

日本語自然言語処理における TASK2VEC の検証

高橋司 櫻井義尚
 明治大学 総合数理学部

1. はじめに

機械学習では、解きたいタスクに対する学習データが少ない場合、類似するタスクで学習したモデル（学習済みモデル）を使ってモデルの精度を上げる転移学習という手法が使われる。その際、適切な学習済みモデルを選ぶ手法として TASK2VEC が提唱されている。TASK2VEC はフィッシャー情報行列を用いてタスクをベクトル化することにより、タスクの類似度を計算できる。先行研究では英語自然言語処理における TASK2VEC の有効性検証が行われていたが、利用言語の違いによりその有効性が変化する可能性がある。そこで、本研究では日本語自然言語処理タスクに対する有効性を検証した。

2. 関連研究

本研究において参考にした TASK2VEC に関する研究は以下の2つが挙げられる。

“TASK2VEC: Task Embedding for Meta-Learning”[1]では、タスクをベクトル化する手法 TASK2VEC が提案された。probe network と呼ばれる ImageNet で事前学習させたモデルによって得られるフィッシャー情報行列を計算し、その行列の対角成分のみを取り出して固定長ベクトルを出力、その後、出力された各タスクのベクトル間でコサイン類似度を計算し最適なソースタスクとしてどのタスクが有効かを検証している。各タスクでの転移学習を踏まえてほとんどのタスクにおいて最適なソースタスクを TASK2VEC によって選択できることを明らかにした。

“Exploring and Predicting Transferability across NLP Tasks”[2]では、TASK2VEC を用いた英語自然言語処理におけるタスクベクトル化手法が提案された。

3. 実験

本研究では日本語自然言語処理タスクに対する TASK2VEC の有効性を検証する。本章では実験の流れ、実験詳細および結果について述べる。

3.1 実験の流れ

1. データセットを用いて学習済みモデルの作成
2. 1で作成した学習済みモデルから TASK2VEC に

従ってタスクをベクトル化、タスク間でのコサイン類似度を計算。

3. モデルから得られるパラメータからそれぞれのタスクにおいて転移学習し、2 で得られた結果とそれぞれのモデルにおける精度から TASK2VEC の有効性を検証し、および日本語データセットにおける特性を分析する。

3.2 実験詳細

3.2.1 データセットについて

表1 データセット

Table 1 Data-set

データセット名	データ数
jrte	3293
chABSA	2813
livedoor	1771
rakuten.1	17128
rakuten.2	17263

本実験ではじゃらんネットロコミ(jrte)[3][7]、有価証券報告書(chABSA)[4]、livedoor ニュース(livedoor)[5]、楽天トラベルデータ(rakuten.1, rakuten.2)[6]の4種類を用いる。楽天トラベルデータは rakuten.1 を 2018 年度のデータセットとし、rakuten.2 を 2019 年度のデータセットとしている。全てのデータセットは2値分類としているが jrte, chABSA, rakuten.1, rakuten.2 をポジティブ・ネガティブによる分類とし、livedoor は家電チャンネルと Sports Watch のトピック 2 値分類とする。

3.2.2 学習済みモデル

本実験では、日本語 Wikipedia で事前学習を行った BERT 事前学習モデルに訓練用データを入力し、ファインチューニングを行う。その際タスク固有の最終層を持たないそのモデルを特徴抽出器とする。

3.2.3 Task embedding

最初に与えられたタスクの訓練データセットで BERT をファインチューニングする。最終的なタスク固有の層を持たないそのモデルは、特徴抽出器を形成する。

次に訓練データセット全体をモデルに入力し、特

徴抽出器のパラメータ (重み) θ のフィッシャー情報行列に基づいてタスクベクトルを計算する.

3.2.4 コサイン類似度

学習済みモデルから計算されたタスクベクトル (768 次元) をそれぞれのタスク同士でコサイン類似度の計算をしていく.

3.3 実験結果

表2 ファインチューニングのみの精度

Table 2 Accuracy based on fine-tuning only

データセット名	精度
jrte	0.836065
chABSA	0.876777
livedoor	0.99322
rakuten.1	0.824518
rakuten.2	0.817518

表3 コサイン類似度

Table 3 Cosine similarity

	jrte	chABSA	livedoor	rakuten.1	rakuten.2
jrte		0.999625	0.989611	0.999641	0.999426
chABSA	0.999625		0.724287	0.999817	0.999729
livedoor	0.989611	0.724287		0.746847	0.807695
rakuten.1	0.999641	0.999817	0.746847		0.999742
rakuten.2	0.999426	0.999729	0.807695	0.999742	

表4 転移学習後の精度

Table 4 Accuracy after transfer learning

↓ trg,src→	jrte	chABSA	livedoor	rakuten.1	rakuten.2
jrte		0.850637	0.843351	0.865209	0.861566
chABSA	0.869668		0.845971	0.879146	0.879146
livedoor	0.99322	0.98305		0.99322	0.98983
rakuten.1	0.825569	0.823817	0.824868		0.82767
rakuten.2	0.820298	0.814389	0.815085	0.81717	

表2 ではそれぞれのタスク同士の転移学習を行っていない, すなわち BERT モデルを各データセットだけでファインチューニングした結果である.

表3, 表4 の結果を見るとほとんどのタスクにおいて類似度が高いソースタスクで転移学習するほど精度が高くなっているのが確認できる.

また, 表3 のポジティブ・ネガティブ分類のドメイン同士におけるコサイン類似度が高くなっていることから, タスクベクトルはそれぞれのタスクのラベルとデータによるパラメータからドメイン本来の類似性を捉えているのが確認できた.

4. おわりに

本研究では日本語自然言語処理タスクに対する有効性を検証した. 結果は, ほとんどのタスクにお

いてタスクベクトルの類似度で選んだソースタスクによって高い精度の値を出すことができた. また, タスクベクトルはそれぞれのタスクのラベルとデータによるパラメータからドメイン本来の類似性を捉えているのが確認できた.

一方, TASK2VEC によって類似度が高いと判断されたデータセットでもデータ数が少ないことから転移学習後の最適なソースタスクとならない可能性がある. 今後の展望としてはデータセット数を均一にして TASK2VEC の計算, 精度検証を詳細に行う実験をしていきたい.

謝辞

本研究では, 国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」

(https://rit.rakuten.com/data_release/) を利用した.

参考文献

- [1] Alessandro Achille, Michael Lam, Rahul Tewari, Avinash Ravichandran, Subhansu Maji, Charless Fowlkes, Stefano Soatto, Pietro Perona. “TASK2VEC: Task Embedding for Meta-Learning” Proceedings of the IEEE International Conference on Computer Vision (ICCV 2019), 6430–6439, (2019).
- [2] Tu Vu, Tong Wang, Tsendsuren Munkhdalai, Alessandro Sordani, Adam Trischler, Andrew Mattarella Micke, Subhansu Maji, and Mohit Iyyer. 2020. “Exploring and predicting transferability across NLP tasks” Proceedings of EMNLP, 7882–7926, (2020).
- [3] Y. Hayashibe. “Japanese Realistic Textual Entailment Corpus”. Proceedings of The 12th Language Resources and Evaluation Conference, pp.6829-6836. (2020).
- [4] Takahiro Kubo, Hiroki Nakayama, Junya Kamura “chABSA: Aspect Based Sentiment Analysis dataset in Japanese”, <https://github.com/chakki-works/chABSA-dataset>, 2018 (accessed 2022/1/7)
- [5]”RONDHUIT”, ダウンロード, <https://www.rondhuit.com/download.html>
- [6] 楽天グループ株式会社 (2020): 楽天トラベルデータ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.2>
- [7] Recruit Co., Ltd., “Japanese Realistic Textual Entailment Corpus”, <https://github.com/megagonlabs/jrte-corpus> (accessed 2022/1/7)