Comments for ai-2642202

# Reviewer 1

## 1. Comment

An explanation of the effectiveness and characteristics of the proposed

## Response

The following text has been added at the end of Introduction section:

*“A common method of calculating the parameters in these neural networks uses a technique to calculate the centers of the functions \phi( x ) and then the weight vector w &xrarr; is calculated as a solution of a linear system of equations. Typically, the method used to calculate the centers is the well - known k-means method [*[*kmeans*](#LyXCite-kmeans)*]. In many cases this way of estimating the parameters of the neural network leads to over-fitting of the model so that it cannot generalize satisfactorily to unknown data. Furthermore, since there is no range of values for the parameters, there is the possibility that they will take extremely large or extremely small values, with the result that any generalizability of the model is lost. This work suggests a two phase method to minimize the error of equation (*[*eq:eqrbf*](#eq_eqrbf)*). During the first phase, an attempt is made to bound the parameter values to intervals in which the training error is likely to be significantly reduced. The identification of the most promising intervals for the parameters is performed using a technique that utilizes Grammatical Evolution[*[*ge1*](#LyXCite-ge1)*], that collects information from the training data. The first phase attempts to create a small interval of values for the neural network parameters by applying a series of of division rules, with the assistance of the Grammatical Evolution. The determination of the value interval is done in such a way that it is faster and more efficient to train the parameters of the neural network with some optimization method during the second phase of the method. In general, if the value intervals for the parameters from the first phase are small in range, the second phase of the technique is expected to be significantly accelerated as well. During the second phase, the parameters of the RBF network can be trained within the optimal range found in the first phase using some global optimization method [*[*rbfSA*](#LyXCite-rbfSA)*,* [*rbfPSO*](#LyXCite-rbfPSO)*]. In the proposed approach, the widely used method of genetic algorithm [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] was used for the second phase of the process.”*

## 2. Comment

Suggest using a flowchart to illustrate the method for readers to understand

## Response

The flowchart of the proposed algorithm has been added at the end of subsection 2.3

## 3. Comment

Add a comparison with existing methods to illustrate their advantages

## Response

Two methods have been added:

1) The Adam optimizer, used to train a neural network with 10 hidden nodes

2) The NeuroEvolution of Augmenting Topologies

# Reviewer 2

## 1. Comment

The abstract must be re-written, focusing on the technical aspects of the proposed model, the main experimental results, and the metrics used in the evaluation. Briefly discuss how the proposed model is superior.

## Response

The new abstract now reads:

*“Radial Basis Function networks are used in a variety of real-world applications such as medical data or signal processing problems.The success of these machine learning models lies in efficiently finding values for the model parameters. In this work, a new method of finding these values is formulated which is divided into two phases. In the first phase, with the use of Grammatical Evolution, an attempt is made to find value intervals for the model parameters using partition rules. In the second phase of the proposed method, an intelligent optimization algorithm such as a genetic algorithm, locates the optimal values of the parameters within the best value interval which is the output of the first phase. The proposed technique has been applied to a wide range of classification or data fitting problems and there has been a significant reduction in error exceeding 40% on most datasets.”*

## 2. Comment

A methodological manuscript, like the current one, must formulate what problem needs to be solved and why existing techniques are insufficient. Why not Autoencoders or Variational Autoencoders?

## Response

The following paragraph has been added in the Introduction section:

*“In the same direction of research, other researchers propose to handle problems of categorization or data fitting, techniques such as Support Vector Machines (SVM) [*[*svm*](#LyXCite-svm)*,* [*svm2*](#LyXCite-svm2)*], decision trees [*[*dt1*](#LyXCite-dt1)*,* [*dt2*](#LyXCite-dt2)*] etc. Also, Wang et al suggested an auto - encoder reduction method, applied on a series of large datasets[*[*nn\_autoencoder*](#LyXCite-nn_autoencoder)*]. ”*

## 3. Comment

The contribution of the current study must be briefly discussed as bullet points in the introduction. And motivation must also be discussed in the manuscript.

## Response

The following text has been added at the end of Introduction section:

*“A common method of calculating the parameters in these neural networks uses a technique to calculate the centers of the functions \phi ( x ) and then the weight vector w &xrarr; is calculated as a solution of a linear system of equations. Typically, the method used to calculate the centers is the well - known k-means method [*[*kmeans*](#LyXCite-kmeans)*]. In many cases this way of estimating the parameters of the neural network leads to over-fitting of the model so that it cannot generalize satisfactorily to unknown data. Furthermore, since there is no range of values for the parameters, there is the possibility that they will take extremely large or extremely small values, with the result that any generalizability of the model is lost. This work suggests a two phase method to minimize the error of equation (*[*eq:eqrbf*](#eq_eqrbf)*). During the first phase, an attempt is made to bound the parameter values to intervals in which the training error is likely to be significantly reduced. The identification of the most promising intervals for the parameters is performed using a technique that utilizes Grammatical Evolution[*[*ge1*](#LyXCite-ge1)*], that collects information from the training data. The first phase attempts to create a small interval of values for the neural network parameters by applying a series of of division rules, with the assistance of the Grammatical Evolution. The determination of the value interval is done in such a way that it is faster and more efficient to train the parameters of the neural network with some optimization method during the second phase of the method. In general, if the value intervals for the parameters from the first phase are small in range, the second phase of the technique is expected to be significantly accelerated as well. During the second phase, the parameters of the RBF network can be trained within the optimal range found in the first phase using some global optimization method [*[*rbfSA*](#LyXCite-rbfSA)*,* [*rbfPSO*](#LyXCite-rbfPSO)*]. In the proposed approach, the widely used method of genetic algorithm [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] was used for the second phase of the process.”*

## 4. Comment

The overall organization of the manuscript is not discussed anywhere in the manuscript. Please add the same in the introduction section of the manuscript.

## Response

At the end of introduction there is the following phrase:

*“The rest of this paper is divided in the following sections: in section* [*sec:Method-description*](#sec_Method_description) *the proposed method is fully described, in section* [*sec:Experiments*](#sec_Experiments) *the datasets used in the experiments are listed as well as the experimental results and finally in section* [*sec:Conclusions*](#sec_Conclusions) *some conclusions are provided.”*

## 5. Comment

According to the authors what is the objective problem?

## Response

We have changed this phrase to the more accurate: *“**The vector x &xrarr; is the input pattern from dataset describing the problem. For the rest of this paper the notation d will be used to represent the number of elements in x; .* ”

## 6. Comment

Authors may provide the architecture/block diagram of the proposed model for better comprehensibility of the proposed model concerning various aspects of the proposed model.

## Response

The flowchart of the proposed algorithm has been added at the end of subsection 2.3

## 7. Comment

How are the weights and biases considered in the current study?

## Response

The proposed method is used to train RBF networks with weights, centers and variances.

## 8. Comment

“The first phase of the proposed algorithm” does the authors mean to say “Crossover phase”.

## Response

We have added the following phrase at the beginning of subsection 2.2

*“The purpose of the first phase is to initalize the bounds of the RBF network and discover a promising interval for the corresponding values.”*

## 9. Comment

More comparative analysis with state-of-art models is desired.

## Response

Two methods have been added:

1) The Adam optimizer, used to train a neural network with 10 hidden nodes

2) The NeuroEvolution of Augmenting Topologies

## 10. Comment

By considering the current form of the conclusion section, it is hard to understand by AI Journal readers. It should be extended with new sentences about the necessity and contributions of the study by considering the authors' opinions about the experimental results derived from some other well-known objective evaluation values if it is possible.

## Response

The conclusion part has been rephrased and more material has been added for some guidelines for future improvements of the proposed method.

## 11. Comment

Authors should use more alternative models as the benchmarking models, authors should also conduct some statistical tests to ensure the superiority of the proposed approach, i.e., how could authors ensure that their results are superior to others? Meanwhile, the authors also have to provide some insightful discussion of the results.

## Response

We have updated the statistical tests including the added methods for comparison

1) The Adam optimizer, used to train a neural network with 10 hidden nodes

2) The NeuroEvolution of Augmenting Topologies

# Reviewer 3

## 1. Comment

Abstract needs to modify and to be revised to be more quantitative. You can absorb readers' consideration by having some numerical results in this section. It is suggested to discuss more about the findings of this study in the abstract.

## Response

The new abstract now reads:

*“Radial Basis Function networks are used in a variety of real-world applications such as medical data or signal processing problems.The success of these machine learning models lies in efficiently finding values for the model parameters. In this work, a new method of finding these values is formulated which is divided into two phases. In the first phase, with the use of Grammatical Evolution, an attempt is made to find value intervals for the model parameters using partition rules. In the second phase of the proposed method, an intelligent optimization algorithm such as a genetic algorithm, locates the optimal values of the parameters within the best value interval which is the output of the first phase. The proposed technique has been applied to a wide range of classification or data fitting problems and there has been a significant reduction in error exceeding 40% on most datasets.”*

## 2. Comment

The aim/objective of the paper is not mentioned in the Introduction section, a brief discussion required.

## Response

The following text has been added at the end of Introduction section:

*“A common method of calculating the parameters in these neural networks uses a technique to calculate the centers of the functions \phi ( x ) and then the weight vector w &xrarr; is calculated as a solution of a linear system of equations. Typically, the method used to calculate the centers is the well - known k-means method [*[*kmeans*](#LyXCite-kmeans)*]. In many cases this way of estimating the parameters of the neural network leads to over-fitting of the model so that it cannot generalize satisfactorily to unknown data. Furthermore, since there is no range of values for the parameters, there is the possibility that they will take extremely large or extremely small values, with the result that any generalizability of the model is lost. This work suggests a two phase method to minimize the error of equation (*[*eq:eqrbf*](#eq_eqrbf)*). During the first phase, an attempt is made to bound the parameter values to intervals in which the training error is likely to be significantly reduced. The identification of the most promising intervals for the parameters is performed using a technique that utilizes Grammatical Evolution[*[*ge1*](#LyXCite-ge1)*], that collects information from the training data. The first phase attempts to create a small interval of values for the neural network parameters by applying a series of of division rules, with the assistance of the Grammatical Evolution. The determination of the value interval is done in such a way that it is faster and more efficient to train the parameters of the neural network with some optimization method during the second phase of the method. In general, if the value intervals for the parameters from the first phase are small in range, the second phase of the technique is expected to be significantly accelerated as well. During the second phase, the parameters of the RBF network can be trained within the optimal range found in the first phase using some global optimization method [*[*rbfSA*](#LyXCite-rbfSA)*,* [*rbfPSO*](#LyXCite-rbfPSO)*]. In the proposed approach, the widely used method of genetic algorithm [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] was used for the second phase of the process.”*

## 3. Comment

Figures 1 should be converted to plain text (poor sharpness, and the text is not a drawing).

## Response

The grammar is placed in a Figure to refer to it from the paper.

## 4. Comment

The Conclusion section must be more extensive. The future scope of the work should be provided.

## Response

The conclusion part has been rephrased and more material has been added for some guidelines for future improvements of the proposed method.

## 5. Comment

As a minor formal remark, the use of acronyms should be avoided in both the title and the abstract.

## Response

Corrected.

# Reviewer 4

## 1. Comment

I cannot detect from the Introduction section what is presented in the paper.

## Response

The following text has been added at the end of Introduction section:

*“A common method of calculating the parameters in these neural networks uses a technique to calculate the centers of the functions φ ( x ) and then the weight vector w &xrarr; is calculated as a solution of a linear system of equations. Typically, the method used to calculate the centers is the well - known k-means method [*[*kmeans*](#LyXCite-kmeans)*]. In many cases this way of estimating the parameters of the neural network leads to over-fitting of the model so that it cannot generalize satisfactorily to unknown data. Furthermore, since there is no range of values for the parameters, there is the possibility that they will take extremely large or extremely small values, with the result that any generalizability of the model is lost. This work suggests a two phase method to minimize the error of equation (*[*eq:eqrbf*](#eq_eqrbf)*). During the first phase, an attempt is made to bound the parameter values to intervals in which the training error is likely to be significantly reduced. The identification of the most promising intervals for the parameters is performed using a technique that utilizes Grammatical Evolution[*[*ge1*](#LyXCite-ge1)*], that collects information from the training data. The first phase attempts to create a small interval of values for the neural network parameters by applying a series of of division rules, with the assistance of the Grammatical Evolution. The determination of the value interval is done in such a way that it is faster and more efficient to train the parameters of the neural network with some optimization method during the second phase of the method. In general, if the value intervals for the parameters from the first phase are small in range, the second phase of the technique is expected to be significantly accelerated as well. During the second phase, the parameters of the RBF network can be trained within the optimal range found in the first phase using some global optimization method [*[*rbfSA*](#LyXCite-rbfSA)*,* [*rbfPSO*](#LyXCite-rbfPSO)*]. In the proposed approach, the widely used method of genetic algorithm [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] was used for the second phase of the process.”*

## 2. Comment

Equations in the text should be cited as they are noted, so eq. (3) not just 3.

## Response

Done.

## 3. Comment

I do not get the notation with an arrow above some parameters, do they mean a vector? Then maybe simply a bold font would be appropriate? Of course an arrow above a variable means a vector but rather as a Euclidean vector and it seems to not fit in this meaning. Besides g in eq. (3) seems to also be a vector, why it does not have an arrow? Please go through the equations and correct this issue.

## Response

Yes, you are correct. We have use the arrow notation now.

## 4. Comment

c\_i is not a "so-called center", it does not explain what it actually is, so please clarify it.

## Response

We have added the following paragraph in Section 2

*“Every RBF network with k weigths is constructed by the following series of parameters:*

1. *A series of vectors c\_i ,i=1, … ,k called centers.*
2. *For every Gaussian unit an additional parameter σ\_i is required*
3. *The output weight vector w. ”*

## 5. Comment

Grammatical Evolution is a type of genetics algorithm; in the Abstract I got confused, thinking that two methods will be used in the paper.

## Response

The new abstract now reads:

*“Radial Basis Function networks are used in a variety of real-world applications such as medical data or signal processing problems.The success of these machine learning models lies in efficiently finding values for the model parameters. In this work, a new method of finding these values is formulated which is divided into two phases. In the first phase, with the use of Grammatical Evolution, an attempt is made to find value intervals for the model parameters using partition rules. In the second phase of the proposed method, an intelligent optimization algorithm such as a genetic algorithm, locates the optimal values of the parameters within the best value interval which is the output of the first phase. The proposed technique has been applied to a wide range of classification or data fitting problems and there has been a significant reduction in error exceeding 40% on most datasets.”*

## 6. Comment

The section 2.1 starts with the sentence: "Grammatical evolution is a genetic algorithm where the chromosomes stand for the production rules of any given BNF (Backus–Naur form) grammar[47]."  
It is confusing for so many reasons. Please explain at first what are genetic algorithms, what are chromosomes.

## Response

## 7. Comment

Figure 1 has a caption above the Figure, but first of all I do not understand it, should it explain anything?

## Response

We have rephrase the caption to the following:

*“The BNF grammar used in the current work, to produce intervals for the RBF parameters.”*

## 8. Comment

What does % mean in Tab. 1? Is it also a mod function?

## Response

We have added the following phrase in the corresponding caption:

*“Every number in cells denotes average classification error as measured on the test set.”*

## 9. Comment

Results do not seem "promising" as promised in the Abstract... Table 3 and 5 gives results in %, Table 4, also? What does the datasets contain? Maybe, because Authors use so many datasets it would be appropriate to describe those a bit more, give some examples? After reading the paper I am still confused with its contents. I understand a little bit what was done, but not enough. Authors write about RBFNN, but there is no NN in this paper described, no structure, no training. I also do not understand what was achived in the paper.

## Response

1) We have extended the captions in result tables to define what the numbers denote.

2) The datasets are fully described in subsection 3.1

3) The RBF networks are described in Introduction section.

4) The flowchart of the proposed algorithm has been added at the end of subsection 2.3

5) We have updated the statistical tests including the added methods for comparison

5.1) The Adam optimizer, used to train a neural network with 10 hidden nodes

5.2 ) The NeuroEvolution of Augmenting Topologies

6) The conclusion part has been rephrased and more material has been added for some guidelines for future improvements of the proposed method.

# Reviewer 5

## 1. Comment

An interesting study proposing an adaptation of the RBF network via genetic algorithms and specifically Grammatical Evolution. The manuscript is well suited to the journal and well structured without any linguistic issues detected.

The authors do analyze the algorithmic and mathematical background of the proposed method and have tested it on around 40 different data sets to evaluate its performance. Which was found superior compared to other traditional methods such as artificial neural networks and traditional the RBF method.

Did not detected any issues in this manuscript and in my opinion can be accepted as is.

## Response

Dear reviewer, thank you for your comments and the efforts to review this manuscript.