A novell genetic algorithm method to train RBF networks

Ioannis G. Tsoulos, Alexandros Tzallas, Evangelos Karvounis

Department of Informatics and Telecommunications, University of Ioannina, Greece

1 Introduction

2 Method description

On each client a genetic algorithm with the termination rule described in [1] accompanied with an additional local search operator is applied. The steps of the algorithm are given below:

1. **Initialization step**
   (a) **Set** iter=0, where iter is the current number of generations
   (b) **Set** N_f the number of desired features.
   (c) **Set** N_c as the total chromosomes.
   (d) **Initialize** chromosomes $X_i, i = 1 \ldots N_c$ The chromosomes are initialized randomly as vectors of integers.
   (e) **Set** ITERMAX as the maximum number of allowed generations.
   (f) **Set** p_s as the selection rate and p_m the mutation rate. Both rates are in the range $[0, 1]$
   (g) **Set** $f_l = \infty$, the best discovered fitness
   (h) **Set** L_I the number of generations that should pass before the local search procedure is applied.
   (i) **Set** L_c the number of chromosomes that will participate in local search procedure.

2. **Termination check.** At every generation the variance $\sigma^{(\text{iter})}$ of f_l is calculated. If there was no improvement of the genetic algorithm for a number of generations, then the algorithm should terminate. The stopping rule has as follows:

$$|f_h - f_l| \leq e \text{ OR } \sigma^{(\text{iter})} \leq \frac{\sigma^{(\text{last})}}{2} \quad \text{OR iter>ITERMAX} \qquad (1)$$

Where last denotes the generation number where f_l was produced initially. **If** equation 1 is true **then Goto** step 9.

3. **Calculate** the fitness f_i for every chromosome of the population:
   (a) **Transform** the original train data using grammatical evolution and create N_f features.
   (b) **Train** a classification model C_i obtaining the train error E
   (c) **Assign** the train error E to fitness value f_i
4. **Genetic Operators**
   (a) **Selection procedure**: The chromosomes are sorted in descending order according to their fitness value. The first $(1-p_s)\times N_c$ chromosomes are transferred to the next generation. The rest of the chromosomes are substituted by offsprings created through one point crossover procedure.
   (b) **Mutation procedure:** For every element of each chromosome a random number r in range $[0,1]$ is produced. If $r\le p_m$ then the corresponding element is randomly altered.
   (c) **Replace** the $p_s\times N_c$ worst chromosomes in the population with the offsprings created by the genetic operators.
5. **Set** iter=iter+1
6. **Local Search Step**
   (a) **If** iters $\mod L_i=0$ **Then**
      i. **Select** randomly L_C chromosomes from the genetic population and create the set L_S from these chromosomes
      ii. **For** every chromosome X_i **in** L_S
         A. **Select** randomly another chromosome Y from the population
         B. **Create** an offspring of $X_i$and Y using one point crossover. Denote the offspring as Z
         C. **Obtain** the fitness $f(Z)$ of chromosome Z**.** **If** $f(z)<f_i$ then $X_i=Z,\ f_i=f(Z)$
   (b) **Endif**
7. **Obtain** the best value in the population, denoted as f_l for the corresponding chromosome x_l
8. **Send** (x_l,f_l) to Server machine and **Goto** step 3
9. **Send** (x_l,f_l) to Server machine and **Terminate**

3 Experiments

3.1 Experimental setup

(nention the Optimus)

3.2 Experimental datasets

All the experiments were conducted 30 times using different seed for the random generator each time and averages were taken. The following datasets were used

1. **Wine** dataset. The wine recognition dataset contains data from wine chemical analysis. It contains 178 examples of 13 features each that are classified into three classes.

2. **Glass** dataset. The dataset contains glass component analysis for glass pieces that belong to 6 classes. The dataset contains 214 examples with 10 attributes each.

3. Liverdisorder: This dataset, which will be denoted by LIVER in the following tables, contains blood analysis data from people with liver disorders. It consists of 345 examples of 6 features each.

4. **Tae** dataset. The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison.

5. Liverdisorder dataset.

6. **Spiral** dataset: The spiral artificial dataset contains 1000 two-dimensional examples that belong to two classes (500 examples each). The number of the features is 2. The data in the first class are created using the following formula: $x_1 = 0.5t\cos\left(0.08t\right),\ x_2 = 0.5t\cos\left(0.08t + \frac{\pi}{2}\right)$ and the second class data using**:** $x_1 = 0.5t\cos\left(0.08t + \pi\right),\ x_2 = 0.5t\cos\left(0.08t + \frac{3\pi}{2}\right)$

7. **Pima** dataset. The Pima Indians Diabetes dataset contains 768 examples of 8 attributes each that are classified into two categories: healthy and diabetic.

8. **Ionosphere** dataset. The ionosphere dataset (ION in the following tables) contains data from the Johns Hopkins Ionosphere database. The two-class dataset contains 351 exam\ples of 34 features each.

9. **Appendictis** dataset, proposed in [3].

10. **Australian** dataset, the dataset concerns credit card applications.

11. **Hayes roth** dataset. This dataset[4] contains **5** numeric-valued attributes and 132 patterns.

12. **Alcohol** dataset, a dataset about Alcohol consumption [5].

13. **Dermatology** dataset. Dataset used for differential diagnosis of erythemato-squamous diseases. The dataset has 366 instances of 34 features each.

14. **Balance** dataset. This data set was generated to model psychological experimental results. It contains 625 patterns of 4 features each.

15. **Regions2** dataset. It is created from liver biopsy images of patients with hepatitis C [7]. From each region in the acquired images, 18 shape-based and color-based features were extracted, while it was also annotated form medical experts. The resulting dataset includes 600 samples belonging into 6 classes.

16. **Parkinsons** dataset. This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD)[6].

17. **Wdbc** dataset. The Wisconsin diagnostic breast cancer dataset (WDBC) contains data for breast tumors. It contains 569 patterns of 30 features each.

18. **Popfailures** dataset. This dataset contains records of simulation crashes encountered during climate model uncertainty quantification (UQ) ensembles. It contains 540 patterns of 18 features each.

19. **Heart** dataset. The task is to detect the absence or presence of heart disease. It contains 270 patterns of 13 features each.

20. **Haberman** dataset. A dataset about breast cancer from a study at the University of Chicago's Billings Hospital. It contains 306 patterns of 3 features each.

21. **HouseVotes** dataset. This data set includes votes for each of the U.S. House of Representatives Congressmen on the 16 key votes.

22. **Lymography** dataset. The aim here is to detect the presence of a lymphoma in patients. It contains 148 patterns of 18 features each.

23. **Mammographic** dataset. This dataset be used to identify the severity (benign or malignant) of a mammographic mass lesion from BI-RADS attributes and the patient's age. It contains 830 patterns of 5 features each.

24. **OptDigits** dataset. Optical Recognition of Handwritten Digits data set. It contains 5620 patterns of 64 features each.

25. **Page Blocks** dataset. The dataset contains blocks of the page layout of a document that has been detected by a segmentation process. It has 5473 patterns with 10 features each.

26. **Penbased** dataset. This is a Pen-Based Recognition of Handwritten Digits data set with 10992 patterns of 16 features.

27. **Saheart** dataset. The dataset is about to categorize persons if have a coronary heart disease. The dataset contains 462 patterns with 9 features each.

28. **Segment** dataset. This database contains patterns from a database of 7 outdoor images (classes). The dataset contains 2310 patterns with 19 features each.

29. Eeg dataset. As an real word example, consider an EEG dataset described in [8] is used here. The dataset consists of five sets (denoted as Z, O, N, F and S) each containing 100 single-channel EEG segments each having 23.6 sec duration. Sets Z and O have been taken from surface EEG recordings of five healthy volunteers with eye open and closed, respectively. Signals in two sets have been measured in seizure-free intervals from five patients in the epileptogenic zone (F) and from the hippocampal formation of the opposite hemisphere of the brain (N). Set S contains seizure activity, selected from all recording sites exhibiting ictal activity. Sets Z and O have been recorded extracranially, whereas sets N, F and S have been recorded intracranially.

30. Fertility Data Set (FERT): 100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration is related to socio-demographic data, environmental factors, health status, and life habits. It contains 100 examples of 10 features each.

The regression datasets are available from the Statlib URL `ftp://lib.stat.cmu.edu/datasets/index.html`:

1. **Abalone** dataset. This data set can be used to obtain a model to predict the age of abalone from physical measurements.

2. **Airfoil** dataset. This is a NASA data set, obtained from a series of aerodynamic and acoustic tests [9]. The dataset has 1503 instances of 6 features each.

3. **Anacalt** dataset. This contains information about the decisions taken by a supreme court. The dataset has 4052 instances of 7 features each.

4. **BK** dataset. This dataset comes from Smoothing Methods in Statistics [14] and is used to estimate the points scored per minute in a basketball game. The dataset has 96 patterns of 4 features each.

5. **BL** dataset: This dataset can be downloaded from StatLib. It contains data from an experiment on the affects of machine adjustments on the time to count bolts. It contains 40 patters of 7 features each.

6. **Concrete** dataset. This dataset is taken from civil engineering[10]. The dataset has 1030 instances of 8 features each.

7. **Housing** dataset. This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University and it is described in [11].

8. **Laser** dataset. Dataset used in laser experiments. The dataset has 993 instances of 4 features each.

9. **MB** dataset. This dataset is available from Smoothing Methods in Statistics [12] and it includes 61 patterns. The number of features is 2.

10. **NT** dataset. This dataset contains data from [13] that examined whether the true mean body temperature is 98.6 F. The number of patterns is 131 and the number of features 2.

11. **Quake** dataset. The objective here is to approximate the strength of a earthquake. The dataset has 2178 instances of 3 features each.

12. **Wankara** dataset. This dataset is used to predict the mean temperature for the Ankara. The dataset has 1609 datasets of 9 features each.

13. **FA** dataset. The FA dataset contains percentage of body fat, age, weight, height, and ten body circumference measurements. The goal is to fit body fat to the other measurements. The number of the features is 18. The total number of patterns is 252.

14. **PY** dataset (Pyrimidines problem). The source of this dataset is the URL: `https://www.dcc.fc.up.pt/~ltorgo/Regression/DataSets.html` and it is a problem of 27 attributes and 74 number of patterns. The task consists of Learning Quantitative Structure Activity Relationships (QSARs) and provided by [15].

References

[1] I.G. Tsoulos, Modifications of real code genetic algorithm for global optimization, Applied Mathematics and Computation **203**, pp. 598-607, 2008.

[2] J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. Journal of Multiple-Valued Logic and Soft Computing 17, pp. 255-287, 2011.

[3] Weiss, Sholom M. and Kulikowski, Casimir A., Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems, Morgan Kaufmann Publishers Inc, 1991.

Table 1: Experimental results for classification problems.

DATASET	KRBF	GENRBF(3)	GENRBF(5)
Alcohol	46.63%		
Appendicitis	12.23%		
Australian	34.89%		
Balance	33.42%		
Bands	37.22%		
Cleveland	67.10%		
Dermatology	62.34%		
Ecoli	59.59%		
Glass	50.16%		
Haberman	25.10%		
Hayes Roth	64.36%		
Heart	31.20%		
HouseVotes	6.13%		
Ionosphere	16.22%		
Liverdisorder	30.84%		
Lymography	25.31%		
Mammographic	21.38%		
Page Blocks	10.09%		
Parkinsons	17.42%		
Pima	25.78%		
Popfailures	7.04%		
Regions2	38.29%		
Ring	21.65%		
Saheart	32.19%		
Segment	59.68%		
Sonar	27.85%		
Spiral	44.87%		
Tae	60.07%		
Thyroid	10.52%		
Wdbc	7.27%		
Wine	31.41%		
Z_F_S	13.16%		
Z_O_N_F_S	48.71%		
ZO_NF_S	9.02%		
ZONF_S	4.03%		
ZOO	21.93%		

Table 2: Experiments for regression datasets.

DATASET	KMEANS RBF	GENRBF(3)	GENRBF(5)
ABALONE	7.371		
AIRFOIL	0.037		
ANACALT	11.628		
BK	0.165		
BL	0.05		
CONCRETE	1.15		
HOUSING	57.682		
LASER	2.35		
MB	1.915		
NT	2.755		
QUAKE	0.071		
WANKARA	0.18		

[4] B. Hayes-Roth, B., F. Hayes-Roth. Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior **16**, pp. 321-338, 1977.

[5] Tzimourta, Katerina D. and Tsoulos, Ioannis and Bilero, Thanasis and Tzallas, Alexandros T. and Tsipouras, Markos G. and Giannakeas, Nikolaos, Direct Assessment of Alcohol Consumption in Mental State Using Brain Computer Interfaces and Grammatical Evolution, Inventions **3**, pp. 1-12, 2018.

[6] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng. 2009;56(4):1015. doi:10.1109/TBME.2008.2005954

[7] Giannakeas, N., Tsipouras, M.G., Tzallas, A.T., Kyriakidi, K., Tsianou, Z.E., Manousou, P., Hall, A., Karvounis, E.C., Tsianos, V., Tsianos, E. A clustering based method for collagen proportional area extraction in liver biopsy images (2015) Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-November, art. no. 7319047, pp. 3097-3100.

[8] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E **64**, pp. 1-8, 2001.

[9] T.F. Brooks, D.S. Pope, and A.M. Marcolini. Airfoil self-noise and prediction. Technical report, NASA RP-1218, July 1989.

[10] I.Cheng Yeh, Modeling of strength of high performance concrete using artificial neural networks, Cement and Concrete Research. **28**, pp. 1797-1808, 1998.

[11] D. Harrison and D.L. Rubinfeld, Hedonic prices and the demand for clean ai, J. Environ. Economics & Management **5**, pp. 81-102, 1978.

[12] J.S. Simonoff, Smooting Methods in Statistics, Springer - Verlag, 1996.

[13] Mackowiak, P.A., Wasserman, S.S., Levine, M.M., 1992. A critical appraisal of 98.6 degrees f, the upper limit of the normal body temperature, and other legacies of Carl Reinhold August Wunderlich. J. Amer. Med. Assoc. 268, 1578–1580

[14] J.S. Simonoff, Smooting Methods in Statistics, Springer - Verlag, 1996.

[15] R.D. King, S. Muggleton, R. Lewis, M.J.E. Sternberg, Proc. Nat. Acad. Sci. USA **89**, pp. 11322–11326, 1992.