

Title	コールセンターに寄せられる顧客の声からの印象判別アルゴリズムによる有益情報の抽出
Sub Title	
Author	高木, 宏明(Takagi, Hiroaki) 林, 高樹(Hayashi, Takaki)
Publisher	慶應義塾大学大学院経営管理研究科
Publication year	2015
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2015年度経営学 第3061号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002015-3061

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

慶應義塾大学大学院経営管理研究科修士課程

学位論文（ 2015 年度 ）

論文題名

コールセンターに寄せられる顧客の声からの
印象判別アルゴリズムによる有益情報の抽出

主 査	林 高樹 教授
副 査	岡田 正大 教授
副 査	中村 洋 教授
副 査	住田 潮 特任教授
副 査	岩本 隆 特任教授

学籍番号	81430663	氏 名	高木 宏明
------	----------	-----	-------

論 文 要 旨

所属ゼミ	林 高樹 研究会	学籍番号	81430663	氏名	高木 宏明
(論文題名) コールセンターに寄せられる顧客の声からの 印象判別アルゴリズムによる有益情報の抽出					
(内容の要旨) ■研究背景 21 世紀に入り、ICT 技術の向上、情報ネットワークインフラの普及、記憶装置・センサー装置などの低価格化により膨大なデータを取得・分析することが可能となった。今日では、「ビッグデータ」「人工知能」「IoT」のキーワードが新聞紙面でも頻出するようになり、データを経営に活かすことが大きな企業課題となっている。					
■研究要旨 本研究では、産学共同プロジェクトを行う総合住宅設備機器メーカーから提供されたコールセンターに寄せられる顧客の声(以下 VOC : Voice of Customer)データを対象にデータマイニング・テキストマイニング分析を行い、有益な情報を抽出することを目指している。コールセンターは日本に 3 拠点あり、1 ヶ月当り計約 11 万件の電話を受けている。問合せ内容は製品に対する苦情や保守サービス、製品詳細情報のリクエスト、カタログの請求など多様性に富む。電話での対応内容はソフトウェアによる音声認識とオペレーターの手入力によってテキストデータで保存されている。これらは膨大な非構造化データであり、その中から有益情報を抽出することが課題となっている。					
コールセンターの VOC データのテキストマイニング分析にはマーケティング、オペレーション改善、品質改良など特定の目的を持ってキーワードの頻出性、共起性などを分析することが行われているが、 <u>本研究ではそれらの分析を行う「前段階での効率的処理」を目指しており、有益情報の VOC データを汎用性高く多目的に利用できることを目的としている。</u>					
本研究では「売上貢献性」「ネガティブ性」「緊急性」という 3 つの印象評価軸を設定し、VOC データから有益情報を自動的に抽出する手法の構築・評価を行った。3 軸を組み合わせることで異なる特徴をもった VOC を膨大なデータの中から目的に合わせて抽出できると考える。					

目次

1. イントロダクション	4
1.1 研究背景	4
1.2 研究目的	4
1.3 問題意識	5
1.4 先行研究	5
2. VOC 印象判別アルゴリズム	7
2.1 VOC 問合せ内容の印象評価軸定義	7
2.2 VOC 判別アルゴリズム構築の概要	7
2.3 判別用評価辞書作成と文章スコア	8
2.4 判別アルゴリズム	10
3. データの概要	12
3.1 VOC データ概要	12
3.2 目視結果	14
4. 分析結果	17
4.1 判別用辞書	17
4.2 判別アルゴリズム	18
4.3 判別精度検証	20
5. 結論	23
5.1 まとめ	23
5.2 今後の課題	23
6. 謝辞	25
7. 参考文献	26
8. 参考資料	27
8.1 組み合わせ内容	27
8.2 R コマンド	30

1. イントロダクション

1.1 研究背景

21 世紀に入り、ICT 技術の向上、情報ネットワークインフラの普及、記憶装置・センサー装置などの低価格化により膨大なデータを取得・分析することが可能となった。今日では、「ビッグデータ」「人工知能」「IoT」のキーワードが新聞紙面でも頻出するようになり、データを経営に活かすことが大きな企業課題となっている。日本政府としてもビッグデータ活用を日本の新たな成長戦略として位置づけ、その推進に力を入れている。北米の株式公開企業 330 社を対象とした調査では、データ重視を自認する企業ほど、財務・営業両面の客観的指標に優れており、特にデータ重視度で業界の上位三分の一を占める企業は、平均すると競合他社を生産性で 5%、収益性で 6%上回っていたという調査結果もある[1]。

また、データと一口に言っても、活用が比較的に容易である POS データや顧客データなどの「構造化されたデータ」と電話・音声放送などの音声データや SNS・BLOG のテキストデータやセンサー・カメラなどの画像・動画データといった ICT の進展に伴い急激に増加している「非構造化データ」もあり、今後は非構造化データの有効活用も注目されている[2]。

1.2 研究目的

本研究では、産学共同プロジェクトを行う総合住宅設備機器メーカー(以下、X 社)から提供されたコールセンターに寄せられる顧客の声(以下 VOC : Voice of Customer)データを対象にデータマイニング・テキストマイニング分析を行い、有益な情報を抽出することを目指している。コールセンターは日本に 3 拠点あり、1 ヶ月当り計約 11 万件の電話を受けている。問合せ内容は製品に対する苦情や保守サービス、製品詳細情報のリクエスト、カタログの請求など多様性に富む。電話での対応内容はソフトウェアによる音声認識とオペレーターの手入力によってテキストデータで保存されている。これらは膨大な非構造化データであり、その中から有益情報を抽出することが課題となっている。

コールセンターの VOC データのテキストマイニング分析にはマーケティング、オペレーション改善、品質改良など特定の目的を持ってキーワードの頻出性、共起性などを分析することが行われているが[3][4][5]、本研究ではそれらの分析を行う「前段階での効率的処理」を目指しており、有益情報の VOC データを汎用性高く多目的に利用できることを目的としている。本研究では「売上貢献性」「ネガティ

ブ性」「緊急性」という 3 つの印象評価軸を設定し、VOC データから有益情報を自動的に抽出する手法の構築・評価を行った。3 軸を組み合わせることで異なる特徴をもった VOC を膨大なデータの中から目的に合わせて抽出し、マーケティング・品質管理など様々な部門の者が VOC データを有効活用することができるようになると考える。

1.3 問題意識

研究背景に記載した通りデータの経営への活用は近年、社会的にその重要性が認識されてきている。しかし、実務として膨大なデータを経営に生かす分析リテラシーをもった企業は非常に少数であり、データ分析事例として取上げられる企業も一部の先進的企業が繰り返し登壇しているのが現状である。KBS へもデータ分析の共同研究依頼が相次いでおり、リテラシーをもったデータサイエンティスト人材のニーズは高い。

データサイエンティストは、データを読み解き分析する「統計スキル」、実務としてデータを操る「IT スキル」、データを経営に生かす「経営スキル」という 3 軸の素養を持つ必要があると考える。さらに、非構造データの解析に当たっては、画像認識処理や自然言語処理など工学的専門知識も必要となってくる。

本研究を通じて著者自身が上記素養を身に着けることもあるが、一般的な各部門の担当者であってもデータ分析を行うことが可能な土台を構築することは社会的意義があると考ええる。

1.4 先行研究

「単語感情極性対応表」

高村ら(2006)[6][7]は、単語ごとに感情極性値を計算し、「単語感情極性対応表」(以下、高村式辞書)を作成した。感情極性とは、その単語に対して一般的に良い印象を持つか (positive) 悪い印象を持つか (negative) を表した二値属性である。感情極性値は positive と negative の二値属性を+1 から-1 の実数値へ割り当てを行っている。例えば、「良い」、「美しい」のような positive な極性を持つ単語が+1 に近く、「悪い」、「汚い」のような negative な極性を持つ単語が-1 に近くなる。高村式辞書の単語スコア例を Table 1.1 に示す。

Table 1.1 Sample of Word Scores in Takamura Dictionary

No	Term	Pos	Score
1	優れる	動詞	1.00
2	価値	名詞	0.89
3	凜と	副詞	0.53
4	目立つ	動詞	0.36
5	切捨て	名詞	-0.35
6	反逆	名詞	-0.79
7	悪い	形容詞	-1.00

「テキスト自動解析に基づく正負印象判別アルゴリズム」

曹(2015)[8]は高村式辞書と X 社特有の単語を合わせて極性判別用辞書の作成を行った。そして、コールセンターに寄せられる VOC データの問合せ内容に含まれる単語に対して、判別用辞書を用いて問合せ内容のスコア付けを行い、スコアをもとに各問合せを positive か negative かの判別を行った。

「テキストカテゴライゼーション」

その他、テキストの自動分類の手法としては上述の極性判別辞書を用いた文書スコア化以外にもサポートベクターマシーン(SVM)、ロジスティック回帰、決定木、ナイーブベイズ分類器など様々な手法が考案されている[9][10][11]。

「テキストからの特徴語抽出 TF-IDF 法」

文章の特徴を表す単語を抽出する手法には TF-IDF(Term Frequency - Inverse Document Frequency)法がある[12]。ある特定の文章中に繰り返し出現する単語は、その文章の特徴を示していると考えられる。しかし、助詞「は」「を」や主語を表す名詞「私」「あなた」などは普遍的に様々な文章に表れるため、ある文章中の出現回数が多くとも文章の特徴を示す単語としては不適である。

上記問題に対応する手法として、TF-IDF 法では単語に重みづけを行う。一般的に頻出される単語は文章の特徴を示す単語としては重みを小さくし、頻出度が低い単語は重みを大きくすることで、適切に文章の特徴を示す単語を抽出するものである。

2. VOC 印象判別アルゴリズム

2.1 VOC 問合せ内容の印象評価軸定義

曹(2015)[8]による先行研究では、コールセンターへの問合せ内容を positive か negative かの判別を行った。本研究では X 社含めて検討を行い、VOC データの問合せ内容を「売上貢献性」「ネガティブ性」「緊急性」という独立した 3 つの評価軸を設けて評価を行うことにした。Fig. 2.1 に評価軸の定義を示す。VOC の問合せ内容をそれぞれ 3 軸で評価を行い、3 軸の該当/非該当を組み合わせることで様々な目的に合わせて VOC データより有益な情報を多目的に汎用性高く効率的に抽出することを最終的に目的している。

「評価軸定義」	
『売上貢献性』(Sales Potential)	
: リフォームの検討、ショールーム・営業店の紹介、費用確認、修理依頼など将来の売上につながる可能性のある問合せ	
『ネガティブ性』(Negative Attitude)	
: クレーム、対応者・製品に関する不満・不具合、点検・修理など顧客がネガティブな印象をもつ可能性がある問合せ	
『緊急性』(Urgency)	
: 至急の仕様確認、本日中の回答要求など緊急に対応を要する可能性がある問合せ	

Fig. 2.1 Definition of Three Standards

2.2 VOC 判別アルゴリズム構築の概要

本研究の VOC 判別アルゴリズム構築の流れを Fig. 2.2 に示す。VOC データは件数が非常に多いため、目視評価用データセットをその中からランダムにピックアップを行う。VOC データの問合せ内容を人手による目視し 3 軸の定義に該当する(評価 1)か、非該当(評価 0)かを判別する。

目視評価した VOC データを学習用データ(以下、LD: Learning Data)と検証用データ(以下、TD: Test Data)に分割する(Fig. 2.3 参照)。分割した目視評価の内、LD をもとに評価辞書の作成を行う。そして、評価用辞書をもとに各 VOC データの問合せ内容をスコア化する。VOC スコアと LD を教師データとして判別アルゴリズムを構築する。また、TD は判別アルゴリズムの精度検証に用いる。

そして、最終的には構築した評価用辞書と判別アルゴリズムを用いて未知データに対して、3 軸それぞれ該当/非該当の判別を行うことを目指している。

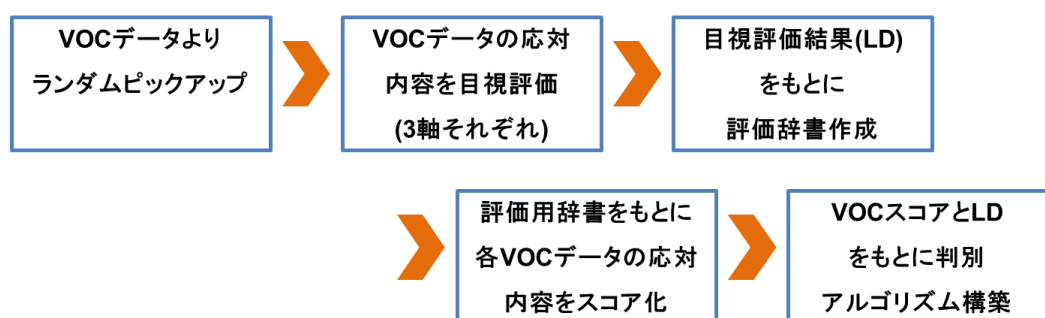


Fig. 2.2 Analytical Procedures in This Study

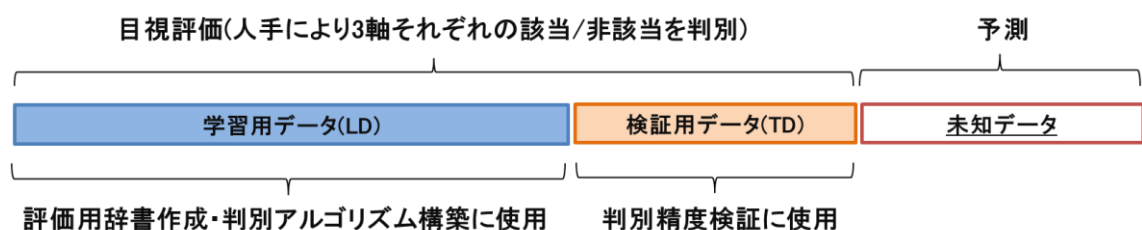


Fig. 2.3 Division of Data Set

2.3 判別用評価辞書作成と文章スコア

判別用評価辞書の作成の流れを説明する。LD の問合せ内容を対象に統計解析ソフトウェアによって『分かち書き』を行い「単語」「品詞」情報を抽出する。尚、統計解析ソフトウェアには R(バージョン 3.2.2)、分かち書きを行う形態素解析には MeCab(R の RMeCab ライブラリー、バージョン 0.99991)を使用した。分かち書きの例を Table 2.4 に示す。(分かち書きとは文章を単語に分割する処理)

Table 2.4 Sample of Separating Words in Japanese with Spaces

Term	「私」	「は」	「友達」	「へ」	「電話」	「を」	「掛け」	「た」	「。」
Pos	名詞	助詞	名詞	助詞	名詞	助詞	動詞	助動詞	記号

分かち書きを行った単語より、助詞や記号などの無意味単語を取り除き、判別用評価辞書として単語スコアを計算する。単語スコアの算出のため、単語毎に評価軸定義に該当する(評価 1)の割合： $r(w)$ を計算する。

$$r(w) = \frac{\text{評価 1 の VOC の数}}{\text{該当単語を含む VOC の数}}$$

その割合に応じて以下の採点基準をもとに単語スコア： $ws(w)$ へ「0」/「0.5」/「1.0」のいずれかの値を付与する。例えば $r(w)=0.2$ の場合は $ws(w)=0$ 、 $r(w)=0.55$ の場合は $ws(w)=0.5$ 、 $r(w)=0.79$ の場合は $ws(w)=1$ となる。

$0 \leq r(w) \leq \frac{1}{3}$	$ws(w) = 0$
$\frac{1}{3} < r(w) \leq \frac{2}{3}$	$ws(w) = 0.5$
$\frac{2}{3} < r(w) \leq 1$	$ws(w) = 1$

尚、単語スコアは3軸それぞれ計算を行う。単語スコアは Table. 2.5 のイメージのように3軸独立に算出される。

Table 2.5 Image of Word Scores

No	Term	Pos	Score		
			Salse Potential	Negative Attitude	Urgency
1	優れる	動詞	1.0	0.0	0.5
2	価値	名詞	0.5	0.0	0.5
3	凜と	副詞	0.5	0.0	0.0
4	目立つ	動詞	0.5	1.0	0.5
5	てらてら	副詞	0.0	0.5	0.5
6	悪い	形容詞	0.5	1.0	1.0

単語の文章出現頻度の逆数: IDF (Inverse Document Frequency)を以下の式より計算して求める。 IDF は単語の VOC 文章全体の頻出度が高くとなる値は小さくなり、頻出度が低いと小さくなる。単語スコアの重みとして使用する。

$$IDF = \log_2 \frac{\text{全学習データ(LD)の VOC 数}}{\text{該当単語を含む VOC の数}} + 1$$

各 VOC の各単語の出現頻度: TF (Term Frequency)を求める。

$$TF = \frac{\text{単語の出現回数}}{\text{対象 VOC に含まれる総辞書登録単語数}}$$

最後に、VOC の文章スコア: $vs(v)$ は以下によって計算する。

$$vs(v) = \sum ws(w) \times TF \times IDF$$

文章スコア: $vs(v)$ は、「単語スコア」×「単語の出現回数」×「単語の重み (IDF)」の合計を「計算対象の VOC に含まれる総辞書登録単語数」で割ったものとなる。文章が長くなると当然、合計した文章スコアは高くなるが、「対象 VOC に含まれる総辞書登録単語数」で割ることで文章の長さを正規化する形となる。

$$vs(v) = \frac{\sum ws(w) \times \text{単語の出現回数} \times IDF}{\text{対象 VOC に含まれる総辞書登録単語数}}$$

2.4 判別アルゴリズム

試験であれば 70 点以上なら合格、70 点未満なら不合格といったように、各 VOC の文章スコア: $vs(v)$ に対して該当/非該当を判断する「判別水準」が必要となる。適切な判別水準を得るため、ロジスティック回帰分析と Confusion Matrix 用いて検討を行う。

まず、「LD の目視結果(評価 0-1)」を教師データとして、各 VOC の「文章スコア: $vs(v)$ 」に対してロジスティック回帰分析を行う。その結果、文章スコアより「評価 1 となる確率: $p(x)$ 」という形に変換できる(ロジスティック回帰結果は以下の式で表され、0 から 1 の値をとる)。

$$p(x) = \frac{e^{\beta(x)}}{1 + e^{\beta(x)}} = \frac{1}{1 + e^{-\beta(x)}} \quad \beta^*(x) = \beta_0^* + \beta_1^* x$$

次に Confusion Matrix を用いて判別水準の検討を行う。Confusion Matrix および Recall(再現率)と Precision(適合率)の定義を以下に示す。Recall は真実(今回は目視結果)で評価 1 であったものが予測でも評価 1 と判別される割合、Precision は評価 1 と予想したものが真実評価 1 であった割合を意味している。

判別水準を変化させると Recall と Precision はそれぞれ変化する。判別水準を高く設定すると(例えば確率 90%以上を評価 1 の判別水準を設定すると)、Precision は高くなるが Recall の値は小さくなる。判別水準を低く設定すると Recall の値は

高くなるが **Precision** は小さくなるなど、**Recall** と **Precision** は相反関係にある。
 そのため、この両者のバランスをとりながら最適な判別水準の検討を行う。

	$I(v)$ 真実		Total
	0	1	
$I(v z)$ 0	$x_{00}(z)$	$x_{01}(z)$	$X_0(z)$
予測 1	$x_{10}(z)$	$x_{11}(z)$	$X_1(z)$
Total	$Y_0(z)$	$Y_1(z)$	$N(z)$

$$\text{Recall}(z) = \frac{x_{11}(z)}{Y_1(z)} \quad \text{Precision}(z) = \frac{x_{11}(z)}{X_1(z)}$$

3. データの概要

3.1 VOC データ概要

本章では本研究で使用するデータの概要について説明する。使用するデータは産学共同プロジェクトを行う総合住宅設備機器メーカーX社によって提供されたものであり、X社のコールセンター国内3拠点に寄せられる電話内容を保存したものである。電話音声は音声認識ソフトウェアとオペレーターによる手入力によってテキストデータとして保存されている。一ヶ月当たり約11万件の問い合わせ(VOC)があり、2014年4月から2015年3月のデータが提供された。問合せ件数の推移をTable 3.1に示す。本研究では2015年3月のデータを中心に研究を行った。2015年3月のVOCデータは、112,618件の問合せ情報がある。各VOCデータには、受付No、顧客分類、問合せ目的、商品区分、問合せ内容、対応オペレーターなど、計91の項目がある。本研究では、効果的にテキストマイニング分析を行うために重要な問合せと項目内容をX社とともに検討し、VOCデータの中から絞込みを行った。

電話を掛けてきた顧客タイプをTable 3.2に示す。VOCの中で最も大きな顧客タイプは「エンドユーザー」52.9%である。そして各地域で住宅設備の施工・販売を行っている「工務店・工事店」23.6%、「販売店・特約店・代理店」6.5%と続いている。それ以外の顧客タイプもX社と関連してビジネスを行っている企業であり、問合せ内容の多くは製品仕様や後継品情報に関するものが多い。エンドユーザーからの問合せが一般的な消費者からの問合せを表しており、本研究では一般消費者からの顧客の声を分析することを目的としており、そこにフォーカスする。

コールセンターへ電話を掛けてきた問合せ目的をTable 3.3に示す。大部分のVOCは「相談」目的であり80.8%を占めている。次に「修理依頼」5.9%、「部品依頼」4.5%が続いている。「相談」目的以外は電話の意図が明確であり、本研究では重要な「相談」にフォーカスした。

さらにどの商品に関する問合せが多いのかについて検討を行った。商品タイプ別の問合せをTable 3.4に示す。件数の最大のものは「トイレ」の29.3%であり、「キッチン」9.0%、「ユニットバスルーム」8.6%、「玄関ドア・引戸」6.5%、「洗面化粧室」6.1%、「窓・シャッター」6.2%と続いている。数多くの商品タイプの中で本研究では室内施設を構成する商品についてフォーカスを行った。そこで商品タイプとして「トイレ」「キッチン」「ユニットバスルーム」「洗面化粧室」「窓・シャッター」にフォーカスした。

以上のような分析対象の絞込みにより VOC の件数は 29,777 件となった。そこから音声認識エラー、機密情報、内容情報が僅かなものなど不完全な VOC を除外すると最終的に研究対象とする VOC データは 22,838 件となった。

Table 3.1 Trends in Calls

Month	No. of VOCs	Data Size(MByte)
2014年4月	110,620	138
2014年5月	107,308	136
2014年6月	113,569	145
2014年7月	117,855	154
2014年8月	100,726	135
2014年9月	112,533	151
2014年10月	114,904	154
2014年11月	103,973	136
2014年12月	107,646	137
2015年1月	102,615	128
2015年2月	101,692	125
2015年3月	112,618	129

Table 3.2 Composition of VOCs by Types of Callers

Types of Callers	No. of VOCs	%
エンドユーザー	59,531	52.9%
工務店・工事店	26,603	23.6%
販売店・特約店・代理店	7,295	6.5%
設計事務所	4,355	3.9%
ハウスメーカー	2,271	2.0%
量販店・ホームセンター	2,132	1.9%
リフォーム会社	1,391	1.2%
グループ会社(社内)	864	0.8%
同業他社	217	0.2%
ゼネコン	179	0.2%
その他	7,780	6.9%
合計	112,618	100.0%

Table 3.3 Composition of VOCs by Reasons for Calls

Reasons for Calls	No. of VOCs	%
相談	91,023	80.8%
修理依頼	6,603	5.9%
部品依頼	5,107	4.5%
その他依頼	2,109	1.9%
カタログ依頼	1,867	1.7%
苦情・不満	826	0.7%
お褒め・お礼	154	0.1%
その他	4,929	4.4%
合計	112,618	100.0%

Table 3.4 Composition of VOCs by Product Types of Concern

Product Types of Concern	No. of VOCs	%
トイレ	32,956	29.3%
キッチン	10,176	9.0%
ユニットバスルーム	9,655	8.6%
玄関ドア・引戸	7,336	6.5%
窓・シャッター	6,970	6.2%
洗面化粧室	6,908	6.1%
インテリア建材	3,134	2.8%
その他	35,483	31.5%
合計	112,618	100.0%

3.2 目視結果

人手により目視した VOC の問合せ内容の評価結果を説明する。3.1 節で絞込みを行った 22,837 件の VOC データは件数として非常に多いため、目視評価用にその中からランダムに 1,956 件のピックアップを行った。VOC データの問合せ内容を人手により目視し 3 軸の定義をもとに該当する(評価 1)か、非該当(評価 0)かを評価する。尚、判別は読者 3 名の多数決によって、該当(評価 1)/非該当(評価 0)の判別を行った。3 軸の判別結果を Table 3.5 に示す。売上貢献性は定義に該当する(評価 1)が 60.3%、それぞれネガティブ性 48.1%、緊急性 41.9%という割合となった。

また、3 軸の組み合わせを Table 3.6 および Fig. 3.7 に示す。組み合わせは「売上貢献性」、「ネガティブ性」、「緊急性」の順に並んでいる。例えば 1,0,1 であれば、売上貢献性(評価 1)、ネガティブ性(評価 0)、緊急性(評価 1)である。3 軸の組み合わせの典型的な問合せ内容を Table 3.8 に示す。組み合わせにより異なる内容の

VOC があることが分かる。また、「緊急性」は問合せ内容の特徴そのものは表していないが、各問合せ内容の重要度の高低を表す傾向があると思われる。将来はこれらを組み合わせることで、例えばマーケティングの部署であれば「組み合わせ 1,0,1」の VOC のみを抽出し、その内容を精査・分析することで営業活動の改善に活かすことができると考えられる。

Table 3.5 Results of the Evaluation

		0(Not Satisfied)	1(Satisfied)	Sum
Sales Potential	件数	776	1,180	1,956
	割合	39.7%	60.3%	100.0%
Negative Attitude	件数	1015	941	1,956
	割合	51.9%	48.1%	100.0%
Urgency	件数	1137	819	1,956
	割合	58.1%	41.9%	100.0%

Table 3.6 Combination of Sales Potential, Negative Attitude and Urgency

Combination	1,1,1	1,1,0	1,0,1	1,0,0	0,1,1	0,1,0	0,0,1	0,0,0	Sum
件数	473	202	131	375	127	140	88	420	1,956
割合	24.2%	10.3%	6.7%	19.2%	6.5%	7.2%	4.5%	21.5%	100.0%

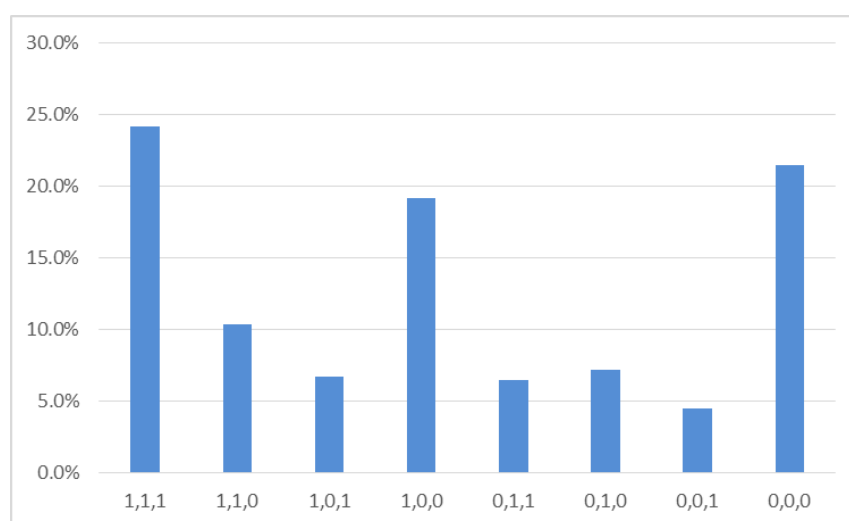


Fig 3.7 Combination of Sales Potential, Negative Attitude and Urgency

Table 3.8 Typical Examples

Combination	Reasons for Calls
1,1,1 / 1,1,0	定期点検ランプ点滅、修理依頼、製品不具合(経年後)
1,0,1 / 1,0,0	購入前相談、営業店紹介、カタログ請求、部品購入、製品仕様確認
0,1,1 / 0,1,0	製品不具合(設置直後)、水漏れ時など使用方法確認、お手入れ方法確認
0,0,1 / 0,0,0	内容不明瞭、軽微な各種問い合わせ

4. 分析結果

4.1 判別用辞書

判別用辞書の単語絞り込みと単語スコアについて説明する。まず、3.2 節で目視を行った 1,956 件より学習用データ(LD)として 1,304 件を抽出する。学習用データ:LD 1,304 件に対して、問合せ内容の分かち書き処理を行い、単語と品詞情報を抽出した。その結果、合計 10,634 単語が抽出された。その中で判別に意味ある品詞として「名詞」「形容詞」「動詞」「副詞」のみに絞り込みを行った。また、あまりに出現頻度の低い単語は辞書として不適切と考え出現頻度 3 件以上のみを抽出し、「名詞」「形容詞」「動詞」「副詞」と出現頻度 3 件以上という条件により 2,564 単語に絞り込みされた(尚、1 つの VOC に同じ単語が複数回含まれていたとしても 1 件とカウントする。1,304 件の VOC の中で 3 件以上の VOC に該当単語が出現しているという意味である)。そこから、内容を精査し「http、TEL、@@@」など判別に意味をなさないとされる単語や記号・符合と思われるものを除外し、最終的に 1,647 単語に絞り込みを行った。単語の絞り込みの流れを Fig. 4.1 に示す。

次に LD の目視結果(評価 0-1)をもとにその単語が含まれている VOC の評価 1 とする割合: $r(w)$ を求め、そこから採点基準にもとづき単語スコア: $ws(w)$ に「1.0」「0.5」「0.0」を付与した。そして、出現頻度の単語スコアの重み:IDF の算出を行った。単語スコア例を Table 4.2 に示す(評価軸「売上貢献性」 - 品詞「形容詞」の単語スコア例)。尚、形容詞「酷い」「悪い」は一見すると売上貢献性に関連がないものとも思えるが、修理依頼・点検に関する問合せなども製品の買換えなどに寄与すると考え、当該単語が含まれる VOC が売上貢献性あり(評価 1)と目視判別されたことによるものと考えられる。

名詞	5,583		名詞	2,203		名詞	1,286
形容詞	97		形容詞	48		形容詞	48
動詞	627		動詞	262		動詞	262
副詞	147		副詞	51		副詞	51
その他	4,180		合計	2,564		合計	1,647
合計	10,634						

Fig 4.1 Procedure of Choosing Useful Words for Creating the Dictionary

Table 4.2 Sample of Word Scores for Sales Potential

Adjective	$r(w)$	$ws(w)$	No. of Appearances in LD	IDF
すごい	1.000	1.0	3	9.764
暖かい	0.750	1.0	4	9.349
酷い	0.800	1.0	5	9.027
多い	0.364	0.5	11	7.889
悪い	0.633	0.5	30	6.442
宜しい	0.569	0.5	51	5.676

・単語スコア算出例:形容詞「暖かい」

学習データ(LD)1,304件のVOCの内、4件のVOCに該当単語「暖かい」を含む(内、評価1は3件のVOC)

$$r(w) = \frac{3}{4} \quad \text{採点基準に基づき} \quad ws(w) = 1 \quad IDF = \log_2 \frac{1304}{4} + 1 = 9.349$$

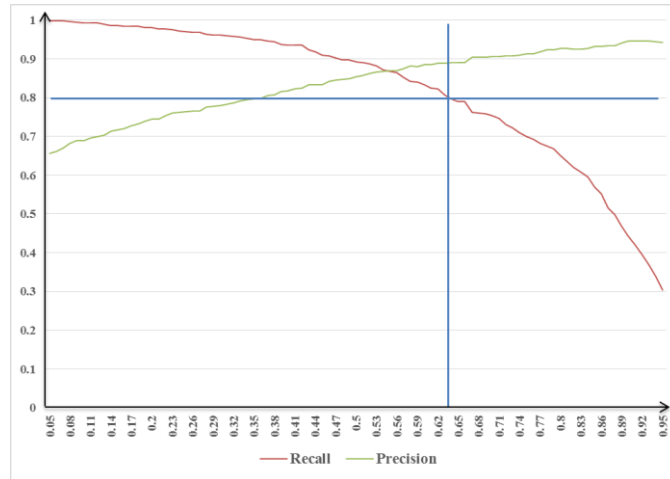
4.2 判別アルゴリズム

4.1 節で算出された単語スコア: $ws(w)$ および単語重み: IDF を用いて LD1,304 件の VOC の VOC スコアの算出を行った。そして「VOC スコア」と教師データとなる「LD の目視結果(評価 0-1)」をもとにして、ロジスティック回帰分析を行った。回帰係数 β_0^* と β_1^* の計算結果を Table 4.3 に示す(3 軸の各係数 β_0^* と β_1^* の有意水準はいずれも 0.001)。

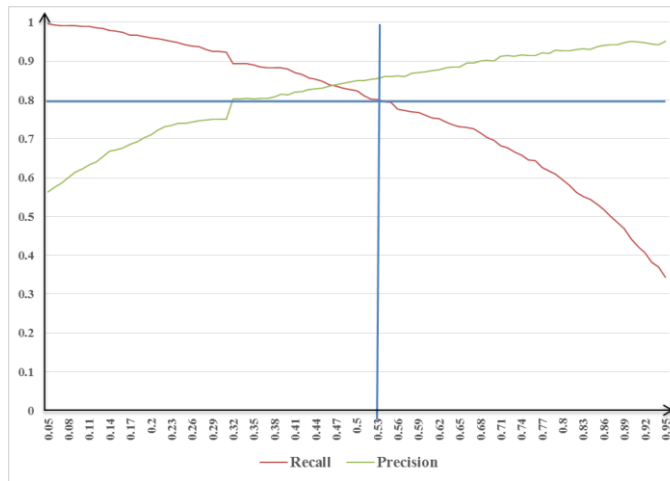
Table 4.3 Estimated Coefficients (with confidence level at 0.001)

	β_0^*		β_1^*	
	value	z - value	value	z - value
Sales Potential	-14.468	-17.70	20.225	18.17
Negative Attitude	-10.473	-18.58	19.430	18.51
Urgency	-9.753	-17.29	17.489	17.25

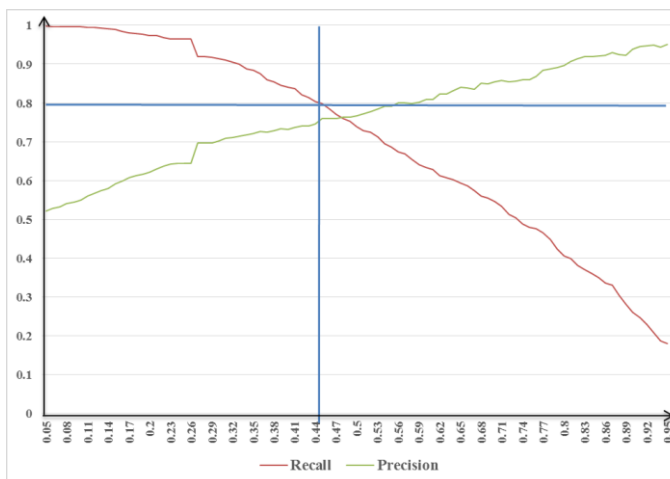
次に評価 0 と評価 1 の区切りとなる適切な判別水準: z^* を検討する。LD のロジスティック回帰結果について、Confusion Matrix を作成し、判別水準: z を変化させたときの Recall(再現率)と Precision(適合率)の推移の描写を行った。Recall と Precision は相反関係にあり、両者のバランスをとる必要がある。本検討では「Recall=0.8 以上」という条件下で「Precision が最大」となる判別水準を最適な判別水準: z^* と定めた。Recall と Precision の推移を Fig. 4.4 に示す。Fig. 4.4 より最適な判別水準: z^* は 3 軸それぞれ、a)売上貢献性: $z^*=0.63$ 、b)ネガティブ性: $z^*=0.53$ 、c)緊急性: $z^*=0.44$ と定めた。



a) Sales Potential : $z^* = 0.63$



b) Negative Attitude : $z^* = 0.53$



c) Urgency : $z^* = 0.44$

Fig. 4.4 Recall and Precision as a Function of Segmentation Level z

4.3 判別精度検証

3.2 節で目視を行った 1,956 件より学習用データ(LD)1,304 件を除いた、残り 652 件の VOC を検証用データ(TD)として用いる。4.2 節で構築を行った判別アルゴリズム(「判別用辞書」「ロジスティック回帰係数」「判別水準: z^* 」)を用いて、652 件の TD に関して評価(0-1)の判別予測を行った。

売上貢献性の Confusion Matrix を Table 4.5 に示す。判別アルゴリズム判別予測: $I(v|z)$ と目視結果: $I(v)$ の組み合わせを示している。Table 4.5 a)および b)は $z^*=0.63$ のときの LD および TD それぞれの Confusion Matrix である。LD の Recall=0.806、Precision=0.889 は、TD の Recall=0.771、Precision=0.755 よりも判別精度が高いが、判別アルゴリズムは LD をもとに構築されているため当然である。しかし、TD の Recall=0.771、Precision=0.755 の値も十分高いと考えられる。例えば、ランダムに VOC をピックアップしそれを評価 1 と予測した場合の正答率は $380/652=0.583$ となり、判別アルゴリズムによって予測を行った場合の正答率は $293/388=0.755$ となる。

また、3.1 節で絞込みを行った 22,837 件の VOC データから目視対象 1,956 件を除いた 20,881 件より、4,000 件の VOC を未知データとしてランダムに抽出を行った。そして、未知データ 4,000 件について判別アルゴリズムにより判別予測を行った。「判別予想の評価 1 の割合」と「目視を行った 1,956 件の評価 1 の割合」を比較したものを Table 4.5 c)に示す。判別予測: $I(v|z)=56.8\%$ 、目視結果: $I(v)=60.3\%$ と判別する割合に大きな乖離はなかった。(参考に各データの件数を Fig. 4.8 に示す。)

同様にネガティブ性の Confusion Matrix と割合比較を Table 4.6 に示す。TD において Recall=0.836、Precision=0.749 と良い予測結果となった。また、未知データに対する判別予測: $I(v|z)=42.6\%$ 、目視結果: $I(v)=48.1\%$ と判別割合にも大きな乖離はなかった。

同様に緊急性の Confusion Matrix と割合比較を Table 4.7 に示す。TD において Recall=0.785、Precision=0.407 となり、良い予測結果とはならなかった。また、未知データに対する判別予測: $I(v|z)=51.2\%$ 、目視結果: $I(v)=41.9\%$ と判別割合にも若干の乖離があった。これは、緊急性に関しては表面的な文章に現れる単語だけでなく、問合せ内容の「文脈まで読み取り」判断する必要があるのではないかと考察する。また、テキスト情報のみで緊急性を読み解くことそのものも難度が高い行為であるとも考えられる。緊急性に関しては今後検討が必要であると考ええる。

Table 4.5 Confusion Matrices for VOCs with Sales Potential

a) Confusion Matrix with LD

<i>LD</i>	<i>I (v)</i>		Total	Precision
	0	1		
<i>I (v z)</i>	422	155	577	
	81	646	727	0.889
Total	503	801	1304	
Recall		0.806		

b) Confusion Matrix with TD

<i>TD</i>	<i>I (v)</i>		Total	Precision
	0	1		
<i>I (v z)</i>	177	87	264	
	95	293	388	0.755
Total	272	380	652	
Recall		0.771		

c) Portion of VOCs with 1: $I(v|z)$ vs $I(v)$

$I (v \setminus z)$	56.8%
$I (v)$	60.3%

Table 4.6 Confusion Matrices for VOCs with Negative Attitude

a) Confusion Matrix with LD

<i>LD</i>	<i>I (v)</i>		Total	Precision
	0	1		
<i>I (v z)</i>	606	123	729	
	83	492	575	0.857
Total	689	615	1304	
Recall		0.800		

b) Confusion Matrix with TD

<i>TD</i>	<i>I (v)</i>		Total	Precision
	0	1		
<i>I (v z)</i>	277	82	359	
	48	245	293	0.749
Total	325	327	652	
Recall		0.836		

c) Portion of VOCs with 1: $I(v|z)$ vs $I(v)$

$I (v \setminus z)$	42.6%
$I (v)$	48.1%

Table 4.7 Confusion Matrices for VOCs with Urgency

a) Confusion Matrix with LD

LD	$I(v)$		Total	Precision
	0	1		
$I(v z)$	491	125	616	
	175	513	688	0.746
Total	666	638	1304	
Recall		0.804		

b) Confusion Matrix with TD

TD	$I(v)$		Total	Precision
	0	1		
$I(v z)$	264	39	303	
	207	142	349	0.407
Total	471	181	652	
Recall		0.785		

c) Portion of VOCs with 1: $I(v|z)$ vs $I(v)$

$I(v \setminus z)$	51.2%
$I(v)$	41.9%

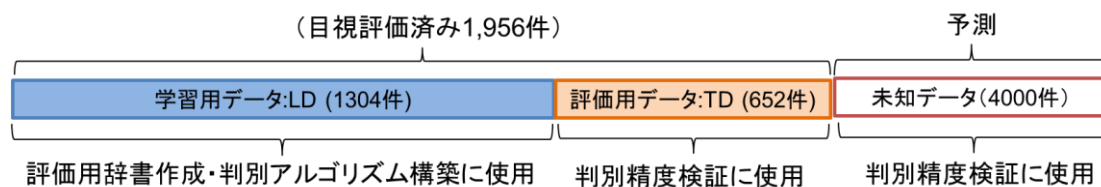


Fig. 4.8 No. of Data Set

5. 結論

5.1 まとめ

本研究では、産学共同プロジェクトを行う総合住宅設備機器メーカーX社から提供されたコールセンターに寄せられる顧客の声(VOC)に対して、データマイニング・テキストマイニング分析を用いて判別アルゴリズムの構築・評価を行った。

「売上貢献性」「ネガティブ性」「緊急性」の軸を設定し、これらの3軸を組み合わせることで汎用性高く多目的にVOCデータを抽出する土台ができると考える。

例えばマーケティングの部署であれば「売上貢献性(評価1)、ネガティブ性(評価0)、緊急性(評価1)」と判別予測されたVOCを抽出することで、その中に含まれていると思われる「購入前相談・カタログ請求・営業店紹介」の問合せ情報から営業活動の改善に生かすこと。コールセンター部署では、「売上貢献性(評価0)、ネガティブ性(評価0)、緊急性(評価0)」と判別予測されたVOCを抽出することで、顧客から寄せられる件数の多い「軽微な問合せ」情報を抽出し、HPや説明書にその対応方法を記載することで寄せられる電話件数の数を減少させることなど、構築した判別アルゴリズムは多目的な用途が今後考えられる。

5.2 今後の課題

判別精度検証では「売上貢献性(検証用データにおける Precision=0.755)」「ネガティブ性(Precision=0.749)」の判別精度は高くなったが、「緊急性(Precision=0.407)」に関しては良い精度が得られず、今後の検討が必要である。また、軸ごとに判別辞書に採用する単語を精査することで判別精度の向上が得られると考えられる。ただ、採用する単語の絞りこみを行うと汎用性が低くなると考えられるため、トレードオフ含め検討が必要である。併せて季節によっては問合せ内容のトレンドが異なると考えられる。本研究では行えなかったが季節性を考慮した判別アルゴリズムの調整なども必要になると考える。

本研究ではX社含め3つ軸の設定を行ったが、今後何らかの評価軸を追加することでさらなる汎用性・多目的対応が可能になると考える。(例えば、感情表現を網羅した感情表現辞典[13]には驚き・悲しみ・喜びなど様々な表現が網羅されている。この辞典の「驚き」表現に着目してVOCより情報抽出するなども考えられる。驚きがあるということは、エンドユーザーが何らかの商品に対する深い感動もしくは大きなネガティブ体験が起きたということであり、そのVOCには興味深い内容が含まれている可能性がある。)

また現在は、ディープラーニング技術をテキスト解析に用いる手法も検討されており、今後のテキストマイニング技術の発展が予想される。今後は文章の単語レベルの意味理解ではなく文章全体の意味解釈まで考慮した内容にしていきたい。今後の技術発展を取り込み研究を発展させていきたいと考える。

6. 謝辞

本論文の作成にあたり熱心にご指導いただきました林 高樹教授、住田 潮特任教授、岡田 正大教授、中村 洋教授、岩本 隆特任教授に深謝いたします。また、ゼミの活動を通じて多くの意見や示唆をくださった林ゼミ同期・先輩の皆様、学校生活を支えてくれた KBS の同期や両親にも深く感謝いたします。また最後に、共同研究の場を提供いただいた X 社担当者の皆様にも深く感謝いたします。

7. 参考文献

- [1] アンドリュー・マカフィー;エリック・ブリニョルフソン;有賀裕子 訳.「ビッグデータで経営はどう変わるか」 DIAMOND ハーバードビジネスレビュー 2013 年 2 月号, pp41-53, ダイヤモンド社, 2013.
- [2] 総務省 平成 25 年度情報通信白書 第 3 節「ビッグデータの活用が促す成長の可能性」 2013.
- [3] 中川純一; 橋本久; 中島大明. 企業を取り巻く声活用の新たな展開: 集まる声から, 集める声, 仕掛ける声へ (特集> 声・質的データを経営・マーケティングに活かす). オペレーションズ・リサーチ: 経営の科学, 2013, 58.8: 455-461.
- [4] 輪島幸治; 小河誠巳; 古川利博. テキスト評価分析を用いたヘルプデスク効率化手法の提案. 経営情報学会 全国研究発表大会要旨集, 2013, 2013.0: 141-144.
- [5] 那須川哲哉. コールセンターにおけるテキストマイニング (< 特集> 「テキストマイニング」). 人工知能学会誌, 2001, 16.2: 219-225.
- [6] 高村大也, 乾孝司, 奥村学. スピンモデルによる単語の感情極性抽出. 情報処理学会論文誌ジャーナル, Vol.47, No.02 , pp.627-637, 2006.
- [7] 高村大地. 「単語感情極性対応表」
http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html (2016 年 1 月現在)
- [8] 曹睿. テキスト自動解析に基づく正負印象判別アルゴリズムの開発とその効果の実証的研究 筑波大学大学院博士課程 システム情報工学研究科修士論文, 2015.
- [9] 高橋和子; 高村大也; 奥村学. 複数の分類スコアを用いたクラス所属確率の推定. 自然言語処理, 2008, 15.2: 3-38.
- [10] 乾孝司; 奥村学. テキストを対象とした評価情報の分析に関する研究動向. 自然言語処理, 2006, 13.3: 201-241.
- [11] F.Sebastiani. "Machine learning in automated text categorization." ACM computing surveys (CSUR) 34.1, pp.1-47, 2002.
- [12] 徳永健伸. 「情報検索と言語処理」 東京大学出版会, pp.27-28, 1999.
- [13] 中村明(編). 「感情表現辞典」 東京堂出版, 1993.

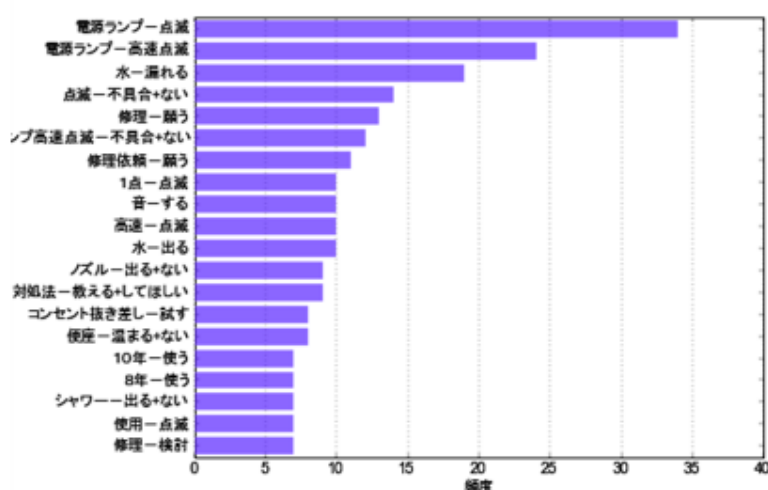
8. 参考資料

8.1 組み合わせ内容

目視結果の組み合わせ毎の問合せ内容について、その傾向の分析を行った。参考として組み合わせ(111/110)と組み合わせ(011/010)の「単語の掛かり受け頻度」「単語ネットワーク図」「クラスター」を以下に示す。各組み合わせともまったく異なる特性をもっていることが分かる。

■組み合わせ(110/110)

組合せ(111/110)-①掛かり受け頻度

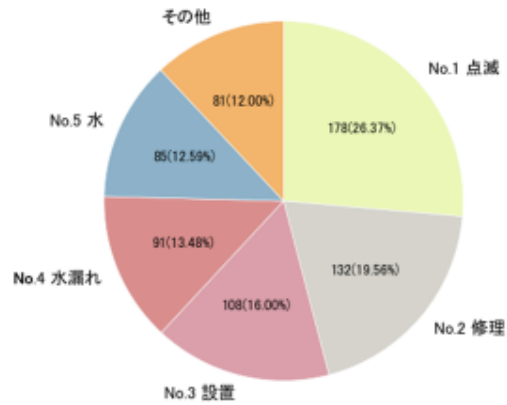


65

組合せ(111/110)-②ネットワーク図



組合せ(111/110)-③文章クラスタ分類



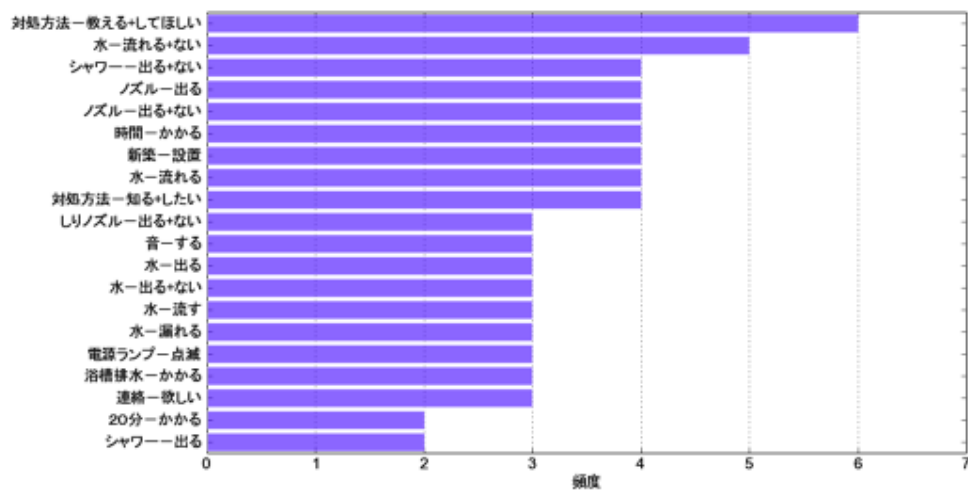
クラスタID	代表語	VOC数
No.1 点滅	点滅 製造番号 電球ランプ 改善+ない 製番	178
No.2 修理	修理 便座 品番 連絡 確認+したい	132
No.3 設置	設置 部品 交換+したい 交換 破損	108
No.4 水漏れ	水漏れ 使用 修理+してほしい	91
No.5 水	水 便器	85
その他	代表語がありません	81

※k-meansクラスタリング
※指定クラスタ数5
※N=675

67

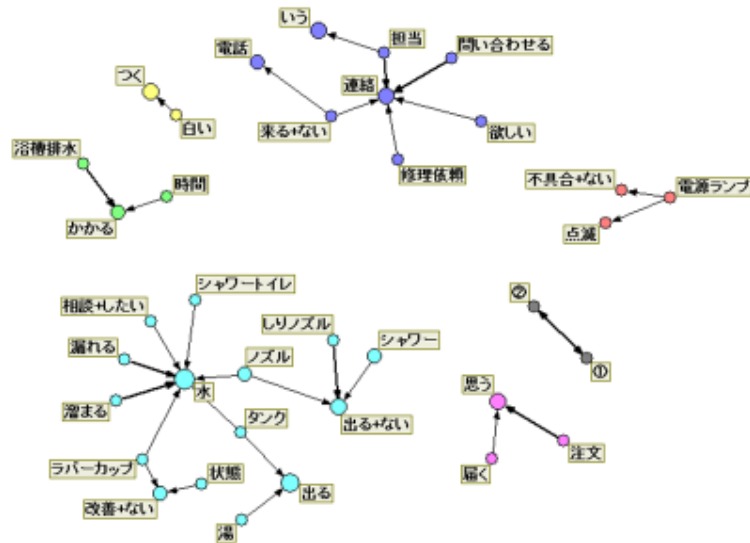
■組み合わせ(011/010)

組合せ(011/010)-①掛かり受け頻度

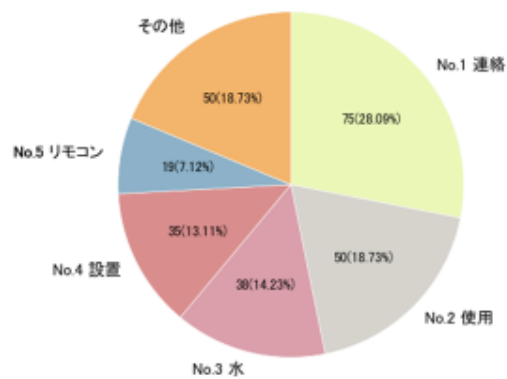


71

組合せ(011/010)-②ネットワーク図



組合せ(011/010)-③文章クラスタ分類



クラスID	代表語	VOC数
No.1 連絡	連絡 便座 電話 修理 対処方法	75
No.2 使用	使用 品番 原因 掃除 排水	50
No.3 水	水 ノズル	38
No.4 設置	設置 新築	35
No.5 リモコン	リモコン ボタン 音 平成	19
その他	代表語がありません	50

※k-meansクラスタリング
※指定クラス数5
※N=267

8.2 R コマンド

参考にロジスティック回帰の R コマンドと実行結果を示す。

#売上貢献性

```
> resT <- read.csv("~/txt/keisan/resT-add3-4.csv")
```

```
> View(resT)
```

```
> s1<-subset(resT,resT$S==1)
```

```
> s0<-subset(resT,resT$S==0)
```

```
lm(S~S.S.point,temp)
```

```
glm(S~S.S.point,temp,family = binomial)
```

```
s.glm<-glm(S~S.S.point,temp,family = binomial)
```

```
train.glm<-glm(S~S.S.point,temp[1:1304,],family = binomial)
```

```
#-----
```

```
> train.glm
```

```
Call: glm(formula = train.temp$S ~ train.temp$S.S.point, family = binomial,  
data = train.temp)
```

Coefficients:

(Intercept)	train.temp\$S.S.point
-14.47	20.23

Degrees of Freedom: 1303 Total (i.e. Null); 1302 Residual

Null Deviance: 1739

Residual Deviance: 1024 AIC: 1028

```
> summary(train.glm)
```

Call:

```
glm(formula = train.temp$S ~ train.temp$S.S.point, family = binomial,  
data = train.temp)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.3943	-0.5463	0.2338	0.5860	2.9556

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-14.4677	0.8175	-17.70	<2e-16 ***
train.temp\$S.S.point	20.2253	1.1131	18.17	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1739.0 on 1303 degrees of freedom

Residual deviance: 1023.9 on 1302 degrees of freedom

AIC: 1027.9

Number of Fisher Scoring iterations: 5