AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Computer Engineering and Systems

Metagenomic data analysis using deep learning

A Thesis submitted in partial fulfillment of the requirements of
Master of Science in Electrical Engineering
(Computer Engineering and Systems)

by
Aly O. Abdelkareem
Bachelor of Science in Electrical Engineering
(Computer Engineering and Systems)
Faculty of Engineering, Ain Shams University, 2016

Supervised By
Prof. Hazem M. Abbas
Dr. Mahmoud I. Khalil

Cairo, 2018

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Computer Engineering and Systems

Metagenomic data analysis using deep learning

by

Aly O. Abdelkareem
Bachelor of Science in Electrical Engineering
(Computer Engineering and Systems)
Faculty of Engineering, Ain Shams University, 2016

Examiners' Committee

Name and affiliation **Signature**

Prof.
Computer Engineering and Systems
Faculty of Engineering, Ain Shams University. . . . .

Prof.
Computer Engineering and Systems
Faculty of Engineering, Ain Shams University. . . . .

Dr.
Choose Department
Faculty of Engineering, University. . . . .

Date:25 Feb 2019

Statement

This thesis is submitted as a partial fulfillment of Master of Science in Electrical Engineering, Faculty of Engineering, Ain Shams University. The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

Aly O. Abdelkareem

Signature

....

Date: 25 Dec 2018

Researcher Data

Name: Aly Osama Aly Ibrahim Abdelkareem (Aly O. Abdelkareem)

Date of Birth: 29/08/1992

Place of Birth: Cairo, Egypt

Last academic degree: Bachelor of Science in Electrical Engineering

Field of specialization: Computer Engineering and Software Systems (Credit Hours)

University issued the degree : Ain Shams University

Date of issued degree : 15/6/2016

Current job : Teaching and Research Assistant at Faculty of Engineering

Thesis Summary

Summary

The thesis is divided into seven chapters as listed below:

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Key words: bioinformatics, classification, deep learning, metagenomics

Acknowledgment

Aly O. Abdelkareem
Computer Engineering and Systems
Faculty of Engineering
Ain Shams University
Cairo, Egypt
Dec 2018

I would like to express my sincere gratitude to my mother and my father for the continuous support of my study, my wife Mayada for her patience and motivation, and my daughter Aisha as she brought happiness to my life.

Contents

List of Figures

List of Tables

Abbreviations

LAH	**L**ist **A**bbreviations **H**ere

Symbols

a	distance	m
P	power	W (Js^{-1})
ω	angular frequency	rads^{-1}

Chapter 1

Introduction

1.1 Metagenomic analysis

Welcome to this LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on LaTeX to make them look stunning.

1.1.1 Definition

1.1.2 Microorganism

1.2 Viruses

LaTeX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for LaTeX is actually a simple, plain text file that contains *no formatting*. You tell LaTeX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the '`\textit{}`' command and put the text I want in italics in between the curly braces. This means that LaTeX is a "mark-up" language, very much like HTML.

1.2.1 Definition

If you are new to LaTeX, there is a very good eBook – freely available online as a PDF file – called, "The Not So Short Introduction to LaTeX". The book's title is typically shortened to just "lshort". You can download the latest version (as it is occasionally updated) from here:

`http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf`

It is also available in several other languages. Find yours from the list on this page:

`http://www.ctan.org/tex-archive/info/lshort/`

It is recommended to take a little time out to learn how to use LaTeX by creating several, small 'test' documents. Making the effort now means you're not stuck learning the system when what you *really* need to be doing is writing your thesis.

1.2.2 Importance in clinical and environment

If you are writing a technical or mathematical thesis, then you may want to read the document by the AMS (American Mathematical Society) called, "A Short Math Guide for LaTeX". It can be found online here:

http://www.ams.org/tex/amslatex.html

under the “Additional Documentation” section towards the bottom of the page.

1.2.3 Identification

1.3 Next Generation Sequencing

1.3.1 Tools

1.3.2 Data

1.4 Our Contribution

Guide written by —

Sunil Patel: www.sunilpatel.co.uk

Chapter 2

Related Work

2.1 Similarity tools

Welcome to this LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on LaTeX to make them look stunning.

2.2 Statistical tools

LaTeX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for

LaTeX is actually a simple, plain text file that contains *no formatting*. You tell LaTeX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the '`\textit{}`' command and put the text I want in italics in between the curly braces. This means that LaTeX is a "mark-up" language, very much like HTML.

Guide written by —
Sunil Patel: www.sunilpatel.co.uk

Chapter 3

Deep neural networks for identification

3.1 Convolution neural networks

LaTeX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for LaTeX is actually a simple, plain text file that contains *no formatting*. You tell LaTeX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the '`\textit{}`' command and put the text I want in italics in between the curly braces. This means that LaTeX is a "mark-up" language, very much like HTML.

3.1.1 Triplet loss mechanism

3.2 Sequence neural networks

3.2.1 Attention mechanism

Guide written by —

Sunil Patel: www.sunilpatel.co.uk

Chapter 4

Experimental results

4.1 Dataset generation

Welcome to this LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on LaTeX to make them look stunning.

4.2 Simulated Metagenome

4.3 Real metagenome casestudy

Guide written by —

Sunil Patel: www.sunilpatel.co.uk

Chapter 5

Conclusion and Future Work

5.1 Summary

Welcome to this LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on LaTeX to make them look stunning.

5.2 Conclusion

5.3 Future Work

Guide written by —

Sunil Patel: www.sunilpatel.co.uk

Appendix A

Appendix Title Here

Write your Appendix content here.

Bibliography

[1] A. S. Arnold, J. S. Wilson, and M. G. Boshier. A simple extended-cavity diode laser. *Review of Scientific Instruments*, 69(3):1236–1239, March 1998. URL http://link.aip.org/link/?RSI/69/1236/1.

[2] Carl E. Wieman and Leo Hollberg. Using diode lasers for atomic physics. *Review of Scientific Instruments*, 62(1):1–20, January 1991. URL http://link.aip.org/link/?RSI/62/1/1.

[3] C. J. Hawthorn, K. P. Weber, and R. E. Scholten. Littrow configuration tunable external cavity diode laser with fixed direction output beam. *Review of Scientific Instruments*, 72(12):4477–4479, December 2001. URL http://link.aip.org/link/?RSI/72/4477/1.