

MODEL-BASED CLUSTERING OF DIGITAL PCR DROPLETS USING EXPECTATION MAXIMIZATION

A Thesis Proposal
Presented to
the Faculty of the College of Science
De La Salle University Manila

In Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Statistics

by

GUIAO, Joyce Emlyn B.

Frumencio F. Co
Adviser

March 6, 2020

Abstract

From 150 to 200 words of short, direct and complete sentences, the abstract should be informative enough to serve as a substitute for reading the thesis document itself. It states the rationale and the objectives of the research.

In the final thesis document (i.e., the document you'll submit for your final thesis defense), the abstract should also contain a description of your research results, findings, and contribution(s).

Keywords can be found at <http://www.acm.org/about/class/class/2012?pageIndex=0>. Click the link "HTML" in the paragraph that starts with "The full CCS classification tree...".

Keywords: Keyword 1, keyword 2, keyword 3, keyword 4, etc.

Contents

1	Research Description	1
1.1	Introduction	1
1.2	Background of the Study	1
1.3	Statement of the Problem	1
1.4	Significance of the Study	1
1.5	Scope and Limitations	1
2	Review of Related Literature	2
2.1	dPCR Droplet Classification Methods	2
2.1.1	Threshold	2
2.1.2	Population Detection of Gaussian Kernel Densities	2
2.1.3	K Nearest Neighbors	2
2.1.4	Non-parametric Mixture Models	2
2.2	Model-Based Clustering	2
3	Theoretical Framework	3
3.1	G-component Finite Mixture Density	3
3.2	Expectation Maximization Clustering	3

4	Methodology	4
4.1	Data	5
4.1.1	Rain Experiment Dataset	5
4.1.2	DNA Quantification Dataset	5
4.2	Model Fitting and Classification	5
4.3	Performance Evaluation	5
A	Diagrams and Other Documentation Tools	6
B	Theoretical and/or Conceptual Framework	7
C	Resource Persons	8
	References	9

List of Figures

List of Tables

Chapter 1

Research Description

1.1 Introduction

1.2 Background of the Study

1.3 Statement of the Problem

1.4 Significance of the Study

1.5 Scope and Limitations

Chapter 2

Review of Related Literature

2.1 dPCR Droplet Classification Methods

2.1.1 Threshold

2.1.2 Population Detection of Gaussian Kernel Densities

2.1.3 K Nearest Neighbors

2.1.4 Non-parametric Mixture Models

2.2 Model-Based Clustering

Chapter 3

Theoretical Framework

3.1 G-component Finite Mixture Density

3.2 Expectation Maximization Clustering

Chapter 4

Methodology

4.1 Data

4.1.1 Rain Experiment Dataset

Plate 2 - Primer and Probe Concentration Gradient

Plate 4 - PCR Enhancers Experiment

Plate 5 - Cycle Gradient

Plate 6 - Sonication Gradient

Plate 7 - Annealing Temperature Gradient

4.1.2 DNA Quantification Dataset

Plate 3 - Rain Dilution Series

Albumin

4.2 Model Fitting and Classification

4.3 Performance Evaluation

Appendix A

Diagrams and Other Documentation Tools

This appendix may consist of proposed architectural design, algorithms, scientific formula for MSCS and Data Flow Diagrams, Fishbone for MSIT.

Appendix B

Theoretical and/or Conceptual Framework

Discusses the basic framework/foundation the thesis is based on. This section is normally referred to when discussing Scope and Limitations, and Research Methodology

Appendix C

Resource Persons

Dr. Firstname1 Lastname1

Adviser

College of Computer Studies

De La Salle University-Manila

emailaddr@dlsu.edu.ph

Mr. Firstname2 Lastname2

Role2

Affiliation2

emailaddr2@domain.com

Ms. Firstname3 Lastname3

Role3

Affiliation3

emailaddr3@domain.net

References