

ベイズアンネットワークによる設問間の因果関係の推定

浅井 孝太 大枝 真一

木更津工業高等専門学校 情報工学科

1. まえがき

近年，教育の現場では学習者の成績や試験結果など大量のデータが蓄積されており，これらのデータに対してデータマイニングの技術を活用しようという試みが活発に行われている[1,2].

教育の現場で行われる試験は学習者が習得したスキルを測るために行われるが，試験結果から試験問題の設問間の因果関係を知ることにより，学習者の習得したスキルの推定を行うことや設問の適切な出題順序を提案することができると考えられる.

そこで本研究では，観測されたデータから探索的に因果モデルを構築できるベイズアンネットワーク[3,4]を用いて，試験問題の設問間の因果関係を推定することを目的とする.

2. ベイズアンネットワーク

ベイズアンネットワークとは，過去の事象が現在の事象の確率に影響を与えるというベイズ推定に基づく確率的なグラフィカルモデルである．観測されたデータから事象間の関係性を探り，それをモデル化し，そこから未知な情報に関して推論するために利用される．図 1 にベイズアンネットワークにおけるグラフ構造の例を示す．このようなグラフを非循環有向グラフ(Directed acyclic graph: DAG)と呼ぶ．このグラフにおいて枝が存在しないノードどうしは，条件付き独立の関係を表している．

さらに，ベイズアンネットワークでは確率変数を表す節点 X に，条件付き確率表(Conditional probability table: CPT)が付けられており，ベイズアンネットワークは DAG で示されるネットワーク構造と CPT によって構成される．

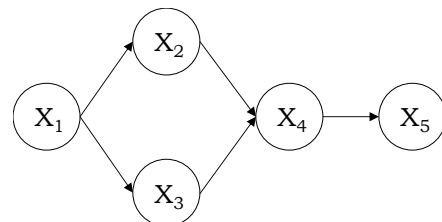


図 1: ベイズアンネットワークにおけるグラフ構造の例

3. 計算機実験

実践の応用場面においては，観測したり測定したりして得られた多変量データからベイズアンネットワークモデルを構築する必要がある．

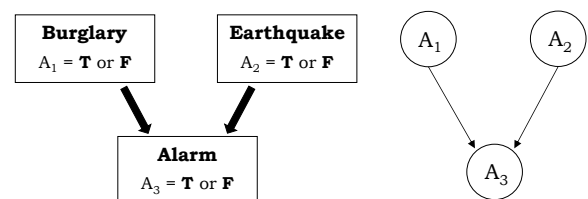
本実験では統計解析環境 R を用いて実際に多変量データ行列からモデルを構築する．

3.1 Burglary-Alarm problem

ベイズアンネットワークの説明でよく用いられる Burglary-Alarm problem を用いて分析を行った．これは次のような問題である．

図 2(a) の 3 つの事象はそれぞれに「泥棒が入る(Burglary)」 「地震が発生する(Earthquake)」 「アラームが作動する(Alarm)」という命題が与えられており，それぞれの命題に対応して真(True; T)と偽(False; F)という 2 つの状態をとる事象をそれぞれ A_1 , A_2 , A_3 により表すものとする．

「泥棒が入る」または「地震が発生する」確率がそれぞれ表 1, 表 2 で与えられている．そして「泥棒が入った」かまたは「地震が発生した」ときに「アラームが動作する」条件付き確率が表 3 で与えられている．



(a) 事象に割り当てられた (b) 事象をノードとして命題を表示した表現 単純化したグラフ表現

図 2: Burglary-Alarm Problem

The inference of question items relationship with Bayesian network

Kouta ASAI, Shinichi Oeda,
Department of Information and computer Engineering,
Kisarazu National College of Technology

表 1: $P(A_1)$ の確率表

$A_1 = T$	0.1
$A_1 = F$	0.9

表 2: $P(A_2)$ の確率表

$A_2 = T$	0.1
$A_2 = F$	0.9

表 3: $P(A_3|A_1, A_2)$ の条件付き確率表

	$A_1 = T$		$A_1 = F$	
	$A_2 = T$	$A_2 = F$	$A_2 = T$	$A_2 = F$
$A_3 = T$	0.607	0.368	0.135	0.001
$A_3 = F$	0.393	0.632	0.865	0.999

ここで、表 1～3 の確率になるようにデータ行列を作成し、このデータからベイジアンネットワークモデルの構築を行った。

図 3 に生成されたモデルを示す。これは図 2(b) のモデルと一致しており、正しくモデルの構築ができていることがわかる。

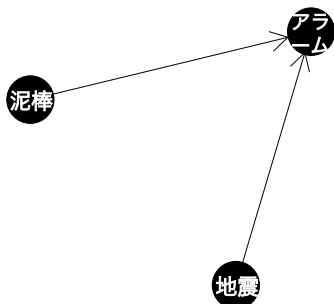


図 3: Burglary-Alarm Problem のデータから構築されたベイジアンネットワークモデル

3.2 低出生体重児データ

低出生体重児に関する医療データを用い、低出生体重児が生まれてくる原因についてベイジアンネットワークで分析を行った[5]。図 4 に生成されたモデルを示す。

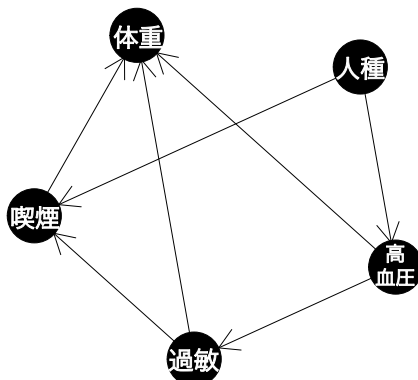


図 4: 低出生体重児のデータから構築されたベイジアンネットワークモデル

図 4 から低出生体重児が生まれてくる原因として挙げられる要素は、「喫煙」「過敏」「高血圧」の可能性が高いと推測することができる。

3.3 分数の引き算問題のデータ

実際の教育データとして、20 問の分数の引き算の問題を 536 人に解答してもらった結果のデータ[6]を用いてネットワーク構造がどのようなのかを実験した。図 5 に生成されたベイジアンネットワークモデルを示す。

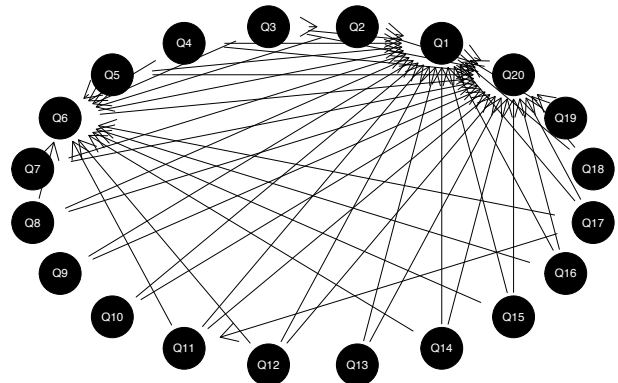


図 5: 分数の引き算問題のデータから構築されたベイジアンネットワークモデル

図 5 を見ると、Q1, Q6, Q20 に有向辺が集中していることがわかる。Q6 は単純な引き算のみであり、ほかの問から有向辺が連結されることが説明できる。また、Q1, Q20 についても他の問に必要な基礎的なスキルを要しており、他の問から有向辺が連結されているのではないかと考えることができる。

4. まとめ

本研究では教育現場で用いられる試験における設問間の因果関係を推定することを目的とした。手法としてベイジアンネットワークを用いることにより、事象間の因果関係を推定することができた。

参考文献

- [1] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions", *Journal of Educational Data Mining*, Vol.1, Issue 1, pp.3-17, 2009.
- [2] Toon Calders and Mykola Pechenizkiy, "Introduction to The Special Section on Educational Data Mining", *SIGKDD*, Vol.13, Issue 2, pp.3-5, 2011.
- [3] 田中和之, "ベイジアンネットワークの統計的推論の数理", コロナ社, 2010.
- [4] 植野真巨, "ベイジアンネットワーク", コロナ社, 2013.
- [5] 豊田秀樹, "データマイニング入門 -Rで学ぶ最新データ解析-", 東京書籍, pp.228-240, 2012.
- [6] Tatsuoka, K. "Analysis of errors in fraction addition and subtraction problems", Final Report for NIE-G-81-0002, University of Illinois, Urbana-Champaign, 1984.