

就職ポータルサイトにおける企業紹介文評価のための有用単語評価法

1X09C104-4 古山 亮
指導教員 後藤正幸

1 研究背景と目的

近年、多くの学生がインターネット上の就職ポータルサイトを用いて就職活動を行なうようになった。就職ポータルサイトは、学生を採用する立場にある多くの企業にとって、採用申込を受付ける場であるとともに、自社を広く学生に知ってもらう広報の場でもある。各企業は多くの学生にエントリーしてもらい、有望な人材を獲得することを期待しており、その一環として企業は個社ページを持ち、個社情報とともに紹介文を掲載している。しかし、どのような内容の企業紹介文が学生のエントリー行動に結びつくかを定量的に把握することは難しい。

一方、就職ポータルサイトのデータベースには企業情報その他、学生のサイト上での行動履歴といったデータも蓄積されている。そのため、これらのデータを用いてエントリー行動との関係を分析することにより、大量の企業紹介文に対する客観的な評価を統一的に進める可能性がある。

そこで本研究では、データベース上のデータを活用して、学生のエントリー行動に効果的な企業紹介文の内容を定量的に評価する手法を提案する。そのために、企業紹介文に含まれる有用な単語を抽出し、その重要度を計算するモデルを構築する。テキストマイニングの分野では、文書を構成する単語の重要度を測る尺度が存在するが [1]、単語と学生のエントリー行動との関係性を考慮した尺度とはなっていない。そこで本研究では就職ポータルサイトに特化し、学生のエントリー行動に有効であるかという点に主眼を置いた単語評価手法を示す。具体的には就職ポータルサイト大手（以下、サイト A）での各企業の企業情報、紹介文、エントリー情報を用いて、紹介文中の各単語が学生のエントリー行動にどれだけ効果的であるかを意味する評価値を算出する。さらに算出した単語の評価値を用いて、各企業紹介文の評価を行う方法を示す。

提案手法の有効性を検証するため、得られた紹介文の評価値が高い企業および低い企業の例を抽出し、学生による一対比較実験を行う。一対比較により得られた紹介文の順位と提案手法により得られた紹介文の評価値順位とを比較することで、提案手法の有効性を検討する。

2 提案手法

本研究は、各企業紹介文が学生のエントリー行動に与える影響を定量的に測ることを目的としている。そこで企業紹介文の評価指標として、「個社ページを閲覧後、個別エントリー¹を行った学生の数」を用いる。しかし、従業員数などの企業規模や株式公開の有無など、紹介文の内容以外でエントリー数に大きく寄与している要因の存在が考えられる。そこで、紹介文以外の要因を説明変数とし、エントリー数を目的変数として回帰分析 [2] を行い、エントリー数に影響を与える他要因の効果を除外した上で、紹介文の効果の評価することを考える。具体的には回帰分析によって得られたエントリー数の予測値と実測値の残差に紹介文の内容による効果が含まれると仮定し、この残差を用いて企業の紹介文に含まれる各単語について、学生のエントリー行動にどれだけ効果的であるかの評価値を算出する。

¹ 検索結果で表示された複数の企業に対して同時にエントリーを行う一括エントリーに対し、個社ページから 1 社ずつエントリーを行うことを個別エントリーと呼ぶ。

さらに各企業紹介文の単語構成から、算出した単語の評価値を用いて企業紹介文そのものの評価値を算出する。

2.1 重回帰モデルの設定

本研究では、重回帰モデルを以下のように設定した。目的変数 個社ページを閲覧後、個別エントリーを行った学生の数

説明変数 次の 4 変数

- 株式公開の有無
- 従業員規模（8 段階）
- 本社所在地（18 地域）
- サイト A に掲載を開始してからの経過日数

回帰分析の実施に際し、サイト A では少数の人気企業に多くのエントリーが集中していることを勘案し、本研究では目的変数に対して対数変換を行なっている。説明変数に用いた株式公開の有無、従業員規模、本社所在地はいずれも学生が企業を検索する際に指定できるクエリである。

これらはエントリー数に大きく寄与すると考えられ、全ての企業に設定されており欠損が存在しないことから説明変数に採用した。本社所在地については、都市部と地方の学生では就職活動の挙動に違いが見られたため、サイト A が設定している 8 地方区分に加え、大都市を擁する都道府県を個別に扱った 18 地域を本社所在地として設定した。またサイト A 運用会社の担当者より、企業情報の掲載開始からの経過日数もエントリー数に大きく寄与するとの知見を得たため、これを説明変数に加えた。

2.2 企業紹介文の分析

分析対象の企業集合を $C = \{c_1, c_2, \dots, c_j, \dots, c_N\}$ とする。これらの企業紹介文を分析するため、サイト A に掲載されている全企業紹介文から抽出した名詞、動詞により単語集合 $\mathcal{W} = \{w_1, w_2, \dots, w_i, \dots, w_M\}$ を生成する。

本手法では、単語 w_i が学生のエントリー行動をどの程度促すかを示す新たな指標として、単語評価値 s_i を算出する。そのためにまず、 M 種類の単語と N 件の紹介文を用いて、単語文書行列を構成する。単語文書行列の要素は、単語 w_i が企業 c_j の紹介文に含まれるか否かを示す 1,0 のダミー変数 e_{ij} とする。そのもとの、企業 c_j のエントリー数についての残差を r_j とし、単語 w_i の評価値 s_i を以下のように定義する。

$$s_i = \frac{\sum_{j=1}^N r_j e_{ij}}{\sum_{j=1}^N e_{ij}}. \quad (1)$$

評価値 s_i は w_i を含む企業の残差の平均である。評価値 s_i が高い単語を見ることで紹介文における学生のエントリーに効果的な単語を把握できる。

さらに、各企業に対し、紹介文中の単語の平均評価値を用いて企業紹介文そのものを定量的に評価することが可能となる。企業 c_j の評価値 t_j を以下のように定義する。

$$t_j = \frac{\sum_{i=1}^M s_i e_{ij}}{\sum_{i=1}^M e_{ij}}. \quad (2)$$

3 有用単語抽出実験

提案手法の有効性を検証するため、実データを用いて有用単語の抽出実験を行った。紹介文中の内容は業種間での差が大きいため本手法は業種ごとに適用する。

3.1 実験条件

業種 ソフトウェア (463社)・福祉 (189社)・食品 (263社) の3業種

期間 2011年12月1日～2012年3月31日間のエントリー単語集合 サイトAに掲載されている全企業紹介文中で50回以上現れた単語(4209語)のうち,当該業種5社以上の紹介文に掲載されている単語(ソフトウェア:1867語,福祉:1050語,食品:1329語)

3.2 実験結果

(a) 回帰式の評価

各業種における回帰式の評価を表1に示す.いずれの業種も寄与率は0.5前後に留まった.説明変数に用いた4要因のみでは高い予測精度が得られていないが,本研究では紹介文の影響が残差に含まれると仮定し分析を行う.

表1: 各業種における回帰式の評価

業種	重相関係数	寄与率
ソフトウェア	0.653	0.427
福祉	0.692	0.478
食品	0.730	0.534

(b) 単語の評価

それぞれの業種において,単語評価値 s_i が高かった上位10件の単語を表2に示す.

表2: 各業種における上位単語10件

順位	ソフトウェア	福祉	食品
1	制作	商品	予防
2	位	用具	タイ
3	有する	叶える	CM
4	クオリティ	創業	中国
5	挑む	世の中	応用
6	妥協	飛達	盛ん
7	優位	要素	飲料
8	ゲーム	進出	スローガン
9	勝負	例	NO
10	リーディング	基準	溢れる

また,上位30語を含む紹介文を分析し,上位の単語が概ね次の4区分に分類されることが明らかとなった.

- 1) 事業内容に関わる単語
個社の事業内容について述べている単語であり,学生に人気の分野が表されていると考えられる.
- 2) 実績を訴える単語
業界での順位や取引先等の実績を訴える単語が見られる.
- 3) 企業理念に関わる単語
姿勢や社風など煽り文句に含まれる単語が多く現れる.
- 4) 社内制度等を訴える単語
福利厚生や社内教育制度等の単語が多く見られる.

4 学生ユーザによる紹介文評価実験

本手法で提案した企業紹介文の評価指標の妥当性を検証するため,実際の学生ユーザによる評価実験を行った.

予測エントリー数が一致し,紹介文の評価値が上位の企業グループ,下位の企業グループを各 m 件ずつ取り出し,それぞれ $\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_m$ 及び $\tilde{c}_{m+1}, \tilde{c}_{m+2}, \dots, \tilde{c}_{2m}$ とする.社名を伏せて紹介文を表示したもとの,被験者にどちらの企業が魅力的であるかを対比較法を用いて評価させた.各比較には n 人の被験者を割り当てる.また,学生が企業を選択する際の着目点に関する知見を得るため,各企業の紹介文についての意見も付させた.

$2m$ 件の企業を紹介文評価値の降順で並べたとき,企業 \tilde{c}_k の一致率 p_k を以下のように定義する.

$$p_k = \begin{cases} \frac{\sum_{l=m+1}^{2m} v_{kl}}{mn} & k \leq m \text{ (上位企業) のとき,} \\ \frac{\sum_{l=1}^m v_{lk}}{mn} & k > m \text{ (下位企業) のとき.} \end{cases} \quad (3)$$

ここで v_{kl} は,企業 \tilde{c}_l よりも企業 \tilde{c}_k にエントリーしたいと考えた被験者数を表す.一致率 p_k は,企業 \tilde{c}_k について対比較の結果と提案手法の評価が一致した割合を示す.

4.1 実験条件

- ソフトウェア企業
- 2012年12月1日に会社ページを公開
- 株式未公開
- 従業員数500人以上1000人未満
- 本社所在地が東京都
- $m = 5, n = 4$

4.2 実験結果

各企業についての一一致率を表3に示す.

表3: 各企業の一一致率

上位企業	一致率	下位企業	一致率
\tilde{c}_1	0.65	\tilde{c}_6	0.95
\tilde{c}_2	0.80	\tilde{c}_7	0.40
\tilde{c}_3	0.70	\tilde{c}_8	0.55
\tilde{c}_4	0.40	\tilde{c}_9	0.60
\tilde{c}_5	0.55	\tilde{c}_{10}	0.60

4.3 考察

紹介文評価値が高い企業群のうち,企業 $\tilde{c}_1 \sim \tilde{c}_3$ については比較的高い一致率が得られた.高い一致率の要因として,企業 $\tilde{c}_1 \sim \tilde{c}_3$ に対して被験者よりグローバル展開に惹かれたといった意見があった.一方で,評価値の高い単語にも「海外」「グローバル」等が存在し,被験者の意見と評価値の高い単語の一致が見られた.

企業 \tilde{c}_4 や企業 \tilde{c}_7 は,低い一致率となった.企業 \tilde{c}_4 では,全体を通して企業理念や社員の精神についての表現が多く算出評価値は高くなったが,そのため事業内容が不明瞭である点が被験者から不評であった.これらの結果から,先述した単語の4区分のうち1つの内容に偏った紹介文は学生のエントリーを促しにくいと考えられる.内容のバランスを考慮できるような評価指標を改善する必要がある.

企業 \tilde{c}_7 は紹介文評価値が低いグループに属すが,新卒採用に関するこだわりが被験者から好評であった.しかし,本手法では採用基準に関わる単語に高い評価値を与えられなかったため,一致率の低下に繋がったと思われる.したがって,採用基準に関わる単語にも高い評価値を与えられるよう評価指標を改善する必要がある.

特筆すべき点として企業 \tilde{c}_6 の一致率が挙げられる.企業 \tilde{c}_6 は非常にくだけた文体の紹介文であるが,これが被験者から不評であった.高い一致率を示したのは文体による影響が大きく,必ずしも本手法で提案した評価指標が有効に機能したとは言えない.

5 まとめと今後の課題

本研究では,就職ポータルサイトにおける企業の紹介文を定量的に評価する手法を提案し,ユーザ実験を行なってその有効性を検証した.その結果,紹介文の評価値が高い企業については比較的高い一致率が得られた.

しかし,紹介文の文体や構成の偏りなど単語を基盤とした本手法では対応できていない点があるため,今後はこれらの問題点に対応した評価手法を検討する必要がある.

参考文献

- [1] Salton, G. and McGill, M.J., "Introduction to Modern Information Retrieval," McGraw-Hill, 1983.
- [2] 永田靖, 棟近雅彦, "多変量解析法入門,"サイエンス社, pp. 61-86, 2010.