such a manner that the Euclidean distance between them will allow a distance between labels to be defined.

The extension of l to $S^k$ is:

$$L: S^k \rightarrow Z^{2k}$$
$$\mathbf{X}=(X_1,\ldots,X_k)\rightarrow L(\mathbf{X})=(l(X_1),\ldots,l(X_k))=$$
$$=(l_1(X_1),l_2(X_1),\ldots,l_1(X_k),l_2(X_k))$$

This function provides the relative position of a k-tuple of qualitative labels with respect to the basis of $S^k$.

**Distance in $OM(n)^k$**

The location function allows similarity between patterns to be measured. Once the k-tuples of qualitative labels have been codified by 2k-tuples of integers, conditions to define distances in the space $OM(n)^k$ are achieved. Let us define:

$$D: S^k\times S^k\rightarrow[0,+\infty)$$
$$(\mathbf{X},\mathbf{Y})\rightarrow\sqrt{(L(\mathbf{X})-L(\mathbf{Y}))^T R\,(L(\mathbf{X})-L(\mathbf{Y}))}$$

where R represents any metric in $R^{2k}$, that is to say, if $\mathbf{X}=(X_1,\ldots,X_k)$ and $\mathbf{Y}=(Y_1,\ldots,Y_k)$:

$$D(\mathbf{X},\mathbf{Y})=$$
$$\sqrt{(l_1(X_1)-l_1(Y_1),\ldots,l_2(X_k)-l_2(Y_k))^T R\,(l_1(X_1)-l_1(Y_1),\ldots,l_2(X_k)-l_2(Y_k))}$$

This function D inherits all properties of the distance in $R^{2k}$, and therefore satisfies the three axioms of a distance.

The distance D between two k-tuples of qualitative labels measures the similarity between them, in the sense that the more similar labels are, the smaller the distance between their codifications is, and so the smaller their distance is.

## Qualitative radial functions in RBF

In this section a methodology, allowing Radial Basis Function networks (RBF) to be used with orders of magnitude input data, is proposed.

Before building appropriate radial functions for this kind of discrete spaces, let us remind ourselves of the basic concepts of Radial Basic Function networks, introduced by Broomhead and Lowe [4].

**Radial basis function networks**

RBF are a type of artificial neural network for application to problems of supervised learning. RBF can be applied to problems of regression, classification and times series prediction. In this paper only the case of classification is considered.

Several reasons make RBF especially interesting. On the one hand, they are universal classifiers, and on the other, the training process associated to these kind of neural networks is usually much faster than for other neural architectures, such as MLP or SVM, and, in addition, it is possible to extract rules from RBF architecture.

RBF are associated with a simple architecture of three layers [4] as in Fig 1:

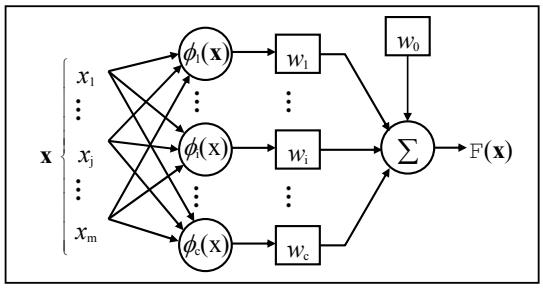


**Fig. 3:** Radial basis function network architecture

Each layer is fully connected to the following one and the hidden layer is composed of a number of nodes with radial activation functions called radial basis functions. Each of the input components fits forward to the radial functions, whereas the outputs of these functions are linearly combined with a set of weights into the output of the network.

The characteristic feature of radial functions is that their response decreases, or increases, monotonically with distance from a central point named center of the radial function. These functions involve two parameters, the center and a distance scale.

Radial functions in the hidden layer have a structure represented as follows:

$$\Phi_i(\mathbf{x}) = \varphi((\mathbf{x}-\mathbf{c_i})^T R\,(\mathbf{x}-\mathbf{c_i}))$$

where φ is the radial function used, $\{\mathbf{c}_i \mid i=1,\ldots,c\}$ is the set of radial function centers and **R** is a metric. So, the term $(\mathbf{x}-\mathbf{c})^T\mathbf{R}\,(\mathbf{x}-\mathbf{c})$ denotes the square of the distance between the input **x** and the center **c** according to the metric defined by **R**.

Usual radial functions are Gaussian, Cauchy's, multiquadric, inverse multiquadric, spline and logarithm. The most widely used radial function is the Gaussian one, in the particular case when the metric is $\mathbf{R} = \frac{1}{r^2}\mathbf{I}$, with **I** the identity matrix and r the radius of the radial function:

$$\Phi(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x}-\mathbf{c})^T(\mathbf{x}-\mathbf{c})}{r^2}\right) = \exp\left(-\frac{d^2(\mathbf{x},\mathbf{c})}{r^2}\right)$$

where d is the Euclidean distance in $R^n$. Gaussian function monotonically decreases with distance from the centre **c**, and in this sense, it is said that its response is local. The output of the RBF network is:

$$F(\mathbf{x}) = w_0 + \sum_{i=1}^{c} w_i \exp\left(-\frac{d^2(\mathbf{x},\mathbf{c_i})}{r^2}\right)$$

**Qualitative radial functions**

Note that all the formulae in previous subsection are applied to the case in which the input **x** and the centers **c** are numerical vectors.

When input data are described by their orders of magnitude, a new kind of radial functions, with a qualitative domain, have to be defined. Then, since both patterns and centers are described by their orders of magnitude in a space OM(n), the distances between them will be measured by using the definition previously introduced.

Radial functions in the hidden layer become:

$$\Phi_i(\mathbf{x}) = \varphi((L(\mathbf{x})-L(\mathbf{c_i}))^T R\,(L(\mathbf{x})-L(\mathbf{c_i}))),$$

and the output of the RBF network is:

$$F(\mathbf{x}) = w_0 + \sum_{i=1}^{c} w_i \varphi\left((L(\mathbf{x})-L(\mathbf{c_i}))^T R\,(L(\mathbf{x})-L(\mathbf{c_i}))\right),$$

where φ is the radial function used.

As in the numerical case, the response of these radial functions decreases, or increases, monotonically with distance from the center. Qualitative radial functions, defined via the distance between the qualitative descriptions of the inputs, can be graphically represented by means of stepwise functions.

As a first example, let us consider a space OM(3) with basic labels $\{N_3,N_2,N_1,0,P_1,P_2,P_3\}$, patterns described by only one input variable, and four centers with the four corresponding Gaussian radial basis functions with radius equal to 1. The centers are $c_1 = P_2$, $c_2 = [N_2,N_1]$, $c_3 = [N_2, P_1]$, and $c_4 = N_3$. The explicit expression of all radial basis functions is then:

$$\Phi_i(x) = e^{-D^2(x,c_i)}$$

By taking weights $w_0= 0$, $w_1= 1$, $w_2 = -3$, $w_3= 2$ and $w_4= -1$, the output of the qualitative RBF network is:

$$F(x) = e^{-D^2(x,c_1)} - 3e^{-D^2(x,c_2)} + 2e^{-D^2(x,c_3)} - e^{-D^2(x,c_4)}$$

Fig. 4 shows the stepwise graphical representation of the output of this qualitative RBF:

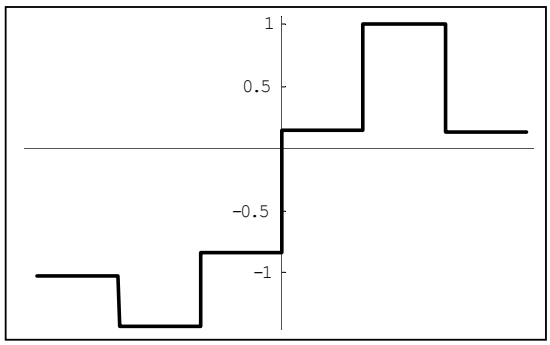


**Fig.4:** The output of a one-dimensional qualitative Gaussian RBF

Let us consider now an example in the case in which patterns are described by two input qualitative variables, again in a space OM(3) with basic labels $\{N_3,N_2,N_1,0,P_1,P_2,P_3\}$, only one center $\mathbf{c} = (P_1,P_2)$, and the corresponding radial basis function with a Gaussian expression with radius equal to 1. The explicit expression of the radial basis function is then:

$$\Phi(\mathbf{x}) = e^{-D^2(\mathbf{x},\mathbf{c})}$$

By taking weights $w_0$= 0.25 and $w_1$= 1, the output of the obtained qualitative RBF network is:

$$F(\mathbf{x}) = 0.25 + e^{-D^2(\mathbf{x},\mathbf{c})}$$

Fig. 5 shows the stepwise graphical representation of the output of this two-dimensional qualitative RBF:

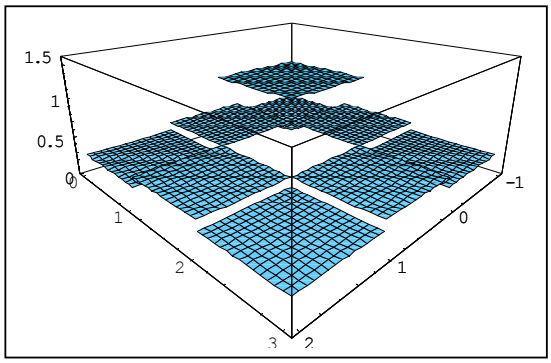


**Fig.5:** The output of a two-dimensional qualitative Gaussian RBF

**Forward selection algorithm to determine centers to a qualitative RBF**

There are several methods to determine the parameters for a RBF [4]; it is the set of centers for the basis functions within the entire set of patterns, their radii and the weights. Standard ways of optimizing networks involve the optimization by gradient descent, which needs differentiability and a fixed architecture in advance (number of layers and nodes). Nevertheless RBF has a simpler alternative based in subset selection. Subset selection compares the errors produced when different subsets of centers are chosen. Usually the entire set of patterns has too many subsets, so a heuristic method must be implemented. One of these is the forward selection method. It consists in starting with an empty subset, to which is added one center at a time, the one that most reduces the approximation error. Forward selection ends when some chosen criterion, such as Generalized Cross Validation, stops decreasing. In forward selection method, weights are not selected, because they are directly determined by the chosen centers and radii.

The forward selection advantages in front of standard ways of optimizing networks are that it does not need a fixed number of centers in advance, and, moreover, it is a model easily tractable in the case of qualitative input data, and its computational requirements are lower.

In addition, forward selection does not require differentiability, and this is crucial when considering a qualitative patterns description by means of orders of magnitude. In such a discrete structure the radial functions are neither continuous nor differentiable.

The location function and the distance in $OM(n)^k$ allow forward selection algorithm and generalized cross validation criterion to be implemented in qualitative radial basis functions networks over orders of magnitude spaces.

## Experiments and results

In this section the defined method is used to classify firms into investment and speculative grades in an example involving some companies and their financial ratios. Results obtained from quantitative and qualitative methods are compared. To this end 638 U. S. firms, defined by 21 of their financial ratios and classified by Standard & Poor's, have been considered. Financial data used in this work were calculated considering the financial statements presented by the companies at the end of 2001. Information has been collected from two databases, acquired by the MERITO project (supported Spanish Ministry of Science and Technology): The Thompson's WorldScope Database and Standard & Poor's Database.

First, let us introduce the problem of predicting financial risk of a firm by means of its financial rating.

**Predicting Financial Rating**

The rating is a qualified assessment about the risk of bonds issued by a company. The specialized rating agencies, such as Standard & Poor's, classify firms according to their level of risk, using both quantitative and qualitative information to assign ratings to issues. The final rating is the agency's judgment, and reflects the probability of issuer default. Predicting the rating of a firm therefore requires a thorough knowledge of the ratios and values that indicate the firm's situation and, also, a deep understanding of the relationships between them and the main factors that can alter these values.

The processes employed by these agencies are highly complex. Decision technologies involved are not based on purely numeric models. Experts use the information given by the financial data, as well as some qualitative variables, such as the industry and the country or countries where the firm operates, and, at the same time, they forecast the possibilities of the firm's growth, and its competitive position. Finally, they use an abstract global evaluation based on their own expertise to determine the rating. Standard & Poor's ratings are labeled AAA, AA, A, BBB, BB, B, CCC, CC, C and D. From left to right these rankings go from high to low credit quality, i.e., the high to low capacity of the firm to return debt.

Nevertheless, a first classification into two classes is derived from these labels. Under present commercial bank regulations, bonds rated in the top four categories ('AAA', 'AA','A', 'BBB') are generally regarded as eligible for bank investment, it is with *investment grade*, meanwhile debt obligation of an issuer with a lower than investment grade rating is considered with *speculative grade*. Standard & Poors criteria to define groups of risk can be found in [14].

The model presented is especially adequate when the goal is to measure the magnitude of a result, based on the qualitative descriptions of the variables that participate. The qualitative descriptions appear when either numerical values are unknown or the experts use only their orders of magnitude.

**Training methodology**

Supervised learning over the available data of the 638 companies has been implemented, using both the numerical

values of their ratios and their orders of magnitude, by a standard RBF network.

Variables employed in this study are extracted from [3] and [8].

All 21 input variables are real-valued while the rating, i.e. the output variable, is a nominal variable with 2 different classes: investment grade and speculative grade, which have been represented using -1 and +1, respectively.

Simulations have been carried out following the PROBEN1 standard rules [12]. The data set available has been sorted by the company name before partitioning it into three subsets: training set, validation set and test set, with a relation of 50%, 25% and 25% respectively. Table 1 shows the pattern distribution in each data subset.

| **Rating** | **Investment grade** | **Speculative grade** | **Total** |
|---|---|---|---|
| Training | 163 | 156 | 319 |
| Validation | 85 | 75 | 160 |
| Test | 77 | 82 | 159 |
| **Total** | **325** | **313** | **638** |

**Table 2:** Pattern distribution over data subsets

To study and analyze the effect that input data discretization has over RBF generalization, two different kinds of training have been done. First training (referred to as *quantitative* training) rescales input values to mean 0 and standard deviation 1. Second training (*qualitative* training) performs the input transformation described, i.e. it considers the discretization on orders of magnitude and the codification of the values given by the location function.

Data discretization has been carried out following the modification of CAIM methodology, with 6 qualitative descriptions: $\{N_3,N_2,N_1,0,P_1,P_2,P_3\}$. In order to determine the landmarks, for each input variable, only the values with a distance to the mean less than three times the standard deviation have been considered. It is necessary to remark that this has been made to avoid the outliers' strength only during the first step of the process. Nevertheless, the firms with extreme values not considered in the landmarks determination process, have been later qualitatively labeled by $N_3$ or $P_3$.

**Results and discussion**

In both kinds of training processes, networks are initially trained on the training set and tested on the validation set. The results are used to adjust the radial function width (r). To perform this adjustment of the radial width, a total of 4000 simulations have been done for each class. Widths checked are from 0.0001 to 0.1 with increments of 0.0001, from 0.101 to 1.1 with increments of 0.001, from 1.11 to 11.1 with increments of 0.01 and 11.2 to 111.1 with increments of 0.1.

The final width (see Table 3) is selected among the 4000 widths trained by applying a standard criterion [10, 10b].

Once the radial width is determined, networks are trained on training and validation sets while the test set is used to assess the generalization ability of the final solution.

| | *Quantitative* training | | | | |
|---|---|---|---|---|---|
| | *r* | $CA_{tr}$ | $CA_{va}$ | $CA_{tr+va}$ | $CA_{te}$ |
| | 12,5 | 77,43% | 75,62% | 76,83% | 74,84% |
| | *Qualitative* training | | | | |
| | *r* | $CA_{tr}$ | $CA_{va}$ | $CA_{tr+va}$ | $CA_{te}$ |
| | 17,8 | 85,27% | 78,12% | 83,92% | 78,62% |

**Table 3:** Classification accuracy for training data set ($CA_{tr}$), validation data set ($CA_{va}$), training + validation data set ($CA_{tr+va}$) and test data set ($CA_{te}$)

As can be seen in Table 3, classification accuracy for the qualitative training is better than for the quantitative training. Since the only difference is the use of qualitative labels during the input conditioning, it confirms experts' intuition when saying that orders of magnitude of the ratios are more relevant than their exact numerical values. Therefore, the use of qualitative values instead of real ones during training and test process not only does not carry a lost of information, but leads to better results, due to the fact that it implies working with the appropriate level of precision.

In table 4, a threshold of 0.25 has been imposed in the test, avoiding classification of the firms with activation function value between -0.25 and 0.25. It can be seen that in this case classification for the qualitative case is again more suitable.

The number of patterns correctly classified is more than 22% better for the qualitative training. At the same time, qualitative training has a lower indetermination in the classification (25.16% in front of the 36.48% of the quantitative training). The number of patterns incorrectly classified is the same in both tranings.

| | *Quantitative* training | | | | | |
|---|---|---|---|---|---|---|
| | Investment | | Speculative | | Global | |
| Correctly classified | 41 | 53,25% | 40 | 48,78% | 81 | 50,94% |
| Incorrectly classified | 4 | 5,19% | 16 | 19,51% | 20 | 12,58% |
| Not classified | 32 | 41,56% | 26 | 31,71% | 58 | 36,48% |
| | *Qualitative* training | | | | | |
| | Investment | | Speculative | | Global | |
| Correctly classified | 49 | 63,64% | 50 | 60,98% | 99 | 62,26% |
| Incorrectly classified | 6 | 7,79% | 14 | 17,07% | 20 | 12,58% |
| Not classified | 22 | 28,57% | 18 | 21,95% | 40 | 25,16% |

**Table 4:** Final classification for test data set (threshold = 0.25)

## Conclusions and future research

The present work aims at motivating, defining, and analyzing the viability of the use of artificial neural networks in structures defined in orders of magnitude spaces.

The focus of this paper is the construction of a radial basis function, necessary in RBF networks, to be used in problems for which the input variables are described in terms of qualitative values on orders of magnitude. For this reason radial basis functions have been built from a distance defined over OM(n) via a location function for qualitative labels.

It has been proved that, using variables defined on orders of magnitude, the network generalization is enhanced.

The system is applied in the financial domain to evaluate and simulate credit risk. But this approach may also be applicable to problems in other areas where the involved variables are described in terms of orders of magnitude. The limitations of the method presented cannot be evaluated until the implementation is completed and sufficiently tested.

The results obtained by using input variables defined over orders of magnitude spaces have been compared with the ones obtained by using numerical values in the example given. The good results obtained with a dataset of 638 U.S. firms from the industrial sector show the suitability of the presented methodology.

Although this paper has focused on a binary classification problem (i.e. classes considered are investment and speculative grade), the methodological aspects considered and given can be used in a multi-classification problem (considering the nine Standard & Poor's ratings from AAA to D).

This work belongs to a wider project, MERITO project, supported by the Spanish Ministry of Science and Technology, in which the methodology given in this paper is going to be used. The project addresses the prediction and measurement of financial credit risk. The proposed method is currently being implemented to be applied to available data referring to the most important American and European firms, whose Standard & Poor's rating is known.

With regard to future work and open problems, the following comments can be made:

- To look for other appropriate distances between qualitative descriptors.
- To define new distances combining numeric and qualitative data.
- To choose different criteria to select centers instead of forward selection.
- To test the new criteria to obtain landmarks by using the information given by categorical variables in a wider dataset with firms from different countries and sectors.
- To implement the given method to be applied in problems of multi-classification.

**Acknowledgements**

This work was supported by the MCyT (Spanish Ministry of Science and Technology) MERITO project (TIC2002-04371-C02).

**References**

[1] Agell, N. *Estructures matemàtiques per al model qualitatiu d'ordres de magnitud absoluts*. Ph. D. Thesis Universitat Politècnica de Catalunya, 1998.

[2] Agell, N. Rovira, X. Ansotegui, C., Sánchez, M. & Prats, F. Homogenysing references in Orders of Magnitude Spaces : An Application to Credit Risk Prediction. QR2000 Proc., pp. 1-8, 2000

[3] Altman, E.I. "Predicting Financial Distress of Companies: Revisiting the Z-score and ZETA® Models," http://pages.stern.nyu.edu/ealtman/Zscores.pdf. Working Paper Series. The Department of Finance New York University, 2000.

[4] Broomhead, D.S. and Lowe, D.: Multivariable Functional Interpolation and Adaptive Network, Complex Systems, Vol. 2, pp. 321-355, 1988.

[5] Campos, R., Ruiz, F., Agell, N & Angulo, C. Financial Credit Risk Measurement Prediction Using Innovative Soft-computing Techniques. To appear in International Conference on Computational Finance & its Applications. Bologna, Italy, 2004.

[6] Ching, J.Y., Wong, A.K.C. & Chan, K.C.C. Class-Dependent Discretization for Inductive Learning from Continuous and Mixed Mode Data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (7), pp. 641-651, 1995.

[7] Forbus, K. D. Commonsense physics. A: *Annals Revue of Computer Science*, pp. 197-232, 1988.

[8] Fryman, H., Altman, E.I. and Duen-Li, K. Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. Journal of Finance, Mar85, Vol. 40, Issue 1, pp. 269-292, 1985.

[9] Kurgan, L. & Cios, K.J. Discretization Algorithm that Uses Class-Attribute Interdependence Maximization, Proc. of the 2001 International Conference on Artificial Intelligence (ICAI-2001): Las Vegas, pp.980-987, 2001.

[10] Parra, X., Agell, N. & Rovira, X. Generalisation Improvement of Radial Basis Function Networks based on Qualitative Input Conditioning for Financial Credit Risk Prediction. Lecture Notes in Computer Science, Springer Verlag, pp 127-134, 2001.

[10b] Parra, X., Agell, N. Qualitative Input Conditioning to Enhance RBF Neural Networks Generalization in Financial Rating Classification. 15th International Workshop on Qualitative Reasoning QR2001. San Antonio, Texas,2001.

[11] Piera, N. Current Trends in Qualitative Reasoning and Applications. Monograph CIMNE, núm. 33. International Center for Numerical Methods in Engineering, 1995.

[12] Prechelt, L.: PROBEN1: Set of Neural Network Benchmark Problems and Benchmarking Rules. Technical Report 21/94, University of Karlsruhe (1994).

[13] Travé-Massuyès, L.; Dague, P. Modèles et raisonnements qualitatifs. Hermès, 2003.

[14] "Public Finance Criteria 2000", de Diciembre 20 de 2000 (en www.standardandpoors.com).