

卒業論文

Deep Knowledge Tracing を用いた クラウドソーシングの動的タスク割当て

指導教官 村上 陽平 教授

立命館大学 情報理工学部
先端社会デザインコース 4 回生
2600200455-0

山本 涼太郎

2023 年度（秋学期）卒業研究 3（CH）
令和 6 年 1 月 31 日

Deep Knowledge Tracing を用いたクラウドソーシングの 動的タスク割当て

山本 涼太郎

内容梗概

低資源言語の言語資源を作成するために、クラウドソーシングが用いられ、不特定多数の作業者に言語資源の作成タスクが割り当てられている。このようなクラウドソーシングでは、作業者の能力にばらつきがあるため、品質管理が重要な課題となる。事前に作業者の能力は不明なため、作業者の作成した言語資源をそのまま利用するのではなく、複数の言語資源を一つにまとめた評価タスクセットを複数の作業者に割り当て、評価タスクセットの多数決をとる方法が取られている。これにより能力の高い作業者が少ない場合でも評価精度を向上させることができる。しかしながら、作業者と作成タスクのミスマッチにより誤った言語資源が作成されると、作成および評価作業のやり直しが多発する。

そこで、本研究では、作業者に作成可能なタスクを割り当てるために、作業者ごとのタスク難易度を考慮したタスク割り当ての手法を提案し、作業時間を最小化することを目指す。具体的には、Deep Knowledge Tracing モデルを各作業者の作業履歴と作業結果のデータでトレーニングすることで、作業者のタスクの成功可否を推定する。作業者が新たなタスクに取り組むたびに、タスクの成功確率を更新し、次に取り組むタスクの割り当てに考慮する。この手法をインドネシア語とミナンカバウ語の対訳辞書の作成に用いることで対訳辞書の正確性と作業時間の短縮を実現する。

本手法の実現にあたり、取り組むべき課題は以下の2点である。

作業者のタスクの成否予測

作業者ごとのタスク難易度を考慮して対訳作成タスクを割り当てるために、作業履歴を元に、各作業者の対訳作成タスクの正解確率を推定する必要がある。ただし、作業者は過去に同じ対訳を作成しているとは限らないため、対訳作成タスクを原語ごとに分類する必要がある。

動的なタスク割当て

作業者の正答率を向上させるために、クラウドソーシングを行いながら、その作業結果を元に次のタスクの成否予測を用いて、次のタスク割当てに動的に反映していく必要がある。

一つ目の課題に対しては、深層学習モデルである Deep Knowledge

Tracing(DKT)モデルを使用することで、回答者、原語の種類、対訳の正誤からなる作業履歴の時系列データを処理し、将来の対訳作成タスクの成否予測を行った。原語の分類については、Word2Vecを用いインドネシア語の単語の分散表現を獲得した後に、最適なクラス数を見つけ、k-means法でクラスタリングを行った。これにより、単語が共に使われることの多い単語、対の意味の形容詞、類似した意味の接続詞や前置詞などでクラスタリングすることができた。このクラスタリング結果に基づき、原語に問題の種類を識別子を割り振り DKT の入力データとして使用した。

二つ目の課題に対しては、タスクの作業結果をその都度 DKT モデルに入力し、次のタスクの正解確率を予測し、新しい正解確率に基づいてハンガリアン法により逐次的にタスク割当てを行う。最初に全タスクの成否を予測し、静的なタスク割当て手法と、毎回のタスク結果を反映して次タスクを動的に割り当てる手法で比較評価を行い、提案手法の有効性を検証した。本研究の貢献は以下のとおりである。

作業者のタスクの成否予測

インドネシア語とミナンカバウ語の対訳辞書作成タスクの作業履歴を用いて、DKT で成否予測をすることができた。モデルの Accuracy は約 0.6~0.7 の間であった。

動的なタスク割当て

DKT から得られた作業者の正解確率を使い、ハンガリアン法で逐次的にタスク割当てをした場合、140 タスクを作業者 20 人に割当てることによって、各作業者の正答率は平均で 71.0710125%となった。また、ランダムに割当てるとより約 20%正答率が高くなった。

Dynamic Crowdsourcing Task Assignment Using Deep Knowledge Tracing

Ryotaro Yamamoto

Abstract

Crowdsourcing is used to create language resources for low-resource languages, and tasks for creating language resources are assigned to an unspecified number of workers. Quality control is an important issue in such crowdsourcing. Since the abilities of workers are not known in advance, an evaluation task set that combines multiple language resources into one is assigned to multiple workers, and a majority vote is taken on the evaluation task set. This method improves the accuracy of evaluation. However, when an incorrect language resource is created due to a mismatch between workers and translation creation tasks, the creation and evaluation tasks are often redone.

Therefore, in order to assign tasks that can be created to each workers, this study proposes a task assignment method that takes into account the task difficulty level of each worker and aims to minimize the work time. Specifically, the Deep Knowledge Tracing model is trained with each worker's work history and work result data to estimate the success or failure of the worker's task. Each time a worker works on a new task, the task success probability is updated and taken into account in assigning the next task to work on. This method is used to create a bilingual dictionary of Indonesian and Minangkabau to improve the accuracy of the bilingual dictionary and reduce the work time.

The following two issues needed to be addressed in order to realize this method.

Predicting the success or failure of a worker's task

In order to assign a bilingual task in consideration of the difficulty level of each worker's task, it is necessary to estimate the probability of correct answers for each worker's bilingual task based on his/her work history. However, because workers have not necessarily created the same translation in the past, it is necessary to categorize the translation creation tasks by source language.

Dynamic Task Assignment

In order to improve the correct response rate of the workers, it is necessary

to dynamically reflect the results of their work in the next task assignment using the success/failure prediction of the next task based on the results of their work while crowdsourcing.

For the first task, we used a deep learning model, the Deep Knowledge Tracing (DKT) model, to process time-series data on the respondent, source language type, and bilingualism, and to predict the success or failure of future bilingualism tasks. For source word classification, we used Word2Vec to obtain a distributed representation of Indonesian words, found the optimal number of classes, and clustered them using the k-means method. This allowed us to cluster words that are often used together, adjectives with paired meanings, conjunctions and prepositions with similar meanings, etc. Based on the results of this clustering, we found that the clustering of words that are often used in the original language is more likely to be clustered. Based on the clustering results, the source words were assigned an identifier of the problem type and used as input data for DKT.

For the second task, the results of the task work were input to the DKT model each time, the probability of correct answers for the next task was predicted, and task assignment was performed sequentially by the Hungarian method based on the new probability of correct answers. The effectiveness of the proposed method was verified through comparison and evaluation. It contrasted two approaches. One predicts the success or failure of all tasks and assigns them all at once. The other dynamically assigns the next task based on the outcomes of each preceding task. The contributions of this study are as follows:

Predicting the success or failure of a worker's task

Using the work history of the task of creating a bilingual dictionary of Indonesian and Minangkabau, the model's Accuracy ranged from about 0.6 to 0.7.

Dynamic Task Assignment

Using the workers' probabilities of correct answers obtained from the DKT, the sequential task assignment using the Hungarian method resulted in an average correct response rate of 71.0710125% for each worker by assigning 140 tasks to 20 workers. The correct response rate was about 20% higher than that of random assignment.

Deep Knowledge Tracing を用いた クラウドソーシングの動的タスク割当て

目次		
第 1 章	はじめに	1
第 2 章	クラウドソーシング	3
第 3 章	Deep Knowledge Tracing	5
第 4 章	タスクの難易度推定	7
4.1	タスク分類	7
4.2	作業者のタスクの成否予測	9
第 5 章	タスクの難易度を考慮したタスク割当て	11
5.1	タスク割当ての定式化	11
5.2	ハンガリアン法	11
5.3	動的なタスク割当て	12
第 6 章	評価	14
6.1	評価データ	14
6.2	評価方法	16
6.2.1	動的なタスク割当て手法	16
6.2.2	静的なタスク割当て手法	17
6.3	結果	18
6.4	考察	19
第 7 章	おわりに	21
	謝辞	23
	参考文献	24

第1章 はじめに

インドネシアとその周辺地域では、700以上の言語が話されており、今後50年から100年の間で話される言語の数が700からおよそ50に減少すると推定されている。現在でも、147の地方言語が消滅の危機に瀕しており、これらの保護支援や地方言語間のコミュニケーション支援のために、対訳辞書が必要である。このような、対訳辞書など言語資源の作成には、クラウドソーシングが用いられている[1]。

クラウドソーシングとは、インターネットを通じて不特定多数の人に仕事を依頼することで、情報や労力を集める手法である。特に、コンピュータでは困難だが、人間にとって比較的簡単なタスクに用いられることが多い。

このようなクラウドソーシングでは、不特定多数の作業者にタスク割り振ることから、作業者の能力にばらつきがある。特に、言語資源作成タスクは、専門性を必要とし、作業者全員が同じレベルでタスクを処理できるとは限ら無いため成果物の品質を保証することは困難である。よって、クラウドソーシングにおける品質管理は重要な課題となる。作成タスクだけでなく、作成した言語資源の評価タスクも組み合わせることで品質を保証している。しかし、タスクを実行するまで作業者の能力は不明であるため、能力が高く、信頼のできる作業者にのみタスクを割り当てることはできない。そのため、作業者の作成した言語資源をそのまま利用するのではなく、複数の言語資源を一つにまとめた評価タスクセットを複数の作業者に割り当て、評価タスクセットの多数決をとる方法が取られている。これにより能力の高い作業者が少ない場合でも評価制度を向上させることができる。しかしながら、作業者と作成タスクのミスマッチにより誤った言語資源が作成されると、作成および評価作業のやり直しが多発する。

そこで、本研究では、作業者に作成可能なタスクを割り当てるために、作業者ごとのタスク難易度を考慮したタスク割り当てを行うことで、成果物の正確性を向上、作業時間を最小化するというアプローチをとった。Deep Knowledge Tracingを用いて、作業者の作業履歴から次タスクの成否予測を行う。作業者が新しいタスクに取り組むごとに成否予測を行い、この成否予測を用いて、逐次的にタスク割り当てをする。この手法の実現にあたり、以下の課題に取り組む必要がある。

作業者のタスクの成否予測

作業者ごとのタスク難易度を考慮して対訳作成タスクを割り当てるために、作業履歴を元に、各作業者の対訳作成タスクの正解確率を推定する必要がある。ただし、作業者は過去に同じ対訳を作成しているとは限らないため、対訳作成タスクを原語ごとに分類する必要がある。

動的なタスク割当て

作業者の正答率を向上させるために、クラウドソーシングを行いながら、その作業結果を元に次のタスクの成否予測を用いて、次のタスク割当てに動的に反映していく必要がある。

本稿の残りは以下のような構成となっている。第2章では、クラウドソーシングに関する基本概念について紹介し、第3章では、Deep Knowledge Tracingの原理について紹介する。第4章では、タスクの難易度を推定する方法に焦点を当て、具体的な手法と基本的な理論について説明する。第5章ではタスク難易度を考慮した動的なタスク割当ての手法について具体的に説明する。さらに、第6章では、提案手法の評価方法と、提案手法を使った際の正答率を提示する。そして、第7章で本稿をまとめる。

第2章 クラウドソーシング

クラウドソーシングとは、インターネットを介して不特定多数の人に仕事を依頼すること、もしくはその仕組みのことを指す。クラウドソーシングは、画像のラベリングや文章の翻訳など計算機では作成が比較的困難であり、人間の持つ能力を用いればそれほど難しくはないタスクにおいて一般的に使用されている。Amazon Mechanical Turk(AMT)のような大規模なクラウドソーシングプラットフォームの存在により、インターネットを介して大量の作業者を容易に確保することができる。現在、多言語データの収集にクラウドソーシングが広く用いられており、これにより多言語話者が用例対訳を作成するプロジェクトや、複数の言語におけるテキストの精度を評価する研究が進められている。

クラウドソーシング分野で注目されている研究として、品質管理の方法がある。人間が作業を行うため、能力が低い作業人や意図的に品質を落とす作業者(スパムワーカー)が存在するリスクがあり、常に正確な結果が得られるとは限らない。この問題に対処するため、品質管理手法の研究では主に以下の2つのアプローチがある[2][3][4]。

作業結果を集約して全体の品質を向上させる方法

複数の作業者に同じタスクを割当て、冗長性を利用し最終的には多数決で結果を決める方法が主に取られている。この方法では、作業者の能力が高い場合には正しい答えを導けるが、作業者の能力が低い場合では正しい答えを導き出すことは困難である。

個々の作業結果の品質を向上させる方法

報酬の設定、タスクのデザイン、作業者の選定などを行うことで、作業者によるタスクの実行結果そのものの品質を向上させる方法。特に、高い能力を持つ作業者を事前に抽出し、タスクを割当てる方法は、低い能力の作業人やスパムワーカーの排除することができるため、作業結果の品質向上が期待される。

また本実験で使用した評価データは、クラウドソーシングを用い対訳評価をしたものである。このクラウドソーシングのタスクの割当てから、作業結果の獲得までは、図1のような流れで行われる。依頼者が作成者に対訳作成タスク、評価者に対訳評価タスクを割当て、対訳作成を作業者に依頼する。作成者が対訳を作成した後、評価者がその対訳を評価し「正しい」と評価するとそ

の対訳を依頼者は獲得し、 評価者が「間違い」と評価した際はそのタスクをもう一度依頼者が作成者、 評価者に割当て対訳を獲得していく。

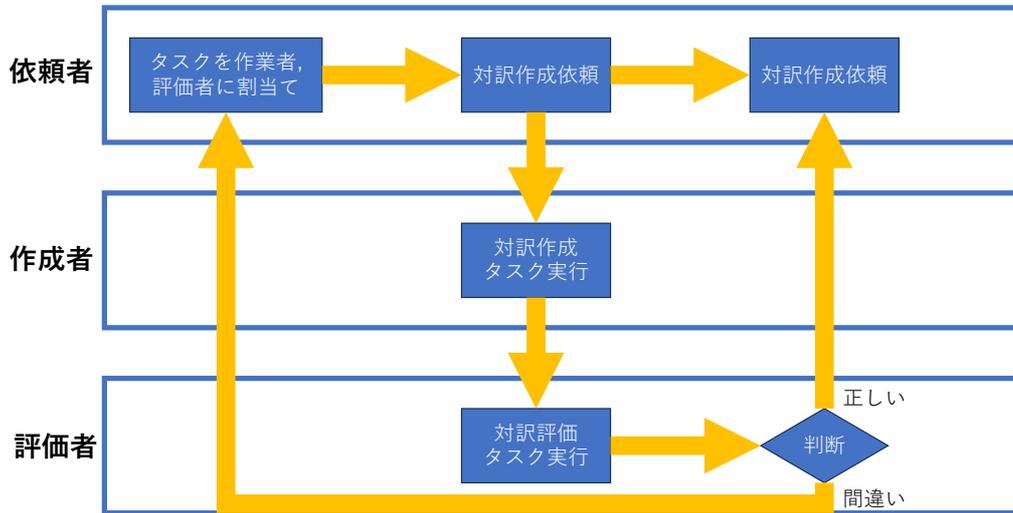


図 1: 対訳作成評価タスクのワークフロー

第3章 Deep Knowledge Tracing

Knowledge Tracing

Knowledge Tracing(KT)は、過去の学習者の回答履歴を基に、各学習者がどの程度知識を獲得しているかを予測することである。主に、教育分野で利用され学習者が特定のスキルに対しての知識獲得状況を理解するのに役にたつ。KTの手法前半には、スキル分類が必要とされる。スキル分類とは、学習内容を細分化したカテゴリを各問題のラベルとしたものである。問題ごとに割り振られており、その問題を解くためのスキルや知識を表す。生徒の回答履歴をスキル分類に紐付けてKTモデルに入力することで、知識獲得状況を予測することができる[5].

Deep Knowledge Tracing

Deep Knowledge Tracing(DKT)[6]とはKTに対して、Recurrent Neural Networks(RNN)を用いることにより、ドメイン知識を手動で入力せずに知識獲得予測を行えるようにしたものである。特に、RNNの一種であるLong Short Term Memory(LSTM)を用いることで長期的な依存関係を学習することができ、過去の作業履歴から、将来のタスクへの作業者のパフォーマンスを予測する。具体的には、作業者、タスクのスキル分類、タスクの正誤から成る作業履歴の時系列データを用い、作業者の作業毎に、すべてのスキル分類にスコアリングを0から1.0で行い、スコア分類ごとに成否予測を行うことができる。

DKTモデルの基本的な式は以下の通りだ

$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b_n) \quad (1)$$

$$y_t = g(W_{yh}h_t + b_y) \quad (2)$$

x_t は入力であり、時刻 t における学習者の問題への回答を表すベクトル。式(1)は隠れ層の更新を表し、 h_t は時刻 t の隠れ層の状態、 W_{hx} 、 W_{hh} は重み行

列, b_n はバイアス項, f は活性化関数(\tanh や ReLU など)である. h_t は, 時刻 t までの学習者のすべての活動の要約を保持することができる.

式(2)より, y_t は時刻 t における出力であり, すなわち次の問題への正解確率のことである. W_{yh} は隠れ層から出力への重み, b_y はバイアス項, g は活性化関数である. 活性化関数 g により, 出力値を0から1.0の間に制限することによって出力を確率として表すことができる[6].

第4章 タスクの難易度推定

4.1 タスク分類

DKT を用いて、 作業者の将来タスクへの成否予測をするためには、 タスクのスキル分類(知識タグ)が必要である。 DKT において知識タグは、 作業者や学習者が取り組むタスクに関連付けられた分類タグのことを指し、 目的に合わせ、 様々な基準で決めることができる。 例えば、 英語の文法教育では、 「現在進行形」や「過去完了時制」など特定の文法項目を知識タグとして用いられることがある。 同じ知識タグを付与されたタスクは、 タスクを正当するためのスキルが同じであるとして、 成否予測を行う。

本研究で今回用いた対訳作成タスクでは、 インドネシア語の原語をミナンカバウ語の訳語を作成した対訳データを評価する作業が行われている。 表 1 は、 そのタスクをクラウドソーシングを用いて、 一人当たり約 5617 件の対訳ペアを作業者に割当て、 対訳として正しいか、 **Correct**, **Wrong** で答えた結果の一部である。 このデータではタスクがスキル分類されておらず、 知識タグが存在していない。 そのため、 DKT で成否予測をするためにスキル分類を行い、 知識タグ[7]を作成する必要がある。

そこで、 **Word2Vec** を用いて単語を用いて単語の分散表現を獲得した後、 k-means 法でクラスタリングを行い、 そのクラスタをタスクのスキル分類として使用した。 クラスタ数は、 どのクラスもクラスタサイズが 5 以上となる最大クラスタ数を使用する。

これにより、 クラスの要素が少なすぎるとモデルの学習のためのデータが少なくなりそのクラスの知識タグについての予測がうまくいかないという問題を解消している。

本研究では、 **Word2Vec** をインドネシア語の文章 1, 000, 000 文で学習させ分散表現を獲得した。 分散表現を獲得する必要がある単語数は 1002 単語、 そしてクラスタリングの際のクラス数は、 11 個がどのクラスもクラスタサイズが 5 以上となる最大クラス数であった。 表 2 より、 各クラスタサイズは 11 から 263 の間となり、 クラスタサイズの大きさにばらつきのある結果となった。 表 3 は、 クラスタリング結果をアルファベット順で先頭から最大 30 個抜粋したものである。 各クラスに主に含まれている単語の規則性は、 次のようになった。

class0 には時間や順序を表す単語を含み, class1, class7 は動詞, class2 は公正さや可能性を問う単語, class3 は物事の理由や結果に関する単語, class4 は方向や位置に関する単語, class5 は必要性や義務に関する単語, class6 は存在や状態に関する単語, class8 は人の関係性を表す単語, class9 には物事の特徴や状態を説明する単語, class10 は数や量を示す単語が多く含まれている結果となった. あくまで主に含まれている単語であるため, 中にはうまくクラスタリングされなかった単語もあったが, 本研究ではこのクラスを使い手法を進めていく.

表 1: 対訳評価タスク結果の一部

sourceword	targetword	evaluation	worker	level	evaluated
ada	ado	CORRECT	Fitri Amanah	5	2020-12-10 08:37:38
ada	lai	CORRECT	Fitri Amanah	5	2020-12-08 10:53:25
ada	lai, ado	CORRECT	Fitri Amanah	5	2020-12-08 10:56:09
adalah	adalah	WRONG	Fitri Amanah	5	2020-12-08 10:56:12
adalah	adalah	CORRECT	Fitri Amanah	5	2020-12-08 10:56:17
adalah	artinyo	WRONG	Fitri Amanah	5	2020-12-08 10:56:20
adalah	iko	WRONG	Fitri Amanah	5	2020-12-08 10:56:23
adalah	inyotu	WRONG	Fitri Amanah	5	2020-12-07 23:40:29
adalah	iyolah	CORRECT	Fitri Amanah	5	2020-12-07 23:40:30
adalah	yaitu	WRONG	Fitri Amanah	5	2020-12-07 23:40:30
adil	adia	CORRECT	Fitri Amanah	5	2020-12-07 23:40:32

表 2: クラスタリング結果の要素数

クラス	class 0	class 1	class 2	class 3	class 4	
要素数	57	105	124	53	102	
クラス	class 5	class 6	class 7	class 8	class 9	class 10
要素数	10	108	60	263	104	16

表 3: クラスタリング結果(各クラスの代表する単語を抜粋)

Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
akhirnya	ambil	adil	alasan	alih	butuh	ada	bangun	anaknya	adalah	beberapa
awal	anggap	akan	bantuan	bakar	diberikan	berbicara	bekerja	aneh	alam	delapan
belum	ayo	apakah	hal	balik	membayar	berbuat	berada	anggur	aman	dua
berapa	baca	apapun	hasil	bawah	membeli	berharap	berakhir	angkat	antara	empat
berikutnya	bantu	artinya	janji	belakang	membunuhmu	bertindak	berangkat	astaga	asing	enam
belum	bicara	bagaimana	karena	bawah	mesti	hilang	beri	ayah	baik	lima
kemudian	buat	bisa	kecuali	dalam	perlu	jadi	berikan	bapak	benar	satu
lalu	coba	boleh	meskipun	dekat	harus	seperti	berlari	teman	besar	sepuluh
saat	dengar	cukup	namun	jauh	wajib	tetap	bermimpi	sahabat	buruk	seribu
sebelum	datang	harus	sehingga	kanan	yakin	tinggal	bersama	musuh	cepat	tiga

4.2 作業者のタスクの成否予測

20 人分, 一人あたり約 5617 件のインドネシア語からミナンカバウ語のクラウドソーシングを用いた対訳評価タスクの時系列データに, スキル分類によって知識タグを付与し, 作業者ごとに DKT で成否予測を行った. 成否予測の精度評価は, モデルが正しくタスクの成否を予測した割合である Accuracy で測定する.

表 4: ある作業者 ID が 40 問目を解いた後の成否予測確率

問題 ID	成否予測確率
0	0. 580700815
1	0. 571043849
2	0. 586631298
3	0. 567974389
4	0. 588099718
5	0. 575584114
6	0. 573873878
7	0. 577137113
8	0. 588856816
9	0. 588933706
10	0. 544296682

表 5: 作業員 20 人別の DKT モデルの Accuracy

作業員 ID	Accuracy	作業員 ID	Accuracy	作業員 ID	Accuracy	作業員 ID	Accuracy
1(4)	0.59848	6(2)	0.66748	11(2)	0.61295	16(1)	0.65947
2(3)	0.74938	7(1)	0.62764	12(5)	0.88448	17(4)	0.70776
3(3)	0.63432	8(5)	0.70643	13(2)	0.68439	18(4)	0.67304
4(2)	0.63743	9(1)	0.53638	14(5)	0.72957	19(1)	0.67415
5(5)	0.83218	10(3)	0.71511	15(3)	0.68283	20(4)	0.66993

DKT モデルからは、表 4 のように出力され、問題を解いた後の全ての知識タグについての成否予測確率がわかる。また、第 5 章で説明するタスク割当てにおいて、この成否予測確率をハンガリアン法のコストとして用いる。

表 5 は、作業履歴を学習させた各作業員のモデルの Accuracy である。作業員 ID の右の括弧内の 1~5 の数値は言語能力のレベルを表す数値を表している。その数値を用いて、作業員 ID が 5 と 12 の回答を元に正解データを作成しているため、これらの作業員の Accuracy は 0.8 以上と非常に高い結果となった。他の作業員に関しては、おおよそ 0.6~0.7 の間の値となり、未知の作業員に対してもある程度の成否予測が可能であることを確認した。

第5章 タスクの難易度を考慮したタスク割当て

5.1 タスク割当ての定式化

利用可能なタスクを割りてることにより、効率と品質を最大化することを目的としている。入力、タスクのセットと作業者のスキルに関する情報である。作業者を*i*、タスクを*j*とした時、タスク割当ての効用が最大となる式は次のように表すことができる(式3)。

$$\max \sum_{i, j} U_{ij} \times X_{ij} \quad (3)$$

U_{ij} は作業者*i*がタスク*j*を行なった場合の効用、 X_{ij} は作業者*i*がタスク*j*に割り当てられているかのバイナリ変数である。

$$\sum_i A_{i, j} = 1 \quad \forall j \quad (4)$$

$$\sum_i A_{i, j} \leq \text{MaxTasks}_i \quad \forall i \quad (5)$$

式4より、この式は、各タスク*j*が1人の作業者にのみ割り当てられることを表し、複数の作業者に同じタスクが割り振られないことを表す。式5より、この式は、各作業者*i*が割り当てられるタスクの数に上限があることを表す。

5.2 ハンガリアン法

ハンガリアン法は、割当て問題を解決するための手法の一つである。割当て問題とは、リソースをタスクに割当ててを目的とした最適化問題の一種である。コスト行列を作成することで以下の手順により、全てのタスクをリソースに割り振る際に、最も低コストになる割当てを決定することができる(図2)。

- (1) 各作業者(リソース)と各タスク間のコストを表す行列を作成する。
- (2) コスト行列の各行から、その行の最小値を引く。
- (3) コスト行列の各列から、その列の最小値を引く。(2)、(3)の手順により各行、
列に少なくとも一つは0が現れる。
- (4) 0の要素を含む行と列をカバーするのに必要な最小数の水平線または垂直線を引く。この時、引いた線の本数がコスト行列のサイズかそれより大

きければ、終了。そうでなければ、次の(5)の処理に進む。

(5) 線が引かれていない要素の最小値を、線が引かれていない要素から引く。
そして、線が重なっている要素に、線が引かれていない要素の最小値を足す。

(6) (4), (5)の手順を、0をカバーする線がコスト行列と同じかそれ以上の本数になるまで続ける。

各行、各列から重複なく、最終的にコストが0であるセルの組み合わせ全てが、ハンガリアン法における最適な割当て解である[8].

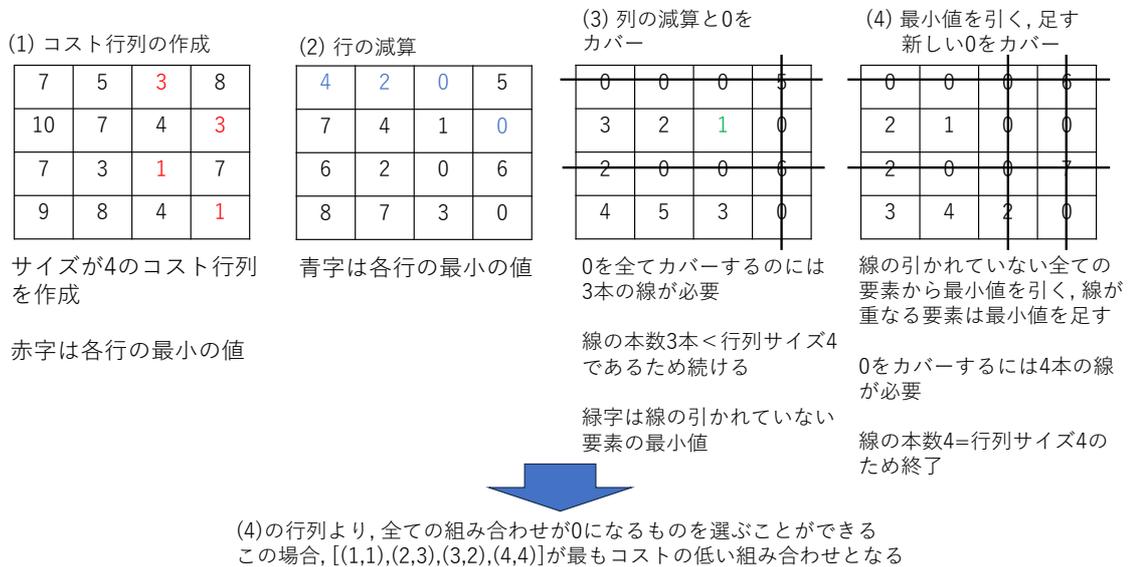


図 2: ハンガリアン法の処理の手順

5.3 動的なタスク割当て

本研究における、動的なタスク割当てとは、クラウドソーシングなど作業環境において、作業者の能力を、作業が進むにつれ理解することができるため、その能力を考慮しながらタスクを作業者に割当てるプロセスである。作業者の過去のパフォーマンスから現在の能力のレベルを予測し、作業者にとってのタスクの難易度を動的に評価することで、最適なタスク割当てを行う。具体的には、作業毎に DKT モデルを用いて、作業履歴を分析することで作業者のタスクに

対する成否予測を行う。次に、これらの成否予測の確率を基にハンガリアン法を適用させるためのコスト行列を作成し割当てを実行する。この際、ハンガリアン法は、全体のコストを最小化するためのアルゴリズムであるため、予測確率は符号反転させることで高い確率(本来、高コストであるもの)が選ばれるように変更する必要がある。ハンガリアン法によって得られた、作業者とタスクの最適な割当てを次タスクの割当てとして使う。この作業を繰り返すことで、作業者の成否予測をより正確に行うことができ、割当ての精度が上がると考えている(図 3)。

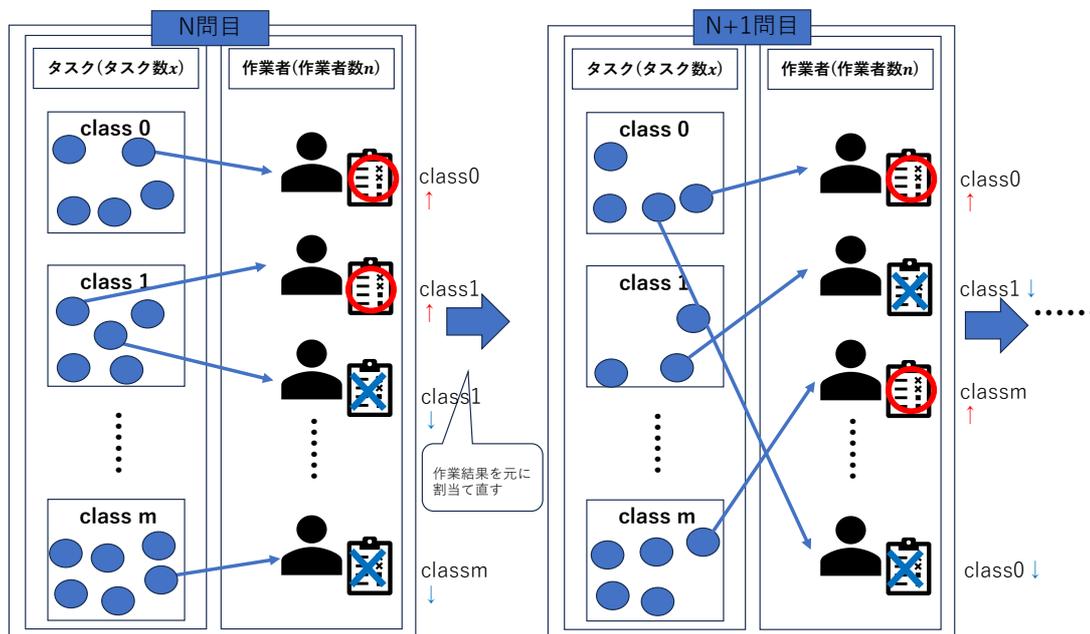


図 3: 動的なタスク割当て

第6章 評価

6.1 評価データ

本研究における手法を評価するために、使用したデータセットについて説明する。今回、用いたデータセットはクラウドソーシングで作成されたインドネシア語とミナンカバウ語間の対訳データの評価を行なった作業履歴である。一人当たり約 5617 件のインドネシア語とミナンカバウ語の対訳ペアを評価者に割当て、対訳として正しいか、**Correct**, **Wrong** で答えるタスクを行っている。本データには、タスクに順番に 0 から数字を割り振った「**Unnamed**」、対訳ペアのインドネシア語である「**sourceword**」、対訳ペアのミナンカバウ語である「**targetword**」、作業者が対訳ペアを評価した結果である「**evaluation**」、作業者の名前である「**worker**」、対訳評価タスクにおける作業者への信頼値である「**level**」、タスクを行った時間である「**evaluated**」で構成されている。(図 1)そして、**Level** の値が最も高い作業者 2 人のデータを基に作成された、本タスクの対訳ペアとして正しいものと推測されるものだけを残したデータセットの二つを使用する。

これらのデータを **DKT** で処理をするには、「作業者 ID」「問題 ID」「作業結果の成否」の時系列データが必要になる。元のデータセットには「作業者 ID」「問題 ID」「作業結果の成否」のデータがないため作成する必要がある。まず、作業者 ID の作成は、作業者が 20 人存在するため 1~20 の数字をそれぞれの作業者に割当て作業者 ID を作成した。次に、問題 ID は第 4 章の 4. 1 で説明したように、データセットの **sourceword** のインドネシア語の単語 1002 単語をベクトル化した後に、最低でもクラスタサイズが 5 以上になるように **k-means** 法を用いてクラスタリングを行った。その結果、11 個のクラスに分類することができたため、それぞれのクラスに対応する単語に 0~10 の数値を問題 ID として割り振った。

最後に、作業結果の成否は正解、不正解の 2 値であり、**Level** 値が高い作業者 2 人を基に作成された、正解と推測される対訳ペアのデータ(正解データと呼ぶ)を使用し、作業履歴のデータにおいて **Correct** と評価した対訳ペアが正解データ存在する、もしくは、**Wrong** と評価した対訳ペアが正解データに存在しない場合、その作業結果は正解しているとして、作業結果の成否のデータは正解

の 1 とする。一方で、作業履歴のデータにおいて **Wrong** と評価した対訳ペアが正解データに存在する、もしくは、**Correct** と評価した対訳データが正解データに存在しない場合、その作業結果は不正解であるとして、作業結果の成否のデータは不正解の 0 とした(図 4)。

作業履歴の **worker** を作業者 ID、**sourceword** を問題 ID に置き換え、新しく作業結果の成否の値を追加したデータを DKT の成否予測に用いた(図 5)。

本研究では、合計約 5617 個のタスクの中から、特定の方法を用いて 140 個のタスクを選択し、20 人の作業者に割当てることによって評価に用いている。具体的には、140 個のタスクを選出する際に、タスクを問題 ID に基づいて、選出したいタスク数(140)を問題 ID の個数(11)で割り、約 12.7 という数値が得られ、この計算に基づき、各問題 ID から 12 個のタスクをランダムに選出する。しかし、全ての問題 ID に 12 個のタスクが含まれているわけではないため、問題 ID 内のタスク数が 12 以下の場合(表 2 の **class5**, **class10**)その不足分を他問題 ID のタスクからランダムに選出し、140 のタスクに問題 ID の分散と多様性を持たせるようにしている。また二つ目の評価方法で使う、成否予測を行う DKT モデルは、選出された 140 タスク以外の履歴(一人当たり 1500 件)のデータで訓練をしておく。

正解データ 評価結果	正解データに存在する場合	正解データに存在しない場合
Correct	1	0
Wrong	0	1

正解データに、りんご→appleがあり、りんご→grapeがない場合

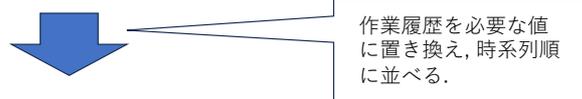
	sourceword	targetword	evaluation
①	りんご	apple	1
②	りんご	grape	0

①, ②のタスクは共に、正解となり「作業結果の成否」の値は1

図 4: 作業結果の成否の作成方法と例

作業履歴例

Unnamed: 0	sourceword	targetword	evaluation	worker	level	evaluated
0	ada	ado	CORRECT	Fitri Amanah	5	2020-12-10 08:37:38
5832	bangga	gadang hati	CORRECT	Irwanto usman	5	2020-12-19 19:44:38



作業者ID	問題ID	作業結果の成否
5	6	0
8	7	0

図 5: 元データを DKT で使うための変更

6.2 評価方法

本研究では、タスクの難易度を考慮した動的なタスク割当て手法の有効性を評価するために、Deep Knowledge Tracing(DKT)とハンガリアン法を用いて、タスク割当てを行った。この手法では、各タスクの完了後に作業者のパフォーマンスを再評価し、次タスクの割当てを毎回更新する。この手法を用いて、インドネシア語とミナンカバウ語の対訳評価タスクのデータでシミュレーションを行う。具体的には、作業者 20 人に合計 140 個のタスクを、本研究提案手法である、成否予測を用いた動的タスク割当てと、成否予測を用いずにランダムなタスク割当ての比較評価を行う。それぞれ、全て割当て終わった時点での、各作業者の正答率を算出する。この手法では、事前に DKT モデルは訓練せずに行う。その正答率を通じて、この手法の有効性を検証する(評価方法 1)。その後、1500 件のデータで訓練した後の DKT モデルを、動的なタスク割当て手法と、静的なタスク割当て手法の二つを用いて比較評価を行う(評価方法 2)。

6.2.1 動的なタスク割当て手法

動的なタスク割当て手法では、一人当たり 7 個のタスク(全体で 140 個のタスク)を、タスクを作業する毎に、DKT を用いて、作業履歴から次タスクへ

の成否予測を、問題 ID ごとに 0~1.0 の間でスコアリングとして出力する。そのスコアリングを用いて、ハンガリアン法でそのスコアリングを付合反転した数値をコストにし、最大利益(最大コスト)の割当てをすることにより、タスクの成功できる人にタスク割当てができるようにする。静的なタスク割当て手法と同じく、140個のタスクを7個の小グループに分割し、6回の成否予測を行い、20人の作業者にタスクを割当てる。一つ目のタスクは、一つ目の評価方法では、予測に使えるデータがないためランダムに割当てる。二つ目の評価方法では、DKTを、事前に選出した140タスク以外の、各作業者一人当たり約1500件の作業履歴でトレーニングを行った後の、成否確率を用いて決定する。

6.2.2 静的なタスク割当て手法

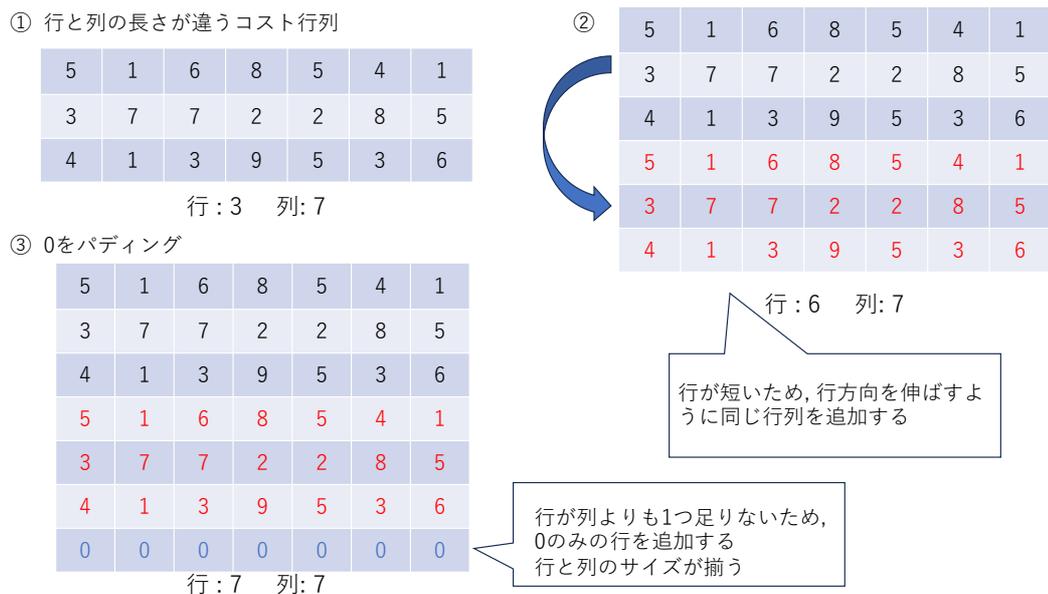


図 6: 行と列のサイズの揃え方

静的なタスク割当て手法では、最初に DKT モデルを割当てに使用する 140 タスク以外の、各作業者一人当たり約 1500 件の作業履歴を使用しその時点での成否確率のみを用いて、140 個のタスクを 20 人に割当てる。割当ては、ハンガリアン法を用いる。コストは、作業履歴から次タスクへの正当確率を、問題 ID ごとに 0~1.0 の間でスコアリングした成否予測を符号反転させることで、

最大利益の割当てを行う。通常、ハンガリアン法のコスト行列は、行と列のサイズが同じでないと使えないため、20行、140列のこのコストタスクではハンガリアン法を用いた割当てができない。そのため、このコストタグを行と列を同じサイズに揃える必要がある。今回の場合、行が列より少ないため元のデータと同じものを6個、行を増やすように追加すると行と列が140のコスト行列が完成する。今回は列の長さが行の長さで割り切れる数であるため行っていないが、割り切れない場合は、残りのセルを0でパディングする必要がある(図6)。

6.3 結果

表6より、140個のタスクを20人の作業者に、タスクの難易度を考慮した動的なタスク割当てを行った正当確率が、平均で71.071%となった。それぞれの正答率に注目すると、正答率は作業者によって大きく異なっている。最低は28.571%、最高は100%となっている。

表7より、140個のタスクを20人の作業者に、ランダムなタスク割当てを行った正当確率が、平均で50.714%となった。最低は14.286%、最高は85.714%という結果になった。

二つの結果から、向上率の最小値は-14.285%、最大値は57.143%、平均値は20.714%になった。これにより、成否予測を用いた動的タスク割当ての方が、正当確率が高いということがわかった。

表6: 評価方法1 動的タスク割当て正答率

作業者 ID	1	2	3	4	5
正当確率(%)	85.714	71.428	57.142	85.714	100
作業者 ID	6	7	8	9	10
正当確率(%)	57.142	42.857	100	28.571	57.142
作業者 ID	11	12	13	14	15
正当確率(%)	57.142	85.714	85.714	100	85.714
作業者 ID	16	17	18	19	20
正当確率(%)	71.428	42.857	57.142	71.428	85.714

表 7: 評価方法 2 ランダムなタスク割当てで正当確率

作業者 ID	1	2	3	4	5
正当確率(%)	57.142	71.428	28.571	71.428	71.428
作業者 ID	6	7	8	9	10
正当確率(%)	42.857	57.142	42.857	14.286	42.857
作業者 ID	11	12	13	14	15
正当確率(%)	42.857	57.142	85.714	57.142	57.142
作業者 ID	16	17	18	19	20
正当確率(%)	28.571	42.857	28.571	42.857	71.428

6.4 考察

DKT の成否予測とハンガリアン法を用いることで、タスクの難易度を考慮した動的なタスク割当てを行うことができた。成否予測を用いた動的タスク割当てと、ランダムなタスク割当ての比較評価から、成否予測や知識タグが正しくできていると考えている。しかし、時間の制約上、タスク割当てを7問目までしか行うことができず、今回の評価が完璧であるとは言えない。特に評価方法1では、予測に用いた履歴が少なすぎるため、ランダムな部分が多くある割当てとなってしまったことや、問題を20問ずつ、7つの小グループに分割したが、小グループ内の問題IDの種類の数がおおよそ4, 5個であったため、次タスクを決めるためのハンガリアン法のコストとなる成否予測確率が不明である部分が、6回の成否予測回数では多く生まれてしまった。そのため、割当てをするための小グループを、予測を次タスクで使えるような設計になるように決定するなどの工夫が必要であると考えた。(図7)

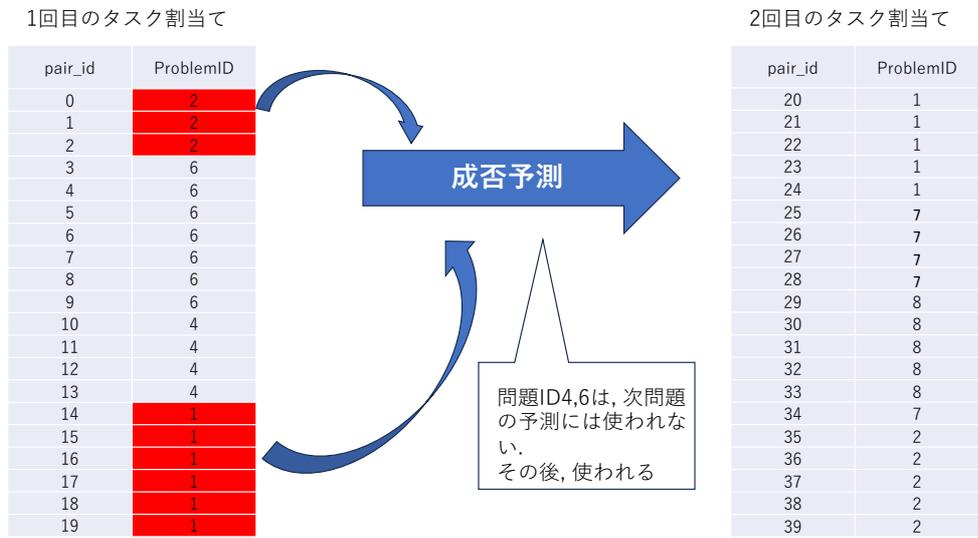


図 7: 小グループと割当てに使えるタスク

第7章 おわりに

クラウドソーシングにおいて、不特定多数の作業者にタスクを割当て、作業させることは、作業者の作業結果には誤りが含まれる可能性が常にある。そのため、誤りを起こさない、減らすということにより、作業結果の品質を管理することが重要な課題となっている。

本研究では、このような背景を踏まえ、成否予測を用いた動的なタスク割当て手法により、能力の高い作業者ばかりでない場合や、作業者の能力にばらつきがある環境下でも、作業結果の正確性を向上させ作業効率を高めることを目指した。具体的には、作業者の作業履歴を、Deep Knowledge Tracing(DKT)モデルを用いて、次タスクの成否予測を行い、その成否予測結果に基づき、次に割り当てるタスクを決定した。その成否予測を用いて、ハンガリアン法で次タスクの割当てを各タスクで行った。またこの手法を用いてシミュレーションを行った。

本研究の貢献は以下の2点である。

作業者のタスクの成否予測

インドネシア語とミナンカバウ語の対訳辞書作成タスクの作業履歴を用いて、DKTで成否予測をすることができた。モデルのAccuracyは約0.6~0.7の間であった。また、評価の際、動的なタスク割り当てによりタスク毎に履歴が長くなっていくことにより精度が徐々に高くなっていくことがわかった。

動的なタスク割当て

DKTから得られた作業者の正解確率を使い、ハンガリアン法で逐次的にタスク割当てをした場合、140タスクを作業者20人に割当てることで、各作業者の正答率は平均で71.0710125%となった。これは、ランダムで割当てるよりも約20%高い結果である。

また、本校では未解決の問題も存在する。評価のプロセスに関して、実験の制約から生じた課題があった。具体的には、評価に利用されたタスクの総数が、時間の制約上、140個に限定してしまったことで、一人当たり7問のみの割当てとなってしまった。その結果、最後に割り当てられた問題でさえ、参照された作業履歴が6問分に限られた。今後の研究では、時間の制約によるタスク割当て回数が少ないという問題を解消し、より大きい作業履歴を用いた

評価を行うことで、DKTモデルを用いた動的なタスク割当ての効果を深く探求する必要がある。もう一つの問題としては、今回は対訳評価タスクについてタスク割当てを実施したが、対訳作成タスクについても同じタスク割当てを行えるのか、行った時の効果を探求する必要がある。

謝辞

本研究を行うにあたり，熱心なご指導，ご助言を賜りました村上陽平教授に深謝申し上げます。また，普段からお世話になっている社会知能研究室の皆様に心より感謝いたします。

参考文献

- [1] Hiroki Chida, Yohei Murakami, Mondheera Pituxcoosuvann, Quality Control for Crowdsourced Bilingual Dictionary in Low-Resource Languages, in Proc. of the 13th Edition of Language Resources and Evaluation Conference (LREC 2022), pp. 6590-6596, 2022.
- [2] 小林 正樹, 人間+AIの相互作用によるクラウドソーシングの品質管理に関する研究, 筑波大学, 2022.
- [3] 鹿島久嗣, 梶野洸, クラウドソーシングと機械学習 (< 特集 > 知識の転移), 人工知能 27. 4, pp381-388, 2012.
- [4] 金地紗里奈, 小坂隆浩, クラウドソーシングにおける品質に対するワーカークの影響, 同志社大学, 研究報告情報システムと社会環境 (IS), 2019-IS-148, 5, pp. 1-6, 2019.
- [5] 潘秋実, 富平準喜, 手塚太郎. Encoder-Decoder DKT モデルによる e ラーニング推薦システム, WebDB Forum 2019 論文集, Vol. 2019, 2019.
- [6] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas Guibas, Jascha Sohl-Dickstein, Deep Knowledge Tracing, Stanford University, Khan Academy, Google, pp. 1-5, 2015.
- [7] 中川大海, 那須野薫, 岩澤有祐, 上野山勝也, 松尾豊, Deep Knowledge Tracing の拡張による擬似知識タグの生成, 人工知能学会論文誌, 33(3), C-H83_1, 2018.
- [8] 亀田 栄一, 篠宮 紀彦, 複数の無線通信サービスが混在した環境における使用アプリケーションを考慮した基地局割当て手法, 情報処理学会研究報告, Vol. 2015-CDS-12, No. 27, pp1-8, 2015.