University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Damian Skrzypiec
Student no.: 320335

Structure Learning Algorithms for Chain Graphs

Master's thesis
in MATHEMATICS

Supervisor:
John Noble, PhD.
Institute of Applied Mathematics and Mechanics

April 2017

Supervisor's statement

Hereby I confirm that the present thesis was prepared under my supervision and that it fulfils the requirements for the degree of Master of Mathematics.

Date Supervisor's signature

Author's statement

Hereby I declare that the present thesis was prepared by me and none of its contents was obtained by means that are against the law. The thesis has never before been a subject of any procedure of obtaining an academic degree. Moreover, I declare that the present version of the thesis is identical to the attached electronic version.

Date Author's signature

Abstract

In this place will be abstract of this project.

Keywords

graphical model, chain graph, structure learning

Thesis domain (Socrates-Erasmus subject area codes)

11.2 Statistics

Subject classification

62 Statistics
62C10 Bayesian Problems

Title of the thesis in Polish

Contents

Chapter 1

Introduction

Chapter 2

Preliminaries

2.1. Graph Theory Terminology

This section provides definitions of graph theory objects required for completeness of further sections. In this section, when is not mention different, V is default notation for set of graph's vertices and E is default notation for set of graph's edges.

Definition 2.1.1. ***(Undirected edge)***
For vertices $u, v \in V$ we say that there is an undirected edge between vertices u and v if $(u, v) \in E$ and $(v, u) \in E$. Undirected edge between u and v is marked as $u - v$.

Definition 2.1.2. ***(Directed edge)***
For vertices $u, v \in V$ we say that there is an directed edge from vertice u to vertice v if $(u, v) \in E$ and $(v, u) \in E$. Directed edge from u to v is marked as $u \to v$.

Definition 2.1.3. ***(Skeleton)***
Skeleton of graph $G = (V, E)$ is a graph $G' = (V', E')$ where $V = V'$ and the set of edges E' is obtained by replacing directed edges of set E by undirected edges.

Definition 2.1.4. ***(Route)***
A route in graph $G = (V, E)$ is a sequence of vertices $(v_0, \ldots, v_k)$, $k \geq 0$, such that

$$(v_{i-1}, v_i) \in E \quad or \quad (v_i, v_{i-1}) \in E$$

for $i = 1, \ldots, k$. The vertices v_0 and v_k are called terminals. A route is called descending if $(v_{i-1}, v_i) \in E$ for $i = 1, \ldots, k$. Descending route from u to v is marked as $u \mapsto v$.

Definition 2.1.5. ***(Path)***
A route $r = (v_0, v_1, \ldots, v_k)$ in graph $G = (V, E)$ is called a path if all vertices in r are distinct.

Definition 2.1.6. ***(Cycle)***
A route $r = (v_0, v_1, \ldots, v_k)$ in graph $G = (V, E)$ is called a pseudocycle if $v_0 = v_k$ and a cycles if further route is a path and $k \geq 3$.

A graph with only directed edges is called an *undirected graph*. A graph without directed cycles and with only directed edges is called a *directed acyclic graph* (DAG).

Definition 2.1.7. ***(Chain graph)***
A graph $G = (V, E)$ is called a chain graph if it does not have directed (pseudo) cycles.

Definition 2.1.8. ***(Section)***
A subroute $\sigma = (v_i, \dots, v_j)$ of route $\rho = (v_0, \dots, v_k)$ in graph G is called section if σ is the maximal undirected subroute of route ρ. That means $v_i - \dots - v_j$ for $0 \leq i \leq j \leq k$. Vertices v_i and v_j are called terminals of section σ. Further vertex v_i is called a head-terminal if $i > 0$ and $v_{i-1} \to v_i$ in graph G. Analogically vertex v_j is called head- terminal if $j < k$ and $v_j \leftarrow v_{j+1}$ in graph G.

A section with two head-terminals is called *head-to-head* section. Otherwise the section is called *non head-to-head*. For a given set of vertices $S \subset V$ in graph G and section $\sigma = (v_i, \dots, v_j)$ we say that section is hit by S if $\{v_i, \dots, v_j\} \cap S \neq \emptyset$. Otherwise we say that section σ is outside set S.

Definition 2.1.9. ***(Intervention)***
A route ρ in graph $G = (V, E)$ is blocked by a subset $S \subset V$ of vertices if and only if there exists a section σ of route ρ such that one of the following conditions is satisfied.

1. *Section σ is head-to-head with respect to ρ and σ is outside of S.*
2. *Section σ is non head-to-head with respect to ρ and σ is hit by S.*

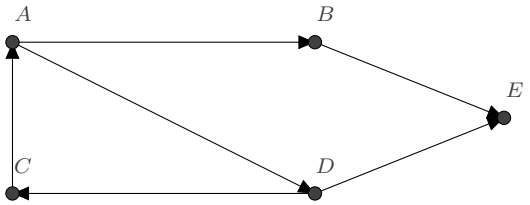

Figure 2.1: Example Graph

In graph 2.1 example of descending route is (A, B, E) and example of nondescending route is (A, D, E, B). Moreover graph 2.1 contains a cycle (A, D, C, A).

TODO: Make a comment about importance of the following definition.

Definition 2.1.10. ***(C-separation)***
Let $G = (V, E)$ be a chain graph. Let A, B, S be three disjoint subsets of the vertex set V, such that A and B are nonempty. We say that A and B are c-separated by S on G if every route within one of its terminals in A and the other in B is blocked by S. We call S a c-separator for A and B and mark as $\langle A, B \rangle_G^{sep}$.

2.2. Graphical Model Terminology

Graphical model therminologies

Definition 2.2.1. *Markov Blanket Markov Blanket is ...*

Chapter 3

Structural Learning of Chain Graphs

3.1. Algorithm

Chapter 4

Undirected Graphical Model Selection

4.1. Algorithm

Bibliography

[1] R. D. Nowak and D. Vats, *A Junction Tree Framework for Undirected Graphical Model Selection*, Journal of Machine Learning Research 15 (2014) 147-191

[2] Z. Ma, X. Xie and Z. Geng, *Structural Learning Of Chain Graphs via Decomposition*, Journal of Machine Learning Research 9 (2008) 2847-2880