

1 はじめに

数学的に定義された概念に基づき、データ分析を行う手法として形式概念分析 [1] がある。形式概念分析では、概念の構造を表す概念束を可視化することで、構造の理解を助けることができる。一方でデータの増大に従い、概念束のサイズが急速に大きくなるため、構造の理解が難しくなる。これまでに概念束の簡素化を行う手法が提案されてきたが、それぞれ問題点がある。本研究の目的は、従来手法の問題点を改善した概念束の簡素化手法を提案することである。

2 形式概念分析

形式概念分析では対象の集合 G と属性の集合 M 、 G と M の間の二項関係 $I \subseteq G \times M$ を扱う。三つ組 $\mathbb{K} = (G, M, I)$ を形式文脈という。また gIm である時、「対象 g は属性 m を持つ」という。ここで二つの集合 $X \subseteq G$ と $Y \subseteq M$ について、 X のすべての対象が共通して持つ属性の集合が Y であり、 Y のすべての属性を持つ対象の集合が X である時、組 (X, Y) を形式概念という。また、対象集合の包含関係により、形式概念の順序が定義される。これを概念束という。

また、形式文脈 $\mathbb{K} = (G, M, I)$ と属性の集合 $P, C \subseteq M$ について、 P の属性をすべて持つ任意の対象が C の属性をすべて持つ時、これを $P \implies C$ と書き、属性の含意関係という。

3 従来手法

概念束の簡素化を行う代表的な手法として、氷山概念束、特異値分解を用いた手法、安定度を用いた手法がある。氷山概念束では、上位の概念を抽出することで簡素化を行うが、これは概念束全体を表していない。特異値分解を用いた手法では、特異値分解による次元縮小を行うが、どのような性質が満たされるかは明らかでない。また、安定度を用いた手法では、安定な形式概念を抽出することで簡素化を行う。しかし、この手法では出力が必ずしも束とはならない。

4 提案手法

本研究では、類似した形式概念間に基づく属性推定を利用して、概念束の簡素化を行う手法を提案する。まず、類似した二つの形式概念の属性集合間の関係を、近似的含意関係として定義する。また、これらの対象集合のサイズの比を近似的含意関係のサポートとする。提案手法では、サポートがしきい値 τ 以上である近似的含意関係と属性の含意関係から属性推定を行い、形式文脈を更新することで概念束の簡素化を行う。

5 実験

学生のエゴセントリックネットワークの構造に関する形式文脈に対して、提案手法と特異値分解を用いた手法を適用した。対象と属性の対応関係が明確に得られないため、他の手法は用いない。提案手法はしきい値 τ を、特異値分解を用いた手法は次元縮小後の次元 k を変化させて適用し、適用後の概念束のサイズと適用後も成立する属性の含意関係の割合を調べた。図 1 からどちらの手法でも概念束のサイズは小さくなるのがわかる。また、適用後に成立する属性の含意関係の割合は、提案手法の方が高い値を保っている。

また、ノイズを加えた形式文脈に対して、提案手法と特異値分解を用いた手法を適用し、形式文脈と概念束の変化を調べた。どちらの手法でも、ノイズを減少する効果は見られなかった。逆に提案手法では、ノイズを増幅するような結果が見られた。

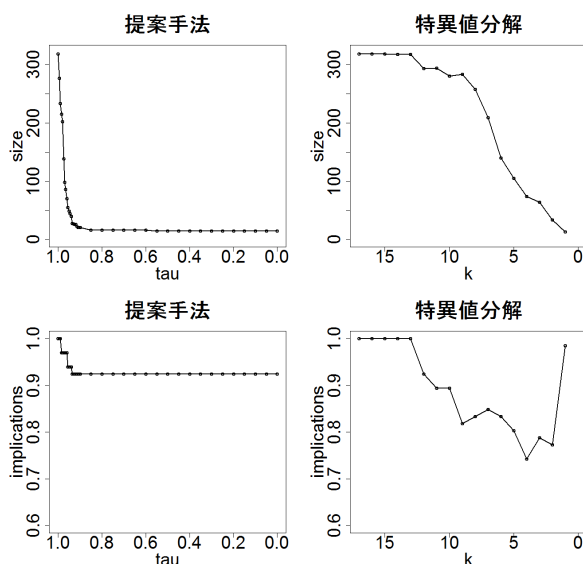


図 1: 概念束のサイズ (上) と成立する含意関係 (下)

6 まとめ

本研究では、類似した形式概念間に基づく属性推定を行うことで、概念束の簡素化を行う手法を提案した。また、実験により提案手法の性質を調べた。提案手法では形式文脈全体を簡素化しているため、出力は元の概念束全体を表していると考えられる。また、出力は必ず束となる。さらに、属性の含意関係の多くが簡素化後も成立することを、実験で確認した。

一方で、提案手法はノイズの影響を強く受ける。また、計算コストに関する考察が必要である。

参考文献

- [1] Ganter, B. and Wille, R.: Formal Concept Analysis: Mathematical Foundations, Springer, 1998.
- [2] 石樽 隼人, 武藤 敦子, 松井 藤五郎, 犬塚 信博: 形式概念間に基づく属性推定を利用した概念束の簡素化, 第 77 回情報学大会, 2R-05, 2015 (発表予定)