

深層学習を用いた MOOCs の学習者の知識構造の分析

37-146856 那須野薫

指導教員 松尾豊特任准教授

1 研究背景および目的

伝統的に、知識は学習や指導の設計において重要な役割を担っており、心理学では知識について多角的に議論が行われていた。知識の属性や性質から知識を比較する研究や、学習者の知識の獲得や活用の仕組みを説明する研究、また、それらを基に効率的な知識獲得の方法を開発する研究、など幅広い。

知識の属性や性質という点では、一般知識と領域知識、宣言的知識と手続き的知識、具体的知識と抽象的知識、暗黙知、メタ知識など、さまざまな属性や性質、あるいはその対比を切り口に議論が行われてきたが、このなかでも、宣言的知識と手続き的知識を対比し知識を議論する研究は多い。宣言的知識は内容や概念を表現する知識であり、*knowing that* という言葉でも表現され、対象(A)がどう(B)であるかに関する知識である。例えば、「東京は日本の首都である」という知識が該当し、この場合、A が東京であり、B が日本の首都である。一方で、手続き的知識は、手続きを表現する知識であり、*knowing how* という言葉でも表現され、タスクをどう達成するかに関する知識である。特に、数学やプログラミングを対象に議論されることが多い。例えば、「1 次方程式 $ax + b = 0$ を解く」ための知識が該当し、この場合、 b を移行し a で両辺を割るという手続きにより問題を解くことができる。知識を宣言的知識と手続き的知識に分けて議論することで、人間の知識の獲得や活用を内容と内容を基にした手続きでよく記述できる。

特に、学習者の知識獲得の点でも、宣言的知識と手続き的知識を対比的に扱い説明する研究は多い。宣言的知識は、学習者が保有する知識に結合する形で獲得されるため、獲得対象の知識に既に獲得しているものが含まれている方が獲得されやすいといわれている。例えば、「東京は日本の首都である」という知識であれば、東京、首都、日本のうち、より多くについて既知であるほど獲得しやすく、逆に、「地球は丸い」という知識が獲得していたとしても必ずしも獲得しやすいわけではない、ということである。一方で、手続き的知識は、タスクが複数の細かいサブタスクに分解されるため、学習者が分解後の全てのサブタスクを達成するための知識を保有し

ている場合に獲得されるといわれている。「1 次方程式 $ax + b = 0$ を解く」ための知識は、「項を移行する」ための知識と「両辺を定数で割る」ための知識に分解され、さらに、「ある定数を引く」ための知識や「ある定数で割る」ための知識に分解され、というように階層的に分解された知識を全て保有している場合に、はじめて「1 次方程式 $ax + b = 0$ を解く」ための知識を獲得しているといえる、ということである。宣言的知識と手続き的知識はその獲得のされやすさが学習者が既に獲得している知識に依るところは共通するが、宣言的知識は関連の強い知識間で局所的にその獲得に影響を与え合い、一方で、手続き的知識は分解された知識がその掛け合わせで構成される知識の獲得に影響を与えるため、知識獲得における知識構造は、手続き的知識のものは宣言的知識のものと比べると、階層的になると考えられていた。

心理学の領域で、この考えを支持する研究報告は少ない。領域知識をよく分析し階層的に知識間関係を構築し、階層構造上より水準の高い知識に着手する前に、予め獲得すべき知識が確実に獲得されるように学習体験を構造化することで、習熟学習を効率化できるという仮説に基づいて、数学やプログラミングの習熟の効率化を狙った研究が報告された。こうした研究は、当初の期待よりは効果が低かったものの、よく分析し階層的に構造化した知識間関係に基づいて学習を設計することで学習を効率化できることを実験で示し、知識獲得における対象科目の知識構造が階層的であることを示唆するというものであった。

しかし、実験より示唆する研究報告はあるものの、宣言的知識と手続き的知識の対比において、知識獲得における知識構造を定量的に比較し分析した研究はまだない。これは、こうした心理学の研究が行われていたのが 20 年以上も前で、当時は、定量評価を実施できる学習者の知識獲得に関するデータの収集が困難であり、知識獲得に関するデータから知識獲得における知識構造を抽出することも難しく、また、抽出した知識構造を分析する方法もなかったからだと推察される。

さて、こうした知識間関係のように対象間の関係を明らかにしようとする分析として、ネットワーク分析というものがある。ネットワーク分析はウェブやそれソー

シャルメディアの普及に伴い注目が高まり、ウェブページ間の関係や人と人の関係を分析するためにしばしば用いられている。ネットワーク分析は対象をノード、対象間の関係をエッジとするグラフを構築し、その構造から関係を評価しようというものである。ネットワークの構造を捉えるための指標や手法が開発されており、例えば、ネットワークのモジュール性を評価するモジュラリティ [5] という指標や階層性を評価するフロー階層 [3] や GRC [4] という指標等、多くの指標がある。

また、学習や教育に関する領域では、近年、Massive Open Online Courses (以下、MOOCs) が注目を集めている。MOOCs は、従来の教室で時間割通りに一斉授業形式で提供される学習機会とは異なり、数多くの多様な講座のオンライン上での提供を通じて、いつでもどこでも自分のペースでさまざまな講座から自分の学習したいものを選択し学習できるというこれまでにない学習機会を提供しており、多くの学習者が利用している。

MOOCs は単にこれまでにない学習機会を提供しているというだけでなく、これまで難しかった大規模な学習効果分析の可能性も高めている。学習者はオンライン上で提供された講義動画や演習問題を通して学習するが、オンライン上で実施されているため学習行動ログをデータとして蓄積でき、さらに、そのデータを分析に活用できる。また、多くの多様な学習者が利用するため、多様な学習者の大規模な学習行動ログから多様な講座の学習効果の分析が可能となりつつある。特に、演習問題の回答ログはその演習問題により評価される知識を学習者が獲得しているか否かを表現しているため、知識獲得の分析に利用できる。

また、ここ数年、多くの研究領域で深層学習が注目されている。深層学習は多層のニューラルネットワークによる機械学習のことで、従来の機械学習では難しかった対象データの表現抽出を最適化の過程で行うことができる。深層学習の活用により画像認識、機械翻訳、質問応答文生成、音声認識等さまざまな研究領域で飛躍的な進展が報告がされている。

特に、学習者の知識獲得を予測する研究も、深層学習により大きく進展している。2015 年に、数学における知識獲得の予測に深層学習を活用し、高い精度で知識獲得を予測できること、予測モデルを分析することで知識間関係をネットワークとして抽出できることが報告された [6]。学習者が獲得している知識から、ある知識の獲得されやすさをそのまま予測しており、得られた知識間関係から抽出したネットワークは知識獲得における知識

構造を表現しているといえる。

これまで、心理学で議論されていたものの定量的な分析が困難だった宣言的知識と手続き的知識の獲得における知識構造の違いは、MOOCs の登場によりさまざまな科目について問題回答ログがデータとして蓄積されるようになり、その蓄積された学習者の知識獲得の軌跡から深層学習により知識間関係を抽出できるようになり、さらに、ネットワーク分析の発展により抽出した知識構造を分析できるようになった今、定量的に分析できる可能性が高まっている。

本論文の目的は、心理学で議論されていた宣言的知識と手続き的知識の獲得における知識構造の違いを定量的に分析することとで下記の 2 つを検証することである。

- 知識獲得における宣言的知識の知識構造は手続き的知識の知識構造と比べてよりモジュール性が高い。
- 逆に、知識獲得における手続き的知識の知識構造は宣言的知識の知識構造と比べてより階層性が高い。

従来心理学で議論され、実際の教育や学習にも活用されているものの、定量的に検証されていないこれら知識構造について明らかにすることは、学術的な意義が大きいと考える。

2 先行研究

2.1 Massive Open Online Courses

Massive Open Online Courses (以下、MOOCs) はオンライン上で誰もが受講できる大規模な講座群のことで、参加人数が非常に大規模であるという点や、高等教育水準の内容の講座だけでなく初等中等教育水準の内容の講座も含まれるという点でオープンコースウェアやこれまでのオンライン講座とは異なる。従来の教室で時間割通りに一斉授業形式で提供される学習機会とは異なり、いつでもどこでも自分のペースでさまざまな講座から自分の学習したいものを選択し学習できるというこれまでにない学習機会を提供している。

さらに、MOOCs にはこれまで難しかった大規模な学習効果分析の可能性を高めるという側面もある。特に、演習問題の回答ログはその演習問題により評価される知識を学習者が獲得しているか否かを表現しているため、知識獲得の分析に利用できる。

2.2 深層学習

知識獲得の予測に深層学習を用いた手法 [6] に用いられていたニューラルネットワークである Recurrent Neural Networks (以下、RNN) について説明する。伝

統的な RNN の構造は，入力層，隠れ層，出力層の 3 層から構成されている．系列方向を時刻とすれば，時刻 t の隠れ層 \mathbf{h}_t の計算に時刻 $t-1$ の隠れ層の情報を入力する $\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1})$ の式のように，一つ前の情報を繰り返し入力するという構造である．

RNN には異なる活性化関数を利用するという形でいくつかの種類がある．特に，本研究で用いる Gated Recurrent Neural Networks (以下，GRNN) [1] について説明する．GRNN は Gated Recurrent Unit [1] というゲート付き活性化関数を用いる RNN のことで，GRU は，長期的な表現と短期的な表現を捉えるために開発された．GRNN は下記の式により定義される． \mathbf{x}_t は入力ベクトル， $\mathbf{W}_{xr}, \mathbf{W}_{hr}, \mathbf{W}_{xz}, \mathbf{W}_{hz}, \mathbf{W}_{xh}, \mathbf{W}_{hh}$ は重み行列， $\mathbf{b}_r, \mathbf{b}_z, \mathbf{b}_h$ はバイアス項である．

$$\mathbf{r}_t = \sigma(\mathbf{W}_{xr}\mathbf{x}_t + \mathbf{W}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_r) \quad (1)$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{xz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z) \quad (2)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}(\mathbf{r}_t \odot \mathbf{h}_{t-1} + \mathbf{b}_h)) \quad (3)$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t \quad (4)$$

$$\mathbf{y}_t = \sigma(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y) \quad (5)$$

2.3 知識獲得の予測

知識獲得の予測は，学習者が対象の知識を獲得しているか否かを予測するというものである．通常，知識を獲得しているか否かは問題回答の正誤を基に評価されるため，知識獲得の予測のタスクは過去の学習者の問題回答履歴から次に解く問題の正誤を予測するというものである．最初の定式化の事例は，1994 年に報告された Knowledge Tracing [2] である．Knowledge Tracing は，学習者の時刻 t において観測された問題回答結果を q_t とすれば， q_1, q_2, \dots, q_t から時刻 $t+1$ において観測される問題回答結果 q_{t+1} を予測するタスクである．特に，回答の正誤確率を算出する場合は， q_1, q_2, \dots, q_t が観測された場合の時刻 $t+1$ に着手する問題において当該学習者の回答が正解となる事後確率 $p(q_{t+1} = \text{correct} | q_1, q_2, \dots, q_t)$ を求めるタスクである．

Bayesian Knowledge Tracing [2] (以下，BKT) はベイズの定理の事前確率と事後確率の関係に基づいて正解確率をモデル化する手法である．スキルごとに 4 つの確率変数が定義されており，これらを組み合わせてスキルの習熟をモデル化している．習熟，つまり時系列性を捉える手法であり，知識間関係は考慮しない．

Deep Knowledge Tracing [6] (以下，DKT) は RNN を利用し Knowledge Tracing を行う手法である．2015 年 6 月に発表され，高い性能で知識獲得を予測できる

表 1: Deep Knowledge Tracing における回答ログデータと対応する入力ベクトルの例

回答ログ			入力ベクトル	
ログの順番	問題番号	正誤	変数名	値
1	1	0	\mathbf{x}_1	[0000 1000]
2	1	1	\mathbf{x}_2	[1000 0000]
3	2	1	\mathbf{x}_3	[0100 0000]
4	3	0	\mathbf{x}_4	[0000 0010]
5	3	1	\mathbf{x}_5	[0010 0000]
6	4	1	\mathbf{x}_6	[0001 0000]

こと，予測モデルを分析することで知識間関係をネットワークとして抽出できることが報告された．

DKT の構造は伝統的な RNN の構造に基づいている．

$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}) \quad (6)$$

$$\mathbf{y}_t = g(\mathbf{h}_t) \quad (7)$$

モデルは関数 f と g によって定義されており，本研究では，関数 f, g に GRNN の式 1-5 を利用する．

DKT の入力は，学習者の学習行動の観測結果を one-hot ベクトルに符号化したもの \mathbf{x}_t をを用いる．観測結果は演習問題と正誤の組み合わせで表現できるため，演習問題の数を M とすれば， \mathbf{x}_t の長さは $2M$ となる．例えば，回答ログと対応する入力ベクトルは表 1 のようになる．出力 \mathbf{y}_t は問題数と同じ長さのベクトルで，それぞれの要素が当該学習者がそれぞれの問題に正しく回答する確率の予測値となっている．モデルは誤差関数を負の対数尤度とし，確率的勾配降下法で学習させる．

DKT による知識間関係抽出法について述べる．問題 i と j のすべての有向ペアのうち，問題 i が出現した後に残りの問題系列の中で問題 j が出現する系列数が問題 i が出現する問題系列数全体の 1% 以上であるものに対して，影響度 $J_{ij} = \frac{y(j|i)}{\sum_k y(j|k)}$ を割り当てる．ここでは， $y(j|i)$ は，ある学習者が最初に問題 i に正答した場合に，RNN によって割り当てられる次の時刻に問題 j に正答する確率である．ネットワークは，影響度が 0.1 以上であればエッジを引くというようにして構築された．得られたネットワークは，単に学習者の問題 (i, j) 間の遷移率から構築したネットワークや問題 i の正解が観測された後に問題 j の正解が観測される条件付き確率から構築したネットワークよりよく知識間関係を捉えていることが指摘されている．こうして得られた行列 J は，問題 i で評価される知識が既に獲得されている場合に，問題 j で評価される知識の獲得されやすさを表現しており， J は知識間関係行列であるといえ，この知識間関係行列から構築したネットワークは知識獲得における知識構造を表現していると考えられる．

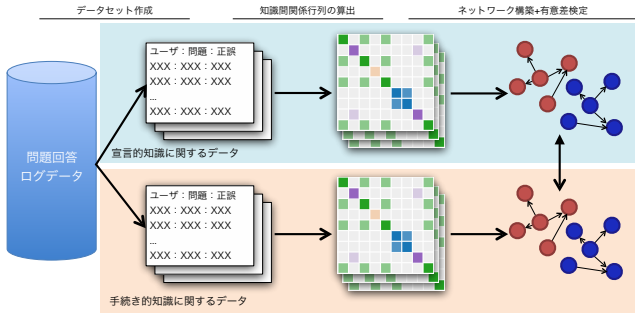


図 1: 分析手法全体の流れ

2.4 ネットワーク分析

ネットワーク構造を評価する指標について述べる。モジュラリティ [5] はネットワークのクラスタへの分割におけるモジュール性を評価するという指標で式 8 で与えられる。ここでは、 e_{ii} はクラスタの隣接行列の (i, i) 成分を、 a_i は隣接行列の i 行の和を示す。この指標はクラスタ内のエッジ割合とクラスタ間のエッジ割合の差を表現した指標だといえる。

$$Modularity = \sum_i (e_{ii} - a_i^2) \quad (8)$$

フロー階層 [3] は式 9 で与えられる。ここでは、 L はネットワークに含まれるエッジの数を、 e_i はエッジ i が環状構造に含まれていたら $e_i = 0$ でそれ以外は $e_i = 1$ となるような指標を示す。この指標は階層性を、環状構造がない、という側面から捉えようとした指標だといえる。

$$FlowHierarchy = \frac{\sum_{i=1}^L e_i}{L} \quad (9)$$

GRC [4] は Global Reaching Centrality のことで、式 10 で与えられる。ここでは、 $C_R(i)$ が Local Reaching Centrality という指標を、 V はネットワークに含まれるノードの集合を、 C_R^{max} は V に含まれる全ノード i について $C_R(i)$ の最大値を、 N は V の大きさであるノード数を示す。 $C_R(i)$ はノード i から出力されるエッジを辿って、ノード i から到達できるノードの割合を指す。この指標は階層性を一部のノードが他のノードより到達できるノードが多い構造にある、という側面から捉えようとして指標だと言える。

$$GRC = \frac{\sum_{i \in V} [C_R^{max} - C_R(i)]}{N - 1} \quad (10)$$

3 分析手法

分析手法全体の流れを図 1 に示す。手法はデータセットの作成、知識間関係行列の算出、ネットワーク構築及び有意差検定の 3 つのブロックから構成される。

まず、問題回答ログデータからデータセットを作成するわけだが、データセットは比較検証に用いるため、知識獲得の予測に利用できる複数のデータセットを作成する。また、データは問題回答ログであれば何でも良いというわけではなく、いくつかの要件を満たすものを利用する。具体的には、1) 比較検証できる複数のデータセットを作成できること、2) 各データセットが大規模であること、さらに、3) 初等中等教育水準の数学、地理、歴史に関する問題回答ログであること、の 3 つである。理由は紙面の都合で省く。

次に、作成した複数のデータセットに DKT を適用し知識間関係行列を算出する。

最後に、行列として定義された知識間関係からネットワークを構築し、ネットワークのモジュール性や階層性を表現する指標について t 検定を実施し、宣言的知識の知識構造と手続き的知識の知識構造の階層性やモジュール性に有意に差があることを検証する。ここでは、モジュール性を評価する指標としてモジュラリティを用い、階層性を評価する指標としてフロー階層および GRC を用いる。ネットワークの構築は、問題 j への影響度の大きい問題群 $i \neq j$ を J_{ij} が大きいものから 3 個選択し、 i から j にエッジを引く、という方法で行う。

4 データセット

データセットは、小学 4 年生から中学 3 年生を対象とする国内最大級の MOOCs「勉強サブリ*1」で提供される 11 の講座の問題回答ログデータから作成する*2。勉強サブリから収集されたログデータは先のデータ要件を満たした本分析に最適なデータである。実験で用いる宣言的知識の獲得を主目的とするデータセットと手続き的知識の獲得を主目的とするデータセットの作成について、特に、歴史や地理（いわゆる暗記系科目）に関する講座の問題が主に宣言的知識の獲得の有無を評価する問題であり、歴史や地理に関する 5 講座（小学 4 年社会、小学 5 年社会、小学 6 年社会、中学地理、中学歴史）から宣言的知識の獲得を主目的とするデータセットとして

*1 <https://benkyosapuri.jp/>

*2 本研究は勉強サブリを運営するリクルートマーケティング（株）との共同研究プロジェクトの一環で行われている。

表 2: 各データセットに対する各手法の予測性能とそれらの関係性

データセット			AUC		DKT - BKT	
分類	科目	学年	BKT	DKT	値	平均
宣言的知識の獲得を主目的とするデータセット	地理・社会	小 4	0.739	0.791	0.052	
		小 5	0.680	0.765	0.085	
		中学	0.681	0.764	0.083	
	歴史・社会	小 6	0.657	0.773	0.116	
		歴史	0.670	0.766	0.096	
		小 4	0.707	0.828	0.121	
手続き的知識の獲得を主目的とするデータセット	算数・数学	小 5	0.724	0.804	0.081	
		小 6	0.749	0.836	0.087	
		中 1	0.750	0.807	0.057	
		中 2	0.696	0.773	0.077	
		中 3	0.735	0.804	0.069	

5 つのデータセットを作成する。また、算数や数学に関する講座の問題が主に手続き的知識の獲得の有無を評価する問題であり、算数や数学に関する 6 講座（小学 4 年算数、小学 5 年算数、小学 6 年算数、中学 1 年数学、中学 2 年数学、中学 3 年数学）から手続き的知識の獲得を主目的とするデータセットとして 6 つのデータセットを作成する。

データセットの作成について述べる。対象期間は 2015 年 4 月から 2015 年 11 月の 8 ヶ月である。1 つの問題に複数の回答欄が存在する場合は当該問題のすべての回答欄が正解の場合に当該問題を正解した、と扱う。同時に複数の問題が提示される場合もあり、その場合、採点は提示された問題群に対して同時に行われる。データセットを作成する際に、前処理として回答行動ログデータから、「同時に回答された複数の問題のログデータ群について、すべての問題の回答欄が空白で投稿されているログデータ」をノイズとみなして除去した。

11 データセットを概観する。ユーザ数は、全体として数千以上が含まれており、問題数は、全体として、数十から数百が含まれており、回答ログ数は、全体として、数十万以上である。ユーザ 1 人あたりの平均回答ログ数は、全体として、数十から百数十となっており、問題 1 問あたりの平均回答ログ数は、全体としては、数百から数千程度の値となっている。いずれのデータセットも非常に大規模な問題回答ログから構成されており、問題あたりのログ数という点でも十分を大きいと考えられる。

5 実験

まず、実験設定について述べる。データセットへの適合のための DKT を拡張するが、紙面の都合で省略する。また、DKT による知識獲得の予測の実験設定について、11 のデータセットそれぞれについて、訓練：検証：テスト = 8 : 1 : 1 となるようにユーザを分けて実験を行う。

表 3: 各ネットワークにおける構造指標

ネットワーク			モジュラリティ		フロー階層		GRC	
分類	科目	学年	値	平均	値	平均	値	平均
宣言的知識の知識間関係ネットワーク	地理・社会	小 4	0.698		0.579		0.787	
		小 5	0.801		0.541		0.679	
		中学	0.759	0.759	0.633	0.627	0.603	0.690
	歴史・社会	小 6	0.808		0.607		0.613	
		中学	0.728		0.774		0.768	
		小 4	0.660		0.590		0.683	
手続き的知識の知識間関係ネットワーク	算数・数学	小 5	0.602		0.757		0.901	
		小 6	0.688		0.705		0.875	
		中 1	0.566	0.602	0.805	0.756	0.803	0.816
		中 2	0.559		0.830		0.818	
		中 3	0.538		0.847		0.861	

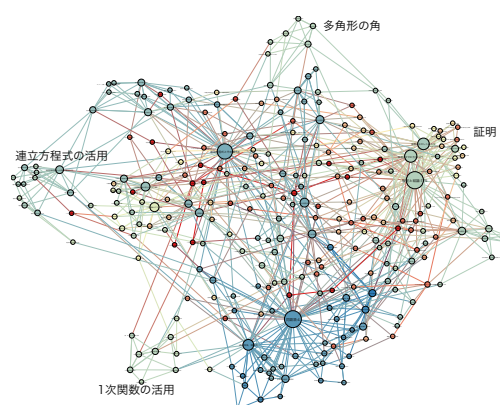
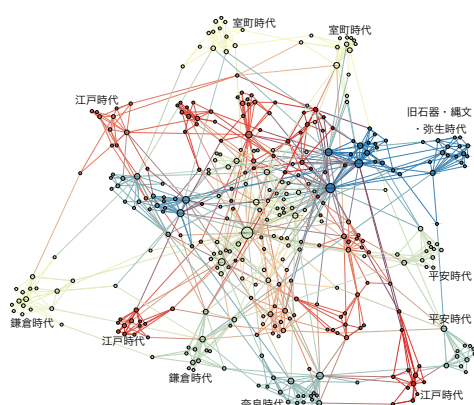
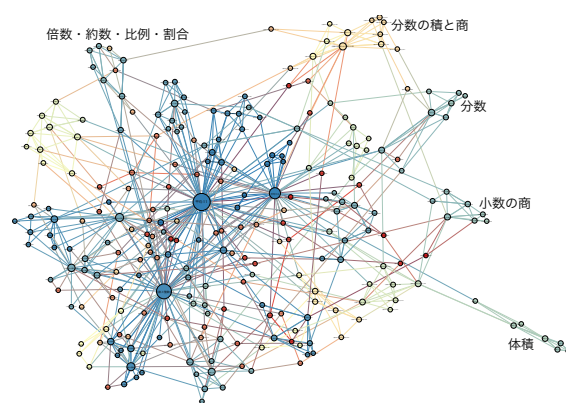
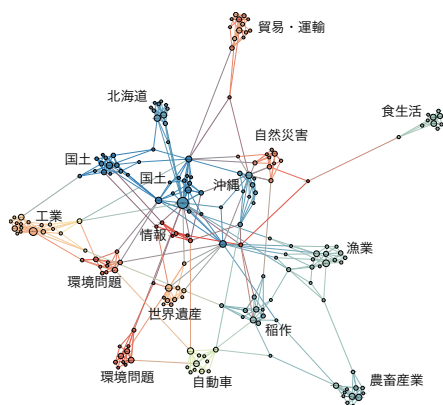
次に、実験結果について述べる。DKT と BKT を 11 データセットに適用した結果を表 2 に示す。いずれのデータセットにおいても、DKT の予測性能が知識間関係を考慮しない BKT よりも大きく高く、DKT により知識獲得における知識間関係を抽出できていることを定量的に確認した。

小学 5 年社会、小学 5 年算数、中学歴史、中学 2 年数学のデータセットについて、構築したネットワークをそれぞれ図 2, 3, 4, 5 に可視化した。ノードの内容から知識間関係を抽出できていることが分かり、ネットワーク全体におけるノード集合の配置やその関係を内容の側面から分析し、知識獲得における知識間関係を抽出できていることを定性的に確認した。

各ネットワークのモジュラリティとフロー階層、GRC を表 3 に示す。宣言的知識の知識間関係ネットワークと手続き的知識の知識間関係ネットワークにおけるモジュラリティ、フロー階層、GRC それぞれについて、t 検定を実施し統計的に有意な差があるかを検証した。モジュラリティにおける p 値は 0.00106 で、フロー階層における p 値は 0.0482 で、GRC における p 値は 0.0470 であった。したがって、いずれの指標においても有意水準 0.05 で有意な差が認められた。このことは知識獲得における知識構造について、宣言的知識のモジュール性が手続き的知識のモジュール性より統計的に有意に高く、逆に、手続き的知識の階層性が宣言的知識の階層性より統計的に有意に高いことを示している。

6 考察

実験では、小学 4 年算数のデータセットから構築したネットワークは、モジュール性が大きく階層性が小さかった。このことは、小学 4 年算数が手続き的知識の獲得を主目的としていないということ意味するのではないと考える。小学 4 年算数のネットワークでは複数のクラ



スタが存在していた．これらのクラスタ内ではある程度の階層性が存在するが，一方で，クラスタ間については，その内容が強く関連しているというわけではなく，その結果，ネットワーク全体としてはモジュール性が高く，階層性が低くなってしまっているのだと推察する．

7 結論

本論文では、心理学で議論されていた宣言的知識と手続き的知識の獲得における知識構造の違いを定量的に分析し、知識獲得における宣言的知識の知識構造は手続き的知識の知識構造と比べてよりモジュール性が高く、逆に、手続き的知識の知識構造は宣言的知識の知識構造と比べてより階層性が高いことを検証した。

本研究は MOOCs の登場，ネットワーク分析の発展，深層学習の躍進等，ここ数年の幅広い領域のさまざまな成果によって，初めてその実施が可能になったものである．本研究が，人間の学習や知能の解明の一助になると信じている．

参考文献

- [1] CHO, K., VAN MERRIËNBOER, B., GULCEHRE, C., BAHDANAU, D., BOUGARES, F., SCHWENK,

H., AND BENGIO, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

- [2] CORBETT, A. T., AND ANDERSON, J. R. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4, 4 (1994), 253–278.
- [3] LUO, J., AND MAGEE, C. L. Detecting evolving patterns of self-organizing networks by flow hierarchy measurement. *Complexity* 16, 6 (2011), 53–61.
- [4] MONES, E., VICSEK, L., AND VICSEK, T. Hierarchy measure for complex networks. *PloS one* 7, 3 (2012), e33799.
- [5] NEWMAN, M. E. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 23 (2006), 8577–8582.
- [6] PIECH, C., BASSEN, J., HUANG, J., GANGULI, S., SAHAMI, M., GUIBAS, L. J., AND SOHL-DICKSTEIN, J. Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (2015), pp. 505–513.