

Comparison of
HYBRID
APPROACHES TO
EVOLUTIONARY
ALGORITHMS
Using Data Clustering

DR. G. MALINI DEVI

Associate Professor

Department of Computer Science & Engineering
G. Narayanamma Institute of Technology and Science
(For women), Hyderabad, Telangana, INDIA

COMPARISON OF HYBRID APPROACHES TO EVOLUTIONARY ALGORITHMS USING DATA CLUSTERING

Copyright© : Dr. G. Malini Devi
Publishing Rights© : VSRD Academic Publishing
A Division of Visual Soft India Pvt. Ltd.

ISBN-13: 978-93-91462-65-9
FIRST EDITION, JULY 2023, INDIA

Printed & Published by:
VSRD Academic Publishing
(A Division of Visual Soft India Pvt. Ltd.)

Disclaimer: The author(s) / Editor(s) are solely responsible for the contents compiled in this book. The publishers or its staff do not take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the Author(s) or Editor(s) or Publishers to avoid discrepancies in future.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the Publishers & Author.

Printed & Bound in India

VSRD ACADEMIC PUBLISHING
A Division of Visual Soft India Pvt. Ltd.

REGISTERED OFFICE

154, Tezab mill Campus, Anwarganj, KANPUR–208003 (UP) (IN)
Mb:9899936803, Web: www.vsrdpublishing.com, Email: vsrdpublishing@gmail.com

MARKETING OFFICE

340, FF, Adarsh Nagar, Oshiwara, Andheri(W), MUMBAI–400053 (MH) (IN)
Mb:9956127040, Web: www.vsrdpublishing.com, Email: vsrdpublishing@gmail.com

PREFACE


Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. Depending on the nature of the data and the purpose for which clustering is being used, different measures of similarity may be used to place items into classes, where the similarity measure controls how the clusters are formed.

The main aim of this project is to find the efficient optimization and similarity measures for Particle Swarm Optimization, Relevance Feedback, Genetic Algorithm, GAPSO and RFBPSO. The performance was measured by the fitness function value, accuracy, and the CPU time. A general thumb rule is that a clustering with lower fitness function value, lower proximity error and lower CPU time is preferable. The effectiveness of these algorithms was tested on 12 datasets, Iris, Wine, Hayes Roth, Glass, Zoo, Contact Lenses, Seeds, Haber man, Teaching Assistant Evolution, Lenses, Concrete, Lung Cancer are taken.

The Genetic algorithm is used to solve a variety of optimization problems that are not well suited for standard optimization algorithms. GA has long been used in different kinds of complex problems, usually with encouraging results. In PSO, the population dynamics resembles the movement of a “*bird flock*” searching for food, where a social sharing of information takes place and individuals can gain from the discoveries and previous experience of their companions Relevance Feedback algorithm which is used to retrieve the data based upon the user tastes and preferences. Along with the neighbors which are found with the help of KNN

algorithm, additional neighbors are added to the existing neighbors because it also takes the user preferences into consideration and thus increasing the accuracy. Accuracy should be measured by taking into consideration the true positives and true negatives. The accuracy should be calculated for all the algorithms applied on different datasets on different distance measures.

Finally, the proposed hybrid algorithm of RFB and PSO i.e., RFBPSO gives better results than the individual algorithms i.e., RFB and PSO. And GA and PSO i.e., GAPSO gives better results than GA and PSO. We hope that " Comparison of Hybrid Approaches to Evolutionary Algorithms Using Data Clustering" will serve as a valuable resource in your journey to explore and understand the captivating realm of data clustering. Our aspiration is that this book will inspire curiosity, foster critical thinking, and empower readers.

 *Author*

CONTENTS

CHAPTER 1 : OVERVIEW OF DATA CLUSTERING ALGORITHM.....	1
1.1. INTRODUCTION.....	1
1.2. DEFINITIONS.....	3
1.3. DATA MINING	4
1.4. TYPES OF CLUSTERING METHODS.....	5
1.5. GENETIC ALGORITHMS	9
1.6. RELEVANCE FEEDBACK	10
1.7. PARTICLE SWARM OPTIMIZATION.....	10
1.8. PROBLEM DEFINITION	10
1.9. OBJECTIVES	12
1.10. METHODOLOGY	14
1.11. STUDY AREA AND CHARACTERISTICS.....	14
CHAPTER 2 : SOFTWARE DESCRIPTION	17
2.1. INTRODUCTION TO MATLAB	17
2.2. THE MATLAB SYSTEM.....	18
2.3. MATLAB WORKING ENVIRONMENT	20
CHAPTER 3 : RELEVANCE FEEDBACK	23
3.1. INTRODUCTION.....	23
3.2. RELEVANCE FEEDBACK ALGORITHM.....	24
3.3. PROBLEMS WITH RELEVANCE FEEDBACK.....	27
3.4. ADVANTAGES OF RELEVANCE FEEDBACK	27

CHAPTER 4 : PARTICLE SWARM OPTIMIZATION 28

4.1. INTRODUCTION..... 28
4.2. NOTATIONS..... 29
4.3. ALGORITHM 30
4.4. APPLICATIONS..... 31
**4.5. RELEVANCE FEEDBACK WITH PARTICLE SWARM
OPTIMIZATION (RFB+PSO)..... 31**

CHAPTER 5 : GENETIC ALGORITHM..... 33

5.1. INTRODUCTION..... 33
5.2. GENETIC ALGORITHMS 34
5.3. HOW GENETIC ALGORITHMS WORK..... 35
5.4. GENETIC ALGORITHMS OPERATORS 36
5.5. GENETIC ALGORITHMS PARAMETERS..... 37
5.6. HOW CROSSOVER AND MUTATION WORK..... 39
5.7. GA CHARACTERISTICS..... 40
5.8. ALGORITHM 45
5.9. APPLICATIONS..... 53

**CHAPTER 6 : PARTICLE SWARM OPTIMIZATION
AND GENETIC ALGORITHM (GA+PSO)..... 55**

6.1. INTRODUCTION..... 55
6.2. ALGORITHM FOR GA+PSO 56
6.3. PROBLEMS WITH PSO AND NEED FOR GA+PSO 56
6.4. ADVANTAGES OF GA+PSO..... 57

CHAPTER 7 : EXPERIMENTAL RESULTS..... 58

7.1. INTRODUCTION..... 58

7.2.	RFB ALGORITHM	58
7.3.	PSO	85
7.4.	RFBPSO ALGORITHM.....	88
7.5.	OUTPUTS GENERATED BY GA WITH PSO ALGORITHM.....	96

CHAPTER 8 : CONCLUSION AND FUTURE ENHANCEMENTS		96
8.1.	CONCLUSION.....	96

CHAPTER 9 : REFERENCES.....		98
------------------------------------	--	-----------

