Comments for PSO – RBF

# Reviewer 1

## 1. Comment

Avoid using acronyms in abstract

## Response

Corrected.

## 2. Comment

Abstract is oddly short. Please expand it by describing the problem in few more details.

## Response

The abstract have changed to the following:

“*In the present work, an innovative two-phase method is presented for parameter tuning in Radial Basis Function artificial neural networks. These kinds of machine learning models find application in many scientific fields in classification problems or in function regression. In the first phase, a technique based on Particle Swarm Optimization is performed to find a promising interval of values for the network parameters. Particle swarm optimization was used as it is a highly reliable method for global optimization problems and in addition it is one of the fastest and most flexible techniques of its class. In the second phase, the network is trained within the optimal interval using a global optimization technique such as a Genetic Algorithm. Furthermore, in order to speed up the training of the network and due to the use of a two-stage method, parallel programming techniques were utilized. The new method was applied to a number of well-known classification and regression datasets, and the results were more than promising.*”

## 3. Comment

Make sure that each acronym has been defined at the first occurrence (RBF for example).

## Response

Corrected.

## 4. Comment

Elaborate in more details why PSO and GA were selected (among other, more recent metaheuristics available). These two algorithms are one of the oldest approaches.

## Response

The following paragraph has been added in the Introduction section:

“*In the current work, a hybrid technique is proposed for the optimal calculation of the parameters of an RBF network. This technique consists of two phases. During the first phase, information is collected from the training data of the neural network and an attempt is made to identify a small interval of values for the neural network parameters. To identify this interval, an optimization method is used, which gradually creates the optimal value interval, which is estimated to give the lowest value for the training error of the network. To locate the optimal interval, the particle swarm optimization (PSO) technique is used [*[*psoTutorial*](#LyXCite-psoTutorial)*]. The PSO method was chosen for the first phase because it is fast and flexible enough for optimization problems, does not require a large number of parameters to be input by the user, and has been successfully used in a variety of problems such as flow shop scheduling [*[*psoApp1*](#LyXCite-psoApp1)*], developing charging stragies for electric vehicles [*[*psoApp2*](#LyXCite-psoApp2)*], emotion recognition [*[*psoApp3*](#LyXCite-psoApp3)*], robot trajectury planning [*[*psoApp4*](#LyXCite-psoApp4)*] etc. The detection of the value interval is performed in order to then make the minimization of the network error faster and more efficient in the second phase optimization method. In the second phase, the parameters of the neural network are optimized within the optimal value interval of the first phase. The optimization can be performed by any global optimization method [*[*goReview*](#LyXCite-goReview)*]. In this work, the genetic algorithms [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] were chosen for the second phase. The main advantages of genetic algorithms are tolerance on errors, easy to implement in parallel, efficient exploration of the search space etc.* ”

## 5. Comment

Expand literature survey by mentioning other state-of-the-art metaheuristics that were used to tune machine learning models in general, such as:

* https://www.hindawi.com/journals/sp/2021/5540024/
* https://www.mdpi.com/2079-9292/11/22/3798
* https://link.springer.com/article/10.1007/s00521-017-2874-2
* https://ieeexplore.ieee.org/abstract/document/8678770

## Response

Done. The following paragraph has been added to the revised manuscript:

“*Recently, many works have been appeared to tune the parameters of of machine learning models such as the work of Agarwal and Bhanot [*[*rbftune*](#LyXCite-rbftune)*] for the adaptation of the RBF parameters, the incorporation of an improved ABC algorithm to adapt the parameters of RBF networks [*[*rbfABC*](#LyXCite-rbfABC)*], the usage of the Firefly algorithm for optimization [*[*fireflyMain*](#LyXCite-fireflyMain)*] along with machine learning models for Cervical cancer diagnosis [*[*fireflyCancer*](#LyXCite-fireflyCancer)*], adaptation of CNN and XGBOOST models by an optimization algorithm for COVID-19 diagnosis [*[*hybridCNN*](#LyXCite-hybridCNN)*] etc*.”

## 6. Comment

Figure 2 layout is broken, please address it.

## Response

Corrected.

## 7. Comment

Figures 3 and 4, caption is partly hidden by images, please address it.

## Response

Corrected.

## 8. Comment

Discussion should be more elaborate.

## Response

The following discussion has been added/ modified:

“As one can see from the experimental results, the proposed method significantly outperforms the other techniques in the majority of cases in terms of the average error in the test set. Moreover, the difference from the established method of training RBF networks is of the order of 40% and in some cases this percentage can be doubled. The statistical difference of the proposed technique against the rest is also shown in graphs [fig:class\_error] and [fig:graph\_regression]. However, the proposed technique is significantly slower than the original training technique, as it is a two-stage technique. In the first stage, an optimal interval of values for the network parameters is created with a modified PSO method, and in the second stage, the network is trained using a genetic algorithm. Of course, this extra time can be significantly reduced by incorporating parallel techniques, as was done experimentally using the OpenMP library. Furthermore, changing the normalization factor \lambda from 100 to 1000 did not have much effect on the mean error in the test set. This induces that the proposed method is quite robust, since it doesn't have much dependence on this parameter.

An additional experiment was performed with different values for the parameter F. The experimental results for this experiment are shown in Table [tab:expsFClass] for the classification datasets and in Table [tab:expsFRegression] for the regression datasets. And for this critical parameter, no large deviations appear in the results of the proposed method. This further enhances the robustness and reliability of the proposed technique.

Also, in the Table [tab:precisionRecall], the metrics of precision, recall and f-score are shown for a series of classification datasets and for the proposed method (mentioned as IRBF-100) and the classic method for training RBF networks (mentioned as RBF-KMEANS) .And in these experimental results, the reader can see the superiority of the proposed technique over the traditional method of training RBF networks.”

## 9. Comment

Other metrics should be presented as well not only the accuracy. Include precision, recall, f-score as well. These additional metrics can be significant for imbalanced datasets (such as credit card frauds dataset).

## Response

An additional table has been added in the revised text, where the metrics of precision, recall and f-score are listed for a series of classification datasets. The added text reads:

“*Also, in the Table* [*tab:precisionRecall*](#tab_precisionRecall) *the metrics of precision, recall and f-score are shown for a series of classification datasets and for the proposed method (mentioned as IRBF-100) and the classic method for training RBF networks (mentioned as RBF-KMEANS) .And in these experimental results, the reader can see the superiority of the proposed technique over the traditional method of training RBF networks*.”

## 10. Comment

Average accuracy isn't reliable indicator when it is calculated as average of 20+ dataset, some of them imbalanced, some not, some with large number of entries, etc. Better, indicate on how many datasets out of total number of datasets the proposed approach obtained the best accuracy, for example 15 out of 24.

## Response

We have extended the AVERAGE line. Now it displays the number of times where the corresponding method achieved the best error.

## 11. Comment

Indicate future work in the conclusion.

## Response

The conclusion section has been re – written. At the end we have included the following paragraph:

“*The method could be extended by the use of other techniques of training the parameters in RBF networks, such as for example the differential evolutionary method [*[*dePaper*](#LyXCite-dePaper)*]. Furthermore, more efficient methods of terminating the first stage of the method could be used as finding a suitable interval of values for the network parameters requires many numerical calculations.*”

# Reviewer 2

## 1. Comment

Page 3: Algorithms 1: Should “Do” be replaced with “do”?

## Response

Corrected.

## 2. Comment

Page 4 Line 95: The reviewer think that “do” has not been deepened.

## Response

Corrected.

## 3. Comment

Page 5: In figure 2, there should be spacing between images and captions.

## Response

Corrected.

## 4. Comment

Page 5 Line 105: It is necessary for the author to check the sequencing.

## Response

It is a format effect of LaTeX.

## 5. Comment

Page 6 Line 156-160: After the URL of the dataset, please mark the specific date of access.

## Response

Added.

## 6. Comment

Page 9 – Page 10: Relevant pictures of the experimental results should be placed in the previous section.

## Response

Fixed.

## 7. Comment

Page 11: I hope the author can adjust the formatting to improve the problem of image obscuring the title.

## Response

The article has been reformatted using the MDPI Template.

## 8. Comment

The appendix section is missing funding, statements, etc. And the title and layout of the references also need further improvement.

## Response

The article has been reformatted using the MDPI Template.

# Reviewer 3

## 1. Comment

There are already many works on the calculation of RBF network parameters. What is the reason for creating another algorithm? The main motivation is not clearly explained in the paper, please explain: What was the motivation for taking up the topic?

## Response

The following paragraph has been added in the Introduction section:

“*In the current work, a hybrid technique is proposed for the optimal calculation of the parameters of an RBF network. This technique consists of two phases. During the first phase, information is collected from the training data of the neural network and an attempt is made to identify a small interval of values for the neural network parameters. To identify this interval, an optimization method is used, which gradually creates the optimal value interval, which is estimated to give the lowest value for the training error of the network. To locate the optimal interval, the particle swarm optimization (PSO) technique is used [*[*psoTutorial*](#LyXCite-psoTutorial)*]. The PSO method was chosen for the first phase because it is fast and flexible enough for optimization problems, does not require a large number of parameters to be input by the user, and has been successfully used in a variety of problems such as flow shop scheduling [*[*psoApp1*](#LyXCite-psoApp1)*], developing charging stragies for electric vehicles [*[*psoApp2*](#LyXCite-psoApp2)*], emotion recognition [*[*psoApp3*](#LyXCite-psoApp3)*], robot trajectury planning [*[*psoApp4*](#LyXCite-psoApp4)*] etc. The detection of the value interval is performed in order to then make the minimization of the network error faster and more efficient in the second phase optimization method. In the second phase, the parameters of the neural network are optimized within the optimal value interval of the first phase. The optimization can be performed by any global optimization method [*[*goReview*](#LyXCite-goReview)*]. In this work, the genetic algorithms [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] were chosen for the second phase. The main advantages of genetic algorithms are tolerance on errors, easy to implement in parallel, efficient exploration of the search space etc.* ”

## 2. Comment

Which algorithm was used in the second phase? In the abstract, the authors write that the second phase uses PSO algorithm and in the "introduction" section that the second phase uses a genetic algorithm. Why? Could the authors explain this?

## Response

Yes, you are correct. We have re – write the abstract in order to clarify this and now it reads:

“*In the present work, an innovative two-phase method is presented for parameter tuning in Radial Basis Function artificial neural networks. These kinds of machine learning models find application in many scientific fields in classification problems or in function regression. In the first phase, a technique based on Particle Swarm Optimization is performed to find a promising interval of values for the network parameters. Particle swarm optimization was used as it is a highly reliable method for global optimization problems and in addition it is one of the fastest and most flexible techniques of its class. In the second phase, the network is trained within the optimal interval using a global optimization technique such as a Genetic Algorithm. Furthermore, in order to speed up the training of the network and due to the use of a two-stage method, parallel programming techniques were utilized. The new method was applied to a number of well-known classification and regression datasets, and the results were more than promising.*”

## 3. Comment

Line 64 - the author write: “In order to locate the best interval for the parameters of the network, a modified PSO algorithm [43] is used”. - Unfortunately, article [43] is a review of many PSO algorithms. Could the author explain which modified algorithm was used?

## Response

To clarify this, we have added the following text:

“*The proposed variant of the PSO method is based on the original technique (algorithm 1 of[*[*pso1*](#LyXCite-pso1)*] ) but the particles are intervals of values and at each iteration a normalization of the velocity vector takes place to avoid generating particles outside the original range of values.*”

## 4. Comment

Figures/algorithms should be placed on the same page as close to the reference as possible.

## Response

Done.

## 5. Comment

In the text, we can refer to figures/algorithms/equations that are in earlier chapters/sections or in the current chapter/section. We should not refer to figures/algorithms/equations that have not yet been explained/are in later chapters because they make the content difficult to understand.

## Response

Corrected.

## 6. Comment

Algorithm 1 point 2 (equation) - sj2  or ẟj2 ?

## Response

We have changed to \sigma\_{j}^{2}

## 7. Comment

Line 86 – Point 8 is unclear. Why should Ng particles be initialized? Shouldn't it be: " initialize Nc particles"? Why "layout of each chromosome.."? - PSO is based on particles not on chromosomes. Could the author please explain this?

## Response

It was a typo. We have corrected it.

## 8. Comment

Line 97 - Why velocity uij=[-rẟ/20, rẟ/20] ? Have other velocities been considered?

## Response

The following paragraph has been added at this point:

“*The velocity is initialized to a small sub-interval of the range of values for the corresponding parameter in order to avoid, as far as possible, excessive values for the velocity. This would result in the particles moving out of their value range very quickly and thus making the optimization process difficult.* ”

## 9. Comment

There are many kinds of inertia weight w including constant inertia weight [1], linearly decreasing inertia weight [2], exponential inertia weight [3], random inertia weight [4], dynamic inertia weight [5], fuzzy inertia weight [6], It would be good to write this and then explain: What  guided the author  when choosing the kind of inertia weight?

1. Constant inertia weight: <https://doi.org/10.1016/S0020-0190(02)00447-7>

**2. Linearly decreasing inertia weight: doi:**10.1109/CEC.1999.785511

3. Exponential inertia weight: https://doi.org/10.1007/978-3-319-46592-0\_23

4. Random inertia weight: https://doi.org/10.1007/3-540-45105-6\_12,   **doi:**10.1109/ISCBI.2013.27

5. Dynamic inertia weight: <https://doi.org/10.1007/978-3-319-70581-1_6>

6. Fuzzy inertia weight: **doi:**10.1109/CEC.2001.934377 ; **doi:**10.1109/JCAI.2009.50

## Response

The following text has been added:

“*Many inertia calculations appeared in the relevant literarature such as constant inertia [*[*psoConstant*](#LyXCite-psoConstant)*], linearly decreasing inertia [*[*psoLinear*](#LyXCite-psoLinear)*], exponential inertia [*[*psoExp*](#LyXCite-psoExp)*], random inertia calculation [*[*psoRandom*](#LyXCite-psoRandom)*], dynamic inertia [*[*psoDynamic*](#LyXCite-psoDynamic)*], fuzzy inertia calculation [*[*psoFuzzy*](#LyXCite-psoFuzzy)*] etc. The present method of calculating the inertia was chosen because it decreases linearly with time and for large values of the inertia it allows a wider search in the search space and for low values it allows a more focused search.* ”

## 10. Comment

Line 104 - Could the authors explain why acceleration coefficients were omitted from the velocity equation?

## Response

The equation for the velocity has been updated to include c1 and c2. Also, the used values for these parameters are added in Table 1.

## 11. Comment

Part of Figure 2 is cut off. The title of figure 2 unnecessarily merges/sticks with the table.

## Response

Fixed.

## 12. Comment

The results (tables, figures) should be placed in the results section. Why are they included in the conclusion?

## Response

Fixed.

## 13. Comment

Why are references in the conclusion section?

## Response

Fixed.

## 14. Comment

The paper lacks comparative analysis of experimental results with other related methods.

## Response

Two additional methods have been added: The use of the RPROP method in an artificial neural network with 10 hidden nodes and the use of a genetic algorithm to an artificial neural network with 10 hidden nodes. Hence, two columns have been added in the experimental results (Tables 2 and 3) in subsection 3.2. Also, graphical comparison between all methods for all datasets have been included on separate graphs.

## 15. Comment

Lack of discussion of results

## Response

The following discussion has been added/ modified:

“As one can see from the experimental results, the proposed method significantly outperforms the other techniques in the majority of cases in terms of the average error in the test set. Moreover, the difference from the established method of training RBF networks is of the order of 40% and in some cases this percentage can be doubled. The statistical difference of the proposed technique against the rest is also shown in graphs [fig:class\_error] and [fig:graph\_regression]. However, the proposed technique is significantly slower than the original training technique, as it is a two-stage technique. In the first stage, an optimal interval of values for the network parameters is created with a modified PSO method, and in the second stage, the network is trained using a genetic algorithm. Of course, this extra time can be significantly reduced by incorporating parallel techniques, as was done experimentally using the OpenMP library. Furthermore, changing the normalization factor \lambda from 100 to 1000 did not have much effect on the mean error in the test set. This induces that the proposed method is quite robust, since it doesn't have much dependence on this parameter.

An additional experiment was performed with different values for the parameter F. The experimental results for this experiment are shown in Table [tab:expsFClass] for the classification datasets and in Table [tab:expsFRegression] for the regression datasets. And for this critical parameter, no large deviations appear in the results of the proposed method. This further enhances the robustness and reliability of the proposed technique.

Also, in the Table [tab:precisionRecall], the metrics of precision, recall and f-score are shown for a series of classification datasets and for the proposed method (mentioned as IRBF-100) and the classic method for training RBF networks (mentioned as RBF-KMEANS) .And in these experimental results, the reader can see the superiority of the proposed technique over the traditional method of training RBF networks.”

## 16. Comment

Lack of details about the weaknesses and strengths of the proposed model

## Response

We have re – write the conclusion section and now it reads:

“*In the present work, a two-stage hybrid method was proposed to efficiently identify the parameters of RBF neural networks. In the first stage of the method, a technique rooted in particle swarm optimization was used to efficiently identify a reliable interval of values for the neural network parameters. In the second stage of the method, an intelligent global optimization technique was used to locate the neural network parameter within the optimal value interval of the first stage. In this work, a genetic algorithm was used in the second phase, however any global optimization method could be used in its place.*

*The method was applied to a multitude of classification and regression problems from the relevant literature. In almost all cases, the proposed method significantly outperforms other machine learning models, and on average the improvement in the error on the test sets was of the order of 40% relative to the established RBF training method. Moreover, the method is quite robust with respect to the basic parameters since any changes in the parameter values do not significantly affect its performance. Furthermore, the method can efficiently locate the value interval of network parameters without any prior knowledge about the type of training data or whether it is a classification or a regression problem. However, the proposed technique is significantly more time-consuming than the traditional training technique as it requires computational time for both its phases. Although, this effect can be overcomed to some extent by the use of modern parallel computing techniques.*

*The method could be extended by the use of other techniques of training the parameters in RBF networks, such as for example the differential evolutionary method [*[*dePaper*](#LyXCite-dePaper)*]. Furthermore, more efficient methods of terminating the first stage of the method could be used as finding a suitable interval of values for the network parameters requires many numerical calculations.*”

## 17. Comment

The paper contains typos and mistakes which should be improved

## Response

(edo na poume pos egine xrisi ispell kai grammar check)

## 18. Comment

References contains typos and mistakes that should be improved e.g. [28],[38], [39], etc.

## Response

Corrected.

# Reviewer 4

## 1. Comment

The article proposes a technique for training RBF NN. Although RBF is relatively known, its abbreviation should be properly introduced in the article.

## Response

Corrected.

## 2. Comment

The paper does not have a proper introduction and it does not present a section with similar attempts to identify parameters of RBF networks or at least for NN in general or even deep learning, if there are not enough recent entries in the literature, although I believe there are quite a few:

-        https://ieeexplore.ieee.org/document/7095057

-        https://www.sciencedirect.com/science/article/pii/S0307904X11006251

-        https://www.mdpi.com/1424-8220/22/11/4204

-        https://www.mdpi.com/2227-7390/9/21/2705

-        https://www.inderscienceonline.com/doi/abs/10.1504/IJAIP.2022.126695

-        https://link.springer.com/chapter/10.1007/978-3-030-39033-4\_3

## Response

The following paragraph has been added in the Introduction section:

“*Recently, many works have been appeared to tune the parameters of of machine learning models such as the work of Agarwal and Bhanot [*[*rbftune*](#LyXCite-rbftune)*] for the adaptation of the RBF parameters, the incorporation of an improved ABC algorithm to adapt the parameters of RBF networks [*[*rbfABC*](#LyXCite-rbfABC)*], the usage of the Firefly algorithm for optimization [*[*fireflyMain*](#LyXCite-fireflyMain)*] along with machine learning models for Cervical cancer diagnosis [*[*fireflyCancer*](#LyXCite-fireflyCancer)*], adaptation of CNN and XGBOOST models by an optimization algorithm for COVID-19 diagnosis [*[*hybridCNN*](#LyXCite-hybridCNN)*] etc.*”

## 3. Comment

The authors should elaborate on why their technique should be preferred to other similar attempts.

## Response

The following paragraph has been added in the Introduction section:

“*In the current work, a hybrid technique is proposed for the optimal calculation of the parameters of an RBF network. This technique consists of two phases. During the first phase, information is collected from the training data of the neural network and an attempt is made to identify a small interval of values for the neural network parameters. To identify this interval, an optimization method is used, which gradually creates the optimal value interval, which is estimated to give the lowest value for the training error of the network. To locate the optimal interval, the particle swarm optimization (PSO) technique is used [*[*psoTutorial*](#LyXCite-psoTutorial)*]. The PSO method was chosen for the first phase because it is fast and flexible enough for optimization problems, does not require a large number of parameters to be input by the user, and has been successfully used in a variety of problems such as flow shop scheduling [*[*psoApp1*](#LyXCite-psoApp1)*], developing charging stragies for electric vehicles [*[*psoApp2*](#LyXCite-psoApp2)*], emotion recognition [*[*psoApp3*](#LyXCite-psoApp3)*], robot trajectury planning [*[*psoApp4*](#LyXCite-psoApp4)*] etc. The detection of the value interval is performed in order to then make the minimization of the network error faster and more efficient in the second phase optimization method. In the second phase, the parameters of the neural network are optimized within the optimal value interval of the first phase. The optimization can be performed by any global optimization method [*[*goReview*](#LyXCite-goReview)*]. In this work, the genetic algorithms [*[*ga1*](#LyXCite-ga1)*,* [*ga2*](#LyXCite-ga2)*,* [*ga3*](#LyXCite-ga3)*] were chosen for the second phase. The main advantages of genetic algorithms are tolerance on errors, easy to implement in parallel, efficient exploration of the search space etc.* ”

## 4. Comment

A diagram with an overview of the methodology should be presented to give a more intuitive presentation of the proposed technique.

## Response

We have added a diagram for the proposed method at the end of subsection 2.3

## 5. Comment

Eq. (7) needs either correction or more explanations. How can an interval be equal with a sum?

## Response

This paragraph changed to “The function E( y ) (equation [eq:eqrbf](#eq_eqrbf)) is modified to an interval one [ E min ( y ) , E max ( y ) ] calculated with the procedure given in Algorithm [alg:Fitness-calculation-for](#alg_Fitness_calculation_for).”

## 6. Comment

The settings from Table 1 should be explained. Actually, all the captions from the tables and figures should contain more explanations.

## Response

The text in all captions have been expanded in the revised manuscript.

## 7. Comment

A (at least one) baseline model should be included in the experiments to indicate the benefits of employing the proposed method.

## Response

Two additional methods have been added: The use of the RPROP method in an artificial neural network with 10 hidden nodes and the use of a genetic algorithm to an artificial neural network with 10 hidden nodes. Hence, two columns have been added in the experimental results (Tables 2 and 3) in subsection 3.2. Also, graphical comparison between all methods for all datasets have been added.

## 8. Comment

Figures 3 and 4 does not bring anything new as compared to the tables. Numerical information about the running times should be presented, as well. The captions from the figures cannot be read well

## Response

An additional table has been added in the revised text, where the metrics of precision, recall and f-score are listed for a series of classification datasets. The added text reads:

“*Also, in the Table* [*tab:precisionRecall*](#tab_precisionRecall) *the metrics of precision, recall and f-score are shown for a series of classification datasets and for the proposed method (mentioned as IRBF-100) and the classic method for training RBF networks (mentioned as RBF-KMEANS) .And in these experimental results, the reader can see the superiority of the proposed technique over the traditional method of training RBF networks*.”