

# トピックモデルとベイジアンネットワークによる 段階評価アンケートの分析

崎 野 三 太 郎\*

Analysis of Stage Evaluation Questionnaire by Topic Models and Bayesian Network

Santaro SAKINO\*

Key words : トピックモデル            Topic Models  
               ベイジアンネットワーク    Bayesian Network  
               段階評価                            Stage Evaluation Questionnaire

## 1. はじめに

本学の児童学科4年次で、前期後期の2期に渡り週1回小学校における学校教育体験実習を実施している。実習終了後、学生に対し5段階評価による前期30設問、後期31設問と自由記述のアンケートを実施している。今回対象にした平成26年度の分析は平成27年に報告（小澤、山崎、崎野、吉田2015）している。しかしながら、この5段階評価の数値を文字として捉えると、自然言語処理の潜在意味解析の文脈で分析することができる。さらに、潜在意味解析のトピックモデルをソフトクラスタリングと捉えると、トピックモデルとベイジアンネットワークからソフトクラスタリング・コミュニティ分析ができる可能性があり、新たな分析方法と知見が期待できる。

本研究は、「5段階評価データをテキストデータとしてトピックモデルで解析し、抽出されたトピックのベイジアンネットワークを求め、トピック間の因果関係を推測する」という一連の処理を、フリーのツールで処理できる方法を探ることが目的である。これにより、潜在意味解析やベイジアンネットワークを容易に使用できる一助になると思われる。

## 2. 調査方法

調査対象 東北女子大学 小学校教員養成課程

4年次履修生 28人

アンケート実施日 平成26年12月

実習後期のアンケート 31設問5段階評価、自由記述

アンケート項目 平成26年度本学紀要参照

使用データ 31設問5段階評価

## 3. 用語の説明

今回使用した手法について簡単に説明する。

- トピックモデル (Topic models)  
1つの文書が複数のトピック（分野と考えればいい）を持つと仮定し、その背後に隠れているトピック（潜在意味）を推定するモデル。
- ベイジアンネットワーク (Bayesian network)  
確率構造に非巡回有向グラフ (Directed Acyclic Graph) の仮定をおいた、ネットワーク構造（重みづけグラフ）である。確率変数をノードで表し、確率変数間の因果関係の推論を非巡回有向グラフ構造により表すことができる。
- TTM (Tiny Text Miner)  
大阪大学大学院経済学研究科 松村研究所の松村真宏氏が公開しているフリーのテキストマイニングツール。形態素解析にMeCabを利用し、単語の出現頻度と出現件数（単語を含むサンプル件数）を出力できる。
- ソフトクラスタリング (Overlapping cluster)  
どのノード（頂点、要素）にも2つ以上のクラスタに属することを許すクラスタリング。

\*東北女子大学

4. 使用ツール

- R：統計解析用のフリーのソフトウェア
- トピックモデル：Rのパッケージ“topicmodels”
- ベイジアンネットワーク：Rのパッケージ“deal”
- 形態素解析：TTM
- Microsoft Office EXCEL

5. 方法と結果

①31設問5段階評価の数値データの個々の数字を一つの文字として捉えると、31設問の回答は31個の文字列と見なすことができる。TTM用に数値データのCSVファイルから、スペース区切りの文字データを作る。このとき、設問と評価が分かるようにデータを作成する。

表1 31設問の5段階評価

設問	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	
1	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
2	4	4	5	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5
3	5	4	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5
4	5	4	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	4	5	5	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
6	5	4	5	5	5	4	4	5	4	5	5	5	5	5	5	5	5	5	5
7	5	5	5	5	4	4	5	4	5	5	3	4	5	5	5	5	5	5	5
8	5	4	5	5	5	5	4	5	5	5	4	5	5	5	5	5	5	5	5
9	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5
10	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
11	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
12	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5
13	5	4	5	4	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5
14	5	4	5	5	4	5	5	5	4	4	4	4	5	5	5	5	5	5	5
15	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
16	5	5	5	4	5	4	3	5	4	5	2	4	3	3	3	5	5	5	5
17	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
18	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

表2 TTM用のスペース区切りデータ

設問	15	24	35	45	55	65	75	85	95	105	115	125	135	145
s01	15	24	35	45	55	65	75	85	95	105	115	125	135	145
s02	14	24	35	44	54	64	75	85	95	105	115	125	135	145
s03	15	24	35	45	55	65	74	85	95	104	115	125	135	145
s04	15	24	35	44	55	65	75	85	95	105	115	125	135	145
s05	15	24	35	45	54	64	75	84	94	105	114	124	135	145
s06	15	24	35	45	55	65	74	85	95	105	114	124	135	144
s07	15	25	35	45	55	64	74	85	94	105	115	123	134	145
s08	15	24	35	45	55	65	75	85	94	105	115	125	134	145
s09	15	25	35	45	55	65	75	85	95	105	114	124	135	145
s10	15	25	35	45	55	65	75	85	95	105	115	125	135	145
s11	15	25	35	45	55	65	75	85	95	105	115	125	135	145
s12	15	25	35	45	55	65	75	85	95	105	115	125	135	144
s13	15	24	35	44	55	64	74	85	95	105	115	125	135	145
s14	15	24	35	45	55	64	75	85	95	105	114	124	134	144
s15	15	25	35	45	55	65	75	85	95	105	115	125	135	145

②一人の学生のデータを一つの文書とし、全学生のデータをTTMで構文解析を行い、学生と数字の出現頻度のクロス集計表を求める。クロス集計表は、0が多い疎（スパース）行列になる。

表3 TTMで出力した設問×学生の頻度表

語	品詞	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10	s11
265	キー	1	1	1	1	1	1	1	1	1	1	1
15	キー	1	0	1	1	1	1	1	1	1	1	1
35	キー	1	1	1	1	1	1	1	1	1	1	1
165	キー	1	1	1	1	1	1	1	1	1	1	0
55	キー	1	0	1	1	0	1	1	1	1	1	1
175	キー	1	1	1	0	0	1	1	1	1	1	1
105	キー	1	1	0	1	1	1	1	1	1	1	1
85	キー	1	1	1	1	0	1	1	1	1	1	1
225	キー	1	1	1	0	1	1	1	1	1	1	1
185	キー	1	1	1	0	1	1	1	1	1	1	0

③はじめにRの“topicmodels”で、TTMの頻度表からトピック数の最適値を求める。パラメータは以下にした。

```
burnin <- -4000, iter <- -2000, thin <- -500
seed <- -list (2003,5,63,100001,765)
nstart <- -5, best <- -TRUE
```

データについて、トピック数を2から50までの対数尤度を求め、対数尤度（Log likelihood）が最大値をとる時のトピック数を採用する。採用トピック数は13である。

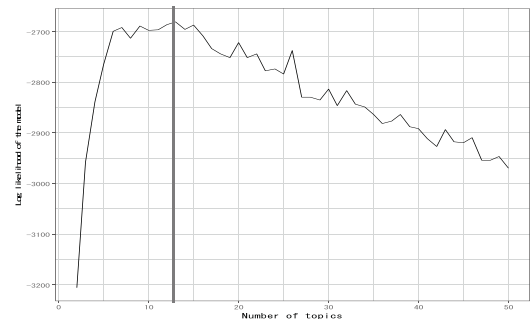


図1 対数尤度とトピック数のグラフ

④Rの“topicmodels”で、TTMの頻度表から13個のトピックを求める。

表4 topicにおける各設問の占める割合

	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
265	0.0013	0.0011	0.0014	0.0015	0.0906	0.0757	0.0014
15	0.0146	0.2048	0.0014	0.0015	0.0794	0.0015	0.0014
35	0.0013	0.0011	0.0014	0.0015	0.0011	0.0015	0.0014
165	0.0013	0.0238	0.0014	0.0015	0.0459	0.0015	0.0014
55	0.0013	0.0011	0.1719	0.1069	0.0011	0.0015	0.0574
175	0.0013	0.0011	0.0014	0.0015	0.0011	0.0015	0.1975
105	0.0013	0.0011	0.0014	0.0015	0.0011	0.2834	0.0014
85	0.0013	0.0011	0.0014	0.0015	0.1353	0.0608	0.0014
225	0.1207	0.0011	0.0014	0.1672	0.0011	0.0163	0.0014
185	0.0013	0.0011	0.0156	0.0015	0.1353	0.0163	0.0154
305	0.1207	0.0690	0.0014	0.0015	0.0011	0.0015	0.0154
195	0.0411	0.0011	0.0582	0.0015	0.0570	0.0015	0.0014
115	0.0146	0.0011	0.1151	0.0015	0.0011	0.0015	0.0014
145	0.0544	0.1595	0.0014	0.0015	0.0011	0.0015	0.0014
75	0.0013	0.0011	0.0014	0.0015	0.1130	0.0015	0.0014
95	0.0146	0.0011	0.0014	0.0015	0.0011	0.0015	0.0014
135	0.0942	0.0011	0.0014	0.1220	0.0011	0.0015	0.0014

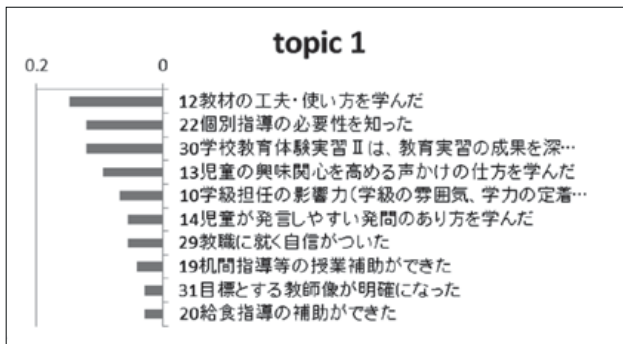


図2 topic 1内の設問の占める割合(上位10)

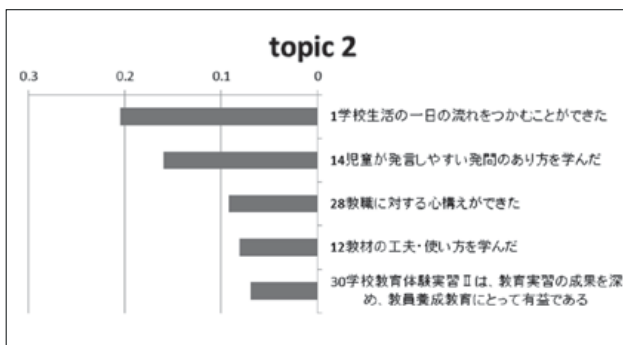


図3 topic 2内の設問の占める割合(上位5)

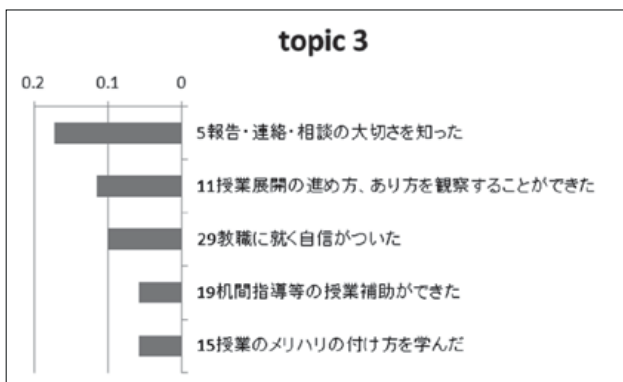


図4 topic 3内の設問の占める割合(上位5)

⑤トピック1～13の上位5を占める設問を示す。トピックの内容から各トピックのラベル付けができる。今回は行わない。

トピック1は「12教材の工夫・使い方を学んだ。22個別指導の必要性を知った。30学校教育体験実習Ⅱは、教育実習の成果を深め、教員養成教育にとって有益である。13児童の興味関心を高める声かけの仕方を学んだ。10学級担任の影響力(学級の雰囲気、学力の定着等)の大きさに気付いた。」

トピック2は「1学校生活の一日の流れをつかむことができた。14児童が発言しやすい発問のあり方を学んだ。28教職に対する心構えができた。12教材の工夫・使い方を学んだ。30学校教育体験実習Ⅱは、教育実習の成果を深め、教員養成教育にとって有益である。」

トピック3は「5報告・連絡・相談の大切さを知った。11授業展開の進め方、あり方を観察することができた。29教職に就く自信がついた。19机間指導等の授業補助ができた。15授業のメリハリの付け方を学んだ。」

トピック4は「22個別指導の必要性を知った。25学年教科の教材の内容・配列を理解できた。13児童の興味関心を高める声かけの仕方を学んだ。5報告・連絡・相談の大切さを知った。20給食指導の補助ができた。」

トピック5は「8児童は教師の特徴・態度等をよく観察していると思った。18掃活動ができた。31目標とする教師像が明確になった。7児童の名前と一人一人の特徴をつかむことができた。26教師には授業以外に多くの仕事があることを知った。」

トピック6は「10学級担任の影響力(学級の雰囲気、学力の定着等)の大きさに気付いた。6担当クラスの様子、実態を理解することができた。26教師には授業以外に多くの仕事があることを知った。8児童は教師の特徴・態度等をよく観察していると思った。27教師の仕事のやりがいとすばらしさを知り、教師になりたいという気持ちが強くなった。」

トピック7は「17示物の手伝いができた20給食指導の補助ができた。24正誤判断は教師の指導を受けてできた。5報告・連絡・相談の大切さを知った。6担当クラスの様子、実態を理解することができた。」

トピック8は「15授業のメリハリの付け方を学んだ。4クラス担当教員とのコミュニケーションができた。8児童は教師の特徴・態度等をよく観察していると思った。26教師には授業以外に多くの仕事があることを知った。17示物の手伝い」ができた。」

トピック9は「16プリント、ドリルの丸付けができた。23正誤判断の難しさを知った。21児童の学力の実態を知ることができた。26教師には授業以外に多くの仕事があることを知った。25学年教科の教材の内容・配列を理解できた。」

トピック10は「2教師がやっている一日の仕事・勤務状態がつかめた。3子どもたちの一日の学習活動・状況がつかめた。9児童一人一人への教師の対応方法を学ぶことができた。27教師の仕事のやりがいとすばらしさを知り、教師になりたいという気持ちが強くなった。11授業展開の進め方、あり方を観察することができた。」

トピック11は「9児童一人一人への教師の対応方法を学ぶことができた。2教師がやっている一日の仕事・勤務状態がつかめた。15授業のメリハリの付け方を学んだ。24正誤判断は教師の指導を受けてできた。14児童が発言しやすい発問のあり方を学んだ。」

トピック12は「4クラス担当教員とのコミュニケーションができた。3子どもたちの一日の学習活動・状況がつかめた。25学年教科の教材の内容・配列を理解できた。7児童の名前と一人一人の特徴をつかむことができた。27教師の仕事のやりがいとすばらしさを知り、教師になりたいという気持ちが強くなった。」

トピック13は「13児童の興味関心を高める声かけの仕方を学んだ。18掃活動ができた。3子どもたちの一日の学習活動・状況がつかめた。17示物の手伝い」ができた。30学校教育体験実習Ⅱ

は、教育実習の成果を深め、教員養成教育にとって有益である」

⑥ 学生の各々のトピックの割合を求める。

表5 学生の各トピックの占める割合

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
topic 1	0.097	0.122	0.097	0.085	0.085	0.047	0.047	0.097	0.060	0.072	0.072	0.072	0.047
topic 2	0.072	0.047	0.072	0.109	0.085	0.109	0.109	0.047	0.060	0.109	0.072	0.072	0.085
topic 3	0.047	0.060	0.085	0.060	0.072	0.109	0.072	0.060	0.060	0.047	0.085	0.072	0.097
topic 4	0.060	0.072	0.060	0.072	0.072	0.060	0.072	0.085	0.072	0.085	0.109	0.097	0.060
topic 5	0.097	0.109	0.085	0.085	0.072	0.072	0.085	0.097	0.122	0.097	0.072	0.097	0.085
topic 6	0.072	0.072	0.109	0.085	0.109	0.072	0.060	0.085	0.060	0.060	0.072	0.072	0.072
topic 7	0.085	0.060	0.072	0.072	0.097	0.097	0.085	0.085	0.072	0.085	0.060	0.072	0.085
topic 8	0.097	0.085	0.072	0.060	0.072	0.060	0.097	0.072	0.097	0.072	0.060	0.060	0.085
topic 9	0.085	0.072	0.097	0.060	0.085	0.097	0.060	0.072	0.097	0.097	0.085	0.109	0.060
topic 10	0.097	0.097	0.072	0.109	0.047	0.072	0.072	0.047	0.085	0.060	0.060	0.072	0.109
topic 11	0.060	0.072	0.060	0.047	0.072	0.060	0.060	0.072	0.060	0.097	0.109	0.060	0.085
topic 12	0.085	0.072	0.072	0.097	0.072	0.060	0.097	0.109	0.109	0.072	0.072	0.060	0.047
topic 13	0.047	0.060	0.047	0.060	0.060	0.085	0.085	0.072	0.047	0.047	0.072	0.085	0.085

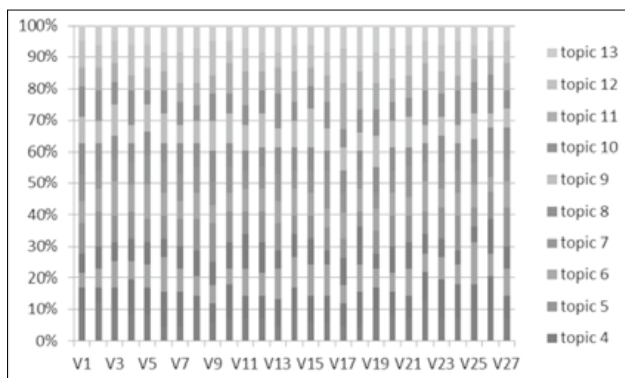


図5 学生の各トピックの占める割合

⑦ Rの“deal”で、トピックの割合の行列からベイジアンネットワークを求める。トピック間の有向グラフが得られる。Topic4,6,7は他のtopicとのつながりはない。“deal”では、データが連続変数のとき、各ノード（頂点）の条件付き確率は出力されず、回帰係数が出力される。

表6 親ノードの一覧

1	topic.1	continuous()	2	13
2	topic.2	continuous()		
3	topic.3	continuous()	12	
4	topic.4	continuous()		
5	topic.5	continuous()	3	11
6	topic.6	continuous()		
7	topic.7	continuous()		
8	topic.8	continuous()	5	
9	topic.9	continuous()		
10	topic.10	continuous()	9	11
11	topic.11	continuous()		
12	topic.12	continuous()		
13	topic.13	continuous()	5	

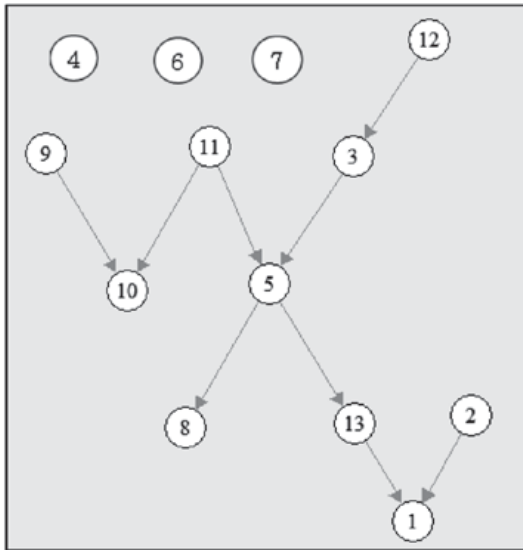


図6 トピックのベイジアンネットワーク

⑧ ベイジアンネットワークから因果関係を推測する。トピック番号の前にTをつける。T1「12教材の工夫・使い方を学んだ。22個別指導の必要性を知った」はT2「1学校生活の一日の流れをつかむことができた。14児童が発言しやすい発問のあり方を学んだ」とT13「13児童の興味関心を高める声かけの仕方を学んだ。18清掃活動」から影響を受けている。T8「15授業のメリハリの付け方を学んだ。4クラス担当教員とのコミュニケーションができた」はT5「8児童は教師の特徴・態度等をよく観察していると思った。18清掃活動」から影響を受けている。T5「8児童は教師の特徴・態度等をよく観察していると思った。18清掃活動」はT3「5報告・連絡・相談の大切さを知った。11授業展開の進め方、あり方を観察することができた」とT11「9児童一人一人への教師の対応方法を学んだ。2教師がやっている一日の仕事・勤務状況がつかめた」から影響を受けている。T10「2教師がやっている一日の仕事・勤務状況がつかめた。3子どもたちの一日の学習活動・状況がつかめた」はT9「16プリント、ドリルの丸付け。23正誤判断の難しさを知った」とT11「9児童一人一人への教師の対応方法を学んだ。2教師がやっている一日の仕事・勤務状況がつかめた」から影響を受けている。T3「5

報告・連絡・相談の大切さを知った。11授業展開の進め方、あり方を観察する」はT12「4クラス担当教員とのコミュニケーションができた。3子どもたちの一日の学習活動・状況がつかめた」から影響を受けている。さらに、T1はT13、T5を介してT3、T11、さらにT3を介してT12から影響を受けている。T8はT5を介して、T3、T11、T12から影響を受け、T5はT3を介してT12から影響を受けている、と推測できる。また、T4「22個別指導の必要性を知った。25学年教科の教材の内容・配列を理解できた」、T6「10学級担任の影響（学級の雰囲気、学力の定着等）の大きさに気付いた。6担当クラスの様子、実態を理解することができた」、T7「17掲示物の手伝い。20給食指導の補助」は他のトピックとつながりが無い。

## 6. トピックに含まれる上位10の設問数

13個のトピックに含まれる上位10の設問項目の数を見ると、6個以上のトピックに含まれる設問は、「7児童の名前と一人一人の特徴をつかむことができた」、「11授業展開の進め方、あり方を観察する」、「19机間指導等の授業補助ができた」、「26教師には授業以外に多くの仕事があることを知った」、「28教職に対する心構えができた」、「29教職に就く自信がついた」である。これらの設問は、いろいろな文脈で使われていることを表している。この中で設問28は上位5の中に入っていない。しかし、設問28は考えの基になっていると推測できる。

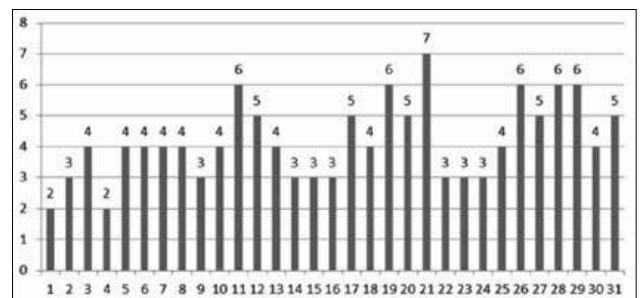


図7 トピックに含まれる上位10の設問数

## 7. まとめ

5段階評価の数値データを、文字データとして扱うとトピックモデルで解析できることが分かった。また、各学生のトピックを求めることでトピックのベイジアンネットワークが求められることが分かった。トピックモデルは、1つの文書が複数のトピックからできていることを仮定している。また、トピックは、複数の単語を含むことを許している。今回の例では、一つのトピックに複数の設問が含まれるが、そのトピックの中で占める割合は各々異なっている。つぎにベイジアンネットワークで、トピック間のつながりを有向矢印で示すことができるのでトピック間の因果関係を推測することができる。今回は行わなかったが、各トピックを、構成する設問でラベル付けすることにより、ネットワークのつながりを把握し易くなる。

今回の分析で、例えば、T1「教材の工夫・使い方を学んだ。個別指導の必要性を知った」は、T2「学校生活の一日の流れをつかむことができた。児童が発言しやすい発問のあり方を学んだ」とT13「児童の興味関心を高める声かけの仕方を学んだ。清掃活動」の影響を受けていることから、一通りの解釈だけでなく、いくつもの解釈ができる。

また、孤立しているトピックT4「個別指導の必要性を知った。学年教科の教材の内容・配列を理解できた」等も、なぜ、孤立しているかを考えることができる。

トピックモデルとベイジアンネットワークをソフトクラスタリング・コミュニティ分析と見ると、各トピックに含まれる割合が小さい設問も、6.の結果の設問28のように複数のトピックに所属している場合、意味を見出せる。

今回述べなかったが、設問を直接ベイジアンネットワークで分析すると、設問間の因果関係を推測できる。これについては、別の機会に述べたいと思う。

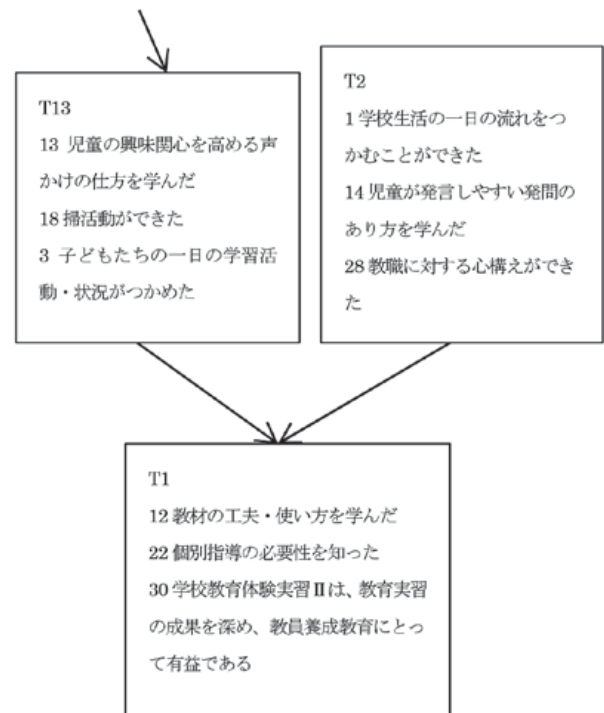


図7 ベイジアンネットワーク（一部）

これらの分析から学生が実習から得た事項の関連付けがどのようになっているのかが分かり、学生の実習の指導に役立てることが期待できる。今後、自由記述との関連も分析できるようにしたい。

## 参考文献

- 小澤、山崎、岩見、崎野、吉田（2013）学校教育体験実習Ⅰ・Ⅱ」に関する実践研究 ～教育実習の事前・事後体験教育の検討～『紀要52』、p.1-10、東北女子大学 東北女子短期大学
- 小澤、山崎、崎野、吉田（2015）学校教育体験実習Ⅰ・Ⅱ」に関する実践研究 2 一学生と教員に対するアンケート調査及びテキストマイニングから一、『紀要54』、p.1-14、東北女子大学 東北女子短期大学