

平成 27 年度

修士論文

Deep Knowledge Tracing の汎用性の実験的評価

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 那須野薰

要旨

TODO

目 次

第 1 章 背景と目的	1
第 2 章 先行研究	4
2.1 Knowledge Tracing の定式	4
2.2 Bayesian Knowledge Tracing	4
2.3 Recurrent Neural Network	6
2.4 Deep Knowledge Tracing	9
第 3 章 データセット	13
3.1 データセットの外観	14
3.2 個々のデータセットの具体的説明	16
第 4 章 実験設定	37
4.1 実験項目	37
4.2 データセットに合わせた手法の修正	39
4.3 最適化	40
第 5 章 結果と分析	41
5.1 実験 1 : DKT と BKT の性能差とデータセットの科目	41
5.2 実験 2 : DKT と BKT の性能差とデータセットの統計量	43
5.3 実験結果の整理	45
第 6 章 考察	46
第 7 章 まとめ	49
参考文献	50

図 目 次

2.1 RNN の構造のイメージ図	6
3.1 小学 4 年国語の問題と回答選択肢の例	17
3.2 小学 4 年国語の平均着手順と回答ログ数の XY プロット	17
3.3 小学 4 年社会の問題と回答選択肢の例	18
3.4 小学 4 年社会の平均着手順と回答ログ数の XY プロット	18
3.5 小学 4 年算数の問題と回答選択肢の例	19
3.6 小学 4 年算数の平均着手順と回答ログ数の XY プロット	19
3.7 小学 4 年理科の問題と回答選択肢の例	19
3.8 小学 4 年理科の平均着手順と回答ログ数の XY プロット	19
3.9 小学 5 年国語の問題と回答選択肢の例	20
3.10 小学 5 年国語の平均着手順と回答ログ数の XY プロット	20
3.11 小学 5 年社会の問題と回答選択肢の例	21
3.12 小学 5 年社会の平均着手順と回答ログ数の XY プロット	21
3.13 小学 5 年算数の問題と回答選択肢の例	21
3.14 小学 5 年算数の平均着手順と回答ログ数の XY プロット	21
3.15 小学 5 年理科の問題と回答選択肢の例	22
3.16 小学 5 年理科の平均着手順と回答ログ数の XY プロット	22
3.17 小学 6 年国語の問題と回答選択肢の例	23
3.18 小学 6 年国語の平均着手順と回答ログ数の XY プロット	23
3.19 小学 6 年社会の問題と回答選択肢の例	23
3.20 小学 6 年社会の平均着手順と回答ログ数の XY プロット	23
3.21 小学 6 年算数の問題と回答選択肢の例	24
3.22 小学 6 年算数の平均着手順と回答ログ数の XY プロット	24
3.23 小学 6 年理科の問題と回答選択肢の例	25
3.24 小学 6 年理科の平均着手順と回答ログ数の XY プロット	25

3.25 中学 1 年国語の問題と回答選択肢の例	25
3.26 中学 1 年国語の平均着手順と回答ログ数の XY プロット	25
3.27 中学 1 年数学の問題と回答選択肢の例	26
3.28 中学 1 年数学の平均着手順と回答ログ数の XY プロット	26
3.29 中学 1 年英語の問題と回答選択肢の例	27
3.30 中学 1 年英語の平均着手順と回答ログ数の XY プロット	27
3.31 中学 2 年国語の問題と回答選択肢の例	28
3.32 中学 2 年国語の平均着手順と回答ログ数の XY プロット	28
3.33 中学 2 年数学の問題と回答選択肢の例	28
3.34 中学 2 年数学の平均着手順と回答ログ数の XY プロット	28
3.35 中学 2 年英語の問題と回答選択肢の例	29
3.36 中学 2 年英語の平均着手順と回答ログ数の XY プロット	29
3.37 中学 3 年国語の問題と回答選択肢の例	30
3.38 中学 3 年国語の平均着手順と回答ログ数の XY プロット	30
3.39 中学 3 年数学の問題と回答選択肢の例	31
3.40 中学 3 年数学の平均着手順と回答ログ数の XY プロット	31
3.41 中学 3 年英語の問題と回答選択肢の例	31
3.42 中学 3 年英語の平均着手順と回答ログ数の XY プロット	31
3.43 中学地理の問題と回答選択肢の例	32
3.44 中学地理の平均着手順と回答ログ数の XY プロット	32
3.45 中学歴史の問題と回答選択肢の例	33
3.46 中学歴史の平均着手順と回答ログ数の XY プロット	33
3.47 中学公民の問題と回答選択肢の例	33
3.48 中学公民の平均着手順と回答ログ数の XY プロット	33
3.49 中学理科 1 の問題と回答選択肢の例	34
3.50 中学理科 1 の平均着手順と回答ログ数の XY プロット	34
3.51 中学理科 2 の問題と回答選択肢の例	35
3.52 中学理科 2 の平均着手順と回答ログ数の XY プロット	35

表 目 次

2.1 回答ログデータのサンプル	10
3.1 26 の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 .	1
4.1 勉強サプリの回答ログデータと対応する入力ベクトルの例	39
5.1 各データセットに対する各手法の予測結果とそれらの関係性	42
5.2 DKT の AUC – BKT の AUC と 各データセット分類に含まれるデータセットの統計量 の	

第1章 背景と目的

Knowledge Tracing は学習者の知識獲得の軌跡をモデリングするタスクである。最初の研究報告は、1995 年の Corbett らによるプログラミング学習についてのものである [1]。Corbett らは、大学で活用するプログラミング学習用の学習支援システムを対象に、プログラミングに用いられる手続き的知識（関数の宣言やリストの要素指定など）の獲得の程度を学習者の問題回答の正誤をもとに推定した。この研究報告から今日まで、約 20 年研究されている。

Knowledge Tracing の学習支援システムへの活用は学習者の成績向上に有効であることが報告されている。XX らは Knowledge Tracing を演習システムに導入することで、学習者の学習速度が NN 倍、最終的な学習成績が XX ばいに向上したことを報告している [xx]。[xx,xx] でも、導入することで学習者の最終的な成績が向上したことが報告がなされている。TODO：要出典

また、近年、大規模オンライン講座 [2–4] の登場により、その有用性はますます拡大していると考えられる。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、Knowledge Tracing の活用可能性も限定的であったと考えられる。しかし、大規模オンライン講座の整備が進み、誰でもいつでも好きな場所で学習支援システムが利用できるようになりつつあり、その有用性はますます拡大していると考えられる。

近年、深層学習の進展により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 [5–9] が、Knowledge Tracing の領域もその例外ではない。Bayesian Knowledge Tracing という Corbett らが提案した手法以降、Bayesian Knowledge Tracing を拡張する形で、個々のユーザの特徴の考慮を試みるもの [10–12] や入力形式を工夫するもの [13] などの手法が提案されていた。しかし、近年、深層学習の進展を受けて Knowledge Tracing の性能は、これまでの研究報告と比べて飛躍的に向上した。

具体的には，Piech らは 2015 年 Deep Knowledge Tracing という手法を開発し，従来手法よりも予測性能 (AUC [14]) が 25% 向上し，また，構築したモデルを分析することで，知識間の関連表現であるコンセプトマップを抽出できることを報告した [15] .

Knowledge Tracing の研究において特徴的な点として，対象科目が主に一部の科目に偏っていることが挙げられる。[16] の対象科目はリーディングであるが，[1, 17] の対象科目はプログラミングで，[10, 12, 13, 15] の対象科目は数学であり，理数系の科目，特に数学とプログラミングに偏っている。これにはいくつかの理由が考えられる。1 つは，最初の Corbett らによる手続き的知識に関する研究であり，その後の研究がプログラミングやそれに近い数学に続いたということ，1 つは米国の方針の 1 つである STEM (Science , Technology , Engineering , Math) [18, 19] という名前に表現されるように数学学習の産業への影響が大きく分析対象科目として数学の重要性が高いということ，1 つは，それらの影響を部分的に受けていると考えられるが，Knowledge Tracing の研究に利用できる公開されているデータセット [20–26] が数学に関するものであることがある。いずれにせよ，対象科目は数学やプログラミングを中心であり，他の科目が取り上げられている研究報告は少ない。

こうした技術の飛躍的進展や対象科目の偏りという研究領域の特徴を踏まえて，本稿では，最新手法をさまざまな科目のデータに適用し結果を整理し，Knowledge Tracing の研究領域の地盤固めを行う。具体的には国語，数学，英語，理科，社会など 26 のデータセットを題材に Deep Knowledge Tracing と Bayesian Knowledge Tracing を適用してそれぞれ予測性能や性能向上の要因を分析する。従来研究の主な対象科目であった数学やプログラミング以外の科目についても Deep Knowledge Tracing が同様に有効であれば，学習支援システムはさらに有用なものになると期待できる。また，構築したモデルの性能に差があれば，その差の要因を分析し有效地に適用できる条件を明らかにすることで，今後の研究の方向性を整理することができるかもしれない。研究が飛躍的に進展した今日，さまざまなデータに適用して結果を整理することは，単に学習支援システムへの活用を通して学習効率の向上に寄与する可能性があるという点だけでなく，研究地盤を固めるという点で学術的意義は大きい。

本稿の貢献は下記の通りである。

- 26のデータセットに対する実験的評価を通じて，[15]で示されたように Deep Knowledge Tracing が確かに数学のデータセットの対して有効であることを示し，また，[15]では扱われていなかった科目である国語，社会，理科，英語などの数学以外の科目においても Deep Knowledge Tracing が Bayesian Knowledge Tracing より予測性能という点で優れていることを示した。
- 問題1問あたりの回答ログ数が十分であれば，問題数が多いほど Deep Knowledge Tracing の性能は高い傾向にあることを示し，Deep Knowledge Tracing への入力次元が多いほど学習者の知識獲得をよくモデリングできる可能性を示唆した。これは，XXX や XXX と同様の結果である。
- スキルタグを用いず，問題 ID を基にして Deep Knowledge Tracing を適用して，高い精度が得られることを示した。
- 学習速度を表現するニューロンを発見した（仮）
- 今後の研究方針として，XXX や XXX を提案した。

本稿の構成について述べる。まず，次節では，Knowledge Tracing の定式と Bayesian Knowledge Tracing の仕組みについて述べ，最新手法で利用される Recurrent Neural Network を説明したのちに，最新手法である Deep Knowledge Tracing を述べる。3 節では，本研究で用いるデータとデータセットの作成方法について述べ，4 節では，データセットに合わせせて Deep Knowledge Tracing に加える修正点について述べる。5 節では 3 つの実験について述べる。6 節で，本稿全体についてのまとめる。

第2章 先行研究

2.1 Knowledge Tracing の定式

Knowledge Tracing は学習者の知識獲得の軌跡をモデリングするタスクである。学習者の時刻 t において観測された学習行動結果を 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学習行動結果 q_{t+1} を予測するタスクである。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。モデル性能の評価は [10, 13] では Accuracy で行っており、[15] では AUC で行っている。

2.2 Bayesian Knowledge Tracing

Bayesian Knowledge Tracing [1] (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 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 $p\text{-guess}$) の 4 つで、これら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。

$$p(L_1)_u^k = p(L_0)^k \quad (2.1)$$

$$p(L_t|obs=correct)_u^k = \frac{p(L_{t-1})_u^k \cdot (1 - p(S)^k)}{p(L_{t-1})_u^k \cdot (1 - p(S)^k) + (1 - p(L_{t-1})_u^k) \cdot p(G)^k} \quad (2.2)$$

$$p(L_t|obs=wrong)_u^k = \frac{p(L_{t-1})_u^k \cdot p(S)^k}{p(L_{t-1})_u^k \cdot p(S)^k + (1 - p(L_{t-1})_u^k) \cdot (1 - p(G)^k)} \quad (2.3)$$

$$p(L_t)_u^k = p(L_t|obs)_u^k + (1 - p(L_t|obs)_u^k) \cdot p(T)^k \quad (2.4)$$

$$p(C_t)_u^k = p(L_{t-1})_u^k \cdot (1 - p(S)^k) + (1 - p(L_{t-1})_u^k) \cdot p(G)^k \quad (2.5)$$

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 2.1 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 2.2 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 2.3 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 2.4 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、理解していないがあ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、身につけ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 2.5 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 [1] 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 [10] 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ということもあり、[10] では勾配法に基づいた識別モデルとして学習させていている。具体的には、[10] 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。本稿で扱うデータは大規模データであるため、学習方法は識別モデルとして扱う場合について説明した。生成モデルとして扱う場合の学習方法については [1] を参照されたい。

2.3 Recurrent Neural Network

Recurrent Neural Network [27] (以下, RNN) について概説する。RNN は深層ニューラルネットワークの一種で主に系列データの解析に利用される。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で活用されるようになり, 例えば, 機械翻訳 [28, 29], 手書き文字認識 [30, 31], 音声認識 [32, 33], ユーザログ解析 [34], 画像説明文生成 [35, 36], 医療診断 [37, 38] 等の領域で高い性能を発揮することが報告されている。

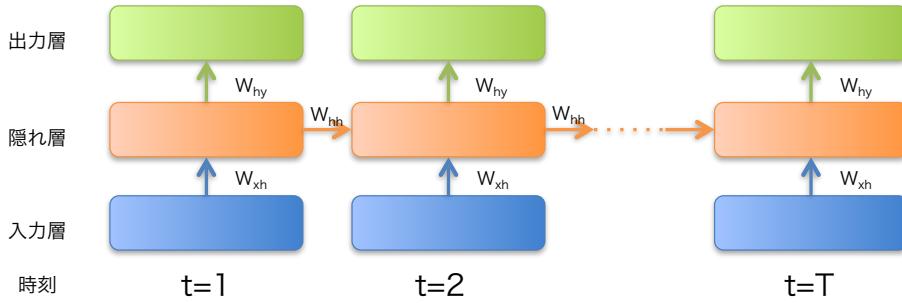


図 2.1: RNN の構造のイメージ図

伝統的な RNN の構造は図 2.1 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。系列方向の時刻 t の隠れ層 $h_t = f(x_t, h_{t-1})$ の計算に時刻 $t - 1$ の隠れ層の情報を入力するというように, 一つ前の情報を再帰的 (recurrent) に入力するという構造である。関数 f は例えば, 入力である x_t や h_{t-1} をアフィン変換して活性化関数にかけるというのが挙げられる。活性化関数はシグモイド関数や \tanh , Relu [39], ELUs [40] 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 [41]。RNN の学習には勾配法に基づいた確率的勾配降下法 [42, 43] や Adam [44], AdaDelta [45] など, さまざまな種類が利用可能であるが, いずれにせよ, 勾配爆発 [41, 46] や勾配消滅 [46, 47] という語が表現するように, 長い系列データを RNN で学習する場合, $h_t = f(x_t, h_{t-1})$ の式に現れるように同じ変換を繰り返し行うため, 勾配が爆発したり消滅したりして, 長期的な特徴量を捉えることが難しい。

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ことが挙げられる。活性化関数については勾配消滅の

緩和に対して有効であるとされるが，詳細は後述する．学習時の勾配に制約を加えることで，勾配爆発を抑制する効果が期待できる．具体的には，[46] 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式 XX に従いノルムを抑制する方法が報告され，[48] 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クリップする方法が報告されている．

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ます．以降では，よく研究報告で取り上げられる Simple RNN (SRNN) [27]，Long Short Term Memory RNN (LSTM-RNN) [49]，Gated Recurrent Unit RNN (GRU-RNN) [50] 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．[51, 52] 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 GRU-RNN 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 によるモデルは下記の式で定義される．

$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (2.6)$$

$$y_t = \sigma(W_{hy}h_t + b_y) \quad (2.7)$$

ここでは， t は時刻を指し， x_t は時刻 t の入力ベクトルを指し， h_t は時刻 t の隠れ層を指し， y_t は時刻 $t+1$ の各問題の正誤確率の予測値を指し， W_{xh} ， W_{hh} はそれぞれ重み行列を指し， b_h ， 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， σ 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，重み行列とバイアス項 (W_{xh} ， W_{hh} ， b_h ， 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 [49]。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。state of the art 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 LSTM-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 で用いられる LSTM-RNN は下記の式で定義される。

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (2.8)$$

$$g_t = \sigma(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \quad (2.9)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2.10)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (2.11)$$

$$m_t = f_t \odot m_{t-1} + i_t \odot g_t \quad (2.12)$$

$$h_t = o_t \odot m_t \quad (2.13)$$

$$y_t = \sigma(W_{my}m_t + b_y) \quad (2.14)$$

ここでは、 i_t はメモリセルへの入力を指し、 f_t は forget ゲートを指し、 g_t は input ゲートを指し、 o_t は output ゲートを指し、 m_t はメモリセルを指し、 W_{xi} 、 W_{hi} 、 W_{xg} 、 W_{hg} 、 W_{xf} 、 W_{hf} 、 W_{xo} 、 W_{ho} 、 W_{my} はそれぞれ重み行列を指し、 b_i 、 b_g 、 b_f 、 b_o 、 b_y はそれぞれバイアス項を指し、 \odot は要素積を指す。

TODO：説明追記

GRU-RNN

GRU-RNN は Gated Recurrent Unit [50] 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為に提案された活性化関数である。Cho ら [50] が 2014 年に発表して以来、GRU-RNN 自体や GRU-RNN を活用した研究が多く報告されている [53–58]。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，GRU-RNNの性能を比較した研究 [53, 54]においてLSTM-RNNとGRU-RNNが同程度の性能であることが報告されている。

GRU-RNNは下記の式により定義される。

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \quad (2.15)$$

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \quad (2.16)$$

$$\tilde{h}_t = \tanh(W_{xh}x_t + W_{hh}(r \odot h_{t-1} + b_h)) \quad (2.17)$$

$$h_t = z_t h_{t-1} + (1 - z_t)\tilde{h}_t \quad (2.18)$$

$$y_t = \sigma(W_{hy}h_t + b_y) \quad (2.19)$$

ここでは， $W_{xr}, W_{hr}, W_{xz}, W_{hz}, W_{xh}, W_{hh}$ は重み行列で， b_r, b_z, b_r はバイアス項である。 r_t がreset gate(LSTMにおけるforget gateに相当する機構)で， z_t がupdate gate(LSTMにおけるmemory cellに相当する機構)である。 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， 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，RNNについて述べた。次に，RNNを活用したKnowledge Tracingの手法であるDeep Knowledge Tracingについて述べる。

2.4 Deep Knowledge Tracing

Deep Knowledge Tracing [15]（以下，DKT）は2015年6月に発表されたRNNを利用しKnowledge Tracingを行う手法である。DKTの特筆すべき点としては，

- Knowledge Tracingに有効な形でRNNを活用しモデルの入出力を定義していること。
- ベンチマーク用のデータセットにおいて従来手法よりAUCが25%以上高いこと。
- 演習問題間の関係や概念間の関係を抽出できること。

が挙げられる。

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 x_1, \dots, x_T を出力のベクトル系列 y_1, \dots, y_T に写像する。この写像は，隠れ状態 h_1, \dots, 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続的に符号化している，とみなせる。図 XX に RNN の簡単な構造を示す。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。

$$h_t = f(x_t, h_{t-1}) \quad (2.20)$$

$$y_t = g(h_t) \quad (2.21)$$

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 2.6，2.7や LSTM-RNN の式 2.8–2.14，GRU-RNN の式 2.15–2.19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 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 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， x_t の長さは $2M$ となる。

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，下記の

表 2.1: 回答ログデータのサンプル

ユーザ ID	ログの順番	問題番号	正誤
A	1	1	0
A	2	1	1
A	3	2	1
A	4	3	0
A	5	3	1
A	6	4	1

のようなベクトル系列となる。

$$x_1 = [0000:1000] \quad (2.22)$$

$$x_2 = [1000:0000] \quad (2.23)$$

$$x_3 = [0100:0000] \quad (2.24)$$

$$x_4 = [0000:0010] \quad (2.25)$$

$$x_5 = [0010:0000] \quad (2.26)$$

$$x_6 = [0001:0000] \quad (2.27)$$

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、 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 $l(\mathbf{y}^T \delta(q_{t+1}), a_{t+1})$ であり、学習者一人のロスは下記の式で与えられる。

$$L = \sum_t l(\mathbf{y}^T \delta(q_{t+1}), q_{t+1}) \quad (2.28)$$

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。^[15] 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 ^[59] 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 ^[60] 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 ^[46] にしたがい、制約を設けている。

問題関係・知識関係抽出法

次に，DKTのモデルを利用した問題関係もしくは知識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 が現れた後に現れる残りの問題系列の中で，問題 j が1%以上を占めること。

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。[15]では，問題間の影響グラフ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グラフを構築した。

以上，Deep Knowledge Tracingの構造，最適化，問題関係・知識関係抽出法について説明した。以降では，これらの背景を踏まえて，本研究で実際に行うことについて述べていく。

第3章 データセット

本節では、実験で用いるデータセットについて説明する。データは勉強サプリ¹というオンライン講座におけるユーザの回答行動ログを利用する²。勉強サプリは主に小学4年生から中学3年生を対象としており、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科目と学年共通で地理、歴史、公民、理科1、理科2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。本稿では、これら26講座におけるユーザのドリル演習ログデータよりKnowledge Tracingのデータセットを作成しBKTとDKTを適用する。対象期間は2015年4月から2015年11月の8ヶ月のデータである。

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。

勉強サプリでは学習者は同時に複数の問題に回答する場合があるが、これは、DKTの原著で利用されていたデータセットの設定と異なり、DKTの手法をやや修正する必要がある。^[15]の設定では同時に回答される問題は1つで、DKTのモデルの入力はその時刻に回答された問題に対応する要素のone-hotベクトルに符号化されたものを用いられていた。したがって、DKTのモデルの入力と出力の部分を勉強サプリのデータの形式に合うように修正する必要があるが、その方法については後述する。

¹<https://benkyosapuri.jp>

²本稿の研究はリクルートマーケティング(株)との共同研究プロジェクトの一環で行われている。

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から下記に該当するレコードをノイズとみなして削除した。

条件：同時に回答された複数の問題のレコード群について、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クリックで簡単にドリル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や、特に回答時間が短く誤って当該ドリル演習ページを開いてしまったと推察されるレコードが多かったからである。

3.1 データセットの外観

26の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3.1に整理した。これらの統計量は実験結果の分析に際して、BKTやDKTのモデルの性能の関係性を評価するのに用いる。

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3年の国語と中学1年の英語）間で差は5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年の数学や英語で、それに続いて、6,000ユーザ以上の中学の地理、歴史、理科1、理科2である。これは、勉強サプリのユーザは中学1年生の割合が多いということが現れていると考えられる。

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、最も大きい2つのデータセット（小学4年の社会と中学2年の英語）間で差は6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、また、小学生のデータセットについては、国語のデータセットは問題数が少なく、算数のデータセットの問題数が多い傾向にある。中学生のデータセットについては、英語のデータセットの問題数は大きい傾向にあるが、その他については概ね同程度の問題数に基づいている。

回答ログ数については、全体として、数万から数十万レコードであり、最も大きい2つのデータセット（中学の歴史と中学3年の国語）間で差は20倍程度である。各学年ごとで見ると算数や数学のレコード数が多い傾向にある。また、中学

表 3.1: 26 の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 .

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	国語	3,534	81	183,192	52	2,262
	社会	3,045	76	227,409	75	2,992
	算数	4,318	182	505,917	117	2,780
	理科	3,552	132	343,731	97	2,604
小学 5 年	国語	2,797	80	144,776	52	1,810
	社会	2,833	197	388,521	137	1,972
	算数	3,380	257	411,957	122	1,603
	理科	3,194	130	302,092	95	2,324
小学 6 年	国語	2,517	80	123,413	49	1,543
	社会	2,891	202	434,324	150	2,150
	算数	3,225	245	395,276	123	1,613
	理科	2,885	132	260,408	90	1,973
中学 1 年	国語	4,664	380	191,050	41	503
	数学	7,137	365	659,237	92	1,806
	英語	7,448	399	504,107	68	1,263
中学 2 年	国語	2,411	275	84,864	35	309
	数学	3,931	278	238,241	61	857
	英語	3,871	461	223,766	58	485
中学 3 年	国語	1,410	295	43,270	31	147
	数学	2,667	343	177,295	66	517
	英語	2,308	449	91,980	40	205
中学	地理	6,499	308	660,882	102	2,146
	歴史	6,381	364	853,419	134	2,345
	公民	2,027	362	253,406	125	700
	理科 1	6,636	438	582,819	88	1,331
	理科 2	6,908	375	624,053	90	1,664

で共通の科目である地理 , 歴史 , 理科 1 , 理科 2 のレコード数は 3 学年のユーザが利用しているということもあり多い傾向にある .

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、最も大きい2つのデータセット（小学6年生の社会と中学3年の国語）間で差は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、最も大きい2つのデータセット（小学4年の社会と中学3年の国語）間で差は20倍程度である。小学生のデータセットは中学生のデータセットより問題1問あたりの平均回答ログ数が大きい傾向にある。本稿で利用するデータは約1年弱の期間のサービスログデータであり、各学年で扱う後半の問題について回答ログ数が少ない傾向にあり、特に、中学生は学校のカリキュラムに沿った形で勉強サプリを利用しているユーザが多いからと考えらえる。これについては、詳細は後述する。

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、ユーザ数は最大で5倍程度の差があり、問題数は最大で6倍程度の差があり、回答ログ数は最大で20倍程度の差があり、ユーザ1人あたりの平均回答ログ数は最大で5倍程度の差があり、問題1問あたりの平均回答ログ数は最大で20倍程度の差があり、各統計量にはある程度ばらつきがあり、BKTやDKTのモデルの性能や性能差を評価することで、モデルの必要なデータ量や科目の特徴を分析できる可能性があると考えられる。

3.2 個々のデータセットの具体的説明

本稿の対象である26の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。要旨は下記の通りである。

- 小学4年国語、小学5年国語、小学6年国語で扱われる内容の項目は近く、また、平均着手順と回答ログ数のXYプロットの外観も近く、いずれも最初の方の問題に回答ログが偏っている。
- 小学生の社会の内容は学年ごとに異なる。おおむね小学4年社会は現代社会、小学5年社会は地理、小学6年社会は歴史となっている。また、平均着手順と回答ログ数のXYプロットの外観も近く、いずれも前半の方の問題から後半の方の問題にかけてほぼ線形に回答ログ数が減少している。
-

小学4年国語

問1 次の文を文節に区切るいくつの文節に分かれますか。後から選び、番号で答えなさい。

今日はとてもよい天気だ。

1 3つ
2 4つ
3 5つ

解答 1 2 3

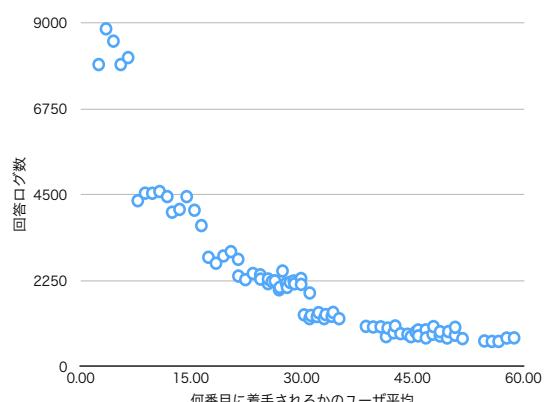


図 3.1: 小学4年国語の問題と回答選択肢の例

図 3.2: 小学4年国語の平均着手順と回答ログ数のXYプロット

小学4年国語では、主に、漢字の読み、あらすじ把握、心情把握、文意把握、接続語、慣用句、分節、文系、指示語などの品詞やその分類、敬語などの基本内容が扱われる。図3.1に、問題と回答選択肢の例を示す。図の問題は日本語の文法のうち文節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3.2に示す。全体として、概ね千件以上の回答ログ数があり、最初の問題に回答ログが大きく偏っていることが分かる。

小学 4 年社会

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

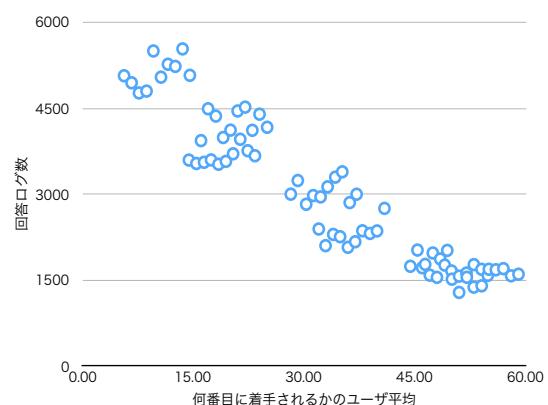


図 3.3: 小学 4 年社会の問題と回答選択肢の例

図 3.4: 小学 4 年社会の平均着手順と回答ログ数の XY プロット

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 3.3 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 3.4 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 4 年算数

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 3.5 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 3.6 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。

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。

99×9

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 3.5: 小学 4 年算数の問題と回答選択肢の例

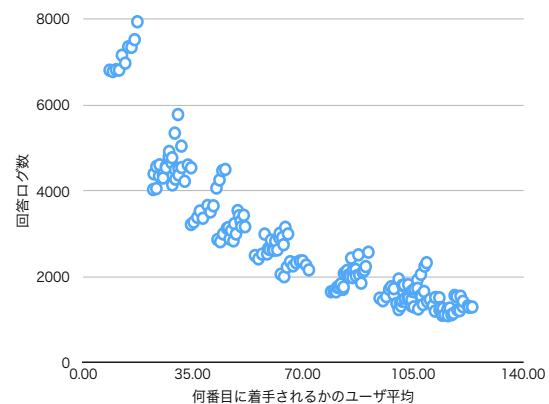


図 3.6: 小学 4 年算数の平均着手順と回答ログ数の XY プロット

小学 4 年理科

問3 右のような試験管にとりつけたガラス管のと中まで水をいれて水面に印をつけました。この試験管を湯の中に入れると、ガラス管の水面の高さはどうなりますか。番号で答えなさい。

ガラス管

① 高くなる。
② 高くなったり低くなったりする。
③ 低くなる。
④ 変わらない。

解答 → 1 2 3 4

図 3.7: 小学 4 年理科の問題と回答選択肢の例

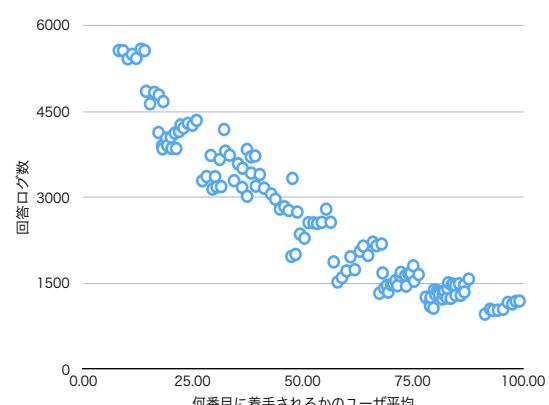


図 3.8: 小学 4 年理科の平均着手順と回答ログ数の XY プロット

小学 4 年理科では、主に、昆虫の種類、植物の育て方、気温、モーター、回路、電池、人体、天体、注射器、気体の体積などの科学の基本内容が扱われる。図 3.7 に、問題と回答選択肢の例を示す。図の問題は温度や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。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 3.8 に示す。全体として、概ね 1,000 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 年国語

問4 次の文の主語を後から選び、番号で答えなさい。

春子さん、今日のあなたの演奏はすばらしかった。

1 春子さん
2 今日の
3 演奏は

解答 → 1 2 3

図 3.9: 小学 5 年国語の問題と回答選択肢の例

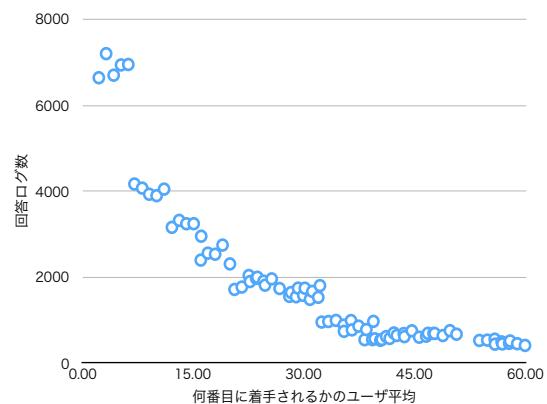


図 3.10: 小学 5 年国語の平均着手順と回答ログ数の XY プロット

小学 5 年国語では、主に、漢字の読み、あらすじ把握、情景把握、論題の把握、趣旨の把握、外来語、詩、動詞、形容詞、副詞、名詞などの品詞やその分類、敬語などの基本内容が扱われる。扱う内容は小学 4 年国語と近いが、小学 4 年のものと比べてより詳細である。図 3.9 に、問題と回答選択肢の例を示す。例えば、図の問題は主語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。この問題の回答には先の小学 4 年国語の分節への理解に加えて、主語とは何か、への理解が必要だと推察されるように、内容がやや詳細となっている。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 3.10 に示す。全体として、小学 4 年国語と同様に最初の問題の方が回答ログが偏っている傾向にある。

小学 5 年社会

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 3.11 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 3.12 に示す。

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 3.11: 小学 5 年社会の問題と回答選択肢の例

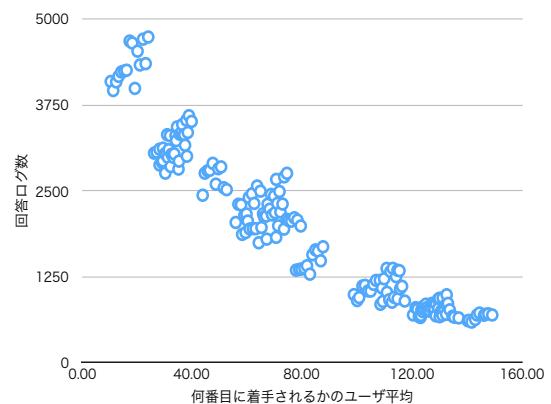
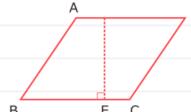


図 3.12: 小学 5 年社会の平均着手順と回答ログ数の XY プロット

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

問3 次のア、イにあてはまることばや式をあとから選び、番号で答えましょう。



平行四辺形の面積を求めるとき、1つの辺を底辺とすると、もう1つの向かい合う辺までのきょりをアといいます。平行四辺形の面積はイで求めることができます。

1 高さ 2 底辺×高さ 3 底辺×高さ÷2

解答 → ア ○ 1 ○ 2 ○ 3
イ ○ 1 ○ 2 ○ 3

図 3.13: 小学 5 年算数の問題と回答選択肢の例

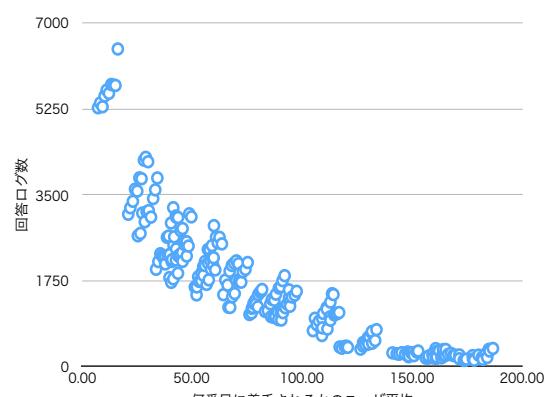


図 3.14: 小学 5 年算数の平均着手順と回答ログ数の XY プロット

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 3.13 に、問題と回答選択肢の例を示す。図の問題は平行四辺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 3.14 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 5 年理科

問4 川の水の量がふえると土や石などを運ぶはたらきはどうなりますか。番号で答えなさい。

① 運ぶはたらきが小さくなり、川底に土や石などがたまりやすく水のようすもきれいになる。

② 運ぶはたらきはあまり変わらず、川のようすにとくに変わりはない。

③ 運ぶはたらきは大きくなり、たくさんの土や石を流すので水のようすもごっている。

解答 1 2 3

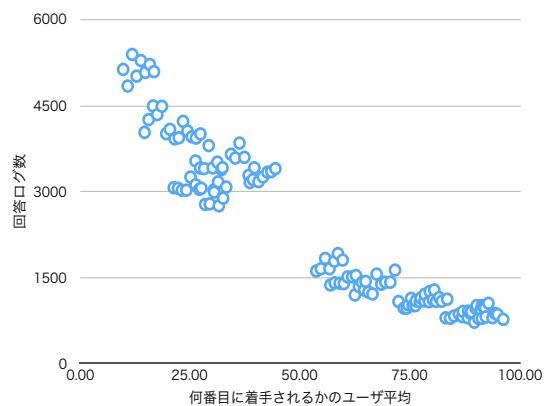


図 3.15: 小学 5 年理科の問題と回答選択肢の例

図 3.16: 小学 5 年理科の平均着手順と回答ログ数の XY プロット

小学 5 年理科では、主に、種子、光合成、プランクトン、花、メダカ、台風、流れの作用、ふりこなどの科学の基本内容が扱われる。図 3.15 に、問題と回答選択肢の例を示す。図の問題は川の流れの作用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 3.16 に示す。全体として、概ね 800 件以上の回答ログ数があり、小学 4 年理科の分布と概ね同じである。

小学 6 年国語

小学 6 年国語では、主に、あらすじの把握、主旨の把握、主題の把握、内容の把握、情景の把握、短歌の意味の把握、文意の把握登場人物の心情の把握、ことわざ、俳句、季語、切れ字、同義語、同訓異字、同音異字、対義語、慣用句、主語、修飾・被修飾、形容動詞、接続語、形容詞、動詞などの内容が扱われる。図 3.17 に、問題と回答選択肢の例を示す。例えば、図の問題は対義語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。対義語は単に後の意味を理解するだけでなく、後のペアの関係性を理解している必

問5 次の反対語を後から選び、番号で答えなさい。

否定

1 決定
2 肯定
3 公定
4 一定

解答 → ○ 1 ○ 2 ○ 3 ○ 4

図 3.17: 小学 6 年国語の問題と回答選択肢の例

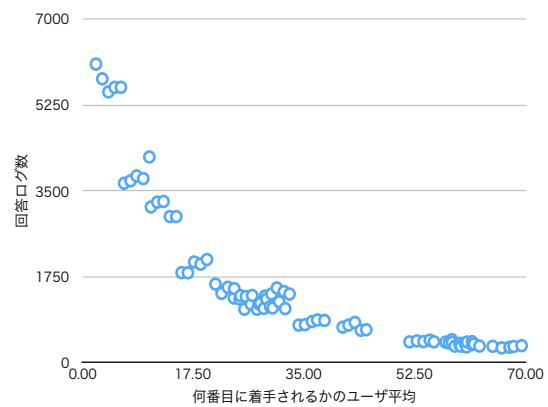


図 3.18: 小学 6 年国語の平均着手順と回答ログ数の XY プロット

要がある発展的な内容と捉えられる。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 3.18 に示す。小学 4 年、小学 5 年国語と同様に最初の問題の方が回答ログが偏っている傾向にある。

小学 6 年社会

問5 次の(ア)(イ)にあてはまる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)が完成了しました。

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

解答 → ○ 1 ○ 2 ○ 3 ○ 4

図 3.19: 小学 6 年社会の問題と回答選択肢の例

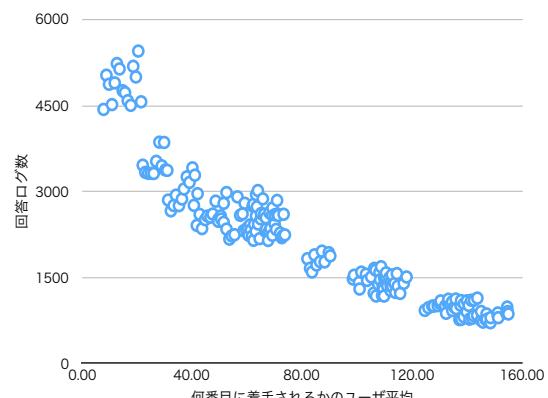


図 3.20: 小学 6 年社会の平均着手順と回答ログ数の XY プロット

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 3.19 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3.20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。

以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] cm³ 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3.21: 小学6年算数の問題と回答選択肢の例

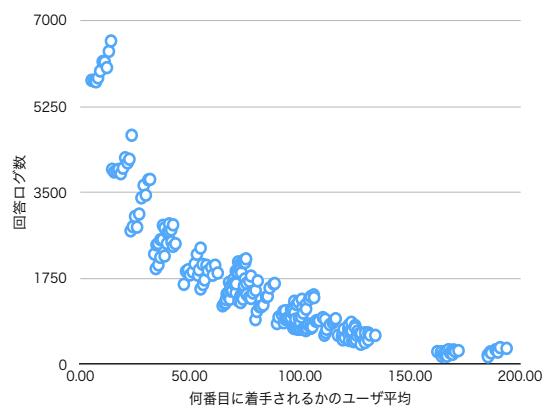


図3.22: 小学6年算数の平均着手順と回答ログ数のXYプロット

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3.21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3.22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．

小学6年理科

小学6年理科では，主に，いろいろな水溶液，てこのはたらき，ふり子，人と環境，動物の呼吸と空気，地層，太陽と月，植物の成長，火山のはたらき，物の燃

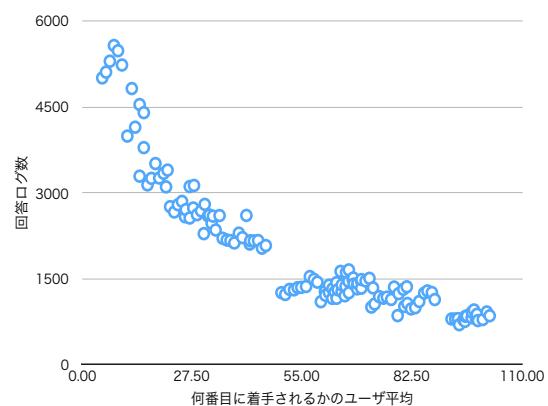
問2 くぎぬきはてこを利用した道具です。右のくぎ
ぬきのア～ウは支点、力点、作用点のどれにあた
りますか。正しい答えの組み合わせを番号で答
えなさい。



① ア 支点 イ 力点 ウ 作用点
 ② ア 作用点 イ 力点 ウ 支点
 ③ ア 力点 イ 作用点 ウ 支点
 ④ ア 力点 イ 支点 ウ 作用点

解答 ○ 1 ○ 2 ○ 3 ○ 4

図 3.23: 小学 6 年理科の問題と回答選択肢の例

図 3.24: 小学 6 年理科の平均着手順と回
答ログ数の XY プロット

え方と空気、血液のはたらき、金属と水よう液、電気のつくり方、電流と熱などの内容が扱われる。図 3.23 に、問題と回答選択肢の例を示す。図の問題はてこのはたらき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 3.24 に示す。全体として概ね 800 件以上の回答ログ数があるが、算数と同様で、小学 4 年、小学 5 年の理科と比べると下に凸になっており、最初の方の問題への回答ログの偏りがやや大きいことが分かる。

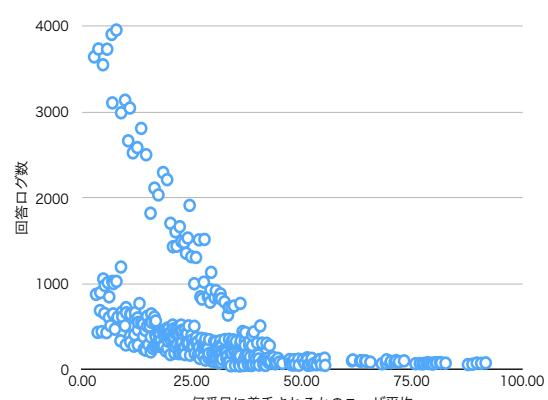
中学 1 年国語

問4 テキスト 192 ~ 193 ページの文章に関する問題です。
 193 ページ 22 行目
 「これは言葉の世界での出来事と同じことではないか」とあります、桜
の木全体にあたるのは次のどれですか。後から選び、番号で答えなさい。

1 言葉
 2 言葉を発する人
 3 言葉を発する季節
 4 言葉を発する場所

解答 ○ 1 ○ 2 ○ 3 ○ 4

図 3.25: 中学 1 年国語の問題と回答選択肢の例

図 3.26: 中学 1 年国語の平均着手順と回
答ログ数の XY プロット

中学1年国語では、大きく分けて古文、漢文、現代文の3つの内容を扱う。図3.25に、問題と回答選択肢の例を示す。例えば、図の問題は文意の把握に関する問題である推察される。テキストはオンライン上でダウンロードできる教材であり、こうした問題に際しては、学習者はテキストをダウンロードし文章を読み、回答をオンラインで行う。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3.26に示す。着手されるのが早い問題でも必ずしも回答ログ数が大きいというわけではなく、2つの集合に分かれていることがわかる。これは、学習者が選択する教材に応じて出題されうる問題群が部分的に異なるためである。また、後ろの方の問題の回答ログ数は非常に少なく、問題によっては数十程度のレコード数しかないことも1つの特徴である。これは、学校のカリキュラムの影響を受けているものと考えられる。つまり、当該データは4月から11月のものであるが、11月の時点では学校の学習カリキュラムは完了しておらず、また、少なうないユーザが学校の学習カリキュラムに合わせて利用しているからであると推察できる。

中学1年数学

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、
あととの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
イ 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

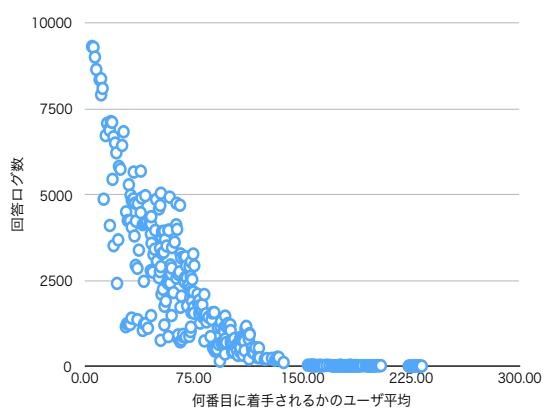


図3.27: 中学1年数学の問題と回答選択肢の例

図3.28: 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3.27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3.28に示す。回答ログは最初の方の問題に大きく偏っていることが分かる。また、中学1年国語と同様に最後の方の問題は回答ログ数が非常に少ない。

中学1年英語

問2 次の下線部を複数形にした文を番号で答えなさい。

This is a blue pen.

1 These are blue pen.
2 This are blue pen.
3 These are blue pens.
4 These is blue pen.

解答 → 1 2 3 4

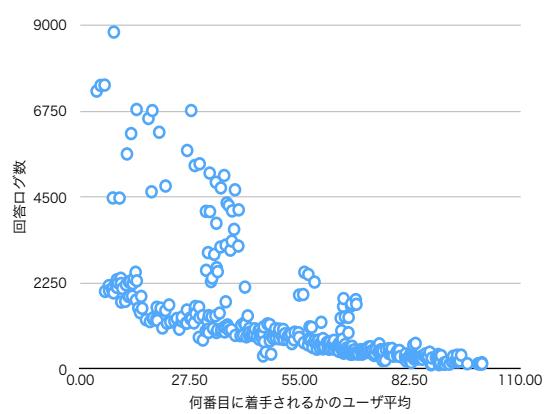


図3.29: 中学1年英語の問題と回答選択肢の例

図3.30: 中学1年英語の平均着手順と回答ログ数のXYプロット

中学1年英語では、主に、文法に関する内容が扱われる。図3.29に、問題と回答選択肢の例を示す。図の問題は文法の特に、代名詞を複数形する問題であり、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3.30に示す。中学1年国語のデータセットと同様に2つの問題集合に分かれているが、英語の方が後半の問題についても回答ログは100以上存在している。

中学2年国語

中学2年国語では、中学1年国語と同様に、古文、漢文、現代文を扱う。図3.31に、問題と回答選択肢の例を示す。図の問題は、語の意味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。必ずしも、語のみから判断することができるというわけではなく、文意の把握が必要であり、難易度

問4 教科書134ページの文章に関する問題です。

134ページ 3行目
「たけき者」の意味を後から選び、番号で答えなさい。

1 富めるもの
2 武に強い人
3 運の良い人

解答 ○ 1 ○ 2 ○ 3

図 3.31: 中学2年国語の問題と回答選択肢の例

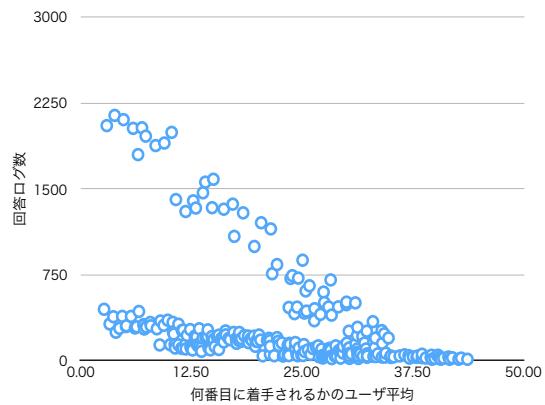


図 3.32: 中学2年国語の平均着手順と回答ログ数のXYプロット

は高いと推察される。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3.32に示す。中学1年国語のデータセットと同様に2つの問題集合に分離しているが、特に、回答ログが少ない方は、数十から200程度しかログがない。

中学2年数学

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 $\boxed{y} = 10 - 3 = 7$ になるので、 $y = \boxed{7}$ 、 $x = \boxed{2}$ 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 3.33: 中学2年数学の問題と回答選択肢の例

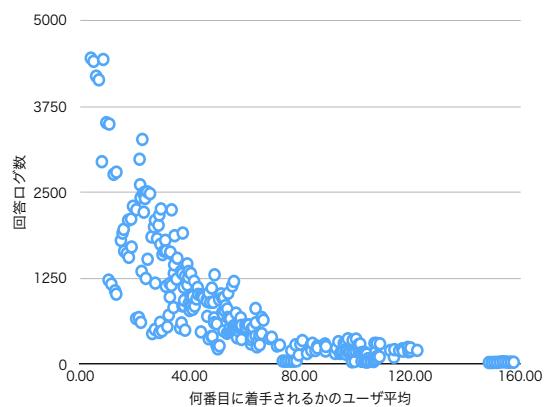


図 3.34: 中学2年数学の平均着手順と回答ログ数のXYプロット

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3.33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3.34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、ログが数十程度と非常に少なくなっている。

中学2年英語

問3 次の下線部を尋ねる文を後から選び番号で答えなさい。

She is going to visit New York next year.

1 When is she going to visit New York?
 2 Who is going to visit New York next year?
 3 Where is she going to visit next year?
 4 How is she going to visit New York next year?

解答 1 2 3 4

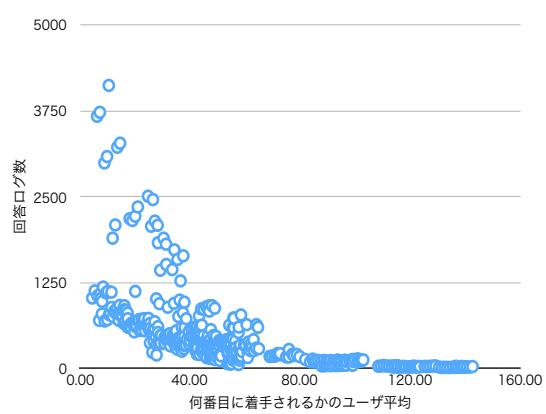


図3.35: 中学2年英語の問題と回答選択肢の例

図3.36: 中学2年英語の平均着手順と回答ログ数のXYプロット

中学2年英語では、主に、英語の文法に関する問題が扱われる。図3.35に、問題と回答選択肢の例を示す。図の問題は文法の特に疑問文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3.36に示す。グラフの外見は中学1年英語に近いが、回答ログ数のスケールが半分程度である。また、一部の問題については、ログが数十程度と非常に少なくなっている。これらの特徴は中学1年数学と中学2年数学の関係と類似している。

問4 テキスト 148 ~ 150 ページの文章に関する問題です。

目の不自由な人が、いろいろなことを実験や体験を通じて理解することで得られる利点は何ですか。後から選び、番号で答えなさい。

- 1 理解するのに時間がかかる分、忘れない点。
- 2 目が見える人と違う感じ方ができる点。
- 3 子どものころの感覚を保つことができる点。
- 4 やる気を失わない点。

解答 → ○ 1 ○ 2 ○ 3 ○ 4

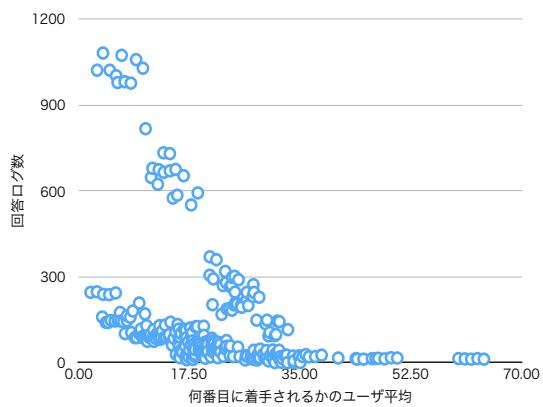


図 3.37: 中学3年国語の問題と回答選択肢の例

図 3.38: 中学3年国語の平均着手順と回答ログ数のXYプロット

中学3年国語

中学3年国語では、古文、漢文、現代文の内容が扱われる。扱う項目は中学1年国語と中学2年国語と近いが、これらと比べて内容がより詳細である。図3.37に、問題と回答選択肢の例を示す。例えば、図の問題は現代文のうち文意把握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3.38に示す。中学1年国語、中学2年国語と類似した外観となっているが、これらよりもログの偏りが大きく、少なくない問題で回答ログ数が100以下となっている。

中学3年数学

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3.39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3.40に示す。グラフの

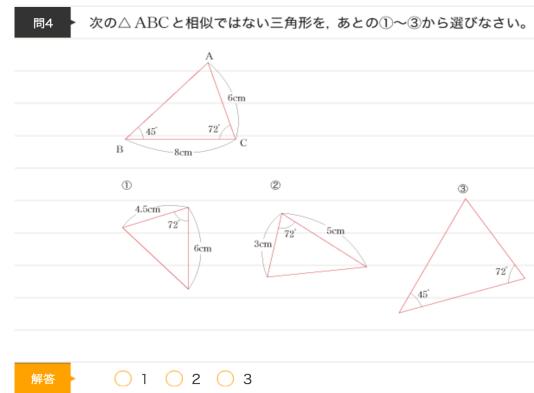


図 3.39: 中学3年数学の問題と回答選択肢の例

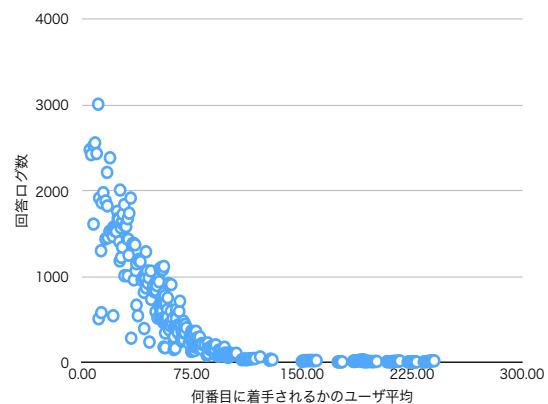


図 3.40: 中学3年数学の平均着手順と回答ログ数のXYプロット

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3年英語

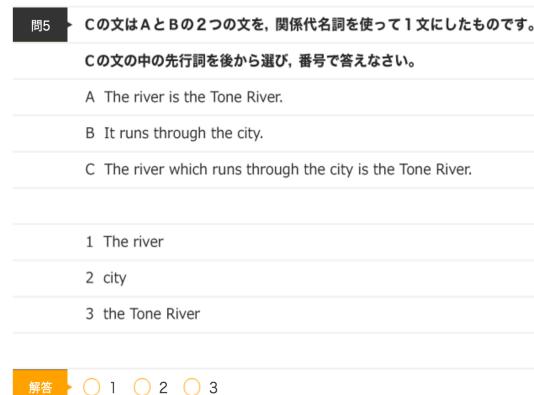


図 3.41: 中学3年英語の問題と回答選択肢の例

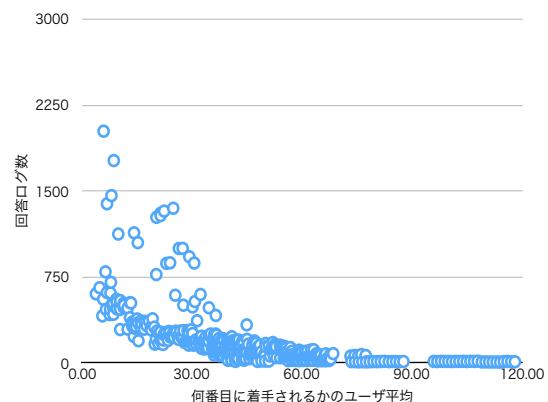


図 3.42: 中学3年英語の平均着手順と回答ログ数のXYプロット

中学3年英語では、主に、英語の文法に関する問題が扱われる。図3.41に、問題と回答選択肢の例を示す。図の問題は文法の特に関係代名詞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3.42に示す。グラフの外観は中学2年英語に強く類似しており、回答ログ数が3分の2程度に減少しているところが特徴的な違いである。

中学地理

問1 次は地域調査の計画を立てるまでの手順を表したものです。(ア)と(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後から選び、番号で答えなさい。

情報を集める→(ア)を決める→(イ)を立てる→調査計画を立てる

1 ア テーマ イ 戰略
2 ア テーマ イ 仮説
3 ア 助手 イ 戰略
4 ア 助手 イ 仮説

解答 ○1 ○2 ○3 ○4

図 3.43: 中学地理の問題と回答選択肢の例

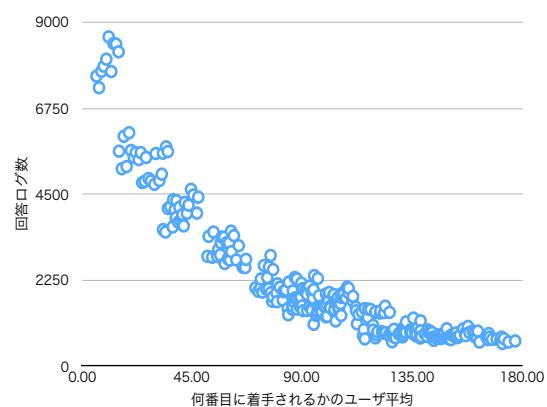


図 3.44: 中学地理の平均着手順と回答ログ数のXYプロット

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 3.43に、問題と回答選択肢の例を示す。図の問題は地域調査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 3.44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

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 3.45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

問7 次の()にあてはまるご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

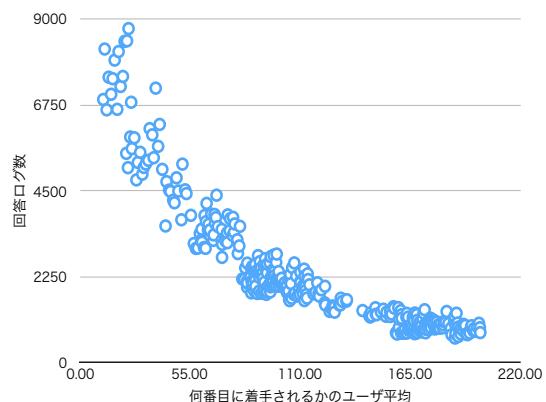


図 3.45: 中学歴史の問題と回答選択肢の例

図 3.46: 中学歴史の平均着手順と回答ログ数の XY プロット

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 3.46 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。

中学公民

問7 次は日本国憲法第14条の条文の一部です。()に入ることばを後から選び番号で答えなさい。

すべて国民は、()の下に平等であって、人種、信条、性別、社会的身分又は門地により、政治的、経済的又は社会的関係において、差別されない。

1 社会
2 法
3 国会

解答 1 2 3

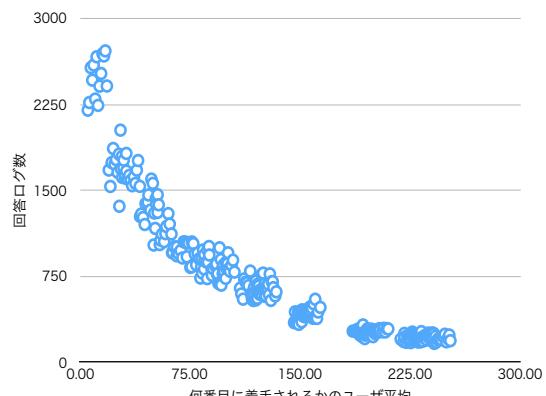


図 3.47: 中学公民の問題と回答選択肢の例

図 3.48: 中学公民の平均着手順と回答ログ数の XY プロット

中学公民では，主に，人権と日本国憲法，内閣（行政権），国会（立法権），国民主権と平和主義，国際社会のしくみ，国際社会の課題，地方自治，基本的人権

の尊重，市場経済，日本経済の課題，消費生活，現代の民主政治，現代の社会，生産のしくみと企業，私たちの生活・文化，裁判所（司法権），財政と福祉，などの内容が扱われる。図3.47に，問題と回答選択肢の例を示す。図の問題は人権と日本国憲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3.48に示す。グラフの外観は中学地理や中学歴史と類似しているが，回答ログ数のが3分の1程度となっている。

中学理科 1

問9 右の図で $R_1 = 20\Omega$, $R_2 = 30\Omega$ です。全体の電圧が $12V$ のとき、 R_1 に流れる電流は何 A ですか。次から選び、番号で答えなさい。



① 0.2A ② 0.3A ③ 0.6A
④ 2A ⑤ 3A

図 3.49: 中学理科 1 の問題と回答選択肢の例

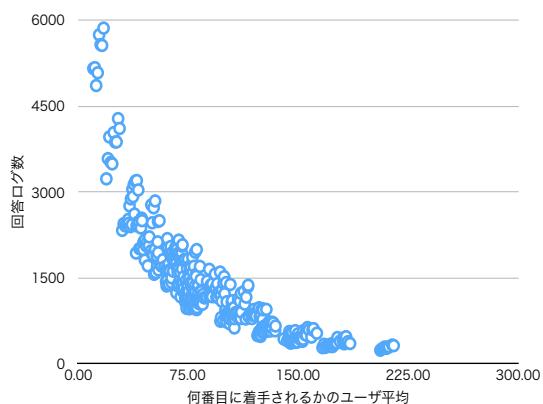


図 3.50: 中学理科 1 の平均着手順と回答ログ数の XY プロット

中学理科1では、主に、中和とイオン、中和計算演習、二酸化炭素 アンモニア 塩化水素の発生とその性質、仕事と仕事率、光、凸レンズ、力のつり合い、力の大きさとその表し方、力学的エネルギーとその移り変わり、化学反応計算演習、化学変化と物質の質量、化学変化の仕組みと化学反応式、回路とオームの法則・電流・電圧、圧力、水溶液の濃さ、物体の運動、物体の運動の計算演習、物質の状態変化、状態変化と体積・質量・温度、窒素 塩素 二酸化硫黄 ・気体の性質のまとめ、酸とアルカリ・中和、酸・アルカリとイオン、酸化と還元・化学変化と熱、酸素・水素の性質と発生、電気とそのエネルギー、電気分解・電池とイオン、電流と磁界、電流が磁界から受ける力、電磁誘導と発電機、電解質・非電解質とイオン、静電気と電流、音、などの内容が扱われる。図3.49に、問題と回答選択肢の例を示す。図の問題は回路とオームの法則・電流・電圧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3.50に示す．全体として，概ね300件以上の回答ログ数があり，やや最初の問題の方が回答ログが大きい傾向にあるが，前半の問題から後半の問題にかけて概ね線形に回答ログ数が減少している．

中学理科2

問9 次の〔ア〕, 〔イ〕にあてはまることはの組み合わせを後から選び, 番号で答えなさい。

生物は1つの細胞でできている〔ア〕と, 複数の細胞でできている〔イ〕に分類することができます。

① ア 単細胞生物 イ 多細胞生物
 ② ア 単細胞生物 イ 複細胞生物
 ③ ア 多細胞生物 イ 単細胞生物

解答 → ○ 1 ○ 2 ○ 3

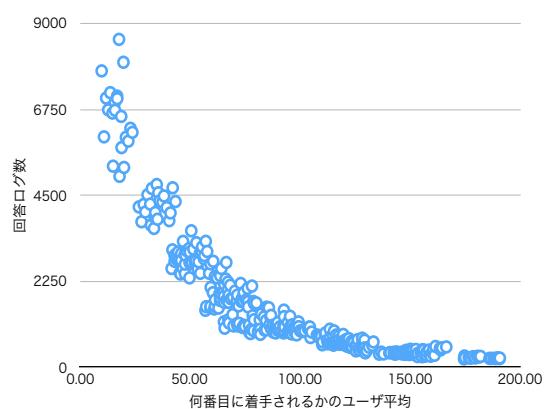


図3.51: 中学理科2の問題と回答選択肢の例

図3.52: 中学理科2の平均着手順と回答ログ数のXYプロット

中学理科2では，主に，光合成と呼吸，刺激と反応，前線と天気の変化，動物のなかま，呼吸器官とそのはたらき・血液の成分，土の中の生物とはたらき，地層，地震のゆれとその伝わり方，大気の働き，大気の動きと日本の天気，天体の1年の動きと地球の公転，天体の1日の動きと地球の自転，天体計算演習，太陽と月，季節の変化，惑星・恒星と太陽系，有性生殖と遺伝の規則性，根・茎のつくり，植物全体の分類，消化と吸収，火山，環境問題，生物と細胞，生物の変遷と進化，生物の成長と無性生殖，空気中の水蒸気，胞子でふえる植物，花のつくりとはたらき，葉のつくりとそのはたらき，血液の循環と排出，身近な生物の観察，雲・霧の発生と水の循環，震度とマグニチュード，地震が起こる原因，骨格と筋肉・感覚器官，などの内容が扱われる．図3.51に，問題と回答選択肢の例を示す．図の問題は生物と細胞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3.52に示す．全体として，概ね300件以上の回答ログ数があり，

やや最初の問題の方が回答ログが大きい傾向にあるが、前半の問題から後半の問題にかけて概ね線形に回答ログ数が減少している。特に中学理科1と比べると最初の方の問題の回答ログ数が大きい。

第4章 実験設定

4.1 実験項目

まず，実験の目的について，本実験では下記の項目を検証することが主な目的である．

目的1 DKT のモデルが数学のデータセットに有効であることを再現し，また，他の科目的データセットに対する有効性を検証すること．

目的2 DKT のモデルと BKT のモデルの予測性能の差の要因を明らかにすること．

そこで，これらの目的を達成するために，先に説明した 26 のデータセットを題材に下記の実験を行う．まず，目的1を達成するために，DKT と BKT を 26 のデータセットに適用して，それぞれの予測精度を比較する．以降では，この実験を実験1と呼ぶこととする．数学に相当するデータセットは小学生の算数 3 データセットと中学性の数学の 3 データセットの合計 6 データセットを用いる．他の科目ごとの分析については，国語に相当するデータセットとして小学生の国語の 3 データセットと中学生の国語の 3 データセットの合計 6 データセットを用い，地理に相当するデータセットとして小学 5 年社会のデータセットと中学地理のデータセットの合計 2 データセットを用い，歴史に相当するデータセットとして小学 6 年社会のデータセットと中学歴史のデータセットの合計 2 データセットを用い，理科に相当するデータセットは小学性の理科の 3 データセットと中学理科 1 と中学理科 2 の合計 5 データセットを用い，英語に相当するデータセットとして中学生の英語の 3 データセットの合計 3 データセットを用い，その他はそれぞれ個別に評価する．こうして，分けて評価することで科目ごとの DKT の適用可能性を検証できる可能性がある．

実験は 26 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 [61] 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，大規模オンライン講座の学習コンテンツもまた，更新頻度は低い。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ドスタート問題 [62] 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，目的 2 を達成するために，実験 1 で得られた DKT と BKT の AUC の差とデータセットの統計量との相関を評価する。以降では，この実験を実験 2 と呼ぶこととする。統計量には先のデータセットの章で取り上げたユーザ数，問題数，回答ログ数，回答ログ数 ÷ ユーザ数，回答ログ数 ÷ 問題数の 5 つの指標を用いる。ユーザ数と AUC の差の相関が強ければ，DKT のモデルが捉える問題間の関係性の網羅性とユーザ数の関連が強いということであり，ユーザにとって問題間の関連性が多様であることを示唆する可能性がある。問題数と AUC の差の相関が強ければ，DKT のモデルの予測性能が DKT のモデルが捉える問題間の関係性に基づいているという推察と合致する。回答ログ数と AUC の差の相関が強さからは，DKT のモデルを十分学習されるために必要な回答ログ数に達しているかどうかを推察できる。回答ログ数 ÷ ユーザ数と AUC の差の相関が強ければ，DKT のモデルの予測性能と回答ログ系列の長さの関連が強いことを示唆する。回答ログ数 ÷ 問題数と AUC の差の相関が強ければ，特に問題あたりのログ数という観点で DKT のモデルを十分学習されるために必要な回答ログ数に達しているかどうかを推察できる。TODO : AUC 差の要因特定になるように文言を書き換える。

以上の分析を対象のデータセットに応じて 3 つの実験を実施する。この実験を 26 のデータセットを用いて行う実験（実験 2-1）と小学生の 12 データセットを用いて行う実験（実験 2-2），中学生の 14 データセットを用いて行う実験（実験 2-3）の 3 つの実験である。これは，小学生と中学生で扱われる問題の特性が異なる可能性を考えて分けた実験するという意図である。

以上，実験 1 と実験 2 の実験内容について述べた次に，実験で用いるデータセットの問題回答が同時に複数されうるという特徴を考慮するために，DKT の手法に加える修正点について述べ，修正を加えた DKT の手法のハイパーパラメタの設定

について述べる。

4.2 データセットに合わせた手法の修正

実験で用いるデータセットの問題回答が同時に複数されうるという点で [15] の問題設定と異なるため、特に、精度向上を狙った修正ではなく、尤もらしい形で対象データセットに適用できるようにすることを狙って DKT に修正を加える。BKT については問題間のインタラクションは考慮していないため修正は行わない。

DKT では 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 x_t とし、演習問題の数を M とすれば、 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環境下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、[15] に近い形で RNN に情報を入力できる。

表 4.1: 勉強サプリの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	x_1	[1000:0100]
A	1	2	0		
A	2	1	1	x_2	[1100:0000]
A	2	2	1		
A	3	3	0	x_3	[0000:0010]
A	4	3	0		
A	4	4	1	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、 x_t の長さは 8 である。ある学習者が、表 4.1 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 x_t は表 4.1 の入力ベクトルのように符号化される。

出力 y_t は DKT 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 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。

4.3 最適化

最適化手法は [15] のものとほとんど同じである。

$$\log(p_1 \times p_2 \times \cdots \times p_{m_t}) = \sum_k^{m_t} \log(p_k) \quad (4.1)$$

であるため， $\tilde{\delta}(q_{t+1})$ を時刻 $t + 1$ にどの問題が回答されたかの m_t -hot ベクトルとすれば，ロス関数は

$$L = \sum_t l(\mathbf{y}^T \tilde{\delta}(q_{t+1}), q_{t+1}) \quad (4.2)$$

である。学習時は [15] 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，RNN は GRU-RNN を学習率の初期値を 200，モーメントを 0.98，1epoch ごとに，減衰率を 0.8 で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 [46] に従い勾配に制約を設けた。dropout は [15] と同様に 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，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 [63] といった探索手法が提案されてるが，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 [63, 64]。

第5章 結果と分析

5.1 実験1：DKTとBKTの性能差とデータセットの科目

DKTとBKTを26データセットに適用した結果を表5.1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。[15]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ついてDKTのAUCはBKTのAUCよりも高かった。これは、数学だけでなく、国語、社会、理科、英語等のさまざまな科目において問題間の関連性を考慮することで程度に差はあるものの知識獲得のモデリング性能が向上することを示している。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、また、DKTのAUCとBKTのAUCの差の平均も0.08となっており、数学のデータセットにおけるDKTの有効性を再現できていることを示している。他の科目的データセットについては、特に、歴史と地理の社会科目について、DKTのAUCがBKTのAUCを大きく上回っていた。このことから、歴史や地理の知識獲得において関連知識間で相互に影響を与えることが示唆されている。国語の6つのデータセットはいずれもDKTのAUCが0.8を下回っており、また、DKTのAUCとBKTのAUCの差も0.054と他の科目群と比べて低かった。理科はDKTのAUCとBKTのAUCの差についても他の科目と比べて高いというわけではなかったが、低いというわけでもなかった。英語の3つのデータセットについて、いずれもDKTのAUCとDKTのAUCとBKTのAUCの差は低かった。

全体の結果を踏まえて、まず、数学のデータセットにおいてDKTが有効であることを確認した。次に、他の科目的データセットについてもDKTがBKTよりも有効であることを、特に、地理と歴史に関する社会科目的データセットにおいて

表 5.1: 各データセットに対する各手法の予測結果とそれらの関係性 .

科目	学年	AUC			DKT – BKT	
		Marginal	BKT	DKT	値	平均
算数・数学	小学4年	0.696	0.707	0.828	0.121	
	小学5年	0.714	0.724	0.804	0.081	
	小学6年	0.734	0.749	0.836	0.087	
	中学1年	0.738	0.750	0.807	0.057	0.082
	中学2年	0.675	0.696	0.773	0.077	
	中学3年	0.729	0.735	0.804	0.069	
国語	小学4年	0.680	0.724	0.769	0.045	
	小学5年	0.663	0.695	0.778	0.083	
	小学6年	0.693	0.733	0.790	0.058	
	中学1年	0.680	0.716	0.766	0.050	0.054
	中学2年	0.675	0.681	0.718	0.037	
	中学3年	0.639	0.657	0.709	0.052	
理科	小学4年	0.749	0.777	0.811	0.035	
	小学5年	0.664	0.684	0.768	0.084	
	小学6年	0.684	0.704	0.772	0.068	0.070
理科1	中学	0.665	0.696	0.765	0.069	
理科2	中学	0.639	0.669	0.765	0.096	
英語	中学1年	0.685	0.691	0.757	0.065	
	中学2年	0.627	0.635	0.734	0.099	0.074
	中学3年	0.645	0.656	0.714	0.057	
地理・社会	小学5年	0.647	0.680	0.765	0.085	
	中学	0.655	0.681	0.764	0.083	0.084
歴史・社会	小学6年	0.642	0.657	0.773	0.116	
	歴史	0.635	0.670	0.766	0.096	0.106
社会	小学4年	0.701	0.739	0.791	0.052	-
公民	中学	0.617	0.645	0.770	0.125	-

有効性が高いことを，国語と英語に関するデータセットにおいて有効性が低いことを示した．これらのことと踏まえると，日常生活で用いる知識や技術というよりは，何に役に立つのだろうか，と学習者が思ってしまうような科目の方が DKT の有効性は高いかもしだれない．

以上，実験 1について述べた．

5.2 実験2：DKTとBKTの性能差とデータセットの統計量

表 5.2: DKT の AUC – BKT の AUC と 各データセット分類に含まれるデータセットの統計量 の相関係数 .

データセットの分類	データセットの統計量				
	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
全体	0.04	0.20	0.37	0.67	0.07
小学生	-0.08	0.74	0.54	0.66	-0.44
中学生	0.11	0.22	0.35	0.74	0.35

DKT の AUC と BKT の AUC の差と各データセット分類に含まれるデータセットのユーザ数 , 問題数 , 回答ログ数 , 回答ログ数 ÷ ユーザ数 , 回答ログ数 ÷ 問題数の 5 つの指標の相関係数を表 5.2 に示す . データセット分類の全体は 26 データセットを対象とした統計量であり , 小学生は小学生の 12 データセットを対象とした統計量であり , 中学生は中学生の 14 データセットを対象とした統計量である . 太字の部分は相関係数の絶対値が 0.4 以上のものである .

まず , 回答ログ数 ÷ ユーザ数と AUC の差の相関係数はいずれのデータセット分類でも正で 0.65 から 0.75 程度と大きかった . このことから , ユーザ一人あたりの回答ログ数が大きいデータセットに基づいてモデルを構築することが重要であると言える . また , 回答ログ数が多いということは , 同じ問題を繰り返し回答するという復習が多いということか , 前半の問題だけでなく後半の問題を着手しているということであり , いずれにせよ , データの時間方向の系列長が長いことを示している . つまり , 過去の学習行動の結果に基づいた知識獲得モデリングは長期的なデータセットに基づいてモデリングされることが望ましいということであり , また , 過去の学習行動の結果に起因する学習効果に基づいて将来の回答行動の予測結果が変わるという点で知識獲得モデリングが学習効果のモデリングの一環であることから , 学習効果の評価は長期的なスパンで行う方が学習効果をよくモデリングできる可能性があるということを同時に示唆していると考えられる .

次に , ユーザ数と AUC の差の相関係数について , 相関係数の絶対値はいずれ

のデータセット分類においても小さかった。DKT と BKT の主要な違いの 1 つが DKT では問題間の関連性を捉えて予測に活用しているということであるが、この結果から、DKT のモデルが捉える問題間の関係性とデータセットのユーザ数の大小はほとんど関係がないことがわかる。したがって、問題間の関係性はユーザごとに多様であるというよりはユーザで概ね共通しているのかもしれない。

回答ログ数と AUC の差はいずれのデータセット分類でも相関係数が正であり弱い相関から中程度の相関であった。回答ログ数 ÷ ユーザ数の相関係数と比較してみると回答ログ数の大きさが重要であるというよりは、ユーザー人あたりの回答ログ数の大きさが重要であり、その影響が受けているのではないかと考えられる。

最後に、問題数と AUC の差について、と、回答ログ数 ÷ 問題数と AUC の差について、の 2 つについて述べる。データセット全体としては、2 つのいずれも相関は弱いが、小学生と中学生の分類においては、相関係数は大きく異なる。小学生のデータセットでは、問題数と AUC の差については、相関係数が 0.74 と正の強い相関を示している一方で、回答ログ数 ÷ 問題数と AUC の差については、-0.44 と負の中程度の相関を示している。BKT では繰り返し同じ問題や同じ知識を扱う問題に回答するなかで習得するという学習過程をモデリングしており、問題あたりの回答ログ数がより大きいデータセットであれば、BKT の予測性能に有利であると考えられ、その影響により回答ログ数 ÷ 問題数と AUC の差が負の中程度の相関となっていると推察できる。一方で、そのような問題 1 問あたりのログ数が十分大きいようなデータセットに対しては、問題数が大きいものの方が予測精度が高くなっているという結果になっており、データセットが十分大きのであれば、予測に使える情報が多ければ多いほど知識獲得をモデリングしやすくなることが示唆されている。一方で、中学生のデータセットでは、問題数と AUC の差と回答ログ数 ÷ 問題数と AUC の差がいずれも正の弱い相関となっており、問題 1 問あたりの回答ログ数が当該データセットをモデリングするのに足りていないことを示唆しているが、このことは、中学生のデータセットの一部の問題に対して回答ログが非常に少なかったことと合致している。

以上、実験 2 について述べた。

5.3 実験結果の整理

それぞれの実験の目的について、目的が十分達成されたかどうかを結果を踏まえて整理する。

目的1 DKT のモデルが数学のデータセットに有効であることを再現し、また、他の科目的データセットに対する有効性を検証すること。

まず、数学のデータセットにおいて DKT が有効であることを算数と数学に関する 6 つのデータセットへの実験と通じて再現した。次に、他の科目的データセットについても DKT が BKT よりも有効であることを示した。特に、地理と歴史に関する社会科目的データセットにおいて有効性が高いこととまた、英語と国語のデータセットにおいて有効性が低いことを示した。

目的2 DKT のモデルと BKT のモデルの予測性能の差の要因を明らかにすること。

DKT と BKT の性能差とデータセットの統計量の分析から回答ログ数 ÷ ユーザ数が DKT と BKT の性能差と強い正の相関を示していたが、回答ログ数 ÷ ユーザ数の値が大きいデータセットから順に、小学 6 年社会、小学 5 年社会、中学歴史、中学公民、小学 6 年算数、小学 5 年算数、小学 4 年算数、中学地理となっており、いずれも科目ごとの分析において DKT と BKT の性能差が大きかった科目群に含まれるデータセットであった。したがって、必ずしも、数学や地理、歴史の科目で扱われる内容の内容間の関係性が他の科目で扱われる内容の内容間の関係性よりも DKT によって捉えやすいから予測精度が高かったとは限らないと考えられる。したがって、DKT のモデルと BKT のモデルの予測性能の差の要因については、DKT は長期的な学習行動のモデリングに効果的であり、データセットの回答ログ数 ÷ ユーザ数が大きいことが DKT のモデルと BKT のモデルの予測性能の差の要因の 1 つである、という考え方や DKT は特にに数学、地理、歴史の科目で扱われる内容の内容間関係性を特に捉えやすいことが、DKT のモデルと BKT のモデルの予測性能の差の要因の 1 つである、という考えができそうであるが、それ以上については、本実験からわからないため、データセットを整えて、再実験することが必要であろう。

以上、実験結果について整理した。

第6章 考察

心理学では知識は手続き的知識と宣言的知識に分けられて議論されることがある。手続き的知識とは、XXXに関する知識であり、宣言的知識は「AはBである」という形式で与えられる知識である。宣言的知識と手続き的知識の違いは手続き的知識が目的に基づいた知識であり汎用性は低い、と説明されることもある。例えば、数学の問題を例に挙げて考えてみる。1次方程式 $2x + 1 = 0$ を x について解く場合には、まず「1を移行する」という手続きを行い、その後に「両辺を2で割る」という手続きを行う。答えは $x = -0.5$ である。数学における移行という作業は定義されているが、実際には、1次方程式を解くという目的を達成するためには、手続きを適用することが求められている。したがって、宣言的知識として知っているということと実際に適用するための知識をもっているということが条件であり、また、2つの手続き的知識を持っていない場合には、この問題は解けない。したがって、一般に数学のある問題を解くためには、その問題を解くために必要な手続き的知識を持っている必要があり、このことから、数学においては知識間の関係性やそれが現れる問題間の関係性を考慮することが重要であることは理解が難しい話ではないと考えられる。Corebettらがはじめて発表した Knowledge Tracing についての論文のタイトルは *Knowledge tracing: Modeling the acquisition of procedural knowledge* であり、手続き的知識 (Procedual Knowledge) に関するものであった。しかし、本稿の実験から、数学だけでなく特に宣言的知識が中心であると推察される地理や歴史の社会科目についても問題間の関係性や知識間の関係性を考慮することが知識獲得モデルに有効であることが示唆されており不思議である。手続き的知識だけでなく宣言的知識の領域においても概念間の関連性を想起できるような形式で問題が提示されることが望ましいのかもしれない。

人間は視覚、聴覚、触覚、嗅覚、味覚を通して、現実世界の膨大のアナログ情報から有用な情報を選び分け認識し判断し行動する。こうした一連の活動を支える脳の中では、大規模かつ高次元の現実のデータをニューラルネットワークで処

理しているとされている。人間のニューラルネットワークが複数の知覚からの情報を処理することでより複雑でより高度な処理を行えると信じられているように、深層学習の研究においても複数の情報源を基にモデルを学習させるマルチモーダル学習という研究領域があり、近年の深層学習の発展を受けて注目が集まっている。モデルの入力デバイス（目や耳など）の増加や（その結果としての）入力次元の増加によってより高い精度で対象をモデリングできるという考えに基づいたものである。本稿の実験では、問題あたりの回答ログ数が十分多いデータセットについては、問題数が多いデータセットの方がモデリング性能が高い傾向にあった。知識獲得のモデリングというタスクは人間という対象からの出力のモデリングであるという点で、より大きな入力次元（問題数）からの情報のデータセットの方がモデリング性能が高い傾向にあるということは、違和感はない。対象を十分モデリングできるだけのレコード数が存在するのであれば、地理と歴史や社会と算数のように直接関係のない科目であっても（マルチモーダル学習のように）同時に利用することで、対象をよりよくモデリングできるかもしれない。また、この知識獲得モデリングの面白いところは、人間が対象であると同時に人間自身が学習モデルであるという再帰的な構造になっている点である。究極的には、人間という対象からのあらゆる出力を完全にモデリングできた時、人間という学習モデルを再現できたと言える可能性があり、そうした側面をささやかながら垣間見せた今回の実験結果は非常に興味深い。

本稿の実験では、個々の問題を基にして、Deep Knowledge Tracing を適用しても高い精度で予測できることを示した。^[15] では Excercise Tags もしくは Skill Tags と呼ばれるものが利用されており、それらは専門家が各問題に対して当該問題が扱うもしくは、当該問題の回答に必要な知識を適当な粒度にして割り当てたものである。こうした割り当て作業は非常にコストが大きく、今回個々の問題を基にして予測できることができたこと自体は面白い結果だったと考えられる。しかし、実際に予測に利用するという点では、Tag は必要であると考える。なぜなら、例えば、XX のデータセットではスキルの数は XX であるが、ユニークな問題の数は XX であり、また、XX のデータセットでもスキルの数は XX であるが、ユニークな問題の数は XX となっており、問題の数が非常に大きい。多くの問題を利用する場合にはそれを学習させるだけのデータセットが必要であるが、それだけの規模のデータを大きな入力次元で学習させることは計算コストの観点で簡単ではない

と考えられる。タグを利用せずに問題を基にモデルを構築したい場合には、RNNに入力する前に入力次元を Embedding などで圧縮することが有効であると考えられる。また、言語モデリングの領域でしばしば活用される階層的 softmax の手法が出力 y_t の計算に有効であるかもしれない。

TODO：系列が長いということ。

TODO 今後の課題：回答ログ数が十分多い複数のデータセットについて同一のモデルで学習させることで、入力次元を増やすということがひとつ。専門家によるタグではなく、予測の過程で生成したタグを用いること。

第7章　まとめ

TODO

参考文献

- [1] Albert T Corbett and John R Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, Vol. 4, No. 4, pp. 253–278, 1994.
- [2] Alexander McAuley, Bonnie Stewart, George Siemens, and Dave Cormier. The mooc model for digital practice. 2010.
- [3] Laura Pappano. The year of the mooc. *The New York Times*, Vol. 2, No. 12, p. 2012, 2012.
- [4] George Siemens. Massive open online courses: Innovation in education. *Open educational resources: Innovation, research and practice*, Vol. 5, , 2013.
- [5] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *arXiv preprint arXiv:1503.03832*, 2015.
- [6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.
- [7] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- [8] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

- [9] Jun Yin, Xin Jiang, Zhengdong Lu, Lifeng Shang, Hang Li, and Xiaoming Li. Neural generative question answering. *arXiv preprint arXiv:1512.01337*, 2015.
- [10] Michael V Yudelson, Kenneth R Koedinger, and Geoffrey J Gordon. Individualized bayesian knowledge tracing models. In *Artificial Intelligence in Education*, pp. 171–180. Springer, 2013.
- [11] Zachary A Pardos and Neil T Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *User Modeling, Adaptation, and Personalization*, pp. 255–266. Springer, 2010.
- [12] Jung In Lee and Emma Brunskill. The impact on individualizing student models on necessary practice opportunities. *International Educational Data Mining Society*, 2012.
- [13] MH FALAKMASIR, Michael Yudelson, Steve Ritter, and Ken Koedinger. Spectral bayesian knowledge tracing. In *Proceedings of the 8th International Conference on Educational Data Mining., OC Santos, JG Boticario, C. Romero, M. Pechenizkiy, A. Mercerón, P. Mitros, JM Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, and M. Desmarais, Eds. Madrid, Spain*, pp. 360–364, 2015.
- [14] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, Vol. 143, No. 1, pp. 29–36, 1982.
- [15] Chris Piech, Jonathan Bassan, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. Deep knowledge tracing. In *Advances in Neural Information Processing Systems*, pp. 505–513, 2015.
- [16] Joseph E Beck and Kai-min Chang. Identifiability: A fundamental problem of student modeling. In *User Modeling 2007*, pp. 137–146. Springer, 2007.

- [17] Jussi Kasurinen and Uolevi Nikula. Estimating programming knowledge with bayesian knowledge tracing. In *ACM SIGCSE Bulletin*, Vol. 41, pp. 313–317. ACM, 2009.
- [18] Will Tyson, Reginald Lee, Kathryn M Borman, and Mary Ann Hanson. Science, technology, engineering, and mathematics (stem) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, Vol. 12, No. 3, pp. 243–270, 2007.
- [19] Jeffrey J Kuenzi. Science, technology, engineering, and mathematics (stem) education: Background, federal policy, and legislative action. 2008.
- [20] Mingyu Feng, Neil Heffernan, and Kenneth Koedinger. Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, Vol. 19, No. 3, pp. 243–266, 2009.
- [21] Zachary A Pardos, Ryan SJD Baker, Maria OCZ San Pedro, Sujith M Gowda, and Supreeth M Gowda. Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 117–124. ACM, 2013.
- [22] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R. Koedinger. Algebra i 2005-2006. development data set from kdd cup 2010 educational data mining challenge. <http://pslcdatahop.web.cmu.edu/KDDCup/downloads.jsp>, 2010.
- [23] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R. Koedinger. Algebra i 2006-2007. development data set from kdd cup 2010 educational data mining challenge. <http://pslcdatahop.web.cmu.edu/KDDCup/downloads.jsp>, 2010.
- [24] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R. Koedinger. Algebra i 2008-2009. development data set from kdd cup 2010 educational data mining challenge. <http://pslcdatahop.web.cmu.edu/KDDCup/downloads.jsp>, 2010.

- [25] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R. Koedinger. Bridge to algebra 2006-2007. development data set from kdd cup 2010 educational data mining challenge. <http://pslcdatashop.web.cmu.edu/KDDCup/downloads.jsp>, 2010.
- [26] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R. Koedinger. Bridge to algebra 2008-2009. development data set from kdd cup 2010 educational data mining challenge. <http://pslcdatashop.web.cmu.edu/KDDCup/downloads.jsp>, 2010.
- [27] Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, Vol. 1, No. 2, pp. 270–280, 1989.
- [28] Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pp. 3104–3112, 2014.
- [29] Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, and Haifeng Wang. Multi-task learning for multiple language translation. ACL, 2015.
- [30] Alex Graves and Jürgen Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in Neural Information Processing Systems*, pp. 545–552, 2009.
- [31] Jérôme Louradour and Christopher Kermorvant. Curriculum learning for handwritten text line recognition. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*, pp. 56–60. IEEE, 2014.
- [32] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, Vol. 29, No. 6, pp. 82–97, 2012.

- [33] Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, and Yoshua Bengio. End-to-end attention-based large vocabulary speech recognition. *arXiv preprint arXiv:1508.04395*, 2015.
- [34] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*, 2015.
- [35] Kelvin Xu, Jimmy Ba, Ryan Kiros, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint arXiv:1502.03044*, 2015.
- [36] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *arXiv preprint arXiv:1411.4555*, 2014.
- [37] Edward Choi, Mohammad Taha Bahadori, and Jimeng Sun. Doctor ai: Predicting clinical events via recurrent neural networks. *arXiv preprint arXiv:1511.05942*, 2015.
- [38] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.
- [39] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807–814, 2010.
- [40] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.
- [41] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *Neural Networks, IEEE Transactions on*, Vol. 5, No. 2, pp. 157–166, 1994.

- [42] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pp. 400–407, 1951.
- [43] Harold J Kushner and George Yin. *Stochastic approximation and recursive algorithms and applications*, Vol. 35. Springer Science & Business Media, 2003.
- [44] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [45] Matthew D Zeiler. Adadelta: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.
- [46] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks.
- [47] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 6, No. 02, pp. 107–116, 1998.
- [48] Tomáš Mikolov. Statistical language models based on neural networks. *Presentation at Google, Mountain View, 2nd April*, 2012.
- [49] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [50] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [51] Quoc V Le, Navdeep Jaitly, and Geoffrey E Hinton. A simple way to initialize recurrent networks of rectified linear units. *arXiv preprint arXiv:1504.00941*, 2015.
- [52] David Krueger and Roland Memisevic. Regularizing rnns by stabilizing activations. *arXiv preprint arXiv:1511.08400*, 2015.

- [53] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [54] Wojciech Zaremba. An empirical exploration of recurrent network architectures. 2015.
- [55] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Gated feedback recurrent neural networks. *arXiv preprint arXiv:1502.02367*, 2015.
- [56] Andrej Karpathy, Justin Johnson, and Fei-Fei Li. Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*, 2015.
- [57] Shamim Biswas, Ekamber Chadda, and Faiyaz Ahmad. Sentiment analysis with gated recurrent units.
- [58] Mohammad Pezeshki. Sequence modeling using gated recurrent neural networks. *arXiv preprint arXiv:1501.00299*, 2015.
- [59] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, 2014.
- [60] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, Vol. 78, No. 10, pp. 1550–1560, 1990.
- [61] 文部科学省初等中等教育局教育課程課. 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,
2011.
- [62] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253–260. ACM, 2002.

- [63] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, Vol. 13, No. 1, pp. 281–305, 2012.
- [64] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th international conference on Machine learning*, pp. 473–480. ACM, 2007.

謝辞

本稿の執筆にあたり，指導教員の松尾豊特任准教授から丁寧かつ熱心なご指導を賜りました。ここに感謝の意を表します。論文執筆の方向性についての議論に際して、ご指導くださった中山先生に感謝いたします。

また、データのご提供並びにサービス理解にご協力くださったリクルートマーケティングパートナーズ(株)の萩原静巖様には深く感謝いたします。そして、検証実験に協力して下さった多くの方々には、この場を借りて感謝致します。論文執筆生活のなかで、様々な配慮をして支えて下さった家族に感謝の意を表します。最後に、研究生活において共に競い合い切磋琢磨して質の高い研究成果を目指して下さった松尾研究室の皆様に改めて感謝致します。

東京大学大学院工学系研究科
技術経営戦略学専攻
松尾豊研究室 修士二年
那須野薫

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, 萩原静巖, & 井上綾香. (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, & 萩原静巖. (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 (マイクロブログ, 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.
- 柿剛史, 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.