

形式概念分析を用いた友人関係パターンと授業出欠の分析

夏目 稔[†] 石博 隼人[†] 武藤 敦子[†] 森山 甲一[†] 犬塚 信博[†]

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

近年、社会ネットワーク分析においてネットワークの詳細を分析するにあたって、1つの観点であるエゴセントリックネットワークがある。また、データから形式的に概念を抽出する形式概念分析もデータマイニング等で注目される。Kuznetsovら[1]は、グラフ集合を正事例と負事例に分け、2つの事例それぞれにパターン構造を用いたグラフを扱う形式概念分析に関連する学習モデルを適用する手法を提案した。

教育支援において、大学は学生に対してのセグメンテーションの充実を図っている。そのためには、学生の性格や学習意欲などの定性情報と筆記試験の成績や出席率などの定量情報の両方が必要である。

しかし、学生の定性情報はアンケートを行うなど収集するのに手間と時間のコストがかかる。また、収集したい定性情報が心理的に答えにくいものであれば正しく答えてくれる可能性が低くなるなどの問題を含んでいる。そこで、情報の収集が容易で客観性が高い定量情報から定性情報を導くことができれば、低コストで客観性がある定性情報を得ることができる。

夏目ら[2]は、教育上の問題をパターン分析で扱っている。一人一人の人間を形式文脈における対象、その人のエゴセントリックネットワークに含まれる部分ネットワークを属性とした形式概念分析を行っている。これと別に打刻傾向を表す打刻属性と形式概念を照合することで、打刻属性を探るために有用なネットワーク構造を探っている。

本研究では、夏目らの打刻属性に関して、打刻データから得た各エゴセントリックネットワークを正事例と負事例に分け、Kuznetsovらの手法を適用する。適用した結果から、学生の友人ネットワークの構造と打刻傾向との関係性を示し、定量的である打刻データから学生の性格や学習意欲を読み取ることを目指す。

2 社会ネットワーク分析とエゴセントリックネットワーク

社会学における複雑ネットワーク分析とは、頂点と繋がり(tie)という観点から社会的な隣接性について考察することである。ここでの頂点とは行為者のことを指し、繋がりとは行為者同士の結びつきを表す。これらをネットワークの構造から法則性や特徴を発見することがこの分野の研究である。そこで、社会ネットワーク分析において、各個人のネットワークの構造からわかる特徴の詳細な分析に注目した。以下で分析するにあたって、1つの観点であるエゴセントリックネットワークについて説明する。

各行為者を頂点とし、行為者間の影響関係を辺で表した無向グラフを考える。社会ネットワーク分析では、行

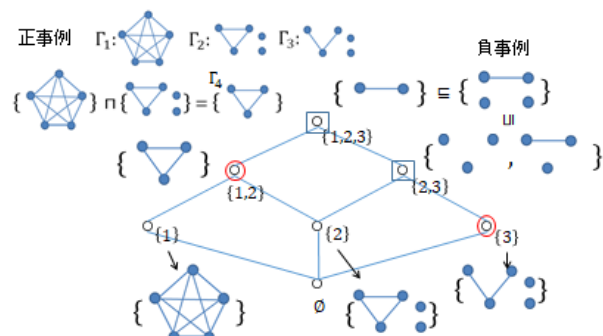


図1: 正事例のパターン概念束の例

為者の1人1人に注目するときに、各行為者をエゴと呼ぶ。各エゴを中心としたローカルネットワーク、つまりエゴと直接つながる行為者(オルター)の集合から誘導される部分グラフをエゴセントリックネットワークという。エゴセントリックネットワークはネットワークの構造からわかる特徴の詳細な分析に利用される。

3 パターン構造

データが活用される場合、大抵は数値や時間幅、グラフなどを含んでおり、形式文脈に概念的な指標を取り入れることが必要となる。そのためデータを変換する代わりに、表現や計算上の違いから、数値や時間幅、グラフなどを含むデータを直接扱う手法が考えられる。そこで、Ganter & Kuznetsovはパターン構造を提唱した[3]。

パターン構造はデータの値から半束を導く類似演算子を用いた対象の記述から構成される。対象の集合を G 、対象の記述の集合を D 、 (D, \cap) を D の交わり半束、対象からその記述への写像を $\delta: G \rightarrow D$ とする。つまり、 \cap は類似の演算子を表す。これらの $(G, (D, \cap), \delta)$ の3つ組をパターン構造という。対象のある集合 $A \subseteq G$ と、対象の記述 $d \in D$ (以後パターン)に対して、 A のすべての対象の共通する記述が d であり、 d をすべて記述とする対象の集合が A であるとき、組 (A, d) をパターン概念という。

二つのパターン概念 $(A, d), (A', d')$ に対し、 d が d' の誘導部分グラフであるか($d \sqsubseteq d'$ と表す)によって順序を定めると、パターン概念は束となる。

例として、グラフ $\Gamma_1, \Gamma_2, \Gamma_3$ を正事例とした対象の集合 $G = \{1, 2, 3\}$ としたときのパターン概念束を図1に示した。 $A = \{1\}$ において、 A のすべての対象の共通する記述は $d = \{\Gamma_1\}$ であり、 d をすべて記述とする対象の集合は $\{1\}$ である。したがって、組 $(\{1\}, \{\Gamma_1\})$ はパターン概念である。他に、 $A = \{1, 2\}$ において、 A のすべての集合の共通する記述の集合は Γ_1 と Γ_2 の共通する極大の誘導部分グラフ $d = \{\Gamma_4\}$ であり、 d をすべて記述とする対象の集合は $\{1, 2\}$ である。したがって、組 (A, d) はパターン概念である。

正事例(負事例)から出したパターン概念 (A, d) の

Analysis of friendship patterns and class attendance records using formal concept analysis

[†]Minoru Natsume

[†]Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

表 1: 生成した打刻属性 (平均より少ない、平均より多い属性の計 12 個)

出席回数	最初から最後まで授業をうけた数
遅刻回数	遅刻した授業の数
平均入室時間差	開始打刻時刻と授業開始時刻の差の平均時間差
平均退室時間差	終了打刻時刻と授業終了時刻の差の平均時間差
前打刻回数	ある教室において前にある IC カードリーダーに打刻した回数
後ろ打刻回数	ある教室において後ろにある IC カードリーダーに打刻した回数

内、負事例 (正事例) のパターン d_0 の誘導部分グラフとなるパターン $d \subseteq d_0$ をもつパターン概念を除いて、極小となるパターンをもつパターン概念が正事例 (負事例) の特徴を表すパターン概念といえる。

4 打刻データを用いた分析方法の提案

夏目らは、ネットワーク構造を形式概念分析を用いて抽出した。そして、生成した打刻属性との照合を検討した。本研究では、夏目らの打刻属性に関して、学生のエゴセントリックネットワークを正事例と負事例に分け、Kuznetsov らの 2 つの事例それぞれにパターン構造を用いたグラフを扱う形式概念分析に関連する学習モデルを適用する。夏目らの打刻属性を表 1 に示す。例えば、出席回数が多い打刻属性をもつ学生のエゴセントリックネットワークを正事例とした場合、出席回数が少ない打刻属性をもつ学生のエゴセントリックネットワークを負事例とする。

学生は各教室に設置されている IC カードリーダーに授業の開始打刻時間内と終了打刻時間内に学生証を提示することで出席となる。打刻データには学生証である IC カードから学生 ID と打刻した時間、それらを読み取り記録する IC カードリーダーの情報が記録されている。本研究では対象の学生が履修している科目で、休講でない授業の打刻を利用している。また、開始 (終了) 打刻時間内に 2 回以上打刻がある場合、その中で最初の打刻のみを利用している。

「出席回数」、「遅刻回数」、「前打刻回数」、「後ろ打刻回数」は、回数が平均より多い学生を回数が多いという属性、回数が平均より少ない学生を回数が少ないという属性をもつとする。「平均入室時間差」、「平均退室時間差」には、平均時間差が平均より大きい学生を平均時間差が大きいという属性、平均時間差が平均より小さい学生を平均時間差が小さいという属性をもつとする。すなわち、計 12 個の打刻属性を利用する。

5 実験

平成 24 年度入学の名古屋工業大学の 1 クラス合計 57 名による 6 月の打刻データを利用した。打刻データから、下村ら [4] の手法を用いて友人ネットワークを生成した。生成した友人ネットワークのノード (学生) 数は 57、友人関係数は 282 であった。打刻属性は表 1 で挙げたものを利用する。パターン構造を用いたグラフを扱う形式概念分析に関連する学習モデルは、各打刻属性においてそれぞれ 20 個の事例に適用する。頂点数 1 のエゴセントリックネットワークはエゴセントリックネット

表 2: 12 個の打刻属性ごとの抽出された極小となるパターンをもつパターン概念の個数

出席回数		遅刻回数		平均入室時間差	
多	少	多	少	大	小
4	3	5	3	4	5

平均退室時間差		前打刻回数		後ろ打刻回数	
大	少	多	少	多	少
4	4	3	5	2	5

ワークなしとする。

分析によって抽出した打刻属性ごとのパターン概念の個数を表 2 に記す。12 個すべての打刻属性においてパターン概念が抽出された。したがって、すべての打刻属性において、その打刻属性をもつ学生の友人ネットワークの構造的特徴を表すパターンを得た。

本研究で抽出したパターン概念のパターンの頂点数と辺数の平均は、12.1 個と 7.6 本となった。夏目らの手法と同様にネットワーク構造を抽出する林ら [5] の手法で抽出できるネットワーク構造は頂点数 4 までのネットワークであるため、本研究は夏目らの手法に比べ、十分なサイズのネットワークを扱える手法であるといえる。

6 まとめと今後の課題

本研究では、夏目らの打刻属性に関して、打刻データから得た各エゴセントリックネットワークを正事例と負事例に分け、2 つの事例それぞれにパターン構造を用いたグラフを扱う形式概念分析に関連する学習モデルを適用した。適用した結果、各打刻属性において学生の友人ネットワークの構造的特徴を表すパターンを得た。

夏目らはアンケートデータから生成した友人ネットワークを利用しているが、本研究では打刻データから生成した友人ネットワークを利用している。したがって、本研究では、定量的である打刻データのみから、学生の性格や学習意欲などの定性情報を導いているといえる。

今後の課題としては、計算量の削減が挙げられる。計算時間を考慮した結果、夏目らの手法と比べて対象数が少なくなった。アルゴリズムを改良し、計算量を削減する必要がある。また、各打刻属性をもつ学生のエゴセントリックネットワークの構造的特徴を表せているかを、夏目らの手法と比較する検定を実施する必要がある。

参考文献

- [1] S. O. Kuznetsov, "Machine Learning and Formal Concept Analysis" in LNCS 2961, 2004.
- [2] 夏目稔, 石樽隼人, 武藤敦子, 森山甲一, 犬塚信博 「友人ネットワークの構造属性と出退席打刻属性による形式概念」, 29 年度信学会東海卒研発表会
- [3] B. Ganter, S. O. Kuznetsov, "Pattern structures and their projections", in LNCS 2120, 2001.
- [4] 下村幸作, 中野智文, 犬塚信博 「学生の出欠時間を活用した学生の友人関係分析」, in SIG-DMSM, 2008.
- [5] 林宏紀, 伊東樹希, 西尾典晃, 武藤敦子, 犬塚信博 「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), 2013.