

# 単語共起関係を用いた求人情報の分析事例について

石嶋 秀太, 宇田川 佳久<sup>†</sup>

東京工芸大学工学部 コンピュータ応用学科<sup>†</sup>

## 1. はじめに

2013年春の大学卒業生の求人倍率は1.27倍であり、この指標を見る限り大卒者の就職は困難な状況ではない。しかし、最近の文部科学省の調査によれば、卒業生の2割以上が安定雇用には就いていない。この原因として大学生の業界・企業研究が不足していることが指摘されている。本文では、企業が求める人材像についてデータマイニング技法を使って分析する。分析技法としては、単語の発生頻度に加え、単語間の共起頻度を対象とする。単語間の共起を分析することにより、企業が求める人材像、企業がアピールする教育・研修制度などを抽出し、視覚的に表示する。

本研究では商用のテキストマイニングツールを用いた。

## 2. 企業データの収集と加工

分析対象としたデータは、本学の企業説明会に参加した企業のプロフィール資料から、事業内容の紹介と学生へのメッセージ文を編集したものである。分析対象とした事業内容とメッセージは通常使われている日本語の文章である。機械的な分析を意識せずに書かれているため、下記のような編集、修正をした。

- (1) 企業固有の製品名、地名、部署名、会社名などは、データ作成時に削除した。
- (2) 標語のように、名詞形で終了している場合は、文章になるよう動詞を補った。
- (3) 文章の最初で、発話の主体（主語）が不明確な場合、「当社は」などを適宜補った。

## 3. データマイニング機能

### 3.1 形態素解析機能

計算機で自然言語の文章を処理するためには、形態素（言語として意味を持つ最小単位）の列に分割し、それぞれの品詞を判別する形態素解析を行う。

### 3.2 形態素解析結果

表1は、形態素の品詞別の発生件数を表している。分析対象とした79社のデータを形態素解析した結果、事業内容に関しては726個、メッセージに関しては2375個の形態素に分解された。品詞別では、事業内容とメッセージ文とも、名詞が圧倒的に多い。これは、内容を説明する文章が多いことを反映しているためと考えられる。

一方、動詞の発生比率は、事業内容の文章では2.20%であるのに対し、メッセージの文章では11.62%である。形容詞、副詞、連体詞の発生比率も、メッセージ文が、事業内容の文章を上回っている。これは、動きのある、感情を込めた表現が、メッセージ文に多く含まれていることを反映しているためと考えられる。

表1. 形態素の品詞別発生件数

	事業内容		メッセージ	
	発生数	発生数の比率	発生数	発生数の比率
名詞	687	94.63%	1900	80.00%
動詞	16	2.20%	276	11.62%
形容詞	1	0.14%	74	3.12%
副詞	0	0.00%	61	2.57%
連体詞	2	0.28%	31	1.31%
接続詞	19	2.62%	29	1.22%
記号	1	0.14%	3	0.13%
間投詞	0	0.00%	1	0.04%
合計	726	100.00%	2375	100.00%

### 3.3 類義語辞書機能

解析した文章に、「会社」、「当社」、「弊社」という用語が現れるが、これらは同じ意味に使われていると解釈できる。作成した類義語辞書を適用することにより、類義語欄に記載した単語が、代表語欄に記載した単語に統合される。図1は情報業界のメッセージ文章に対し、類義語辞書を適用して得られた単語の発生頻度の多いものから順に表示したものである。

同じ略語が異なる分野で使用されることがある。例えば、PCは情報分野ではパーソナル・コンピューターであるが、建築分野ではプレキャスト・コンクリートである。本研究では、業種ごとに分析対象とする文章データを作成し、その文章データに対して類似語辞書を定義している。

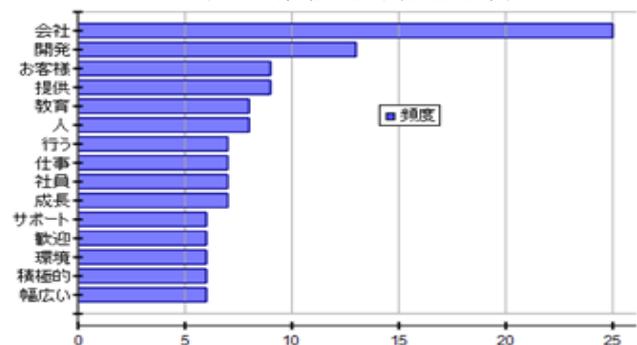


図1. 発生頻度の多い単語

