

## エントリーシートを用いた深層学習による採用企業推定

西川 浩平

新卒の就職活動では、就職活動生(以下、就活生)が企業にエントリーシート(以下、ES)を提出し、企業がそのESを元に採用を判断するという過程が一般的となっている。近年では、人によるESの振り分けの前に、機械学習を用いた自動振り分けを取り入れることで、大量のESから自社の採用傾向に合ったESを絞り込み、採用活動を効率的に行おうとする企業が増えている。これに対し、学生が就職活動をより効率的に行い、活動の中でマッチした企業を見つけるには、学生個人の技量に頼らざるを得ない。そのため、企業の採用活動と学生の就職活動との間に技術的な差が生じているのが現状である。この技術的な差により、就活生が自身に最適な企業選択ができないことは、就活生だけでなく、企業にとっても損失である。

本研究は就活生のESから深層学習を用いて、複数の企業から就活生自身にマッチした企業を推定し、就職活動を効率的に行うための支援を目的としたESの採用企業推定の手法を提案する。

本研究では、Web上で公開されている就職活動支援サービスに蓄積されている52社の内定者のES、13,456枚をデータセットとして文書分類を行う。このデータセットは1枚のESに対して1社の企業名が内定した企業としてラベル付けがされたものである。そのため、複数の企業から就活生にマッチした企業を推定するために1枚のES内で質問と回答のペアを訓練データ、検証データ、テストデータに分けた。訓練データに含まれていない質問と回答のペアを検証データ、テストデータとして精度を測る際に用いることで、他の企業に対して書いたESから別の企業を推定する状況を再現する。ESは複数の質問と回答のペアがまとめられたテキストデータであるため、ESを非構造化テキストとして扱うのではなく、質問と回答のペアごとに処理することによって、この構造を保持したまま推定を行う。具体的には、ESの質問と回答のペアを単語の分散表現を用いてベクトル化し、質問と回答それぞれを時系列データを扱うことができる再帰型ニューラルネットワークであるLSTM(Long short-term memory)に入力する。その後、LSTMから出力された質問と回答のベクトルを組み合わせ、質問と回答のペアを考慮した上で全結合層に入力し、SoftMax関数を用いてどの企業のESか推定を行う。

実験の結果、質問と回答のペアを1つの文書として扱った比較手法と比べて、質問と回答をペアとしつつも、それぞれを独立した文として扱った提案手法が高い精度を出した。このことから、質問と回答のペアをそれぞれ独立した文として文ベクトルに変換した上で結合させることで、質問内容と回答内容を保持しつつ、質問と回答の繋がりを表現することが可能であることが明らかとなった。しかし、各企業のエントリーシート内における質問間に強い相関があったため、質問内容のみで企業の推定が行われており、データセットに問題があった。

今後の課題は、質問内容の抽象化やOpenESなどの別のデータセットを利用して、提案手法の有効性を検証することである。

(指導教員 松村敦)