

不要発話特定を導入した通話とコールメモ間のトピック対応付け —差分マイニングの性能改善に向けて—

田村 晃裕† 石川 開† 安藤 真一†

†日本電気株式会社 共通基盤ソフトウェア研究所

1. はじめに

近年、多くの企業が、コールセンターに蓄積されたデータを分析する事で、製品やサービスの問題、顧客ニーズなどを抽出し、製品やサービスの改善に役立っている。これまでの分析対象は、応対者により作成されたコールメモが中心だった[1]。しかし、最近では、音声認識技術の発展もあり、コールメモより多くの情報を含む、顧客と応対者間の通話を直接分析する方式の研究が盛んになってきている[2]。しかし、通話には多種多様な情報が混在するため、分析が困難である。そこで、我々は、特に、コールメモには含まれない、顧客の生の声や応対者の応対上の問題といった通話特有の情報を効率良く獲得する『差分マイニング』を提案している[3]。

差分マイニングは、通話をテキスト化した通話テキストからコールメモに書かれていない通話特有の部分（通話とコールメモの差分）を求め、この差分を対象にテキストマイニングを行う方式である。この差分マイニングを実施する上で、差分を適切に生成する事が重要である。そこで差分生成に際しては、差分を生成する前にまず、通話テキストとコールメモの各部分をトピックの同一性に基づいて対応付けし、その後トピックが同一である部分同士の比較から差分を生成する方法を提案している[4]。このようにトピックの異なる部分同士を誤って同一視しないように工夫することで、差分マイニングの性能を向上することができる。

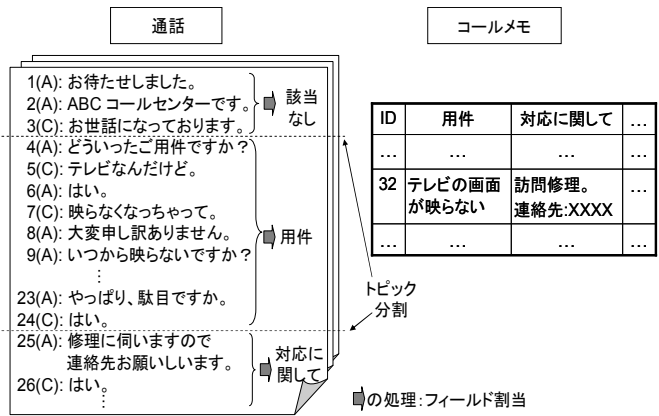
一般に、トピックが同一である通話中のセグメントは語彙的結束性を基に計算される。しかし、通話には、トピックに依存せずに出現する相槌などの冗長な発話（以降、「不要発話」と呼ぶ）が多数存在する。そのため、手がかりとなる語彙が離れてしまい、その連なりが捉えづらくなる。これが、トピック分割結果やその分割結果を用いるトピック対応付けの結果に誤りを生じる原因となっていた。

そこで、本稿では、不要発話を自動で特定し、除去した後、トピック対応付けを行う新たな手法を提案する。

以降、2 章でトピック対応付けの従来手法とその課題を説明し、3 章で提案手法を提案する。そして、4 章で提案手法をコールセンターの実データに適用して有効性を検証する。そして、5 章で本稿をまとめる。

2. 従来手法

本章では、トピック対応付けの従来手法[4]と問題点を述べる。トピック対応付けとは、通話の各発話に対し、該当するコールメモのフィールドを特定する処理である。図 1 にトピック対応付け結果の例を示す。図 1 では、通話の発話 ID4~24、ID25~が、それぞれコールメモの「用件」、「対応に関して」フィールドに該当し、ID1~3 がどのフィールドにも「該当なし」に対応している。



※左の数字は発話ID, 括弧内はA:応対者, C:顧客

図 1 通話とコールメモのトピック対応付け

通話の一発話毎に、該当するコールメモのフィールドを特定すると、一発話中の情報量が不十分なため、適切に特定できない事が多い。そこで、トピック対応付けは、

(1) 通話テキストを話題が同じセグメント毎に分割する「トピック分割」を行った後、(2) 分割された各セグメントに対し、コールメモのどのフィールドに記述されるかを判定する「フィールド割当」で実現する。このトピック対応付けは、トピック分割結果が真のセグメントより粗い程、1つのセグメントが複数のフィールド部分に対応するため、誤りを生じる。また、真のセグメントより細かい程、個々のセグメントの情報量が少なくなるため、誤りやすくなる。このように、トピック分割で、真のトピック境界の再現率を高く、かつ、細かく分割し過ぎないようにする事が重要である。そこで、本稿では、トピック分割手法に着目して課題の分析と改良を行い、トピック対応付けの性能改善を試みる。以降では、トピック分割、フィールド割当の手法を順に説明した後、従来のトピック分割手法の問題点を説明する。また、以降の説明では、 i 番目の事例の通話を D_i 、 D_i に対応するコールメモを M_i とする。そして、コールメモは b 個のフィールドを持ち $M_i = M_i^1, M_i^2, \dots, M_i^b$ とする。

2.1 トピック対応付け手法

2.1.1 トピック分割手法

通話 D_i を、 $D_i^1, D_i^2, \dots, D_i^a : D_i^k (1 \leq k \leq a)$ 内では話題が同じ、となるように分割する。越仲らの手法[5]は、テキスト中の各単語に対して、話題を指定する隠れ変数を用意し、テキスト中の各単語がどの話題に属するかを確率的に表現する。そして、テキストの生成モデルとして HMM を仮定し、トピック分割を入力テキストを用いたモデルパラメータ推定とモデル選択の問題として扱う。

また、従来手法[4]は、越仲らの手法[5]にコールセンターの通話の特徴を表したヒューリスティクスを加え、トピック分割を行う。具体的には下記の 2 ステップである。

- **ステップ1**：応対者が話題を変える際に用いる発話を検出し、その直前で分割する。検出は、予め人手で作成したパタンとのマッチングにより実現する。
- **ステップ2**：ステップ1による分割後の各セグメントに対して、越仲らの手法を適用する。その際、質問文の直後は分割境界の候補としないという制約を用いる。

2.1.2 フィールド割当手法

単純には、解析対象 D_i^k とそれに対応するコールメモの各フィールド $M_i^1, M_i^2, \dots, M_i^b$ を比較し、類似度最大のフィールドに割り当てる事で実現できる。しかし、コールメモは、応対者により、使われる単語や記述される事柄の粒度などが異なるため、個々の事例の状況により正しくフィールドを割り当てられない可能性がある。そこで、従来手法[4]は、解析対象以外の通話とコールメモも利用し、解析対象 D_i^k と類似のセグメントが、コールメモのどのフィールドに記載されるかを考慮する事で、個々の事例の状況により生じる誤りを緩和する。具体的には、下記の2ステップである。

- **ステップ1**： D_i^k と D_i^j 以外の通話に存在する全セグメントとの類似度を各々計算する。そして、その類似度が閾値 θ_1 以上となったセグメントを要素とする、 D_i^k と類似のセグメント集合 $Cluster(D_i^k)$ を求める。
- **ステップ2**：次式で該当するフィールドを決定する。

$$\arg \max_{1 \leq j \leq b} \{x \cdot \text{sim}(D_i^k, M_i^j) + y \cdot \sum_{D_x \in Cluster(D_i^k)} \text{sim}(D_i^k, D_x) \cdot \text{sim}(D_x, M_x^j) / N\}$$

ただし、 $\arg \max$ 中の評価関数の値が全て、閾値 θ_2 以下の場合、 D_i^k に該当するフィールドはコールメモに存在しないと判定する。ここで、 N は $Cluster(D_i^k)$ の要素数、 $\text{sim}(D_i^k, M_i^j)$ は D_i^k と M_i^j の単語ベクトル間のコサイン距離である。また、第一項目は解析対象のセグメントの評価値で、第二項目は解析対象と類似のセグメントに関する評価値である。また、 x と y は第一項目と第二項目の重みを調節するパラメータである。

2.2 従来手法の問題点

通話には、トピックに依存せずに出現する不要発話（図1中では「はい」や「大変申し訳ありません」など）が多く存在する。従来のトピック分割手法[4][5]では、不要発話がトピックの連続を示唆する発話の間に挿入されると、その発話同士の距離が離れるため、誤って分割される可能性がある。例えば、図1では、発話ID7と9の語彙「映らない」の連鎖は「用件」の話題が連続する事を表す有用な手がかりである。しかし、その間に発話ID8の不要発話が挿入される事で、その間で誤って分割される可能性がある。また、トピックが変化する発話箇所に同一の不要発話が存在する事で、語彙の連鎖によって正しく分割できない可能性もある。例えば、図1では、発話ID23と25の間でトピックが変化する。しかし、発話ID24と26に同一の不要発話が存在するため、正しく分割できない可能性がある。このように、不要発話の存在は、トピック分割結果に分割誤りを生じさせる。

この問題は、[4]や[5]の手法に限らず、一般的なトピック分割手法に共通する。従来、人手でストップワードを

設定する、という対処はあるが、人手による設定のコストと網羅性の問題がある。特に、コールセンターの通話は人間同士の対話であるため、全ての不要発話を網羅するのは困難である。

3. 提案手法

提案手法は、まず、自動で不要発話を特定する。その際、同時に、トピックを変える為に発せられる発話（以降、「トピック変化発話」と呼ぶ）も特定する。そして、特定した不要発話やトピック変化発話を考慮してトピック分割、フィールド割当の順で処理を行う。以降、各処理の手法を説明する。

3.1 不要発話特定手法

「はい」や「大変申し訳ありません」などのあらゆるトピックで出現する不要発話は、あらゆる通話テキストにも出現しやすい、という仮定に基づき不要発話を特定する。具体的には、「発話が出現した通話テキスト数/全通話テキスト数」で定義される文書頻度が高い発話を不要発話として特定する。ただし、コールセンターの通話は、ある程度決まった流れを持ち、応対者主導で聞き出すべき話題を変えながら対話がすすむ場合が多い。このような通話テキストでは、応対者がトピックを変える為に発する発話「トピック変化発話」（図1中の発話ID4や25など）も文書頻度が高くなる。そこで、文書頻度の高い発話の前後でトピックの変化を調べ、高い割合でトピックが変化しない発話を不要発話、高い割合でトピックが変化する発話をトピック変化発話として特定する。具体的には、下記の3ステップからなる。

- **ステップ1**：各発話の文書頻度を計算し、文書頻度が閾値 θ_3 以上の発話を高頻度発話と特定する。
- **ステップ2**：高頻度発話の各出現位置の前後でトピックが変化しているかを判定する。具体的には、予めコールメモの各フィールドを教師データとして作成した識別器を、高頻度発話の各出現位置の前 m 文と後 m 文に適用し、前後のトピックを判定し、その一致をみる。
- **ステップ3**：各高頻度発話の前後でトピックが変化する割合を「出現位置の前後でトピックが変化した数/出現数」で求める。そして、高頻度発話の内、その割合が閾値 θ_4 以下の発話を「不要発話」、閾値 θ_5 以上の発話を「トピック変化発話」と特定する。

3.2 トピック分割手法

3.1節の手順で特定した(1)不要発話や(2)トピック変化発話を考慮してトピック分割を行う。具体的には、(1)不要発話に対する処理として、通話テキストから不要発話を除く。これにより、不要発話が原因で離れていたトピックの連続を示唆する発話間の距離を縮め、トピックの連続を正しく捉えられるようにする。また、不要発話の連続により、トピックが連続すると誤って判定される可能性を除く。また、(2)トピック変化発話に対する処理として、トピック変化発話の直前で分割する。これにより、発話の連鎖の情報からは検出できない境界を補充する。以上の2つの処理を行った後、2.1.1節の従来手法[4]を適用する。

表1 トピック分割手法の性能比較

手法	精度	再現率	CoAP	平均境界数
従来	0.240	0.313	0.605	8.4
提案(不要発話のみ)	0.267	0.330	0.618	8.2
提案(トピック変化発話のみ)	0.252	0.386	0.620	9.8
提案	0.272	0.400	0.623	9.6

3.3 フィールド割当手法

3.2節の手順で分割した各セグメントに対して、2.1.2節の従来手法[4]でフィールド割当を行う。ただし、フィールド割当の際にも、不要発話は手がかりとして重要ではないと考え、通話テキストから不要発話を除く。

また、従来手法[4]では、単語ベクトルの要素として、各次元の単語の tf 値を用いた。しかし、単語の中には、フィールドの識別に寄与するものとしなないものがある。そこで、提案手法では、各単語に対し、コールメモの各フィールドに対する尤度比を、そのフィールドへの寄与度として計算し加味する。具体的には、コールメモの各フィールドの単語ベクトルの要素を「tf×各フィールドに対する尤度比」の値とする。また、コールメモと通話テキストでは使用される単語が異なるため、コールメモでは特定フィールドに特徴的に出現するが、通話テキストではあらゆるトピックで出現する単語が存在し、悪影響を及ぼす可能性がある。例えば、コールメモの「対応に関して」フィールドには、顧客の希望が書かれるため、単語「お願い」が特徴的に出現するが、通話テキストでは、オペレータや顧客により、あらゆるトピックで使われる。そこで、このような単語を手でストップワードとして設定する（単語ベクトルの要素から除く）。

4. 評価実験

4.1 実験データ

実験データは、国内コールセンターにおけるコールメモ 4,596 件と、各コールメモに対応する顧客と応対者間の通話音声を書き起こした通話テキストを用いた。コールメモは、「顧客情報」、「用件」、「対応に関して」の情報を記述する3つのフィールドを用いた。

4.2 トピック分割手法の性能評価

評価セットは、実験データより無作為に抽出した 40 事例から人手で作成した。具体的には、各事例の通話テキストの各発話に対して、その発話の情報がコールメモの3つのフィールドの内、どこに該当するか、或いは、どこにも該当しないかを判定し、該当するフィールドが変化する発話の間をトピック境界とした。1 通話の平均境界数は 6.5 である。この評価セットに対して、下記のトピック分割手法を適用した結果を比較する。なお、各手法では単語は自立語のみを使用する。

トピック分割手法

- 従来：2.1.1節で説明した従来手法[4]である。
- 提案（不要発話のみ）：2.1.1節の従来手法[4]に加え、3章で説明した不要発話の特定及び除去を行う手法である。なお、3.1節のステップ1で各発話の文書頻度を求める際、発話ごとの微細な表現の違いを吸収するた

表2 トピック対応付け手法の性能比較

トピック分割		従来		提案		
フィールド割当	不要発話除去	×	×	○	○	○
	単語ベクトル	tf	tf	tf	tf×尤度比	tf×尤度比+stop word
正解率		0.574	0.594	0.598	0.603	0.623

め、発話を形態素解析し付属語を除き、自立語の原形の連続に変換して、統計量を計算する。また、ステップ2で行うトピックの判定は、SVM[6]と One-versus-Rest 法を組み合わせて、コールメモのフィールドのいずれかに分類する多値分類問題として解く。素性は自立語の表層と品詞の組み合わせを用いた。また、各パラメータは、 $\theta_3=0.5$ 、 $m=3$ 、 $\theta_4=0.5$ を用いた。

- 提案（トピック変化発話のみ）：2.1.1節の従来手法[4]に加え、3章で説明したトピック変化発話の特定及びその直前での分割を行う手法である。なお、3.1節のステップ1と2の処理は「提案（不要発話のみ）」と同じである。また、ステップ3の θ_5 は0.8を用いた。
- 提案：3章で提案した提案手法である。なお、本手法の設定は、「提案（不要発話のみ）」や「提案（トピック変化発話のみ）」と同じである。

評価は、精度（手法が特定したトピック境界の内、正しい割合）、再現率（真のトピック境界の内、手法が検出した割合）、CoAP[7]（k 発話離れた2つの発話が、同一セグメント或いは異なるセグメントかを正しく判定する確率）で行う。k は、[7]に倣い、1セグメントの平均発話数の半分の値（ $k=5$ ）を用いた。結果を表1に示す。

表1より、3章の手順で特定した不要発話やトピック変化発話を考慮する事で、トピック分割性能を改善できる事が分かる。また、トピック変化発話の考慮は、境界数を増やし、特に再現率の向上に寄与し、一方、不要発話の考慮は、境界数を保ったまま、精度と再現率をバランスよく改善できる事が分かる。

4.3 トピック対応付け手法の性能評価

4.2節の実験と同じ評価セットに対し、4.2節のトピック分割手法（従来と提案）を行った後、2.1.2節や3.3節のフィールド割当手法を適用した結果を比較する事で、トピック対応付け全体の性能比較を行う。なお、フィールド割当手法の各パラメータは、 $\theta_1=0.5$ 、 $\theta_2=0.1$ 、 $x=1$ 、 $y=1$ を用いた。

評価は、各発話のトピック対応付け結果に対する正解率（＝フィールドが正しく割り当てられた発話数／全発話数）で行う。結果を表2に示す。なお、表2において、不要発話除去の行は、「○」がフィールド割当において不要発話除去を行い、「×」が行わない事を示す。また、不要発話除去「×」で単語ベクトルの要素「tf」が従来フィールド割当手法であり、不要発話除去「○」で単語ベクトルの要素「tf×尤度比+stop word」が提案のフィールド割当手法である。

表2より、提案したトピック分割手法を用いる事で、トピック対応付け結果が向上する事が分かる。これは、トピック分割結果が改善され、真のトピックのセグメントに近い単位で、フィールド割当を行う事ができるよう

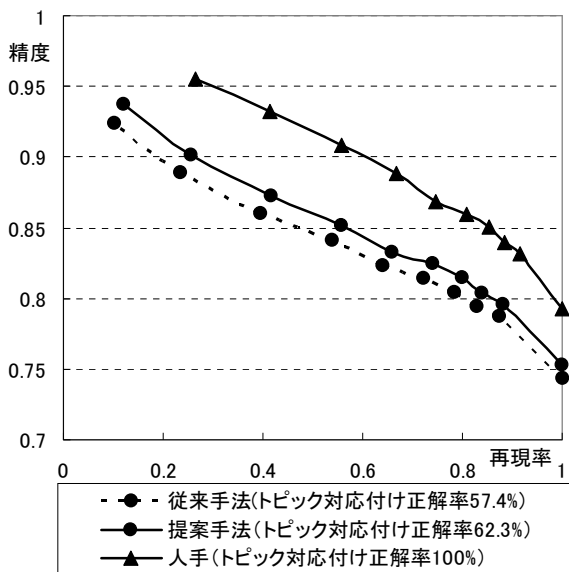


図2 差分の品質評価

になったためである。また、フィールド割当て時の不要発話除去は、それ程効果がない事が分かる。これは、不要発話のほとんどがコールメモに存在しないため、通話中のセグメントとコールメモの各フィールドとの類似度の計算には、不要発話の悪影響が少ないためと考えられる。また、単語ベクトルの要素として「tf×尤度比+stop word」の提案手法が最も正解率が良い事が分かる。これより、3.3節の単語ベクトルの改良の有効性も確認できる。

4.4 差分生成手法の性能評価

本節では、提案手法により改善されたトピック対応付け結果を用いる事で、生成する差分の質が向上するかを検討する。なお、差分生成手法は、語の共起関係に基づく差分生成法[4]を用いた。具体的には、通話テキストとそれに対応するコールメモで同一トピック部分同士を比較し、全事例を通して所定の閾値以上共起する関係にある単語を共通部分と判定し、残りを差分とする方法である。

評価データは、4.2節の実験と同じ40事例を用いた。この評価データに対して、通話テキストとそれに対応するコールメモを比較し、通話テキストの中でコールメモに書かれていない部分を特定した。そして、その部分に含まれる自立語リストを、差分の正解データとした。

各差分生成法により、差分と判定された自立語に対して、精度と再現率(40事例の平均値)をプロットした結果を図2に示す。縦軸は精度、横軸は再現率を表す。なお、図2は、共通部分と差分とを分ける前記閾値を変えて、再現率-精度曲線を描いた。また、丸印の点線は、従来手法によるトピック対応付け結果を用いた時の差分生成の性能であり、丸の実線は、3章の提案手法によるトピック対応付け結果を用いた時の差分生成の性能である。また、三角の実線は、人手で行った100%正解率のトピック対応付け結果を用いた時の差分生成の性能である。

図2より、提案手法によるトピック対応付け結果を用いた差分の方が、従来手法によるトピック対応付け結果

を用いた差分よりも質が良い事が分かる。これは、提案手法によりトピック対応付けの結果を改善する事で、誤って同一トピック同士と判定される箇所が減少し、差分生成の際に誤った箇所同士の比較が少なくなるため、差分の質の向上につながったと考えられる。

5. おわりに

本稿では、通話の各発話に対し、該当するコールメモのフィールドを特定するトピック対応付け手法に関して、発話の文書頻度を利用して不要発話及びトピック変化発話を自動で特定し、それらを利用する新たな手法を提案した。そして、コールセンターの実データを用いた実験を通じて、提案手法により、不要発話を除去する事で、トピック分割の手がかりとなる語彙の連なりを捉える事ができるようになり、また、トピック変化発話の直前で分割する事で、発話の連鎖の情報からは検出できないトピック境界を補充でき、トピック分割の性能が向上する事を確認した。また、トピック分割結果を用いるトピック対応付けや差分生成の結果が改善する事から、提案手法の有効性を示した。

本稿では、差分マイニングの性能改善のためにトピック分割手法を改良したが、複数トピックを含むテキストの情報検索など、その他の分析においても、トピック分割が有効である事が知られている[8]。したがって、本稿の手法が、そのような分野にも有効であると期待できる。

今後は、音声認識誤りを含む通話音声データに対しても同様の効果が得られるかを確認する予定である。また、質の向上した差分を用いる事で、その後の分析の精度がどの程度向上するかを確かめる予定である。

参考文献

- [1] 那須川 哲哉, “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, Vol.16 No.2, 219-225, 2001.
- [2] H. Takeuchi, L.V. Subramaniam, T. Nasukawa, and S. Roy, “Automatic Identification of Important Segments and Expressions for Mining of Business-Oriented Conversations at Contact Centers”, In *Proceedings of EMNLP-07*, 458-467, 2007.
- [3] 田村 晃裕, 石川 開, 安藤 真一, “コールセンターのコールメモと通話を対象とした差分マイニング”, 第7回情報科学技術フォーラム(FIT 2008), 2008.
- [4] 田村 晃裕, 石川 開, 安藤 真一, “トピック分割を用いた通話とコールメモの差分生成法の提案”, 言語処理学会第15回年次大会(NLP2009), 2009.
- [5] 越仲 孝文, 奥村 明俊, 磯谷 亮輔, “HMMの変分ベイズ学習によるテキストセグメンテーション及びその映像インデキシングへの応用”, 電子情報通信学会論文誌, J89-D(9), 2113-2122, 2006.
- [6] V. Vapnik, “Statistical Learning Theory”, Wiley-Interscience, 1998.
- [7] D. Beferman, A. Berger, and J.D. Lafferty, “Statistical Models For Text Segmentation”, *Machine Learning*, 177-210, 1999.
- [8] David M. Blei, and Pedro J. Moreno, “Topic Segmentation with an Aspect Hidden Markov Model”, In *Proceedings of SIGIR-01*, 343-348, 2001.