

# 実習型 e-learning システムを対象とした 形式概念分析による成績データ分析手法の検討

朝見義樹<sup>\*1</sup>, 本吉達郎<sup>\*1</sup>, 澤井圭<sup>\*1</sup>, 増田寛之<sup>\*1</sup>, 高木昇<sup>\*1</sup>

<sup>\*1</sup> 富山県立大学大学院

## Analysis Method for Grade Data of Practice-based e-learning System Using Formal Concept Analysis

Yoshiki Asami<sup>\*1</sup>, Tatsuo Motoyoshi<sup>\*1</sup>, Kei Sawai<sup>\*1</sup>, Hiroyuki Masuta<sup>\*1</sup>, Noboru Takagi<sup>\*1</sup>

<sup>\*1</sup> Toyama Prefectural University

本研究では実習型 e-learning システムである「ナレロー」の成績データに対して形式概念分析を適用する手法について検討する。具体的には、Office 系ソフトウェアの学習システムである「ナレロー」の成績データを対象として、形式概念分析の含意論理によって捉えた正解問題/不正解問題の含意関係と統計的な分析手法から得た全体的な傾向を比較し、形式概念分析から得られる知見の統計的な分析へのトリガとしての利用可能性について考察する。

キーワード: e-learning システム, 成績データ, 形式概念分析, データ分析

### 1. はじめに

近年、コンピュータの普及に伴い、多くの教育機関において e-learning が導入されている。e-learning では、Learning Management System (LMS)を用いて、学習所要時間、訪問時間、学習者の正解した問題の記録などの学習履歴データを取得することができる。蓄積された学習履歴データを分析することで個々の能力に応じた出題の実現ができると期待されている<sup>(1)</sup>。個々の能力に応じた問題の出題を実現するには、学習履歴データからユーザが正解した問題の記録などを分析し集団における解答の傾向などの特徴を捉えることが重要となる。本稿では、ユーザが正解した問題の記録のことを「成績データ」と呼ぶ。成績データから解答状況の特徴を捉える方法として相関分析を用いて問題同士の正答率や誤答率の相関を見る方法やカテゴリデータの検定を用いてある問題の正誤が他の問題の正誤と関連を持つか検証する方法がある。しかし、学習コンテンツが多い教材の場合、

問題同士の関連を上記の方法で網羅的に分析するのは困難であり非常に労力がかかる。このような分析を行う際は、事前にユーザ全体の解答の傾向などを把握することが重要である。本研究では実習型 e-learning システムである「ナレロー<sup>(2)</sup>」の成績データを対象に属性間の包含関係を網羅的に抽出する形式概念分析を適用する手法について検討する。「ナレロー」の成績データを対象に、形式概念分析の含意論理によって捉えた正解問題/不正解問題の含意関係と統計的な分析手法から得た全体的な傾向を比較し、形式概念分析から得られる知見が統計的な分析への利用可能性について考察する。

### 2. 実習型 e-learning システム「ナレロー」

「ナレロー<sup>(2)</sup>」は学校教育内での Microsoft Office の指導管理を支援するクラウド型成績管理システムを備えており、課題の進行状況や習熟度を把握することが可能な e-learning システムである。

ユーザは、ナレローにより提示された問題を Office 系ソフトウェアを実際に操作し学習する。問題内容は、「上書き保存」などの簡単なレベルから「宛名ラベルの作成」などのある程度の経験者にしか操作方法がわからないレベルまで幅広く設定されている。ユーザが問題に正解したら「1」を入力し、不正解ならば「0」が入力されることで成績データが蓄積される。

### 3. 形式概念分析

形式概念分析<sup>(3)(4)</sup>は、オブジェクトと対象が内包する属性の対から構成されたコンセプトを数学的に定義されたデータとして扱い、コンセプトの階層関係を表現するコンセプトラティスの生成や属性間の包含関係を論理的帰結関係である含意論理として抽出する手法である。縦軸をオブジェクト、横軸を属性としてそれらの対応関係を示すコンテキスト表からは、属性間の全ての包含関係を包含する含意関係を読み取ることは困難であるが、形式概念分析を適用することでそれらを抽出することが可能である。表 1 に動物を例としたコンテキスト表を示す。

表 1 コンテキスト表の例

	陸上	水中	空中	羽	足
イヌ	×				×
カエル	×	×			×
コイ		×			
ダチョウ	×			×	×
ワシ	×		×	×	×

このコンテキスト表からは、

- 足 → 陸上
- 陸上 → 足
- 空中 → 陸上 羽 足
- 羽 → 陸上 足

の 4 つの含意論理が得られる。ここで「空中」に着目すると「空中→羽根」、「空中→足」、「空中→陸上」といった包含関係がコンテキスト表からは読み取れるが形式概念分析では、このような包含関係を「空中→陸上 羽 足」といった属性間の全ての包含関係を包含する含意関係を含意論理の形で記述することができる。ただし形式概念分析は、コンテキスト表の中でのみ成り立つ包含関係を全て含意論理の形で抽出するもので

あり、ここで得られた関係が普遍的に成立するものとは限らない。

## 4. 形式概念分析を用いた成績データの分析

本節では、実習型 e-learning システム「ナレロー」の成績データを対象に形式概念分析を適用して抽出した含意論理の結果と Fisher の正確確率検定を比較し、考察したことについて述べる。対象とした成績データは、2016 年から 2019 年の富山県立大学工学部知能ロボット工学科 1 年次生が講義中に取り組んだデータを用いる。対象としたのは全問題を完了したユーザ 183 人で対象とした問題は 110 問である。

### 4.1 分析データ

各年で成立する含意論理を抽出するために 1 年ごとに成績データを分割した。成績データからオブジェクトをユーザ、ユーザが正解した問題を属性とするコンテキスト表と不正解となった問題を属性とするコンテキスト表を作成した。また、正解した問題を属性とするコンテキスト表を作成する際に 4 年間全ての年で正解率が 90% 以上の問題を属性から削除した。ほとんどのユーザが正解となり含意論理の結論部に頻繁に出現することが自明なためである。また、不正解となった問題を属性とするコンテキスト表を作成した際に 4 年間とも正解率が 20% 以下の問題を属性から削除した。削除した問題はほとんどのユーザが不正解となり、含意論理の結論部に頻繁に修験することが自明なため分析対象から外した。各年のオブジェクト数と属性数を表 2 に示す。

表 2 オブジェクト数と属性数

	2016	2017	2018	2019
オブジェクト数	41	51	35	56
属性数	正解	109	109	109
	不正解	89	89	89

### 4.2 含意論理の抽出結果

Concept Explorer<sup>(5)</sup>を使用して各年度ごとのコンテキスト表から含意論理を抽出する。ある年で得られた含意論理を  $I_n^p$  とする。 $I_n^p$  の上付き文字  $n$  は成立した

年を示し、下付き文字  $c$  は正解(correct)の含意論理を示し、不正解(incorrect)の含意論理の場合は  $I_{inc}^n$  と表示する。例えば 2016 年で成立する正解の含意論理  $I_{c1}^{2016} : P105 \rightarrow P112 (33/41)$

は、「2016 年において P105 の問題に正解したユーザは結論部の P112 も必ず正解する」と読み取ることができる。なお、含意論理の横の ( ) 内の数字は含意論理の対象となるユーザの数とその年のユーザ全体の数との割合を示している。形式概念分析を適用し、得られた含意論理に出現した問題を表 3 に示し、得られた含意論理の数を表 4 に示す。今回は得られた含意論理の数が多いため含意論理が成立する対象となるオブジェクト数が多いものに着目し、各年で成立数が多いものを 5 個ずつ抜粋したものを統計的な分析との比較に用いる。

表 3 含意論理に出現した問題

問題番号 (出題順)	問題内容
P6	再変換
P7	名前を付けて保存
P9	文書を開く
P10	新規作成
P11	画面を拡大する
P12	編集記号の表示切り替え
P13	文章を入力する
P14	頭語と結語
P20	コピーする
P21	移動する
P23	フォント、フォントサイズ
P24	フォントスタイル
P27	中央揃えと右揃え
P29	箇条書き
P30	ページ設定
P33	文字を入力する
P35	表内の文字の配置
P37	表の配置を変更する
P46	表紙と空白ページの挿入
P47	ワードアートの挿入
P50	図の水平方向の配置
P54	図形の書式を変更
P65	見出しとスタイル① (スタイルの適用)
P69	段落番号の適用 (番号種類指定)
P71	ブックマーク
P78	段落後の間隔
P79	インデントの設定
P83	段組の適用
P86	表の編集 (線の太さ、色、背景色)
P87	表のクイックスタイルと並べ替え
P90	用紙サイズの変更
P91	余白と印刷の向きの変更
P97	コメントの挿入
P98	コメントの表示 ([変更履歴]ウィンドウ)
P102	最終版として保存
P105	部数を指定して印刷
P111	OneDriveに保存
P112	OneDriveの文書を開く

表 4 含意論理の数

	2016 年	2017 年	2018 年	2019 年
正解	3253	6389	6715	18332
不正解	4469	7489	6690	16741

#### 4.2.1 正解問題の含意論理

正解した問題を属性としたコンテキスト表から抽出した含意論理を以下に示す。

##### (1) 2016 年

$I_{c1}^{2016} : P105 \rightarrow P112 (33/41)$

$I_{c2}^{2016} : P29 \rightarrow P112 (32/41)$

$I_{c3}^{2016} : P111 \rightarrow P112 (30/41)$

$I_{c4}^{2016} : P90 \rightarrow P112 (27/41)$

$I_{c5}^{2016} : P10, P33 \rightarrow P112 (26/41)$

##### (2) 2017 年

$I_{c1}^{2017} : P20, P112 \rightarrow P6 (35/51)$

$I_{c2}^{2017} : P20, P54 \rightarrow P6 (29/51)$

$I_{c3}^{2017} : P20, P105 \rightarrow P6 (29/51)$

$I_{c4}^{2017} : P11, P21, P23 \rightarrow P10 (28/51)$

$I_{c5}^{2017} : P9, P27 \rightarrow P7 (28/51)$

##### (3) 2018 年 全員 P10 に正解

$I_{c1}^{2018} : P9 \rightarrow P6 (32/35)$

$I_{c2}^{2018} : P54 \rightarrow P112 (32/35)$

$I_{c3}^{2018} : P7 \rightarrow P6 (30/35)$

$I_{c4}^{2018} : P111 \rightarrow P112 (30/35)$

$I_{c5}^{2018} : P105 \rightarrow P20, P112 (29/35)$

##### (4) 2019 年

$I_{c1}^{2019} : P29 \rightarrow P9, P10 (52/56)$

$I_{c2}^{2019} : P7 \rightarrow P9 (49/56)$

$I_{c3}^{2019} : P54 \rightarrow P9 (48/56)$

$I_{c4}^{2019} : P20 \rightarrow P9 (47/56)$

$I_{c5}^{2019} : P14 \rightarrow P112 (47/56)$

上記に示した 20 個の含意論理のうち複数年で共通して成立する含意論理について調査した。全ての含意論理を対象に調べた結果 2016 年、2018 年、2019 年で「P111 が正解ならば P112 も正解する」が成立し 2016 年、2018 年で「P105 正解ならば P112 も正解する」が成立した。その他の含意論理は全て単年度でのみ成立する含意論理である。

#### 4.2.2 不正解問題の含意論理

不正解となった問題を属性としたコンテキスト表から抽出した含意論理を以下に示す。

(5) 2016年 P71 P78 P83 P86 は全員不正解

$I_{inc 1}^{2016}$  : P87 → P102 (39/41)

$I_{inc 2}^{2016}$  : P79 → P87, P102 (37/41)

$I_{inc 3}^{2016}$  : P69 → P102 (37/41)

$I_{inc 4}^{2016}$  : P91 → P47 (37/41)

$I_{inc 5}^{2016}$  : P98, P102 → P87 (37/41)

(6) 2017年

$I_{inc 1}^{2017}$  : P65 → P78 (48/51)

$I_{inc 2}^{2017}$  : P91 → P71 (47/51)

$I_{inc 3}^{2017}$  : P78, P83 → P86 (46/51)

$I_{inc 4}^{2017}$  : P71, P97 → P91 (46/51)

$I_{inc 5}^{2017}$  : P47, P78 → P65 (45/51)

(7) 2018年

$I_{inc 1}^{2018}$  : P37, P50 → P46 (26/35)

$I_{inc 2}^{2018}$  : P13 → P37 (24/35)

$I_{inc 3}^{2018}$  : P12, P46 → P30, P50 (24/35)

$I_{inc 4}^{2018}$  : P35, P46 → P30 (24/35)

$I_{inc 5}^{2018}$  : P14 → P24 (23/35)

(8) 2019年

$I_{inc 1}^{2019}$  : P78 → P83 (53/56)

$I_{inc 2}^{2019}$  : P71 → P83 (52/56)

$I_{inc 3}^{2019}$  : P87 → P83 (52/56)

$I_{inc 4}^{2019}$  : P102 → P83 (50/56)

$I_{inc 5}^{2019}$  : P65 → P83 (48/56)

上記に記した含意論理には複数年で共通して成立する含意論理を抽出された全ての含意論理を対象に調査した結果, 上記に記した含意論理は全て単年度で成立するものであり複数年で共通したものは発見されなかった。

#### 4.3 Fisher の正確確率検定との比較分析

2変数の間に統計学的に有意な差があるか検定する手法のひとつである Fisher の正確確率検定を用いて含意論理の前提部の正誤と結論部の正誤に関連性があるか分析し, 結果を報告する。4.3節で取りあげた含意論理のうち, 前提部および結論部の正解と不正解を 2

変数とし, Fisher の正確確率検定を適用することで前提部の正誤と結論部の正誤に関連があるか検証した。前提部が 1 問のみで構成されている含意論理では, 表 5 に示す分割表を作成し, 分析を行なった。前提部が複数の問題で構成された含意論理は, 表 6 に示す分割表を作成し, 分析を行なった。

表 5 分割表の例  $I_{inc 1}^{2016}$

		P112		合計
		正解	不正解	
P105	正解	107	5	112
	不正解	26	4	30
合計		133	9	142

表 6 前提部に複数の問題を含む分割表の例  $I_{inc 5}^{2016}$

		P112		合計
		正解	不正解	
P10, P33	正解	89	44	133
	その他	5	4	9
合計		94	48	142

表 6 のその他とは「片方の問題だけ正解あるいは両方不正解」のことを指す。なお, 含意論理が発見された年以外の成績データを対象に検定を行なった。例えば, 2016 年で成り立つ含意論理  $I_{c 2}^{2016}$  について検定を行う際は, 2017 年, 2018 年, 2019 年の 3 年分のデータを分析の対象とする。今回は有意水準を 0.05 とした両側検定によって検定を行った。

##### 4.3.1 検定結果

検定の結果を表 7, 8 にそれぞれ示す。表の見方は 縦軸に含意論理, 横軸を年とし, 対応箇所は 4.3 節で作成した分割表から求めた p 値を記す。

表 7 各年の含意論理と p 値 (正解)  $p < 0.05$

n	2016	2017	2018	2019
$I_{c 1}^n$	0.374	0.114	0.056	※3
$I_{c 2}^n$	0.219	0.277	0.013	0.005
$I_{c 3}^n$	1.000	0.271	0.266	1.000
$I_{c 4}^n$	0.278	0.148	1.000	0.051
$I_{c 5}^n$	0.486	0.208	※2	1.000

※2 P105 → P20 [p = 0.173], P105 → P112 [p = 0.374]

※3 P29 → P9 [p = 0.310], P29 → P10 [p = 0.067]

表 8 各年の含意論理と p 値 (不正解)  $p < 0.05$

n	2016	2017	2018	2019
$I_{inc 1}^n$	0.000	0.218	0.073	0.002
$I_{inc 2}^n$	※4	0.071	0.168	0.335
$I_{inc 3}^n$	0.000	0.000	0.059	0.001
$I_{inc 4}^n$	0.168	0.066	0.000	0.001
$I_{inc 5}^n$	0.000	0.015	0.124	0.705

※4 P79 → P87 [ $p = 0.006$ ], P79 → P102 [ $p = 0.001$ ]

検定結果より, 正解の含意論理では, 4.2.1 節に列挙した 20 個の含意論理のうち, 2 つの含意論理において, 前提部の正誤と結論部の正誤の間に統計的な関連性が認められた. 不正解の含意論理では, 4.2.2 節に列挙した 20 個の含意論理のうち 11 個の含意論理で前提部の正誤と結論部の正誤の間に関連性があることが認められた.

#### 4.3.2 含意論理との比較と考察

形式概念分析によって得られた含意論理の前提部の正誤が結論部の正誤と関連しているか調べるために Fisher の正確確率検定を行なった. 本稿では, 抽出した含意論理の中でも対象となるオブジェクト数が多いものを各年度からそれぞれ 5 個ずつ抜粋し, 正解の含意論理 20 個, 不正解の含意論理 20 個を Fisher の正確確率検定との比較に用いた. 検定の結果, 正解の含意論理では, 20 個のうち 2 個の含意論理で前提部と結論部の正誤の関連性が認められた. 不正解の含意論理では, 20 個のうち 10 個の含意論理で前提部と結論部の正誤の関連性が認められた. 正誤の関連性が認められたのは全て単年度で成り立つ含意論理であった. 複数年度で共通して成り立つ含意論理  $I_{c1}^{2016}, I_{c5}^{2018}$  と  $I_{c3}^{2016}, I_{c4}^{2018}$  を対象とした Fisher の正確確率検定の結果, それぞれの含意論理の前提部と結論部の間には関連性が確認されなかった.  $I_{c3}^{2016}, I_{c4}^{2018}$  は, 2019 年も含め 3 年間で共通して成立したもので今回は成立した年以外のデータを対象にしたため, 2017 年のみのデータが対象となった. データの頻度によっては, 関連性を見つけられない可能性がある. そこで 4 年分のデータを対象に検定を行ったところ  $I_{c3}^{2016}, I_{c4}^{2018}$  の p 値は [ $p = 0.004$ ] となり, 有意水準を満たす. データ全体で見ると前提部と結論部の正誤に関連性が認められた.  $I_{c1}^{2016}, I_{c5}^{2018}$  は 2017 年と

2019 年の 2 年分のデータを対象に分析を行なった. そこで 4 年分のデータを対象に分割表を作成し検定を行った結果,  $I_{c1}^{2016}, I_{c5}^{2018}$  の p 値は [ $p = 0.011$ ] となり, データ全体を対象にした際には, 前提部と結論部の正誤に関連性が認められた.

形式概念分析によって得られた含意論理のうち, 成立数の多いものに着目したところ前提部と結論部に統計的な関連性が認められる含意論理を発見することができた. 以上から形式概念分析を用いて成績データを分析した際に得られる正解問題/不正解問題の含意論理の中には統計的分析への足がかりとして有用である可能性が示唆された.

## 5. まとめ

本稿では, 「ナレロー」の成績データを対象に形式概念分析を適用し, 含意論理によって捉えた正解問題と不正解問題の包含関係と統計的な分析によって得た全体的な傾向の比較を行った. まず, 2016 年から 2019 年の富山県立大学知能ロボット工学科 1 年生のナレローの成績データからユーザをオブジェクト, 正解した問題あるいは不正解となった問題を属性とするコンテキスト表を作成した. さらに 1 年ごとにデータを分割し, 形式概念分析を適用することで年ごとに成立する正解問題/不正解問題の包含関係を含意論理として抽出した. 各年で成立数が多い順に 5 つの含意論理を抜粋し, 合計で 40 個の含意論理を統計的分析との比較対象とした. 抽出された含意論理の中から「P111 が正解ならば P112 も正解する」が 2017 年を除く全ての年で成り立ち「P105 が正解ならば P112 も正解する」が 2016 年と 2018 年で共通して成り立つことがわかった. 一方, 他の含意論理は全て単年度でのみ成り立つものであった. Fisher の正確確率検定を行い得られた含意論理のうち, 前提部と結論部の正誤が関連しているか検証した. 検証の結果, 成立数の高い正解の含意論理 20 個のうち, 2 つが統計的に見て前提部と結論部の正誤に関連性があることが認められた. また不正解の含意論理においては成立数が多いもの 20 個のうち 11 個の含意論理で前提部と結論部の正誤の関連が認めら

れた.形式概念分析によって得られた含意論理のうち成立数の多いものに関しては,統計的な観点からも前提部と結論部に関連を持つものが見られ,統計的手法へのトリガとして有用であると考えられる.

本稿では,形式概念分析によって得た含意論理のうち,各年で成立数の多いものを順に5つ選びそれらを統計的な手法との比較に使用していた.今後は,含意論理の成立数した割合に着目し,成立数の小さいものから大きいものまでを統計的な分析との比較に用いることで統計分析への利用可能性について検討していく必要がある.また今回は,富山県立大学知能ロボット工学科1年生のデータを対象としたており,富山県立大学知能ロボット工学科の間でのみ成り立つ関係性であることを否定できないため,他の教育機関にてナレローの成績データを収集し,本稿の結果と比較する必要があると考えられる.

#### 参 考 文 献

- (1) 石川晶子,小川賀代:"学習履歴データを活用した学習者の特性抽出手法の検討"教育システム情報学会誌, Vol.31, No.2, pp.185-196, (2014)
- (2) ナレロー, [https://narero.com/index\\_main.html](https://narero.com/index_main.html) (2021年11月25日確認)
- (3) 鈴木治,室伏俊明:"形式概念分析-入門・支援ソフト・応用."知能と情報 Vol.19, No.2, pp.103-142, (2007)
- (4) R.Wille : "Restructuring lattice theory: an approach based on hierarchies of concepts." International conference on formal concept analysis. Springer, Berlin, Heidelberg, (2009)
- (5) The Concept Explorer, <http://conexp.sourceforge.net> (2021年11月25日確認)
- (6) R.A.Fisher : "On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P." Journal of the Royal Statistical Society, Vol.85, No.1, pp.87-94, (1922)