ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

Classification and Prediction in Highly Dimensional Spaces An Application to Tribology

Leonardo M. Reyneri
Politecnico di Torino - Dept. Electronics
e.mail leonardo.reyneri@polito.it

Abstract. This paper presents a few applications of neuro-fuzzy systems to tribology. Classification and prediction act on a highly dimensional input space, posing severe problems of generalization capability and reliability of results. The paper shows how the major problems have been solved for the specific application domain.

1. Introduction

Tribology (namely, the science and technology of *friction*) is a tough discipline, due both to the large number of parameters involved in technological processes and to the strongly non-linear influence of the various chemical components and technological parameters onto the characteristics and performance of final product(s) (for instance, car brakes). This work has been sponsored by ITT Galfer (I).

2. Problem Description

A typical manufacturing process of a *friction material* involves:

- mixing a large number of basic components (among which: different types of rubbers, different types of lubricants, different types of metals and a few other materials);
- processing the mixture through a number of technological steps (among which: pressing, heating, baking, cooling, etc.);
- giving the material the desired shape and testing.

The actual composition (that is, the percentage of each component) and the actual sequence of technological steps (for instance, time, pressure and temperature at each step) heavily affect performance of final product, with a heavily non-linear and mostly unknown relationship. The knowledge about this relationship is mostly acquired through direct experience and is usually traded among experts.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

As a consequence, finding the best composition is mostly left to the know-how of human experts and is based on a trial-and-error approach, which is rather time-consuming (manufacturing and testing of a new material may take up to several months), quite costly, error-prone and often misleading. Furthermore, the high dimensionality of the problem causes such a tremendously high size of input space, which is far from being explored completely.

2.1. Problem Encoding and State of the Art

A given friction material is encoded (that is, completely defined) with a 67-dimensional input pattern $\vec{X}'$, which contains the concentration of 25 basic components and 12 technological parameters, as described above. Performance of a material is expressed as a 2-dimensional output pattern $\vec{Y}$, which contains the friction coefficient and wear of the manufactured material (there are assessed after a two-days long testing).

At present, the development of a new material is based on a trial-and-error approach, which requires: i) to empirically define a new chemical composition and technological parameters (based on prior know-how and experience of a few human experts); ii) to manufacture the material and iii) test it according to a standardized procedure; iv) modify either chemical composition or process parameters with the aim of increasing performance; v) loop until requirements are met.

The major problem is that manufacturing and test is time consuming (about three days per each new composition), therefore costly, while the probability of success (that is, the chance that a given new composition meets the specifications) is rather low, therefore a large number of trials and tests has to be done, causing a very long time-to-market and high development costs.

In addition, one of the performance figures (namely, friction coefficient) is not a unique value, as friction heavily depends on working conditions and on the history of the material (namely, the sequence of past working conditions). In most cases, an average value is used, but this is far from identifying the performance unambiguously.

2.2. Aims of the Work

Aims of the activity were manifold:

1. to develop an accurate empirical model to estimate, with a reasonable confidence, the performance of a given material, without manufacturing it. This would significantly reduce the development time and cost of any new material, by substituting a three-days manufacture-and-test step with a few-minutes model estimation, provided that the reliability of the model is high enough;

2. to use and validate the model as an aiding tool in a decision-making process, so that the expert who is willing to try a new material may have a quick feedback about the estimated performance of the new compound;

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

3. to use the model in an automatic optimization process, to automatically find a composition which improves a given cost function (mostly, either performance figure, or both of them);

4. to extract knowledge from the model to understand the effects onto performance of each chemical component and process parameter and to interpret the physics which lays behind tribology;

5. to find a parametric model of friction as a function of working conditions, such that the model parameters can be used instead of the average value to identify the characteristics of the material more accurately. A side-effect of this would be to produce a more reliable identification of the material, which is a further step toward manufacturing quality;

6. to extract knowledge from the latter model, in order to predict the behavior of the material under critical working conditions;

7. to force standardization, therefore quality, of the manufacturing process, as it is required to have homogeneous and comparable process parameters and testing procedures.

3. Preprocessing

As a first step, input/output patterns are preprocessed to:

1. reduce dimensionality of input space, by manually gathering together "similar" chemical components into homogeneous groups (for instance, all *soft*, *medium* and *hard rubbers*) and summing up the concentration of components in the same group. A 19-dimensional reduced input vector $\vec{X}$ is obtained. In the preliminary study, a crisp grouping has been done manually, based on human experience and on reasonable assumptions. Later, an automatic process has led to a fuzzy grouping with better performance, as described in sect. 4.3..

   Dimensional reduction is often achieved by means of PCA, which is used to find the directions in input space where most patterns are located. Yet we have decided not to use PCA for dimensional reduction, as this has been used for other purposes (see sect. 5.1.) and a double use of it might cause undesired effects. We will try in a future work to apply PCA also for preprocessing;

2. the first step of component grouping has caused some input patterns to become similar to each other (for instance, two patterns with a different percentage of two components from the same group may have an identical sum of partial concentrations). Yet a number of similar (or identical) patterns would cause an undesired stronger effect of these on training. We have therefore decided to remove all similar patterns, by analyzing the relative distance from each other;

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

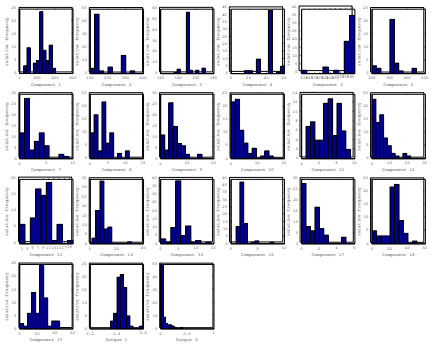

Figure 1: Frequency histogram of the 19 components of input vectors and the 2 components (last two plots) of output vectors vectors, both before preprocessing.

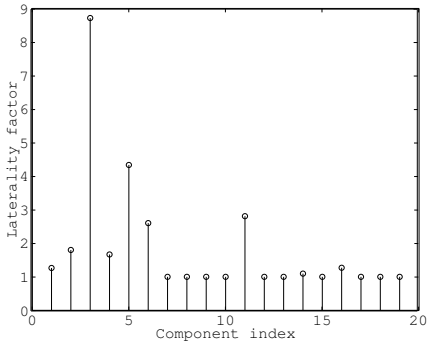

Figure 2: Laterality factor of the 19 components of input patterns.

3. removing outliers, which are likely due to defective measurements or anomalous process runs. Outliers are identified by verifying whether all components of input/output pattern lie within reasonable bounds. These bounds (called $x_{\mathrm{i,max}}$ and $x_{\mathrm{i,min}}$) have been determined by manually analyzing the frequency histograms of all components (see fig. 1);

4. computing, for each component of input/output vector, the *laterality factor* (see fig. 2):

$$\alpha_{\mathrm{L}i} = \frac{x_{\mathrm{i,max}} + x_{\mathrm{i,min}}}{x_{\mathrm{i,max}} - x_{\mathrm{i,min}}}$$

   which expresses how much each component of input and output patterns is distributed symmetrically around the origin. If values are distributed symmetrically around origin, then $\alpha_{\mathrm{L}i} \approx 0$ (we have no such cases), while if they are distributed between $x_{\mathrm{i,min}} = 0$ and an upper bound, then $\alpha_{\mathrm{L}i} = 1$ (most of our components), but if they are too biased (that is $x_{\mathrm{i,max}} \approx x_{\mathrm{i,min}}$, then $\alpha_{\mathrm{L}i} \to \infty$ (for instance, components 3, 5, 6, 11).

5. as all components are unilateral (that is, $\alpha_{\mathrm{L}i} \geq 1$), we normalize each component of input/output pattern to the range $[0, 1]$ (static normalization); bilateral components (that is, $\alpha_{\mathrm{L}i} \approx 0$) would rather be normalized to the range $[-1, 1]$;

6. removing immaterial input components, that is, all those with nearly null standard deviation (that is $\alpha_{\mathrm{L}i} \gg 10$; a few of them were present in our study, but these were removed *before* plotting fig. 1 and fig. 2).

By doing so, we were left with a set of 19-dimensional, normalized, unilateral, input pattern and the corresponding set of 2-dimensional, normalized, unilateral, output pattern.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

After preprocessing, the number of available patterns (about 67) was tremendously low compared with the dimensionality of input space (19!). This caused heavy criticalities in training and bounds to accuracy. A mixture of neural networks and fuzzy logic has been used to achieve the best results, even with such a limited number of training patterns, as described in next sections.

4. Modeling Friction and Wear

The first aim of this work has been to develop an empirical model of friction and wear versus chemical composition and technological parameters. We have used a $19 \times H \times 2$ two-layers neural network of order $n = 9$, Gaussian, on the first layer and order 0, linear, on the second layer [1], with a normalization layer in between. More precisely, an WRBF-9(Gaussian)-NORM-1/1-WRBF-0(linear), that is:

$$\hat{y_j} = f(\vec{X}) = \frac{\sum_k w_{jk}^2 h_k}{\sum_k h_k} \quad \text{where} \quad h_k = \mathrm{e}^{-\left(\sqrt[9]{\sum_i \left(w_{ki}^1 (x_i - c_{ki})\right)^9}\right)^2} \tag{1}$$

where w_{jh}^1 and w_{hi}^2 are the weights of the first and second layer, respectively, while c_{hi} and the centers of the first layer. Activation functions are Gaussian on the first layer and linear on the second layer, while the order n^1 of the first layer has been chosen equal to 9 to reduce the curse of dimensionality, as commented in sect. 4.4..

As the number of training patters was so low, we decided not to train the network (it would likely not be successful with so few patterns), but to use a *direct initialization*. We therefore:

1. centered one first-layer neuron on each training pattern (that is, as many hidden neurons as training patterns). Therefore:

   $$\underline{\underline{C}}^1 = \underline{\underline{X}}_S$$

   where $\underline{\underline{X}}_S$ is a matrix whose columns are all the training patterns;

2. first-layer weights were fixed: $w_{ki}^1 = w_0$;

3. the second-layer weights were initialized by means of linear optimization, as in RBF networks:

   $$\underline{\underline{W}}^2 = \underline{\underline{Y}}_S \cdot \underline{\underline{H}}_S{}^{-1}$$

   where $\underline{\underline{Y}}_S$ is a matrix whose columns are all the training output patters, while $\underline{\underline{H}}_S$ is a matrix whose columns are all the corresponding hidden vectors.

This builds up a non-linear interpolator, which interpolates output values in between the known values (training patterns). By applying a new pattern (that is, a new compound) at the input of the network, this tries to predict the corresponding values of friction and wear.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

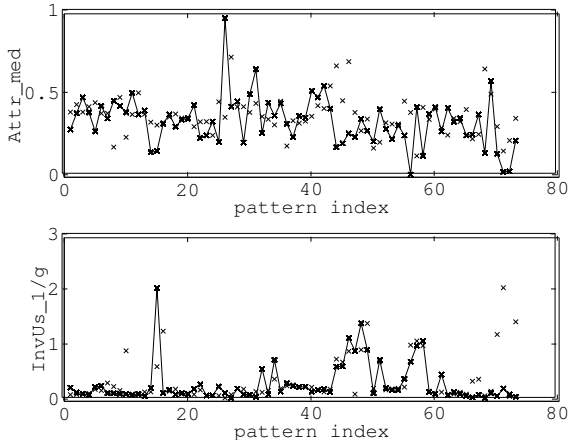

Figure 3: Results of the leave-one-out validation of neural model. Continuous line joins the 67 values in the training set, while the crosses show the estimated value of each left-out pattern.

4.1. Establishing Model Reliability

In the beginning, model reliability has been assessed with a leave-one-out approach, by: i) removing one pattern from the training set; ii) building the network based on the remaining patterns; iii) testing the network on the removed pattern, by computing the estimation error; iv) put back the removed pattern and remove next one; v) repeat for all patterns and average quadratically the corresponding estimation errors. Results are shown in fig. 3, while the average estimation error is shown in tab. 1 for different values of the first-layer weights w_0. It can be seen that $w_0 = 11$ is a reasonable compromise.

It is clear that most patterns would be predicted correctly only on the basis of the knowledge acquired from the others. Results were considered excellent, except for the last few ones (at the right hand side of plot), for the reason outlined in sect. 5.1..

4.2. Process Sensitivity

Once the model has been validated (see sect. 4.1.), we have tried to use it to assess the effect of each concentration or process parameter on performance, by trying to evaluate the *sensitivity* of each performance index (j) versus each component (i) of input pattern:

$$\hat{S}_{ji}(\vec{X}) \triangleq \frac{\partial \hat{y}_j(\vec{X})}{\partial x_i}$$

Since the model is rather non-linear, we computed the sensitivities around a given composition. We have selected one of the best available materials, as there was no reason to optimize a poor material. Figure 4 (respectively, fig. 5) shows the

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

w_0	avg. error (friction)	avg. error (wear)	avg. error (friction)	avg. error (wear)	avg. error (friction)	avg. error (wear)
	$n = 2$		$n = 9$		$n = 37$	
1	0.6964	0.3981	2.9200	15.4713	1.3542	5.3678
3	0.2527	0.3447	1.2763	3.3762	6.7618	2.7368
5	0.2393	0.3651	2.1615	0.7584	0.4563	0.6840
7.5	0.2283	0.3632	0.1917	0.4159	0.2050	0.4264
9	0.2258	0.3638	0.1870	0.4165	0.1933	0.4273
11	0.2233	0.3655	0.1831	0.4174	0.1854	0.4313
15	0.2184	0.3700	0.1787	0.4241	0.1775	0.4402
20	0.2149	0.3755	0.1762	0.4350	0.1736	0.4452
30	-	-	0.1741	0.4505	0.1736	0.4480

Table 1: Normalized estimation error for different values of first-layer weights.

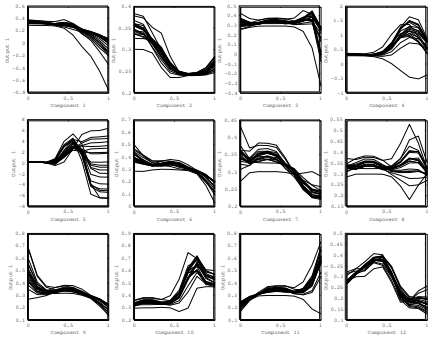

Figure 4: Estimated friction as a function of the first 12 components of input patterns, computed around a given material, which is at center of each plot.

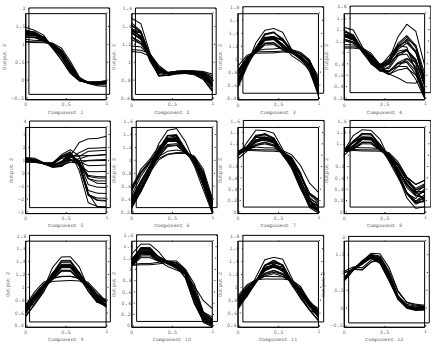

Figure 5: Estimated wear as a function of the first 12 components of input patterns, computed around a given material, which is at center of each plot.

sensitivities of friction (respectively, wear) with respect to the first 12 components of input pattern. These plots have been used to point out which were the parameters which most affected each performance. For instance, it can be seen that third component significantly affects wear (showing an extreme around a given value), but has nearly no effect on friction.

4.3. Automatic Grouping

We mentioned in sect. 3. that dimensionality of input patterns has been reduced by manually gathering chemical elements into *homogeneous groups*. Homogeneity was defined according to reasonable human considerations, but there was no clue that this was the optimal grouping, that is, that which increases model accuracy,

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

Grouping can be defined by means of a *grouping matrix* $\underline{\underline{\mathrm{M}}}$ such that

$$\vec{X} = \underline{\underline{M}}\vec{X}'$$

where $\vec{X}'$ is the ungrouped vector containing concentrations of *all* chemical element and process parameters, while $\vec{X}$ is the input pattern for the neural model, with grouped components.

For manual grouping, grouping matrix $\underline{\underline{\mathrm{M}}}$ has a 1 on j^{th} row and i^{th} column if and only if chemical element i is placed into group j. There has to be one and only one "1" in each column, as each chemical element can be associated with one and only one group.

We have then developed three automatic techniques to group chemical elements together:

- *crisp empirical grouping*, which consists of generating a number of random grouping matrices, with the only constraint that there has to be one and only one "1" in each column (see above). The best matrix is then chosen as described later;
- *fuzzy empirical grouping*, which consists of generating a number of random grouping matrices, with the only constraint that the sum along each columns shall always be 1. Each chemical element can therefore be a partial member of more groups. The best matrix is then chosen as described later;
- *training grouping matrix*: the grouping matrix $\underline{\underline{\mathrm{M}}}$ can be seen as an additional layer at the input of the neural model, building up a $(25+12)\times 19\times H\times 2$ three-layers neural network of type WRBF-0(linear)-WRBF-9(Gaussian)-NORM-1/1-WRBF-0(linear). This three-layer neural network (therefore also the grouping matrix) can be trained with a traditional backpropagation rule, using the RMS error on training set as cost function.

The best matrix is then chosen (except for the third approach), namely the one which reduces estimation error with the leave-one-out approach (see sect. 4.1.). Generalization capability of the resulting grouping matrix will be verified with empirical results, after a longer process of manufacture and test of samples, which is still going on.

4.4. Reducing the Curse of Dimensionality

The curse of dimensionality is reduced by finding the optimal order n of the Gaussian layer. A theoretical analysis (not reported here) pointed out that a good candidate was $n = 9$, as used so far. We have also verified this assumption by running the leave-one-out approach with different values of n, as reported in tab. 1, which confirmed the theoretical assumptions.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

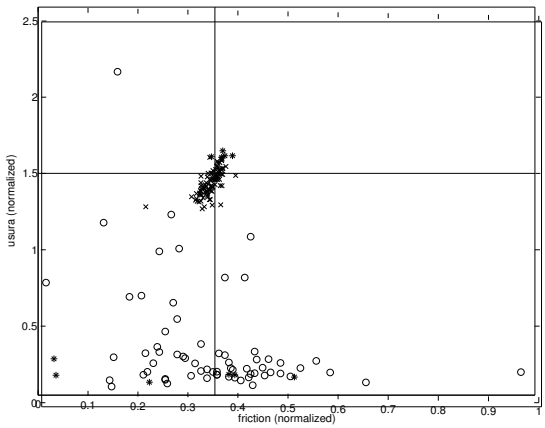

Figure 6: A subset of the Pareto surface (wear vs. friction) estimated around a user-selected material. The circles indicate the existing materials. The intersection of the horizontal and vertical axes indicate the user-selected material (one of the best available), while the crosses indicate the estimated performance on a small number of random point (we plotted only 100 points to improve clarity, although in practice we computed about 10,000 points). All points which are in the first quadrant with respect to the axes are potential candidates for an improved material.

5. Pareto Optimization

Another application of the proposed model, somehow similar to the analysis of sensitivity described in previous section, is the portion of *Pareto surface* shown in fig. 6. The idea is: i) to supply to the neural model a large number of random input patterns $\vec{X}^p$; ii) to estimate performance $y_j(\hat{\vec{X}}^p)$, based on the model of formula (1); iii) to plot *estimated wear* (in practice, the *inverse of wear*) versus *estimated friction*, for each random input pattern; iv) to select the random point(s) which maximizes both performance figures.

If the model were accurate and an infinite number of points covering the whole input space were supplied to the network, the resulting plot would be the Pareto surface of friction materials, which could then be used for multi-objective optimization, as shown in fig. 6. For instance, random points which have a better friction coefficient and a reduced wear (note that the plot shows the inverse of wear, which should be maximized) are candidates for an improved material. The best ones can then be manufactured and tested.

In practice, the model has a limited accuracy and the number of points is necessarily limited, therefore the resulting plot is nothing but a very small subset of the Pareto surface, mostly centered around a user-selected point (usually, one of the best available materials), for the reasons which will be explained in sect. 5.1..

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

5.1. Bounds on Modeling Reliability (Explored Subspace)

A few new materials have been manufactured and tested according to the results of the Pareto-like neural optimization described above. Results are shown in fig. 3, where the first 70 materials were those manufactured before starting this work, while the last four were those manufactured according to the Pareto-line neural optimization.

Results were not good, as the prediction capability of the model for these "new" points is much worse, which means that the model was very reliable to predict previously existing materials but it was not so reliable in predicting performance of new materials (namely, poor generalization capability), despite the indications of the leave-one-out validation (see sect. 4.1.).

We have deeply analyzed the reasons for this behavior, namely under what respect the "new" points were different with respect to the "old" ones. We pointed out that available materials could definitely not cover all input space, due to its large dimensionality, therefore we analyzed the *principal components*, that is, the regions in the input space where input pattern were more concentrated, therefore where the available "knowledge" was concentrated. Figure 7 shows standard deviation of the principal components of training set, in decreasing order, from which is can be seen that six components contain most of the knowledge.

We also pointed out that, since the new points were randomly distributed, they were likely to be in an "unknown" direction, that is, a direction in which no knowledge (namely, no old material) was available. It is reasonable to believe that generalization capability might be high along principal components and lower along the other directions in input space.

We therefore used results of PCA for three purposes:

1. to outline the *unknown areas* of input space, that is, the *least principal components*. No knowledge was available along these directions, that is, either no or very few materials have ever been manufactured and tested (by this industry) with those particular compositions. This analysis therefore provided valuable indications on *unexplored compositions*, which were worth being tried, although no indication whatsoever could be available on the performance which might be expected;

2. to outline the *reliability bounds* of the model, that is, the directions in input space where the neural model is reliable. This indications has been used in a second phase of the work to restrict the subspace in which random points were generated for the Pareto-like neural optimization. This explains why we have decided to restrict random points within a hyper-ellipsoid centered around a user-selected point (usually one of the best available material) with the axes oriented along the principal components and the radiuses proportional to the corresponding standard deviations, as shown in fig. 6 and described in sect. 5.. This has significantly improved the reliability of the suggested compositions (results are not shown in fig. 3);

3. to reduce the size of the neural model, although results are not yet available.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

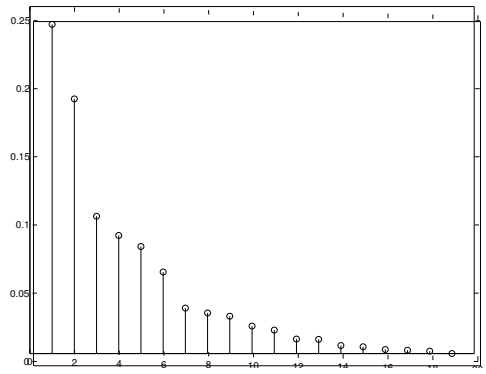

Figure 7: Standard deviation along the principal components of training set, in decreasing order.

6. Fuzzy Interpretation of Acquired Knowledge

Once the neural model has been validated via a number of experimental tests, the knowledge which the model has acquired has been converted into a set of fuzzy rules, which were readable by a human expert. We therefore converted the neural networks into a set of Mamdani fuzzy rules of the form:

$$R^1 : \text{IF } (x_1 \text{ is } small) \text{ AND } (x_2 \text{ is } medium) \text{ AND } \ldots$$

$$\ldots \text{ THEN } (y_1 \text{ is } large), (y_2 \text{ is } medium)$$

by means of the automatic conversion method described in [1]. We cannot report the whole set of rules, as they are confidential.

It is clear that the fuzzy model has the same validity bounds of the neural model (as the two are identical), therefore knowledge is mostly concentrated within the hyper-ellipsoid defined by PCA analysis (see sect. 5.1.).

7. Conclusion

This work has presented an application of neuro-fuzzy networks to tribology, showing how the many problems caused by the large dimensionality of input space have been solved. The paper has also outlined how the same neural model, which has been developed, has also been used for a number of different purposes.

In conclusion, results were considered encouraging, mostly for the following reasons:

- no reliable model was available so far for the problem under consideration. The model we have developed in this work has therefore increased the knowledge on the manufacturing process;

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 239-250

- the model developed significantly reduces development time of new materials, by giving quick indications whether a new composition has some chance to offer good performance;
- the model will be used in a decision-making process, to assist the human expert in the process of optimizing material performance;
- this work and all surrounding activities have outlined a few weaknesses of the testing and database management procedures, which are currently being solved for the future.

Acknowledgments

The author wishes to thank Drs. Pietro Buonfico, Alberto Portesani and Luigi Regis for their valuable help and comments in carrying out this work.

References

[1] L.M. Reyneri, "Unification of Neural and Wavelet Networks and Fuzzy Systems", in *IEEE Trans. on Neural Networks*, Vol. 10, no. 4, July 1999, pp. 801-814.

[2] S. Haykin, "Neural Networks: A Comprehensive Foundation", *Mc Millan College Publishing Company*, New York, 1994.

[3] V. Colla, L.M. Reyneri, M. Sgarbi, "Neuro-Wavelet Characterization of Jominy Profiles of Steels", in *Journal of Integrated Computer-Aided Engineering*, 1999, John Wiley & Sons, New York (NY).