2K-08

形式概念束の簡約化のための概念間の距離に関する検討

寺町太貴 † 石榑隼人 ‡ 武藤敦子 ‡ 森山甲一 ‡ 犬塚信博 ‡

† 名古屋工業大学工学部情報工学科 ‡ 名古屋工業大学大学院工学研究科情報工学専攻

1 はじめに

形式概念分析とは形式概念と呼ばれる概念データを思考単位とするデータ解析手法である。そこでは概念束と呼ばれる概念構造が得られるが、データの増大に従って概念束が複雑化するため、簡約化が必要である。簡約化では簡約化の度合いやノイズ耐性が検討されてきた。本研究は簡約化に対する概念の分布の影響を見るため簡約化前後の概念間の距離の分布を調査する。

2 形式概念とその簡約化

形式概念分析 [1] は、対象集合 G、属性集合 M、対象と属性の二項関係 $I \subseteq G \times M$ を扱う。$gIm \Leftrightarrow (g,m) \in I$ である時、対象 g は属性 m を持つという。三つ組 $\mathbb{K} = (G, M, I)$ を形式文脈という。$\mathbb{K}$ において任意の $X \subseteq G$ と $Y \subseteq M$ に次の写像を定義する。

$$X \mapsto X^I = \{m \in M \mid gIm \text{ for all } g \in X\}$$
$$Y \mapsto Y^I = \{g \in G \mid gIm \text{ for all } m \in M\}$$

$A^I = B$ かつ $B^I = A$ である $A \subseteq G$ と $B \subseteq M$ の組 (A, B) を $\mathbb{K}$ の形式概念といい、A を外延、B を内包という。また次の順序によって定まる $\mathbb{K}$ の形式概念の束を概念束と呼ぶ。

$$(A_1, B_1) \geq (A_2, B_2) \Leftrightarrow A_1 \supseteq A_2 \Leftrightarrow B_1 \subseteq B_2$$

隣接する 2 つの概念 $(A_1, B_1), (A_2, B_2)$ 間の距離 d_{12} を以下の数式 [2] で定義する。

$$d_{12} = 1 - \{\frac{1}{2}\frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} + \frac{1}{2}\frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}\}$$

本稿では概念束の簡約化として氷山概念束 [3]、特異値分解を用いた手法 [4]、安定度による手法 [5]、属性推定を用いた手法 [6] を取り上げ、簡約化による距離の分布の変化を調べる。

A study of distance between formal concepts for deduction of formal concepts lattice

Taiki Teramachi†, Hayato Ishigure‡, Atsuko Mutoh‡, Koichi Moriyama‡, and Nobuhiro Inuzuka‡

Dept. of Computer Science, Faculty of Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan †

Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan ‡

3 形式概念における距離の分布

簡約化はその目的に応じて必要な性質がいくつか考えられる。1 つの目的は適切な間隔による概念の再配置であり概念間の距離は均一が求められるかもしれない。他方で、簡約化はデータの分布を反映した概念束の概要を得ることを目的とするなら、簡約化は分布を維持したまま間引くことが必要かもしれない。本稿は、目的を定めず各手法の性質を観察する。

図 1 に距離分布を算出する R 言語のプログラムを示す。各関数の入力はともに形式概念の集合である。distance は概念と対象の帰属を行列で表した外延と、概念と属性の適否を表す内包を e と i とした後、行列 $e \cdot e^t$ を計算して概念間の外延の積集合の要素数を E に、同様に内包の積の要素数を I に得る。また、e 及び i の 0/1 を反転して和集合の要素数を計算し、これらから距離を得る。neighbor は積集合の要素数を得た後、$|A| = |A \cap B|$ であれば $A \subseteq B$ である事実を用い、包含関係を incl に得る。包含関係を行列として自乗すると、隣接した包含関係のみがなくなることを利用し、隣接関係を出力する。nbr_dist は隣接関係と距離を合わせることで、隣接した概念間の距離を出力する。

4 実験と結果

実験に友人関係データ [7] と人工データ [6] を用いた。

2 つのデータとそれに 4 つの簡約化を適用した結果の概念束の隣接概念への最大距離の分布を図 2,3,4 に示す。概して距離について期待する良い性質をどの方法も明確には示さなかった。その中で特異値分解と属性推定は比較的分布を維持する傾向があり、安定度は分布を均一方向に修正する傾向が見られた。

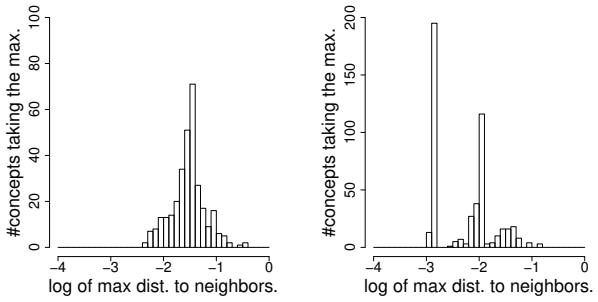
5 まとめと今後の課題

概念束簡約化に求める性質として隣接概念への距離分布に関する性質を提案し、4 つの簡約化についてこの性質を調べた。距離のバリエーションや求める距離分布のさらなる検討、さらに距離について必要な性質を達成する簡約化については今後の課題である。

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.

```
distance <- function(concepts){
  e<- matrix(unlist(concepts$extents),
   nrow=length(concepts$extents),byrow=T)
  i<- matrix(unlist(concepts$intents),
   nrow=length(concepts$intents),byrow=T)
  E<- e%*%t(e)  #外延の積集合
  I<- i%*%t(i)  #内包の積集合
  RE<- ncol(e)-((!e)%*%(t(!e))) #和集合
  RI<- ncol(i)-((!i)%*%(t(!i))) #和集合
  (1-(E/RE))/2 + (1-(I/RI))/2 #距離を出力
}
neighbor <-function(concepts){
  e<- matrix(unlist(concepts$extents),
  nrow=length(concepts$extents),byrow=T)
  E<- e%*%t(e)          # 外延の積集合
  incl <- E == diag(E)
  diag(incl)<- FALSE    #対角成分を FALSE に
  incl2 <- incl %*% incl # 包含関係の経由数
  (incls == 0) & incl2
}  #経由なしで包含ありを出力
nbr_dist<-function(concepts){ #main 関数
  H<- neighbor(concepts)
  PH<- H|t(H)       # 隣接関係対称化
  R<- distance(concepts)
  ifelse((PH==TRUE),R,NA) #隣接する距離を出力
}
```

図 1: 使用したプログラム



(a) 友人関係データ　(b) 人工データ

図 2: 実験に用いた概念束の隣接概念への最大距離

参考文献

[1] Wille, R.: Restructuring lattice theory : An approach based on hierarchies of concepts, in Rival, I., ed., Ordered Sets, D. Reidel Publishing, pp. 445–470, 1982.

[2] Blachon, S., *et al.*: Clustering Formal Concepts to Discover Biologically Relevant Knowledge from Gene Expression Data, In Silico Biology **7**, 4–5, pp. 467–483, 2007.

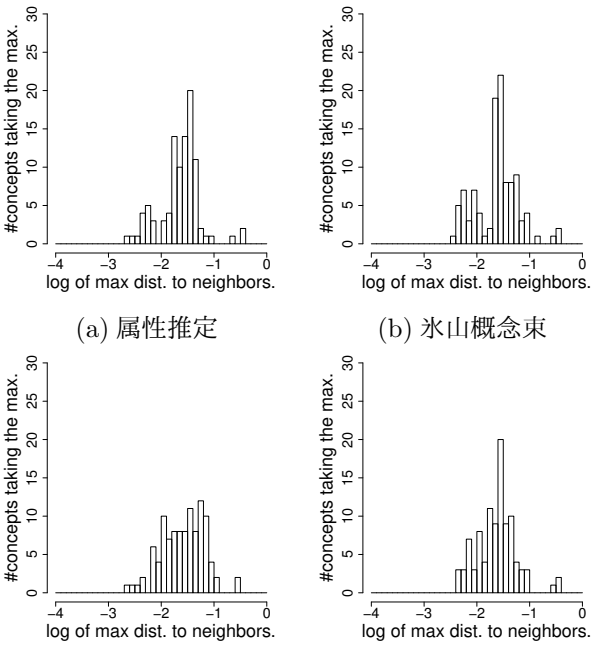
[3] Stumme, G.,*et al.*: Conceptual clustering with iceberg concept lattices, FGML 01, 2001.

[4] Gajdoš, P.,*et al.*: Concept lattice generation by singular value decomposition, CLA2014, pp.102–110, 2014.

[5] Kuznetsov, S. O.: On stability of a formal concept, Ann. Math. Artif. Intell. **49**.1–4, 2007.

[6] Ishigure, H., *et al.*: Concept Lattice Reduction Using Attribute Inference, GCCE2015, pp. 108–111, 2015.

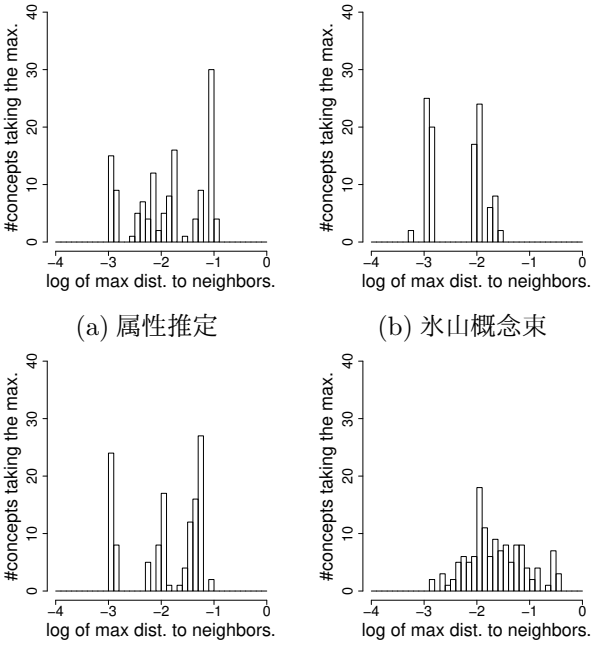
[7] 林宏紀，他：エゴセントリックネットワークと形式概念分析を利用した社会ネットワーク分析，人工知能学会論文誌 **29**, 1, pp. 177–181, 2014.



(a) 属性推定　(b) 氷山概念束

(c) 特異値分解　(d) 安定度

図 3: 友人関係データ簡約後の隣接概念への最大距離



(a) 属性推定　(b) 氷山概念束

(c) 特異値分解　(d) 安定度

図 4: 人工データ簡約後の隣接概念への最大距離

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.