

企業トップのメッセージに基づく 企業風土の特徴分析

佐藤裕也

掛谷英紀

s1420791@u.tsukuba.ac.jp

kake@iit.tsukuba.ac.jp

筑波大学大学院システム情報工学研究科

概要 本論文では自然言語処理を用いて社長のメッセージをもとにその企業の平均勤続年数が長いかどうかを推定する手法を提案する。機械業の上場企業のうち、1975年以前に設立された206件のメッセージを収集し、平均勤続年数によって2つのカテゴリに分類する。使用するメッセージは株主向けかどうか、文字数、社長の就任している年数、使用するメッセージ数によって場合分けを行う。機械学習には最大エントロピー法を用いて、結果の検証には10分割クロスバリデーションを使用する。十分な正解率を得ることはできなかったが、平均勤続年数が長い企業と短い企業の素性データには特徴がみられた。また、平均勤続年数が特に長い企業と特に短い企業の正解率は他に比べ高かった。

1. はじめに

近年、若者を大量に採用し、過重労働、違法労働をさせて離職に追い込むいわゆるブラック企業が注目を浴びている。2013年にはブラック企業がユーキャン新語流行語大賞のトップテンに入り、ブラック企業対策プロジェクト共同代表の今野晴貴氏が受賞した[1]。学生が就職する企業を決定する上でもブラック企業であるか否かは一つの重要な指標であるといえる。ブラック企業は若者を大量に採用し使い捨てるため、離職率が高くなる。また、若者が大量に採用されることから平均勤続年数も短くなる。

厚生労働省はブラック企業を「若者の使い捨てが疑われる企業」と称して、ブラック企業対策の取り組みを強化している。その中で厚生労働省は2015年春卒業予定の大学生、大学院生を雇用する企業に対し過去3年間の採用者数と離職者数を求人票に明示するよう要請した[2]。また、それを受けてハローワークは過去3年間の採用者数と離職者数の記入欄を設置した。しかし、企業側にはそれらを記入する義務がない。そこで離職率の次にブラック企業を判定する指標として平均勤続年数もあげられる。上場企業は有価証券報告書内で記載の義務がある。しかし、上場企

業でなければ記載の義務はなく、平均勤続年数を公開している企業であっても会社設立から間もない企業のそれは離職者数を推定するデータとして妥当性があるとはいえない。実際に企業の内部に入って視察を行えばその企業がブラック企業であるか否かの判断はつくと考えられるが、全ての企業に対して視察を行うことは現実的でなく、また就職活動を行う学生には就職を希望する企業全てを訪問する時間はない。そこで、離職率と平均勤続年数以外の指標を用いてそれらを推定する必要があると考えられる。

企業風土の特徴分析を行う手段として、戸塚は自然言語処理によって経営トップのメッセージを分析する手法[3]を提案した。企業の経営トップのメッセージを株価が安定である企業と不安定である企業に分類して、最大エントロピー法を用いた機械学習とクロスバリデーションによる判定からメッセージの特徴分析をした。その結果、2分類での正解率は65.28%となり、3分類での正解率は46.28%であった。クロスバリデーションでの正解率は高くないが、機械学習の結果得られた α 値の高い素性から、それぞれのカテゴリに関して経営方針や特徴を表す素性を得ることができている。したがって、経営トップのメッセージと企業風土の間には何らかの因果関係があるといえる。

そこで、本研究では自然言語処理を用いて、平均勤続年数の長い企業と短い企業の経営トップのメッセージを分析し、離職者数や平均勤続年数の公開されていない企業およびそのデータの有益性が疑われる企業の企業風土を推定することを目的とする。

2. 手法

2.1 概要

本研究では企業サイトから収集した経営トップのメッセージを形態素解析ツール(ChaSen)[4]を通して、それぞれの単語すなわち素性ごとに分割してその際に品詞から使用する素性データを決定する。その素性データを最大エントロピー法を用いた機械学習[5]にかけることでカテゴリ別分類を行う。カテゴリ別分類とは学習データから文章のカテゴリを判定する機械学習の手法である。判定させたい文章を入力し、機械学習によって得られた特徴からその文章がどのカテゴリに属する確率が高いかを算出し、最も確率の高いカテゴリに判別する。また、カテゴリ分類に使用される指標である素性の α 値を見ることで素性の分析を行う。

2.2 使用するデータ

本節では機械学習に使用する経営トップのメッセージについて説明する。

まず、収集するデータについて説明する。カテゴリ分類をする指標として離職者数と平均勤続年数が考えられるが、上場企業は有価証券報告書内で平均勤続年数の記載の義務があるので平均勤続年数を指標としてカテゴリ分類を行う。20歳で入社した社員が定年になるまでに40年かかることから会社設立後40年未満の企業の平均勤続年数には離職率を推定するデータとしての妥当性がない。以上のことから、分析に用いる経営トップのメッセージとして

- ① 上場企業であること
- ② 会社設立後40年以上経過していること

の2つの条件に当てはまるものを収集した。本研究では1974年以前の上場企業を収集する。また業種間でトップメッセージの特徴に違いがある可能性を考慮し、今回は一

つの業種に限定して実験を行う。そこで

③ 企業の業種が機械業であること

も条件に加える。本研究では、2014年6月から7月にかけて各企業内で公開されていた、上記の3つの条件に当てはまる206件の企業トップのメッセージを用いる。

次に、収集したデータの中から実験に使用するデータについて説明する。使用するデータの組み合わせをあらゆる条件で変更してどの条件において高い判別率が得られるかを調べる。経営トップのメッセージには一般向けのメッセージと株主向けのメッセージがあるが、株主向けのメッセージしか公開していない企業がある。そこで

A 全てのメッセージを使用する場合と一般向けのメッセージのみを使用する場合

で場合分けをする。機械業において、株主向けのメッセージのみを公開している企業は46件ある。また、極端に短いメッセージが機械学習の結果に影響を及ぼす可能性を考慮し、トップメッセージの文字数に制限をつける。そこで

B 全てのメッセージを使用する場合、150字未満のメッセージを使用しない場合と200字未満のメッセージを使用しない場合

で場合分けをする。機械業において、150字未満のメッセージを公開している企業は9件、200字未満のメッセージを公開している企業は150字未満のメッセージを公開している企業を含め16件ある。さらに、社長が就任してから企業風土に影響を与えるまで数年かかる可能性を考慮し、社長が就任してからの年数にも制限をつける。そこで

C 全てのメッセージを使用する場合、社長が就任してから3年以上の企業のメッセージのみを使用する場合と社長が就任してから5年以上の企業のメッセージのみを使用する場合

で場合分けをする。2014年10月時点で機械業において、

社長が就任してから3年未満の企業は56件、5年未満の企業は3年未満の企業を含め85件ある。また、平均勤続年数の長い企業から順に並べたときに、中央付近の企業のカテゴリ分類は困難であると考えられる。そこで特徴が顕著に現れると考えられる平均勤続年数の長さが上位の数十社と下位の数十社のみで比較してカテゴリ分類を行うことでクロスバリデーションの再現率が上昇するか検証する。そこで

D 全てのメッセージを使用する場合、上下70社のみを使用する場合、上下60社のみを使用する場合、上下50社のみを使用する場合と上下40社のみを使用する場合

で場合分けを行う。以上の4つの条件で場合分けを行い再現率が高くなる条件を探索する。ただし複数の条件を組み合わせた際に企業数の合計が140社を下回った場合上下70社のみを使用する条件は除外する。

3. 結果

前章の条件で使用するデータを変更し名詞のみの素性データから機械学習をし、10分割のクロスバリデーションをした結果、株主向けのメッセージを使用し、200字未満のメッセージを使用せず、社長が就任してから3年以上のメッセージのみを使用し、平均勤続年数が上下40社のみメッセージを使用した場合の再現率が最も高くなった。平均勤続年数が上位の企業の再現率は65%であり、平均勤続年数が下位の企業の再現率は75%であった。平均勤続年数の長さで全体を8分割した時の再現率を図1に示す。平均勤続年数が長い10社および短い10社の再現率は90%である。また、この判定において重要な手がかりになった、つまり α 値の高い上位20個の素性を平均勤続年数が長い、短い企業それぞれについて表1に示す。

また、使用するデータに関する4つの条件に関して、各条件を固定して他の条件全てを使った場合の10分割クロスバリデーションの再現率の平均を求めた。株主向けのメッセージに関して、全てのメッセージを使用した場合の再現率の平均は54.47%で、株主向けのメッセージを除外して機械学習をした場合は56.34%であった。メッセージの文字数に関して、全てのメッセージを使用した場合の再現

率の平均は55.40%で、150字未満のメッセージを使用しない場合は56.54%、200字未満のメッセージを使用しない場合は55.65%となった。社長の就任している期間に関して、全てのメッセージを使用した場合の再現率の平均は55.86%で、3年以上に限定した場合は59.70%、5年以上に限定した場合は57.12%となった。使用するデータ数に

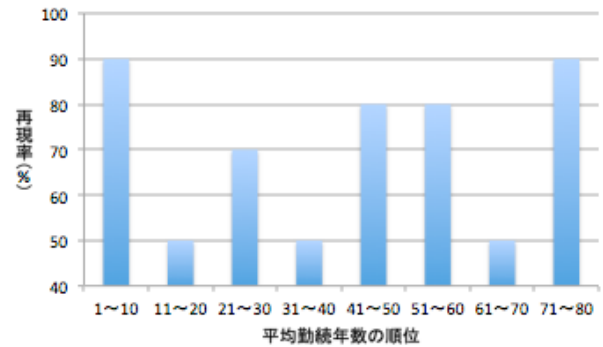


図1 平均勤続年数の順位別の再現率

表1 判定の上位素性

	上位	下位
1	活動	業界
2	づくり	お客様
3	年	多様
4	効率	ニーズ
5	現在	中心
6	時代	力
7	昭和	人々
8	国	販売
9	形	商品
10	機械	生活
11	建設	私
12	地域	たち
13	製品	開発
14	可能	もと
15	分野	拡大
16	品質	情報
17	もの	これ
18	価値	努力
19	挑戦	確率
20	産業	地球

に関して、全てのメッセージを使用した場合の再現率の平均は 52.63%で、平均勤続年数の長い 70 社と短い 70 社に限定した場合は 54.29%、60 社ずつに限定した場合は 55.90%、50 社ずつに限定した場合は 61.92%、40 社ずつに限定した場合は 55.63%であった。

4. 考察

使用するメッセージの組み合わせを変更して実験を行った結果、最良の結果が再現率 70%であった。ランダムに答えても再現率は 50%になることを考慮すると十分な結果とはいえないが、平均勤続年数の特に長い企業と特に短い企業においては再現率が高くカテゴリの傾向をつかめているといえる。株主向けのメッセージと使用する文字数の条件に関しては、条件を固定しても再現率に大きな差は見られなかった。一方、社長就任年に関しては 3 年以上に限定した方が、再現率が高くなるといえる。また、使用するデータ数に関してはメッセージに特徴が現れやすいと推測される上下数十社に限定した方が高くなるといえる。ただし、データ数を限定しすぎると機械学習を行う上で十分なデータ数を得られなくなるため再現率が低くなったと考えられる。

また、平均勤続年数の長い企業と短い企業の経営トップのメッセージの素性には明らかな傾向が見られる。平均勤続年数が長い企業に関しては、「もの」「づくり」や「機械」「建設」「製品」といった製造する過程を重視する企業が多いと見られる。一方で、平均勤続年数が短い企業に関しては、「お客様」「ニーズ」や「人々」「商品」「販売」といった顧客を相手に商品を販売することを重視する企業が多いと見られる。図 1 のようにこの傾向は平均勤続年数が特に長い企業や短い企業に多く見られると考えられる。

5. まとめ

本研究では、企業の平均勤続年数が長いか短いかを判別する新たな指標として経営トップのメッセージを提案した。経営トップのメッセージを平均勤続年数によりカテゴリ分類をして分析をした。その結果、10 分割クロスバリデーションによる再現率は平均勤続年数が特に長い企業

と特に短い企業では高く、その結果得られた α 値の高い素性から、各カテゴリに関して特徴的な素性を得ることができた。

本研究で用いたデータは数多く存在する企業の中で機械業という一つの業種に限定したものであるため、本研究での結果の一般性を証明するために他の業種のデータも収集して分析する必要がある。また、機械業においても再現率は十分であるとはいえないので今後再現率を高めるための手法を探す必要がある。

参考文献

- [1]新語・流行語大賞
<http://singo.jiyu.co.jp>
- [2]リーフレット・厚生労働省
<http://aichi-hellowork.jsite.mhlw.go.jp/var/rev0/0110/3821/201422819540.pdf>
- [3]戸塚貴之, (2008), 自然言語処理による経営トップのメッセージ分析, 筑波大学工学システム学類卒業論文
- [4]奈良先端科学技術大学院大学, 松本研究室, ChaSen
<http://cl.aist-nara.ac.jp/>
- [5]内本清貴, 村田真樹, 関根聡, 居佐原均, (1999), 日本語係り受け解析に用いる ME モデルと解析精度, 言語処理学会第 5 回年次大会併設ワークショップ論文集
- [6]内山将夫, maxent
http://www2.nict.go.jp/univ-com/multi_trans/member/mutiyama/index-ja.html