

営業日報のテキストマイニング

～ お客様視点での新たな期待探索の試み～

石原聖子

富士ゼロックス株式会社 総合企画部

1. はじめに

企業の業務を通じて収集・蓄積されたテキストデータは、質問紙調査で収集されたテキストデータと比べ、マイニングに難しさがある。その理由として業務データはメーカーのコールセンター情報であればお客様とのコミュニケーションに、新聞記事DBであれば日々のニュース発信に活用することを第一目的として収集されており、あらかじめ分析者の仮説や問題意識を組み込んでデータ収集が可能なアドホック調査とは成り立ちが異なるからである。

表1 テキストマイニングに活用するデータ種類と特性

データの種類	データ収集方法	データ量	データの代表性	収集コスト
蓄積データ (コールセンター情報、営業日報など)	■一次目的で活用(クレーム・問い合わせ対応など) ■一次のデータの二次的活用になる	■時系列に沿って、大規模データの収集が可能	■母集団からの確率標本ではないが ■「自社の顧客」という視点では母集団そのものという場合もある	■二次活用で使用する場合は安価にデータ収集が可能
調査データ (アンケート調査、インタビュー調査など)	■一次データとして新規収集が可能 ■分析目的に適した変数を収集できる	■調査規模による	■無作為抽出 ■有意抽出 上記の双方が可能	■新規収集コストが発生

富士ゼロックスでは営業員がお客様企業を訪問し収集してきた要望や苦情に迅速かつ適切に対処できるように「Voice of Customer (VOC)」という年間18万件以上のテキスト情報が入るデータベースを構築している。

これらのテキスト情報の一次活用は個々のお客様に対する営業部門や保守サービス部門の活動をサポートしており、二次的な活用として商品開発部門への開発要望、月次・年間の情報分析などに活用している。しかしながらこの大規模なデータベースの内容は多岐に渡り、特に

新たなビジネステーマの探索など、少数意見ではあるが、「今後の期待」を感じさせる情報をいかに探索するかが課題となっている。

本研究報告ではこうした新たなビジネステーマに関わる話題を、効率的に抽出するための試みを報告する。

2. 分析方法

2004年1月～12月末までにVOCに書き込まれた営業日報18万8,576件を解析対象とした。本データベースは営業が書いた「お客様の声」以外に、表2に示した入力者が選択した声の分類コードを持っている。

表2 営業日報188,576件の声の内容分類

声の分類		2004年
1	お客様の関心/期待すること/経営課題/満足/クレーム	48,136
2	お客様の現状	52,147
3	社内連絡事項	82,340
4	その他	52,170
5	未選択	129
合計(複数選択可の延べ件数)		234,922

このうち今回は社内連絡や、お客様の現状報告などを除き、分析目的である今後の期待や変化に関連するカテゴリー「お客様の関心、期待すること、経営課題、満足、クレーム」48,136件に絞って分析対象とし、実際にはここからサンプル抽出したデータ4,550件を用いた。

まず単語単位の頻出ランキングを出し、同義語・類似語をとりまとめ、統一を行った。具体的な分析には野村総合研究所のTrue Teller、およびSPSSのClementineを利用した。

3 分析結果

主な結果は下記の通り。その他の詳細は口頭報告する。

頻出名詞の除外による特徴強調

単語や、名詞・形容詞・動詞の係り受けを、出現頻度の高さ = 意見の特殊性の欠如とみなして分析対象から除外した。

表3はデータ全体に含まれる全ての単語を用いて算出した構文(係り受け)一覧と、上位30個の名詞を除外して算出した構文一覧である。左表に含まれている「機械 - 使う」「機械 - 新しい」など出現頻度の高い係り受けを含むテキストデータは右の表では除外され、よりお客様の今後の期待や変化が抽出しやすくなること分かる。

表3 全ての単語を用いた一覧(左)と上位30個の名詞を除外して算出した一覧(右)

全ての単語を使用した構文ランキング					上位30の名詞を除外した構文ランキング				
No.	名詞	形容詞・動詞	タイプ	テキスト件数	No.	名詞	形容詞・動詞	タイプ	テキスト件数
1	機械	使う	名・動	101	1	興味	ある	名・動	12
2	機械	新しい	名・形	74	2	スキャナー	使う	名・動	9
3	調子	悪い	名・形	69	3	表示	出る	名・動	8
4	興味	ある	名・動	58	4	カレンダー	持つ	名・動	7
5	プリンター	使う	名・動	46	5	発生	伺う	名・動	7
6	機械	ある	名・動	44	6	これから	お願いする	名・動	6
7	今	使う	名・動	44	7	依頼	ある	名・動	6
8	機械	入れる	名・動	43	8	調子	悪い	名・形	6
9	機械	交換する	名・動	42	9	セミナー	ある	名・動	5
10	話	聞く	名・動	41	10	可能性	ある	名・動	5
11	必要	ある	名・動	38	11	言葉	頂く	名・動	5
12	修理	来る	名・動	37	12	時間	かかる	名・動	5
13	言葉	頂く	名・動	36	13	色々	教える	名・動	5
14	料金	かかる	名・動	36	14	必要	ある	名・動	5
15	機械	古い	名・形	34	15	DW	使う	名・動	4

テキスト属性を活用したキーワード抽出

次にテキストの属性を活用したキーワード抽出を試みた。

この抽出には分析に使用したサンプルデータの中でも比較的件数の多い、関心事、満足、期待の3属性についてキーワードを抽出した。これはそれぞれのグループの特徴をよく表すキーワード(全体に比べて当該グループに偏って出てくる単語や係り受け)をランキング形式で表示するものである。表4は「ゼロックスに対する満足」と、「ゼロックスに対する期待」

のキーワード抽出結果である。スコア値の定義は1に近づくほど特徴的、0に近づくほど他グループでも平均的に出現する特徴のない単語となる。スコア値の算出式は、

$$\pm (ad-bc)2 / ((a+b) * (c+d) * (a+c) * (b+d))$$

a=係り受けの出現件数、 b=非出現件数、 c=グループ外出現件数、 d=グループ外非出現件数

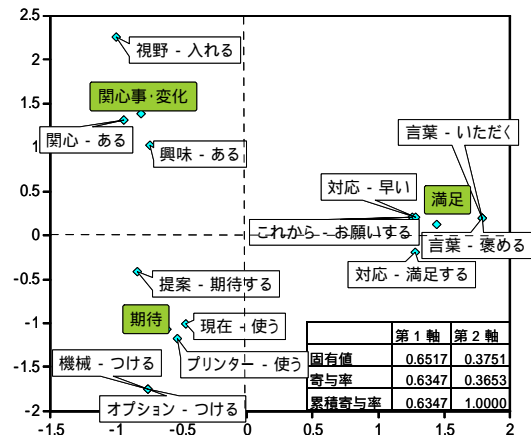
「まだそれほど件数が多いが特徴的」なテキストデータを探索するにはこのスコア値が利用できると思われる。

また、件数データをコレスポネンス分析してグループ特性の違いを表現したのが図1である。

表4 テキスト属性グループによるキーワード抽出

No.	ゼロックスへの満足	スコア	件数	ゼロックスへの期待	スコア	件数
1	これから お願いする	0.0119	16	プリンター 使う	0.0047	25
2	言葉 褒める	0.0113	10	機械 つける	0.0044	7
3	対応 早い	0.0112	17	提案 期待する	0.0037	6
4	言葉 いただく	0.0086	8	現在 使う	0.0032	18
5	対応 満足する	0.0075	8	オプション つける	0.0029	6
6	対応 助かる	0.0072	7	シール 作る	0.0029	6
7	言葉 頂く	0.0071	16	ゼロックス やる	0.0028	7
8	機械 満足する	0.0059	6	機械 使う	0.0025	39
9	早い 助かる	0.0059	6	欲しい もらう	0.0025	4
10	便利だ 使う	0.0057	5	相談 乗る	0.0023	7

図1 テキスト属性グループによるコレスポネンス分析



[謝辞]本研究には首都大経営学系の朝野熙彦先生にご助言いただいた。御礼申し上げます。

[参考文献]

上田・黒岩・戸谷・豊田(2005)「テキストマイニングによるマーケティング調査」講談社

連絡先:

富士ゼロックス株式会社 総合企画部マーケティング・リサーチG
seiko.ishihara@fujixerox.co.jp