

## 企業の口コミを用いた企業の特徴の抽出方法

山田一輝<sup>†</sup> 村田皓教<sup>†</sup> 炭竈桂輔<sup>‡</sup> 遠藤正隆<sup>‡</sup> 中嶋裕一<sup>‡</sup> 三浦哲郎<sup>‡</sup> 田中大地<sup>‡</sup> 菱田隆彰<sup>†</sup>  
愛知工業大学<sup>†</sup> 株式会社リオ<sup>‡</sup>

### 1 はじめに

自身が勤めたことのない企業の優れた点や問題点を知ることは難しい。就職・転職を望む人々にとってできる限り詳細な企業の実態を知ることが人生を決める上で重要な作業と言えるだろう。インターネットが普及し、ユーザが発するさまざまな情報が公開されるようになった。企業に対する口コミもその一つであり、その口コミは企業の実態を知る有用なツールとして活用されている。しかし、他人が書いた膨大な量の口コミを整理して読むことは容易ではなく、ユーザが短時間で企業の特徴を把握することは困難な作業である。企業の口コミから特徴となる情報を自動的に抜き出す仕組みを構築することができれば、ユーザは口コミを全て読まなくても、企業のあらましを知ることができる。本研究では、転職サイトなどの企業に対しての口コミから、企業の特徴に相当する情報を抽出する手法を提案する。

### 2 関連研究

消費者の口コミ情報への関心の高さに着目しインターネット上のブログから評判情報を抽出するシステムが開発されている。長野ら[1]は、ブログ記事に含まれている口コミを自然言語処理(形態素解析, 構文解析)で解析し, 主語(品名), 属性(評価視点), 属性値(評価表現)の抽出を行った。そして, ある分野で使われる語彙及び, 語彙間の関係を体系化したオントロジーの辞書を用いて照合し, 属性, 属性値の組がオントロジー上に存在するかどうかを調べることによって, その対象商品に対する評価の判定を行い, 口コミが“良い”もしくは“悪い”のかを分類して提示する手法を提案している。

企業の就業環境に関する口コミサイト, いわゆる転職サイトの口コミは, 整った文章で書かれており, 比較的長文で, 複数の話題に触れて

いる場合が多く, 投稿者の経験を踏まえた明確な評価が述べられているという特徴を持つ。ユーザは口コミ内の大量の文から, それぞれ必要とする話題に着目し, その企業の評価を抽出する作業が必要になるが, その作業は簡単ではない。

本研究では, 特定の企業に対する口コミから企業の特徴を分析する手法を提案する。特徴は口コミ内の用語の傾向を分析することで企業の傾向とする手法と, 口コミ内に含まれる投稿者の企業に対する評価を述べた文を抽出し, その内容の良し悪しを分類する手法の2つの手法を提案する。

企業の傾向を得る全体の手順は,

- (1) 事前準備: SVMによる学習
- (2) 特徴の分類

企業の評価を抽出する手順は,

- (1) 事前準備: 企業の特徴に関する辞書の作成
- (2) 口コミの構文解析と係り受け文の作成
- (3) 辞書との照合と企業評価文の作成

のように行う。手順の詳細はそれぞれ3節, 4節に示す。

### 3 企業の傾向の分析

- (1) 事前準備: SVMによる学習

口コミから企業の傾向を得るためには, ある口コミがあるカテゴリに属するかを分類する必要がある。口コミを分類するために, SVM (support vector machine)を用いる。SVMとはパターン認識器の一つであり, 与えられた未知のデータを, 学習モデルを元に分類することができる。

今回は企業に対する特徴的な4つのカテゴリ(残業, スキル, 女性, 給料)を用意し, それぞれのカテゴリで出現する単語を抽出した辞書を元に学習モデルを作成する。辞書の作成にはVorkers[2]を含む企業の口コミサイトから収集した口コミを利用する。それぞれのモデルデータには, 3つのラベル(属さない, 話題を含む, 評価が得られる)を付加する。学習に必要な入力データは, カテゴリごとに用意したサンプルの口コミに対して形態素解析を行って得られた単語の出現頻度を用いた。

An extraction method of the corporate features from reviews of the employees

Kazuki Yamada<sup>†</sup> Hironori Murata<sup>†</sup> Keisuke Sumigama<sup>‡</sup>  
Masataka Endo<sup>‡</sup> Yuichi Nakashima<sup>‡</sup> Tetsuro Miura<sup>‡</sup>  
Daichi Tanaka<sup>‡</sup> Takaaki Hishida<sup>†</sup>

<sup>†</sup> Aichi Institute of Technology University

<sup>‡</sup> RIO CORPORATION

(2) 特徴の分類

(1)で作成したそれぞれの学習モデルを元に、収集した企業の口コミ全てに対してカテゴリの分類と口コミの抽出を行う。各口コミはラベルに紐付けた数値（1（属さない）、2（話題を含む）、3（評価が得られる））が得られるため、口コミ全ての値の総和を口コミ数で割った値（代表値）を求める。カテゴリ毎の代表値を集めたベクトルを企業の傾向とする。また、“話題を含む”と判断された口コミは何らかの企業の重要な情報を得られる可能性が高いため、結果として元の口コミをそのまま出力する。

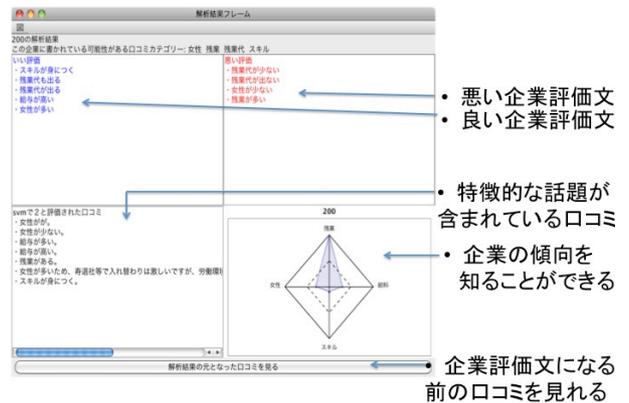


図1 企業の分析結果の例

4 企業の評価の抽出

(1) 事前準備：企業の特徴に関する辞書の作成

口コミの文章から企業の特徴を表す文章を抽出するため、関連する用語の辞書を作成する。

企業の口コミには、“給料”、“残業”など企業の特徴を示す単語が多く含まれている。このような口コミは利用者にとって、企業の状態を判断する重要な文章と言える。また、その特徴語を説明する形容詞，“多い”や“良い”などが含まれる場合、その内容は企業の評価を表す文章になると考えられる。

作成する辞書には、企業の特徴となる単語（特徴語）とその単語を形容することで評価を表す単語（評価語）のペアを登録した。例えば、辞書に登録した語句のペアとしては、（“残業”，“多い”），（“離職率”，“高い”）などがある。加えて、各ペアの評価の良し悪しを区別するため，“良い”，“悪い”のラベルを付加する。

(2) 口コミの構文解析と係り受け文の作成

口コミの構文解析を行い、連結処理を行う。構文解析には日本語係り受け解析器CaboCha[3]を利用した。口コミを句点で文に分解し、さらにその文を読点で分解して分析を行う。CaboChaの出力には文節番号と係り先番号が含まれるため、これらを利用して係り受け関係にある文節同士を繋げた文（以下、係り受け文と呼ぶ）を作成する。

(3) 辞書との照合と企業評価文の作成

(2)で作成した係り受け文と(1)で作成した辞書との照合を行い、企業の特徴を表す内容かどうかを判別する。まず、作成した係り受け文のうち、辞書に含まれている特徴語と評価語のペアが共に含まれているものを抽出し、企業評価文の候補とする。候補の中には、係り受けの関

係や辞書の組み合わせによって、評価文とならないものや、重複して抽出してしまう文があるため、それらを取り除いた結果を企業評価文とする。選出した企業評価文は評価の良し悪しを区別し結果として出力する。

5 結果とまとめ

提案手法を導入した企業分析システムの施策を行った。構築システムの分析結果の例を図1に示す。3節で求めた口コミ内の特定のカテゴリに含まれる話題の出現度を企業の傾向としてグラフ表示と特徴的な話題が含まれている口コミを列挙し、4節に示した企業の評価文を企業の“良い点”，“悪い点”として区別して列挙した。ユーザは口コミを全て読むことなく全体の話題の傾向を一覧可能であり、企業のあらましを知ることができる。

本研究では、転職サイトの企業に対する口コミに対して、SVMを用いた文章の分類と特徴語と評価語の辞書を用いた特徴的な文章の抽出を行うことで、企業の特徴に相当する端的な情報が抽出可能であることを示した。今後は適切な分類を行うために、コーパスに利用する語の見直しや、辞書の充実を図り、分析精度を向上させたいと考えている。

参考文献

[1]長野伸一，溝口由美子，稲葉真純，インターネットから評判情報を抽出する口コミ分析技術ユビ de コミミハサンダー，東芝レビュー，Vol.63, No.10, pp.35-38, 2008.  
 [2]Vorkers, <http://www.vorkers.com/>  
 [3]CaboCha/南瓜，Yet Another Japanese Dependency Structure Analyzer, <http://code.google.com/p/cabocha/>