A Feature Selection Method based on Binary Black Hole Algorithm for Spam E-Mail Filtering

Prepared by
Amal Mahmood Abdullah Abdulla

Supervised by
Dr. Adnan Hadi Mahdi Al-Helali

A Thesis
Submitted to Faculty of Information Technology as a Partial Fulfilment of the Requirement for Master Degree in Software Engineering

December 2019
Authorization Form

I, Amal Mahmood Abdullah Abdullah, authorize the Isra University to supply copies of my Thesis to libraries or establishments or individuals on request.

Signature: [Signature]
Date: [Date]
نموذج التفويض

أنا امّال مهمرد عبد الله، أفوض جامعة الامراء بتوزيع نسخ من رساليتي للمكتبات أو المؤسسات أو الهيئات أو الأشخاص عند طلبها.

التوقيع:
التاريخ:
This Thesis (A Feature Selection Method based on Binary Black Hole Algorithm for Spam E-Mail Filtering) was successfully defended and approved on

Examination Committee  

Dr. Adnan Hadi Mahdi Al-Helali

Dr. Bassam Al-Shargabi

Dr. 'Mohammed Ali’ Al-Janini

Signature
DEDICATIONS

-To who I carry his name and I miss his since childhood, my dear father.

-To the path of guidance, patience and hope, my ideal mother.

-To alone son of my mother, who his support me after my father death despite of young age, my beloved brother (Abdulla), also his wife and daughter.

-To all who have place for me (my brothers, sisters, their wives, their husbands and all their sons) and to all my friends and Relatives.
ACKNOWLEDGMENTS

-First of all, my great thanks to the god who give me the ability to achieve this work.

-I would like to express my appreciation to my supervisor Dr. Adnan Al-Helali for this help and guidance to complete this thesis.

-To all my family, friends and everyone who help me and care for me.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER ONE INTRODUCTION</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>1.2 PROBLEM STATEMENT</td>
<td>3</td>
</tr>
<tr>
<td>1.3 RESEARCH QUESTIONS</td>
<td>4</td>
</tr>
<tr>
<td>1.4 THE AIM AND OBJECTIVES</td>
<td>4</td>
</tr>
<tr>
<td>1.5 THE SCOPE</td>
<td>4</td>
</tr>
<tr>
<td>1.6 RESEARCH METHODOLOGY</td>
<td>5</td>
</tr>
<tr>
<td>1.7 THESIS ORGANIZATION</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER TWO THEORETICAL BACKGROUND</td>
<td>8</td>
</tr>
<tr>
<td>2.1 INTRODUCTION</td>
<td>8</td>
</tr>
<tr>
<td>2.2 ELECTRONIC MAIL (E-MAIL)</td>
<td>8</td>
</tr>
<tr>
<td>2.2.1 HISTORY OF E-MAILS</td>
<td>8</td>
</tr>
<tr>
<td>2.2.2 SPAM E-MAIL</td>
<td>9</td>
</tr>
<tr>
<td>2.2.3 SPAM FILTER SYSTEM</td>
<td>11</td>
</tr>
<tr>
<td>2.3 FEATURE SELECTION AND CLASSIFICATION</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1 FEATURE SELECTION METHODS</td>
<td>16</td>
</tr>
<tr>
<td>2.3.2 NAIVE BAYSIAN CLASSIFIER (NBC)</td>
<td>20</td>
</tr>
<tr>
<td>2.4 BLACK HOLE (BH) ALGORITHM</td>
<td>21</td>
</tr>
<tr>
<td>2.4.1 THE SOURCE OF INSPIARTION AND MATHEMATICAL MODEL</td>
<td>21</td>
</tr>
<tr>
<td>2.5 RELATED WORK</td>
<td>24</td>
</tr>
</tbody>
</table>
2.5.1 BLACK HOLE ALGORITHM FOR DIFFERENT OPTIMIZATION PROBLEM

2.5.2 MACHINE LEARNING MODELS FOR E-MAIL SPAM FILTERING

CHAPTER THREE THE PROPOSED ALGORITHM

3.1 INTRODUCTION

3.2 THE PROPOSED EMAIL SPAM FILTERING SYSTEM

3.3 THE PROPOSED FEATURE SELECTION ALGORITHM

3.3.1 INFORMATION GAIN (IG)

3.3.2 BINARY BLACK HOLE (BBH) ALGORITHM

3.4 MATRICES OF EVALUATION

CHAPTER FOUR RESULTS AND DISCUSSION

4.1 INTRODUCTION

4.2 EXPERIMENTAL SETTINGS

4.3 RESULTS AND DISCUSSION

4.3.1 RESULTS OF INFORMATION GAIN (IG)

4.3.2 RESULTS OF BBH

4.4 COMPARISON WITH RELATED WORKS

CHAPTER FIVE CONCLUSION AND FUTURE WORKS

5.1 CONCLUSION

5.2 SUGGESTION FOR FUTURE WORKS

REFERENCE
LIST OF FIGURES

FIGURE 1-1 RESEARCH METHODOLOGY ......................................................... 6
FIGURE 2-1 A SAMPLE OF EMAIL MESSAGE............................................. 10
FIGURE 2-2 GENERAL BLOCK DIAGRAM FOR THE FEATURE SELECTION
ALGORITHMS ................................................................................................ 17
FIGURE 2-3 TYPES OF FEATURE SELECTION METHODS BASED ON THE
TYPES OF FEATURE AND DATASETS .......................................................... 18
FIGURE 2-4 GRAPHICAL REPRESENTATION FOR FILTER MODEL ............. 19
FIGURE 2-5 GRAPHICAL REPRESENTATION FOR WRAPPER MODEL ........ 19
FIGURE 2-6 BLACK HOLE ALGORITHM (Hatamlou, 2013) ......................... 22
FIGURE 3-1 THE MAIN BLOCK DIAGRAM OF THE PROPOSED FILTERING
SYSTEM ........................................................................................................... 33
FIGURE 3-2 SOLUTION REPRESENTATION FOR THE STARS ................. 36
FIGURE 3-3 THE FLOWCHART OF THE PROPOSED ALGORITHM ............. 37
FIGURE 4-1 RESULTS OBTAINED USING 10 STARS ................................. 44
FIGURE 4-2 RESULTS OBTAINED USING 20 STARS ................................. 45
FIGURE 4-3 RESULTS OBTAINED USING 30 STARS ................................. 46
FIGURE 4-4 RESULTS OBTAINED USING 40 STARS ................................. 47
FIGURE 4-5 RESULTS OBTAINED USING 50 STARS ................................. 48