A Hybrid Approach using Fuzzy Logic and MapReduce to Achieve Meaningful using of Big Data

By
Ikhlas Hassan Almukahel
Supervisor
Prof. Dr. Mohammed Alfayomi
Co-supervisor
Dr. wael Alzyadat

This thesis was submitted in partial fulfillment of the requirements for the Master’s Degree in Software Engineering
Faculty of Graduate Studies
ISRA University
August 2018
Authorization statement

I Ikhlas Hasan Almukahel, authorize Isra University to provide hard copies or soft copies of my thesis to libraries institutions or individuals upon their request.

Name: Ikhlas Hasan Almukahel
Signature:
Date:
أنا إخلاص حسن المكحل أفوض جامعة الأسرايا للدراسات العليا بتزويد نسخ من رسالتي ورقية والكترونية للمكتبات أو المنظمات أو الهيئات والمؤسسات المعنية بالأبحاث والدراسات العليا عند طلبها.

الاسم: إخلاص حسن المكحل
التوقيع:
التاريخ:
The undersigned have examined the thesis entitled Hybrid Approach using Fuzzy Logic and MapReduce to Achieve Meaningful used Big Data presented by Ikhlas Almukahel, a candidate for the degree of Master of Software Engineering and hereby certify that it is worthy of acceptance.

Date ____________________________ Date ____________________________
Supervisors name

Date ____________________________ Date ____________________________
Co-supervisors name

Date ____________________________ Date ____________________________
committee name 1

Date ____________________________ Date ____________________________
committee name 2
DEDICATION

This thesis is dedicated:

-To my great parents, who never stop presenting support and being the constant source of love and encouragement.

-To my brilliant adviser Dr. Wael Alzyadat, who leads me through the valley of darkness with light of hope and support.

Ikhlas Hassan Almukahel
August 2018
ACKNOWLEDGMENT

First and foremost, Praise is to God the Almighty for giving me strength, knowledge, ability and opportunity. Without His blessings, this achievement would never have been possible.

I would kindly present my thanks to my supervisor Dr. Mohammad Al-Alfayomi, who has been so helpful and cooperative in giving support and advice.

I'd like to express massive appreciation with very profound gratitude and personal thanks to my elegant co-supervisor Dr. wael Alzyadat who encouraged, supported, and took the time to suggest changes, improvements or clarifications. Although his suggestions caused more work for me, I believe that the end result is the best that I could achieve.

I owe everything to my family for being patient and understanding while I set the normal flow of life aside in order to focus on the research. This research would also have never been possible without the support of a number of people at ISRA University, who are too many people to list all of them. Those people helped me and provided me with information, in addition to answering many of my questions to achieve my goal. They all deserve to be acknowledged as well.
Table of content

AUTHORIZATION ........................................................................................................... II
ARABIC AUTHORIZATION ............................................................................................ III
COMMITTEE’s DECISION ............................................................................................... IV
DEDICATION .................................................................................................................. V
ACKNOWLEDGMENT ....................................................................................................... VI
LIST OF TABLES ............................................................................................................. VII
LIST OF FIGURES .......................................................................................................... VIII
LIST OF ABBREVIATIONS .............................................................................................. IX

ABSTRACT .................................................................................................................... X

CHAPTER ONE INTRODUCTIONS .................................................................................... 1
1.1 Overview .................................................................................................................. 1
1.2 Motivations .............................................................................................................. 1
1.3 Problem Statement .................................................................................................. 2
1.4 Research Questions .................................................................................................. 2
1.5 Research Objectives ............................................................................................... 3
1.6 Significance ............................................................................................................. 3
1.7 Scope ...................................................................................................................... 3
1.8 General Methodology .............................................................................................. 3
1.9 Conceptual Hybrid Approach using Fuzzy Logic and MapReduce ......................... 4

CHAPTER TWO Literature Review ................................................................................ 7
2.1 Overview .................................................................................................................. 7
2.2 Keywords ................................................................................................................. 7
2.2.1 Big Data .............................................................................................................. 8
2.2.2 MapReduce ........................................................................................................ 8
2.2.3 Classification ...................................................................................................... 9
LIST OF TABLES

Table 2.1: Comparison among Approaches .............................................................. 13
Table 2.2: Comparison among Frameworks and Techniques .................................... 14
Table 3.1: Attributes of dataset .............................................................................. 18
Table 3.2: Normal reading of some attribute .............................................................. 21
Table 3.1: Abbreviation of Attributes ...................................................................... 22
Table 3.4: Rule Base ............................................................................................... 23
Table 4.1: Comparison between R and Weka in monitoring data .............................. 28
Table 4.2: Comparison between R and Weka in Preprocessing ................................. 29
Table 4.3: Comparison between R and Weka with missing value ............................. 30
Table 4.4: Comparison between R and Weka in map-function ................................. 31
Table 4.5: Comparison between R and Weka in reduce-function .............................. 32
Table 4.6: Comparison between R and Weka in predict and evaluation .................. 33
Table 4.7: Evaluation comparison between R and Weka ......................................... 33
LIST OF FIGURES

Figure 1.1: General Methodology of research design………………………………………………………… 3
Figure 1.2: Proposed Conceptual Hybrid Approach using Fuzzy Logic and MapReduce to achieve Meaningful used Big Data……………………………………………………………………………… 5
Figure2.1: Architecture of MapReduce ................................................................. 8
Figure2.2: Main structure of fuzzy Logic................................................................. 10
Figure 3.1: Design of Hybrid Approach using Fuzzy Logic and MapReduce to achieve Meaningful used Big Data……………………………………………………………………………… 17
Figure3.2: Fuzzy value design model………………………………………………………… 21
Figure3.3: Flow chart of build fuzzy rule. ............................................................... 21
Figure3.4: Remove duplicated rules………………………………………………………… 22
Figure 3.5 Confusion matrix………………………………………………………………… 24
Figure 4.1 Data collections component…………………………………………………… 27
Figure 5.1 Confusion matrixes for R packages and Weka steps .............................. 34
Figure 5.2: Measurement for R packages and Weka …………………………………… 34
Figure 5.3: Result difference among R packages and Weka ................................. 35
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>#</th>
<th>Abbreviation</th>
<th>Full Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>FLC</td>
<td>Fuzzy Logic Controller</td>
</tr>
<tr>
<td>2.</td>
<td>DT</td>
<td>Decision Tree</td>
</tr>
<tr>
<td>3.</td>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>4.</td>
<td>4V's</td>
<td>volume, velocity ,variety and value</td>
</tr>
<tr>
<td>5.</td>
<td>Weka</td>
<td>Waikato Environment for Knowledge Analyze</td>
</tr>
<tr>
<td>6.</td>
<td>UCI</td>
<td>University of California Irvine</td>
</tr>
<tr>
<td>7.</td>
<td>HDFS</td>
<td>Hadoop Distributed File System</td>
</tr>
<tr>
<td>8.</td>
<td>Npreg</td>
<td>Pregnancies</td>
</tr>
<tr>
<td>9.</td>
<td>Glu</td>
<td>Glucose</td>
</tr>
<tr>
<td>10.</td>
<td>BP</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>11.</td>
<td>Skin</td>
<td>Skin Thickness</td>
</tr>
<tr>
<td>12.</td>
<td>PED</td>
<td>Diabetes Pedigree Function</td>
</tr>
</tbody>
</table>
Abstract

Big data is facing many challenges in different aspects, which appears through characteristics such as: volume, velocity and variety. Processing and analyzing big data are challenging issues to acquire quality information to support accurate values for decision making. Quality data taxonomy point to three basic elements which are meaningfulness, predicatively, and decision making, where all of these elements encourage investigating previous work that focused on the same challenges of big data. Consequently, the Hybrid Approach using Fuzzy Logic and MapReduce produces a new version of MapReduce consisting of four layers with data collection in the first layer, while the second layer preprocesses and treats semi-structured data clean up. Map function to acquire relationships, applying fuzzy controller and classification to generate rules among occur data in the third layer, whereas the last layer includes reduce and classification to achieve a predicative and meaningful outcome.

Experiment A Hybrid Approach uses Diabetes patient data and symptoms in R (Readr, Dplyr, TidyR, PreProcess, HadoopStreaming, HiveR and FuzzyR) packages and Weka; Evaluation confirmed significant efficiency of the proposed approach through precision, recall and F-measure.