就職活動求職者向けの企業情報推薦システムの提案

鮎 若愚†

白井 康之‡

湊 真一†‡

†北海道大学, 大学院 情報科学研究科

‡(独) 科学技術振興機構, ERATO 湊離散構造処理系プロジェクト

1 はじめに

昨今の就職活動において多くの求職者は就職情報サイトを活用しており, 就職情報サイトも企業情報の提供を目的としたり, 求職者間の情報交換を主な目的としたりと様々なものが存在する. こういった就職情報サイトでは企業情報の獲得や企業へのエントリーを簡単に行えるだけでなく, サイトによっては求職者のプロフィールに応じてお薦め企業のリストを配信するなどの支援も行っている.

同時に, 企業にとって不都合な情報を提供できない. 提供する情報が非常に多くなり, 求職者が提供情報の全てをチェックできない場合が多いといった問題も存在する.

本研究では就職情報サイトからメール配信された情報をもとに取得できる企業情報や, 企業に関連するニュース記事から取得できる対象企業の話題性(記事中の評価表現の極性や割合)に基づく情報推薦システムの提案を行う.

これまでも, ニュース記事や口コミなどのテキストから評価データを抽出し, 利用する研究[金山 07]は盛んに行われてきた. 商品やレストラン, 宿泊施設の消費者への推薦, 株価収益率の予測[前川 13]などもある. しかし, 既存の就職情報サイトは, 多くの場合, 採用する企業側がスポンサーとなっており, 基本的に企業側の要望に応える形で設計されていて, 求職者にとっての利便性については, かなり改善の余地があると感じているので, 筆者は本研究を行うことに意味があると考えた.

2 準備(意見抽出ツール)

本研究では, ニュース記事中に存在する評価表現を抽出するために, 「意見(評価表現)抽出ツール」[NICT 11]を用いる. 「意見(評価表現)抽出ツール」は独立行政法人情報通信研究機構(旧知識処理グループ情報信頼性プロジェクト)によって2011年9月に公開されたものである. 1行につき1文が書かれたテキストファイル

を入力とし, 何らかの事象に対する評価表現が文中に存在するかどうかの判定を行ない, 存在する場合その評価表現, 評価表現のタイプ, 評価表現の極性, 評価保持者を出力する.

抽出される評価表現タイプは「感情±」, 「批評±」, 「メリット±」, 「採否±」, 「出来事±」, 「当為」, 「要望」がある. これらのうち「当為」, 「要望」意外はポジティブ(+)/ネガティブ(-)の極性を持つ.

処理過程のうち評価表現の抽出, 評価保持者の抽出, 評価タイプの分類, 評価極性の判定において, 機械学習用のモデルデータが必要となる. 本研究では, 高度言語情報融合フォーラム(ALAGIN)によって公開された「C-3 意見(評価表現)抽出ツール用モデル(Version 1.1)」[ALAGIN 11]を使用する.

3 提案方法

本研究で提案する企業情報推薦の方法はまず, 既存の就職情報サイトから大量に流れてくるニュース記事を元にして, その中に含まれる企業それぞれについて, インターネット上のweb情報を検索し, 情報推薦を行う. 推薦手順は以下の通りである.

(1) 企業リストを作成

ここでは就職情報サイトに登録し, 配信される電子メールを利用して, 登録された企業ID, 企業名, 企業ページURLを含む企業リストを作成する. 就職情報サイト理由としては, 就職情報サイトでは自分の学歴などの個人情報や, 希望業種, 勤務地などのデータを登録することにより, それらのデータに基づいたお薦め企業のリストを電子メールで配信されるからである.

(2) 企業情報の取得

(1)で得られた企業ページURLを利用して, 就職情報サイトから対象企業の基本情報や採用情報を取得, 整理する.

(3) 企業の話題性の取得

ここでは企業に関連するニュース記事中の評価表現を企業の話題性として抽出する.

まず Google News Search API を用いて, それぞれの企業についての検索結果(ヒット数, タイトル,

A Company Information Recommendation System for Job Hunting Students

†Ruoyu Bao ‡Yasuyuki Shirai †‡Shin-ichi Minato

†Hokkaido University, Graduate School of Information Science and Technology

‡Japan Science and Technology Agency, ERATO Minato Discrete Structure Manipulation System Project

ニュース記事の URL など) を取得する。次にそれぞれのニュース記事の本文を取得、整形してから「意見(評価表現)抽出ツール」を適用することにより、記事中の評価表現を抽出し、各企業、各記事ごとにまとめる。

(4) 求職者への提示

全企業については、ポジティブな評価表現が多い企業、ネガティブな評価表現が多い企業というような評価表現の割合に基づくランキング形式で求職者に提示する。

また各企業については、(2) で得られた企業情報と、ポジティブ評価が多いニュース記事、ネガティブ評価が多いニュース記事をピックアップして提示する。

4 実験と考察

本研究では、株式会社リクルートホールディングスが提供する就職情報サイト「リクルートナビ」[リクナビ 15]を利用して、2013年7月31日から2013年12月31日までに送信された電子メールから企業リスト(計368社)を作成し、実験を行った。また、Google News Search API を用いたニュース記事検索及び記事本文取得において、ニュース記事を取得出来た企業は157社である。APIの使用制限により、ヒット件数にかかわらず取得できたニュース記事は各企業について最大64件である。

そこから抽出された評価表現を各企業、各記事についてまとめ、図4、図4のように提示する。

このように、メール配信された企業について、求職者に対し企業情報や関連ニュース記事から取得できる話題性を提供することで、企業の選択、研究に役立つことができる。

図1: 提示例 - ポジティブ評価順

順位	企業名	ポジ評価	ネガ評価
1	企業A	97%	3%
2	企業B	95%	5%
3	企業C	94%	6%
...	...	...	...

図2: 提示例 - 企業別

企業N					
業種: 通信 職種: 事務系, IT系, ... 勤務地: 北海道, 東京, ...					
ポジティブ寄りニュース記事			ネガティブ寄りニュース記事		
タイトル	ポジ割合	ネガ割合	タイトル	ポジ割合	ネガ割合
「泣ける」と話題...	87%	13%	企業N「未承諾広...	50%	50%
企業Nの参入で...	73%	27%	企業Nは高値更新...	67%	33%
...	...	...	...	...	...

5 まとめ

現在の課題は

- ・ 企業名
本研究の実験では出来るだけ多くのニュース記事を取得するために、企業名から「株式会社」を取り除いたうえで、フレーズ検索を行った。しかし、中には「縁」、「無限」のような一般的な単語や人名などを企業名とする例も多い。こういった企業については無関係なニュース記事を多く取得してしまうため、正確な結果が得られなかった。
- ・ 企業数
本研究で作成した企業リストに含まれる計368社のうち、実際に関連ニュース記事を1件以上取得できたのは157社のみである。
- ・ 「意見(評価表現)抽出ツール」
本研究ではニュース記事中の評価表現抽出に「意見(評価表現)抽出ツール」を使用したが、精度があまり高くない、導入が難しいといった問題がある。

本稿では、企業に関連するニュース記事に含まれる評価表現に着目し、それらをまとめて求職者に提示することで、求職者企業選択や企業研究に役立てることを目的とした。これに加え、求職者の情報や志向に基づく企業との自動マッチング、企業間の類似度に基づく企業推薦といった推薦方法も考えられる。

参考文献

[金山 07] 金山 博: テキストを用いた評判と嗜好の分析, 情報処理 48(9), 1001-1007, 2007-09-15.

[前川 13] 前川浩基, 中原孝信, 岡田克彦, 羽室行信: 大規模ニュース記事からの極性付き評価表現の抽出と株価収益率の予測, オペレーションズ・リサーチ: 経営の科学 58(5), 281-288, 2013-05-01.

[NICT 11] 「意見(評価表現)抽出ツール」, <http://alaginrc.nict.go.jp/opinion/#reference>.

[ALAGIN 11] 意見(評価表現)抽出ツール用モデル (Version 1.1), <http://alaginrc.nict.go.jp/resources/nictmaster/resource-info/abstract.html#C-3>.

[リクナビ 15] リクナビ 2015, <http://job.rikunabi.com/2015/>.