

会話データを対象にした有効な分析視点の設定と対象概念の自動抽出

Automatic Identification of Viewpoints and Expressions for Mining of Business-Oriented Conversations

竹内 広宜 那須川 哲哉 渡辺 日出雄
Hironori Takeuchi Tetsuya Nasukawa Hideo Watanabe

日本アイ・ビー・エム株式会社 東京基礎研究所
IBM Research, Tokyo Research Laboratory.

Textual records of business-oriented conversations between customers and agents need to be analyzed properly to acquire useful business insights that improve productivity. For such analysis, it is critical to identify appropriate textual segments and expressions to focus on, especially when the textual data consists of complete transcripts, which are often lengthy and redundant. In this paper, we propose a method to identify important segments from the conversations by looking for changes in the accuracy of a categorizer designed to separate different business outcomes. We extract effective expressions from the important segments to define the various viewpoints. In text mining a viewpoint defines the associations we will find and it is crucial that the correct viewpoints are identified. We show the effectiveness of the method by using real datasets from a car rental service center.

1. はじめに

近年、蓄積されたテキストデータを活用するテキストマイニング技術が研究・開発されている。特に CRM(Customer Relationship Management) の分野では蓄積された顧客の声をテキストマイニング技術を通して分析し、ビジネスに活用することが行われている [10]。従来、分析の対象となっていた顧客の声はエージェントが顧客との会話の後に会話内容を記したコールメモが中心であった。一方で、音声認識技術の向上により自動的にテキストに書き起こされた生の会話をテキストマイニングの対象にすることが可能となっており、会話データの分類などの研究がなされている [4][8]。電話による会話データの場合、音声認識の精度が課題となっているが、コールセンターにおける会話データについても特定の文書集合に注目しその中に頻出する特徴的な表現の傾向を見つめるといった分析は認識精度に対して頑健であるという報告がされている [11]。

コールセンターで行われているテキストマイニング技術を用いた分析には、コールメモに含まれる製品名や問題表現を辞書として登録し、製品名と問題表現という観点間の関係を2次元の表形式で表示するといったことがあげられている [9]。このような分析によって、ある製品のみの特徴的に出現している問題表現を概観することができ、知見につながる傾向を容易に得ることが可能となっている。

このようなコールメモの分析では各文書が対話中の重要な内容のみで比較的短く記述されている。そのため、観点(カテゴリ)の設定や各観点ごとに該当する概念(キーワード、表現)を集めた辞書の構築はサンプルデータ中に頻出する語を手がかりにすることで比較的容易に人手で行うことができた [9]。しかしながら会話データの場合、各会話のデータサイズが大きくなり冗長な表現も多く含まれる。そのため、高頻度語の情報だけから、どういった観点が分析に有効であるのかを事前に設定したり適切な辞書を準備することは難しい。例えば営業活動や問題解決のように目的と結果を伴う会話において何が成功に寄与しているかといった要因分析は、生産性の向上が期待できることから、テキストマイニングの魅力的なアプリケーションである。この要因分析においては、冗長性の高い会話の一体どこに着目すれば有益な知見の獲得につながるかの判断が重要で

あるが、分析者の勘に依存しながら試行錯誤しては効率が悪い。たとえ要因が存在しても、そこに気づけるとは限らない。

本研究では、コールセンターにおける目的を伴う会話はある事前に定義した流れに沿って進むという性質を用い、分析目的に応じて有用と思われる会話の発言箇所および特徴表現をデータから自動的に抽出する方法を提案する。そして、自動的に抽出された特徴発言箇所および特徴表現の情報を用いて結果を伴う会話データを分析することで、要因を半自動的に認識することができることを実際のデータを用いて検証する。

2. 目的をもった会話と特徴表現の抽出

本研究では企業のコールセンターなどに寄せられる目的をもった会話を対象とする。電話による商品の予約・購入といったものが例として挙げられる。このような会話データは以下の特徴がある。

- 各会話は顧客とエージェントとの間のやりとりで構成される。
- 会話の流れはある程度事前に定義されている。
- 一定数のビジネス結果が存在し、各データにはそのうちのひとつが割り当てられる。

例えば、レンタカーの予約センターでは以下のような会話の流れが定義されエージェントの研修などで用いられている。

- 開始：ブランド名および担当者名を言及する
- 予約詳細：借りる日時・場所および返却日時・場所を伺う。
- 提案：車種と料金を提案する。
- 個人情報：顧客の名および連絡先を伺う。
- 予約再確認：予約内容を再度確認する。
- 必須事項の確認：年齢、免許などについての必須事項を確認する。
- 終了：予約番号を発行し、顧客に対し感謝を述べ電話を終了する。

また、以下のコール種別が各データに結果として付与される。

- reservation(予約成立)
 - pick up: 予約後に顧客がカウンターに車を取りに来る。
 - not picked up: 予約後に顧客はカウンターに車を取りに来ない。
- unbooked(予約不成立)

このような会話の分析においては、“会話中の何が結果に影響を与えるのか”といった分析が必要となることが多い。

本研究ではそれぞれが一見すると同じような内容である会話でありながら結果が異なる会話データの要因分析において、有効な発言箇所を同定し特徴表現を抽出する手法を提案する。会話データの中から重要箇所の抽出は [3] などで行われているが、これは会話の主題を見つける手法であり、本研究で扱う同じ流れに従う会話の分析に直接用いることは難しい。また従来の特定の文書集合における特徴語の抽出方法 [1] では、特徴語のみを抽出するため会話中のどの箇所の表現が有効であるかということが具体的にわからないため、要因分析のために注目すべき観点の定義が困難となる。

以下の章では提案手法について述べ、提案手法によって観点と辞書を半自動的に構築することができ有効な分析を効果的に行うことができることを、実データを用いた実験を通して示す。

3. 会話データからの特徴発言箇所同定と特徴表現抽出

3.1 データモデル

図 1 に対話データ d_i を示す。本研究で扱う会話データは、

```

1: AGENT: Welcome to CarCompanyA. My name is Albert. How may I
help you?
2: CUSTOMER: Aah ok I need it from New York.
3: AGENT: For what date and time.
.....
j-1: AGENT: Wonderful so let me see ok mam so we have a 12 or
15 passenger van available on this location on those dates and
for that your estimated total for those three dates just 300.58$
this is with taxes with surcharges and with free unlimited free
milleage.
j: AGENT: That is fine.
.....
Mi-2: AGENT : Tehe confirmation number for your booking is 221
384 699.
Mi-1: CUSTOMER: OK ok thanks you.
Mi: AGENT: Thank you for calling CarCompanyA and you have a
great day good bye.
    
```

図 1: 会話データ (例)

顧客とエージェントとの交互のやり取りで構成され、前章で述べた性質を持つ。会話 d_i におけるやり取りの数 (turn 数) を M_i とすると各会話データは以下のように表される。

$$d_i = d_i^1 + d_i^2 + \dots + d_i^{M_i} \quad (1)$$

ここで会話の最初から j 番目の発言までを考慮したデータを考え、 $\tilde{d}_i^j = d_i^1 + d_i^2 + \dots + d_i^j$ で表す。そして m_k 番目の発言までを考慮した各会話データ ($\tilde{d}_i^{m_k}$) を集め、 D_k とする。図 2 は d_i 、 $\tilde{d}_i^{m_k}$ 、 D_k を図示したものである。そして、 m_k をいくつか設定し、図 3 に示すような時系列累積データ $D_1, D_2, \dots, D_k, D_{all}(= D)$ を作成する。例えば図 3 で D_1 は会話の先頭の発言だけを集めたデータ、 D_2 は 2 番目の発言までを集めたデータ、 D_3 は 5 番目の発言までを集めたデータとなる。

3.2 特徴発言箇所同定

本研究では結果が付随している会話を想定している。ここでは結果は 2 値 (A または not A) であるとする。時系列に沿っ

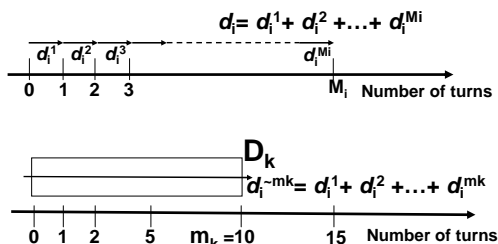


図 2: 会話データのモデル

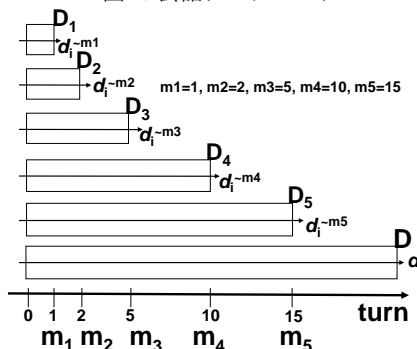


図 3: 時系列に沿って累積したデータセット

て分割された各データ集合 D_k に対してデータを学習データ (D_k^{train}) とテストデータ (D_k^{test}) に分ける。 D_k^{train} を用いて分類器を構成しその性能を D_k^{test} を用いて求める。分類器の性能は精度 [6] を用いて求める。

こうして得られた分類器の精度 $acc(categorizer(D_k))$ を時系列に沿ってプロットする。 D_k は時系列に沿って累積的にデータを分割したものであるため、 $acc(categorizer(D_k))$ は、文書全体 D に対する分類器の精度に収束する。本研究で対象とする会話データは話しの流れや内容が事前に定義されているため、どの会話もほぼ同期して会話が進んでいると考えられる。会話中の発言の中には結果に影響を与える表現が出現する一方、結果に影響を与えない表現も多く出現する。会話の流れ全体を考えた場合、結果に影響を与える発言箇所と影響を与えない箇所がある。したがって、結果に影響を与える表現を含んでいる箇所が D_k に加わった場合には $acc(categorizer(D_k))$ は増加する。一方、結果に影響を与えない表現はノイズとなるため、そのような表現を多く含む箇所が D_k に加わった場合には $acc(categorizer(D_k))$ は減少する。これを模式的にあらわしたのが図 4 である。

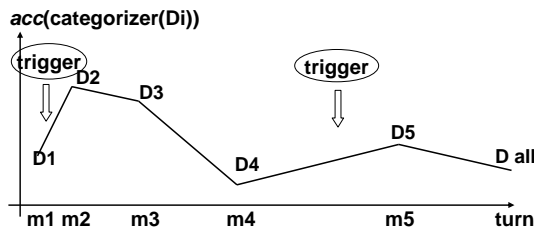


図 4: 特徴発言区間の同定

ここで分類器精度が増加する区間を trigger 区間 seg (開始地点, 終了地点) として抽出とする。図 4 の例では $seg(m_1, m_2)$ および $seg(m_4, m_5)$ が trigger 区間として同定される。trigger 区間の同定は分類器の精度の増減を元に行うため、使用する分類器には大きく依存しないと考えられる。本研究では SVM を用いた文書分類器を用いた [2]。

3.3 特徴表現抽出

trigger 区間として抽出された区間では何らかの表現が結果に影響を与えていると考えられる。本節では、trigger 区間における特徴表現の抽出を考える。trigger 区間 $seg(m_{k-1}, m_k)$ における特徴表現は D_{k-1} にデータが加わり D_k になってはじめて特定のクラス (A または not-A) の文書に多く出現する表現と定義する。文書集合における特徴表現の抽出はさまざまな方法が提案されている [1][7]。本研究では特徴的に出現する表現を抽出する尺度として χ^2 統計量を用い、 D_k 中の各 kwd に対して統計量を求める。また、注目している区間に特徴的に偏って出現している表現を抽出するため以下の尺度を用いる。

$$new(kwd) = \frac{freq_{D_{k-1}}(kwd) / m_{k-1}}{freq_{D_k}(kwd) / m_k} \times sign(freq_{D_k}^A - freq_{D_k}^{notA}) \quad (2)$$

ここで $freq_{D_k}(kwd)$ は D_k における kwd の出現頻度、 m_k は D_k の turn 数 (発言のやりとりの数)、 $freq_{D_k}^A(kwd)$ は D_k でクラス "A" に属する文書における kwd の出現頻度、 $sign(\cdot)$ は符号関数をあらわす。 D_k において kwd がクラス "A" に属する文書においてはじめて出現した場合、本尺度によるスコアは 1 以上になる。これら 2 つの尺度を組み合わせた $\chi^2(kwd) \cdot new(kwd)$ を用いて特徴表現を抽出する。この尺度によるスコアを用いることで、 D_k 以前にすでに出現している特徴的な表現を取り除くことができる。

4. 実験

4.1 データ

レンタカー会社の電話予約センターで収集された会話を人手で書き起こしたデータを実験に用いた。図 1 のデータはその一例である。会話データのうち予約が成立した 137 会話を分析の対象とした。予約成立会話には結果として顧客が車を取りに来た (pick up) および取りに来ない (not-pick up) が付与されており、実験データ中の内訳はそれぞれ 72 会話、65 会話となっている。同じ予約成立会話にも関わらず結果が異なる理由は何かという分析 (要因分析) に対して、特徴発言箇所の同定と特徴表現の抽出を実験データで試みた。

4.2 結果

$m_1=1, m_2=2, m_3=5, m_4=10, m_5=15, m_6=20$ を設定し、 D_1, \dots, D_6 および D を作成した。各会話データは、名詞、複合名詞、形容詞+名詞といった特定の名詞句および動詞を属性としたベクトルで表されている。それぞれの D_k において会話データを pick up/not-picked up に分類する分類器を作成し精度を求めた。図 5 は SVM を用いた分類器 [2] によって得られた $acc(categorizer(D_k))$ の推移を示している。この結果から

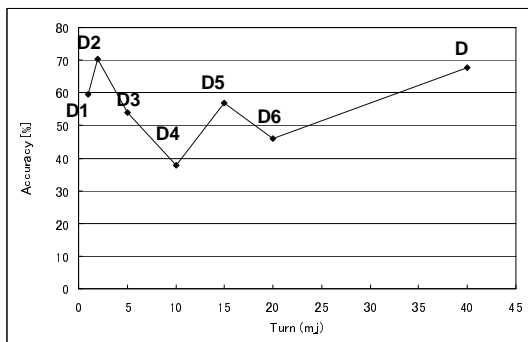


図 5: 各 D_k における $acc(categorizer(D_k))$

$seg(1, 2)$ と $seg(10, 15)$ が trigger 区間として同定される。これらの区間が会話の結果に影響を及ぼしていることになる。

各 trigger 区間ごとに 3.3 で定義した尺度を用いて特徴表現を抽出する。表 1 は各 trigger 区間における高いスコアを持つ表現を示している。抽出された結果から、顧客の最初の発言

表 1: 各 trigger 区間ごとに抽出された特徴表現

Trigger	Selected expressions	
	pick up	not picked up
$seg(1, 2)$	make, return, tomorrow, day, airport, look, assist, reservation, tonight	rate, check, see want, week
$seg(10, 15)$	number, corporate program, contract, card, have, tax surcharge, just NUMERIC dollars, discount, customer club, good rate, economy	go, impala

($seg(m_1, m_2)$) およびディスカウント表現やよい提案内容であることをアピールする表現が結果に影響を与えることがわかる。顧客の最初の発言については、実データの該当箇所を見ることで、"would like to make a reservation" や "just check the rate" といった表現が特徴表現であることがわかった。

比較のため、会話データ全体 D に対して従来の特徴語抽出手法を適用する。以下は χ^2 統計量を用いて抽出された特徴 20 語である。

corporate program, contract, counter, September, mile, rate, economy, last name, valid driving license, BRAND NAME, driving, telephone, midsize, tonight, use, credit, moment, airline, recap, afternoon

この結果からでは、ディスカウントに関連した表現が結果に影響を与えるということがわかる程度である。従来手法に比べ、本提案手法は分析観点の設定に有効であることがわかる。

4.3 テキストマイニングシステムを使った分析への適用

4.2 の結果より、顧客の最初の発言、ディスカウント表現および提案内容のアピールが結果に影響を表すことがわかった。そこで、これらを元に以下のような分析観点と該当する概念 (表現) を辞書として準備し、予約成立会話間の差異を分析するテキストマイニングシステム [5] を構築し、エージェントの生産性を改善する知見の取得を試みた。

- Customer intention at start of call: 以下の表現パターンを辞書で定義し、顧客の最初の発言から該当表現を抽出する。
 - strong start: would like to make a booking, need to pick up a car, ...
 - weak start: would like to check the rates, want to know the rate for vans, ...

strong start の顧客は実際に予約する意思を持っており、weak start の顧客は値段を調べているだけである可能性が高いという仮説を立て、それぞれの顧客タイプを **booking customer** および **rates customer** と定義する。

- Discount-related phrases: discount, corporate program, motor club, buying club といった表現をディスカウント表現として辞書に登録する。
- Value selling phrases: 以下のような提案している車種や値段が魅力的であることをアピールする表現をそれぞれ辞書に登録する。
 - mentions of good rates: good rate, wonderful price, save money, just need to pay this low amount, ...
 - mentions of good vehicles: good car, fantastic car, latest model, ...

表2は137の予約データにおける2次元の相関分析の結果であり、顧客の最初の発言内容から抽出された顧客タイプと pick up 情報 (顧客が予約した車を取りに来たかどうか) との関係を示している。この表から booking customer の 67% (47/70) が予約

表2: 顧客タイプと pick up 情報との関係

Customer types extracted from texts based on customer intent at start of call	Pick up information	
	pick up	not-picked up
booking customer (w/ strong start) (70)	47	23
rates customer (w/ weak start) (37)	13	24

した車を取りに来ている (picked up) が、一方で rates customer は 35% (13/37) しか車を取りに来ない (not-picked up) ということがわかる。この結果から、顧客の最初の発言内容から顧客が予約後に車を取りに来るかどうかを予測することが可能だということがわかる。表3は、rates customer および booking customer それぞれの場合におけるエージェントによるディスカウントの言及と pick up 情報との関係を示している。この結果

表3: ディスカウント表現の言及と pick up 情報との関係

<i>Rates customer</i>		Pick up information	
Mention of discount phrases by agents		pick up	not-picked up
yes (21)		10	11
no (16)		3	13
<i>Booking customer</i>		Pick up information	
Mention of discount phrases by agents		pick up	not-picked up
yes (40)		30	10
no (30)		17	13

より、ディスカウントに言及する方が rates customer が車を取りに来る確率が高い ($P(\text{pick-up})=0.476$) ことがわかる。value selling phrase に関しても同様の傾向が得られた。

テキストマイニングによる分析で得られた結果から以下のような知見を得ることができた。

- 最初の発言に注目することで顧客の種類を判断することができる。予約する顧客には2種類 (booking customer と rates customer) あり、それぞれ車を取りにくる傾向が異なる。
- rates customer の場合には積極的にディスカウントを提案することで予約後に実際に車を取りに来る確率を改善することができる可能性がある。

得られた知見を元にアクションアイテムを作成し業務に適用したところ、レンタル成功率 (予約した顧客が車を取りに来る比率) が約3%ほど改善し、売り上げの増加が確認された [5]。

なお、分析にあたって別途、専門家がレンタル場所、車種などの観点を設定し辞書を作成したが、これらの分析観点は本要因分析においては有効ではなかった。お客様が第一声でどう発言したかという、通常コールメモにも残さないような一見些細な情報が分別要因になっていたという意外性のある知見を得ることができたことは、生の会話を分析対象とする会話マイニングの有用性を示している。しかしながら業務に精通した専門家の知識を利用したとしても、このような些細な情報を分析観点として事前に定義するのは難しい。本提案手法では、分析に有効な観点の設定や登録すべき表現の情報が対象データから自動的に抽出される。そのため専門家でなくても分析に有効な観点の設定や表現の登録を容易に行うことができるようになると考えられる。

5. まとめ

本研究では、コールセンターなどにおけるビジネス会話はほぼ同じ流れに沿っているという性質を利用して結果に影響を与える特徴的な発言区間を同定し、特徴表現を抽出する方法を提案した。実データに提案手法を適用し、テキストマイニング

による分析に有効と思われる分析観点 (カテゴリ) および該当表現を容易に取得できることが検証した。また、構成された分析観点をを用いることでビジネスを改善することができる知見が得ることができた。実験では人手で書き起こした会話データを用いており、音声認識の結果を用いた本手法の有効性の検証が今後の課題である。

参考文献

- [1] T. Hisamitsu and Y. Niwa. A measure of term representativeness based on the number of co-occurring salient words. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 1–7, 2002.
- [2] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 133–142, 2002.
- [3] G. Mishne, D. Carmel, R. Hoory, A. Roytman, and A. Soffer. Automatic analysis of call-center conversations. In *Proceedings of ACM Conference on Information and Knowledge Management (CIKM)*, pp. 453–459, 2005.
- [4] S. Roy and L. V. Subramaniam. Automatic generation of domain models for call centers from noisy transcriptions. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (COLING/ACL)*, pp. 737–744, 2006.
- [5] H. Takeuchi, L. V. Subramaniam, T. Nasukawa, S. Roy, and S. Balakrishnan. A conversation-mining system for gathering insights to improve agent productivity. In *Proceedings of IEEE Joint Conference on E-Commerce Technology and Enterprise Computing, E-Commerce and E-Services (CEC/EEE)*, 2007.
- [6] Y. Yang and X. Liu. A re-examination of text categorization methods. In *Proceedings of the 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 42–49, 1999.
- [7] Y. Yang and J. O. Pedersen. A comparative study on feature selection in text categorization. In *Proceedings of the 14th International Conference on Machine Learning (ICML)*, pp. 412–420, 1997.
- [8] G. Zweig, O. Shiohan, G. Saon, B. Ramabhadran, D. Povey, L. Mangu, and B. Kingsbury. Automatic analysis of call-center conversations. In *Proceedings of IEEE International Conference of Acoustics, Speech and Signal Processing (ICASSP)*, pp. 589–592, 2006.
- [9] 那須川. コールセンターにおけるテキストマイニング. 人工知能学会誌, 16(2):219–225, 2001.
- [10] 那須川. テキストマイニングを使う技術/作る技術—基礎技術と適用事例から導く本質と活用法. 東京電機大学出版局, 2006.
- [11] 那須川, 宅間, 竹内, 荻野. コールセンターにおける会話マイニング. 言語処理学会第13回年次大会発表論文集, pp. 590–593, 2007.