

時間変動するタスクを多様な処理ユニットに割り付ける強化学習での報酬探索 Reward search in reinforcement learning to allocate time-varying tasks to various processing

森 渉[†] 島川 博光[†]
Wataru Mori Hiromitsu Shimakawa

1. はじめに

タスク割り当て問題は NP-hard 問題であり、解を見つけるために非常に時間がかかる。さらに、時間変動するタスクの場合、処理ユニットに固有の特徴があればその変動により最適な割り当てが変化するため、計算量が増大する。したがって現在タスク割り当てに用いられている手法ではこのような状況の問題を解く際の計算量が膨大になるため解を求めることが困難である。そこで機械学習、特に強化学習を用いれば最適な割り当てを提案する AI が作成できると考えた。

効果的な強化学習のためには、エージェントの行動やプロパティを決定し、行動に対してエージェントが受け取る報酬をシミュレートする環境の構築が必要である。この環境の構築には、自然言語で書かれた要件から必要なエージェントの行動やプロパティを決めるプロセスが含まれる。しかし、文献[1]によると、これまでの手法では環境は常に定常的であることを仮定しているものが多いため、非定常状態における環境では適応することが困難であるとしている。そのため、強化学習を非定常状態に適応させるためには環境構築の過程で多くの手動調整が必要である。特に複雑な環境においては時間と労力が大きな課題となっていた。

本研究では、自然言語で記述された非定常状態の環境構築の要件から、エージェントの行動やプロパティとして導入すべき単語を提案する AI モデルを開発する。この AI モデルを利用することで、環境の作成時に提案された単語を基に行動やプロパティを作成できる。これによって、環境構築にかかる時間が大幅に短縮されるだけでなく、最適な行動とプロパティを環境内に定義することが可能となる。

自然言語で記述された非定常状態の環境要件から効果的に強化学習の環境を構築できるようにすることにより、強化学習の適用範囲が広がり、さまざまな分野での問題が迅速に解決できることが期待される。

2. 文書中からの単語抽出

2.1 単語抽出に使用される技術

近年では、文脈を理解して単語を抽出・分類する技術が提案されてきている。

文献 [2]では、BERT と呼ばれる言語表現モデルが提案されている。BERT は、ニューラルネットワークの全層で左右両方の文脈を考慮することで、ラベル付けされていないテキストから深い双方向表現を事前学習するモデルである。従来の片方向モデルとは異なり、BERT は単語の前後の文脈を同時に理解する。このため、文脈の理解が大幅に向上し、より精度の高い自然言語処理が可能となる。さらに、BERT は特定のタスクにファインチューニングすることで、少ないデータでも高いパフォーマンスを発揮する。

2.2 文章からの単語抽出

BERT を用いた単語抽出の先行研究として研究論文内から要約生成のためのキーワードを抽出する研究[3]がある。この研究で提案された手法ではキーワード抽出のためにチューニングされた BERT を作成し、単語を抽出している。また、入力文書には n-gram と Count Vectorizer を用いて文書内の頻度から n-gram をランク付けすることで要約になり得るキーフレーズを抽出している。

また、文献[4]では、BRRT に注意機構を持たせた分類モデルの手法を提案している。注意機構は入力シーケンスの中で「どの部分が重要か」を動的に判断し、モデルがその重要な部分に重点を置くようにする機能のことである。文献[4]の手法では注意層を持つファインチューニングされた BERT がベースモデルの BERT よりもより良い精度で単語分類ができることが示されている。

さらに、文脈表現の少数サンプルでの BERT のチューニング方法に関する研究として文献[5]がある。この研究では、少数サンプルでの BERT のチューニングでの現在の方法の問題点と解決プロセスが示されている。これによると、BERT で用いられているオプティマイザのバイアス補正の追加処理、事前トレーニングされた BERT の上位のレイヤーの再初期化、エポックの訓練時間の増加により少ないサンプルでの BERT のファインチューニングを安定させ、学習を早めることが示されている。

3. プロパティになり得る単語を推薦する AI モデル

本研究では、環境構築の要件文書から BERT を用いてそれぞれのエージェントのプロパティとして導入すべきキーワードを抽出し、環境作成者に提案する AI モデルを作成する。要件文書からプロパティになり得る単語を抽出するための機械学習モデルの構築とその評価手順を以下の図 1 に示す。

分類モデルに入力するデータを作成するために、入力デ

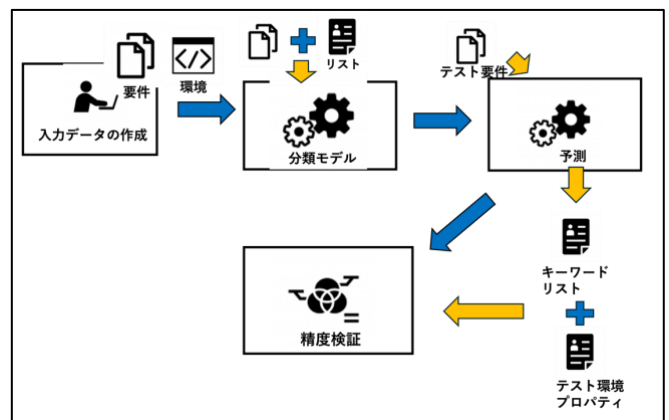


図 1

[†]立命館大学 情報理工学部

ータである環境と環境の構築要件を n 件作成する。また環境からプロパティに対応する単語を取り出したテキストファイルを作成する。これらを分類モデルに入力し学習させる。学習させたモデルにテストデータの要件を入力し、キーワードリストを取得する。得られたキーワードリストとテスト環境のプロパティから作成したモデルを評価する。

3.1 実験で用いるデータ

本研究で用いるデータはタスクが時間的に変動し、それぞれの処理ユニットが固有の特性を持つ場合を考える。その状況で処理ユニットに対してタスクを割り当てるシミュレーションができる環境を作成するための要件文書を対象のデータとする。

文献 [1]によると、非定常状態における環境の変化の仕方に規則性がなければ、解決が不可能な問題であるため、一定の規則性を仮定するべきである。そのため、対象とする環境のタスクの時間変動は一定の規則性を持っていると仮定する。また、本研究では、特徴を持つ処理ユニットの特徴は時間変動しないものとする。

このような実際のシミュレーション環境と環境の構築要件の組みを n 個作成し、環境の要件文書を今回の AI モデルの学習データとして使用する。

テストデータのために 学習データとは異なるシミュレーション環境と要件文書を用意しておき、作成した AI モデルの評価指標の導出のために用いる。

3.2 分類モデル作成の手順

分類モデルの作成では、3.1 で述べたように時間変動するタスクを複数の固有の特徴を持つ処理ユニットに割り当てるシミュレート環境を想定する。

環境の構築要件と環境を構築した際にそれぞれのエージェントのプロパティとなった単語・フレーズを取り出し、それらをまとめたファイルを作成する。このファイルを以後、単語ファイルと呼ぶ。要件文書を説明変数、単語ファイルを目的変数とする。それを入力データとして BERT に環境要件の文章と環境内のプロパティになり得る単語との文脈を学習させ分類モデルを作成する。分類モデルでは入力された要件から単語・フレーズごとにプロパティになり得るクラスタとそうでないクラスタのそれぞれに属する確率を計算し、分類を行う。

本研究ではデータである環境と環境要件を自身で作成するため、学習させるデータが少ない。したがって文献[5]に示されたファインチューニングの方法を取り入れる。それにより、少数のデータでもファインチューニングの安定化と学習の高速化を図る。

また、分類モデルによる精度の向上を目指して文献[4]で使用されていた注意層を BERT 内のニューラルネットワークに取り込むことで注意機構を持った BERT を実現する。学習データのプロパティとその周辺に共起する単語が環境の中でプロパティになり得るクラスタに属する確率が高くなるように注意機構に学習させる。

以上の処理により、少ない学習データで高い分類精度を実現することを目指す。

3.3 モデルの評価方法

文献[2]では、単語の分類問題の精度評価として正解率と F 値が用いられている。正解率はモデルが正しく分類し

た単語の割合を示し、F 値は適合率と再現率の調和平均を表す。適合率はモデルが正しく分類した単語の中で、実際に正解であった単語の割合を示し、再現率はモデルが正しく分類した単語の中でモデルが正しく分類した単語の割合を示す。F 値はこれらのバランスを取るために用いられ、高い F 値はモデルが適合率と再現率の両方で優れていることを示す。

本研究では、モデルから出力されたテスト要件のキーワードリストの単語が実際にテスト要件から作成された環境のプロパティに含まれるかを重要視する。したがって、適合率と再現率の両方を重要視するため、モデルの評価には F 値を用いる。本研究で F 値の計算に用いる際の適合率、再現率は以下の式で定義する。

$$\text{適合率} = \frac{\sum \text{リスト中でプロパティとして使用された単語}}{\sum \text{キーワードリストに含まれる単語}}$$

$$\text{再現率} = \frac{\sum \text{キーワードリスト, プロパティに含まれる単語}}{\sum \text{プロパティとして使用された単語}}$$

3. おわりに

本手法では、タスクが時間的に変動し、それぞれの処理ユニットに固有の特性がある環境を構築する場合を想定した。その環境の要件を BERT に基づく分類モデルに学習させることで、環境構築時にプロパティになり得るキーワードを予測させる。また、モデルの評価には F 値を用いることで適合率、再現率の両方のバランスを取りつつスコアを評価することができる。これによって、環境構築の効率化を図り、強化学習や最適化問題に取り組む際の環境構築の助けになることを目指す。

参考文献

- [1] Samuel P. Choi, Dit-Yan Yeung, and Nevin Lianwen Zhang. 2000. An environment model for nonstationary reinforcement learning "In Advances in Neural Information Processing Systems. 987-993.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. "Bert: pre-training of deep bidirectional transformers for language understanding" in Proc of NAACL-HLT, pp.4171-4186, 2019
- [3] Mushammad Qasim Khan, Abdul Shahid, M, Irfan Uddin, Muhammad Roman, Abdullah Alharbi, Wael Alosaimi, Jameel Almalki, and Saeed M. Alshahrani. 2022. Impact analysis of keyword extraction using contextual word embedding. PeerJ Computer Science 8 (May 2022), e967. <https://doi.org/10.7717/peerj-cs.967>
- [4] M. Tang, P. Gandhi, M. A. Kabir, C. Zou, J. Blakey, X. Luo, Progress notes classification and keyword extraction using attention-based deep learning models with bert, arXiv preprint arXiv:1910.05786 (2019).
- [5] Tianyi Zhang, Felix Wu, Arzo Katiyar, Kilian Q. Weinberger, and Yoav Artzi. 2021a. Revisiting few-sample BERT fine-tuning. In 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net.