

テキスト含意認識による分類軸発見と分類の自動化—多様なチャネルからの顧客の意見分析による実践—

土田 正明^{†1} 山本 康高^{†1}

^{†1}日本電気（株）

本稿では、テキストデータの概要を把握するために開発した分類軸の発見と分類を自動で行うシステムとその実践例について述べる。本システムは、テキスト含意認識技術を用いて、同じ意味を含む文をまとめ上げ、各まとまりが共通に含む意味を「代表文」として付与することで、分類軸の発見と分類を自動化している。実践例として、さまざまなチャネルのテキストデータを横断的に分析した結果、着目する製品・サービスに共通する意見を集計でき、また、意見の見落としの削減に有効であることを確認できた。

1. はじめに

自然言語で書かれたテキストデータ（以降、テキストデータとする）は、人にとって価値のある情報の宝庫である。なぜなら、基本的には、人が、他人や自分自身のために、伝えるべき、残すべき情報を記述したものであるためである。

テキストデータを分析することで、さまざまな知見が得られ、その知見を活用できる。たとえば、商品改善のためにとったアンケートの自由記述欄を分析することで、現状の問題や顧客の要望を把握できるので、改善のための仮説を立てることができる。また、たとえば、保守サービスの作業報告書を分析することで、保守員が過去の類似ケースを参考に円滑に対応できるようになるであろうし、故障の原因の傾向を把握できれば、特定の製品の不具合に気付くこともできるであろう。

テキストデータの分析は有用であるが、数値・定型データと異なり、分析の基本機能である検索や集計が難しい。理由は1) 同じ意味でもさまざまな表現が可能であることと、2) 似た表現でも異なる意味のことがある、の2点である。たとえば以下の3つの文を考える。

- A) 突然エンジンが止まった
- B) 急にエンストした
- C) エンジンから突然異音が出た

人が自然に解釈すればAとBが同じ意味でCは異なると考えるであろう。一方、機械的な処理としては、AとBは同じ単語が使われていないが、AとCは「エンジン」や「突然」が同じであるため、AはBよりもCに似ている、と判断しがちである。

筆者らは、この問題を解消すべく、テキスト含意認識技術を開発している。テキスト含意認識とは、2つのテキスト（T, Hとする）が与えられたとき、Tが真なら、常識的にHも真といえるか、すなわち、TがHの意味を含むか（含意関係）を判定するタスクである。たとえば、T:「商品Aは高価格だけどデザインが良い」とH:「商品Aは値段が高い」の場合、TはHを含意する。一方、T:「商品Aは値段が高い」とH:「商品Aは高価格だけどデザインが良い」の場合は、Hの中の「商品Aのデザインが良い」という意味がTに含まれないため、含意関係とはいえない。これが一定のレベルで実現できれば「テキストの意味」で検索したり、意味の出現頻度の集計をすることが可能となる。

本稿では、テキストデータの内容の概要を把握するために開発した、分類軸の発見と分類の自動化を可能にするシステムについて述べ、その実践例を報告する。本システムは、テキスト含意認識技術を用いて、同じ意味を含む文をまとめ上げ、各まとまりが共通に含む意味を「代表文」として付与することで、分類軸の発見と分類を自動化している。この結果により、分析対象のテキストデータの中に、どのような意味が、どれくらい書かれているかが把握できるようになる。

以降、業務で分析している方々（以降、分析者と呼ぶ）との議論を通して把握できた、分析上の機能要件について述べ（第2章）、それを解決するために開発したシステムとそれを構成する技術の概要を説明する（第3章）。次に、実践例（第4章）、経験から分かってきたノウハウや注意点を紹介する（第5章）。

2. 課題から抽出した分析機能の要件

テキストデータの概要を把握するためには、テキストマイニングツールを利用することが多い。テキストマイニングツールは、基本的には、単語や文節間の係り受け関係（以降、係り受け関係と呼ぶ）といった単純な構造に変換し、検索、集計を行う[1]。

一方、そういったテキストマイニングツールを評価・利用しつつも、導入を見送っていたり、十分に活用されていないことがあった。その理由を探るため、現場ではどのような分析結果を出すために、こういった作業を行っているかを分析者にヒアリングし、議論した結果、大きく3つの課題が抽出できた。

1. 単語や係り受け関係は、集計用の情報としては粗いため、結局原文を読み込み、カテゴリ定義とフラグ立てを行っている
2. 単語や係り受け関係では、文脈によって意味が変わってしまうため、カテゴリを検討するための数量的情報としては使いにくい
3. ツールで分析する前に、テキストデータから必要な情報を抽出（あるいは、不要な情報の除去）することが必要だが、その処理が難しい

以下、具体的に説明する。

1について、実際の作業としては、分析者が、経験に基づきカテゴリ定義を行い、先頭から原文を読み込みフラグ立てを行っていくが、読み込むにつれ、カテゴリの追加や細分化をした方が良いことに気が付き、また先頭からやりなおす、ということを行っている。カテゴリの定義が固まるために、何度も先頭からの読み込みが発生するため、多大なコストがかかっていた。単語や係り受け関係では、カテゴリ定義よりも情報として粗いため、その単位での集計は助けにならないことが多いようであった。

2については、単語や係り受け関係では、原文での文脈によっては、想定する意味とは異なる場合があるため、数量的な情報としては参考にできないということであった。たとえば、顧客の意見を把握するために「デザイン→良い」と「デザイン→悪い」の量を比較することを考える。この場合、前者には「デザインが良いと聞いていたがそうでもない」「デザインが良かったら欲しい」など「デザインが良い」という意味とは異なる情報が入り込んでしまう。また、「良い感じのデザイン」など同じと見なしたくとも、係り受け関係が異なると別扱いになるため、数量的な情報としては信頼できない、というこ

とであった。

3については、蓄積されているテキストデータには、分析目的に関係がない情報も含まれるため、目的に合った情報を抽出した上でツールで分析する必要があるが、現実的な人手コストで、必要な情報を抽出することは難しいということであった。基本的には、1) 人手で抽出する、2) 絞り込みのためのキーワードを選定して機械的に抽出する、3) キーワードで絞り込んだ候補からさらに人手で抽出する、の3つの方法が採られている。人手による抽出は、基本的には重要な情報を取り逃すことがないと考えられるため、現実的なコストに納まれば、最も良いアプローチである。一方、作業量には限界があるため、小規模なアンケート以外では現実的に難しい。一方、2のようにキーワードで絞り込む場合は、重要な情報を見落とすおそれがある。3のアプローチでバランスを取ることが望ましいが、これには経験や試行錯誤が必要となる。

これらの課題を解消するために、筆者らは以下の要件を定め、システムを開発した。

1. 単語や係り受け関係よりも詳細な「文」を分析の単位とする
2. 「文」を表示ラベルにして、その「文」の意味を含むかどうかで集計する
3. 分析目的に合わせ、特定の種類の内容が書かれた文の抽出機能をカスタマイズ可能とする

要件1により、「文」は1つの情報や主張として完結している可能性が高いため、「単語や係り受け関係では集計用の情報として粗い」という問題に対処している。要件2では、「文」の意味が含まれる原文の数で集計することで、文脈によってラベルの意味と原文で意味が変わってしまう問題に対処し、数量的情報の信頼性を向上させる。また、要件3にて、分析目的に合わせて必要な情報を抽出する（不要な情報を除去する）作業を自動化する機能を提供することで、分析の前準備の作業負荷を緩和できる。

3. 分類軸発見と分類の自動化による概要把握システム

3.1 システム概要

第2章で述べた3つの課題に対応し、かつ、さまざまなテキスト分析・活用ソリューションを提供可能にする概要把握システムを開発した(図1)。本システムは、1) 分析目的に合わせて定義された種別の情報を抽出する

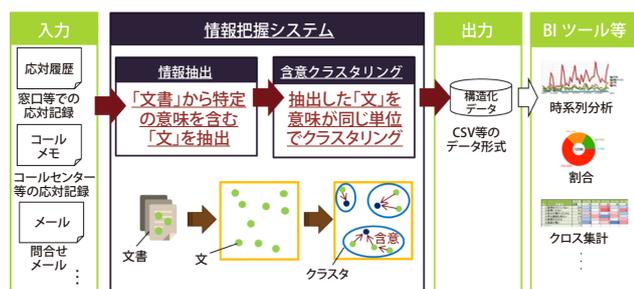


図1 システム概要

「文抽出」と、2) 同じ意味を含む文をまとめ上げてグループ化し、グループが共通に含む意味を代表文として付与する「含意クラスタリング」から構成される。含意クラスタリングは分類軸（代表文）の発見と分類を自動実行する技術と見なせる。本システムは、分析部分のみを共通化して開発されているため、他システムとのデータ連携やBIツール等への分析結果の連携も柔軟に対応できる。

3.2 文抽出

本機能は、テキストデータをその作成時と異なる目的で分析・活用する際に重要になる。筆者らの経験上、アンケートなど分析目的で作成されたテキストデータでない限りは、分析対象に使えるテキストデータは必ずしも分析目的に沿ったものになっていないことが多い。たとえば、営業日報から市場の動向を分析することを考えた場合、営業日報は市場の動向を分析するために作成されているわけではない。そのような場合は、営業日報から市場の動向に関連する情報を抽出する必要がある。狙いは、この作業の自動化である。

文抽出は、教師あり機械学習を利用している。すなわち、文書の中から、抽出すべき文（正例）とそうでない文（負例）を選別し、正例のみを抽出するモデルを学習する。詳細は割愛するが、機械学習の特徴量には、形態素解析、フレーズ（Nグラム）、係り受け関係などを用いるが、さらに、同義語辞書などを利用して、表層表現は異なるが、意味的には同一視したい情報を同じ特徴量とするためのロジックも備えている。

活用イメージの例として、業務やサービスの改善を目的に、営業の対応記録から顧客の意見を分析することを考える。対応記録には、面談時の状況や、後続対応者への情報連携・引き継ぎ事項など、さまざまな情報が記載されているであろう。一方で、アンケートとは異なり、設問によるバイアスのない、生の意見が含まれることも期待できる。このような場合に、本機能により、文書

中から「顧客の意見」が含まれる文を抽出する。詳細は4.2節に譲るが、実際には「顧客の意見」のような曖昧な定義ではなく、教師あり機械学習の学習データ作成には、人による解釈の揺らぎが許容範囲になる程度に基準を設ける必要がある。

3.3 含意クラスタリング

先にも述べたが、含意クラスタリングは、分類軸（代表文）の発見と分類を自動実行する機能である。入力された文の集合を対象に、同じ意味を含む文を同じグループ（以降クラスタ）に分類しつつ、各クラスタに対して、そのクラスタ内の文が共通的に含む意味を「代表文」として付与する。代表文のリストを見ることで、どのような分類軸があるかを一覧でき、また、各クラスタに属する文の数を見ることで、代表文の意味がどれくらい出現しているかを把握できる。また、クラスタに含まれる文の一覧を読み込むことで、効率的により詳細な内容を把握することもできる。

含意クラスタリングでは文間の含意関係の認識処理を網羅的に行う必要がある。これを現実的な時間で処理するために、テキスト含意認識技術（3.3.1節）の高速処理（3.3.2節）が求められる。また、文間の含意関係認識を実行するのみでは、同じ意味のクラスタが多数生成されてしまうため、クラスタの併合処理（3.3.3節）が必要である。

3.3.1 テキスト含意認識技術

テキスト含意認識は、2つのテキストを入力に、片方のテキストがもう片方のテキストの意味を含む場合に含意関係と認識するタスクである。筆者らの技術は、テキストとして「文」を前提にしており、国際評価ワークショップで1位[2],[3]になった実績がある。

特徴は、正確性と高速性を両立するために、2段階判定を行っている点にある。具体的には、2つの文の間の単語の意味の被覆率を計算し、おおまかに含意関係がありそうかを判定（1段目の判定）し、次に、文の構造的な情報を用い、含意関係にはないものをふるい落とす（2段目の判定）。1段目の判定は単純な処理であるため高速に動き、この段階で含意関係にないと判定された場合は、高速に処理が終わる。1段目で含意関係と判定された場合は、2段目でより詳細な判定を行うため、正確性も両立している。

1段目の判定：文T、文Hが与えられたとき、同義語など意味的な一致も考慮して、文Hの中の重要な単語が、文Tにも十分に出現しているかを重み付き単語被覆率

として計算し、その値が閾値以上の場合に、「含意関係」と判定する。重みは、低頻度な単語ほど意味の特定性が高いと考え計算する。単語重要度はデータベース化しておく。

2段目の判定：主語・述語、断定・否定・推測など、1段目の判定で活用していない情報を手掛かりに「非含意関係」であるかを判定している。経験的に、1段目の重み付き単語被覆率が十分に高い場合は、含意関係を持つ可能性が高い。そこで、2段目では、基本的には「含意関係」と考えつつ、「非含意関係」の場合に持つ特徴に着目して判定する。

3.3.2 高速含意文列挙

文間の網羅的な含意認識の実現のために、大量の文の集合から、クエリ文を含意する文を列挙する問題を高速に実行する技術を開発した。各文に対して、この処理を実行することで、文間の網羅的な含意認識が可能となる。

開発した技術では、3.3.1節の2段階判定の特質を活かし、精度を落とさずに含意文列挙を行うことができる。具体的には、1段目の判定の「重み付き単語被覆率」が閾値を超える文を、大量の文の集合から列挙するタスクと見なし、この高速処理を実現している。

開発技術は、岡崎らの方法[4]を、本タスク設定に合わせて拡張したものであるため、まず岡崎らの方法を説明する。岡崎らは、入力された集合（以降、入力集合）に対して、4種類の集合間類似度（ダイス係数、ジャックカード係数、コサイン類似度、シンプソン係数）が閾値以上となる集合を、大量の集合の内から列挙する高速なアルゴリズムを示している。詳細は文献[4]に譲るが、主には2つの工夫がなされている。

工夫1：入力集合と全要素が共通していても、類似度が閾値以上にならないサイズ（要素数）の集合が存在する。この条件を利用するために、検索対象の集合について、同じサイズの集合のみが検索されるようにサイズ別に検索インデックスを作成し、入力集合のサイズに基づき、類似度が閾値以上になりうるサイズに対応する検索インデックスのみを検索する。

工夫2：入力集合の要素を順にインデックスと照合して、入力集合と要素が共通する集合を取得していく過程で、ある時点で、未照合の要素がすべて共通しても、類似度が閾値を超えられない、という状況になる。それ以降は、その時点までに候補となっている集合が未照合の要素を含むかを確認すれば良いため、検索インデックスのIDリストの二分探索で確認する。

以下、改良点について説明する。筆者らが計算した

い重み付き単語被覆率は以下の式で与えられる。 $w(.)$ は単語を引数に重みを返す関数、 H はクエリの文の単語の集合（入力集合）、 T は判定対象の文の単語の集合（大量に存在する検索対象の個々の集合）である。

$$\frac{\sum_{c \in |T \cap H|} w(c)}{\sum_{h \in H} w(h)}$$

すなわち、 H が与えられたとき、この値が閾値以上となる T をすべて列挙することが目標となる。

「集合のサイズ」を要素の重みの総和と定義し、重み付き単語被覆率の閾値を α とすると、 $\alpha \sum_{h \in H} w(h)$ （以降、 β とする）未満のサイズの集合は、原理的に閾値を超えないことが分かる。この性質を利用し、同じ上限と下限の範囲のサイズを持つ集合でグループを作り、そのグループの集合のみが検索可能なインデックスを作成しておくことで、各グループの上限値が β より小さい場合、そのグループへの検索が不要と判断できる（工夫1の拡張）。また、検索が必要なグループに対しては、検索中に、入力集合の要素の重みの和から「その時点までに検索した要素の重みの和」を引いた値が β よりも小さくなった段階で取得済みの集合のみを候補とする（工夫2の拡張）。

集合のグループは、分割数をパラメタとして、各インデックスから検索可能な集合の数ができるだけ同じになるようにサイズの上限と下限を設定して作る。インメモリで処理できる場合は、分割数を多めにとってよいが、ディスク上にインデックスを配置する場合は、検索回数が多くなるとランダムアクセスの回数が多くなってしまいうため、実験的に適切な値を設定する。

この高速列挙技術により、1段目の判定の基準を満たす候補を列挙でき、それらの候補に対してのみ2段目の判定を行えばよくなるため、全体として高速な含意文の列挙が可能となる。

3.3.3 クラスタの併合処理

網羅的な文間の含意関係が認識できても、同じ意味のクラスタが複数存在してしまうため、併合処理が必要である。たとえば、文AとBが同じ意味の場合、Aを代表文とするクラスタとBを代表文とするクラスタが両方存在していても、クラスタが表す意味は同じであるため、1つに統合すべきであろう。

本アルゴリズムでは、2つのクラスタが同義であるか、含意関係にある場合に併合する。同義や含意の関係は、

2つのクラスタに属する文から判定できる。具体的には、含意関係がすべて正しく判定できれば、クラスタAとBが同義関係の場合、AとBに属する文は基本的には同じになる。また、BがAを含意する場合は、含意関係の推移律よりBに属する文はAにも属する。

併合するか否かはクラスタ間の同義・含意の関係を集合間類似度により判断する。類似度により、含意関係が完全に正しく認識できなくとも、多少のミスは吸収して併合処理を行う狙いである。同義関係のクラスタのみ併合する場合は、2つのクラスタの文集合のジャカード係数（集合間の要素が完全一致のときに1となる）を計算し、十分に高い値であれば併合する。含意関係の場合は、シンプソン係数（集合間の要素に包含関係がある場合に1となる）を用いる。

サイズの大きいクラスタに小さいクラスタを併合していき、最終的にどのクラスタ間でも併合できなかつたら終了する。

4. 実践例：顧客の意見分析

本章では、顧客の意見分析の適用事例を紹介する。製品・サービスに関する要望は、企業が次の打ち手を検討する重要な材料となる。そのため、要望の集計による課題選定、施策を立案・実行、顧客の意見の変化を評価、というPDCAは、業種を問わず実施されている。顧客の意見分析では、自動集計が基本となるが、その用途は広い。たとえば、コールセンター業務では、オペレータの回答の効率化に利用でき、品質管理業務では、ほかの数値情報と組み合わせてBIツールによる傾向分析の材料に使える。また、営業が記載した顧客の意見（営業日報）から顧客に響いた提案をまとめることで営業支援にも利用できる。筆者らは、金融業、製造業、交通、流通、キャリアなど20社を超える企業の顧客の声の分析を実施するとともに、システムとして本機能を納め、実運用にも利用されているという実績を有する。本章で紹介する事例は、守秘義務のため業種を伏せ部門名も架空のものを用いているが、数値情報、分析結果は実例に近い内容である。

4.1 分析の目的

本節では、まず企業Aで行っていた分析業務を概説した上で、それをどう変えたいか、目的を示す。

企業Aの主力事業は、一般消費者向けへの製品、サービス提供であり、消費者からは、月に数十万件に及

ぶ意見が寄せられている。意見は、コールセンターへの問合せ、店舗スタッフへの申し入れ、Webサイトへの投稿といったチャンネルから寄せられ、それぞれが、テキストデータとしてシステムに蓄積されている。なお、テキストデータの他、性別、年齢層、製品・サービスの利用歴など消費者に関する属性情報も分かる範囲で記録されている。

顧客満足度向上を目指す部門（以降CS部門とする）では、これらのデータを用いて、日々の意見の確認や定期的な傾向分析を行うことで、突発的な不具合の発見や製品・サービス改善のための分析結果を関係各部にフィードバックすることをミッションとしている。これらの分析は、いくつかのサービスや製品単位で分析チームを作り行われており、基本的には、どのような意見が何件程度届いているかを整理していた。

企業Aでは、意見を発見・整理するために、導入済みのツールを用いて頻出キーワードの見える化を行っていた。しかしながら、キーワードだけでは内容が把握できず、結局、個々の意見の目視が必要となり意見の分析に時間を要していた。

個々の意見の目視確認には量的限界があるため、確認できる量までキーワードや属性で意見を絞りこんだり、特定の種類のデータを調査対象から外すといったことを行わざるを得なかった。特に、Webで投稿される意見は長文になる傾向があるため、分析に用いておらず、意見の見逃しのリスクがあった。また、意見全体を通して見える製品・サービスの不具合に気付きにくいといった問題も生じていた。たとえば、複数の製品やサービスで共通して利用している別サービスがある場合、この仕様の变化や不具合で、企業Aの製品・サービスに共通の問題が生じることがある。しかしながら、チーム別の分析では、個々のチームで見られる報告件数は少なく、その把握が遅れることがあった。

筆者らはこれらの問題に対して、どのような意見が届くか想定しきれない中で分類軸を事前に設定することが難しい、という共通の要因を見出した。筆者らのシステムは、テキストデータの内容に応じて、分類軸の発見と分類を実行できるため、有効であると考えた。また、さまざまなチャンネルのデータを横断して分析することで、これまで気が付けなかった意見を把握できる可能性があると考えた。そこで、以下2点を目指し、分析を行うことにした。

1. 全データを分析対象とし、意見分析として妥当な分類軸を洗い出す。

2. 全データを分析対象とし、製品・サービス別では見逃しやすい意見を捉える。

4.2 分析方法

筆者らは分析目的に合った結果が得られるかを確認するために2つの課題を設定し、企業Aの顧客の意見分析を実施した。分析の途中経過を企業Aの分析者と共有することで、現場の感覚とずれないように注意を払いながら進めた。対象データは特定期間の約5万件とした。

記載形式にロバストな意見抽出：全データを利用するために、さまざまな記載形式のテキストデータを統合して分析しやすいように、主要な意見のみを抽出する。

分類軸の粒度決定：業務側で理解・活用しやすいレベルに意見のまとめ上げの粒度をコントロールする。

まず、記載形式にロバストな意見抽出では、主要な意見を言い表している個所だけを抽出し、短文として意見を扱うようにした。製品・サービスの利用者が直接コメントを記載するチャンネルのテキストは、長文で冗長であるケースが多い。一方、コールセンターのオペレータは端的な表現を用いる。短文として意見抽出することでこれらを統合しやすくなる。このとき、企業Aにとって有益な意見の条件は、CS部門や彼らからレポートを受け取る部門へのヒアリングを行い定義した。具体的には、企業Aと筆者らで共通見解が持てるようにガイドラインとして、必要な意見の要件を明文化するとともに例文を列挙した。意見抽出には3.2節で説明した文抽出機能を用い、学習データ作成でも複数人でズレが生じないかを確認して進めた。

分類軸の粒度決定では、いくつかの結果を作り、企業Aの分析者に確認を取りながら、粒度を調整し、顧客の意見を把握しやすい粒度を決定した。含意クラスタリングは分類軸を自動発見する利点はあるが、その意味の抽象度が高すぎると中身を推測しにくくなり、逆に具体的すぎるとその意味を含む意見が少なくなるため、こういった粒度の調整が必要となる。調整のためには、最低限含まれる名詞や述語項の数を設定したり、文とみなす最低の文字数などで調整した。また、企業Aに特化した同義語の拡充も行った。この拡充には同意語抽出技術[5]を用いた。

4.3 分析結果

本節では、本実践で目指した、1) 全データを分析対象とし、意見分析として妥当な分類軸を洗い出す、2) 全データを分析対象とし、製品・サービス別では見逃し

やすい意見を捉える、のそれぞれの結果と考察を紹介する。

4.3.1 分類軸の洗い出し

全データを用いるために行った意見抽出結果の例を表1に示す。表1では、チャンネルごとに同じ内容でも記載形式の異なるテキストデータが蓄積されていることが分かる。また、意見を端的に表した部分文字列が抽出できていることが分かる。

含意クラスタリングの結果を表2に示す。表2では件数の多かった上位10件を示している。

本結果を企業Aに提示したところ、違和感のない結果であるとのコメントがあり、定性的にも実用できるという評価を得た。表3に示す通り、表2中の下線で示した2つの意見が分かれてまとまっている点は、従来は人手で再分類作業が必要だった内容であるため、自動的に提示される点は特筆できる、とのことであった。これは、筆者らのシステムの狙いが確認できた例ともいえる。

また、過去に人手で分類した結果と、今回の自動分類の結果のうち、可能な範囲で人手で対応付け、関係係

表1 意見抽出結果の例

チャンネル	抽出された意見
コールセンター	〇〇サービスの更新のアナウンス届いていない。
店舗	【対象】 〇〇サービス 【内容】 アップデートのアナウンスがなかった 【対処】 お詫びと再発防止策を説明
Web 投稿	5/3に〇〇サービスの新しいバージョンが出たと思いますが、アップデートの案内が、全然届いていないので、非常に不満です。

表2 含意クラスタリングによる顧客の意見分析結果

代表文	件数
反応が悪い	257
通知機能が使えない	212
電源入れ直して改善せず	177
アップデートのアナウンスなし	144
データ保存に失敗していた	121
電源入れ直して改善	115
ボタン操作がきかない	109
設定方法が分からない	89
設定が戻せない	88

表3 まとめ上げられた意見の例

代表文：電源入れ直して改善
電源を再起動すると改善したが再発が心配
電源の入れ直しを行ったら改善
店頭にて電源の入れ直し、設定を直したところ改善
代表文：電源入れ直して改善せず
電源を入れ直してみたが改善なし
電源の再起動を案内してみたが直らない
電源の再起動や設定の初期化などいろいろ試してみたが効果なし

表4 含意クラスタリングによる意見分析結果

代表文	件数
反応が悪い	209
電源入れ直しで改善せず	190
:	:
ダウンロードができない	45
:	:
キャンペーン商品がもらえない	36

数を計算したところ、0.7を超える値を示した。これは、人手で集計した場合と比べて強い相関があることを意味し、数量の大小関係の傾向を把握するという視点では、人の感覚に近い結果が得られていることを示唆する。

以上より、観点の洗い出しの自動化が一定のレベルで実現できていると評価された。

4.3.2 見逃しやすい意見の捕捉

特定の期間に絞った場合の分類結果(表4)を提示したところ、ダウンロードに関する不具合を示唆する分類軸が目目された。当然、「ダウンロードができない」に類する問合せが存在することは分かっているが、思ったよりも量がある、ということであった。

確認したところ、製品・サービスごとの集計ではそれぞれ10件未満となることから、気がつきにくい意見であった。この期間には、ダウンロード用サーバの異常があったとのことであった。また、窓口販売において、キャンペーン商品が手渡されていないことへのクレームが一定量存在することも想定外の発見であった。

これらの発見が即座に顧客満足度向上につながるとはいえないが、たとえば、ダウンロードサーバの障害を速報的にコンタクトセンターに連携すれば、問い合わせしてきたユーザに的確な情報が提供できるようになるであろうし、キャンペーン商品の品切れに注意することで、残念な思いをする顧客の数を減らすことができたり、品切れの際には顧客の感情をケアするといった施策も考えられる。

このように、全データを対象に、製品・サービスを横断で分析し、分類軸を自動提示することは、意見の見落としを防ぐ手段として有効であることを示唆する結果が得られ、企業Aから高く評価された。

5. 顧客の意見分析に関する知見

5.1 顧客の意見分析の知見

前章で紹介した顧客の意見分析やその他の経験してきた事例を踏まえて得た知見を述べる。ここでの知見とは、顧客が解消したい問題と技術で実施できることとの差異

を埋めるためのノウハウを指す。

この差異を埋めていくためには、解くべき問題の認識を顧客と合わせることで技術的な工夫が求められる。技術的工夫は第3章で説明済みであるため、本節では顧客との認識合わせの工夫を述べる。ポイントは、目的に合った評価基準、正解データに対する合意、顧客の技術に対する正確な理解と考える。この3つを説明する。

目的に合った評価基準: 分析が適切に行えているかは、分析結果の用途を理解しないと測れない。前章の事例では、顧客の要望をレポートとしてまとめること、また、時系列的な変化が大きい意見は、即座にアラートを上げることが用途であった。そのため、分類の観点が正しく挙げられており大局的に数字が合うこと、また急増している意見もまとめられることが本分類結果の評価基準になる。顧客の意見分析となると、単純には、分類項目が正しく洗い出せ、各分類結果の数値が正確に合うことを評価基準としがちであるが、これは上記目的からすると、オーバースペックであり、時間の浪費となる。そのほか、何十年も利用できる耐久消費財を扱っている製造業では、それら製品に対するノウハウを継承していくことが難しいため、たまにしか来ない故障情報でも、過去の同じ故障情報と同じ分類項目に割り当てられることが重要な評価基準となり、また、食品や美容業界であれば、アレルギーを匂わす意見は、少数でも見逃さずアラートが出せるかが評価基準となる。このように目的に合った評価基準を設定し、技術をそれに合わせチューニングしていくことが運用時に使われる技術となる。

正解データに関する合意: 顧客の意見分析では、望ましい結果の客観的な基準がない。そのため、上記目的を鑑みた上で、実運用に資する分類結果を企業と合意することが重要となる。具体的には、企業と取り決めたデータの範囲で実運用に資する分類結果を人手で作ってもらい、その結果を正解として用いることに合意を得る。これを分析のプロセスの中を含め、最初に企業に提示しておく。4.3.1節において、人手分析との相関により評価をしているのは、事前に上記合意がなされているためである。特に、正解がない状況では、総合的判断ができず、各分類結果の間違いに議論が集中してしまい、本来の目的を見失うことが多いため注意が必要である。

顧客の技術に対する正確な理解: 分析技術を伴うシステムでは、技術的に解決できることと、そうでないことが混同されやすい。そのため、その差異を丁寧に説明していくことが重要となる。たとえば、顧客の意見の中から新しい意見を見つけないという要望は非常に多い。筆

者らの技術は、含意すれば、少数の意見でも1つのクラスを形成する点において、新しい意見として現れた少数意見を発見しやすい性質はある。しかしながら、そういった意見があるか否かはデータに依存するものであり、技術が保証できる範囲ではない。これを技術の問題と混同されないように、技術を提供する側が、正確に伝えていく必要がある。こうした技術が行えることと解くべきスコープのズレは、経験上、課題解決に専念する当事者では気づきづらい。そのため、筆者らは、課題と解決策が一致しているかを客観的に観察するステージゲートを設ける工夫をしている。

5.2 分析に関するフレームワーク

最後に意見分析に限ったことではないが、筆者らがデータ分析を成功させるために活用しているフレームワーク[6]を紹介したい。本フレームワークは文献[6]の著者と筆者らが所属する日本電気(株)とで検討・考案したもので、DIVAと呼んでいる。データ(Data)を分析することで、情報(Information)に変換し、それによって価値(Value)につながる行動を利用者に起こさせ、利用者の目的を達成(Achievement)する。DIVAはこの頭文字をとったものであり、データ処理の手順を示している。企業へのヒアリングを行うと、このDからAの順序で分析を検討していることが多い。これは、データがあるので活用したいという発想の場合にしばしば起こる。一方、分析することが目的化しやすいという落とし穴もある。

DIVAは、分析設計するにあたり、AからDの逆順で検討することがポイントである。すなわち、成功状態を達成するのに資する活動は何か、それを起こさせるために必要な情報は何か、そのために必要な分析は何か、どんなデータを用意すべきか、の順に考える。これにより、達成条件を顧客企業と共有した上で、分析が目的ではなく手段とする議論が可能となる。筆者らが、顧客の意見分析でも分析結果が生む価値を最初に検討しているのは、この考え方に則ったものである。

6. おわりに

本稿では、テキストデータの概要を把握するために開発した分類軸の発見と分類を自動化するシステムとその実践例について述べた。

テキストマイニングツールの評価・利用を行っている顧客との議論を通し、1) 単語や係り受け関係が分析

の観点の粒度としては粗い、2) 単語や係り受け関係は原文での出現文脈によっては意図しない意味で使われている、3) 分析目的に合わせて必要な情報を抽出する作業が難しい、という課題を見出し、それぞれに対応するために、i) 単語や係り受け関係よりも詳細な「文」の単位で分析可能で、ii) 「文」の意味の出現に基づき集計でき、iii) 分析目的に合った「文」を抽出するためのモデルを作成し自動抽出するシステムを開発した。

実践例として、コールセンターへの問合せ、店舗スタッフへの申し入れ、Webサイトへの投稿といったさまざまなチャネルのテキストデータを分析する、顧客の意見分析を紹介し、分類軸の洗い出しを自動化することで、製品・サービスで共通する意見を集計でき、また、意見の見落としの削減に有効であることを示唆する結果が得られたことを述べた。

本システムを有効活用するためには、目的の達成(Achievement)のために、価値(Value)につながる行動を起こさせるために必要な情報(Information)をデータ(Data)の分析によって作り出す、という全体設計を行い、その設計で求められる分析要件を満たせるかを判断することが重要である。これを適切に行うためには、経験や知識が必要であるが、この意識を持つだけで筋の悪い分析・活用案を回避することができる。

本稿が、テキストデータの分析・活用を実行したり、検討されている方々の一助になれば幸いである。

参考文献

- 1) 脇森浩志：ビッグデータに対するテキストマイニング技術とその適用例，UNISYS 技報，Vol. 32, No. 4, pp.337-349 (2013).
- 2) Tsuchida, M. and Ishikawa, K.: IKOMA at TAC 2011: A Method for Recognizing Textual Entailment using Lexical-level and Sentence Structure-level features, TAC 2011 Workshop Notebook (2011).
- 3) Andrade, D., Tsuchida, M., Onishi, T. and Ishikawa, K.: Detecting Contradiction in Text by Using Lexical Mismatch and Structural Similarity, Proceeding of the 10th NTCIR Conference, pp.512-517 (2012).
- 4) 岡崎直観，辻井潤一：集合間類似度に対する簡潔かつ高速な類似文字列検索アルゴリズム，自然言語処理，Vol. 18, No. 2, pp.89-117 (2011).
- 5) 土田正明，石川 開，久寿居大，楠村幸貴，中尾敏康：ビッグデータ活用のためのテキスト分析技術，NEC 技報，Vol. 65, No. 2, pp. 77-80 (2012).
- 6) 鈴木良一：データ活用仮説量産 フレームワーク DIVA，日経BP 社(2015).

土田 正明 (正会員) m-tsuichida@cq.jp.nec.com

2005年東京理科大学理工学研究科経営工学専攻修士課程修了。同年日本電気(株)に入社。以来、自然言語処理に関する研究開発に従事。2009～2011年情報通信研究機構専門研究員。現在、データアナリストとしてデータ分析事業を推進しつつ、研究開発にも従事。2015年度情報処理学会喜安記念業績賞。博士(工学)。

山本 康高 (非会員) k-yamamoto@jx.jp.nec.com

2006年名古屋大学大学院工学研究科博士後期課程修了。博士(工学)。同年日本電気(株)に入社。近年は、分析コンサルタント/データアナリストとして、分析技術の実課題適用による分析事業の拡大に従事。

採録決定：2016年7月23日

編集担当：福島俊一 ((独) 科学技術振興機構)