

平成 27 年度  
修士論文

深層学習を用いた  
MOOCs の学習者の知識構造の分析

平成 28 年 3 月  
指導教員 松尾豊 特任准教授  
東京大学大学院工学系研究科技術経営戦略学専攻  
37-146856 那須野薫

# 概要

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が多い。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が多い。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、この考え方を支持する研究報告は少なくない。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という指標や階層性を評価するフロー階層やGRC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，特に学習や教育に関する領域に限らず，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．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する。

まず、データセットを小学生と中学生を対象とした国内最大級の MOOCs である「勉強サプリ」における問題回答ログより作成する。特に、地理と歴史に関する 5 つの講座の問題回答ログから作成した 5 つのデータセットを宣言的知識の獲得を主目的とするデータセットとし、算数や数学に関する 6 つの講座の問題回答ログから作成した 6 つのデータセットを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、合計 11 のデータセットに対して分析を行う。次に、各データセットに深層学習を活用した知識獲得の予測手法を適用し、それぞれ知識間関係をネットワークとして抽出する。そして、抽出したネットワークのモジュール性をモジュラリティで評価し、階層性をフロー階層と GRC で評価する。

検証実験の結果、地理と歴史に関する講座から抽出したネットワークのモジュラリティは算数や数学に関する講座のものより有意水準 0.05 で有意に高かった。逆に、算数や数学に関する講座から抽出したネットワークのフロー階層と GRC は地理と歴史に関する講座のものより有意水準 0.05 で有意に高かった。したがって、心理学で議論されていたように、知識獲得において宣言的知識の知識構造は手続き的知識のものと比べて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ものと比べて階層性が高いことが検証された。

検証結果を踏まえ、本研究で用いた知識構造の分析手法の他 MOOCs への適用可能性や、研究の拡張として適用対象や知識間関係抽出手法の拡張を考察した。

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。

# 目 次

<b>第 1 章 序論</b>	<b>1</b>
1.1 背景 . . . . .	1
1.2 研究目的 . . . . .	6
1.3 本論文の構成 . . . . .	6
<b>第 2 章 先行研究</b>	<b>9</b>
2.1 心理学における知識 . . . . .	9
2.1.1 知識の属性や性質 . . . . .	9
2.1.2 知識の獲得や活用 . . . . .	11
2.2 Massive Open Online Courses . . . . .	14
2.3 深層学習 . . . . .	17
2.4 知識獲得の予測 . . . . .	23
2.4.1 Knowledge Tracing の定式化 . . . . .	24
2.4.2 Bayesian Knowledge Tracing . . . . .	25
2.4.3 Performance Factor Analysis . . . . .	27
2.4.4 Deep Knowledge Tracing . . . . .	27
2.5 ネットワーク分析 . . . . .	30
<b>第 3 章 分析手法</b>	<b>34</b>
3.1 知識獲得における知識間関係の抽出 . . . . .	36
3.2 データの要件 . . . . .	38
3.3 知識間関係ネットワークの構造評価 . . . . .	40
<b>第 4 章 データセット</b>	<b>43</b>
4.1 勉強サプリ . . . . .	43
4.2 データセットの作成 . . . . .	45

4.3 データセットの概観	46
4.4 個々のデータセットの具体的説明	48
<b>第5章 実験</b>	<b>58</b>
5.1 Deep Knowledge Tracing の拡張および最適化	58
5.2 設定	61
5.3 結果	62
5.3.1 各データセットにおける予測性能	62
5.3.2 抽出した知識間関係ネットワークの可視化	63
5.3.3 ネットワーク構造の分析	72
<b>第6章 考察</b>	<b>75</b>
6.1 データセットと知識構造	75
6.2 知識構造分析手法の他 MOOCs への適用可能性	76
6.3 今後の展望	77
6.3.1 対象データの多様化，結合，長期化	77
6.3.2 知識間関係抽出手法の拡張	78
<b>第7章 結論</b>	<b>80</b>
<b>参考文献</b>	<b>81</b>
<b>謝辞</b>	<b>91</b>
<b>発表文献</b>	<b>94</b>

# 図 目 次

2.1 ACT*理論の構成	12
2.2 Coursera のイメージ	15
2.3 KhanAcademy のイメージ	15
2.4 RNN の構造のイメージ	19
2.5 フロー階層の適用事例	33
3.1 分析手法全体の流れ	35
4.1 勉強サブリのマイページ	44
4.2 勉強サブリの講義一覧ページ	44
4.3 勉強サブリの講義視聴ページ	44
4.4 勉強サブリの問題演習ページ	44
4.5 小学 4 年社会の問題と回答選択肢の例	49
4.6 小学 4 年社会の平均着手順と回答ログ数の XY プロット	49
4.7 小学 4 年算数の問題と回答選択肢の例	49
4.8 小学 4 年算数の平均着手順と回答ログ数の XY プロット	49
4.9 小学 5 年社会の問題と回答選択肢の例	50
4.10 小学 5 年社会の平均着手順と回答ログ数の XY プロット	50
4.11 小学 5 年算数の問題と回答選択肢の例	51
4.12 小学 5 年算数の平均着手順と回答ログ数の XY プロット	51
4.13 小学 6 年社会の問題と回答選択肢の例	51
4.14 小学 6 年社会の平均着手順と回答ログ数の XY プロット	51
4.15 小学 6 年算数の問題と回答選択肢の例	52
4.16 小学 6 年算数の平均着手順と回答ログ数の XY プロット	52
4.17 中学 1 年数学の問題と回答選択肢の例	53

4.18 中学 1 年数学の平均着手順と回答ログ数の XY プロット . . . . .	53
4.19 中学 2 年数学の問題と回答選択肢の例 . . . . .	54
4.20 中学 2 年数学の平均着手順と回答ログ数の XY プロット . . . . .	54
4.21 中学 3 年数学の問題と回答選択肢の例 . . . . .	54
4.22 中学 3 年数学の平均着手順と回答ログ数の XY プロット . . . . .	54
4.23 中学地理の問題と回答選択肢の例 . . . . .	55
4.24 中学地理の平均着手順と回答ログ数の XY プロット . . . . .	55
4.25 中学歴史の問題と回答選択肢の例 . . . . .	56
4.26 中学歴史の平均着手順と回答ログ数の XY プロット . . . . .	56
5.1 小学 4 年社会の知識間関係ネットワーク . . . . .	64
5.2 小学 5 年社会の知識間関係ネットワーク . . . . .	65
5.3 小学 6 年社会の知識間関係ネットワーク . . . . .	65
5.4 中学地理の知識間関係ネットワーク . . . . .	66
5.5 中学歴史の知識間関係ネットワーク . . . . .	67
5.6 小学 4 年算数の知識間関係ネットワーク . . . . .	68
5.7 小学 5 年算数の知識間関係ネットワーク . . . . .	68
5.8 小学 6 年算数の知識間関係ネットワーク . . . . .	69
5.9 中学 1 年数学の知識間関係ネットワーク . . . . .	70
5.10 中学 2 年数学の知識間関係ネットワーク . . . . .	71
5.11 中学 3 年数学の知識間関係ネットワーク . . . . .	72

# 表 目 次

2.1 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	28
4.1 11 データセットの統計量	47
5.1 拡張した 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	59
5.2 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	63
5.3 各ネットワークにおける構造指標	72

# 第1章 序論

本章では、本論文の背景、研究目的および本論文の構成について述べる。

## 1.1 背景

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 [Ryle, 1945, De Jong and Ferguson-Hessler, 1996] や、学習者の知識の獲得や活用の仕組みを説明する研究 [Anderson, 1982, Anderson, 1993]、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 [Corbett and Anderson, 1994]、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、戦略的知識、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いた [De Jong and Ferguson-Hessler, 1996] が、このなかでも、宣言的知識と手続き的知識の 2 つを対比し知識を議論する研究が多い [Anderson, 1982, McCormick, 1997, Ten Berge and Van Hezewijk, 1999]。宣言的知識は内容や概念を表現する知識であり、*knowing that* [Ryle, 1945] 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* [Ryle, 1945] 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。

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 [Anderson, 1982, Anderson, 1990]。宣言的知識は連想による意味ネットワークの形式で保持され、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。

心理学の研究領域で、この考え方を支持する研究報告は多い。習熟学習に関する研究としては、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、予め獲得するべき知識が確実に獲得されるように学習体験を設計することで、ほとんど全ての学習者がスキルを十分に習熟できるという考え方を報告するもの [Keller, 1968, Bloom, 1968] もある。また、その報告を受けて実験を行い、当初の期待よりは効果は小さかったものの、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と報告する研究も少なくない [Block et al., 1971, Cohen and Hyman, 1979, Resnick, 1977]。[Anderson, 1982, Anderson, 1990] では、ACT\*理論 (Adaptive Character of Thought Theory) という認知構造に関する理論が提案され、手続き的知識の獲

得を，解釈を必要とする宣言的知識の適用から生成規則の適用への置き換えの過程（手続き化）として表現し，より複雑な手続きの獲得を生成規則の系列の結合として位置付けた。

また，こうした習熟学習の研究や ACT\*理論をプログラミングの学習支援システムに応用して知識獲得を予測し学習の効率化を狙った研究も報告されている。Corbett ら [Corbett and Anderson, 1994] は，知識獲得の予測のタスクとして Knowledge Tracing というタスクを提案した。Knowledge Tracing は過去の問題回答ログデータから次に解く問題の正誤を予測するというタスクである。次に解く問題の正誤を予測し，スキルを獲得しているか否かを推定する仕組みを学習支援システムに導入することで，階層構造上より水準の高いプログラミングスキルに着手する前に，予め獲得するべきプログラミングスキルが確実に獲得されるように学習体験を設計し，学習の効率化を狙った研究である。

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，その構造から個々の対象や関係あるいは全体を評価しようというものである。ネットワーク分析はウェブやそれを活用するソーシャルメディアの普及に伴い注目が高まり，ウェブページやユーザを分析するためにしばしば用いられている。応用事例として，ウェブページ検索エンジンの性能向上のためノードの重要度を評価するページランクという指標を開発する研究 [Page et al., 1999] やソーシャルメディアにおけるユーザ間の関係性から友人関係を推定し推薦への活用を試みる研究 [Dong et al., 2012] 等がある。

ネットワークを評価する指標という点でも多くのものが存在する。ネットワークにおける個々のノードやエッジの特徴を評価する指標として，中心度という点でノードを評価する次数中心性や近接中心性，経路における中心度という点でエッ

ジを評価する媒介中心性等がある。また、ネットワークの構造を捉えるための指標も開発されており、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という指標 [Newman, 2006] や階層性を評価するフロー階層という指標 [Luo and Magee, 2011] や GRC という指標 [Mones et al., 2012]、ネットワークの大きさを評価するネットワーク直径という指標等がある。

学習や教育に関する領域では、近年、Massive Open Online Courses(以下、MOOCs) [McAuley et al., 2010, Pappano, 2012, Siemens, 2013] が注目を集めている。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ある。講座を運営するプラットフォームサービスを指して MOOCs と呼ぶこともある。MOOCs はオンライン上でさまざまな講座を提供し、各講座ごとに講義動画や演習システム、掲示板等が提供されている。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、数多くの多様な講座のオンラインでの提供を通じて、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。例えば、有名な MOOCs の 1 つである Coursera<sup>1</sup> では 2016 年 1 月の時点で、コンピュータサイエンス、数学や論理、社会科学などに関する 1500 以上の講座を提供し、1700 万人以上が利用している<sup>2</sup>。

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することができ、さらに、そのデータを分析に活用することができる。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。例えば、MOOCs の演習問題の回答ログを利用して知識獲得の予測を行う研究 [MacHardy and Pardos, 2015] では、有名な MOOCs の 1 つである Khan Academy<sup>3</sup> から収集したデータを利用していたが、その問題回答ログ数は 100 万件

<sup>1</sup><https://www.coursera.org/>

<sup>2</sup> 講座数と利用者数はトップページの記載より引用。

<sup>3</sup><https://www.khanacademy.org/>

以上であり、大規模なデータを対象に分析がなされていた。

さて、ここ数年、特に学習や教育に関する領域に限らず、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 [Schroff et al., 2015, Szegedy et al., 2014]、機械翻訳 [Sutskever et al., 2014, Dong et al., 2015]、質問応答文生成 [Yin et al., 2015]、画像説明文生成 [Xu et al., 2015, Vinyals et al., 2014]、音声認識 [Hinton et al., 2012, Bahdanau et al., 2015] 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。画像認識では深層学習により例えば犬の表現として目や鼻、口の表現が抽出できると報告する研究 [Zeiler and Fergus, 2014] や、人間より高い精度で人の顔を見分けられたと報告する研究 [Schroff et al., 2015] もある。機械翻訳の領域では、可変長の英語を可変長のフランス語に翻訳する研究 [Sutskever et al., 2014] やフランス語とオランダ語、スペイン語の共通表現を抽出し英語から1つのモデルで翻訳する研究もある [Dong et al., 2015]。また、これらを掛け合わせた画像説明文生成の領域では、説明文の単語1つずつを生成する際に画像のどの部分に着目すべきかという表現を抽出し、説明文を生成する研究 [Xu et al., 2015] もある。このように、深層学習は多くの研究領域で大きな進展をもたらしている。

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た。Piechらは Knowledge Tracing に深層学習を活用する Deep Knowledge Tracing という手法を発表した [Piech et al., 2015]。数学における知識獲得の予測に時系列分析でよく用いられる深層学習モデルである Recurrent Neural Networks を活用し、高い性能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。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され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、主に心理学の研究領域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

構造の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と考えられる。

以上、本研究の背景について述べた。次に、上記の背景を踏まえた研究目的について説明する。

## 1.2 研究目的

本論文の目的は、主に心理学の研究領域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、主に心理学の研究領域で議論され、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。

以上、本研究の目的について述べた。次に、背景と目的を踏まえて本論文の構成について述べる。

## 1.3 本論文の構成

以降の本論文の構成について述べる。

2章では、先行研究について述べる。まず、心理学における知識について詳細に述べる。次に、MOOCsについて、比較事例を挙げ、種類や効果、問題点等について述べる。さらに、深層学習について概説したのうち、特に本論文と関わりが深いRecurrent Neural Networksについて詳細に述べる。そして、知識獲得の予測につ

いて伝統的な手法と最先端の手法について整理し、最後に、ネットワーク分析について、特に、モジュール性を捉える指標と階層性を捉える指標について述べる。

3章では、宣言的知識と手続き的知識の知識獲得における知識構造の違いを分析するための手法について説明する。まず、分析手法全体の流れを概説し、手法全体が3つのブロックから構成されることを述べる。その後、まず、既存の知識間関係の抽出手法の中で、Deep Knowledge Tracing が最適であることを説明する。次に、分析対象の問題回答ログデータについて要件があることを述べ、その要件として、比較検証できる複数のデータセットを作成できること、各データセットが大規模であること、初等中等教育水準の数学や地理、歴史に関する科目であること、について説明する。最後に、抽出した知識間関係ネットワークのモジュール性と階層性を評価するために、ネットワーク構造指標の中でもモジュラリティとフロー階層および GRC を用いること、また、これらの各指標について、宣言的知識によるネットワークと手続き的知識によるネットワーク間で有意差検定を実施し、構造の違いを検証することを述べる。

4章では、実験で用いるデータセットについて述べる。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 「勉強サプリ<sup>4</sup>」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。まず、勉強サプリから収集されたログデータが分析に利用するためのデータ要件を満たした本分析に最適なデータであることを述べる。次に、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述べる。特に、歴史や地理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ることを指摘し、歴史や地理に関する5講座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ることを指摘し、算数や数学に関する6講座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。

5章では、実験について述べる。まず、データセットへの適合のための Deep Knowledge Tracing の拡張法を述べたのち、次に、実験設定について述べ、最後に実験結果について述べる。実験結果においては、まず、11のデータセットのいずれ

<sup>4</sup><https://benkyosapuri.jp/>

のデータセットにおいても，Deep Knowledge Tracing の予測性能が知識間関係を考慮しない Bayesian Knowledge Tracing よりも高いことを示し，Deep Knowledge Tracing 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する．さらに，11 の知識間関係ネットワークを可視化し，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する．そして，宣言的知識の知識構造を表現する 5 つのネットワークと手続き的知識の知識構造を表現する 6 つのネットワークのモジュラリティとフロー階層，GRC を算出し，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す．

6 章では，実験結果を踏まえた考察を述べる．検証過程で得られた知見に基づいて，本研究で用いた知識構造の分析手法の他 MOOCs への適用可能性について議論し，講座の科目や問題へのタグの有無に依らず知識構造を分析できる可能性があること，および，特に，大学水準の難易度を扱う MOOCs については検証実験を行う必要があることを指摘する．また，本研究の拡張として，適用対象の拡張という点で対象データの多様化，結合，長期化の 3 つの拡張を述べる．また，知識間関係抽出法である Deep Knowledge Tracing の拡張という点で，カリキュラム学習による拡張，および，スキルタグ自動抽出による拡張の 2 つについて述べる．最後に，7 章で結論を述べる．

以上，序論について述べた．次に，先行研究について述べる．

# 第2章 先行研究

本章では、先行研究について述べる。まず、心理学における知識について詳細に述べる。次に、MOOCsについて、比較事例を挙げ、種類や効果、問題点等について述べる。さらに、深層学習について概説したのうち、特に本論文と関わりが深いRecurrent Neural Networksについて詳細に述べる。そして、知識獲得の予測について伝統的な手法と最先端の手法について整理し、最後に、ネットワーク分析について、特に、モジュール性を捉える指標と階層性を捉える指標について述べる。

## 2.1 心理学における知識

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なされていた。知識の属性や性質から知識を比較する研究 [Ryle, 1945, De Jong and Ferguson-Hessler, 1996] や、学習者の知識の獲得や活用の仕組みを説明する研究 [Anderson, 1982, Anderson, 1993]、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 [Corbett and Anderson, 1994]、など幅広い。それぞれ、説明していく。

### 2.1.1 知識の属性や性質

まず、知識の属性や性質から知識を比較する研究について説明する。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、戦略的知識、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されてきた。属性や性質は多様であるが、例えば、[De Jong and Ferguson-Hessler, 1996] では物理学を事例とし、これまで先行研究で議論してきたさまざまな名称の知識を質と種類の 2 つの軸を導入して

整理した。質の軸は深さ、構造性、自動性、様相、一般性の4軸を設け、種類の軸は、状況的、概念的、手続き的、戦略的の4軸を設けた。例えば、浅い手続き的知識であれば、ルールや操作が該当し、構造化された戦略的知識であれば、一貫した系列の行動が該当する、というように知識を  $32 = 2 \times 4 \times 4$  種類に分けて整理している。

このような知識を整理する研究の中でも、宣言的知識と手続き的知識の2つを対比し知識を議論する研究が多い [Anderson, 1982, McCormick, 1997, Ten Berge and Van Hezewijk, 1999]。宣言的知識は内容や概念を表現する知識であり, *knowing that* [Ryle, 1945] 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* [Ryle, 1945] 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。

また、宣言的知識と手続き的知識の対比は、しばしば類似した他の名前の知識の対比によって表現されることがある。特に、宣言的知識について、他に事実的知識と概念的知識等の複数類似した内容をさす用語が存在して紛らわしい。そこで、それぞれ違いをここで整理しておく。宣言的知識が概念や命題、系列を表現し、意味ネットワークという形式を取る [Anderson, 1990] が、事実的知識は事実に関する知識であり、必ずしも系列や概念を含まない。この場合、宣言的知識は事実的知識を含む。この他にも、事実的知識は対象が事実であるかどうか判断する知識を指す場合もある [Unger, 1968]。概念的知識は特に数学や理科、プログラミングに関する内容を扱う場合によく使われる用語である [McCormick, 1997, Rittle-Johnson and Alibali, 1999, Hiebert, 2013]。数学や理科、プログラミングの対象は多くの場合、数や記号で表現される概念を対象としているからだと考えられる。宣言的知識との違いとして、概念的知識は知識間の関連に焦点を当てている点、概念的知識は活発なプロセスの一部であるという点が挙げられている [McCormick, 1997]。宣言的知識と手続き的知識の対比に加えて、構造化や抽象化を行う論理的知識の

特異性を主張し、3つを対比する研究もある [Cauley, 1986]。しかし、宣言的知識と手続き的知識を対比する研究 [Anderson, 1982, Anderson, 1990] では、手続き的知識に推論や構造化に関する知識が含まれているおり、議論が終わっていないような側面もある。本研究は、特に、ACT\*理論 [Anderson, 1982, Anderson, 1990] における宣言的知識と手続き的知識の定義を採用している。ACT\*理論の詳細は後述する。

### 2.1.2 知識の獲得や活用

次に、学習者の知識の獲得や活用の仕組みを説明する研究について説明する。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が多い [Anderson, 1982, Anderson, 1990]。宣言的知識は連想による意味ネットワークの形式で保持され、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獲得している知識との関係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、この考え方を支持する研究報告は多い。習熟学習に関する研究としては、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、予め獲得するべき知識が確実に獲得されるよう学習体験を設計することで、ほとんど全ての学習者がスキルを十分に習熟できるという考え方を報告するもの [Keller, 1968, Bloom, 1968] もある。また、その報告を受けて実験を行い、当初の期待よりは効果は小さかったものの、よく分析し階層的に構造化した知識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と報告する研究も少なくない [Block et al., 1971, Cohen and Hyman, 1979, Resnick, 1977]。

また、[Anderson, 1982, Anderson, 1990] では、ACT\*理論 ( Adaptive Character of Thought Theory ) という知識の獲得と活用を含む認知構造に関する理論が提案された。本研究では、特に、ACT\*理論における宣言的知識と手続き的知識の定義を採用している。そのため、ACT\*理論について詳細に述べる。まず、ACT\*理論の構成について述べる。ACT\*理論の構成を図 2.1 に示す。ACT\*理論には 3 種類

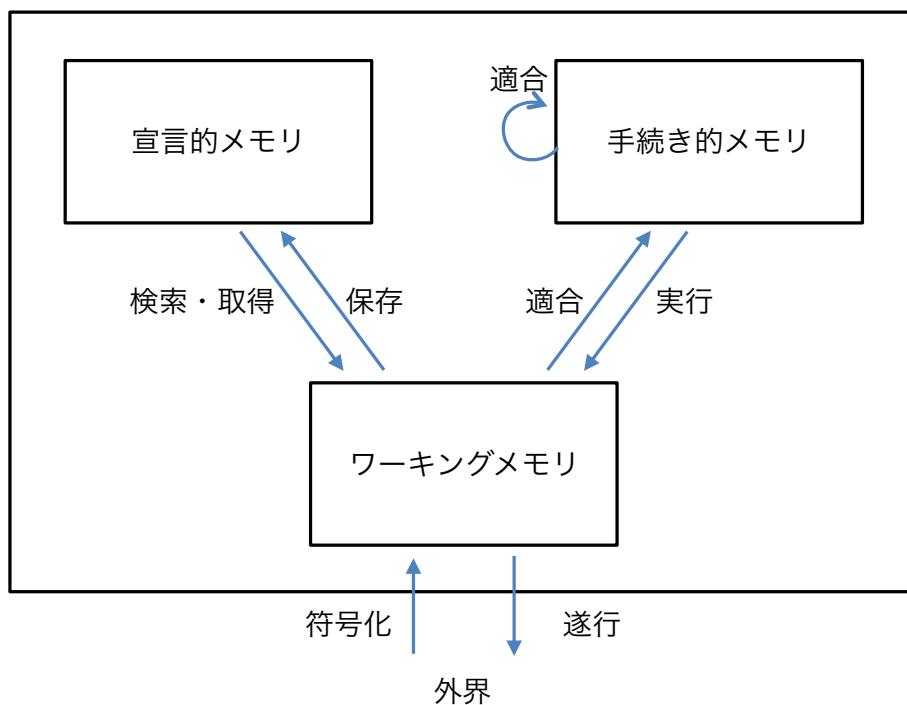


図 2.1: ACT\*理論の構成

のメモリがある。宣言的知識を格納する宣言的メモリ (Declarative Memory) と

手続き的知識を格納する手続き的メモリ (Procedural Memory), ワーキングメモリ (Working Memory) の3つのメモリである。宣言的メモリは連想により概念や命題、系列を結合する意味ネットワークの形式をとる。手続き的メモリは生成規則 (Production Rule) の形式で表現される。それぞれの生成規則は宣言的メモリに格納されている条件と行動のペア (IF-THEN-) をもつ。手続き的メモリはそれぞれで活性化の程度があり、ワーキングメモリが最もよく活性化されるメモリである。ACT\*理論には基本的な学習の仕組みが3つある。1つは「一般化」で、生成規則が適用される幅が広がるというものである。1つは「差別化」で、生成規則の適用される幅がせばまるというものである。1つは「強化」で、いくつかの生成規則が適用される頻度が向上し、また、既存の生成規則の結合あるいは分離により新しい生成規則が形成される、というものである。

次に、ACT\*理論における知識の獲得と活用について述べる。まず、全ての知識は宣言的な情報から始められる。手続き的知識は既に存在する宣言的知識から推論を行うという問題解決の過程で学習される。あるタスクが実行されると、解釈を必要とした宣言的知識の適用が徐々に、当該タスクを直接実行する生成規則によって置き換えられていく。この過程は手続き化と呼ばれる。例えば、初めて  $2+4$  や  $1+5$  等の数字の足し算を実行するときには、右手と左手の指を立てたり折ったりして、最終的に残った指の本数を数えるということがしばしば行われる [Gersten and Chard, 1999] が、繰り返し行うことによって、当該生成規則を認知し実行する生成規則に徐々に置き換えられていく。つまり、明示的な宣言的知識の適用が手続き的知識の直接的適用に置き換えられていくということである（習熟することもある）。生成規則の系列が結合して、ひとつの生成規則になることもある。また、知識が生成規則の形式に変換されると適用の早さや安定性が向上しワーキングメモリへの負担が減る。特に、新しい生成規則の獲得が既存の生成規則との掛け合わせである場合、その掛け合わせが知識獲得における手続き的知識の階層性を表現しているといえる。

また、こうした習熟学習の研究や ACT\*理論をプログラミングの学習支援システムに応用して知識獲得を予測し学習の効率化を狙った研究も報告されている。Corbettら [Corbett and Anderson, 1994] は、知識獲得の予測のタスクとして Knowledge Tracing というタスクを提案した。Knowledge Tracing は過去の問題回答ログ

データから次に解く問題の正誤を予測するというタスクである。次に解く問題の正誤を予測し、スキルを獲得しているか否かを推定する仕組みを学習支援システムに導入することで、階層構造上より水準の高いプログラミングスキルに着手する前に、予め獲得するべきプログラミングスキルが確実に獲得されるように学習体験を設計し、学習の効率化を狙った研究である。また、この研究の拡張は多いが、知識獲得の予測に関する詳細については後述する。

以上、心理学における知識について述べた。次に、MOOCsについて述べる。

## 2.2 Massive Open Online Courses

MOOCs は Massive Open Online Courses [McAuley et al., 2010, Pappano, 2012, Siemens, 2013] の略称で、特に日本語で表記する場合は大規模公開オンライン講座と記述することがある。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ある。講座を運営するプラットフォームサービスを指して MOOCs と呼ぶこともある。MOOCs という概念の登場は 2008 年に主に履修登録をした学生向けに講座をオンラインで開催したら、学生だけでなく 2000 人以上の人々がその講座に参加したことがきっかけだと言われている [Yuan et al., 2013]。以前から大学はオープンコースウェア [Abelson, 2008] という形で講義の動画や資料を公開していたが、MOOCs は、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とは異なる。また、これまでもオンライン講座というものは存在していたが、MOOCs は、参加人数が非常に大規模であるという点や公開している講座の数が大規模である点、また、その内容が多様であるという点、利用が無料、あるいは無料に近いという点で、これまでのオンライン講座とは異なる。

MOOCs ではオンライン上でさまざまな講座を提供され、また各講座ごとに講義動画や演習システム、掲示板等が提供されている。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、数多くの多様な講座のオンラインでの提供を通じて、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 は場所や時間に学習ペース , 内容に縛られないこれまでにない学習機会を提供しているという点で , 社会や産業への影響が期待されている . 例えば , 大学生だけでなく社会人も自分の専門領域に関する講座を受講することで理解を深めたり , あるいは , 専門領域とは大きく関わりのない幅広い講座を受講することで教養を高めたりすることができる . また , 特に , 公教育の整備がまだ追いついていない発展途上国においては MOOCs の影響は大きく , その影響や位置付け , 可能性を分析する報告は多い [Trucano et al., 2013, Liyanagunawardena et al., 2013] .



図 2.2: Coursera のイメージ

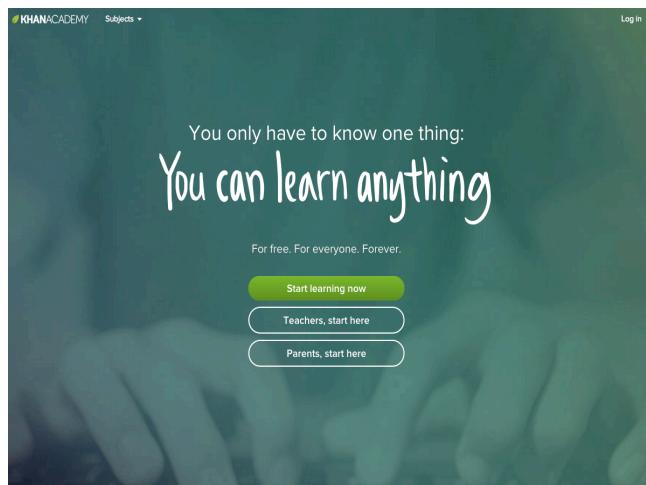


図 2.3: KhanAcademy のイメージ

MOOCs の有名な事例としては , Coursera<sup>1</sup> や Khan Academy<sup>2</sup> が挙げられる . MOOCs と Khan Academy のイメージを図 2.2 , 2.3 に示す . Coursera は主に大学の講座を提供する MOOCs で 2016 年 1 月の時点で , コンピュータサイエンス , 数学や論理 , 社会科学などに関する 1500 以上の講座を提供し , 世界中から 1700 万人以上が利用している<sup>3</sup> . 図 2.2 では , データサイエンス , ビッグデータ , 機械学習 , ビジネスアナリティクス等の講座が人気講座として表示されているが , このように提供されている講座は専門性が高いものも多い . Khan Academy は 2008 年にサービスを開始し , 主に初等中等教育水準の講座を提供する MOOCs で 2016 年 1 月の時点で , 数学 , 生物 , 芸術史 , 地理などに関する多くの講座を提供し , また ,

<sup>1</sup><https://www.coursera.org/>

<sup>2</sup><https://www.khanacademy.org/>

<sup>3</sup> 講座数と利用者数はトップページの記載より引用 .

演習問題については10万以上が提供されており<sup>4</sup>、世界中の学習者が利用している。初等中等教育水準の講座への提供ということもあり、図2.3にあるように保護者や先生と連携する仕組みもある。

MOOCsはそのプラットフォームや運営者に応じて内容や性質が異なり、それを体系化する研究も多い。[Yuan et al., 2013]では、MOOCsを4つの項目1) サービスが営利目的か否か、2) アクセスが有料か否か、3) 修了証が有料か否か、4) 単位取得できるか否か、で整理した。この整理では、例えば、Courseraは営利目的のサービスであり、アクセスは無料で、修了証の発行は有料で、単位取得は一部可能というように整理される。また、MOOCsの設計の背後にある教育学的立場の違いからしばしばxMOOCsとcMOOCsに分けて議論される[Daniel, 2012]。cMOOCsはconnectivist MOOCsの略称で学習者間のやり取りを通した学習を重視したMOOCsであり、xMOOCsはコンテンツベースの学習を重視したMOOCsである。

このように、MOOCsは、その登場より多くの学習者が利用しており期待も高かったが、MOOCsはその登場当初に期待されていたほど、教育や学習に効果があったというわけではない。特に、MOOCsにおける大きな課題の1つとして学習の継続性が非常に低いということが挙げられる。さまざまなMOOCsで提供されている講座の受講完了率を集計した報告[Jordan, 2013]によると、MOOCsの講座完了率は15%程度であるという。また、[Rivard, 2013]によると、Courseraで提供されたデューク大学の生体電気に関する講座の履修状況は当該講座に履修登録した12,700人の受講者の中で、最終テストを受けた受講者は350人と全体の3%程度だったという。これまでにない学習の機会を提供しているという点で非常に期待されていたが、継続的に勉強しなければ継続的な学習効果は期待できず、したがって、学習の継続性が低いということはMOOCsにおける大きな課題であるといえる。

また、初等中等教育への影響に懐疑的な意見もある。学習者のやる気や多様性の観点から考察して報告した研究[Bock and O'DEA, 2013]では、オンライン学習はやる気の強い学習者にとってはよく機能するがそうでない学習者は辞めてしまうことや、異なる教育背景を持つ多様な学習者に対応できるだけの柔軟性がま

<sup>4</sup>トップページの記載より引用。

だ MOOCs にないが、特に初等中等教育水準の学習ではそれが重要であることが指摘された。

MOOCs はこうした学習者に学習の機会を提供するという側面だけでなく、これまで難しかった大規模な学習効果分析の可能性を高めるという側面もある。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することができ、さらに、そのデータを分析に活用することができる。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。例えば、MOOCs の演習問題の回答ログを利用して知識獲得の予測を行う研究 [MacHardy and Pardos, 2015] では、Khan Academy から収集したデータを利用していたが、その問題回答ログ数は 100 万件以上であり、これまでにないほど大規模なデータを対象に分析が実施されたといえる。

また、MOOCs は多くの学習者が利用しており、こうした分析に基づいた学習の効率化や学習の継続を促進する教材推薦システムの開発は効果大きいと考えられる。従来、e ラーニングによる学習支援システムは、例えば大学であれば大学に所属する学生が利用者の中心で大学に所属していない人の利用が難しかったように、学習支援システムの利用者が限定されており、したがって、教材推薦による学習の効率化の活用可能性も限定的であったと考えられる。しかし、MOOCs における学習効果の改善は、単に一部の限られた学習者がその恩恵を享受できるということに留まらず、世界中の幅広い利用者がその恩恵を享受できるため、その社会的な影響は小さくないと考えられる。

以上、MOOCs について述べた。次に、深層学習について述べる。

### 2.3 深層学習

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、従来の機械学習では難しかった対象データの抽象的表現の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 [Schroff et al., 2015, Szegedy et al., 2014]、

音声認識 [Hinton et al., 2012, Bahdanau et al., 2015] , 会話認識 [Sak et al., 2015] , 機械翻訳 [Sutskever et al., 2014, Dong et al., 2015] , 質問応答文生成 [Yin et al., 2015] , 画像説明文生成 [Xu et al., 2015, Vinyals et al., 2014] 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 .

画像認識では深層学習により例えば犬の表現として目や鼻 , 口の表現が抽出できると報告する研究 [Zeiler and Fergus, 2014] や , 人間より高い精度で人の顔を見分けられたと報告する研究 [Schroff et al., 2015] もある . また , 機械翻訳の領域では , 翻訳前の文章の長さと翻訳後の文章の長さが可変の場合でも利用でき , 文章の長さに依らない特徴を抽出する機械翻訳モデルを開発し , 英語からフランス語への翻訳を試みる研究 [Sutskever et al., 2014] やフランス語とオランダ語 , スペイン語の共通表現を抽出し英語からただ 1 つのモデルで 3 言語に翻訳する研究もある [Dong et al., 2015] . また , これらを掛け合わせた画像説明文生成の領域では , 説明文の単語 1 つずつを生成する際に画像のどの部分に着目すべきかという表現を抽出し , 説明文を生成する研究 [Xu et al., 2015] もある . このように , 深層学習は多くの研究領域で大きな進展をもたらしている .

深層学習モデルを学習させるときには , 大規模な訓練データが必要となる . 深層学習モデルの内部変数は非常に膨大で数十万や数百万以上となることも多く , 膨大な変数を学習させるためには , 通常 , 大規模な訓練データが必要である . 例えば , 人間より高い精度で人の顔を見分けらると報告する顔認識の研究 [Schroff et al., 2015] では , 数百万人の 2 億枚以上の顔画像を訓練データに利用している . 英語からフランス語に翻訳する機械翻訳の研究 [Xu et al., 2015] では , 1200 万もの文章を訓練データとして利用している .

深層学習のネットワークには , いくつかの種類があるが , 特に , 画像処理に利用される Convolutional Neural Networks [LeCun et al., 1998] というネットワークと系列データの処理に利用される Recurrent Neural Networks [Williams and Zipser, 1989] ( 以下 , RNN ) というネットワークがよく利用される . ここでは , 知識獲得の予測に深層学習を用いた手法 [Piech et al., 2015] に用いられていたニューラルネットワークである RNN について説明する .

近年 , RNN はデータの大規模化や計算機性能の向上などにより幅広い領域の系列データに対して適用されるようになった . 具体的には , 機械翻訳 [Sutskever

et al., 2014, Dong et al., 2015] , 手書き文字認識 [Graves and Schmidhuber, 2009, Louradour and Kermorvant, 2014] , 音声認識 [Hinton et al., 2012, Bahdanau et al., 2015] , ユーザログ解析 [Hidasi et al., 2015] , 画像説明文生成 [Xu et al., 2015, Vinyals et al., 2014] , 医療診断 [Choi et al., 2015, Lipton et al., 2015] 等の領域で高い性能を発揮することが報告されている .

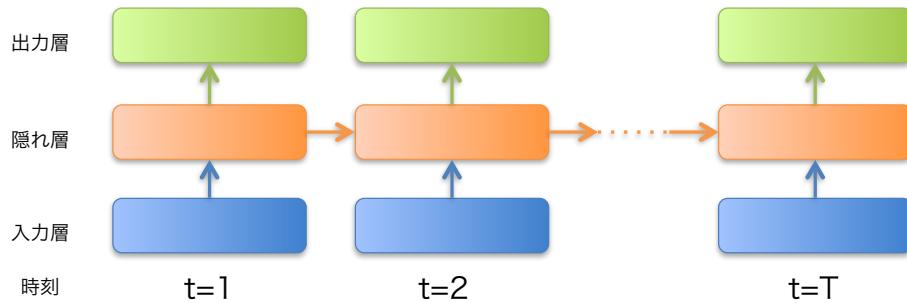


図 2.4: RNN の構造のイメージ

伝統的な RNN の構造は図 2.4 のように , 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  $h_t$  の計算に時刻  $t - 1$  の隠れ層の情報を入力する  $h_t = f(x_t, h_{t-1})$  の式のように , 一つ前の情報を繰り返し (recurrent) 入力するという構造である . 関数  $f$  は , 入力である  $x_t$  や  $h_{t-1}$  をアフィン変換して足しあわせた後 , 活性化関数にかけるというのがよく利用される . 活性化関数はシグモイド関数や  $\tanh$  (Hyperbolic Tangent 関数) , Relu [Nair and Hinton, 2010] , ELUs [Clevert et al., 2015] など多く提案されており , 通常 , 非線形関数である .

RNN の 1 つの特徴として , 効果的に長期的な表現を学習させることが難しいということが挙げられる [Bengio et al., 1994] . RNN の学習には勾配法に基づいた確率的勾配降下法 [Robbins and Monro, 1951, Kushner and Yin, 2003] や Adam [Kingma and Ba, 2014] , AdaDelta [Zeiler, 2012] など , さまざまな手法が利用可能である . しかし , いずれの勾配法を用いるにせよ , 勾配が爆発して学習モデルが壊れてしまうという勾配爆発 [Bengio et al., 1994, Pascanu et al., 2013] という問題や , 勾配が消滅して対象データの長期的な特徴量を捉えることができないという勾配消滅 [Pascanu et al., 2013, Hochreiter, 1998] という問題がしばしば発生する . これ

は， $\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1})$  の式に表れるように同じ変換を繰り返し行うためであり，このため，特に，長い系列データを RNN で学習する場合，効果的に長期的な表現を学習させることが難しい

こうした問題を解決もしくは緩和するため，ゲート付き活性化関数の利用や学習時の勾配に制約を加える方法が提案されている。勾配消滅の緩和に対しては，ゲート付き活性化関数の利用が有効であるが，詳細は後述する。勾配爆発の緩和に対しては，学習時の勾配に制約を加える方法が有効である。具体的には，[Mikolov, 2012] では学習させるパラメタの勾配の絶対値の最大値を予め決めておき，最大値以上の場合には，勾配の最大値になるように勾配の値を置き換えることで勾配爆発の影響を緩和する方法が報告された。また，[Pascanu et al., 2013] では学習させるパラメタの勾配のノルムの最大値を予め決めておき，最大値以上の場合には，ノルムが最大値以下になるように疑似コード 1 に従いノルムを抑制することで勾配爆発の影響を緩和する方法が報告された。

---

#### Algorithm 1 勾配爆発を防ぐための勾配ノルム抑制の疑似コード

---

```

 $\hat{\mathbf{g}} \leftarrow \frac{\delta \varepsilon}{\delta \theta}$ 
if  $\|\hat{\mathbf{g}}\| \geq threshold$  then
     $\hat{\mathbf{g}} \leftarrow \frac{threshold}{\|\hat{\mathbf{g}}\|} \hat{\mathbf{g}}$ 
end if
```

---

先に，言及したが，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。うまく設計された活性化関数を利用することで，データの長期的な特徴をよく捉えられたり，計算コストを削減することができたりする。以降では，よく研究報告で取り上げられる Simple RNN (以下，SRNN) [Williams and Zipser, 1989]，Long Short Term Memory RNN (以下，LSTM-RNN) [Hochreiter and Schmidhuber, 1997]，Gated Recurrent Neural Networks (以下，GRNN) [Cho et al., 2014] の 3 つについて詳細に説明する。

### SRNN

SRNN はゲート付き活性化関数を用いない簡単な構造の RNN である。[Le et al., 2015, Krueger and Memisevic, 2015] で報告される工夫を取り入れることで，データの長期的な特徴を効果的に捉えることができるようになるが，多くの場合で，

LSTM-RNN や GRNN のようにゲート付き活性化関数を用いる RNN の方がモデルの性能という点で優れている。

SRNN によるモデルの定式はいくつか種類が存在するが、シンプルなものは例えば下記の式で定義される。

$$\mathbf{h}_t = \tanh(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h) \quad (2.1)$$

$$\mathbf{y}_t = \sigma(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y) \quad (2.2)$$

ここでは、 $t$  は時刻を指し、 $\mathbf{x}_t$  は時刻  $t$  の入力ベクトルを指し、 $\mathbf{h}_t$  は時刻  $t$  の隠れ層を指し、 $\mathbf{y}_t$  は時刻  $t+1$  の各問題の正誤確率の予測値を指し、 $\mathbf{W}_{xh}$ 、 $\mathbf{W}_{hh}$  はそれぞれ重み行列を指し、 $\mathbf{b}_h$ 、 $\mathbf{b}_y$  はそれぞれバイアス項を指し、 $\tanh$  は  $(e^x - e^{-x}) / (e^x + e^{-x})$  で定義される Hyperbolic Tangent 関数を指し、 $\sigma$  は  $1 / (1 + e^{-x})$  で定義されるシグモイド関数を指す。訓練時には、重み行列  $\mathbf{W}_{xh}$ 、 $\mathbf{W}_{hh}$  とバイアス項  $\mathbf{b}_h$ 、 $\mathbf{b}_y$  を学習する。

## LSTM-RNN

LSTM-RNN は Long Short Term Memory という活性化関数を用いる RNN で、その名前の通り、SRNN では捉えることが難しかったデータの長期的表現と短期的表現の両方の獲得を目的に開発されたものである [Hochreiter and Schmidhuber, 1997]。LSTM-RNN は SRNN と比較すると、モデルの性能という点で優れているが、内部のパラメタの数が非常に大きく学習コストは大きい。最先端の成果を報告する研究でしばしば利用されているが、LSTM-RNN 自体が開発されたのは 1997 年であり LSTN-RNN が新しいというわけではない。

LSTM-RNN によるモデルの定式にはいくつか種類が存在するが、特に、後述する Deep Knowledge Tracing [Piech et al., 2015] で用いられる LSTM-RNN は下記

の式で定義される。

$$\mathbf{i}_t = \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i) \quad (2.3)$$

$$\mathbf{g}_t = \sigma(\mathbf{W}_{xg}\mathbf{x}_t + \mathbf{W}_{hg}\mathbf{h}_{t-1} + \mathbf{b}_g) \quad (2.4)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f) \quad (2.5)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o) \quad (2.6)$$

$$\mathbf{m}_t = \mathbf{f}_t \odot \mathbf{m}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \quad (2.7)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \mathbf{m}_t \quad (2.8)$$

$$\mathbf{y}_t = \sigma(\mathbf{W}_{my}\mathbf{m}_t + \mathbf{b}_y) \quad (2.9)$$

ここでは、 $\mathbf{i}_t$  は Input Gate を指し、 $\mathbf{f}_t$  は Forget Gate を指し、 $\mathbf{g}_t$  はメモリセルへの入力を指し、 $\mathbf{o}_t$  は Output Gate を指し、 $\mathbf{m}_t$  はメモリセルを指し、 $\mathbf{W}_{xi}$ 、 $\mathbf{W}_{hi}$ 、 $\mathbf{W}_{xg}$ 、 $\mathbf{W}_{hg}$ 、 $\mathbf{W}_{xf}$ 、 $\mathbf{W}_{hf}$ 、 $\mathbf{W}_{xo}$ 、 $\mathbf{W}_{ho}$ 、 $\mathbf{W}_{my}$  はそれぞれ重み行列を指し、 $\mathbf{b}_i$ 、 $\mathbf{b}_g$ 、 $\mathbf{b}_f$ 、 $\mathbf{b}_o$ 、 $\mathbf{b}_y$  はそれぞれバイアス項を指し、 $\odot$  は要素積を指す。

式 2.7 にあるように、メモリセルへの入力は 1 つ前のメモリセルの状態  $\mathbf{m}_{t-1}$  と入力  $\mathbf{g}_t$  であり、それぞれの入力に対して、過去のメモリセルからの情報を捨てる Forget Gate と現在からの情報を調整する Input Gate を作用させ、 $\mathbf{m}_t$  をえる。新しい隠れ層  $\mathbf{h}_t$  は式 2.8 のようにメモリセルからの出力を Output Gate で調整したものを受け取る。これらのゲートにより、長期的な特徴の短期的な特徴が捉えられるとされている。

## GRNN

GRNN は Gated Recurrent Unit [Cho et al., 2014] というゲート付き活性化関数を用いる RNN のことで、GRU は LSTM のように、長期的な表現と短期的な表現を捉えるために提案された活性化関数である。Cho ら [Cho et al., 2014] が 2014 年に発表して以来、GRNN 自体や GRNN の活用に関する研究が多く報告されている [Chung et al., 2014, Zaremba, 2015, Chung et al., 2015, Karpathy et al., 2015, Biswas et al., 2015, Pezeshki, 2015]。LSTM よりもゲートの数が少なく学習コストが小さい傾向にあるが、LSTM-RNN, GRNN の性能を比較した研究 [Chung

et al., 2014, Zaremba, 2015]において LSTM-RNN と GRNN が同程度の性能であることが報告されている。

GRNN は下記の式により定義される。

$$\mathbf{r}_t = \sigma(\mathbf{W}_{xr}\mathbf{x}_t + \mathbf{W}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_r) \quad (2.10)$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{xz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z) \quad (2.11)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}(\mathbf{r}_t \odot \mathbf{h}_{t-1} + \mathbf{b}_h)) \quad (2.12)$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t \quad (2.13)$$

$$\mathbf{y}_t = \sigma(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y) \quad (2.14)$$

ここでは、 $\mathbf{W}_{xr}, \mathbf{W}_{hr}, \mathbf{W}_{xz}, \mathbf{W}_{hz}, \mathbf{W}_{xh}, \mathbf{W}_{hh}$  は重み行列で、 $\mathbf{b}_r, \mathbf{b}_z, \mathbf{b}_h$  はバイアス項である。 $\mathbf{r}_t$  が Reset Gate(LSTM における Forget Gate に相当する機構) で、 $\mathbf{z}_t$  が Update Gate(LSTM におけるメモリセルに相当する機構) である。 $\mathbf{r}_t$  が 0 に近いほど前の隠れ層からの入力よりも現在の入力をより強く考慮するようになり、 $\mathbf{z}_t$  が 0 に近いほど前の隠れ層をより大きく更新するようになる。

以上、深層学習について述べた。次に、知識獲得の予測について述べる。

## 2.4 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問題である。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 Corbett らによって報告された Knowledge Tracing [Corbett and Anderson, 1994] である。スキルの習熟学習において、領域知識をよく分析し階層的に知識間関係を構築し、階層構造において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設計することで、ほとんどの学習者がスキルを十分に習熟できるとする仮説 [Keller, 1968, Bloom, 1968] や、コンピュータサイエンスの発展を受けて、予め獲得すべき知識が確実に獲得されるように学習者の知識の獲得有無を予測するというのが主な目的であった。Knowledge Tracing における

学習者とモデルは、学習者が勉強し知識を獲得したら、モデルが学習者が獲得した知識を予測することで学習者の獲得している知識の変化を追跡する（*Knowledge Tracing*），という関係になっている。

伝統的に，知識獲得の予測には知識獲得の時系列性を重視するものと，知識間の関係性を重視するものがある。[Corbett and Anderson, 1994] で報告された Bayesian Knowledge Tracing という手法は知識獲得の時系列性を重視するもので，問題に予めスキルを割り当て個々のスキルに習熟過程に関する 4 つの確率変数を定義しモデル化するというもので，スキル間の関係性は考慮しないが，個々のスキルの習熟，つまり時系列性を考慮する手法である。[Pavlik Jr et al., 2009] で報告された Performance Factor Analysis という手法は知識間の関係性を重視するもので，個々の知識（あるいは，スキル）に関する過去の回答の正誤を重み付けして，次の問題の正誤を予測しようというもので，Performance Factor Analysis は知識獲得の時系列性より知識間の関係性を重視する手法である。いずれの手法も本論文と関連が深い。

以降では，まず，Knowledge Tracing の定式化について述べ，Bayesian Knowledge Tracing と Performance Factor Analysis の 2 つの手法を説明し，最後に，深層学習を活用した Deep Knowledge Tracing について説明する。

#### 2.4.1 Knowledge Tracing の定式化

Knowledge Tracing は過去の学習者の問題回答履歴から学習者が次に解く問題の正誤を予測するというものである。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  $q_{t+1}$  を予測するタスクと表現できる。特に，過去の観測された問題の正誤から将来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といえる。予測性能の評価は [Yudelson et al., 2013, FALAKMASIR et al., 2015] では Accuracy で，[Piech et al., 2015] では AUC で行っており，目的に応じてさまざまである。

なお，モデルの入力次元である問題の粒度はさまざまである。問題はその問題

を回答するのに必要な知識を学習者が獲得しているか否かを評価するという点で，問題は知識集合を表現していて，また，その粒度もさまざまである．個々の問題をそのままモデルの入力次元とするものや，問題に予めタグを割り当て問題により評価される知識の粒度をある程度整え，そのタグをモデルの入力次元とすることもある．例えば，[Piech et al., 2015] ではモデルの入力次元は演習タグもしくはスキルタグと呼ばれるものであり，演習問題に割り当てられ，それぞれの演習問題で扱われる学習要素を説明するものである．通常，こうしたタグは専門家によって設計され，利用される．本論文では個々の問題をそのままモデルの入力次元として用いる．

#### 2.4.2 Bayesian Knowledge Tracing

Bayesian Knowledge Tracing [Corbett and Anderson, 1994] (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  $p(q_{t+1} = \text{correct} | q_1, q_2, \dots, q_t)$  をモデリングする手法である．BKT には下記の 4 つの確率変数がある．

- 初めから当該スキル理解している確率  $p(L_0)$  (もしくは  $p\text{-init}$ )
- 学習者が当該スキルを理解していない状態から理解している状態へ遷移する確率  $p(T)$  (もしくは  $p\text{-transit}$ )
- 学習者が当該スキルを理解しているが誤答する確率  $p(S)$  (もしくは  $p\text{-slip}$ )
- 学習者が当該スキルを理解していないが推測で正解する確率  $p(G)$  (もしくは  $p\text{-guess}$ )

これらの 4 つの確率変数がすべてのスキルに定義されている．つまり，スキル数を  $N$  とすれば，確率変数の合計数は  $4N$  である．学習者  $u$  がスキル  $k$  の問題を時刻  $t$  に解いた場合に正解する確率は下記の式に基づいて更新される．

$$p(L_1)_u^k = p(L_0)^k \quad (2.15)$$

$$p(L_t|obs=correct)_u^k = \frac{p(L_{t-1})_u^k \cdot (1 - p(S))^k}{p(L_{t-1})_u^k \cdot (1 - p(S))^k + (1 - p(L_{t-1})_u^k) \cdot p(G)^k} \quad (2.16)$$

$$p(L_t|obs=wrong)_u^k = \frac{p(L_{t-1})_u^k \cdot p(S)^k}{p(L_{t-1})_u^k \cdot p(S)^k + (1 - p(L_{t-1})_u^k) \cdot (1 - p(G))^k} \quad (2.17)$$

$$p(L_t)_u^k = p(L_t|obs)_u^k + (1 - p(L_t|obs)_u^k) \cdot p(T)^k \quad (2.18)$$

$$p(C_t)_u^k = p(L_{t-1})_u^k \cdot (1 - p(S))^k + (1 - p(L_{t-1})_u^k) \cdot p(G)^k \quad (2.19)$$

右上の  $k$  はスキル番号を示し、右下の  $u$  はユーザ番号を示すことに注意されたい。まず、学習者  $u$  が初めから当該スキル  $k$  を身につけている確率は式 2.15 の通り定義する。正解が観測され、正しく当該スキルを身につけている確率は、式 2.16 で与えられ、不正解が観測されたが、正しく当該スキルを身につけている確率は、式 2.17 で与えられ、それらを合わせて、次の時刻に当該スキルを身につけている確率は、式 2.18 で与えられる。このように定めることで、理解しているがうっかり間違ってしまう場合や、理解していないがてずっぽうで正解してしまう場合を考慮できる。なお当該モデルでは、身についたスキルの忘却は無視している。最後に、学習者  $u$  がスキル  $k$  の問題を時刻  $t$  に解いた場合に正解する確率  $p(C_t)_u^k$  は式 2.19 のように算出され、この値を次の問題の正誤予測に利用する。

上記に説明したモデルの学習にはいくつかの方法が適用され報告されている。1つは [Corbett and Anderson, 1994] にあるように HMM を用いて生成モデルとして学習させる方法であり、1つは [Yudelson et al., 2013] にあるように勾配法を用いて識別モデルとして学習させる方法である。それぞれ長所と短所があるが、特に、大規模データへの適用という観点では HMM に基づいた生成モデルの手法では計算量が大きく学習に非常に多くの時間がかかるてしまうということもあり、[Yudelson et al., 2013] では勾配法に基づいた識別モデルとして学習させている。具体的には、[Yudelson et al., 2013] では、目的関数に負の対数尤度 (Negative Log Likelihood) を利用し、勾配降下法 (Gradient Descent) で学習させている。

### 2.4.3 Performance Factor Analysis

Performance Factor Analysis [Pavlik Jr et al., 2009] (以下, PFA) も過去の学習者の問題回答履歴から学習者が次に解く問題の正誤を予測するための手法である。しかし、知識獲得の時系列性を考慮する BKT と異なり、知識獲得の順番を考慮せず知識間の関係性を考慮して予測する手法である。PFA は下記のように定義される。

$$p(i, j \in KCs, s, f) = \sigma(\beta_j + \sum_{k \in KCs} (\gamma_k s_{i,k} + \rho_k f_{i,k})) \quad (2.20)$$

ここでは、 $s$  は事前に正答した問題回答、 $f$  は事前に誤答した問題回答、 $p$  はユーザ  $i$  が知識  $j$  に正答する確率、 $\beta_j$  は知識  $j$  の簡単さ、 $\gamma_k$  と  $\rho_k$  はそれぞれ知識  $k$  の正答と誤答の重み、 $s_{i,k}$  と  $f_{i,k}$  はそれぞれユーザ  $i$  が知識  $k$  に事前に正答した問題回答、事前に誤答した問題回答である。 $\sigma$  はシグモイド関数、過去の各知識の正誤を重み付けしシグモイド関数にかけ、別の問題の正誤を予測するというものである。

PFA は知識間の関係性を考慮できない場合に複数の知識がないと獲得できない知識のモデルが難しいという Bayesian Knowledge Tracing やその拡張手法の問題を解決するために提案された。PFA は知識間の関係性を重み付けして考慮しているが、問題回答の順番は考慮しない。

### 2.4.4 Deep Knowledge Tracing

Deep Knowledge Tracing [Piech et al., 2015] (以下, DKT) は RNN を利用し Knowledge Tracing を行う手法である。2015 年 6 月に発表された。数学の問題回答ログのデータセットで実験され、高い性能で将来の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。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され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。DKT の構造と最適化、および知識間関係の抽出手法について順に説明していく。

## 構造

まず，DKT の構造について述べる．DKT の構造は伝統的な RNN の構造に基づいている．伝統的な RNN は入力のベクトル系列  $\mathbf{x}_1, \dots, \mathbf{x}_T$  を出力のベクトル系列  $\mathbf{y}_1, \dots, \mathbf{y}_T$  に写像する．この写像は，隠れ状態  $\mathbf{h}_1, \dots, \mathbf{h}_T$  を計算することで達成されるが，一連の写像の過程で過去観測から得られる関連情報を将来予測のために連續的に符号化している，とみなせる．確率変数は下記の式で定義されるネットワークにより関連付けられる．

$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}) \quad (2.21)$$

$$\mathbf{y}_t = g(\mathbf{h}_t) \quad (2.22)$$

モデルは関数  $f$  と  $g$  によって定義されており，これらの関数  $f, g$  には SRNN の式 2.1, 2.2 や LSTM-RNN の式 2.3–2.9, GRNN の式 2.10–2.14 を利用できる．

RNN で学習者の学習行動の観測結果をモデリングするため観測結果を固定長の入力ベクトル  $\mathbf{x}_t$  の系列に変換する必要があるが，DKT ではシンプルな変換を行っている．具体的には，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  $\mathbf{x}_t$  とする，というものである．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．

表 2.1: 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

ユーザ ID	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
A	1	1	0	$\mathbf{x}_1$	[0000:1000]
A	2	1	1	$\mathbf{x}_2$	[1000:0000]
A	3	2	1	$\mathbf{x}_3$	[0100:0000]
A	4	3	0	$\mathbf{x}_4$	[0000:0010]
A	5	3	1	$\mathbf{x}_5$	[0010:0000]
A	6	4	1	$\mathbf{x}_6$	[0001:0000]

具体例を交えて説明する．例えば，演習問題の数が 4 つで，問題回答は 1 つず

つしかできないと仮定する。 $M = 4$ であり、 $x_t$ の長さは8である。ある学習者が、表2.1の回答ログのように問題を回答し正誤が観測されたとする。この時に、例えば、表2.1に記載のような入力ベクトルの系列となる。このようにして、回答行動の観測結果を符号化することで、どの演習問題をいつ正解もしくは不正解したのかをRNNに入力できる。

出力 $y_t$ は問題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。したがって、 $t+1$ の回答 $q_{t+1}$ の正誤予測は $t+1$ に回答される問題 $q_{t+1}$ に対応する $y_t$ の要素から読み取れる。

### 最適化

次に、DKTの最適化について述べる。訓練時に用いられる目的関数は、モデルにおいて学習者の回答行動の観測系列の負の対数有度（Negative Log Likelihood）である。 $\delta(q_{t+1})$ を時刻 $t+1$ にどの問題が回答されたかのone-hotベクトルとし、 $a_{t+1}$ を時刻 $t+1$ に当該問題で正答したか否か（1か0）とし、 $l$ をクロスエントロピーとすれば、当該予測結果に対するロス関数は $l(y^T \delta(q_{t+1}), a_{t+1})$ であり、学習者一人のロスは下記の式で与えられる。

$$L = \sum_t l(y_t^T \delta(q_{t+1}), a_{t+1}) \quad (2.23)$$

学習時はミニバッチごとに確率的勾配降下法で目的関数を最小化する。[Piech et al., 2015]では、モデル学習時には過学習を防ぐため $y_t$ への入力としての $h_t$ にはdropout [Srivastava et al., 2014]を適用している（ $h_{t+1}$ の方向にはdropoutを適用しない）。また、系列方向の誤差逆伝搬 [Werbos, 1990]において勾配が爆発するのを防ぐため、閾値以上のノルムの勾配は [Pascanu et al., 2013]にしたがい、制約を設けている。

### 知識間関係抽出法

次に、DKTのモデルを利用した知識間関係（あるいは、問題間関係）抽出法について述べる。DKTのモデルは、従来ではよく人間の専門家が行っていたデータの潜在的な構造や概念を発見するタスクに応用できる。問題 $i$ と $j$ のすべての有向

ペアのうち下記の条件を満たすものに対して下記の影響度  $J_{ij}$  を割り当てる .

条件 有効ペア  $(i, j)$  について , 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  $V\%$  以上であること .

### 影響度

$$J_{ij} = \frac{y(j|i)}{\sum_k y(j|k)}$$

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 . [Piech et al., 2015] では , 問題間影響行列からのネットワーク抽出には ,  $V = 1$  を用いた . また , ネットワークの可視化に際しては , 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ネットワークを構築した .

さらに , [Piech et al., 2015] は , 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を指摘している .

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 .

以上 , 知識獲得の予測について述べた . 次に , ネットワーク分析について述べる .

## 2.5 ネットワーク分析

ネットワーク分析は現実世界に存在するさまざまな複雑なネットワークの特徴を明らかにするための分析である . 従来 , ネットワーク分析は社会の信頼関係や社会規範といった社会関係資本を分析するために用いられており , 社会学の研究

領域であった [Putnam, 1993, Coleman and Coleman, 1994] が、ウェブやそれを活用するソーシャルメディアの普及に伴い注目が高まり、ウェブページやユーザを分析するためにしばしば用いられるようになった。事例として、ウェブページ検索エンジンの性能向上のためウェブページの重要度を評価するページランクという指標を報告した研究 [Page et al., 1999] や生物学における論文の共著ネットワークの構造を分析する研究 [Sun et al., 2011]、ソーシャルメディアにおけるユーザ間の関係性から友人関係を推定し友人推薦に活用しようとする研究 [Dong et al., 2012] 等がある。

ネットワーク分析の対象となるネットワークには、予め明示的に対象間の関係がネットワークとして与えられているもの (Explicit Networks) と、明示的には与えられておらず対象間の類似性や関係性の強度からネットワークを構築し利用するもの (Implicit Networks) がある。例えば、ウェブページのハイパーリンクによる関係、ソーシャルメディアの友人関係、論文の共著関係が Explicit Networks に該当する。ネットワーク上でエッジとして表現されたこれらの関係はそれぞれウェブページ同士を友人同士を著者同士の明示的な関係を表している。一方で、明示的ではない関係を分析する場合は、対象間の類似性や関係性の強度を表す行列に基づいて閾値以上の強度の対象間にエッジを張るという方法や、各対象から一定数のエッジを強度の大きい対象間に引いていく方法で、ネットワークを構築し、そのネットワークを対象にネットワーク分析を行う。例えば、説明文の語の共起の類似性を表現する行列から説明対象のネットワークを構築するもの [Buzydowski et al., 2002] や、知識間関係の影響度行列からネットワークを構築する [Piech et al., 2015] が該当する。

ネットワーク構造を評価する指標は多くのものが存在する。ネットワークにおける個々のノードやエッジの特徴を分析する指標として、中心度という点でノードを評価するノードの媒介中心性やノードの次数中心性、経路における中心度という点でエッジを評価するエッジの媒介中心性などがある。ノードの媒介中心性とは、当該ノードを通る経路数で表現され、ノードの次数中心性とは当該ノードに接続しているエッジ数で表現され、エッジの媒介中心性は当該エッジがノードの組の最短パスに含まれている割合で表現される。また、ネットワークの構造を捉えるための指標も開発されており、例えば、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モジュラリティという指標 [Newman, 2006] や階層性を評価するフロー階層という指標 [Luo and Magee, 2011] や GRC という指標 [Mones et al., 2012] , 等がある .

$$Modularity = \sum_i (e_{ii} - a_i^2) \quad (2.24)$$

モジュラリティは式 2.24 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 .

$$FlowHierarchy = \frac{\sum_{i=1}^L e_i}{L} \quad (2.25)$$

フロー階層は式 2.25 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 .

また , [Luo and Magee, 2011] では , サンプルネットワークに適用した結果についても紹介しており , 図 2.5 における  $(A, B, C, D, E)$  に対するフロー階層の値は  $(1, 1, 1, 0.40, 0.67)$  であったという . このことからも , フロー階層が環状構造がない方が指標の値が高くなり , したがって , ネットワークの階層性を表現できる指標であることを確認していた .

$$GRC = \frac{\sum_{i \in V} [C_R^{max} - C_R(i)]}{N - 1} \quad (2.26)$$

GRC は Global Reaching Centrality のことで , 式 2.26 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

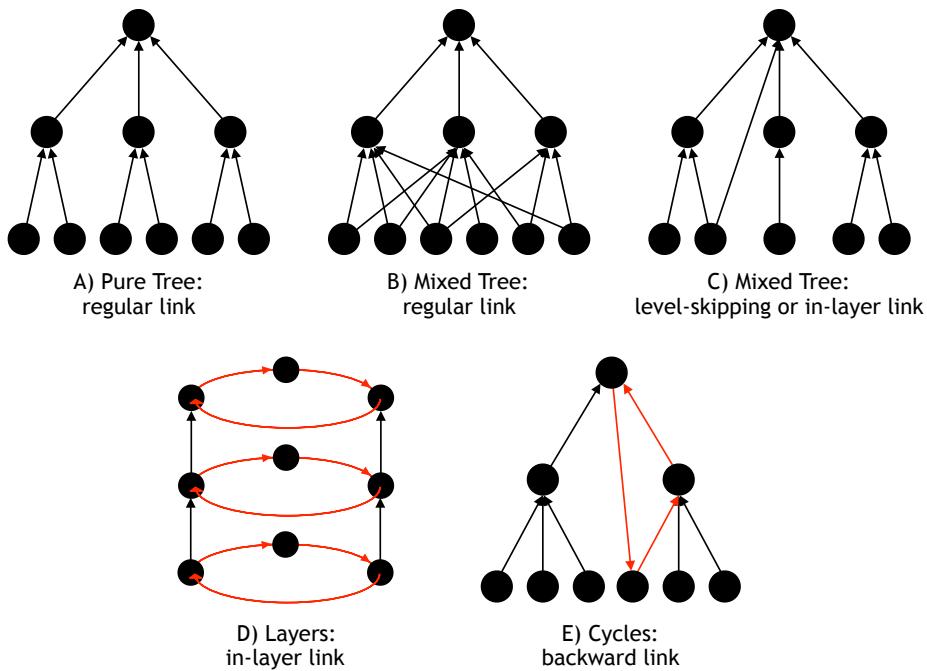


図 2.5: フロー階層の適用事例

辿って，ノード  $i$  から到達できるノードの割合を指す． $GRC = 1$  の場合，Local Reaching Centrality が 0 でないノードがたかだか 1 つしか存在せず，したがって，当該ネットワークはスター状のネットワークということになる．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．

以上，先行研究について述べた．次に，分析手法について述べる．

## 第3章 分析手法

本章では、宣言的知識と手続き的知識の知識獲得における知識構造の違いを分析するための手法について説明する。まず、分析手法全体の流れを概説し、手法全体が3つのブロックから構成されることを述べる。その後、まず、既存の知識間関係の抽出手法の中で、Deep Knowledge Tracing が最適であることを説明する。次に、分析対象の問題回答ログデータについて要件があることを述べ、その要件として、比較検証できる複数のデータセットを作成できること、各データセットが大規模であること、初等中等教育水準の数学や地理、歴史に関する科目であること、について説明する。最後に、抽出した知識間関係ネットワークのモジュール性と階層性を評価するために、ネットワーク構造指標の中でもモジュラリティとフロー階層およびGRC を用いること、また、これらの各指標について、宣言的知識によるネットワークと手続き的知識によるネットワーク間で有意差検定を実施し、構造の違いを検証することを述べる。

まず、分析手法全体の流れを説明する。手法全体の流れを図 3.1 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。本論文では、学習者の知識獲得について分析を行うが、学習者が知識を獲得しているか否かを評価するためには、通常、その学習者の問題への回答を評価する。問題への回答結果は、その問題によって評価される知識を学習者が既に獲得しているか否かを表現しており、回答結果が正解であれば、その問題によって評価される知識を既に獲得しており、回答結果が不正解であれば、その問題によって評価される知識は未だ獲得していないと評価できる。したがって、分析では、問題回答ログデータを対象データに用いる。

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いいというわけではなく、いく

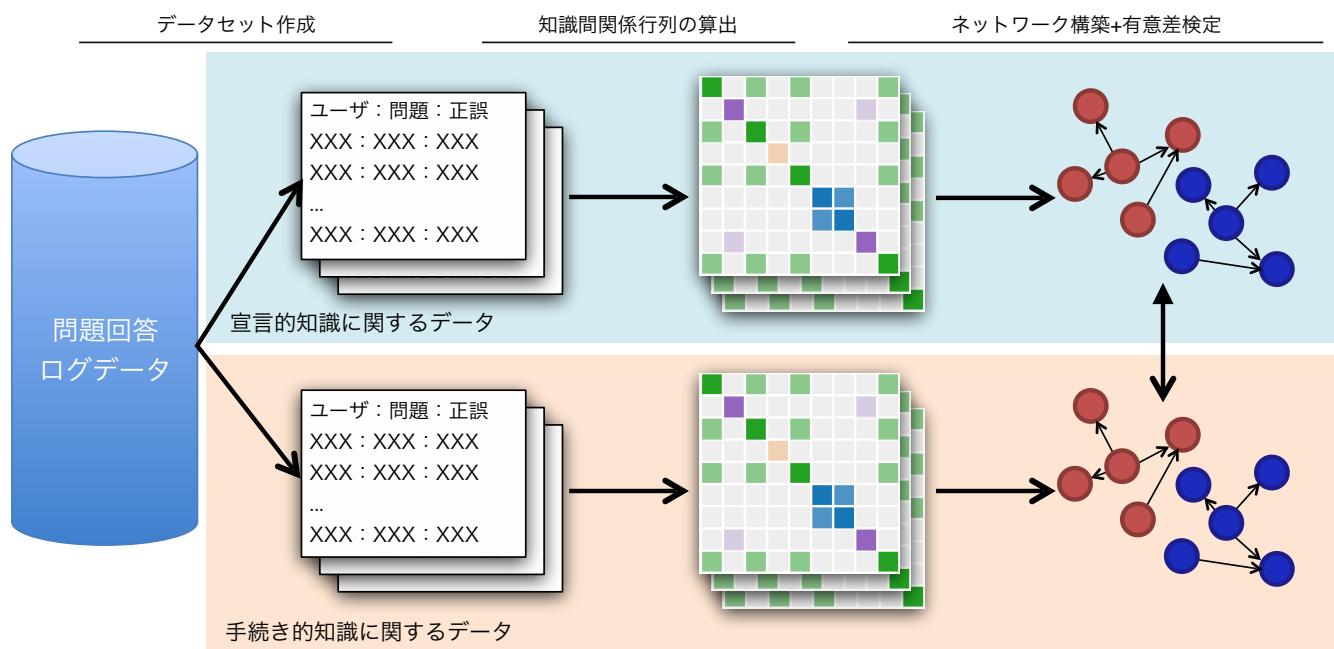


図 3.1: 分析手法全体の流れ

つかの要件を満たすものを利用する。その要件については、後述する。

そして、作成した複数のデータセットに知識獲得における知識間関係を抽出できる手法を適用し知識間関係行列を算出する。また、知識間関係を抽出する手法も特に Deep Knowledge Tracing が最適であるが、その理由について後で詳細に述べる。

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有意差検定を実施し、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。

以上が、分析手法全体の流れである。以降では、知識獲得における知識間関係の抽出、データセットに利用できるデータの要件、知識間関係ネットワーク構築とその構造評価を順に説明していく。

### 3.1 知識獲得における知識間関係の抽出

これまでの知識間関係を抽出する手法の中でも，Deep Knowledge Tracing（以下，DKT）が最適であること説明する。まず，知識間関係を抽出する手法としては，過去に獲得した知識から知識獲得を予測する過程で行うものと，そうでないものがある。後者については，専門家が作成するという手法や，テキスト解析により概念関係ネットワークを構築するという手法 [Chen et al., 2008] がある。しかし，これらはどちらかといえば，専門家や研究者の「こういう関係になっているはずだ」という仮説に基づいたものであり，必ずしも知識獲得における知識と知識の関係を表現しているわけではなく，定量分析への利用は限定的である。専門家が手で知識の構造化を行う方法では，自分自身や個々の生徒の学習過程の観察を通して構造化が試みられた。しかし，得られた知識間関係は定量分析に基づいたものであり，この関係に基づいて定量分析を実施したとしても，手法全体を通して定量分析を行ったとは言えない。教材テキストの解析により知識の構造化を行う方法 [Chen et al., 2008] は，知識の内容が近ければ教材テキストの語が類似するはずである，という考えに基づいて抽出するという方法である。しかし，得られた知識間関係は類似性を語の類似性を考慮したもので，知識の獲得順序や影響関係を抽出したものではない。したがって，いずれの手法も定量分析への活用は限定的である。

一方で，過去に獲得した知識から知識獲得を予測する過程で行うものは，学習者が既に獲得している知識を基にある知識が獲得されやすいかを予測しているという点で，学習者の知識獲得における知識間関係を表している可能性が高い。したがって，知識間関係を定量的に抽出する手法としては，過去に獲得した知識から知識獲得を予測する過程で知識間関係を抽出する手法に絞る。

問題回答正誤の分析により知識の構造化を行う方法では，2つの問題 $i$ と $j$ の間で問題 $i$ が正解後と不正解後の問題 $j$ の正解率の差と着手順序を基に知識は構造化された。しかし，2つの問題 $i$ と $j$ の関係性のみを考慮しており，他の問題との関係性は独立だと見なされているため得られた知識間関係は2つ知識の間についてのものを合算したものであり，複雑で密接に関係している複数の知識の獲得順序や影響関係を捉えているものではない。

次に、知識間関係の抽出を、過去に獲得した知識から知識獲得を予測する過程で行う手法として、先行研究の説明において、Performance Factor Analysis（以下、PFA）とDKTについて述べた。いずれの手法も知識間関係を考慮して予測に利用しているが、DKTの方が有効性が高いと考える。なぜなら、[Piech et al., 2015]では言及されていなかったが、DKTはPFAの拡張になっているためである。DKTはRNNを利用してあり、

$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}) \quad (3.1)$$

$$\mathbf{p}_t = \sigma(\mathbf{h}_{t-1} \cdot \mathbf{W}_{hp} + \mathbf{b}_p) \quad (3.2)$$

で与えられる。一方でPFAは

$$p(i, j \in KCs, s, f) = \sigma(\beta_j + \sum_{k \in KCs} (\gamma_k s_{i,k} + \rho_k f_{i,k})) \quad (3.3)$$

で与えられる。したがって、

$$f(\mathbf{x}_t, \mathbf{h}_{t-1}) = \mathbf{x}_t + \mathbf{h}_{t-1} \quad (3.4)$$

$$\mathbf{h}_0 = [0, 0, \dots, 0] \quad (3.5)$$

とすると、 $\mathbf{h}_t$ がこれまでの各問題についての正答回数を表現するベクトルと各問題についての誤答回数を表現するベクトルを結合したベクトルになるが、これは、ベクトルsとfを結合したものと同じである。したがって、PFAはDKT内部のRNNの繰り返しの部分を表現する関数fを式3.4にした特殊なケースであり、DKTはPFAの拡張になっている。したがって、PFAよりDKTの方がより適している。逆に、DKTを利用するということを考えた時に、[Piech et al., 2015]で、既に知識間関係を抽出できることが報告されており、複雑で密接に関係している複数の知識の獲得順序や影響関係を捉えている可能性があり、DKTを利用する事が最適であると考えられる。

## 3.2 データの要件

データセットは問題回答ログデータから作成するが、問題回答ログデータは、特に、下記の要件を満たすものが適切である。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。

まず、比較検証できる複数のデータセットを作成できることが必要であることについて述べる。本論文では、宣言的知識と手続き的知識の知識獲得における知識構造を比較するが、そのためには、宣言的知識と手続き的知識それぞれ複数のネットワークで比較検証できた方がいい。逆に、1つずつを比較しただけでは、想定していなかったデータセットの特徴やノイズの影響を受け必ずしもよく検証できたとは言えないからである。したがって、データは、複数の宣言的知識あるいは手続き的知識の獲得を主目的とする問題回答ログのデータセットを作成できるものであることが必要である。

次に、各データセットが大規模であることが必要であることに説明する。深層学習は大量のデータを元に特徴量を抽出するため、深層学習モデルを十分学習させるためには、大規模なデータが必要である。これは、RNN を活用する Deep Knowledge Tracing についても同様である [Piech et al., 2015]。したがって、作成した複数の問題回答ログのデータセットはいずれも大規模である必要がある。実際には、十分学習させるだけのデータセットが得られているかは実験を通して構築した予測モデルの性能を評価して判断できることであるため、既存研究のデータセットを目安にできるだけ大規模な問題回答ログデータを利用し、実験にて深層学習が有効であったかを確認する。

最後に、初等中等教育水準の数学や地理、歴史に関する問題回答ログであることが重要であることについて述べる。本論文では、宣言的知識と手続き的知識の知識獲得における知識構造を比較するが、そのためには、複数のデータセットが特に宣言的知識あるいは手続き的知識の獲得を主目的とするように偏りが大きいことが必要である。この要件に合致するデータについて、科目と難易度という2軸で整理する。

まず、科目について、従来より、数学やプログラミング等の問題はその多くが

手続き的知識を獲得しているか否かを評価するものとされており、どちらかといえば、手続き的知識の獲得を主目的とする科目であるといえる。一方で、しばしば暗記科目と呼ばれるような地理や歴史等の問題はその多くがある特定の地域の過去の出来事の事実について知識を扱っており、どちらかといえば、宣言的知識の獲得を主目的とする科目であると考えられる。

次に、難易度について、科目が同じでも難易度が変わると知識の獲得のされ方も変わってくる。歴史を例に挙げて説明する。歴史について学ぶ時に、難易度が低い時は、例えば「江戸幕府を築いたのは誰か。」や「江戸時代の次は何時代か。」という問題では、特に事実に関する知識を獲得するのが中心である。一方で、難易度が高い時は、例えば「なぜ、徳川家康は江戸幕府を築けたのかを文章で説明せよ。」や「江戸時代と明治時代についてまとめよ。」という問題であれば、この問題を解くためには、単に歴史的な事実を知っているということだけでなくそれらを伝わる形式に構造化したり、分からぬ内容を推論したりする作業が必要であろう。

簡単に言うと、科目が同じでも難しい問題はいろいろなことを知っていて、分かりやすく伝える練習をしていないと正解できないということである。知識構造という点では、科目が同じでも難易度が高いと、目的志向の知識である手続き的知識の有無が問題回答の正誤に影響を与える可能性が高く、知識構造もより階層的になる可能性が高い。したがって、例えば、Coursera のように大学水準の内容を扱う MOOCs における問題回答ログデータはあまり適していないと推察される。

科目と難易度の整理を踏まえると、難易度が高すぎないこと、また、その範囲内で数学やプログラミング等の手続き的知識の獲得を主目的とする科目や地理や歴史などの宣言的知識の獲得を主目的とする科目に関するものであることは要件に合致するデータであると考えられる。逆に難易度が高すぎないものでかつ、大規模データが利用できるとなると、数学や地理、歴史のように公教育に導入されており、よく学習されているものに限られてくると推察される。

これらの3つの要件が満たされていた時に、Deep Knowledge Tracing のモデルを十分学習させることで知識獲得における知識間関係を複数のデータセットから抽出することができ、また、抽出した知識間関係を宣言的知識と手続き的知識の対比において比較することができる。

こうした要件に合致するを可能があるデータとしては特に、初等中等教育水準のMOOCsの問題回答ログデータがある。初等中等教育水準のMOOCsであれば、算数や数学、地理や歴史の科目的講座が存在し、かつ、それらが難しすぎない可能性が高く、また、MOOCsでは非常に多くの学習者が利用しているため、多くの問題回答ログデータが収集できる可能性が高い。

### 3.3 知識間関係ネットワークの構造評価

Deep Knowledge Tracingより抽出した知識間関係を分析するために、知識間関係からネットワークを構築する方法を述べ、その構造におけるモジュール性と階層性を評価し違いを明らかにするために、ネットワーク分析指標の中でもモジュラリティとフロー階層、GRCを用いること、また、各データセットから構築した知識間関係ネットワークの指標について、宣言的知識および手続き的知識のネットワーク集合間で有意差検定を実施することを説明する。

まず、ネットワーク構築法について述べる。知識構造はネットワークとして明示的に与えられておらず行列として与えられており、ネットワークを構築する必要があるためである。構築方法には、知識間の影響度に基づいて閾値以上の強度の知識間にエッジを張るという方法や、各知識から一定数のエッジを影響度の大きい知識間に引いていく方法がある。前者は [Piech et al., 2015] で用いられた方法であるが、ここでは、後者の方法を用いる。具体的には、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  $M$  個選択し、 $i$  から  $j$  にエッジを引く、という方法である。各問題が扱う知識の獲得にどの知識が有効かの関係抽出に際して、議論の焦点は、ネットワーク全体における近さ（グローバルな関係）というよりは各問題間の近さ（ローカルな関係）であり、知識結合という局所的な関係から構築したネットワークの方が、どちらかといえば、よく知識構造を表現していると考えられるからである。ここでは、 $M = 3$  とし、つまり、構築するネットワークは 1 つの問題に対して 3 つの問題からエッジが入力される有向グラフである。

次に、ネットワーク構造におけるモジュール性と階層性の評価に、ネットワーク分析指標の中でもモジュラリティとフロー階層、GRC が有効でありそれらを用

いること説明する。本論文では、知識構造を表現するネットワークのモジュール性と階層性を分析するが、ここでいう知識構造のモジュール性とは、特に、ある知識を予め獲得しているか否かが他の知識獲得に及ぼす影響が小さくなるように知識が構造化されている度合いのことであり、また、知識構造の階層性とは、その獲得が順序付けられるように知識が構造化されている度合いのことである。

これらのモジュール性と階層性の評価には先行研究の説明の際に紹介したモジュラリティやフロー階層、GRC が有効である。まず、モジュール性について述べる。例えば、縄文時代に関する知識と江戸時代に関する知識はそれぞれ時代における知識間（石器と土器、江戸幕府と徳川家康等）は相互に関連しあっている可能性が高いと考えられる。一方で、時代をまたいだ知識間（石器と江戸幕府、土器と徳川家康等）の影響は弱いと考えられる。ネットワークに存在するエッジの数が一定の場合、モジュール性は適当に分割した知識集合（江戸時代に関する知識集合）内でのエッジが多く、逆に知識集合間（縄文時代に関する知識と江戸時代に関する知識）のエッジが少ない度合いのことであるが、これは、式 2.24 で表現されるモジュラリティと同じである。したがって、モジュラリティは知識構造のモジュール性の表現に有効であると考えられる。

次に、階層性について述べる。例えば、1 次方程式を解く知識と 2 次方程式を解く知識の 2 つの知識においては、先に 1 次方程式を解くためのさまざまな知識が獲得され、その後、2 次方程式を解く知識が獲得される、というようにその獲得が順序付けられていると考えられる。この関係性をネットワークにした場合、1 次方程式を解く知識ノードと 2 次方程式を解く知識ノードによる環状構造がなく、1 次方程式を解く知識ノードからエッジを辿ることで 2 次方程式を解く知識ノードに到達できるが、その逆はできないネットワークとなっているはずである。したがって、環状構造の少なさを評価するフロー階層や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度合いを評価する GRC は知識構造における階層性の一部を表現できる可能性がある。そこで、階層性の評価については、フロー階層と GRC の 2 つを用いる。

最後に、各データセットから構築した知識間関係ネットワークの指標について、宣言的知識および手続き的知識のネットワーク集合間で  $t$  検定を実施することを説明する。本研究の目的は知識獲得における知識構造を、宣言的知識と手続き的知

識を対比し，知識獲得における宣言的知識の知識構造の方が手続き的知識の知識構造よりもモジュール性が高く逆に，知識獲得における手続き的知識の知識構造の方が宣言的知識の知識構造よりも階層性が高いことを検証することである．これを達成するために3つのネットワーク構造指標の平均が宣言的知識の知識構造と手続き的知識の知識構造の間で異なることを示す．そのため独立した2集合へのt検定により3つのネットワーク構造指標の平均それぞれにおいて統計的に有意差があることを検証する．

以上，分析手法について述べた．次章では，実験で利用するデータセットについて述べる．

# 第4章 データセット

本章では、実験で用いるデータセットについて述べる。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「勉強サプリ<sup>1</sup>」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<sup>2</sup>。まず、勉強サプリから収集されたログデータが分析に利用するためのデータ要件を満たした本分析に最適なデータであることを述べる。次に、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述べる。特に、歴史や地理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ることを指摘し、歴史や地理に関する5講座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ることを指摘し、算数や数学に関する6講座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。

## 4.1 勉強サプリ

ここでは、勉強サプリより収集された問題回答ログデータが本分析に最適なデータであることを述べるために、まず、勉強サプリについて説明する。勉強サプリは主に小学4年生から中学3年生を対象とした国内最大級の大規模オンライン講座である。リクルートマーケティングパートナーズ(株)が運営している。サービスは2015年に開始された。学習者はオンライン上で問題回答形式によるドリル演習や動画視聴形式による授業聴講を通して勉強する。小学生には各学年ごとに国語、社会、算数、理科の4科目が、中学生には各学年ごとに国語、数学、英語の3科目

---

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

<sup>2</sup>本論文の研究は勉強サプリを運営するリクルートマーケティングパートナーズ(株)との共同研究プロジェクトの一環で行われている。

と学年共通で地理、歴史、公民、理科Ⅰ、理科Ⅱの5科目が提供されており合計26の講座が提供されている。無料で利用できるコンテンツもある。

勉強サプリは提供教材、利用者数、学習行動ログ数の点でサービスの規模が非常に大きい。4,000以上の授業動画と7,000近い演習問題を提供しており、サービス開始から1年経たずして、利用者は5万人以上あり、講義視聴ログ数は100万以上、問題回答ログ数は総計で1,000万以上である<sup>3</sup>。



図 4.1: 勉強サプリのマイページ

図 4.2: 勉強サプリの講義一覧ページ



図 4.3: 勉強サプリの講義視聴ページ

図 4.4: 勉強サプリの問題演習ページ

サービスの具体を利用画面を交えて説明する。図4.1, 4.2, 4.3, 4.4にページのイメージを記載する。図4.1はログイン後のマイページを示している。学習に関する

<sup>3</sup>2015年11月31日時点。

るコンテンツだけでなく、利用継続性の向上を狙って、ポイントというシステムやサプモン（勉強サプリのモンスター）というアバターを用いたゲームを大きく表示している。図4.2は中学1年向け数学の講義一覧ページを示している。ユーザは講義視聴のドリル演習の両方あるいはどちらかを勉強できる。図4.3は講義視聴ページを示している。講義視聴ページは視聴しやすいように細かく分割されている。図4.4は問題演習ページを示している。問題演習はテスト形式で提供されており、一度に複数の問題が提供され、また、その採点も同時に行われる。この問題回答から収集されるログデータの仕様は同時に1つの問題しか回答ログが入らないというDeep Knowledge Tracingで想定されている設定と異なるため、Deep Knowledge Tracingを拡張する必要があるが、その拡張法については実験説明の際に合わせて述べる。

現在の学習指導要領 [文部科学省, 2011]によると、小学4年生の社会では社会基盤や地域社会など地理に近い内容を扱い、小学5年生の社会では地域の山地山脈、気候など地理に近い内容を扱い、小学6年生の社会では日本の歴史に関する内容を扱うとしており、また、実際に、勉強サプリの学習内容も概ねそれに従っている。

勉強サプリは小学生から中学生が対象で、その難易度は難しすぎず、収集された問題回答ログデータは1,000万以上と膨大であり、また、先に指摘したような地理や歴史といったいわゆる暗記系の科目と数学に関する科目が複数提供されているため、データの要件を満足している可能性が高い。そこで、以降では、勉強サプリから収集された問題回答ログデータのなかで、特に地理と歴史に関する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）のデータと、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）のデータの11データよりデータセットを作成し、これらのデータセットをそれぞれ、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とする。

## 4.2 データセットの作成

データセットの作成について述べる。データセットは問題回答ログデータから作成する。対象期間は2015年4月から2015年11月の8ヶ月である。

勉強サプリでは学習者の回答は自動採点される。いずれの講座も回答は選択方式で、1つの回答欄には選択肢の中から1つの数字や文字を選択する。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1つの問題しか提示されないという形式だけでなく、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。つまり、同時に複数の問題回答ログが発生しうるデータである。

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下記に該当するログデータをノイズとみなして除去した。

条件 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もの。

これは、オンライン講座ではサービス上の学習者の行動には大きな制約はなく、1クリックで簡単に問題演習ページを開けてしまえる状況にあるということや、すべての回答欄が未記入で投稿された回答は直前に着手した問題群と同じ問題群であることが多いこと、特に回答時間が短く誤って当該問題演習ページを開いてしまったと推察されるログが多かったからである。

以上、データセットの作成について述べた。11データセット全体を概観する。

### 4.3 データセットの概観

作成した11データセットを概観する。11データセットそれぞれについてユーザ数、問題数、回答ログ数とそれらの関係性を表4.1に整理した。それについて順に説明していく。

まず、ユーザ数について、全体として、数千以上のユーザのログデータに基づいたものとなっており、最も差の大きい2つのデータセット（中学1年の数学と中学3年の数学）間で差は2.6倍程度である。特に、ユーザ数が多いのは7,000ユーザ以上の中学1年の数学で、それに続いて、6,000ユーザ以上の中学の地理、歴史である。

次に、問題数について、全体として、数十から数百の問題に基づいたものとなっており、最も差の大きい2つのデータセット（小学4年の社会と中学1年の数学）

表 4.1: 11 データセットの統計量

学年	科目	ユーザ数	問題数	回答ログ数	回答ログ数 ÷ ユーザ数	回答ログ数 ÷ 問題数
小学 4 年	社会	3,045	76	227,409	75	2,992
	算数	4,318	182	505,917	117	2,780
小学 5 年	社会	2,833	197	388,521	137	1,972
	算数	3,380	257	411,957	122	1,603
小学 6 年	社会	2,891	202	434,324	150	2,150
	算数	3,225	245	395,276	123	1,613
中学 1 年	数学	7,137	365	659,237	92	1,806
中学 2 年	数学	3,931	278	238,241	61	857
中学 3 年	数学	2,667	343	177,295	66	517
中学	地理	6,499	308	660,882	102	2,146
	歴史	6,381	364	853,419	134	2,345

間で差は 4.8 倍程度である。小学生のデータセットよりは中学生のデータセットの方が問題数が大きい傾向にあり、特に、問題数が多いのは 360 以上の中学 1 年の数学と中学歴史である。

回答ログ数については、全体として、数十万以上であり、最も差の大きい 2 つのデータセット（中学歴史と中学 3 年の数学）間で差は 4.8 倍程度である。特に回答ログ数が最も大きい中学歴史のデータセットは 85 万以上と大規模な問題回答ログデータからなる。

さらに、これら 3 つの統計量の関係性を捉えるために、ログの密度という観点からユーザ 1 人あたりの平均回答ログ数と問題 1 問あたりの回答ログ数の 2 つの指標を評価した。まず、ユーザ 1 人あたりの平均回答ログ数について、全体として、数十から百数十となっており、最も差の大きい 2 つのデータセット（小学 6 年社会と中学 2 年数学）間で差は 2.5 倍程度である。また、いずれのデータセットでも平均回答ログ数は問題数を下回っている。

次に、問題 1 問あたりの平均回答ログ数について、全体としては、数百から数千程度の値となっており、最も差の大きい 2 つのデータセット（小学 4 年社会と中学 3 年数学）間で差は 5.8 倍程度である。特に、中学 2 年数学と中学 3 年数学が小

さい傾向にある。平均回答ログ数が最も大きい小学4年社会のデータセットは1問あたり3,000近いログデータが存在する。

以上、改めてデータセット間の各統計量について整理すると、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。また、いずれの統計量もデータセット間においても10倍以上の開きはなく、特に、ログ数の偏りによる今後の議論の影響は小さいと考えられる。

以上、11データセット全体を概観した。次に、個々のデータセットを具体的に説明する。

#### 4.4 個々のデータセットの具体的説明

本論文の対象である11データセットについて、それぞれのデータセットの性質について1つずつ述べる。具体的には、問題群の概要、問題とその回答選択肢の具体例、問題ごとの何番目に着手されるかの平均値と回答ログ数の関係の3つである。各問題について何番目に着手されるかの平均値は各ユーザごとに当該講座の各問題を何番目に着手したかを算出し、それらをユーザ全体で平均した値を用いる。各問題について何番目に着手されるかの平均値と回答ログ数の関係を見ることで、ユーザが講座の問題全体を均一に着手しているのか、もしくは、一部に偏っているのかを捉えることができる。XYプロットとして可視化し、図のデータ点はそれぞれひとつの問題に相当する。

##### 小学4年社会

小学4年社会では、主に、警察、消防、浄水場、ダム、汚水処理、発電所、ゴミ処理、リサイクル、水不足、地図の見方などの社会基盤や社会問題に関する基本内容が扱われる。図4.5に、問題と回答選択肢の例を示す。図の問題は社会ルールのうち道路交通法に関する問題であり、その回答を選択肢のなかから選択する、という回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図4.6に示す。全体として、概ね1500件以上の回答ログ数が

問9 次の( )に入ることばを後からえらび番号で答えなさい。

自転車や自動車などの交通のきまりをさだめたものを( )交通法といいます。

1 道路  
2 自動車  
3 自転車

解答 →  1  2  3

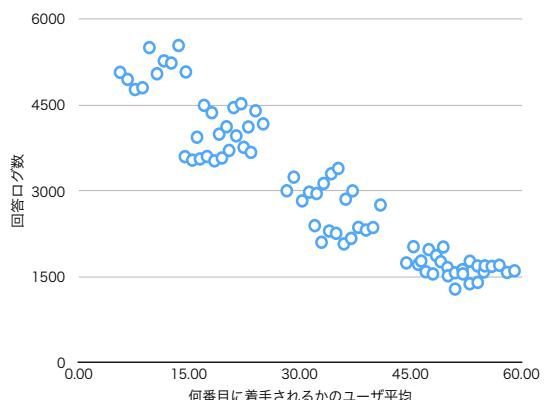


図 4.5: 小学 4 年社会の問題と回答選択肢の例

図 4.6: 小学 4 年社会の平均着手順と回答ログ数の XY プロット

あり、最初の問題の方が回答ログが大きい傾向にあるが大きく偏っているというわけではない。

### 小学 4 年算数

問5 次の計算をくふうしてする式を考えました。ふさわしいやり方をあとから選び、番号で答えましょう。

$99 \times 9$

①  $99 \times (10 + 1) = 1089$   
 ②  $99 \times 9 = (100 - 1) \times 9 = 100 \times 9 - 1 \times 9 = 891$   
 ③  $99 \times 9 = 10 \times 9 \times 9 = 10 \times 81 = 810$

解答 →  1  2  3

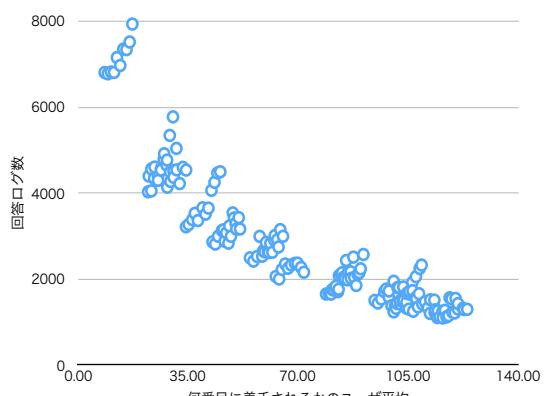


図 4.7: 小学 4 年算数の問題と回答選択肢の例

図 4.8: 小学 4 年算数の平均着手順と回答ログ数の XY プロット

小学 4 年算数では、主に、大きな数や小数など桁の認識、四則演算、数直線、折れ線グラフ、角度、分度器、垂直、並行、面積、単位、などの数理演算の基本内容が扱われる。図 4.7 に、問題と回答選択肢の例を示す。図の問題は分配法則に関する問題であり、その回答を選択肢のなかから選択する、という回答形式である。

次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.8 に示す。全体として、概ね 1500 件以上の回答ログ数があり、最初の問題の方が回答ログが大きい傾向にあるが大きく偏っているというわけではない。

### 小学 5 年社会

問8 次の（ ）に入ることばを番号で答えなさい。

インターネットを利用するとき、（ ）をぬすまれないように注意しなければならない。

1 個人情報  
2 パソコン  
3 電気

解答  1  2  3

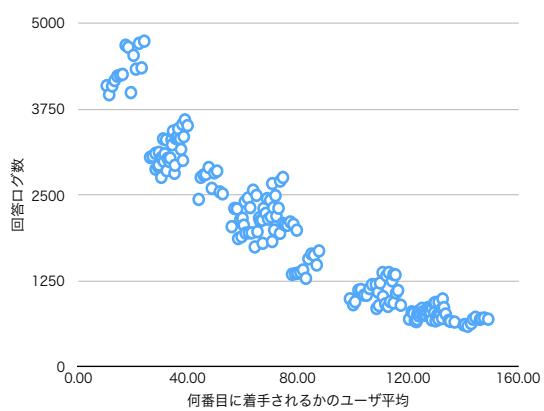


図 4.9: 小学 5 年社会の問題と回答選択肢の例

図 4.10: 小学 5 年社会の平均着手順と回答ログ数の XY プロット

小学 5 年社会では、主に、首都、地球儀、緯度経度、山地山脈、気候、平野、海流、特産物、農林水産業、輸出入、自給率、自動車などの社会基盤や社会問題に関する基本内容が扱われる。図 4.9 に、問題と回答選択肢の例を示す。図の問題は社会問題のうち近年しばしば議論されている個人情報とインターネットに関する問題であり、その回答を選択肢のなかから選択する、という回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.10 に示す。全体として、概ね 500 件以上の回答ログ数があり、回答ログの分布については小学 4 年の社会と概ね同じである。

### 小学 5 年算数

小学 5 年算数では、主に、整数、約数、倍数、平行四辺形、内角、外角、展開図、円、柱、などの基本内容が扱われる。図 4.11 に、問題と回答選択肢の例を示す。図の問題は小数の割り算に関する問題であり、その回答を選択肢のなかから

問1 次のア～ウにあてはまる数を答えましょう。

0.8mの長さが9.6kgのはりがね(あ)と1.2mの長さが9.6kgのはりがね(い)があります。はりがね(あ)の1m当たりの重さは  
 $9.6 \div 0.8 = \boxed{12}$ kgになります。  
 ここで、1より小さい小数でわるとわられる数より大きくなり、1より大きい小数でわるとわられる数より小さくなることがわかります。

解答 ア ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9  
 イ ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9  
 ウ ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

図 4.11: 小学 5 年算数の問題と回答選択肢の例

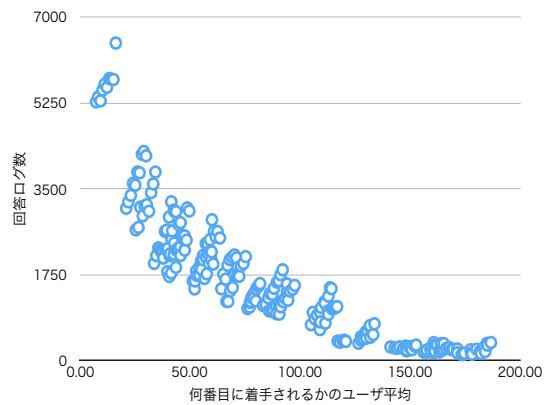


図 4.12: 小学 5 年算数の平均着手順と回答ログ数の XY プロット

選択する、という回答形式である。特に、回答の桁数が指定されているだけで、回答欄への自由記述に近い回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.12 に示す。最後の方の問題他の問題と比べると回答ログ数がかなり小さい。

## 小学 6 年社会

問5 次の(ア)(イ)にあてはまることばの組み合わせとして正しいものを番号で答えなさい。

1641年に平戸のオランダ商館を移しました。この結果、外国との交渉は(ア)が独占する(イ)が完成しました。

1 ア 藩 イ 鎮国  
 2 ア 幕府 イ 開国  
 3 ア 幕府 イ 鎮国  
 4 ア 朝廷 イ 鎮国

解答 ○ 1 ○ 2 ○ 3 ○ 4

図 4.13: 小学 6 年社会の問題と回答選択肢の例

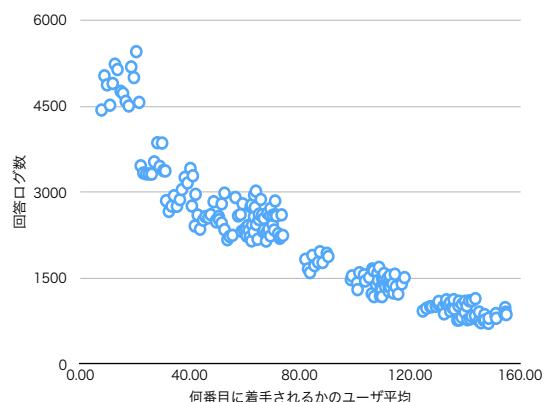


図 4.14: 小学 6 年社会の平均着手順と回答ログ数の XY プロット

小学 6 年社会では、主に、旧石器・縄文・弥生・古墳時代から昭和時代までの幅広い期間の歴史に関する内容が扱われる。図 4.13 に、問題と回答選択肢の例を示

す。図の問題は江戸時代の鎖国に関する問題であり、その回答を選択肢のなかから選択する、という回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.14 に示す。小学 4 年、小学 5 年の社会と同様に、概ね一定の割合で後半の問題の回答ログ数が減っていっていることがわかる。

## 小学 6 年算数

問7 次のア～エにあてはまる数を答えましょう。

下の図は底面の直径が 8cm、高さが 10cm の円柱をななめに半分に切り落としたものです。



この立体の体積は [ア] [イ] cm<sup>3</sup> です。

ただし、円周率は 3.14 とします。

**解答**

ア ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
イ ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
ウ ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
エ ○ + ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

図 4.15: 小学 6 年算数の問題と回答選択肢の例

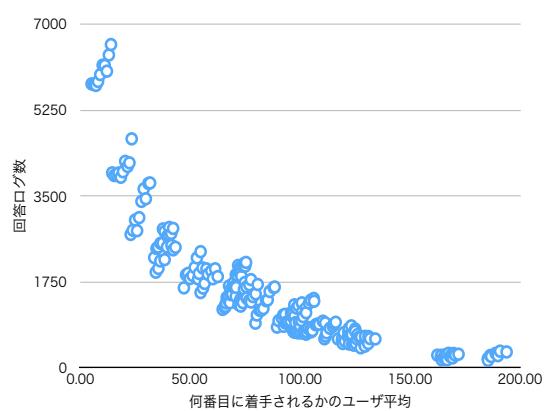


図 4.16: 小学 6 年算数の平均着手順と回答ログ数の XY プロット

小学 6 年算数では、主に、ならべ方と組み合わせ方、倍と割合、円の面積、分数のかけ算、分数のわり算、小数と分数の計算、拡大図と縮図、文字と式、比との利用、比例と反比例、点対称、立体の体積、線対称、資料の調べ方、速さ、量と単位の基本内容が扱われる。図 4.15 に、問題と回答選択肢の例を示す。図の問題は立体の体積のうち特に円柱の体積に関する問題であり、その回答を選択肢のなかから選択する、という回答形式である。特に、回答の桁数が指定されているだけで、回答欄への自由記述に近い回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.16 に示す。小学 4 年、小学 5 年の算数と比べ、分布が下に凸の形になっており、最初の方の問題への回答ログの偏りが大きいことが分かる。

問3 次の各数量を文字式で表したとき、ア、イにあてはまる数を、  
との①～⑥から選びなさい。

4kgのみかんを  $a$  人で分けたときの1人分の重さ

$\frac{\text{イ}}{\text{ア}}$  (kg)

① 2 ② 3 ③ 4 ④  $a$  ⑤  $b$  ⑥  $c$

解答 ア ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6  
イ ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

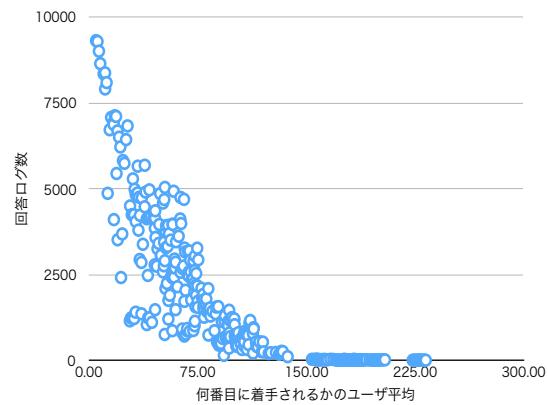


図 4.17: 中学 1 年数学の問題と回答選択肢の例

図 4.18: 中学 1 年数学の平均着手順と回答ログ数の XY プロット

## 中学 1 年数学

中学 1 年数学では、主に、1 次式、1 次方程式、2 平面の関係 面の動き、代入と式の値、作図のしかた、円、反比例、図形の移動、対称移動、平面上の 2 直線、度数の分布、座標、数の集合と四則計算、文字式の活用、方程式、正負の数の利用、比と比例式、立体の体積、角錐と円錐などに関する内容が扱われる。図 4.17 に、問題と回答選択肢の例を示す。図の問題は文字式を活用した問題であり、その回答を選択肢のなかから選択する、という回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.18 に示す。回答ログは最初の方の問題に大きく偏っており、特に後半の一部は数十程度のログしかないことが分かる。

## 中学 2 年数学

中学 2 年数学では、主に、1 次方程式、1 次関数、三角形の合同条件、二等辺三角形、仮定と結論、単項式と多項式、合同な三角形、平行四辺形、証明、連立方程式などを扱う。扱う内容は中学 1 年数学のデータセットのものと概ね同様であるが、難易度が高い。図 4.19 に、問題と回答選択肢の例を示す。図の問題は数連立方程式に関する問題であり、その回答を選択肢のなかから選択する、という回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.20 に示す。グラフの外見は中学 1 年数学に近いが、回答ログ数の

**問1** 連立方程式の解き方を説明したものについて、□にあてはまる数を答  
えなさい。

$$\begin{cases} 2x-3y=3 & \cdots ① \\ x+2y=5 & \cdots ② \end{cases}$$

この連立方程式を解くのに、 $x$  の係数をそろえることを考える。②の方程  
式を2倍すると、 $x$  の係数は2になる。

したがって

$$\begin{cases} 2x-3y=3 & \cdots ① \\ 2x+4y=10 & \cdots ③ \end{cases}$$

となるので、③から①を引くと  $\boxed{2y}=10-3=7$  になるので、  
 $y=\boxed{1}$  、 $x=\boxed{2}$  と求められる。

**解答**

ア	○ -	○ +	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
イ	○ -	○ +	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
ウ	○ -	○ +	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9

図 4.19: 中学2年数学の問題と回答選択肢の例

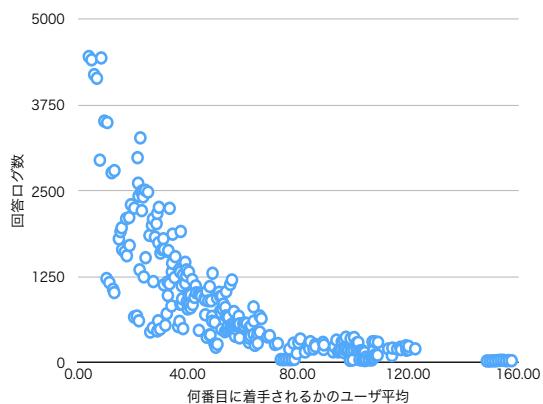


図 4.20: 中学2年数学の平均着手順と回答ログ数のXYプロット

スケールが半分程度である。また、一部の問題については、ログが数十程度と他  
と比べると非常に少なくなっている。

### 中学3年数学

**問4** 次の△ABC と相似ではない三角形を、あと①～③から選びなさい。

①      ②      ③

解答	○ 1	○ 2	○ 3
----	-----	-----	-----

図 4.21: 中学3年数学の問題と回答選択肢の例

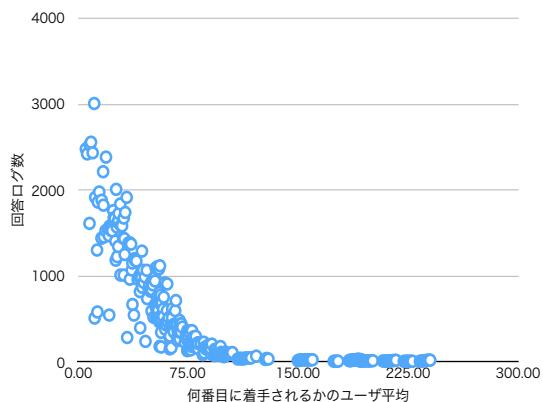


図 4.22: 中学3年数学の平均着手順と回答ログ数のXYプロット

中学3年数学では、たすき掛け、一次関数、三平方の定理、中点連結定理、乱数表、二次方程式、二次関数、円の面積、円周角、分配法則、回転体の体積、因数分解、増加関数、変域、外接円、媒介変数、展開、平方完成、平方根、平行線、循環小数、指数、接線の定義、放物線、文字式、有理数、根号、標本調査、母集団、

減少関数，無理数，相似の三角形，相似比，立方体，等差数列の和，約数，素因数分解，素数，解と係数の関係などの内容が扱われる。図 4.21 に，問題と回答選択肢の例を示す。図の問題は図形のうち相似の三角形に関する問題であり，その回答を選択肢のなかから選択する，という回答形式である。次に，各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.22 に示す。グラフの外観は中学 2 年数学に類似しているが，後半の問題の回答ログ数は非常に少ない。

### 中学地理

問2 次の国々の中で南ヨーロッパに含まれる国を1つ選び，番号で答えなさい。

1 アイスランド  
2 デンマーク  
3 スウェーデン  
4 ギリシャ

解答 →  1  2  3  4

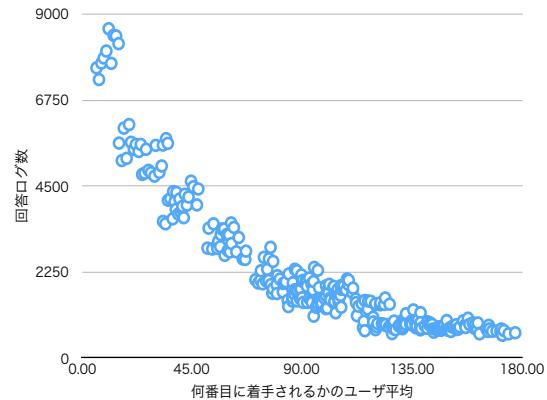


図 4.23: 中学地理の問題と回答選択肢の例

図 4.24: 中学地理の平均着手順と回答ログ数の XY プロット

中学地理では，主に，アジア，アフリカ，オセアニア，ヨーロッパ，北アメリカ，南アメリカ，世界と日本の関係，世界の地域区分と特色，世界の衣食住・宗教，地域の調査，日本の地域区分，日本の工業と商業・サービス業，日本の農林水産業，東北地方，近畿地方，関東地方，中部地方，九州地方，北海道地方，中国・四国地方などが地域や地域の性質，地域間の関係性が扱われる。図 4.23 に，問題と回答選択肢の例を示す。図の問題はヨーロッパに関する問題であり，その回答を選択肢のなかから選択する，という回答形式である。次に，各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.24 に示す。全体として，概ね 500 以上の回答ログ数があり，前半の問題から後半の問題にかけてほぼ線形に回答ログ数が減少している。

## 中学歴史

問7 次の( )にあてはまることばを後から選び番号で答えなさい。

( )は、「モナ＝リザ」や「最後の晩餐」などの絵画を生んだほか、彫刻や建築、数学、生物学などでも業績を残した。

1 レオナルド＝ダ＝ヴィンチ  
2 ミケランジェロ  
3 ラファエロ

解答  1  2  3

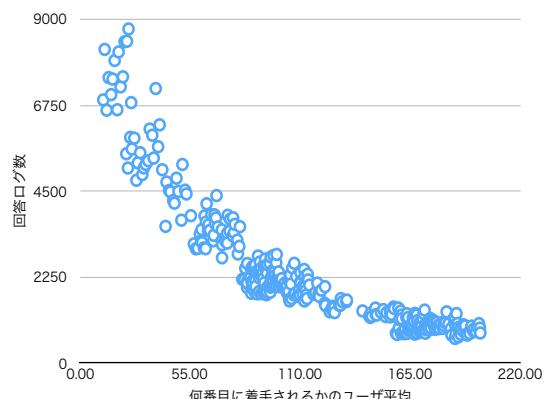


図 4.25: 中学歴史の問題と回答選択肢の例

図 4.26: 中学歴史の平均着手順と回答ログ数の XY プロット

中学歴史では、主に、ルネサンスと大航海時代、世界恐慌と国際情勢の悪化、世界文明の発生、国際協調体制、第一次世界大戦とロシア革命、欧米の市民革命と産業革命、旧石器・縄文・弥生時代、安土桃山時代、大和時代、奈良時代、平安時代、鎌倉時代、室町時代、江戸時代、明治時代、昭和時代、大正時代、平成時代、など内容が扱われる。図 4.25 に、問題と回答選択肢の例を示す。図の問題はルネサンスと大航海時代に関する問題と推察される。その回答を選択肢のなかから選択する、という回答形式である。次に、各問題について何番目に着手されるかの平均値と回答ログ数の関係を図 4.26 に示す。全体として、概ね 500 以上の回答ログ数があり、前半の問題から後半の問題にかけてほぼ線形に回答ログ数が減少している。外観は中学地理と概ね合致している。

以上、個々のデータセットおよびデータが収集された講座を具体的に説明した。具体的な内容から、地理と歴史に関する 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）のデータセットがどちらかというと、宣言的知識の獲得を主目的としているデータセットであり、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）のデータセットがどちらかというと、手続き的知識の獲得を主目的としているデータセットであることが分かる。

また、データセットの一部の問題については他の問題と比べてログ数が数十程

度と少なかったが，それでも数十程度のログデータがあり，そうした問題が一部に限られていることから知識構造を抽出できると考え，このまま全ての問題をデータセットに利用する。

以上，データセットについて述べた。次章では，実験について述べる。

# 第5章 実験

本章では、実験について述べる。まず、データセットへの適合のための Deep Knowledge Tracing の拡張法を述べたのち、次に、実験設定について述べ、その後、実験結果について述べる。実験結果においては、まず、11 データセットのいずれのデータセットにおいても、Deep Knowledge Tracing の予測性能が知識間関係を考慮しない Bayesian Knowledge Tracing よりも高いことを示し、Deep Knowledge Tracing 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する。さらに、11 の知識間関係ネットワークを可視化し、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する。そして、宣言的知識の知識構造を表現する 5 つのネットワークと手続き的知識の知識構造を表現する 6 つのネットワークのモジュラリティとフロー階層、GRC を算出し、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す。

## 5.1 Deep Knowledge Tracing の拡張および最適化

データセットへの適合のための Deep Knowledge Tracing の拡張法を述べ、その後その最適化について述べる。勉強サプリの問題演習はテスト形式で提供されており、一度に 1 つ以上の問題が提供され、また、それらの採点も同時に行われるため、同時に複数の問題の回答ログデータが発生する可能性があり、同時に 1 つの問題しか回答ログデータが発生しないという Deep Knowledge Tracing で想定されている設定と異なる。そのため、Deep Knowledge Tracing をデータセットに適合するように拡張する。

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った。

同時に複数の問題について回答ログが生じることを許容する場合, Knowledge Tracing のタスクは学習者の時刻  $t$  において観測された問題回答結果ベクトルを  $\mathbf{q}_t$  とすれば,  $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_t$  から時刻  $t + 1$  において観測される問題回答結果ベクトル  $\mathbf{q}_{t+1}$  を予測するというものである。

Deep Knowledge Tracing では RNN への入力は one-hot ベクトルに符号化され, 入力ベクトルを  $\mathbf{x}_t$ , 演習問題の数を  $M$  とすれば,  $\mathbf{x}_t$  の長さは  $2M$  であった。同時に複数の問題が回答されるという設定では, その同時に回答された問題数を  $m_t$  とすれば  $m_t$ -hot ベクトルに符号化することで, [Piech et al., 2015] に近い形で RNN に情報を入力できる。

表 5.1: 拡張した 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

ユーザ ID	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
A	1	1	1		
A	1	2	0	$\mathbf{x}_1$	[1000:0100]
-----				-----	
A	2	1	1		
A	2	2	1	$\mathbf{x}_2$	[1100:0000]
-----				-----	
A	3	3	0	$\mathbf{x}_3$	[0000:0010]
-----				-----	
A	4	3	0		
A	4	4	1	$\mathbf{x}_4$	[0001:0010]

具体例を交えて説明する。例えば, 演習問題の数が 4 つと仮定する。 $M = 4$  であり,  $\mathbf{x}_t$  の長さは 8 である。ある学習者が, 表 5.1 の回答ログのように問題を回答し正誤が観測されたとすれば, 入力ベクトル  $\mathbf{x}_t$  は表 5.1 の入力ベクトルのように符号化される。

出力  $\mathbf{y}_t$  は Deep Knowledge Tracing と全く同じであり, 問題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。したがって,  $t + 1$  の回答  $\mathbf{q}_{t+1}$  の正誤予測は  $t + 1$  に回答さ

れる問題  $\mathbf{q}_{t+1}$  に対応する  $\mathbf{y}_t$  の要素から読み取れる。

最適化手法は [Piech et al., 2015] のものとほとんど同じである。

$$\log(p_1 \times p_2 \times \cdots \times p_{m_t}) = \sum_k^{m_t} \log(p_k) \quad (5.1)$$

であるため、 $\tilde{\delta}(\mathbf{q}_{t+1})$  を時刻  $t + 1$  にどの問題が回答されたかの  $m_t$ -hot ベクトルとし、 $\mathbf{a}_{t+1}$  を時刻  $t + 1$  に対応する問題で正答したか否か（1か0）のベクトルとすれば、ロス関数は

$$L = \sum_t l(\mathbf{y}_t^T \tilde{\delta}(\mathbf{q}_{t+1}), \mathbf{a}_{t+1}) \quad (5.2)$$

である。学習時は [Piech et al., 2015] と同様にミニバッチごとに確率的勾配降下法で目的関数を最小化する。

ハイパーパラメタについては、対象データが多いため学習コストの削減を狙い RNN の部分には GRNN を用いる。学習率の初期値を 200、モーメントを 0.98、1 エポックごとに、減衰率 0.8 として学習率を最小学習率 10 まで減衰させる。また、勾配のノルムの最大値を 0.00001 として [Pascanu et al., 2013] に従い勾配に制約を設けた。dropout は [Piech et al., 2015] と同様に  $\mathbf{y}_t$  の方向にのみかけ、dropout 率は 0.5 とした。隠れ層のユニット数は 200 として、各重み行列の初期化は [Glorot and Bengio, 2010] にしたがった。時系列方向の誤差逆伝搬は最長で 200 まで伝搬するように制約を設けた。

これらのハイパーパラメタは実験的に高い予測性能を発揮したため設定しており、網羅的に探索したわけではない。通常、深層学習の手法はハイパーパラメタの数が非常に大きく、また、計算コストが大きいため大規模な探索は行えない。Grid Search や Random Search [Bergstra and Bengio, 2012] といった探索手法が提案されているが、専門家が手で調整した方が優れていることが報告されている [Larochelle et al., 2007, Bergstra and Bengio, 2012]。

実装には Theano を用いた [Bergstra et al., 2010, Bastien et al., 2012]。Theano は多次元行列を含む数学的表現の定義や計算、最適化を効率的に行える Python のライブラリで、深層学習の研究ではよく利用される。

以上、Deep Knowledge Tracing の拡張およびその最適化について述べた。次に、

実験設定について述べる。

## 5.2 設定

ここでは、特に Deep Knowledge Tracing による知識獲得の予測の実験設定、および、構築したネットワークの可視化の設定、ネットワーク構造指標の計算について述べる。

まず、Deep Knowledge Tracing 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、訓練ユーザのデータでモデルを構築し、検証ユーザのデータでハイパーパラメタを調整し、検証ユーザのデータで精度が最も高かったモデルをテストユーザのデータに適用し当該モデルの最終的な精度とする。小学生や中学生が学習する内容は国が定める学習指導要領 [文部科学省, 2011] に従っており、学習内容が高い頻度で入れ替わっていくというわけではないため、勉強サプリの学習教材もまた、更新頻度は高くない。したがって、Knowledge Tracing のモデルは商品推薦や広告推薦の領域でしばしば問題となるアイテムに関するコールドスタート問題 [Schein et al., 2002] の影響を受けにくく、こうして同じ時期のユーザを分割する評価方法でもモデルの性能を十分評価できると考えられる。

次に、構築したネットワークの可視化の設定およびネットワーク構造指標の計算について述べる。ネットワークの可視化には Gephi [Bastian et al., 2009] を用いた。Gephi はネットワーク可視化用のソフトウェアであり、豊富なレイアウト手法やネットワークの統計量計算ツールが提供されている。ノードは問題を、エッジは影響関係を示すように可視化する。ノードの色はデータセットの説明の際に用いた平均問題着手番号について、番号が小さいものから大きいものにかけて青色から赤色になるようにする。ノードの大きさはノードの出力次数に比例するよう可視化する。ノードのラベルはそのノードが指す問題が提供される授業名とその問題に割り振った適当な数字を結合した文字列を利用する。また、一見してクラスタだと判断できそうなノード集合には、ノードやその結合が分かりにくくならない範囲で扱われている内容の説明を近くに記載する。ネットワークのレイアウト手法には、ForceAtlas [Jacomy, 2009] を用いる。また、モジュラリティの計算

は Gephi で行い，フロー階層と GRC の計算は NetworkX [Hagberg et al., 2008] というネットワーク解析や探索及びネットワークアルゴリズムを提供する Python のライブラリを用いた。

以上，実験設定について述べた。次に，実験結果について述べる。

## 5.3 結果

実験結果について述べる。まず，11 データセットのいずれのデータセットにおいても，Deep Knowledge Tracing の予測性能が知識間関係を考慮しない Bayesian Knowledge Tracing よりも大きく高いことを示し，Deep Knowledge Tracing 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する。さらに，11 の知識間関係ネットワークを可視化し，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する。そして，宣言的知識の知識構造を表現する 5 つのネットワークと手続き的知識の知識構造を表現する 6 つのネットワークのモジュラリティとフロー階層，GRC を算出し，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す。

### 5.3.1 各データセットにおける予測性能

Deep Knowledge Tracing (以下，DKT) と Bayesian Knowledge Tracing (以下，BKT) を 11 データセットに適用した結果を表 5.2 に示す。Marginal は各問題についてそれぞれ正解の周辺確率を予測結果とするものである。[Piech et al., 2015] にも記載されていたため，本稿でも同様にベースラインとして記載した。また，値が大きい箇所は太字で記載した。

まず，すべてのデータセットにおいて DKT の AUC は BKT の AUC よりも大きく高かった。また，DKT–BKT の値の平均も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でほとんど差は

表 5.2: 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

分類	データセット		AUC			DKT – BKT	
	科目	学年	Marginal	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	小学4年	0.701	0.739	0.791	0.052	
		小学5年	0.647	0.680	0.765	0.085	
		中学	0.655	0.681	0.764	0.083	<b>0.086</b>
	歴史・社会	小学6年	0.642	0.657	0.773	<b>0.116</b>	
		歴史	0.635	0.670	0.766	0.096	
		小学4年	0.696	0.707	<b>0.828</b>	<b>0.121</b>	
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	小学5年	0.714	0.724	<b>0.804</b>	0.081	
		小学6年	0.734	0.749	<b>0.836</b>	0.087	
		中学1年	0.738	0.750	<b>0.807</b>	0.057	
		中学2年	0.675	0.696	0.773	0.077	
		中学3年	0.729	0.735	<b>0.804</b>	0.069	
							<b>0.082</b>

なかった。つまり、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とともに、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と考えられる。

次に、特に、算数や数学の 6 つのデータセットにおいて、中学2年のデータセットを除く 5 つのデータセットにおいて DKT の AUC は 0.8 を超えており、[Piech et al., 2015] で示されたように、確かに、DKT が数学のデータセットに対して有効であった。地理や歴史の社会関係のデータセットにおいても DKT の AUC は 0.76 を超えており、DKT が地理や歴史の社会関係のデータセットに対して有効であり、宣言的知識の獲得を主目的とするデータセットに対しても有効であることが示された。

以上、11 データセットのいずれのデータセットにおいても、Deep Knowledge Tracing 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確認した。

### 5.3.2 抽出した知識間関係ネットワークの可視化

11 の知識間関係をネットワークとして可視化し、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、これらのネットワークが、知

識獲得における知識構造を表現しているものであることを定性的に確認する。

まず、宣言的知識について知識間関係ネットワークを順に可視化していく。

### 小学4年社会

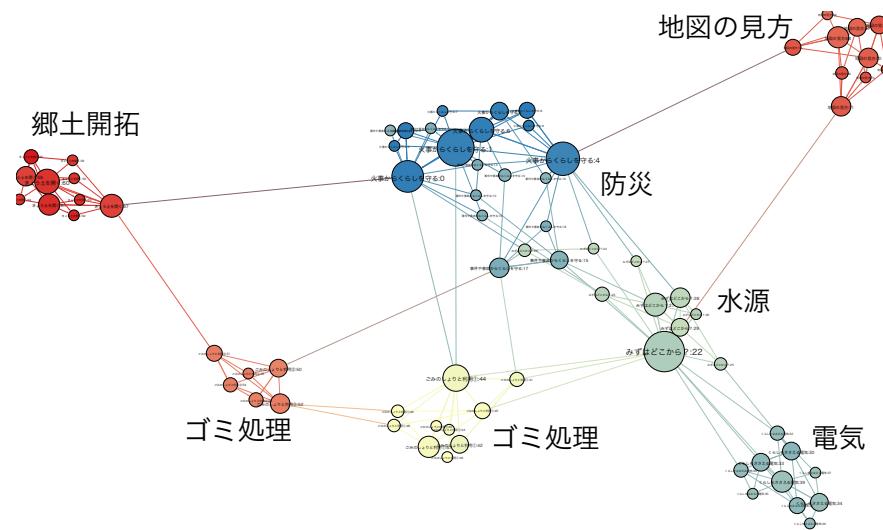


図 5.1: 小学4年社会の知識間関係ネットワーク

小学4年社会のネットワークを図5.1に示す。ノードは内容ごとでクラスタを形成していることが分かる。クラスタ間の関係は弱い関係となっていそうである。特筆すべき点は、防災とラベル付けしたクラスタは火事に関するものと事故や事件に関するものが1つとなってできているということである。

### 小学5年社会

小学5年社会のネットワークを図5.2に示す。ノードは内容ごとでクラスタを形成している。特に、国土、北海道、沖縄のクラスタが近くにあることは、国土における地理関係において北海道と沖縄の関連が強いからだと考えられる。また、工業と近い環境問題と自動車と近い環境問題があることも、それぞれの文脈において環境問題があることと関連がある可能性がある。

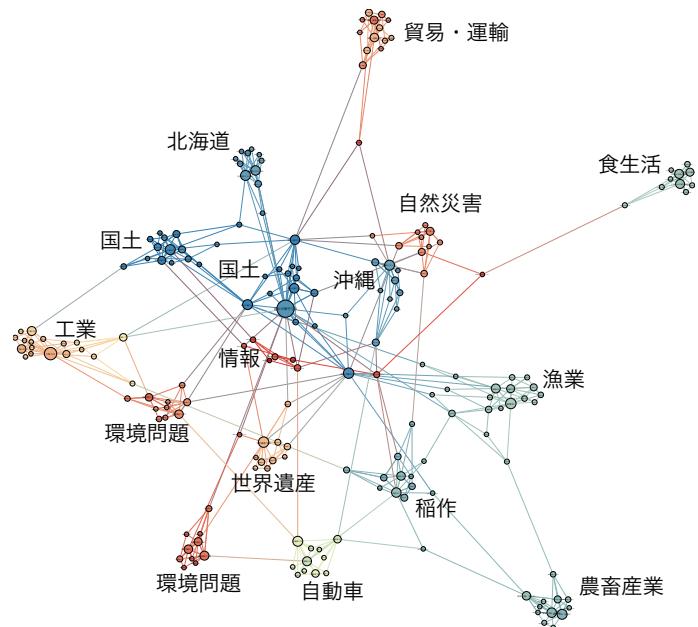


図 5.2: 小学 5 年社会の知識間関係ネットワーク

小学 6 年社会

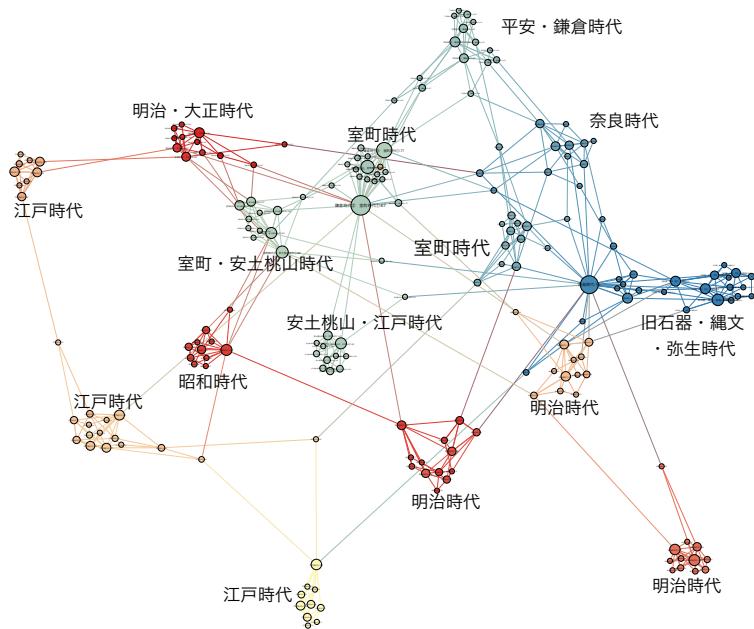


図 5.3: 小学 6 年社会の知識間関係ネットワーク

小学6年社会のネットワークを図5.3に示す。ノードは内容ごとでクラスタを形

成している。特に、クラスタは各時代を表現しているが、逆に、江戸時代や明治時代等の一部の時代は1つのクラスタを形成しているというわけではない。

中学地理

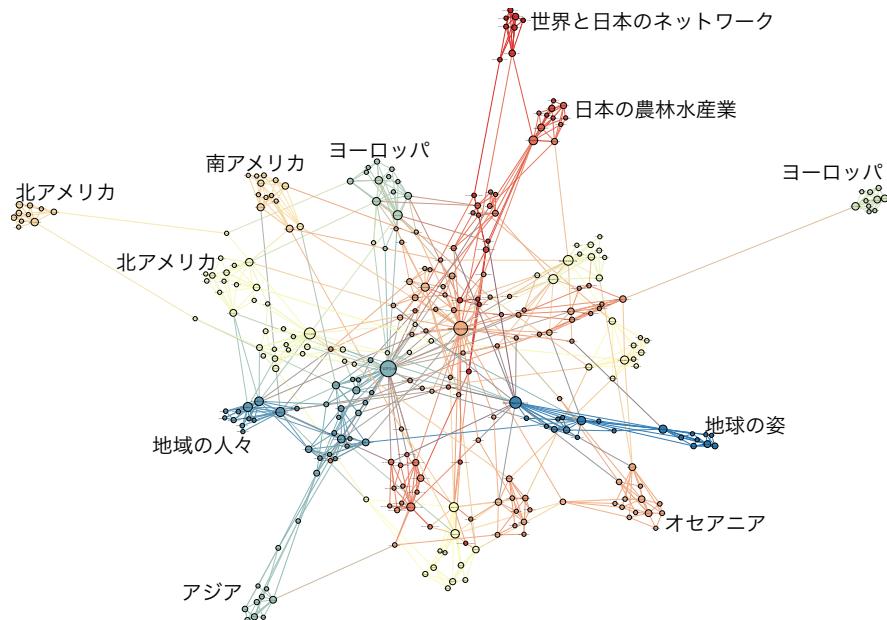


図 5.4: 中学地理の知識間関係ネットワーク

中学地理のネットワークを図5.4に示す。多くのノードが内容ごとでクラスタを形成している。特に、ヨーロッパ、南アメリカ、アジア、オセアニア等、世界の各地域の内容はそれぞれでクラスタを形成している。また、中央のオレンジから右上にかけて緩やかに結合しているノード群は日本あるいは日本と世界との関連についての内容となっている。

中學歷史

中学歴史のネットワークを図5.5に示す。多くのノードが内容ごとでクラスタを形成している一方で、中央にノードが雑多な形で集まっている。中央の雑多なノード群は主に安土桃山時代、江戸時代、ルネサンス時代、大航海時代、市民革命、産業革命等に関するものであるが、日本の安土桃山時代（1573年から1603年）、江

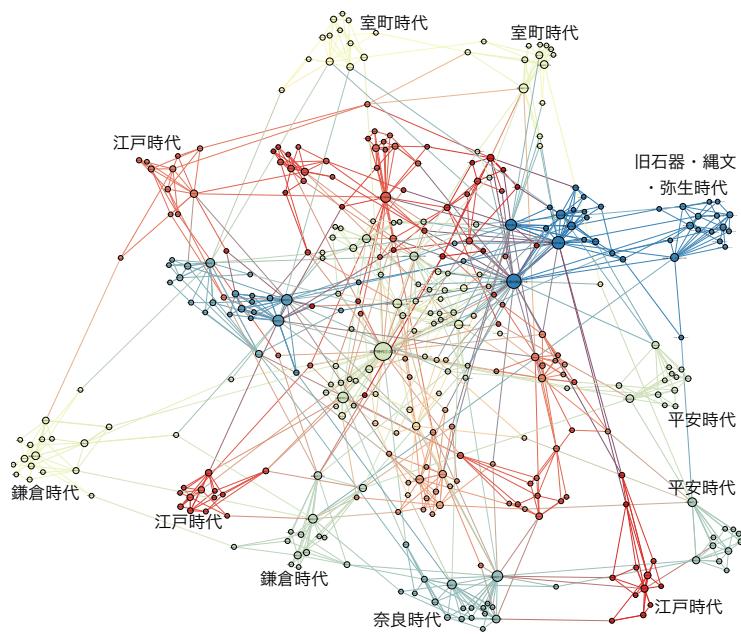


図 5.5: 中学歴史の知識間関係ネットワーク

戸時代（1603年から1867年）の年代と、欧米諸国のルネサンス時代（14世紀から16世紀）、大航海時代（15世紀から17世紀）、市民革命（18世紀）、産業革命（18世紀から19世紀）の年代が概ね合致していることが興味深い。

次に、手続き的的知識について知識間関係ネットワークを順に可視化していく。

#### 小学4年算数

小学4年算数のネットワークを図5.6に示す。一部のノードがクラスタを形成している。小学4年社会や小学5年社会のネットワークと比べると、クラスタの形成度合いは小さいように見られる。特に、青色のノードが中央に寄っていて、外側に出ているノードはオレンジや赤色のノードが多いことも1つの特徴であると考えられる。このことは、着手される順序が遅い問題の方がネットワークの中心になく、つまり影響を与える問題が少ないことを示唆している。

#### 小学5年算数

小学5年算数のネットワークを図5.7に示す。全体としてノードは複雑に結合しており、モジュール性は低そうである。また、ごく一部のノードが他のノードと

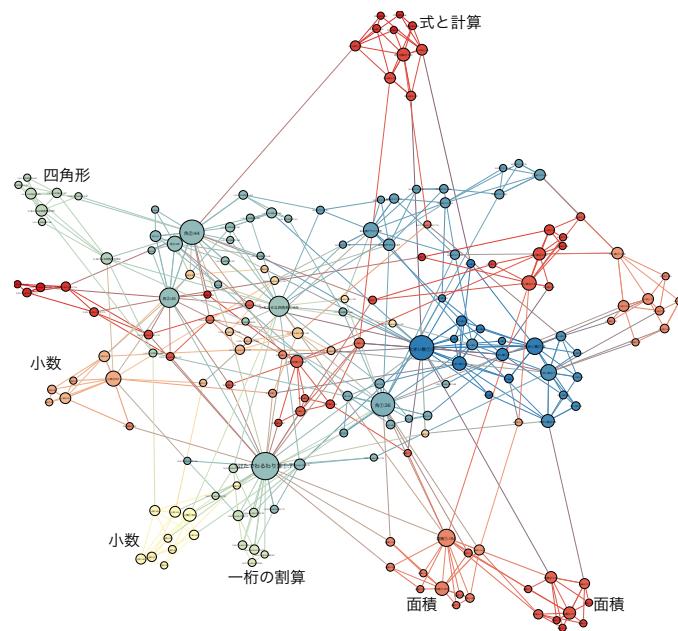


図 5.6: 小学 4 年算数の知識間関係ネットワーク

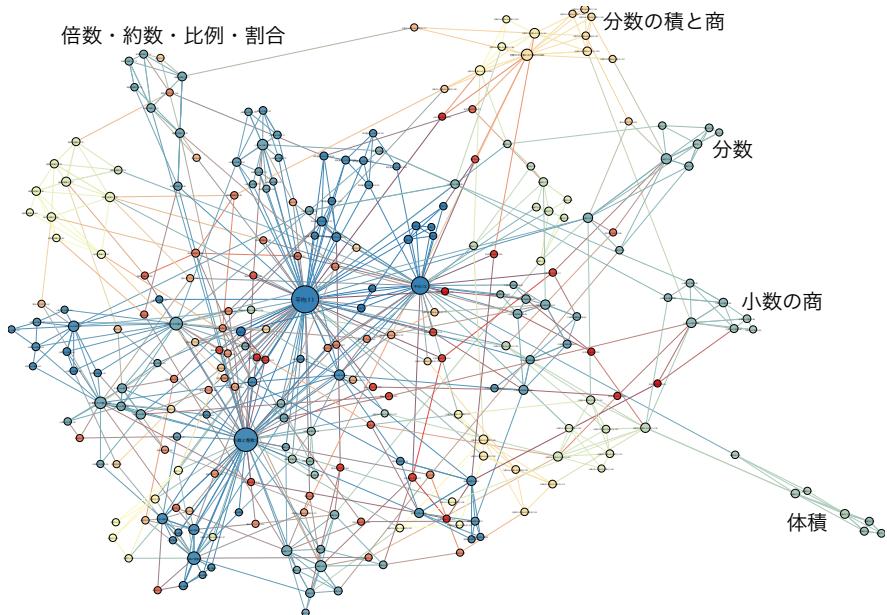


図 5.7: 小学 5 年算数の知識間関係ネットワーク

比べて大きく、つまり、出力次数が大きい。このことは、これらの大きいノードによって表現される知識は多くの知識の獲得に影響を与えていることを示唆している。特に、大きいノードは平均や小数、整数に関する問題で、かつ、着手順序が

早いものであることも大きな特徴である。クラスタという観点では、小さなクラスタを形成しているノード群は小数や分数、あるいは体積に関するノードであることが特徴的である。

### 小学 6 年算数

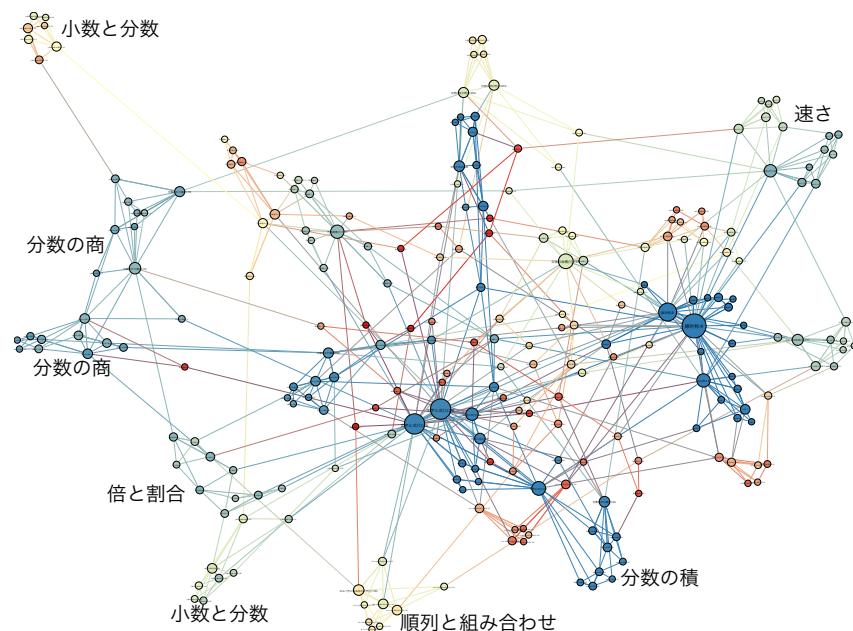


図 5.8: 小学 6 年算数の知識間関係ネットワーク

小学 6 年算数のネットワークを図 5.8 に示す。全体としてノードは複雑に結合しているが、一方で一部は小さなクラスタを形成している。特にクラスタを形成するノード群の内容は左のクラスタのノード群が小数や分数、それらの積や商の活用が求められる内容で、下のクラスタのノード群が順列と組み合わせに関する内容で、右上のクラスタのノード群が速さに関する内容となっている。また、中央のごく一部ノードが青色で他のノードと比べて大きいことも大きな特徴である。このことは、最初に獲得した知識が他の多くの知識の獲得に影響をもつことを示唆している。

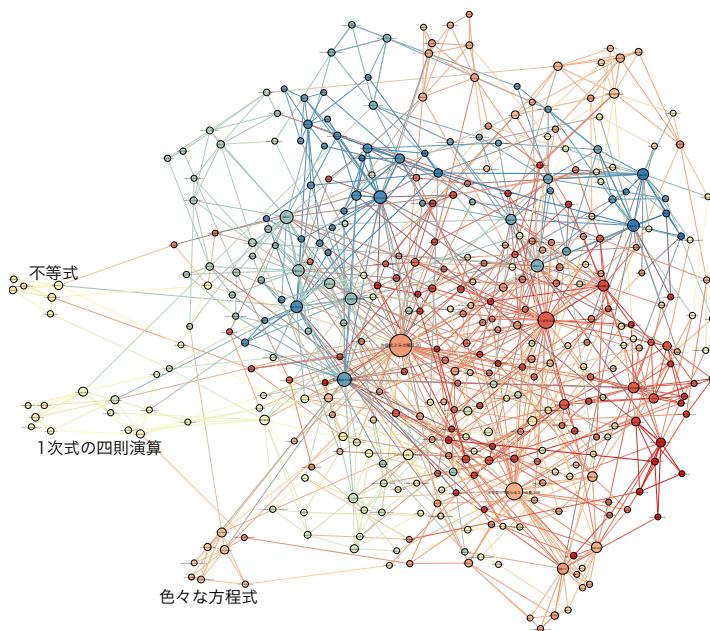


図 5.9: 中学 1 年数学の知識間関係ネットワーク

### 中学 1 年数学

中学 1 年数学のネットワークを図 5.9 に示す。全体としてノードは複雑に結合しており、小さなクラスタはほとんど存在しない。特筆すべき点は中央オレンジ色の「方程式とその解」の内容を扱うノードが非常に大きく、出力次数が大きいことである。このことは、方程式とその解に関する知識が中学 1 年数学の学習において非常に重要な位置付けにあることを示唆している。小さなクラスタを形成しているノード群は不等式、1 次式の四則演算、色々な方程式に関する内容を扱うものであった。

### 中学 2 年数学

中学 2 年数学のネットワークを図 5.10 に示す。全体としてノードは密に結合しているものの、一部で緩やかにクラスタを形成している。連立方程式の活用や 1 次関数の活用がクラスタとなっている。活用に関する知識は、ある種の特殊な事例を扱うためほかへの適用可能性が低く、結果としてモジュール性が高い可能性がある。また、右上の大きなノードが仮定や結論、証明に関する内容を扱うものであり、またその近傍にあるノード群が三角形の合同証明、2 線の並行証明等、証明

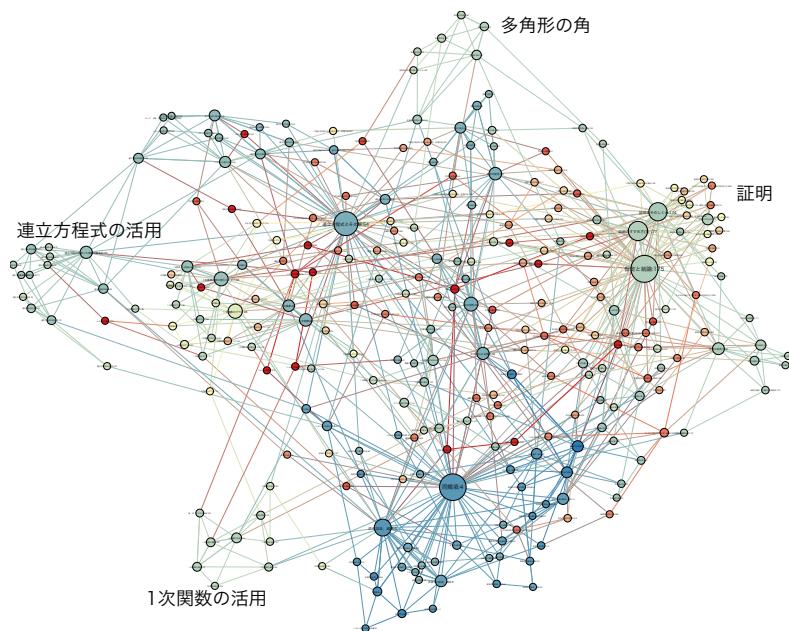


図 5.10: 中学 2 年数学の知識間関係ネットワーク

の活用事例に関する内容を扱うものであることも興味深い。証明という一般的な手続き的知識が合同や並行への理解に大きな影響を与えることを示唆している。

### 中学 3 年数学

中学 3 年数学のネットワークを図 5.10 に示す。全体としてノードは複雑に結合しており、一部が密になっているというようなクラスタはほとんど存在しない。中央の青色のノードが非常に大きく、出力次数が大きい。このノードは単項式と多項式の乗法除法に関する内容を扱うものであり、したがって、中学 3 年数学において単項式と多項式に乘法除法に関する知識は非常に重要な位置付けにあることを示唆している。

以上、個々の知識間関係ネットワークを可視化した。全体として、ノードは内容の関連度が大きいノードと結合しており、これらのネットワークは、確かに、知識獲得における知識構造を表現しているものと考えられる。

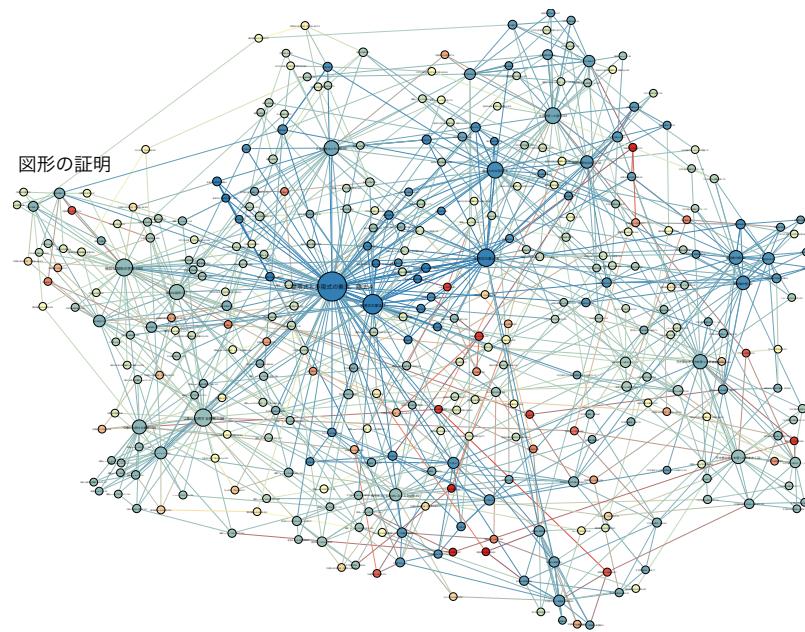


図 5.11: 中学 3 年数学の知識間関係ネットワーク

表 5.3: 各ネットワークにおける構造指標

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

### 5.3.3 ネットワーク構造の分析

宣言的知識の知識構造を表現する 5 つのネットワークと手続き的知識の知識構造を表現する 6 つのネットワークのモジュラリティとフロー階層、GRC を算出し、

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す。

表5.3に各ネットワークにおける各構造指標を整理した。まず、モジュラリティについて、いずれの宣言的知識の知識間関係ネットワークのモジュラリティもいずれの手続き的知識の知識間関係ネットワークのモジュラリティよりも高かった。また、特に、手続き的知識の知識間関係ネットワークにおいては、学年が高いほど、モジュラリティが低い傾向にあり、つまり、難易度が高いほど知識のモジュール性が低い傾向にあることを示唆している。逆説的に、難易度がより高い問題とは扱われる知識のモジュール性がより低いということを示唆しているとも考えられる。

次に、フロー階層について、手続き的知識の知識間関係ネットワークのフロー階層の平均は宣言的知識の知識間関係ネットワークのフロー階層の平均と比べ、高かった。このことは、手続き的知識の知識構造の方が宣言的知識の知識構造よりも循環構造が少ないことを示唆している。また、手続き的知識と宣言的知識に限らず、フロー階層は学年が高いほど高い傾向にあり、このことは、知識はその獲得の難易度が高いほど循環構造が少ない傾向にあるか、あるいは、知識獲得の順序性が強い傾向にあることを示唆している。

最後に、GRCについて、手続き的知識の知識間関係ネットワークのGRCの平均は宣言的知識の知識間関係ネットワークのGRCの平均と比べ、高かった。これはフロー階層と同様である。一方で、GRCにおいてはフロー階層で見られた学年の傾向は現れていないようである。

また、3つの指標において、小学4年算数の知識間関係ネットワークの構造は他の学年の算数や数学の知識間関係ネットワークの構造とは大きく異なる。具体的には、小学4年算数の知識間関係ネットワークのモジュラリティは高く、フロー階層とGRCは低かった。

さらに、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。

以上、実験について述べた。次章では、考察について述べる。

# 第6章 考察

本章では、実験結果を踏まえた考察を述べる。

## 6.1 データセットと知識構造

データセットと知識構造について考察する。データセットの作成に利用する問題回答ログデータの要件を述べる際、科目が同じでも難易度が高いと、目的志向の知識である手続き的知識の有無が問題回答の正誤に影響を与える可能性が高く、知識構造もより階層的になる可能性が高いことを説明し、データの要件として、問題群が難しすぎないことを挙げた。また、それを満足するデータの収集対象として勉強サプリを選択し、手続き的知識の獲得を主目的とする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6データセットを作成した。

しかし、この中でも、特に、小学4年算数のデータセットは難しすぎないという点では十分だったが、難易度が低すぎた、あるいは内容があまり適切でなかつた可能性がある。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、モジュラリティについては宣言的知識についての全てのネットワークのものより小さかったものの、その中で最も小さかった小学4年社会のものと大きな差があるというわけではなく、フロー階層およびGRCについては、宣言的知識についての全てのネットワークのものよりも小さい組み合わせが複数存在する。つまり、モジュール性が大きく階層性が小さかったということである。

このことは、小学4年算数が手続き的知識の学習を主目的としていないということ指しているのではないと考える。小学4年算数の問題例を示した図4.7は四則演算の問題であり、これは数字という宣言的知識への手続きを用いて達成されるものである。小学4年算数の知識間関係ネットワークを示した図5.6ではいくつか

のクラスタが抽出されている。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、したがって、ネットワーク全体としてはモジュール性が高く、また、階層性が低くなってしまっているのだと推察する。

## 6.2 知識構造分析手法の他 MOOCs への適用可能性

知識構造の分析手法の他 MOOCs への適用可能性について考察する。本研究で用いた知識構造の分析手法は 1) データセット作成、2) 知識間関係行列の算出、3) ネットワーク構築および指標による構造評価の 3 つの要素から構成されていた。MOOCs で収集される問題回答ログデータは大規模であると考えられ、また、ネットワーク構築および指標による構造評価は Deep Knowledge Tracing (以下、DKT) で知識間関係を抽出できるかに依存する。したがって、知識構造の分析手法の他 MOOCs への適用可能性は、DKT の適用可能性に依存すると考えられる。

そこで、DKT の予測性能について考察する。[Piech et al., 2015] では、数学に関するデータセットにおいてのみ、DKT の有効性が検証されていた。本研究の実験では、算数や数学に関するデータセットと地理や歴史に関するデータセットは、Bayesian Knowledge Tracing からの精度向上という点では大きな差はなかった。したがって、DKT の適用可能性は科目に依らない可能性がある。

また、[Piech et al., 2015] では、モデルへの入力次元には問題に割り当てられたタグが利用されており、DKT の有効性はタグを用いた場合のみ、検証されていた。本研究の実験では、モデルへの入力次元にはタグを利用せず、個々の問題を利用し特に算数や数学に関する 6 データセットのうち 5 データセットで AUC が 0.8 を超えていた。このことは、DKT の適用には必ずしもタグが必要というわけではないということを示唆している。

一方で、難易度という点では、例えば、高等教育水準の学習内容を理解するためには、初等教育や中等教育水準の学習内容をあらかじめ理解している必要があるため、モデルの訓練と予測の両方において、初等教育や中等教育水準の内容を理解している人と理解していない人を同様のデータとして扱うことは難しい可能性がある。つまり、知識獲得の予測は、当該講座の内容の学習に際してあらかじ

め必要とされる知識を獲得している学習者に対して適用されるべきであり、そうでない学習者がいる割合が相対的に大きいと推察される高等教育水準の内容を扱う MOOCsにおいては、DKT の適用は限定的である可能性もある。

以上、考察を踏まえると、適用できるとする側面と、適用が限定的であるという側面があり、DKT の適用可能性については検証実験を行う必要があると考えられる。

本論文では、知識獲得における宣言的知識と手続き的知識の構造の違いを分析するために、初等中等教育向けの MOOCs を用いた。もし、Coursera をはじめとする大学講座を扱う多くの MOOCs の問題回答ログデータに対しても適用できるということであれば、大学講座を扱う多くの MOOCs では非常に多様な科目が提供されているため、それらの講座で扱われる知識の構造を分析することで、知識獲得と内容や教材についてより詳細に分析できるようになる可能性がある。

## 6.3 今後の展望

本研究の今後の展望について大きく 2 つの方針を述べる。1 つは対象データの多様化、結合、長期化についてであり、1 つは知識間関係の抽出に用いた Deep Knowledge Tracing の拡張についてである。

### 6.3.1 対象データの多様化、結合、長期化

まず、対象データの多様化について述べる。対象データの多様化は多様な科目や難易度についても学習者の知識獲得を予測しその知識構造を分析するという研究方針である。本研究では算数、数学、地理、歴史の知識構造について分析した。科目という点では、論理的思考、外国語、プログラミング等多くの科目については知識構造を分析しておらず、難易度という点では、大学水準のものについては知識構造を分析していない。知識獲得に関する知見は従来より指導や学習の設計に活用されており、他の多様なデータに対してもその知識構造を明らかにすることは重要であると考える。

次に、対象データの結合について述べる。対象データの結合は、異なる科目間で同じ学習者の知識獲得を予測しその知識構造を分析するという研究方針である。本研究では算数、数学、地理、歴史の知識構造についてそれぞれ独立に用いて分析した。例えば、数学や歴史は内容の関連性は低いだろうと考えるが、数学と理科は知識獲得という点で密接に関係していると考える。しかし、こうした科目間の知識構造を定量的に分析した研究もまだない。対象データを結合した研究は科目間の知識構造を定量的に分析することで、科目内の指導や学習の設計だけでなく科目間も考慮した設計に貢献すると考えられる。

最後に、対象データの長期化について述べる。対象データの長期化は、同じ学習者の知識獲得を長期間に渡って測定し、同じ科目で、あるいは、科目をまたいで、知識獲得を予測し、その知識構造を分析するという研究方針である。本研究では、データセットの対象期間は1年弱であった。しかし、学習者の知識獲得は既に獲得している知識に依存しており、できるだけ長い期間のデータを利用する方が、よりよく知識獲得を予測できる可能性が高い。つまり、より知識構造を明瞭に抽出でき、長期的な視点で科目を横断した指導や学習の設計に活用できる可能性が高い。

### 6.3.2 知識間関係抽出手法の拡張

知識間関係の抽出に用いた Deep Knowledge Tracing の拡張について2つ述べる。1つ目は、モデルの学習の際に長期的なログがある学習者の優先度をあげるという研究である。知識獲得は既に獲得している知識に依存しており、したがって、既に獲得している知識について多くわかっていた方が予測しやすい。つまり、長期的なログのある学習者を優先的に考慮した方が抽出される知識構造はいいものになる可能性がある。こうした研究に関連するものとして、カリキュラム学習というものがある。カリキュラム学習は2009年に Bengio らが提案した研究領域 [Bengio et al., 2009] である。カリキュラム学習はモデルの学習に際して、学習に用いるデータをカリキュラムに基づいて制御するという学習方法である。例えば、手書き文字認識では文字列が短い画像から学習させていく徐々に長い文字列の画像を含めて学習させることで、最終的に得られるモデルの性能が向上することが報告され

ている [Louradour and Kermorvant, 2014]。こうした手法でよりよい予測モデルを構築することで、抽出される知識構造も知識間関係をよく表現したものになる可能性がある。

2つ目は、予測の過程でタグを生成し知識の構造化に利用するという研究である。データセットによっては同じ知識を複数の別の問題で繰り返し学習させるという学習支援システムから収集されたものもあると考えられる。例えば、数学の演習問題  $2x + 1 = 0$  と  $3x - 4 = 0$ ,  $-3x + 5 = 0$  はほとんど同じ知識を問う問題であると考えられる。こうした場合、抽出した知識間関係ネットワークは必ずしも、ノード数という点で解釈しやすいものであるとは限らず、また、同じ知識を扱う問題群は相互に影響を与え合うため、環状構造を構築しやすい。したがって、こうしたネットワークでは、階層性の評価は難しくなる可能性があると考えられる。こうした問題を解決するためには、ほとんど同じ内容を扱っている問題にタグを割り当て、タグを1つのノードとして扱うようなネットワークを構築する必要があり、また、そのためには、抽出する知識間関係行列の各行、各列がタグを表現するようになっている必要があると考えられる。

こうした抽象的な関係を捉えるということに深層学習における埋め込みに関する技法が利用できる可能性がある。従来専門家がよいとしてきたタグと予測する過程で深層学習モデルが獲得するタグを比較することで、知識の粒度について新たな知見を得ることができると考えられる。また、タグを自動で抽出できるようになれば、これまで専門家が手で行っていた問題へのタグの割り当てとタグ間の関係の定義の2つの難しいタスクを深層学習の手法によりできるようになるということであり、学習科学の研究はこれまでにないほど進展するだろう。

以上、考察について述べた。次章では、結論を述べる。

## 第7章 結論

本論文では、主に心理学の研究領域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。

検証過程で得られた知見に基づいて、本研究で用いた知識構造の分析手法の他 MOOCs への適用可能性について議論し、講座の科目や問題へのタグの有無に依らず知識構造を分析できる可能性があること、および、特に、大学水準の難易度を扱う MOOCs については検証実験を行う必要があることを指摘した。

また、本研究の拡張として、適用対象の拡張という点で対象データの多様化、結合、長期化の3つの拡張を述べた。また、知識間関係抽出法である Deep Knowledge Tracing の拡張という点で、カリキュラム学習による拡張、および、スキルタグ自動抽出による拡張の2つについて述べた。

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信じている。

# 参考文献

- [Abelson, 2008] Abelson, H. (2008). The creation of opencourseware at mit. *Journal of Science Education and Technology*, 17(2):164–174.
- [Anderson, 1982] Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, 89(4):369.
- [Anderson, 1990] Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.
- [Anderson, 1993] Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, 48(1):35.
- [Bahdanau et al., 2015] Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., and Bengio, Y. (2015). End-to-end attention-based large vocabulary speech recognition. *arXiv preprint arXiv:1508.04395*.
- [Bastian et al., 2009] Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362.
- [Bastien et al., 2012] Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I. J., Bergeron, A., Bouchard, N., and Bengio, Y. (2012). Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop.
- [Bengio et al., 2009] Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM.

- [Bengio et al., 1994] Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *Neural Networks, IEEE Transactions on*, 5(2):157–166.
- [Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1):281–305.
- [Bergstra et al., 2010] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., and Bengio, Y. (2010). Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*. Oral Presentation.
- [Biswas et al., 2015] Biswas, S., Chadda, E., and Ahmad, F. (2015). Sentiment analysis with gated recurrent units. *Advances in Computer Science and Information Technology*.
- [Block et al., 1971] Block, J. H., Airasian, P. W., Bloom, B. S., and Carroll, J. B. (1971). *Mastery learning: Theory and practice*. Holt, Rinehart and Winston New York.
- [Bloom, 1968] Bloom, B. S. (1968). Learning for mastery. instruction and curriculum. regional education laboratory for the carolinas and virginia, topical papers and reprints, number 1. *Evaluation comment*, 1(2):n2.
- [Bock and O ’DEA, 2013] Bock, M. and O ’DEA, V. (2013). Virtual educators critique the value of moocs for k-12. *Education Week*, 32(20):10.
- [Buzydłowski et al., 2002] Buzydłowski, J. W., White, H. D., and Lin, X. (2002). Term co-occurrence analysis as an interface for digital libraries. In *Visual interfaces to digital libraries*, pages 133–144. Springer.
- [Cauley, 1986] Cauley, K. M. (1986). Studying knowledge acquisition: Distinctions among procedural, conceptual and logical knowledge. *the Annual Meeting of the*

*American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).*

[Chen et al., 2008] Chen, N.-S., Wei, C.-W., Chen, H.-J., et al. (2008). Mining e-learning domain concept map from academic articles. *Computers & Education*, 50(3):1009–1021.

[Cho et al., 2014] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

[Choi et al., 2015] Choi, E., Bahadori, M. T., and Sun, J. (2015). Doctor ai: Predicting clinical events via recurrent neural networks. *arXiv preprint arXiv:1511.05942*.

[Chung et al., 2014] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

[Chung et al., 2015] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2015). Gated feedback recurrent neural networks. *arXiv preprint arXiv:1502.02367*.

[Clevert et al., 2015] Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.

[Cohen and Hyman, 1979] Cohen, S. A. and Hyman, J. S. (1979). Learning for mastery: Ten conclusions after 15 years and 3,000 schools. *Educational Leadership*, 37(2):104–9.

[Coleman and Coleman, 1994] Coleman, J. S. and Coleman, J. S. (1994). *Foundations of social theory*. Harvard university press.

- [Corbett and Anderson, 1994] Corbett, A. T. and Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278.
- [Daniel, 2012] Daniel, J. (2012). Making sense of moocs: Musings in a maze of myth, paradox and possibility. *Journal of interactive Media in education*, 2012(3):Art–18.
- [De Jong and Ferguson-Hessler, 1996] De Jong, T. and Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational psychologist*, 31(2):105–113.
- [Dong et al., 2015] Dong, D., Wu, H., He, W., Yu, D., and Wang, H. (2015). Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. ACL.
- [Dong et al., 2012] Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N. V., Rao, J., and Cao, H. (2012). Link prediction and recommendation across heterogeneous social networks. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 181–190. IEEE.
- [FALAKMASIR et al., 2015] FALAKMASIR, M., Yudelson, M., Ritter, S., and Koedinger, K. (2015). Spectral bayesian knowledge tracing. In *Proceedings of the 8th International Conference on Educational Data Mining*, OC Santos, JG Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, JM Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, and M. Desmarais, Eds. Madrid, Spain, pages 360–364.
- [Gersten and Chard, 1999] Gersten, R. and Chard, D. (1999). Number sense re-thinking arithmetic instruction for students with mathematical disabilities. *The Journal of special education*, 33(1):18–28.

- [Glorot and Bengio, 2010] Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *International conference on artificial intelligence and statistics*, pages 249–256.
- [Graves and Schmidhuber, 2009] Graves, A. and Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 545–552.
- [Hagberg et al., 2008] Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA.
- [Hidasi et al., 2015] Hidasi, B., Karatzoglou, A., Baltrunas, L., and Tikk, D. (2015). Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*.
- [Hiebert, 2013] Hiebert, J. (2013). *Conceptual and procedural knowledge: The case of mathematics*. Routledge.
- [Hinton et al., 2012] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, 29(6):82–97.
- [Hochreiter, 1998] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116.
- [Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- [Jacomy, 2009] Jacomy, M. (2009). Force-atlas graph layout algorithm. *URL: http://gephi.org/2011/forceatlas2-the-new-version-of-our-home-brew-layout*.

- [Jordan, 2013] Jordan, K. (2013). Mooc completion rates: The data. *Availabe at: http://www. katyjordan. com/MOOCproject. html.* [Accessed: 27/08/2014].
- [Karpathy et al., 2015] Karpathy, A., Johnson, J., and Li, F.-F. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*.
- [Keller, 1968] Keller, F. S. (1968). Good-bye, teacher... *Journal of applied behavior analysis*, 1(1):79.
- [Kingma and Ba, 2014] Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [Krueger and Memisevic, 2015] Krueger, D. and Memisevic, R. (2015). Regularizing rnns by stabilizing activations. *arXiv preprint arXiv:1511.08400*.
- [Kushner and Yin, 2003] Kushner, H. J. and Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer Science & Business Media.
- [Larochelle et al., 2007] Larochelle, H., Erhan, D., Courville, A., Bergstra, J., and Bengio, Y. (2007). An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th international conference on Machine learning*, pages 473–480. ACM.
- [Le et al., 2015] Le, Q. V., Jaitly, N., and Hinton, G. E. (2015). A simple way to initialize recurrent networks of rectified linear units. *arXiv preprint arXiv:1504.00941*.
- [LeCun et al., 1998] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- [Lipton et al., 2015] Lipton, Z. C., Kale, D. C., Elkan, C., and Wetzel, R. (2015). Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*.

- [Liyanagunawardena et al., 2013] Liyanagunawardena, T., Williams, S., and Adams, A. (2013). The impact and reach of moocs: A developing countries ' perspective. *eLearning Papers*, (33).
- [Louradour and Kermorvant, 2014] Louradour, J. and Kermorvant, C. (2014). Curriculum learning for handwritten text line recognition. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*, pages 56–60. IEEE.
- [Luo and Magee, 2011] Luo, J. and Magee, C. L. (2011). Detecting evolving patterns of self-organizing networks by flow hierarchy measurement. *Complexity*, 16(6):53–61.
- [MacHardy and Pardos, 2015] MacHardy, Z. and Pardos, Z. A. (2015). Toward the evaluation of educational videos using bayesian knowledge tracing and big data. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 347–350. ACM.
- [McAuley et al., 2010] McAuley, A., Stewart, B., Siemens, G., and Cormier, D. (2010). The mooc model for digital practice.
- [McCormick, 1997] McCormick, R. (1997). Conceptual and procedural knowledge. *International journal of technology and design education*, 7(1-2):141–159.
- [Mikolov, 2012] Mikolov, T. (2012). Statistical language models based on neural networks. *Presentation at Google, Mountain View, 2nd April*.
- [Mones et al., 2012] Mones, E., Vicsek, L., and Vicsek, T. (2012). Hierarchy measure for complex networks. *PloS one*, 7(3):e33799.
- [Nair and Hinton, 2010] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814.
- [Newman, 2006] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

- [Page et al., 1999] Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: bringing order to the web.
- [Pappano, 2012] Pappano, L. (2012). The year of the mooc. *The New York Times*, 2(12):2012.
- [Pascanu et al., 2013] Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *arXiv preprint arXiv:1211.5063*.
- [Pavlik Jr et al., 2009] Pavlik Jr, P. I., Cen, H., and Koedinger, K. R. (2009). Performance factors analysis—a new alternative to knowledge tracing. *Online Submission*.
- [Pezeshki, 2015] Pezeshki, M. (2015). Sequence modeling using gated recurrent neural networks. *arXiv preprint arXiv:1501.00299*.
- [Piech et al., 2015] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., and Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems*, pages 505–513.
- [Putnam, 1993] Putnam, R. D. (1993). The prosperous community: social capital and public life. *The american prospect*, (13).
- [Resnick, 1977] Resnick, L. B. (1977). Assuming that everyone can learn everything, will some learn less?
- [Rittle-Johnson and Alibali, 1999] Rittle-Johnson, B. and Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of educational psychology*, 91(1):175.
- [Rivard, 2013] Rivard, R. (2013). Measuring the mooc dropout rate. *Inside Higher Ed*, 8:2013.
- [Robbins and Monro, 1951] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407.

- [Ryle, 1945] Ryle, G. (1945). Knowing how and knowing that: The presidential address. In *Proceedings of the Aristotelian Society*, pages 1–16. JSTOR.
- [Sak et al., 2015] Sak, H., Senior, A., Rao, K., and Beaufays, F. (2015). Fast and accurate recurrent neural network acoustic models for speech recognition. *arXiv preprint arXiv:1507.06947*.
- [Schein et al., 2002] Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM.
- [Schroff et al., 2015] Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *arXiv preprint arXiv:1503.03832*.
- [Siemens, 2013] Siemens, G. (2013). Massive open online courses: Innovation in education. *Open educational resources: Innovation, research and practice*, 5.
- [Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.
- [Sun et al., 2011] Sun, Y., Barber, R., Gupta, M., Aggarwal, C. C., and Han, J. (2011). Co-author relationship prediction in heterogeneous bibliographic networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 121–128. IEEE.
- [Sutskever et al., 2014] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

- [Szegedy et al., 2014] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*.
- [Ten Berge and Van Hezewijk, 1999] Ten Berge, T. and Van Hezewijk, R. (1999). Procedural and declarative knowledge an evolutionary perspective. *Theory & Psychology*, 9(5):605–624.
- [Trucano et al., 2013] Trucano, M., Kendrick, C., and Gashurov, I. (2013). More about moocs and developing countries.
- [Unger, 1968] Unger, P. (1968). An analysis of factual knowledge. *The Journal of Philosophy*, pages 157–170.
- [Vinyals et al., 2014] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2014). Show and tell: A neural image caption generator. *arXiv preprint arXiv:1411.4555*.
- [Werbos, 1990] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560.
- [Williams and Zipser, 1989] Williams, R. J. and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280.
- [Xu et al., 2015] Xu, K., Ba, J., Kiros, R., Courville, A., Salakhutdinov, R., Zemel, R., and Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint arXiv:1502.03044*.
- [Yin et al., 2015] Yin, J., Jiang, X., Lu, Z., Shang, L., Li, H., and Li, X. (2015). Neural generative question answering. *arXiv preprint arXiv:1512.01337*.
- [Yuan et al., 2013] Yuan, L., Powell, S., and CETIS, J. (2013). Moocs and open education: Implications for higher education.

- [Yudelson et al., 2013] Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. In *Artificial Intelligence in Education*, pages 171–180. Springer.
- [Zaremba, 2015] Zaremba, W. (2015). An empirical exploration of recurrent network architectures.
- [Zeiler, 2012] Zeiler, M. D. (2012). Adadelta: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.
- [Zeiler and Fergus, 2014] Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014*, pages 818–833. Springer.
- [文部科学省, 2011] 文部科学省 (2011). 現行学習指導要領・生きる力: 学習指導要領とは何か? [http://www.mext.go.jp/a\\_menu/shotou/new-cs/idea/1304372.htm](http://www.mext.go.jp/a_menu/shotou/new-cs/idea/1304372.htm).

## 謝辞

本研究の遂行や本論文の執筆にあたり，非常に多くの方からご指導，ご支援をいただきました．心より御礼申し上げます．

指導教官である 松尾豊特任准教授 には，研究構想の相談や論文の書き方，本論文の論理構成について，貴重なご指導をいただきました．特に，本論文執筆途中に頂いた厳しいフィードバックは，本論文の完成に不可欠なものでした．ここに，深く謝意を表します．

深層学習の実装に関する多くの相談を応じて下さり，また，分析サーバや GPU 解析環境の用意等，物理的な研究環境の構築に多大なご協力を下さった研究室の教官である 中山浩太郎先生 に，深く感謝致します．

共同研究プロジェクトという形式で本研究の遂行に必要不可欠であった貴重なデータをご提供して下さったリクルートマーケティングパートナーズ(株)の社長の 山口文洋様 には深く感謝申し上げます．特に，共同研究プロジェクトの進行に際して，さまざまな要望・相談に対応して下さったリクルートマーケティングパートナーズ(株)の 萩原静巣様 に，深く感謝致します．

経営共創基盤の 川上登福様 と 浜田貴之様 からは，3年前の共同研究プロジェクトの発足から現在に至るまで，プロジェクト進行に際してあらゆる観点でご協力およびご配慮いただきました．深く感謝致します．

松尾研究室や GCI の皆様には，多大なご協力，ご支援いただきました．秘書の中野佐恵子さん，永本登代子さん，浪岡亮子さんは，日頃から研究室の環境を整えて下さり，研究生活を支えてくださいました．松尾研究室博士課程の 岩澤有祐さん，野中尚輝さん，鈴木雅大さん，金子貴輝さん，Edison Marrese Taylor さん，

---

Pablo Loyola さんは，人工知能という競争が激しい過酷な研究領域において，共に切磋琢磨してくださいました．特に，研究アイディアの相談に乗ってくださいり，また，生成モデルや統計検定等，機械学習の理論方面の理解を手助けしてくださいり，研究遂行上重要な技術的な相談に何度も応じてくださった鈴木雅大さんには，深く感謝いたします．学部生の味曾野雅史君，塩谷碩彬君は，膨大で度々更新されるデータの整備を通して，本研究の下支えをしてくださいました．また，河野慎さん，富山翔司君，金子貴輝さんは，しばしば飲み会を通じて息抜きさせてくださいました．ここに，松尾研究室や GCI の皆様へ謝意を表します．

最後に，両親と兄弟に深く深く感謝を申し上げます．多忙を理由に実家に帰らず連絡もあまりしない私をたまに帰った時にはいつも通り温かく迎えてくれる家族のおかげで，これまでこうして研究生活を続けられているのだと深く感じております．また，今後もこれまでと変わらず見守ってくれるであろう家族の存在を心の支えとして，研究を続けていきたいと考えています．改めて，家族に深く感謝申し上げます．

東京大学大学院工学系研究科  
技術経営戦略学専攻  
松尾研究室 修士二年  
那須野薫  
平成 28 年 3 月

# 発表文献

## 学会誌論文

- 那須野薰，奥山晶二郎，中西鏡子，松尾豊．(2015)．Twitter における候補者の選挙地盤に着目した国政選挙の当選者予測．情報処理学会論文誌，56(10)，2044-2053．

## 学会発表

- 那須野薰，萩原静巣，井上綾香，松尾豊．(2015)．大規模オンライン講座における自己適応学習者に着目した学習項目の理解度予測．人工知能学会全国大会論文集，29，1-4．
- 後藤拓矢，那須野薰，萩原静巣，松尾豊．(2015)．受験向け動画サービスにおける合否結果を加味した教材の推薦手法の提案．人工知能学会全国大会論文集，29，1-3．
- 那須野薰，松尾豊．(2014)．Twitter における候補者の情報拡散に着目した国政選挙当選者予測．第 28 回人工知能学会全国大会論文集．
- 那須野薰，松尾豊．(2013)．2013 年参議院議員選挙における Twitter を用いた当選者予測 (マイクロプログ，第 5 回集合知シンポジウム)．電子情報通信学会技術研究報告．NLC，言語理解とコミュニケーション，113(338)，25-28．
- 那須野薰，上野山勝也，松尾豊．(2013)．次世代プログラミング学習サイト構築の試み．人工知能学会全国大会論文集，27，1-3．
- 樋剛史，那須野薰，柳原正，古賀光，加藤芳隆，那和一成，松尾豊．(2013)．ソーシャルメディアからの予告型の地域イベント及び参加状態の抽出手法の提案．人工知能学会全国大会論文集，27，1-4．