HOKKAIDO UNIVERSITY

Title	Relevance-dependent Biclustering Models for Relational Data Analysis [an abstract of dissertation and a summary of dissertation review]
Author(s)	大濵, 郁
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第13508号
Issue Date	2019-03-25
Doc URL	https://hdl.handle.net/2115/74052
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/4.0/
Type	doctoral thesis
File Information	Iku_Ohama_review.pdf, 審査の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　大濵　郁

審査担当者　主　査　教　授　有村 博紀
副　査　特任教授　原口 誠
副　査　准教授　喜田 拓也

学位論文題名

Relevance-dependent Biclustering Models for Relational Data Analysis
（関係データ分析のための関連度依存型共クラスタリングモデルの研究）

インターネットやセンサーネットワークの発展により、購買履歴や SNS 上の交友関係等の膨大なデータに基づく人間行動の解析が注目されている。このような人やモノの結びつき (リンク) を表すデータは、ネットワーク構造や二値行列で表すことができ、関係データ (relational data) と呼ばれており、そこからの潜在構造発見が重要性を増している。ここでは、関係データからの潜在構造発見問題の一つである共クラスタリング問題を考察する。共クラスタリング (biclustering または co-clustering) は、関係データの行と列を同時クラスタリングすることにより、潜在構造としてデータの要約であるブロック構造を得る問題であり、2000 年代に入って、情報理論や、スペクトル的グラフ理論、統計推論等のさまざまな枠組みにおいて盛んに研究されている。

本研究では、統計モデルに基づく関係データからの潜在構造発見を考察する。他の枠組みと比べて、統計モデルの枠組みの長所として、(i) 潜在変数を用いた柔軟なモデル設計が可能であり、 (ii) ノイズや外れ値を扱いやすい、(iii) ブロック数の自動推定が可能、(iv) 種々の推論手法が利用可能等の利点をもつ。その一方で、(v) モデルが高次元になりやすく、(v) 推論に大きな計算量を要する等の短所をもつ。そのため、実用に際しては、効率良いモデルと推論手法の開発が重要となってくる。

本研究では、統計モデルに基づく共クラスタリング手法の中で、Kemp らが 2006 年に提案した無限関係モデル (Infinite Relational Model、IRM) に注目する。IRM を含む従来のモデルでは、消費者と商品といった行と列の個体は「消費嗜好」や「商品カテゴリ」といったブロック構造への所属に関する潜在変数をもつが、一方で消費者の「購入熱意」や「予算」、商品の「露出度」といった個体とブロック構造の結合度合いは直接には扱えない。そこで、本研究では、結合度合いを表す「関連度」(relevance) という潜在変数を導入して IRM を拡張した関連度依存型共クラスタリングを提案し、その種々の拡張と効率良い推論手法の開発を行なっている。

まず第 3 章では、論理関数を用いて IRM モデルに関連度依存性を導入した関連度依存 IRM (Relevance-dependent IRM, RDIRM) を提案する。行と列の関連度に関する相互作用を表現する論理関数を導入し、ブロックへのデータの所属性が関連度に依存するように IRM を拡張する。さらにクラスタ数の自動推定と、周辺化ギブス乱択による効率的な学習方法を提案している。

次に第 4 章では、RDIRM における相互作用を表す論理関数の一般化を考察する。関連度と所属性の相互作用関数を連続緩和して拡張することで、データからの相互作用関数の自動獲得と、任意個の分布の混合とを可能にした多層 IRM(Multi-Layered IRM、MLIRM) を提案している。これに

より、少数の個体が非常に多数のリンクを張る特異な場合も効率良く扱える。MLIRM に対しても、クラスタ数の自動推定と効率的なモデル推論を実現している。

最後に第 5 章では、先に導入した 3 章と 4 章のモデルがもつ高次元性と解釈困難性の問題に取り組んでいる。これらを解決するために、非負値をとる各個体の関連度とブロックの典型的リンク強度の積を確率に変換し、関連度依存型の二値行列を直接生成する手法を新たに導入することで、相互作用関数と多層の分布選択を不要にした関連度依存型無限双クラスタリング手法 (RD Infinite Biclustering、R-IB) を提案している。実験では、従来の IRM および 3 章と 4 章のモデルと比較して、推論の高速性を実証している。

本論文の成果は次のようにまとめられる。

(1) 関係データからの共クラスタリング問題において、従来の無限関係モデルに関連度の概念を導入した関連度依存無限関係モデル RDIRM を提案し、その効率良い推論アルゴリズムを与えた。

(2) RDIRM を拡張して、MLIRM と R-IB という二つの新しい関連度依存無限関係モデルとそれらの推論アルゴリズムを開発し、関連度の相互作用関数の自動獲得や、複雑な関係モデルの次元削減、効率良い推論のための新しい手法を開発した。

(3) 提案したモデルとアルゴリズムに対して、人工データおよび実データを用いた計算機実験によって、それらの有効性を実証的に示した。

これを要するに、著者は、情報科学における関係データの共クラスタリング問題において、関連度に依存した無限関係モデルを定式化し、これに基づく種々の拡張と効率良いアルゴリズムを提案し、その有効性を実証したものであり、大規模データ解析に関する情報科学に貢献するところ大なるものがある。よって著者は北海道大学博士 (情報科学) の学位を授与される資格あるものと認める。