

テキストマイニングを用いた、CRM ユーザ事例集の表現の検討

Mining Semi-structure for text expression in the customer introduction examples for eCRM

一松 達矢† 辻 光宏†
Tatsuya Ichimatsu Mitsuhiro Tsuji

1. はじめに

企業ビジネスの営業活動の一つとして提供されているユーザ事例の文書テキストの表現は、多種多様であり、その完成度も玉石混合である。我々は、統計的なテキストマイニング技術を利用し、ユーザ事例集のテキスト表現の構造について検討を進めてきた。

プレゼンテーションに優れた製品紹介に定評がある eCRM の Salesforce 社に注目し、そのユーザ事例集テキストが持っていると思われる半構造化された文書テキストを取り上げる。特に eCRM のクラウド企業として提供している最先端サービスの特性にも留意しながら、情報の発見だけでなく半構造化文書テキストの仮説検証を目指して取り組んだので、報告する。

2. ユーザ事例

企業が提供しているユーザ事例は、企業の製品を既に購入した先行ユーザにおいて、どのように役立させているかを紹介するものである。そこでは、いつ何をどのように導入したのかを、その課題や選定理由や導入効果などを含めて報告している。

Salesforce 社のユーザ事例は、ユーザごと 2 ページに分量をコンパクト化し、100 社以上の事例を集大成している。クラウド導入や多様な顧客対応に取り組みたい新規の顧客が導入検討する際に、大いに役立っている。

2.1 段階的な取り組み

本研究の準備段階では、ユーザ事例の潜在的な構造を説明している Salesforce 社のさまざまな説明文の文章テキストを活用して、ユーザ事例の文章テキストの構造に準じたテキスト構造を事前調整しておき、進展に応じて再調整する仕組みを試行しておく予定である。

第一段階では、ユーザ事例表現の初動検証の立場から 3 つの顧客（キャノンマーケティングジャパン株式会社、甲府市役所、株式会社陣屋）をプロトタイプとして、ユーザ事例文書の構造を明らかにしていく。その際に想定した半構造化文書の基幹的な項目は、

- ・ユーザの紹介
- ・対象の業務の紹介
- ・ユーザの課題
- ・Salesforce 選定の理由
- ・Salesforce 導入の効果

の 5 つである。特に後の 3 項目について関連付け（連想）

が、どの程度できているのかに注目した。

第二段階では、Salesforce 社の他の顧客 100 社程度のユーザ事例集に適用することによって、第一段階で定めた初動検証方法の妥当性を確認していく。対象のユーザは、業種ごとに

- ・ハイテク 6 社
- ・ビジネスサービス 22 社
- ・メディア・エンタメ・通信 7 社
- ・医療・製薬 7 社
- ・小売・流通・運輸 15 社
- ・建設・不動産 12 社
- ・教育・人材 5 社
- ・製造 10 社
- ・金融・保険 10 社
- ・飲食・宿泊 10 社

の合計 104 社の事例である。

第三段階では、Salesforce 社の同業他社でのユーザ事例に範囲を拡張して、本研究のアプローチ方法の有効性を確認する予定である。

2.2 準備段階の手がかり

ユーザ事例の表現を構造化するための準備として、ソフトウェア工学の上流過程でのゴール指向の一つである Goldratt の LTP（論理思考プロセス）による研究がある（渡邊, 2013）。そこでは、ユーザの課題を構造的に図式化する現状分析ツリーと、選定理由と導入効果を図式化する未来実現ツリーとを用いている。

現状分析ツリーでは、分析対象の何らかの目標を定めるために、好ましくない課題と原因とをマイナス面から整理している。一方、未来実現ツリーでは、対立する関係を認識しながら、それぞれの立場から要件を見つけてプラス面から整理している。

テキストマイニング手法を、発見的ではなく、マイナス面とプラス面双方からの意味解析の結果に沿って構築し検証することが可能になることを、我々は目指している。

2.3 分析する項目

Salesforce 社の場合、各ユーザ事例には、本文テキストとは別に、概要として、業種、活用分野、課題、選定理由、効果などをまとめている。これらの項目は、ユーザ事例すべてに掲載されている共通した説明テキスト資料であり、構造化されている。

本研究では、これらの項目を指標に用い、本文テキスト内に、それぞれの項目に関係のある単語表現がどの程度現れているかを検証する。これにより、ユーザ事例の本文

†関西大学 総合情報学研究科

テキストが概要に準じた構造を有しているかの確認を試みた。

比較対象である、事例概要内テキストの影響を受けないようにするため、テキストマイニングの分析対象は事例を紹介している本文テキストと見出しのみとし、記事に掲載のある会社の規模と社員数と所在地といった紹介テキスト、対話形式文の質問文、発話者名、取材日、図や画像に含まれるテキストは分析対象外とした。

3 分析

以下の分析を、数理システムのテキストマイニング用のソフトウェア TextMiningStudio を用いて行った

3.1 係り受け分析

(1)分析の指針

テキストマイニングの中で重要な要素を占める単語の出現頻度の分析をさらに進め、係り受け関係のような単語間の関係性を考慮することによって、文テキストが持つ意味構造をより深く分析できる(佐藤ら, 2006)。

各社ごとでの係り受け頻度解析の結果に基づいて、問題の解決方法について、検証した。以下に係り受け頻度解析の上位3つの表現を示し、ユーザ事例内の概要に用いられる表現との関連を調べた。

(2)分析の結果

・株式会社陣屋

「ソーシャルメディア—活用」「紐—づける」「Facebook—活用」が比較的多く、SNSを活用することに重点を置いていることが確認できた。さらに、紐づけるという表現に注目し、原文での使われ方を確認すると、「多彩なデータを紐づけて一元管理」「お客様情報と紐づけて」「項目を紐づけたお部屋カルテ」のように使われており、SNSだけでなく企業内でのCRMを強化する動きに関する記述が示されていた。

CRMの強化に関しては、概要内の課題項目「予約の一元管理やきめ細やかな対応ができなかった。」という記述に原因が現れている。

またSNSの活用に関しては、原文を参照すると顧客との繋がりだけでなく、社員同士の繋がりをも可能にしていた。これは、概要内の効果項目「予約業務と顧客管理の効率化が可能に」が本文内に現れていると推測される。

・キャノンマーケティングジャパン

「CRM—連携」「データ—連携」「顧客データベース—構築」が比較的多く、情報管理のキーワードがみられる。また「連携」という単語は下位の係り受けにも多く現れている。そこで「連携」を係り元の単語フィルタに設定し、頻出語句の使われ方をより詳細に調べてみた。その結果、「Facebook 連携—活用」「ソーシャルメディア連携—活用」「ネット予約連携—開始」のようにSNSやインターネットを使ったサービスとの連携を表すことを確認できる。キャノンの事例ではSNSを用いる目的理由として、概要テキスト内の効果項目「リアルタイム双方向コミュニケーションで会員の満足度向上」が挙げられる。

・甲府市役所

「Force.com—採用」「IT コスト—圧縮」「コスト—稼働」が多く、この他にも現れる表現があり、他の2社より重複表現が少ない。「IT コスト—圧縮」は目標となる表現である。また「コスト—稼働」について、原文を参照すると、当初予定していたコストの5分の1で稼働できたことを確認した。したがってこの表現についてもITコストの圧縮を表している。これらは概要内テキストの課題項目の「IT コスト 38.5%減の目標に向け、低価格なシステムが必要だった」と対応していると推測できる。

本研究の現時点では、事例集テキスト内の概要項目の内容に対して、関連する単語が特徴的に現れているかについて確認した。しかし、今後は概要内のテキストを構造化し、本文テキスト内に現れる統計的な係り受け情報との関連性をも比較する予定である。それによって、複数個ある課題がどの程度記載されているのか、類似する課題を持つ事例は、本文テキスト構造も類似しているかを検証することを目指す。

3.2 連想関係マップ

(1)分析の指針

ユーザ事例集のテキスト情報から知識につながる有用な情報を得るため、統計処理した連想関係から発見しやすくする技術として、連想関係マップを採用する。連想分析は、単語の共起性に基づいて、数値的に得られるテキストや単語間の関係を、辞書作成や関連度計算によって関係抽出するものである(三末ら, 1999)。

単語の間の連想関係を、単語の共起性に基づいて計算し、単語のマップ(ネットワーク図)として可視化する。ここでは、重要単語の検索や関連単語や関連テキストの検索を行うことで、個別テキストでは調べにくい、テキスト群全体が持つ特徴や傾向を、この単語の連想マップを通して読み取ることが可能となる(渡部ら, 1999)。

本研究では、ユーザ事例資料に含まれている、概要の効果、選定理由、結果などに現れる単語の中にどの程度重要なものが含まれているのか、類似した表現がどのように他の単語やテキストと連想しているのかについて分析し可視化して、検証を進める。

(2)分析の結果

単語同士の共起性を用いて、単語と単語の連想(繋がり)をネットワーク図に表した。出現回数を3回以上、行テキスト単位の共起ルール数に基づいて、9つのクラスタに示した。各クラスタそれぞれに、効果的な表現が反映されているのかを確認することとした。以下の様相が導出されることを想定している。

・「Salesforce」には、「CRM」「導入」「基盤」と「連携」から矢印が伸びる。SalesforceがCRMを中心にシステム基盤を整えるサポートを行っていることを確認可能であろう。

・「ソーシャルメディア」は、「手法」「活用」のノードと結びつき、SNSと組み合わせたサービスの展開を表すことになる。

・「手続き」を中心として、「人」「順番」「窓口」が集まっているクラスタが存在し、甲府市役所のサービスから現れるであろう。クラスタは独立して出現し、他の2社と異なる業務特性があると推測できる。

・9つのクラスタのうち、3社すべてから表現が抽出されているのは「業務」「流れ」を結んだクラスタであろう。すべての企業において業務の流れを改善することが問題となっているだろう。

今回の分析では、統計的に処理した単語の共起性をネットワーク図に表した。

例えば、今回分析した3社の結果のうち、キャノンの文章からは、「CRM」「導入」「基盤」「連携」の単語が「Saleceforce」に向かっていることが確認できるだろう。係り受け頻度分析や、単語頻度解析の結果と照らし合わせて考えると、「CRM」が主な課題であると推定できる。同様に「CRM」を主な課題としている企業のユーザ事例を、現在得られている連想関係マップと関連付けて、「CRM」以外の「導入」「基盤」「連携」のノードを介して、「CRM」に関係する単語の集団を形成したい。このことから、CRMを課題としている企業の共通点を見つけることができる。

また、コストのノードを持つ甲府市役所のクラスタに注目すると、「システム」「稼働」のノードが関係することを想定している。これらの結果から、新しくシステム稼働しようとする際にコストの問題が発生していることが確認できる。他のユーザ事例の中でも、「システム」「稼働」といった単語を多く含んでいる場合には、その連想単語同士が紐づき、このクラスタの連想単語を強化し、その根本原因にはコストの問題が含まれている、ということを実証したいと考えている。

また、単語の共起性情報だけでなく、係り受け情報からも「連想関係マップ」を作成し、共起と係り受けの2つの情報から、他のユーザ事例が既存のユーザ事例集に対してどのような連想（関連性）を持っているのか、どのような注目すべき単語と連想する（結びつく）のかを明らかにしたい。これによって新しい関係性で事例集の特徴を可視化することを目指している。

さらに、Saleceforceの104社のユーザ事例すべてに適用して、共通する構造的な根本原因クラスタを発見し、他の企業のユーザ事例がどのクラスタと結びつくのかを分析することで、業種やサービスの枠を超えて、課題、選定理由、導入効果の発見と、ユーザ事例の選定に役立てることも目指している。

4 今後の課題と展望

本研究では、情報の発見だけでなく半構造化文書の仮説検証までを目指す。Salesforceの3つのユーザ事例について、テキストマイニングを利用して、その特徴的な表現についての構造化の検証と、連想関係マップを用いた可視化を試みた。

業種もサービスも異なる3社のみでは、表現を検討する精度を高めるには、不十分な量である。

本研究の手法を応用して、さらにSalesforceで取り扱っている、その他の業種も含めたユーザ事例の検証を進める。

同時に、同業種間での特徴を比較することで、潜在的な構造の評価の定式化を図るだけでなく、「連想関係マップ」の精度向上をめざす。

また、Salesforce同様に、クラウドサービスを提供する企業が、ホームページ上に公開しているユーザ事例を調査し、企業ごとに、ユーザ事例構造の特徴的な要素を比較検証することも検討したい。

5. 参考文献

佐藤一誠，中川裕志（2006）：係り受け関係を考慮したテキストマイニングのための半構造マイニング手法の提案．情報処理学会研究報告 第78号．pp.207-214

那須川哲哉（2006）：テキストマイニングを使う技術／作る技術．東京電機大学出版局

福部兼敏（2010）：テキストマイニングで広がる看護の世界．ナカニシヤ出版

三末和男，渡部勇（1999）：テキストマイニングのための連想関係の可視化技術．情報処理学会研究報告．FI-55-9 DD-19-9，pp.65-72

渡部勇，三末和夫（1999）：単語の連想関係によるテキストマイニング．情報処理学会 第55回 情報学基礎研究会資料．99-FI-55-8/99-DD-19-8，pp.57-64

渡邊大輝（2013）：ゴール指向を用いたビジネス戦略のデザイン．関西大学総合情報学研究科修士論文

キャノンマーケティングジャパン- CRM Salesforce.com 日本（最終閲覧日：2013年7月26日）．

<http://www.salesforce.com/jp/showcase/stories/cmj.jsp>

甲府市役所- CRM Salesforce.com 日本．（最終閲覧日：2013年7月26日）

<http://www.salesforce.com/jp/showcase/stories/kofu.jsp>

陣屋-CRM Salesforce.com 日本．（最終閲覧日：2013年7月26日）

<http://www.salesforce.com/jp/showcase/stories/jinya.jsp>