

コミュニケーションデータにおける不満・要望・原因推定のための確率モデル構築に関する一検討

Preliminary examination of probability model construction for estimation of user frustrations, demands and causes in communication data

小松 恭子[†] 石先 広海[†] 松本一則[†] 服部 元[†] 滝嶋 康弘[†]

Kyoko Komatsu Hiromi Ishizaki Kazunori Matsumoto Gen Hattori Yasuhiro Takishima

1. はじめに

近年、ブログや Twitter, SNS 等の普及によりインターネット上に提供される情報がユーザの購買行動に大きな影響を与えるようになってきた。このような中で、商品レビュー文やブログ記事を対象に、ユーザが必要とする商品の推薦や商品開発などに利用するために、不満や要望を抽出している[1][2][3]。文献[1]では商品レビューから語彙構文パターンを利用することで購買意図を抽出している。文献[2]では評価表現ツールを用いてレビュー文に対してポジティブ/ネガティブを判断することでユーザ意見を抽出している。文献[3]では名詞のみでネガティブな意見を持つ語句とポジティブな動名詞の組み合わせを辞書とし、不満表現を抽出している。このように既存方式ではルールを定義することや、不満に関連する単語の辞書を利用することで不満を検出している。このため、レビュー文とは別の新たな性質のデータを利用することや、辞書には登録されていない曖昧性のある単語や多義語などを考慮していないので、SNS などのコミュニケーションデータでは不満抽出精度が落ちる可能性がある。この課題を解決するために、原因や要望など不満に関連付けられる他の要素を考慮することで、不満検出精度の向上を図る。そこで本稿では、不満・要望・その原因を関連付けるモデル（以下 FDC 関連付けモデル）を仮説し、実際のデータに対して仮説モデルが成立するか検証する。さらに、FDC 関連付けモデルによる不満推定を自動化するために確率モデルを適用することを検討しており、コミュニケーションデータの発話順序に着目し、不満・要望・原因の順序関係を分析することで、HMM や CRF[4]などの確率モデルの適用可能性についても検証する。

2. FDC 関連付けモデル

テキスト文を入力とした際、構文パターンにより要望を抽出し、既存辞書のポジネガ極性により不満を抽出することはできるが、不満の原因を推定する方式や、それぞれの関係性を表す手法はこれまでに研究されていない。そこで本稿では、不満や要望は単独で起こり得るものではなく、なんらかの原因により不満が発生・顕在化し、形を変えて要望となる、という一連の流れを想定し、不満・要望・原因の三要素において、各要素を相互から推定可能とする FDC 関連付けモデルを仮定する(図 1)。また、確率モデルを適用するための指標として不満・要望・原因の関連性と、発話順序における位置関係に着目する。

図 1 では SNS やブログなどのオンライン上にテキストとして現れている不満・要望・原因を上段に、オンライン上に投稿されていない潜在的な不満・要望・原因を下段に配置されている。

3. 実験

3.1 サンプルデータへのラベリング方法

本稿では企業 SNS の日記に該当するエントリとそのコメント応答内容のテキスト文 (100 エントリ 4831 発話) を対象とし、不満・要望・原因を手動でラベリングすることで分析を行う。エントリ内容は、業務内容や調査報告だけでなく、プライベートの話なども含んでいる。テキスト文は、エントリとそのコメントを句読点区切りで分割し、発話単位で時系列順に並べたものをサンプルデータとして使用する。この際、発話には時系列順に発話番号をふっておく。

ラベリング作業では、分析者 (本論文著者 2 名) がサンプルデータを読み、不満・要望・原因に関する分類項目を各 6・4・5 種類作成した (表 1)。表 1 に基づいて、サンプルデータに対して不満・要望・原因のラベルを付与した。この際、分類項目の修正を随時行った。また、各要素のラベル付与は分析者 2 名の意見が合致する発話のみとし、それ以外の発話をその他としてラベル付けた。

3.2 不満・要望・原因の関連性と位置関係

不満・要望・原因の関連性と、発話順序における位置関係に着目し、不満・要望にラベル付けされた発話に対し、その原因となる発話との関連付けを行う。次に、関連付けられた組において、不満・要望を起点とした発話順の原因

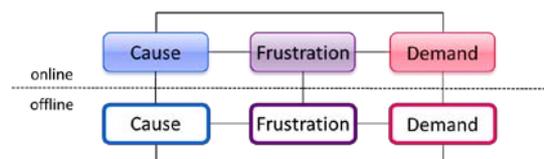


図 1 FDC 関連付けモデル

表 1 要望・不満・原因の分類項目

要素	分類項目	説明
原因	他人の意見	他人の意見が原因となるもの。
	客観的事実	客観的事実が原因となるもの。
	環境	環境が原因となるもの。
	自己主張	自己主張が原因となるもの。
	体験	体験が原因となるもの。
不満	書き手が人や環境、物事に関してもつ不満を含む発言。	
	後悔	書き手が人や環境、物事に関してもつ後悔を含む発言。
	体験による	書き手が体験によって認識した欠点についての記述。
	その他	その他のフラストレーション
要望	返答を求めている状態	読み手に対して返答を求めるような発言。要求、問いかけ
	提案	書き手がアイデアや解決策などを提案するような発言。
	未来	書き手が未来において実現出来たらいいなど希望 (期待) する事柄や願望を含む発言。
	発信	読み手に対して伝えたい事柄。注意喚起や主張を含む思い
	感情	書き手の感情。嫉妬、羨望。
	自己欲求	書き手がその時に感じた実現したい事柄、希望、意思表示

[†] (株) KDDI 研究所 KDDI R&D Laboratories, Inc.

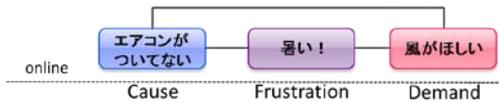


図 2a 3要素の関係例

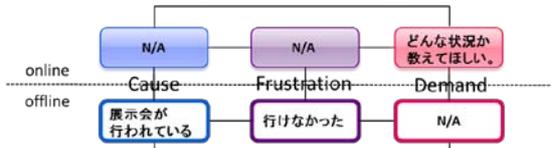


図 2b 要望のみ単独で表れている例

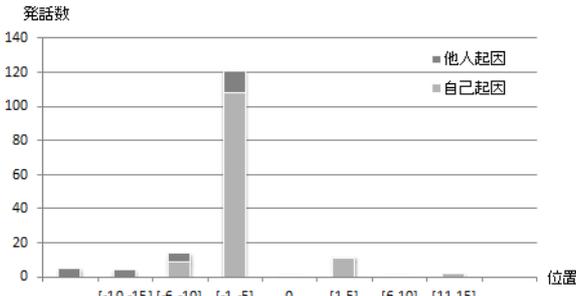


図 3a 要望を起点とした原因位置

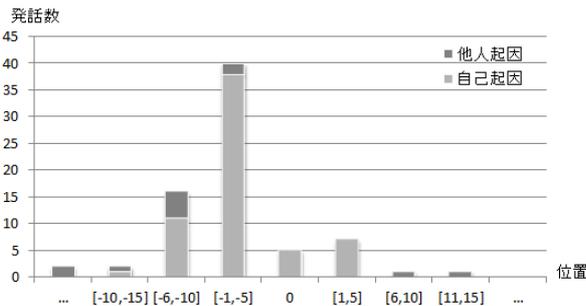


図 3b 不満を起点とした原因位置

位置を調査した。原因位置は、ラベル付けされた不満・要望とその原因の発話番号によって算出される。

3.3 結果と考察

正解データのラベルは、不満 114 件、要望 260 件、検出した不満および要望の原因 219 件、その他 4238 件となった。また、原因と関連付けられた不満は 74 件、要望 260 件であった。不満および要望と原因が関連付けられたケースを以下に示す。

- | | |
|---------------------------------|---|
| 1 : 室温 28.8℃ | |
| 2 : A 氏机上のデジタル温湿度計 29.4℃ 湿度 36% | |
| 3 : エアコンついてないね。 | C |
| 4 : これ、すごいいい情報です。 | |
| 5 : 暑い! | F |
| : : : | |
| 50 : 風がほしいです。 | D |

5 発話目は表 1 における分類項目「不満」に当てはまる不満 F であり、50 発話目は要望 D の「自己欲求」に当てはまる。さらに、3 発話目は原因 C の「環境」に当てはまり、前述の不満・要望の両方に対応付けられる。図 2a はこれらの関連付けを FDC 関連付けモデルに適用した例である。

原因が発話単位ではなくコメント全体を指すケースや、文章中に原因となる発話が存在せず、不満や要望が単独で発生しているケース (不満 38 件、要望 95 件) においては、各要素の関連付けは行うことができない。しかし、図 2b のように、エントリーやそのコメント応答には「どんな状況なのか教えて欲しい」という要望があるが、関連する不満や原因が書かれていない場合にも、前後の発話位置により不満や原因を推定でき、FDC 関連付けモデルに適用可能である。

次に FDC 関連付けモデルにおける発話順序関係について分析した。不満 F および要望 D を起点とした原因 C の位置はそれぞれ $-47(=50-3)$ 、 $-2(=5-3)$ として算出される。原因位置の値は、不満・要望より原因が前述されていれば負の数を、原因が後述されていれば正の数をとる。抽出した不満・要望に関連付けられた原因位置の集計結果を図 3 に示す。縦軸は関連付けられた件数、横軸は原因位置の値を表す。図 3a は要望を、図 3b は不満を起点としており、不満・要望の原因が自身の書き込みによるもの、他人の書き込みによるものを区別して集計した。

グラフより、全体的に原因は前述されることが多いことがわかる。特に起点から 5 発話以内で前述されている割合が不満 54.0%、要望 75.6% であり、不満では自己起因 61.3%、他人起因 16.7%、要望では自己起因 80.5%、他人起因 50.0% であり、サンプルデータでは自己起因の原因は 5 発話以内で前述される傾向があった。一方、他人起因の場合は、自己起因に比べて原因位置がより離れることが多い。さらに、原因が後述される場合は、自己起因で解決する傾向がみられた。これらの結果から、確率モデルが適用可能だといえる。

4. まとめ

本稿では、不満・要望・原因を高精度に推定することを目的とし、不満・要望・原因を関連付ける FDC 関連付けモデルを仮説し、企業 SNS 上のテキストデータを用いた検証により妥当性を確認した。また、不満および要望に関連付けられた原因の発話順序を調査し、原因は不満・要望の 5 発話以内で前述される傾向があることがわかった。この結果から、不満・要望・原因を関連付けるための確率モデルを構築できる可能性がある。

今後、今回作成した分析データをもとに CRF などの確率モデルを適用し、評価検証を行う。また、対象テキストの分割を発話単位に加え、述語項構造を利用するなど、特徴量を選定する。

参考文献

- [1] 石野 亜耶, 村上 浩司, 関根 聡, “商品レビューからの購買意図の抽出とそれを用いた商品検索システムの構築”, 言語処理学会第 20 回年次大会発表論文集, pp622-625 (2014).
- [2] 坂井 俊之, 藤村 孝, “ブログに記述された不満表現からの潜在ニーズの発見”, 情報処理学会論文誌, Vol. 52, No. 12, pp3806-3816 (2011).
- [3] 山本 竜太郎, 藤本 悠, 大原 剛三, “ユーザの重視する不満意見と好評意見を考慮した商品推薦システム”, DEIM Forum 2012 A9-2 (2012).
- [4] J.Lafferty, A.McCallun, and F.Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data”, In Proceedings of the 18th International Conference on Machine Learning (2001).