

## 顧客の声からの低頻度重要情報の抽出 Extraction of Low-Frequency Information from Voice of Customer

森本 康嗣† 内部 こなぎ†  
Yasutsugu Morimoto Konagi Uchibe

### 1. まえがき

企業のコールセンタには、顧客からの問合せや意見要望が電話やメールにより日々多数寄せられ、テキストデータとして蓄積されている。このデータを分析し、顧客のニーズ、製品の使い勝手、不具合の発生状況などを把握することで、マーケティングやサービス向上などに役立てたいというニーズは大きい。そのため、自然言語処理やテキストマイニングを応用することで、顧客の声を分析する試みが数多く行われている。

しかしながら、主に統計的なアプローチで抽出される情報は、その分野の専門家にとっては当たり前であることが多く、コストを掛けて分析するのに値する情報を抽出することは難しい。そこで、本報告では必ずしも高頻度ではないが、価値がある情報を抽出する方法の検討を試みた結果について報告する。提案手法では、問い合わせデータの特徴である、「質問」と「回答」の組み合わせである点に着目し、「質問」と「回答」それぞれに出現する言語表現の組み合わせによって価値がある情報を抽出する。

### 2. 従来技術

#### (1) 統計的なアプローチ

機械学習を含む統計的なアプローチでは、情報の中身ではなく、頻度が高い情報、時間的な変化(急増・急減)、発生の偏り等に着目して価値がある情報を抽出する。統計的に有意である情報を抽出するためには、ある程度の頻度がある情報であることが必要なため、専門家や現場の人間にとっては新味がない情報であることが多い。鈴木らによれば、発見される知見の妥当性と興味深さのトレードオフが発生することが指摘されている[1]。同様の現象は情報推薦技術でも指摘されており、推薦内容の正確さだけでなく、情報の意外性、目新しさを示すセレンディピティ(senendipity)を向上させるための提案が行われている[2]。

#### (2) 言語的なアプローチ

テキストの内容に基づいて、重要な情報を抽出するアプローチである。典型的な方法として、評価極性(ポジティブ/ネガティブ)を表す評価表現を利用するアプローチが知られている[3]。Web テキストを用いたマーケティングでは、非常に有用な手法であるが、コールセンタに集まる情報は、概ねネガティブな情報であることが多いため、適用が困難である。

評価極性より詳細な分析を行う技術としては、意見中の「要望」、「現状認識」、「不満」等を表す部分を抽出する技術が存在する[4]。本報告の提案は、このタイプの技術に属する。

### 3. 提案方法

#### 3.1 抽出対象とアプローチ

問い合わせデータには、要望、苦情、質問、依頼等の情報が混在している。本研究では、弊社のコールセンタ運営部署にヒヤリングを行い、「新製品・新機能につながる情報」(以下、顧客ニーズ)を抽出対象として設定した。

顧客ニーズを抽出する方法としては、まず、「要望」を示すテキストを抽出し、その中から顧客ニーズを含む表現を選別するというアプローチを採った。

#### 3.2 問い合わせデータの分析

問い合わせデータにおいて、要望を示す質問テキストを調査すると、「(エラーの原因や対策方法を)教えてほしい」、「ログを調べてほしい」、「改良版を送付してほしい」、「担当者と話がしたい」、「パスワードが知りたい」など、製品の機能とは関係のない要望の方が多い。製品の機能と関係する問い合わせのみについて更に詳細に調査すると、コールセンタ側の回答テキストが、以下の2種類に分かれることが分かった。

- 既にサポート済みである機能に関する情報提供(「(マニュアルの記載箇所を)確認して下さい」等)
  - 未サポートである旨の回答
- 後者は、「満足されていない要望」を示しており、顧客ニーズとして抽出することが適切であると考えられることから、この種の情報を抽出することとした。

#### 3.3 抽出方法

##### (1) 処理の流れ

分析結果に基づいて、以下の3つのステップからなるツールを試作した。

Step 1 質問選別：要望表現を含む質問のみを抽出

Step 2 機能抽出：機能について述べた質問以外を除外

Step 3 回答選別：質問選別、機能内容抽出ステップで選別された質問の中から、未サポートであるとの回答を含む問い合わせを抽出

各ステップについて以下で詳細に説明する。

##### (2) 質問選別

「希望」の助動詞「～たい」のような表現以外に、可能/不可能を確認する質問文を抽出する。質問選別ルールの例を表1に示す。

表1 質問選別ルールの例

No	分類	抽出対象表現
1	可/不可を問う表現	(できる できない 出来る 出来ない)(か でしょうか のか のでしょうか)
2		(できます できません 出来ます 出来ませ

† (株) 日立製作所中央研究所

		ん) (か でしょうか)
3		(可能 不可能) (か ですか でしょうか なの か なのですか なのでしょうか)
4		(可能 不可能) (である であります) (か で しょうか のか のでしょうか)
5		(可能 不可能) (と に) (なる なります) (か  でしょうか のか のでしょうか)
6	要望を示す 表現	たい。(  が、 のだが、 のですが、 の で、 と考 と思)

### (3) 機能抽出

機能に関連する質問のみを抽出する。実際には、機能に関連する表現を全て列挙することが困難であるため、機能とは無関係の質問に多く出現する表現をルール化し、これらを含む質問を除外するものとした。表 2 に除外ルールを形成する単語の例を示す。コールセンタとして対応して欲しい要望を示す単語(No.1,2)、機能とは直接関係のない事柄(No.3)、要望ではなく不具合対策に関連する質問である可能性が高いことを示す表現(No.4)などを列挙した。

表 2 除外ルールに用いる単語の例

No	分類	除外対象単語
1	動詞	もらう/頂く/話す/尋ねる/お願いする/問合せ せる
2	名詞	契約/回避/調査/質問/相談/対策/適用/要望/ 話/判断/解決/対応/使用/入手
3	機能外	インストール/バージョンアップ/アップグレ ード/ダウンロード
4	不具合 関係	エラー/障害/不具合/異常終了

### (4) 回答選別

提供されていない機能であることを示す表現を抽出する。表 3 に回答選別ルールの例を示す。

表 3 回答選別ルール

No	抽出対象表現
1	(できません 出来ません できかねます 出来かねま す)
2	未サポート サポート対象外
3	サポートして(いません おりません)

## 4. 抽出結果

### (1) 対象データと抽出例

弊社コールセンタに寄せられた問い合わせ履歴データ 98,117 件(2010/6~11)を対象として評価を行った。人手によって評価した結果、適合率 97%、再現率 65%の精度で抽出を行えることを確認した。抽出された結果の例を表 4 に示す。

表 4 新機能候補情報抽出結果例

質問	回答
リカバリージョブを実行する 回数は設定できるのか。	リカバリージョブの実行回数 は設定できません。
画面を表示するバッチジョブ	GUI を表示する等の処理を行

インタフェースの管理/非管 理をコマンドで変化できま すでしょうか。	インタフェースの直接管理モ ード(管理/非管理)をコマン ドで変更することはできかね ます。
BAT ファイルを実行するユー ザ権限を指定することは可 能でしょうか。	実行権限を指定して外部プロ グラムを実行することは出来 かねます。
スケジュールの機能を使っ て夜間だけレプリケーション をする設定にしたい	同期やレプリケーションは時 間帯で制限をかけることはで きません

### (2) 有用性に関する検討

#### (a) 抽出された情報の量

対象データの場合、1ヶ月あたり平均 196 件(全問合せの 1.27%)が該当する問い合わせとして抽出された。抽出される件数が多い場合には、専任者をあてる必要がある等、運用コストが高くなる。今回の評価では、製品別では、1ヶ月 20 件前後であり、十分目視チェックが可能なレベルにまで件数を絞り込むことが出来ている。

#### (b) 抽出された情報の質

分析する価値がある情報の中に絞られているかどうかという点では概ね所望の結果が得られたと考えている。マーケティング部門からは、件数が少ないため、複数のニーズの中でどれを優先すべきか判断できないというコメントがあったが、これについては元々のデータ量不足の問題であり、アンケート調査等別の情報源を利用することが必要だと思われる。

## 4. まとめ

コールセンタで蓄積されている問い合わせデータから、高頻度ではないが、価値がある情報を抽出する方法を検討した結果について述べた。提案手法では、問い合わせデータの特徴である、「質問」と「回答」の組み合わせである点に着目し、「質問」と「回答」それぞれに出現する言語表現の組み合わせによって価値がある情報を抽出する。人手によって作成したルールによって、適合率 97%、再現率 65%の精度で抽出を行え、概ね所望の情報を絞り込むことができることを確認した。

## 参考文献

- [1] 鈴木英之進, 津本周作: 日本・アジアにおけるデータマイニングコンテスト, 情報処理学会誌「情報処理」, Vol. 42, No. 5, pp.457-461, 2001.
- [2] 村上知子: 推薦の意外性向上のための手法とその評価, 人工知能学会誌, Vol.24, No.5, 2009.
- [3] 乾孝司, 奥村学: テキスト評価分析の技術とその応用, 情報処理学会誌「情報処理」, Vol.48, No.9, pp.995-1000, 2007.
- [4] 山本瑞樹, 乾孝司, 高村大也, 丸元聡子, 大塚裕子, 奥村学: 文章構造を考慮した自由回答意見からの要望抽出, 言語処理学会第 12 回年次大会併設ワークショップ「感情・評価・態度と言語」, 2006.