EFFICIENT PARTIALLY OBSERVABLE MARKOV DECISION PROCESS BASED FORMULATION OF GENE REGULATORY NETWORK CONTROL PROBLEM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

UTKU ERDOĞDU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

APRIL 2012

Approval of the thesis:

EFFICIENT PARTIALLY OBSERVABLE MARKOV DECISION PROCESS BASED FORMULATION OF GENE REGULATORY NETWORK CONTROL PROBLEM

submitted by **UTKU ERDOĞDU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Prof. Dr. Faruk Polat
Supervisor, **Computer Engineering Department, METU** _____

Prof. Dr. Reda Alhajj
Co-supervisor, **Computer Science Dept., Univ. of Calgary** _____

Examining Committee Members:

Prof. Dr. Varol Akman
Computer Engineering Department, Bilkent University _____

Prof. Dr. Faruk Polat
Computer Engineering Department, METU _____

Assoc. Prof. Dr. Tolga Can
Computer Engineering Department, METU _____

Assoc. Prof. Dr. Halit Oğuztüzün
Computer Engineering Department, METU _____

Assist. Prof. Dr. Mehmet Tan
Computer Engineering Department, TOBB ETÜ _____

Date:30.04.2012

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Utku Erdoğdu

Signature :

ABSTRACT

EFFICIENT PARTIALLY OBSERVABLE MARKOV DECISION PROCESS BASED FORMULATION OF GENE REGULATORY NETWORK CONTROL PROBLEM

Erdoğdu, Utku
Ph.D., Department of Computer Engineering
Supervisor: Prof. Dr. Faruk Polat
Co-Supervisor: Prof. Dr. Reda Alhajj

April 2012, 9 pages

The need to analyze and closely study the gene related mechanisms motivated the research on the modeling and control of gene regulatory networks (GRN). Different approaches exist to model GRNs; they are mostly simulated as mathematical models that represent relationships between genes. Though it turns into a more challenging problem, we argue that partial observability would be a more natural and realistic method for handling the control of GRNs. Partial observability is a fundamental aspect of the problem; it is mostly ignored and substituted by the assumption that states of GRN are known precisely, prescribed as full observability. On the other hand, current works addressing partially observability focus on formulating algorithms for the finite horizon GRN control problem. So, in this work we explore the feasibility of realizing the problem in a partially observable setting, mainly with Partially Observable Markov Decision Processes (POMDP). We proposed a POMDP formulation for the infinite horizon version of the problem. Knowing the fact that POMDP problems suffer from the curse of dimensionality, we also proposed a POMDP solution method that automatically decomposes the problem by isolating different unrelated parts of

the problem, and then solves the reduced subproblems. We also proposed a method to enrich gene expression data sets given as input to POMDP control task, because in available data sets there are thousands of genes but only tens or rarely hundreds of samples. The method is based on the idea of generating more than one model using the available data sets, and then sampling data from each of the models and finally filtering the generated samples with the help of metrics that measure compatibility, diversity and coverage of the newly generated samples.

Keywords: Gene Regulatory Networks, Partially Observable Markov Decision Process, Control of GRN, Gene Expression Data, Data Enrichment

ÖZ

GEN AĞLARININ KISMİ GÖZLEMLENEBİLİR MARKOV SÜREÇLERİ İLE MODELLENEREK ETKİN OLARAK KONTROLÜ

Erdoğdu, Utku
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Faruk Polat
Ortak Tez Yöneticisi: Prof. Dr. Reda Alhajj

Nisan 2012 , 9 sayfa

Genlerin çalışma prensiplerini inceleme gereksinimi gen düzenleyici ağların (GDA) modellenmesi ve kontrolü üzerine bilimsel çalışmalar yapılmasına yol açmıştır. GDA'ları modellemek için değişik yaklaşımlar mevcuttur ve bu yaklaşımların çoğu genler arasındaki ilişkileri matematiksel modeller vasıtasıyla modellemektedir. Problemi daha zorlaştırmasına rağmen, GDA kontrol problemlerinin daha doğal ve gerçekçi çözülebilmesi için kısmi gözlemlenebilirliğin önerilmesi gerektiğini savunuyoruz. Kısmi gözlemlenebilirlik bu problemin temel bir bileşeni olmasına rağmen çoğunlukla gözardı edilmiş ve problemin çözümünde GDA'nın tüm durumlarının mükemmel olarak bilinebileceği varsayımı yapılmıştır, yani problem tam gözlemlenebilir kabul edilmiştir. Bir yandan da literatürdeki kısmi gözlemlenebilirliği dikkate alan yöntemler sınırlı adımdan oluşan bir problem tanımı ile GDA kontrol problemini çözen algoritmalar üretmeye çalışmaktadır. Bu çalışmada problemin kısmi gözlemlenebilir bir kurgu ile tanımlanması üzerinde çalışılmakta ve Kısmi Gözlemlenebilir Markov Karar Süreçleri (POMDP) bu kurguda kullanılmaktadır. Bu çalışmada problemin sonsuz adım-

dan oluşan bir hali problem POMDP modeline uygun bir şekilde tanımlanarak sunulmaktadır. POMDP problemlerinin boyutlardan kaynaklanan problemler yaşamasından dolayı POMDP problemlerin birbirinden bağımsız parçalarını ayırıp problemi otomatik olarak parçalayan ve bu parçaları çözerek problemin çözümünü bulan bir çözüm yaklaşımı da bu çalışmada sunulmuştur. Bu çalışmada ayrıca POMDP kontrol problemine girdi olarak verilen gen ifade verisini zenginleştirmek için de bir metot sunulmaktadır. Gen ifade verilerinde binlerce gen olmasına rağmen genelde onlarca, nadir olarak da yüz mertebesinde örneklem bulunduğundan böyle bir metot gerekli ve faydalıdır. Sunulan metot varolan veri kümesini kullanarak birden fazla model oluşturur; her modelden yeni veri noktaları üretikten sonra üretilen veri noktalarını veri kümesinin uygunluğunu, farklılığını ve genişliğini ölçen metrikler yardımıyla filtreyerek kullanıma hazır bir veri kümesi oluşturur.

Anahtar Kelimeler: Gen Düzenleyici Ağlar, Kısmi Gözlemlenebilir Markov Karar Süreçleri, GDA'ların Kontrolü, Gen İfade Verisi, Veri Zenginleştirme

To my grandparents

Hatice Oruç, İbrahim Oruç, Necibe Erdoğdu, Şehmus Erdoğdu

ACKNOWLEDGMENTS

I would like to thank my supervisor Professor Faruk Polat for his constant support, guidance and friendship. It was a great honor to work with him for the last twelve years and our cooperation influenced my academical and world view highly. I also would like to thank Professor Reda Alhajj for his support and guidance on my stay at Calgary. While away from my home, he not only supported me on my research but also provided that I feel welcome and personally attended all my needs and problems. He also motivated and influenced me highly in scientific context.

A lot of people influenced and supported this work scientifically and their contribution were most valuable for me. Assist. Prof. Mehmet Tan personally supplied a lot of important material for the real kick off of this work. His ideas and support made it possible that in a short time I were able to build the frame of this work. Members of my thesis committee Professor Varol Akman and Assoc. Prof. Halit Oğuztüzün always gave valuable feedback for the progress of this work, and were not hesitant to warn me of the shortcomings or risks of my work. Assoc. Prof. Tolga Can also provided valuable feedback for the future of this research, which I very appreciate. I would also like to thank Assoc. Prof. Douglas J. Demetrick of UofC, our talks about my research not only inspired me but also helped me to understand the biological aspects of the work.

This work is also supported by TÜBİTAK-BİDEB PhD scholarship (2211) and TÜBİTAK-ARDEB Scientific and Technological Research Project Program (1001) (Project No: #110E179).

My family also provided invaluable support for this work. I would like to thank specially to my uncle Zeki Erdoğdu. He always make me feel loved and cared. I am also thankful for all the love and support by Sevilay Erdoğdu, Emin-Esra-Dicle-Deniz Erdoğdu, Hülya Oruç, Asiye Erdoğdu, Ahmet Erdoğdu, Sebahat-Serhun-Dilhun Ayaydın, Özlem-Metin-İpek Sarıbaş, Aynur-Kenan-Cihan-Berna Baykan. My father Ab-

dulkadir Erdoğdu was not able to see this work completed, however his memory was with me all the way to the end.

And there are a lot of people that were with me in these eight years. They defined me, they made me who I am, they are true owners of this work. It is not possible to write down why each of them is important to me and this work, because it will take more space than the work itself. So I'll just give names of some of them; Neslihan Arslan, Selma Süloğlu, Hande Çelikkanat, Ömer Nebil Yaveroğlu, Gökdeniz Karadağ, Sinan Kalkan, Burçin Sapaz, Turan Yüksel, Aykut Erdem, Erkut Erdem, Ruken Çakıcı, Oğuz Özün, Semra Doğandağ, Burcu Büyükkağnıcı, Ayça Aksu, Mehmet Ersan Topaloğlu, Servet Kızıltaş, Emre Hamit Kök, Sedar Gökbulut, Özgür Kaya, Gültekin Kurtar, Serdar Özcan, Fatih Balkan and İsmet Yalabık.

And finally very special thanks to a group of people who've taught me real meaning of friendship lately. Special thanks to the residents of the most joyful office ever, A-206 : Kerem Hadımlı, Can Hoşgör, Okan Akalın, Merve Aydınlılar and Elvan Gülen.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

A	An abbreviation
AA	Another abbreviation
LONGESTABBRV	Longest Abbreviation is used to balance the columns of list of abbreviations

CHAPTER 1

INTRODUCTION

This is the introductory chapter.

CHAPTER 2

NEXT ONE

This is another chapter. All these chapters can go into seperate .tex files and you can include them with `\input{chapter1.tex}`. This is a citation [1].

REFERENCES

[1] U. Erdoğdu. Sample reference article. *METU Thesis Journal*, 2(2):1–2, 2012.

APPENDIX A

PROOF OF SOME THEOREM

This is appendix text.

CURRICULUM VITAE

Utku Erdoğdu

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.S.	M.S. Institute	M.S. Year
B.S.	B.S. Institute	B.S. Year
High School	High School Name	High School Graduating Year

PROFESSIONAL EXPERIENCE

Year	**Place**	**Enrollment**
Duration 1	Institute/Company 1	Role/Position/Experience 1
Duration 2	Institute/Company 2	Role/Position/Experience 2

PUBLICATIONS

International Conference Publications

Your publications goes here. Do not try to use Bibtex, since Bibtex builds a single bibliography database for the document.