

# エントリーシートを用いた機械学習による採用企業推定

西川 浩平<sup>†</sup> 松村 敦<sup>‡</sup> 宇陀 則彦<sup>‡</sup>

筑波大学 情報学群 知識情報・図書館学類<sup>†</sup> 筑波大学図書館情報メディア系<sup>‡</sup>

## 1. はじめに

新規卒業生の採用活動では、就職活動生（以下、就活生）が企業に自己PRなどを書いたエントリーシート（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、企業側は、機械学習を用いて、ESの振り分けを自動化し、自社にマッチした人材を効率的に獲得しようとして取り組んでいる。一方で就活生は個人の技量に頼らざるを得ない。

そこで、本研究では、就活生のESから深層学習を用いて、そのESの採用企業推定の手法を提案する。具体的には、ある企業に提出するESから別の企業でも内定を獲得できることを就活生に提示するモデルの作成が目的である。これにより、就活生がどの企業の採用傾向にマッチしているかを自分の書いたESから推定することができ、就職活動を効率的に行うための支援につなげることができる。

## 2. 提案手法

提案手法では、ESの質問と回答のペアを単語埋め込みを用いてベクトル化し、質問と回答のそれぞれの文を再帰型ニューラルネットワークであるLSTM(Long short-term memory)[1]に入力する。LSTMとは、時系列データを扱うことができるニューラルネットワークであり、文における単語の並びを考慮するために採用した。その後、LSTMから出力された質問と回答の文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関数を用いて採用企業の推定を行う。

### 2.1. 問題設定

ESの集合を $ES$ とした時、1枚のESを $es \in ES$ と表現する。また、 $es$ は複数の質問 $q$ と回答 $a$ の

ペアからなるテキストデータであるため、 $es = \{(q_1, a_1), (q_2, a_2), \dots, (q_n, a_n)\}$ となる。企業の集合 $C$ をESが分類されるクラスの集合であるとする。

$es$ のテキストデータから採用企業 $c \in C$ を推定する問題とした場合、 $es$ の $q$ の組み合わせのみで $es$ の分類がされてしまう。なぜなら、企業ごとに $es$ の $q$ は統一されているためである。よって、本研究における問題は、 $es$ の $q$ と $a$ の1ペアから、採用企業 $c$ を推定する問題である。

### 2.2. モデル

図1に示す通り、本研究の提案モデルは下記の5つの要素から構成されている。(1) 質問と回答の文中の単語埋め込み、(2) 質問と回答の文単位で表現を生成するためのLSTM、(3) LSTMから出力される質問と回答の最後の隠れ層の結合、(4) 結合された隠れ層を全結合層に入力し、質問と回答のペアの特徴変数の出力、(5) 質問と回答のペアからどの企業に採用された $es$ なのか確率の計算。

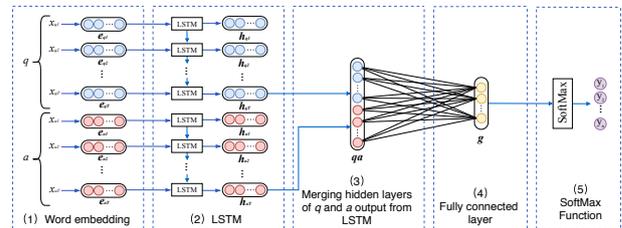


図1: 提案モデルの構造

## 3. データセット

実験では、実際に就職活動で就活生が内定をもらったESを蓄積している就職活動支援サービス「ONE CAREER」[2]のテキストデータを用いた。

### 3.1. データセットの分割

実験で用いるデータは、内定を獲得した企業に対して書かれた $es$ に、その企業名がラベル付けされたものである。

本研究では、ある企業 $c_x$ に提出される $es$ から企業 $c_y$ でも内定をもらうことができることを就活生に提示することができるモデルの作成が目的である。そのため、提案手法の精度を測る上で必要なデータは、ある企業 $c_x$ に対して書かれた $es$ に、

Employer estimation by machine learning using entry sheet

<sup>†</sup> Kohei Nishikawa

College of Knowledge and Library Sciences,  
School of Informatics, University of Tsukuba

<sup>‡</sup> Atsushi Matsumura

Faculty of Library, Information and Media Studies,  
University of Tsukuba

別の企業 $c_y$ が内定企業としてラベル付けされたデータである。データの分割をする際に、訓練データには、企業 $c_x$ に対して書かれ、かつ企業 $c_x$ から内定を獲得した $es$ が存在するようにする。検証データ及び、テストデータには、企業 $c_x$ に対して書かれつつも、企業 $c_y$ から内定を獲得している $es$ が存在するようにする。これにより、企業 $c_x$ のESを学習したモデルに企業 $c_y$ に対して書かれつつも企業 $c_x$ から内定を獲得している訓練データには存在しない未知のESを与えることで、精度を測ることが可能となる。

そのため、本研究のデータセットを用いて上記の状況を再現するには、各 $es$ 内で訓練データ、検証データ、テストデータに分ける必要がある。企業 $c_x$ から内定を獲得している $es_{c_x}$ 内の質問 $q$ と回答 $a$ のペアを 2:1:1 の比率で訓練データ、検証データ、テストデータに分割する。

分割の比率から、本実験では、1枚の $es$ 内に質問と回答が4ペア以上存在し、1社の $es$ 蓄積枚数が100枚以上の企業、52社を対象とした。

## 4. 実験設定

### 4.1. 単語埋め込み

単語を高次元の実数ベクトルで表現する単語埋め込みを得るために、日本語版 Wikipedia を用いて、単語埋め込みの次元数を300次元としてスキップグラムモデルを学習した。また、ES内には「貴社」などの Wikipedia には存在しない未知語が存在するため、日本語版 Wikipedia に加え、本実験で用いるESも単語埋め込みを得るために学習に使用した。

### 4.2. 比較手法

本研究では、Random Forest[3] (以下、RF) と Naive Bayes[4] (NB), LSTM の3つの機械学習アルゴリズムを提案手法の比較手法として用いる。RF, NB では、各単語の出現回数のみを利用する Bag-of-Words を用いて各 $es \in ES$ の質問 $q$ と回答 $a$ のペアを1つの文としてベクトル化し、それを入力として推定を行う。また、Bag-of-Words によって作成される文ベクトルの次元数はデータセット内の語彙数と同じ、36,004次元とした。また、提案手法では質問 $q$ と回答 $a$ のペアをそれぞれ LSTM に入力し、それぞれの最後の隠れ層を結合して文ベクトルを作成するが、比較手法の LSTM では、質問 $q$ と回答 $a$ のペアを一つの文として単語埋め込みによりベクトル化した後、LSTM に入力する。

## 5. 結果

手法別の出力評価 (有効数字3桁で四捨五入した) は、表1に示す。ただし、Precision, Recall, F1 は Macro Average で計算するものとする。表1に示したように、提案モデルの精度がどの比較手法よりも高かった。

表1: 手法別の出力評価

手法	Accuracy	Precision	Recall	F1
RF	0.232	0.201	0.189	0.174
NB	0.305	0.276	0.227	0.200
LSTM	0.320	0.211	0.226	0.207
提案手法	<b>0.380</b>	<b>0.289</b>	<b>0.311</b>	<b>0.279</b>

## 6. 考察・まとめ

本研究は、ESを利用して内定を獲得できる可能性のある企業を就活生に提示することを目的とし、ESの採用企業推定の手法を提案した。実験の結果、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となることが分かった。今後は OpenES[5]のようなデータを用いてモデルの精度を検証することが課題である。

### 参考文献

- [1] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735-1780, 1997.
- [2] 株式会社 ONECARRER. 就活サイト ONECAREER. <https://www.onecareer.jp/experiences>, (参照 2019-09-17)
- [3] Cortes, C., Vapnik, V.: Support-Vector Networks, *Machine Learning*, Vol. 20, pp. 273-297, 1995.
- [4] Breiman, L.: Random Forests, *Machine Learning*, Vol. 45, pp. 5-32, 2001.
- [5] 株式会社リクナビ. OpenES もっと伝わる、自分らしさ. <https://open-es.com>, (参照 2019-12-10)