

A NEW HYBRID EXPANSION FUNCTION BASED MUTUAL INFORMATION FOR A MULTILAYER NEURAL NETWORKS OPTIMIZATION

Kais NCIBI[★], Amor DJENINA^{★★}, Tarek SADRAOUI^{★★★}
Faycel MILI^{★★★★}

Received date: 21/ 10/2017, Accepted date: 26/ 11/2017, Online publication date: 09/ 12/2017

ABSTRACT

Function expansion was used to expand initial features based on a non linear transformation. Many known expansion functions are found such the trigonometric, the polynomial, the Legendre polynomial, the power series, the exponential and the logarithmic transformation. This paper present a comparison between different expansion functions based on mutual information and different performance functions. We propose a new expansion process able to improve the correspondent mutual information and the final performance. The process was tested; using different benchmark databases, and shows his ability to improve results of classification problems

Keyword: Function expansion; multilayer perceptron, Classification, mutual information, features selection.

JEL Code: C15, C61, C38, C41

Citation: NCIBI, K., DJENINA, A., SADRAOUI, T., MILI, F. (2017). A new hybrid expansion function based mutual information for a multilayer neural networks optimization. *Journal of Finance and Corporate Governance*, Vol., 1, No. 2: pp. 7 – 23. (Dec 2017); ISSN: 2602-5655.

[★] PhD University of economics and Management of Sfax MODELIS Lab, Email: quayesn@yahoo.fr

^{★★} Professor at University of Tebessa Algeria, Email: amor.djenina@gmail.com.

^{★★★} University of economics and Management of Mahdia, MODELIS Lab, Email: tarek.sadraoui@gmail.com.

^{★★★★} Associate Professor at University of economics and Management of Mahdia, MODELIS Lab, Email: milisoft@yahoo.fr

1. INTRODUCTION

Newly artificial neural network (ANN) has emerged as a powerful learning technique to perform complex tasks in highly nonlinear environment [1]. This model is a strong non linear approximator that attracts a lot of researchers. But in literature, there is no fix theory that illustrates an ANN construction. Fining the number of hidden nodes is one of constructive tasks in this area. Using a low number can results a bad representatively of the non linear approximation, and using a big number can results over fitting. Fixing hidden nodes, and using expansion function can be a good alternative to improve the non linear representatively of the data.

Functions expansion functions were firstly used in the functional link artificial neural network (FLANN) that was proposed by Pao [2]. The performance of the FLANN models have been experimented during numerous research works and found to be an effective approach both computationally at the same time as performance wise [3, 4].

The FLANN is principally a single layer network, so we didn't need hidden layers since functional expansion of the input pattern was included. The input, extended by expansion function engenders an increase in the input vector dimensionality, which helps the FLANN to resolve difficult classification problems by generating nonlinear decision boundaries. This nonlinearity to the input vector diminishes the number of layers and hence diminishes the computational complexity. And hence the hyper planes created by the FLANN present better discrimination ability in the input pattern space [5].

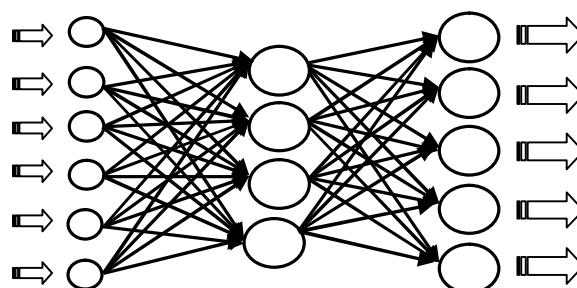
There have been several applications of FLANN including pattern classification and recognition, system identification and control, functional approximation, and digital communications channel equalization. In the research work [6], the writer have been successively applied FLANN for the channel equalization problem. The FLANN also has been applied to the financial forecasting and proved to be computationally efficient [7]. Moreover FLANN was used to forecast the one-day-ahead prediction of three fast growing global stock markets [8]. In this work we proposed a MLP structure, equivalent to a FLANN model, based on different expansion function. In the functional expansion development every input to the model is nonlinearly extended to engender a number of input values. A performance comparison between these expansion functions has also been made.

I. CONCEPTS AND DEFINITION

A. Multilayer perceptron

Multilayer Perceptron (MLP) is one of the most used ANNs. It consists of a series of fully interconnected layers of nodes where there are only connections between adjacent layers. General structure is showed in figure 1.

Figure 1 : A Fully connected multilayer perceptron (MLP)



The first layer (the input layer) takes as inputs the various attribute values. The output of the nodes in the input layer, multiplied with the weights attached to the links, is passed to the nodes in the hidden layer.

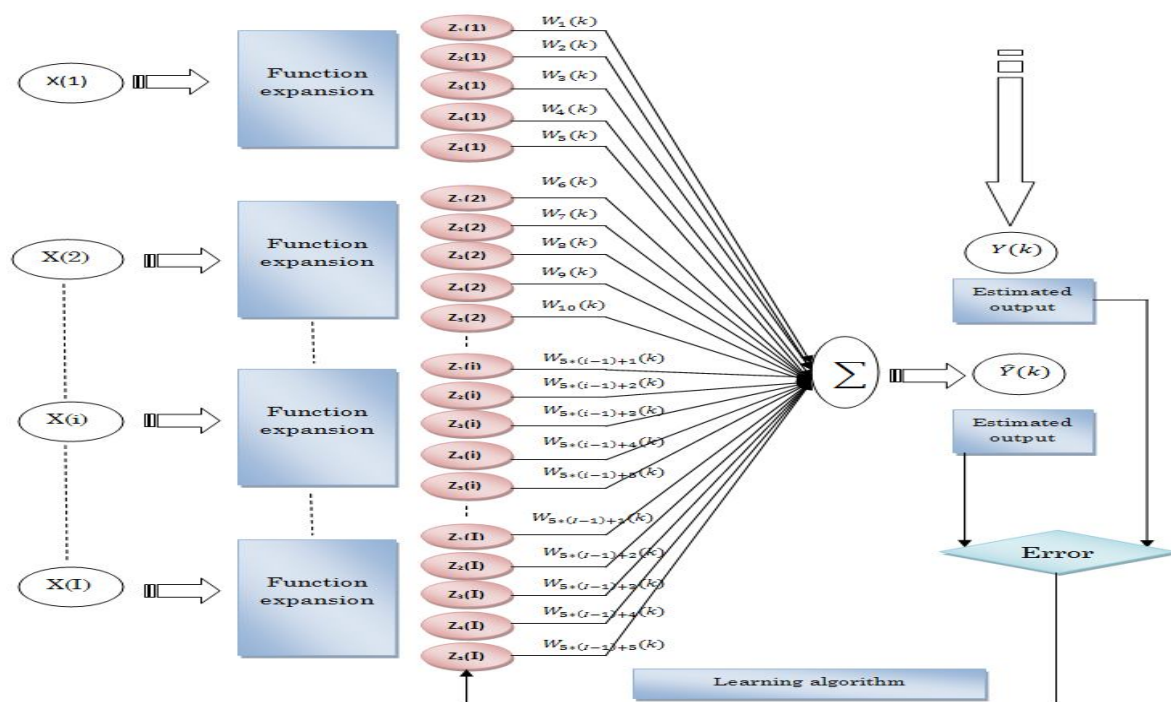
A hidden node collects the incoming weighted output values of the previous layers. Besides that, it receives also the weighted value of a bias node.

The sum of the weighted input values is passed through a nonlinear activation function. The only requirements are that the output values of the function are bounded to an interval and that the nonlinear function can be differentiable. The output of a node in the hidden layer is fed into the nodes of the output layer. To each node in the output layer a class label is assigned.

B. Expansion function

Expansion function was initially used by Pao et al. [2], in the construction of the functional link artificial neural networks (FLANN). The basic idea is to use some non linear transformation in order to increase the input vector dimensionality. A detailed description of FLANN is presented next figure. As presented there, each feature x_i will be expanded many times, giving $Z_n(i)$, and it will be represented by many non linear transformations.

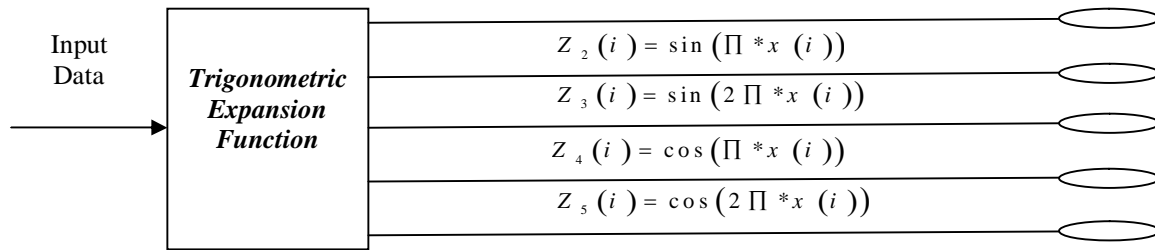
Figure 2 : An example of expansion function uses for the FLANN model



Many classic expansion functions are presented in the literature, such as the trigonometric, the polynomial, the Legendre polynomial, the power series, the exponential and the logarithmic transformation. We illustrate these functions in the next section.

1) Trigonometric Expansion Function

Trigonometric function is employed to increase the input data to higher dimension. After applying trigonometric function expansion, various value input terms are obtained such as :

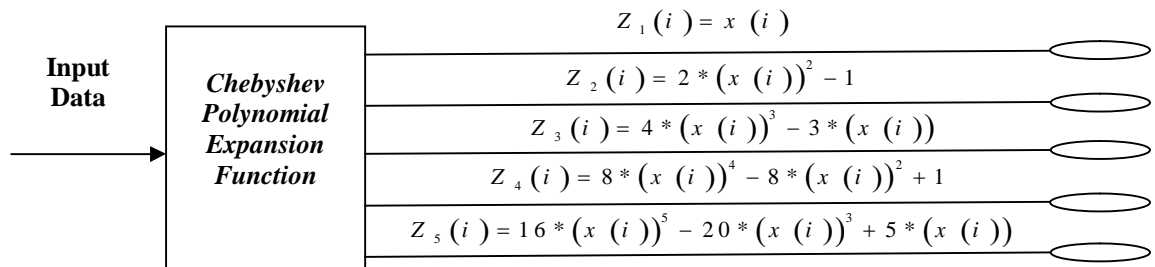


2) *Chebyshev Polynomial Expansion Function*

The basic Chebyshev Expansion Function used is called Chebyshev polynomial function. The Chebyshev polynomials are a set of orthogonal polynomials obtained as a solution of the Chebyshev differential equation.

The higher order Chebyshev polynomials may be generated using the recursive formula as follow: $Z_{n+1}(x) = 2xZ_n(i) - Z_{n-1}(i)$

The first expressions for the first six terms presented as subsequently:

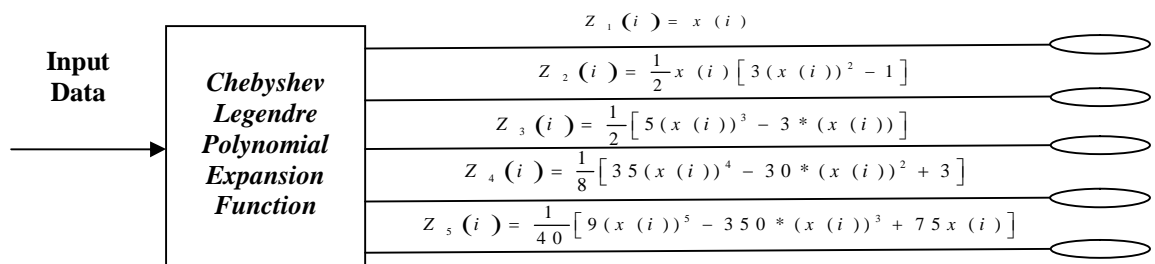


3) *Chebyshev Legendre Polynomial Expansion Function*

Higher order Legendre polynomials can be produced by the recursive formula known by:

$$Z_{n+1}(i) = \left(\frac{1}{n+1}\right) * [(2n+1)x(i)Z_n(i) - nZ_{n-1}(i)]$$

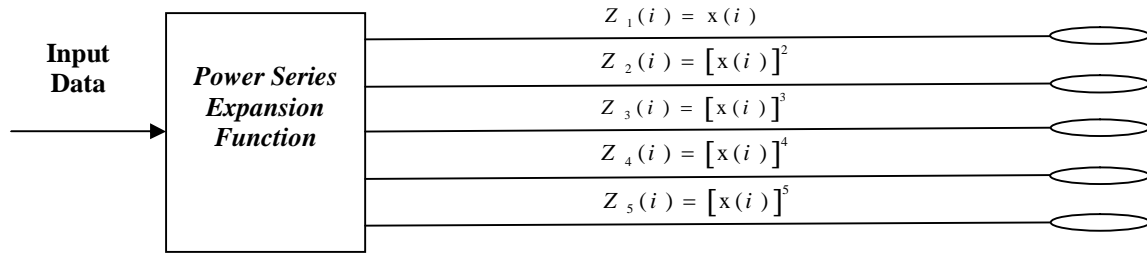
The few first terms presented next:



4) *Power Series expansion function*

Power series can be defined as an infinite series which has the next form:

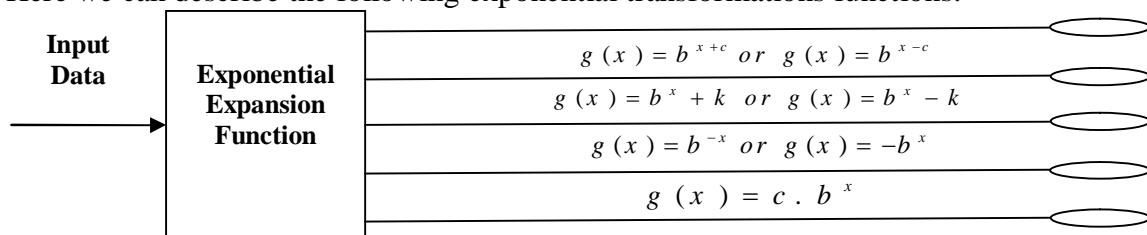
The few first terms presented next:



5) *Exponential expansion function*

The parent's Exponential function has the form: $f(x) = b^x$, $b > 1$ or $0 < b < 1$
 Where b is called the base of the function, and it determine the direction of the graph.
 $b=1$ Means that we have a constant function. $b > 1$; The graph increases beginning the left to right.
 $0 < b < 1$; The graph decreases commencing the left en route for the right.

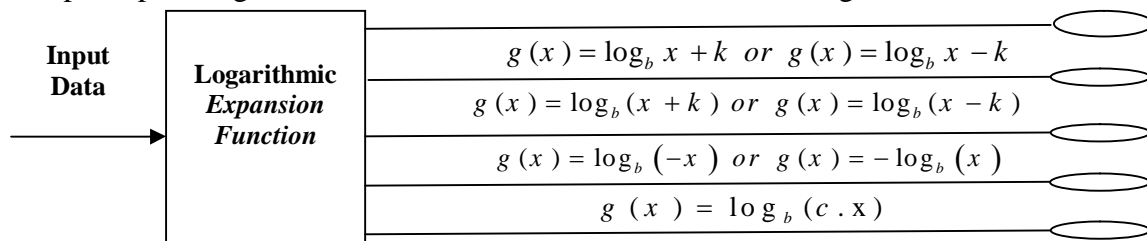
Here we can describe the following exponential transformations functions:



6) *logarithmic expansion function*

The parent's Exponential function logarithmic function with a base b has the following form $y = \log_b x$. b is called the base of the function. b is in general a number superior than 1 (while it need just to be bigger than 0 and not equal to 1). The logarithmic function is defined for all $x > 0$.

The principal's logarithmic transformations functions are showing then:



C. *Features discretization*

Dealing with quantitative features, these features is discretized in order to making data more easy to use and in order to compute some feature evaluation criteria such the mutual information.

At first, we define the mutual information; next we describe the used features discretization method.

1) *Mutual information*

In information theory, mutual information measures the statistical dependence between two random variables and so can be used to evaluate the relative utility of each feature to classification. It is defined as:

$$I(X, Y) = \int_Y \int_X p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx dy$$

Where $p(x, y)$ is the joint probability density function of continual random variables X and Y , and $p(x)$ and $p(y)$ are the marginal probability density functions respectively. It is not difficult to find that mutual information is related to entropy as:

$$I(X, Y) = H(Y) - H(Y/X)$$

Given the Shannon entropy (discrete) defined as:

$$H(X) = - \sum_x p(x) \log(p(x))$$

The use of mutual information for feature selection can be intuitively justified. From the point of view of statistics, the mutual information measures more general dependency in the data, and therefore may lead to better capability to feature selection.

II. MODEL DESCRIPTION

We propose a constructive ANN model for classification task this model is composed with 2 phases

- A preprocessing phase:
- An optimization phase:

A detailed description of these phases will be illustrated next:

A. A preprocessing phase:

The preprocessing phase is described in figure 1. This phase prepare initial data for the optimization one, It is an important task that can influence final result of the proposed model. These steps are described as follows:

1) Data cleaning

It is one of important tasks in data warehouse task. In our model we are limited to outlier's adjustment. An observation is described as an outlier if it is out of some boundary. For example, we can use the interval inter-quartile IIQ, and we can describe an outlier as a point out of $1.5 * IIQ$.

- If this observation is identified, we propose to use regression as follows
- Select the 2 best most correlated features to the correspondent one
- Construct a linear model using the features
- Predict the new value of the outlier and replace the old one

The proposed outliers control can be used for missing data.

2) Feature discretization

Discretization can be a good task for data representation. Data can be more comprehensive, ANN became faster and rules extraction tasks can be executed. Else some evaluation criteria for input features can be computed such mutual information criteria.

3) Data expansion

While we aim to use ANN, as a non linear approximator, we propose to use expansion function in order to represent better non linearity of the data. Many expansion function

was found such the trigonometric, the polynomial, the Legendre polynomial, the power series, and the exponential and the logarithmic transformation. We propose to use many expansions functions in the same time and select the best one based on the principle of mutual information.

4) *Data reduction*

Clustering can be a good alternative for the proposed model. It can reduce the number of features, the required time for convergence and it can improve performance and required time. We propose to use the principal component analysis techniques (PCA) and the self organizing mapping (SOM). These two techniques are very used for clustering task. After prior experiences, we select the PCA as a principal clustering technique, and it is proved to be more compatible for our model.

5) *Feature selection*

Feature selection problem is to select a minimum set of features that perform well the performance of the studied model. In literature, many techniques can be found such, heuristic methods as Step-wise forward selection, Step-wise backward elimination, and decision tree induction. We propose preprocessing approach, based on t-tests, correlation coefficient, and mutual information.

6) *Data normalization*

We use the mix-max normalization function, we choose to use logsig transfer function and the normalization between [0, 1]. This one is the best recommended for ANN models.

7) *Data division and cross-validation*

K-fold cross-validation re-sampling technique is one of the most know technique that can be used in classification in order to avoid over fitting. In this study, 10-fold cross-validation re-sampling technique is used. 80% of data is used for training, 10% for validation and 10% for the test. Experience will be repeated 10 times, and at each time, we change the test fold and the validation fold.

A. *An optimization phase*

The optimization phase of an ANN construction has the aim to find the best combination of weight that improves better performance function. This problem is classed as a continued non linear optimization problem, and it is hard to be optimized. Many algorithms are found in literature. The most used one is the back-propagation algorithm. The last one performs good results, but it can be confronted in a problem of local minimum. For this aims, we add a local search algorithm and a differential evolution algorithms, in the hope to avoid this problem and increase the chance of faster convergence.

A detailed description of the optimization phase is presented in figure 2. This phase is illustrated as follows:

1) *Initial weight and improving initial weights*

The process begins by generating randomly initial weights. We generate 10000 times weights and we select the best one of these generation based on the criteria of minimum MSE.

2) *Apply the back-propagation algorithm*

This algorithm is used at first as a basic optimization algorithm of the model.

3) *Apply the local search*

After applying the back-propagation algorithms, we add to each weight a random walk and we observe the MSE effect. Weight will be updated if adding this random walk improves the MSE function.

4) *Apply the differential evolution*

Differential evolution is an evolutionary algorithm that can have a good ability to improve our weights. If we suppose that we have 3 states of weights:

W_1 : initial weights

W_2 : solution weights after the back propagation process.

W_{best} : is the set of weights after applying local search

Many schemes can be found in literature. While we have only 3 states of weights, we choose the following scheme DE/best/1. This schema updates weights as follows. $W = W_{best} + F * (W_1 - W_2)$

Where F defines the constriction factor generally taken equal to 0.5.

5) *Model evaluation*

3 basic functions are choose for evaluation such :

- **The mean square error function**: noted MSE:

The mean $\frac{1}{n} \sum_{i=1}^n$ squared error $\left((\hat{y}_i - Y_i)^2 \right)$ (MSE) of an estimator calculate the difference between the estimator and what is estimated. The difference happened due to randomness or since the estimator doesn't consider information that can generate a more accurate estimate [9].

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

- **The misclassification error function**: noted MCE.

The misclassification error rate (MCE) can be used too as a criterion for classifier evaluation. The MCE is defined as: $MCE = \frac{s_i}{S} * 100$. Where s_i is the number of incorrectly classified data. And S is the total data size.

- **The area under the ROC curve**: noted AUC.

Once using normalized units, the area under the curve (often referred AUC, or AUROC) correspond to the probability that a classifier will rank randomly selected instances by assuming 'positive' ranks higher than 'negative' one. [10] This can be seen as follows: the area under the curve is calculated by the following integral:

$$A = \int_{-\infty}^{\infty} TPR(T)FPR'(T)dT = \int_{-\infty}^{\infty} I(T' > T) f_1(T')f_0(T)dT' dT = P(X_1 > X_0)$$

Where: X_1 represent the score for a positive instance. X_0 Represent the score for a negative instance.

The criteria that will use for comparison is the MSE. The MCE and the AUC are just displayed to observe their dynamic evolution over the dynamic process.

Others evaluation criteria can be used such:

The required time: Time can be a good alternative for comparison. More we are faster, more we takes quick decision.

Mean of mutual information: each model can have different form if input features, the mean of mutual information of these features can be a good alternative to observe model before and after features transformation. The last one can be resulted by clustering, expansion, features removing etc.

The number of improvement: When dealing with hybrid algorithms, we need to know which algorithm improve better than. In this case we implement an instruction the count the number of improvement of each algorithm. 3 number of improvement are defined such:

- Number of improvement of back-propagation
- Number of improvement of local search
- Number of improvement of differential evolution

6) *Stopping criteria*

The process will run in a cycle while there is MSE improvement. Stopping criteria is an instruction that fix where our process must be stopped. 3 stopping criteria are fixed:

- **We reach a maximum number of epochs**

We fix this number equal to 20 for prior test, and 100 for the final model optimization.

- **Reaching a MSE less than a prior fixed level**

We fix MSE level equal to 0,001.

- **Having a number of successive execution without MSE improvement**

We fix this number level equal to 5 executions.

B. Features discretization method:

We propose to use the principle of binning method. The proposed process is described as follow:

- Normalize the data between [0,1]
- Choose the number of split, such 2, 3, 4 or 5 splits.
- Apply the rule “the nearest take all.

For example, if we aim to divide data to 3 splits: 1 2 and 3. We choose center such 0, 0.5 and 1. For example 0.6 is near 0.5 than, then this value will be arranged to class 2. 0.1 is near 1, this value will be replaced by class 1.

But how to choose the number splits. We propose to

- choose many split in the same time (from 2 to 10)
- Select the best split based on the maximization of mutual information.

C. Proposed function expansion

In the last section, we have presented 7 classic expansion functions that can be used. Each one performs a different non linear transformation, and there is no fixed theory confirms the best one from the six expansion function.

From this idea, and in order to improve results, we propose two new expansion function techniques, which were inspired from the classic ones.

In order to improve results, we propose two new expansion techniques based on the classic ones.

Selected expansion function: each feature will be expanded using all expansions functions, and the best expansions will be used.

Improved selected expansion function: each feature will be expanded using all expansions functions, and while there is improvement, these features will be expanded other times until there is no improvement.

Results are presented next. From this result, we observe that the improved selected expansion function gives better results than other expansion. So this proposed idea can be a good alternative for our model.

Description of these expansion functions are as follows:

1) Selective expansion function

This expansion process is described in figure 3. Description is as follows

Step 1: Multiple expansion function will be done in the same time. That mean we have seven different functions. If we don't account initial feature, each function performs 4 different non linear new features. So we will have $4*7+1=29$ different features.

Step 2: These 29 features will be discretized based on a binning method. So we obtain 29 nominal features.

Step 3: By obtaining 29 nominal features, and based on output classes, mutual information, p-value of chi-square test can be computed

Step 4: The best one of 29 nominal features will be selected. This feature must have the maximum mutual information and the minimum p-value.

Step 5: We replace the initial feature by the new selected one.

We do the same for all features. At the end we obtain a new database containing new feature with maximum mutual information and a minimum p-value of chi-square test.

2) Improved selective expansion function

This expansion process follows the same steps than what we have presented in the last process by presenting the selective expansion function. What we add new in this process is as follows:

- We repeat this selective expansion function many times while there is mutual information and p-value improvement.
- We allow selecting many representation of each feature, if we find many non linear best representation of it.

This process has the advantage

- To expand many time while there is mutual information and p-value improvement.
- To represent a feature with many non linear representation. When used in an ANN optimization, this can improve performance of the model.

But if a feature is expanded many time, and represented with many representation, the process can result a big number of new representation of features. This can result a slowly optimization process. Thus, a next feature selection and data reducing technique can reduce this number and help to takes best of the best representation. The confirmation of the ability of this proposed process will be done in the empirical performance results.

Figure 3 : Improved selective expansion function process

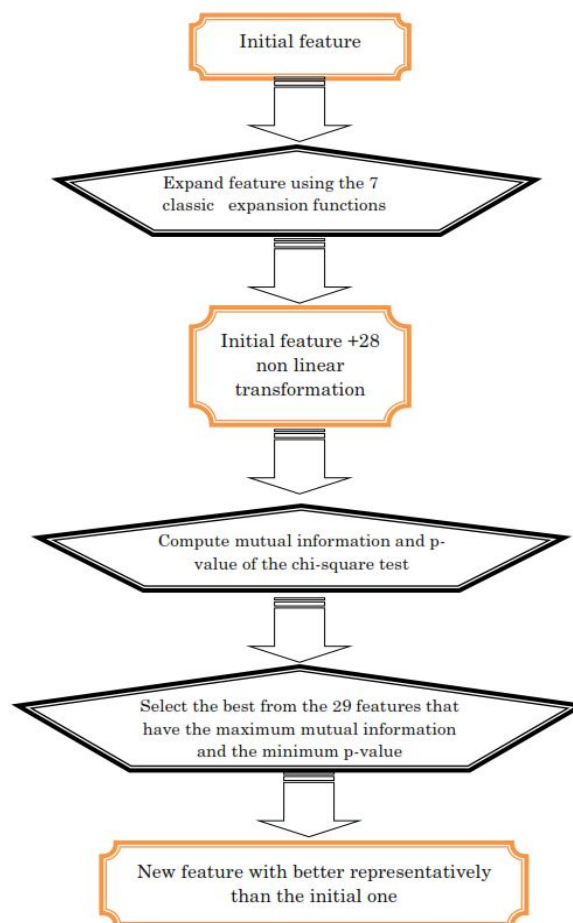
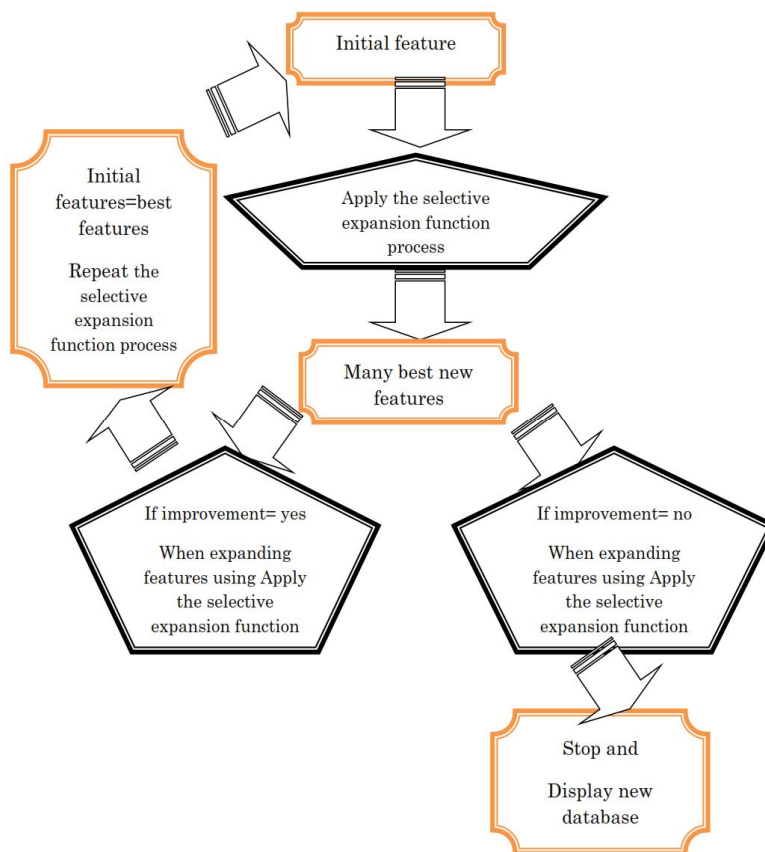


Figure 4 : Selective expansion function



3/ Data reduction

Initial data will be transformed in order to reduce the number of used features and perform better representatively of used features.

We propose 3 data transformations techniques such

3) Principle component analysis

PCA was invented in 1901 by [11], This technique is one of the most used one the data transformation techniques. His principle is to reduce the features space by creating new factors that represent well the initial space.

4) Self organizing mapping organization (SOM)

The Self-Organizing Map was proposed by [12]. The SOM has been confirmed functional in numerous applications [13].

SOM is a neural networks model that was used for clustering problem. This model is based on a non linear projection and can be a good alternative for the data transformation process

III. Experimental results

8 real-world databases were selected there to be used in simulation works. 6 are chosen from the UCI repository machine learning, which is commonly used to benchmark learning algorithms [14]. The “TUNISIAN BREAST CANCER” Database was collected from real Tunisian hospital. The “ISLAMIC VS

CONVENTIONAL BANK” database was extracted from the “banqscop” database source [15].

A. Data bases description

A brief description of used databases for experimental setup is presented in table I. Num. is the numeric features, Bin. is the binary ones, and Nom. is the nominal inputs that mean discrete with three or more distinct labels.

Tableau 1 : Summary of the dataset used in simulation studies

| <i>Datasets</i> | | <i>Inputs</i> | | | | <i>Ex.</i> | <i>Cls</i> | | |
|------------------------------|--------|---------------|-------------|-------------|--------------|------------|--------------------|-----------|-----------|
| | | <i>Num.</i> | <i>Bin.</i> | <i>Nom.</i> | <i>Total</i> | | <i>Numbe r</i> | <i>C1</i> | <i>C2</i> |
| TUNISIAN CANCER | BREAST | 4 | 17 | 0 | 17 | 248 | 2 | 100 | 148 |
| PRIMA | | 8 | 0 | 0 | 8 | 768 | 2 | 268 | 500 |
| CREDIT | | 6 | 4 | 4 | 14 | 690 | 2 | 307 | 383 |
| BREAST | | 9 | 0 | 0 | 9 | 699 | 2 | 241 | 458 |
| VOTING | | 0 | 16 | 0 | 16 | 232 | 2 | 124 | 108 |
| BUPA | | 6 | 0 | 0 | 6 | 345 | 2 | 145 | 200 |
| ISLAMIC CONVENTIONAL BANK | VS | 25 | 0 | 0 | 25 | 706 | 2 | 137 | 569 |
| VISA CARTE | | 7 | 0 | 0 | 7 | 1710 | 2 | 765 | 945 |

B. Experiences Description

The experience consists to run 10 times the proposed process over all the 8 presented databases. For each database, we compute the mean of each performance evaluation criteria. Next, we compute the mean of the 8 means of each database.

Based on the training set, the validation set and the test set, 20 first epochs will be observed. Results will be displayed based on

- MSE, MCE, AUC,
- The number of improvement of BP, LS and DE,
- The mean of mutual information
- The required time for execution

At first we display results of the optimization phase without any preprocessing task. Next we display results with the total proposed process.

Tables and figures will displays MSE, MCE and AUC using this transformation.

$$MSE = (1 - MSE) * 100$$

$$MCE = (1 - MCE) * 100$$

$$AUC = AUC * 100$$

These transformations were done in order to display the 3 criteria in the same figure.

C. Results of expansion function based mutual information

In this part, we compare different expansion function based on the criteria of mutual Information. Results are presented in the next table.

Tableau 2 : Expansion function results based on mutual information

| | <i>trigonometric expansion</i> | <i>chebyshev polinomial expansion</i> | <i>chebyshev legendre polinomial expansion</i> | <i>power series expansion</i> | <i>exponential expansion</i> | <i>log expansion</i> |
|-----------------|--------------------------------|---------------------------------------|--|-------------------------------|------------------------------|----------------------|
| Zribi | 0,653 | 0,656 | 0,613 | 0,649 | 0,498 | 0,381 |
| Voting | 0,463 | 0,478 | 0,358 | 0,557 | 0,594 | 0,443 |
| Credit | 0,537 | 0,487 | 0,481 | 0,535 | 0,521 | 0,456 |
| Bupa | 0,662 | 0,366 | 0,328 | 0,454 | 0,484 | 0,510 |
| Diabetes | 0,477 | 0,493 | 0,593 | 0,437 | 0,517 | 0,232 |
| Islamic | 0,490 | 0,499 | 0,583 | 0,498 | 0,581 | 0,566 |
| Breast | 0,833 | 0,511 | 0,474 | 0,518 | 0,678 | 0,656 |
| Visa | 0,618 | 0,585 | 0,500 | 0,614 | 0,557 | 0,602 |
| Mean | 0,603 | 0,490 | 0,493 | 0,509 | 0,556 | 0,504 |

This table shows that the trigonometric expansion is 4 times the best over 9 database execution. Thus, by observing means of each expansion, the trigonometric one is the best expansion function, based on the mutual information criteria.

In order to improve results, we choose selective expansion methods that select the best expansion for each feature. Results are presented next.

Tableau 3 : Trigonometric versus selective Expansion function results based on mutual information

| | <i>Trigonometric expansion</i> | <i>Expansion with one best selection</i> | <i>Expansion with many best selection</i> |
|-----------------|--------------------------------|--|---|
| zribi | 0,6529 | 0,4852 | 0,5124 |
| voting | 0,4625 | 0,4625 | 0,4625 |
| credit | 0,5368 | 0,58 | 0,6209 |
| bupa | 0,6615 | 0,7229 | 0,8401 |
| diabetes | 0,4769 | 0,3869 | 0,4907 |
| islamic | 0,4901 | 0,5547 | 0,463 |
| breast | 0,8334 | 0,8137 | 0,8401 |
| Visa | 0,618 | 0,7344 | 0,4768 |
| mean | 0,60278333 | 0,6321 | 0,65096 |

By comparing the trigonometric results with the selective one, we find that the selective methods with many features replacement better results. So this expansion will be used as a principal expansion function for our model.

D. Using the expansion function

Many expansion functions can be found in the literature. The 4 most known are the trigonometric, the Chebyshev polynomial, the Chebyshev legendre polynomial and the power series. Results shows that the trigonometric is the most recommended, which gives better results on MSE, MCE and AUC.

Figure 5 : The 4 most known are the trigonometric, the Chebyshev polynomial, the Chebyshev legendre polynomial and the power series.

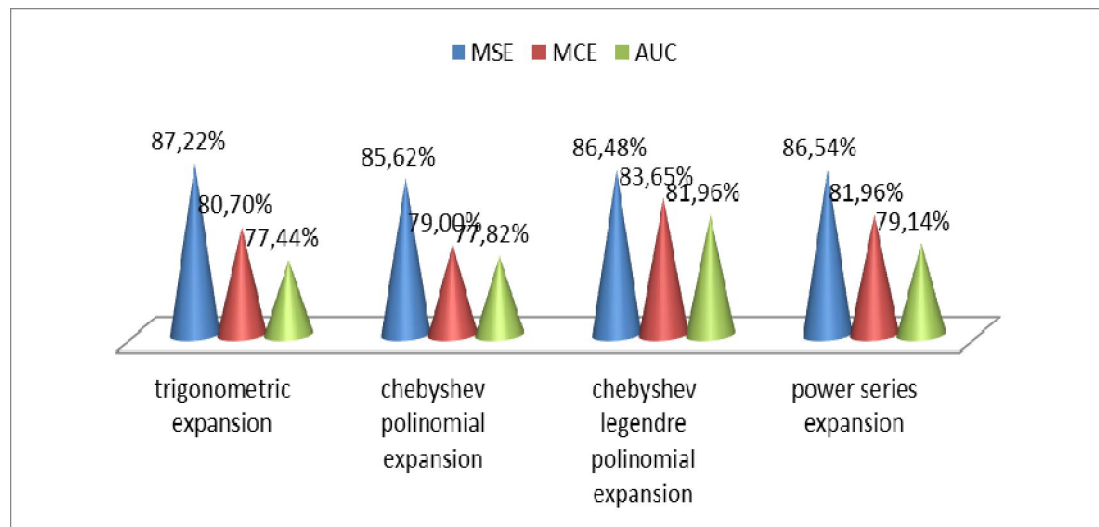


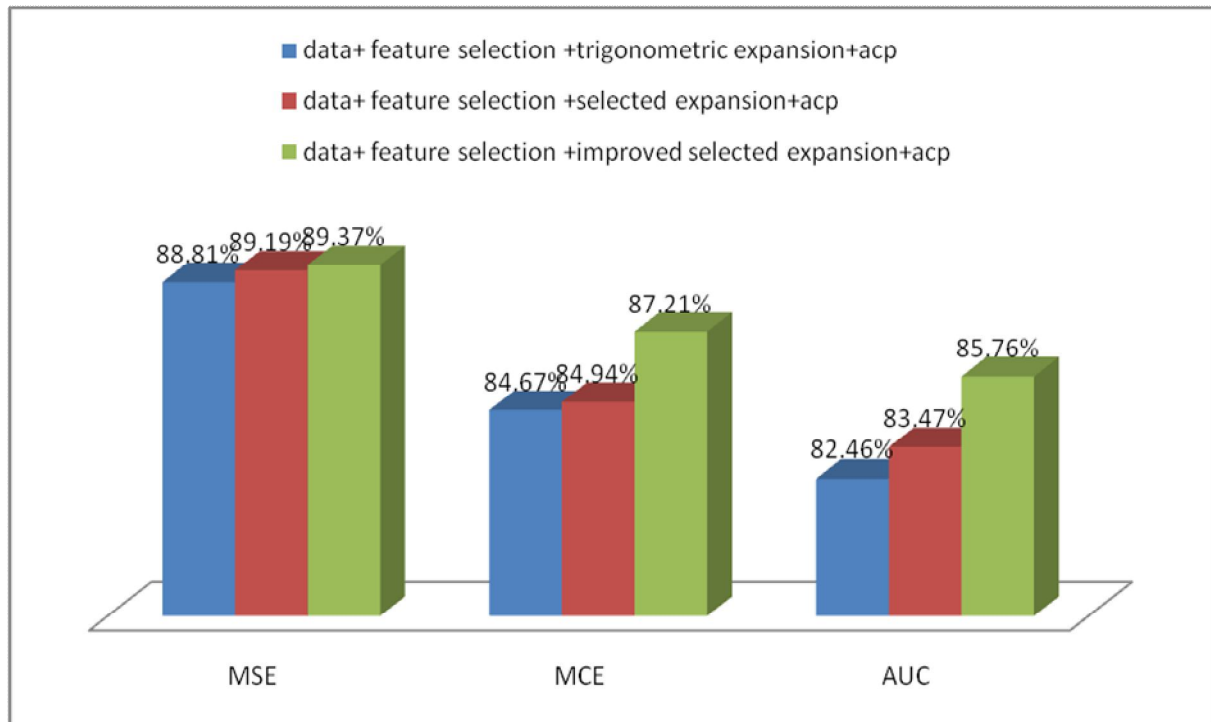
Tableau 4 : Comparing the 4 most known expansion function based on MSE, MCE, AUC

| | <i>MSE</i> | <i>MCE</i> | <i>AUC</i> |
|--|------------|------------|------------|
| trigonometric expansion | 87,216% | 80,695% | 77,441% |
| chebyshev polinomial expansion | 85,615% | 78,999% | 77,821% |
| chebyshev legendre polinomial expansion | 86,483% | 83,649% | 81,955% |
| power series expansion | 86,540% | 81,964% | 79,141% |

E. Comparing the trigonometric vs the proposed function expansion function

Tableau 5 : Comparing the trigonometric vs the proposed function expansion function

| | <i>MSE</i> | <i>MCE</i> | <i>AUC</i> |
|---|------------|------------|------------|
| | 88,81% | 84,67% | 82,46% |
| Data + feature selection + trigonometric expansion + ACP | 89,19% | 84,94% | 83,47% |
| Data + feature selection + selected expansion+ ACP | 89,37% | 87,21% | 85,76% |
| Data + feature selection + improved selected expansion+ACP | | | |

Figure 6 : Comparing the trigonometric vs the proposed function expansion function

CONCLUSION

This paper present a comparison between different expansion functions based on mutual information and different performance functions. We propose a new expansion process able to improve the correspondent mutual information and the final performance. The process was tested; using different benchmark databases, on a preprocessing phase of multilayer neural networks and shows his ability to improve the results of a supervised classification process.

REFERENCES

- [1] Patra. J. C, Pal. R. N, Chatterji. B. N, Panda, G, "Identification of nonlinear dynamic systems using functional link artificial neural networks" IEEE Transactions , Systems, Man and Cybernetics, Part B, Vol 29 , April-1999, pp 254 – 262.
- [2] Pao, Y.H.: Adaptive pattern recognition and neural networks. Addison-Wesley, Boston (1989).
- [3] Patra, J.C., Bos, A. V. D.: Modeling of an intelligent pressure sensor using functional linkartificial neural networks. ISA Trans. Elsevier 39,15–27 (2000)
- [4] Majhi, R., Panda, G., Sahoo, G.: Development and performance evaluation of FLANN based model for forecasting of stock markets. Expert Syst. Appl. 36, 6800–6808 (2009).
- [5] A. R. Foruzan, B. N. Araabi, "Iterative median filtering for restoration of images with impulsive noise." Electronics, Circuits and Systems, 2003. ICECS 2003.Dec 2003 ,PP 14-17 Dec.

- [6] Patra, J.C., Pal, R.N., Baliarsingh, R., Panda, G.: Nonlinear channel equilization for QAM signal constellation using artificial neural network. *IEEE Trans. Syst. Man Cybern. Part B* 29 (2), 262–271 (1999)
- [7] Patra, J.C., Kim, W., Meher, P.K., Ang, E.L.: Financial Prediction of Major Indices Using Computational Efficient Artificial Neural Networks, pp. 2114–2120. *IJCNN, Vancouver* (2006)
- [8] Nayak, S.C., Misra, B.B., Behera, H.S.: Impact Comparison of Performance of Different Functions in Functional Link Artificial Neural Network: A Case Study on Stock Index Forecasting. L.C. Jain et al. (eds.), *Computational Intelligence in Data Mining - Volume 1, Smart Innovation, Systems and Technologies*, pp 478-487 Springer India 2015
- [9] Lehmann, E. L.; Casella, George (1998). *Theory of Point Estimation* (2nd ed.). New York: Springer. ISBN: 978-1-4419-3130-6; Series ISSN: 1431-875X; *Edition Number: 2*.
- [10] Fawcett, Tom (2006); An introduction to ROC analysis, *Pattern Recognition Letters*, 27, 861–874.
- [11] K. Pearson (1901). : On Lines and Planes of Closest Fit to Systems of Points in Space" . *Philosophical Magazine* 2 (11): 559–572.
- [12] Teuvo Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995.
- [13] Liu, Y., and Weisberg, R. H. (2011). “ A review of Self-Organizing Map applications in meteorology and oceanography,” in *Self-Organizing Maps—Applications and Novel Algorithm Design*, edited by J. I. Mwasiagi (InTech, Rijeka, Croatia), pp. 253–272.
- [14] Y.-H. Pao, S.M. Phillips and D.J. Sobajic, “Neural-net computing and intelligent control systems”. *Int. J. Contr.*, 56, pp. 263-289 , 1992.
- [15] R. Majhi, G. Panda and G. Sahoo, “Development and performance evaluation of FLANN based model for forecasting of stock markets”, *Expert Systems with Applications* 36, pp. 6800–6808,2009.