

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

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

コールセンターは、企業にとって顧客と直接コンタクトする窓口であるため、近年、様々な企業が、顧客の声を獲得する絶好の場としてコールセンターを重要視している。実際、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、潜在リスク、顧客ニーズや評価などを抽出し、コールセンターの応対や、製品、サービスの改善を試みている。

これまで、コールセンターにおける分析対象は、定型データやエージェントにより作成されたコールメモが中心であった[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んでいる、顧客とエージェント間の通話を直接分析対象とする方式の研究が盛んになってきている[2]。

このような状況で、我々は、コールメモには含まれない、顧客の生の声やエージェントの応対上の問題といった通話特有の情報に絞り込んで分析する、差分マイニングの方式を提案してきた[3]。これは、通話テキスト（通話をテキスト化したもの）からコールメモに書かれていない通話特有の部分を、通話とコールメモの情報の差分として求め、この差分を対象にテキストマイニングを適用する方法であり、通話テキスト全体を分析する従来手法では得られない事例が、絞込みの効果によって得られることが分かってきた。

しかし、同提案方式で生成される差分は、人手で作成される差分と比較して精度が低いという問題があった。その一因として、様々なトピックを含む冗長な通話全体を一括でコールメモと比較するため、トピックの異なる部分間での無用な比較からノイズが生じ、差分の計算に悪影響を及ぼしていたことが考えられる。

そこで、本稿では、通話中のトピックを考慮した、精度よい差分生成法を提案する。具体的には、まず、通話をトピック分割し、分割したセグメントに対応するコールメモの部分特定する。そして、それらに対応する部分同士から差分を生成する。このように、差分生成の際にノイズとなる、対応関係のないコールメモの部分を予め除外する事で、良質な差分の生成を目指す。

2. 従来手法

本章では、従来の差分生成法とその問題点を述べる。

最も単純な差分生成法として、単語の表層が一致する部分を共通部分として除き、差分を生成する手法が考えられる。しかし、コールメモと通話間には、多くの同義表現が存在するため、この単純な手法では正しい差分が生成できない。そこで、この同義語問題を解決するため、通話とコールメモ間での単語の共起関係を基に共通部分を判定し、差分を生成する差分生成法を提案してきた。

2.1 共起関係に基づく差分生成法

本節では、従来手法である、共起関係に基づく差分生成法について簡単に説明する（詳細は[3]を参照）。ここ

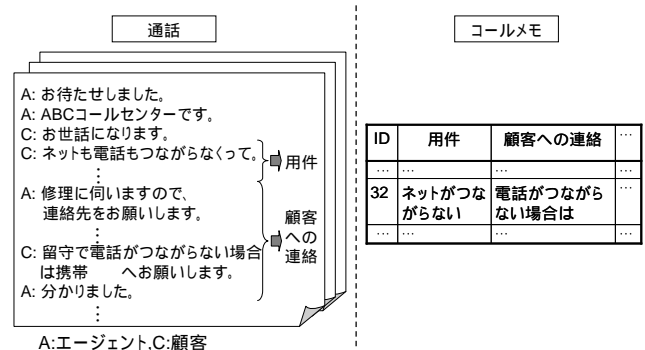


図 1 通話とコールメモの例

では、差分生成の対象データを n 個の事例（通話対応）とし、各事例は、通話テキストとその通話に対するコールメモの 2 つのテキストを持つものとする。また、 i 番目の事例の通話テキストを D_i 、コールメモを M_i 、通話テキストの集合を $D=\{D_1, \dots, D_n\}$ 、コールメモの集合を $M=\{M_1, \dots, M_n\}$ とする。そして、 i 番目の事例に関する差分 $D_i - M_i$ を生成する処理を例に取り、従来の差分生成法を説明する。

本手法は、通話とコールメモの組 (D_i, M_i) の集合 $(i=1, \dots, n)$ において、一定以上共起する関係にある単語を共通部分と判定する。 D_i 中の単語 v が、差分 $D_i - M_i$ の構成要素かどうかを判定する手順は以下の通りである。

(ステップ 1)

通話とコールメモの組 (D_i, M_i) の集合 $(i=1, \dots, n)$ に対して、単語 v, w の出現に関する相互情報量 $I(v; w)$ を次のように計算する。

$$I(v; w) = \sum_{x \in \{v \in D_i, v \notin D\}} \sum_{y \in \{w \in M_i, w \notin M\}} p(x, y) \log \frac{P(x, y)}{P(x)P(y)}$$

ここで、 x は、単語 v が D_i 中出现する、または、出現しないという状態を取り、 y は、単語 w が M_i 中出现する、あるいは、出現しないという状態を取る。また、 $P(x)$ 、 $P(y)$ 、 $P(x, y)$ は、各状態の出現確率であり、全 n 事例における各状態の出現頻度から最尤推定によって求める。

(ステップ 2)

D_i 中の単語 v の差分スコアを次のように計算する。

$$D - \text{Score}(v, M_i) = \exp \left[-\beta \sum_{w \in M_i^N} I(v; w) \right]$$

ここで、 β は正の値を持つパラメータを表す。また、 M_i^N は、 M_i 中の単語の内、相互情報量 $I(v; w)$ の大きい上位 N 個の単語集合である。この差分スコアが、閾値 θ 以上である場合、単語 v を差分 $D_i - M_i$ の構成要素とする。

2.2 従来手法の問題点

本節では、例を用いて従来手法の問題点を説明する。通話とコールメモの例を図 1 に示す。図からわかるように、通話には様々なトピックが含まれる。また、コールメモは、図 1 中の「用件」や「顧客への連絡」のように、応対を通じて聞き出す情報が予めフィールドとして定められている場合が多い。そして、エージェントは、通話

から得られた情報を、該当するフィールド内へ記述する事でコールメモを作成する。

ここで、図1における通話中の「ネットも電話もつながらなくなって」という発言の単語「電話」に着目する。この「電話」は、「電話が繋がらない」という「要件」の情報であり、この情報は、コールメモには記載されていないため、本来、差分となるべきである。

しかし、2.1節で述べた通り、従来手法は D_i 中の単語 v が差分かどうかを、 M_i 中の全単語と比較して判定する。したがって、従来手法では、コールメモの「顧客への連絡」部分の「電話」と、表層的に一致するため、差分と判定されない可能性が高い。また、従来手法は、単語の共起関係に基づいて差分を判定するが、単語 v と同一内容とはならないフィールド中の単語も考慮するため、相互情報量を求める際、及び差分を計算する際の両方に悪影響を及ぼすことが予想される。

3. 提案手法

本章では、2.2節で述べた問題点を解決する手法を提案する。提案手法は、まず、通話中の発言のトピック解析を行う。ここで、「トピック解析」とは、(1)通話をトピック分割する「トピック分割」を行った後、(2)分割された各セグメントが、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「フィールド割り当て」を行う処理であるとする。ただし、通話中の各トピックは、コールメモのいずれかのフィールド1つに割り当てられるか、いずれにも割り当てられないものとする。そして、トピック解析後、通話の各セグメントと、それに対応するコールメモのフィールド部分とから差分を生成する。以降の節より、各処理の具体的手法を提案する。

提案手法を図1に適用すると、通話中の「ネットも電話もつながらなくなって」という発言を含むセグメントが、コールメモの「要件」のフィールドに記述されるセグメントと解析できる。そして、「ネットも電話もつながらなくなって」内の「電話」は、コールメモの「要件」の部分とのみ比較され、差分が否かが判定されるため、提案手法では、正しく差分として生成されると期待できる。

3.1 トピック解析手法

3.1.1 トピック分割手法

通話を意味的なまとまり(話題)ごとに分割する。すなわち、 $D_i = D_i^1, D_i^2, \dots, D_i^a$: $D_i^k (1 \leq k \leq a)$ 内では話題が同じとなるようなセグメントに分割する。

本稿では、越仲らの手法[4]をベースにトピック分割を行う。越仲らは、テキストの生成モデルとしてHMMを仮定し、テキストセグメンテーションを、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定式化する。越仲らの手法は、予め人手で調整、設定するモデルパラメータが存在しないため、実システムへの適用が容易である。また、今回分析対象の通話は、分割数が通話ごとに異なるため、分割数を予め設定できないという状況の中で、越仲らの手法は、分割数の自動推定が可能である。以上の理由より、この手法をベースとする。

越仲らの手法を単純に通話へ適用した場合、以下のコールセンターの通話にみられる特徴を利用できない。

- 質問の発言の後には、その質問に対する発言がされるため、質問文の後では分割されない。
- エージェントは、必要な情報を得るための決まった発言を行う事があり、その発言により話題が変わる。

そこで、以下のようにトピック分割を行う。

(ステップ1)

エージェントが話題を変える際に用いる発言を検出し、その直前で分割する。検出は、予め人手で作成したパター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。

3.1.2 フィールド割当手法

分割した各セグメント D_i^k の話題が、コールメモのどのフィールドに記述されるものかを判定する。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。また、 D_i^k の話題が M_i^j に記述される時、 $field(D_i^k) = j$ とする。 D_i^k の話題に該当するフィールドがコールメモにない場合、 $field(D_i^k) = \phi$ (空集合) とする。

最も単純な手法として、次式(1)の通り、 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、フィールドとの類似度を基に、 $field(D_i^k)$ を決定する手法が考えられる。

$$field(D_i^k) = \arg \max_{1 \leq j \leq b} sim(D_i^k, M_i^j) \dots (1)$$

ただし、 $sim(D_i^k, M_i^j)$ が全て、予め設定した閾値 γ_1 以下の場合、 $field(D_i^k) = \phi$ とする。また、 $sim(D_i^k, M_i^j)$ 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。

しかし、コールメモ内で使用される単語は、通話中で使われた単語とは限らない。また、エージェントの書き忘れ等が原因で、本来ならば、そのフィールドに残すべき情報が含まれていない場合も存在する。そのような場合、前記の単純な手法では、本来書かれるべきフィールドとの類似度が高くなり、正しく $field(D_i^k)$ を決定できない可能性がある。

そこで、解析対象 D_i^k 以外のデータも用いて、 D_i^k 以外の通話に存在する D_i^k と同じようなセグメントが、コールメモのどのフィールドに記載されているかという傾向を考慮する。こうすることで、 D_i^k という個々に特殊な状況により生じる解析誤りを緩和し、精度のよい解析を目指す。提案手法の具体的アルゴリズムを下記に示す。

(ステップ1)

D_i^k 以外の通話に存在する D_i^k と類似のセグメントの集合 $Cluster(D_i^k)$ を求める。具体的には、 D_i^k 以外の通話に存在する全セグメントと D_i^k との類似度を各々計算し、閾値 γ_2 以上のものを $Cluster(D_i^k)$ の要素とする。

(ステップ2)

次式(2)の通り $field(D_i^k)$ を決定する。

$$field(D_i^k) = \arg \max_{1 \leq j \leq b} \{ x \cdot sim(D_i^k, M_i^j) + y \cdot \sum_{D_x \in Cluster(D_i^k)} sim(D_i^k, D_x) \cdot sim(D_x, M_x^j) / N \} \dots (2)$$

ただし、 $\arg \max$ 中の評価関数の値が全て、予め設定した閾値 γ_3 以下の場合、 $field(D_i^k) = \phi$ とする。また、 N は $Cluster(D_i^k)$ の要素数であり、関数 sim は、式(1)と同じである。

表1 トピック解析手法の性能比較

| トピック分割手法 | 提案手法 | | | | 越仲らの手法 | | | 発話単位 | | | 用件 |
|-----------|-------|--------------|-------|-------|--------|-------|-------|-------|-------|-------|-------|
| フィールド割当手法 | 人手 | 提案手法 | 類似度 | SVM | 提案手法 | 類似度 | SVM | 提案手法 | 類似度 | SVM | |
| 正解率 | 0.794 | 0.574 | 0.558 | 0.493 | 0.511 | 0.508 | 0.444 | 0.533 | 0.336 | 0.356 | 0.384 |

式(2)において、一項目が判定対象のセグメントの評価値で、二項目が判定対象に類似したセグメントの傾向を表す評価値である。そして、 x と y は一項目と二項目の重みを調節するパラメータである。

3.2 差分生成法

通話内の各セグメント D_i^k と、それに対応するコールメモのフィールド $field(D_i^k)$ とから、次式のように差分を計算する。なお、いずれのフィールドにも割り当てられなかった通話内のセグメントは差分となる。

$$\text{差分}(D_i - M_i) = \bigcup_{1 \leq k \leq a} \text{差分}(D_i^k - M_i^{field(D_i^k)})$$

差分 $(D_i^k - M_i^{field(D_i^k)})$ を求める手法は、2.1 節の従来手法で実現する。

このように、差分生成の際にノイズとなる、 D_i^k の話題が記述される事のないコールメモのフィールド、を予め除外し差分を生成する事で、良質な差分の生成を目指す。

4. 評価実験と考察

本章では、3 章で提案した各提案手法の有効性を実験を通じて確認する。実験データとして、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エージェント間の通話音声を書き起こした通話データを用いた。コールメモは、「顧客情報」、「用件」、「対応に関して」の情報を記述するフィールドがある。

以降の実験では、差分となった付属語の分析は価値があまりないと考え、単語は、自立語のみを使用する。

4.1 トピック解析に関する実験

4.1.1 実験設定

3.1 節で提案したトピック解析手法の有効性を検証する。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、各事例の各発話に対して、その発話の情報が 3 つのコールメモのフィールドの内、どこに記述されるものか、或いは、どこにも該当しないかを判定した。

この評価データに対して、下記のトピック分割手法（提案手法と 2 つのベースライン）を適用した後、フィールド割当手法（提案手法と 3 つのベースライン）を適用した結果を比較する事で、トピック解析の性能を評価する。評価は、各発話のトピック解析結果に対する正解率（＝コールメモの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。つまり、フィールド割当手法がセグメントに割り当てた結果を、そのセグメント内の各発話に対する結果と捉え、発話単位で評価する。

トピック分割手法

- 提案手法：3.1.1 節で提案した手法。
- 越仲らの手法：提案手法では、話題が変わる発話の検出や、質問直後は分割させない等の改良を行ったが、それらを行わずに越仲らの手法[4]をそのまま適用する。
- 発話単位：一つの発話を一つのセグメントとする。

フィールド割当手法

- 提案手法：3.1.2 節で提案した手法。本手法における各パラメータは、 $x=1$ 、 $y=1$ 、 $\alpha_2=0.5$ 、 $\alpha_3=0.1$ を用いる。
- 類似度：3.1.2 節冒頭の、対応するコールメモの各フィールドとの類似度を基にした式(1)による手法。パラメータは、 $\alpha_1=0.1$ を用いる。
- SVM：本手法は、判定対象以外の通話とコールメモの情報を利用する手法のベースラインとする。

SVM[5]と One-versus-Rest 法を組み合わせ、通話中の各セグメントを、コールメモのフィールド（「顧客情報」、「用件」、「対応に関して」）と「該当なし」のいずれかに分類する多値分類問題として解く。具体的には、まず、学習フェーズにおいて、4,596 件のコールメモの各フィールドを用いて、それらのフィールドを識別する分類器を作成する。あるフィールドの分類器は、そのフィールド内を正例とし、その他の 2 つのフィールドを負例として学習する。テストフェーズでは、通話中の各セグメントに対して、作成した 3 つの分類器を適用して、コールメモのフィールドを割り当てる。素性は、自立語の表層と品詞の組み合わせを用いた。

- 人手：分割されたトピックに対して、人手でフィールドを割り当てる。この際、トピック分割を自動で行うと、セグメント内に複数のフィールドに対応する情報が混在する場合がある。その場合、最もセグメントを占有しているフィールドに割り当てた。

4.1.2 実験結果と考察

上記の各手法によりトピック解析を行った結果を表 1 に示す。表 1 内の「用件」は、評価データの発話に割り当てられるフィールドとして最も多かった「用件」を、全発話に割り当てた結果である。

表 1 より、「人手」を除くと、トピック分割手法、及び、フィールド割り当て手法、各々において提案手法が最も性能がよい事が分かる。そして、それらの提案手法の組み合わせが、最も精度良く自動でトピック解析できる手法である事が分かる。

トピック分割手法の中で、提案手法が発話単位よりも性能が良い事より、同じ種類の情報を一つのまとまりとして扱い、フィールドを割り当てる事が有効である事が分かる。これは、トピック分割の結果が、その後のフィールド割当の単位として扱われる事を考えると、同一セグメント内の他の発話の情報に助けられて、正しいフィールドが割り当てられる発話が存在するためであると考えられる。また、トピック分割手法で越仲らの手法の性能が十分でない理由として、一通話平均分割数が 3.03 と、実際の分割数 6.5 より少なく、粒度の粗い分割となっていた事が考えられる。本来分割されるべき場所で分割されていないと、混同された部分は必ず誤りとなってしまふ。そのため、越仲らの手法では、十分な性能が出ていないと考えられる。以上より、トピック解析にはセグメントの適切な単位の設定が必要である事が分かる。

また、フィールド割当手法の中で提案手法が最も有効である事より、3.1.2 節で説明した提案手法の核である、解析対象と類似したセグメントが、コールメモ中のどの

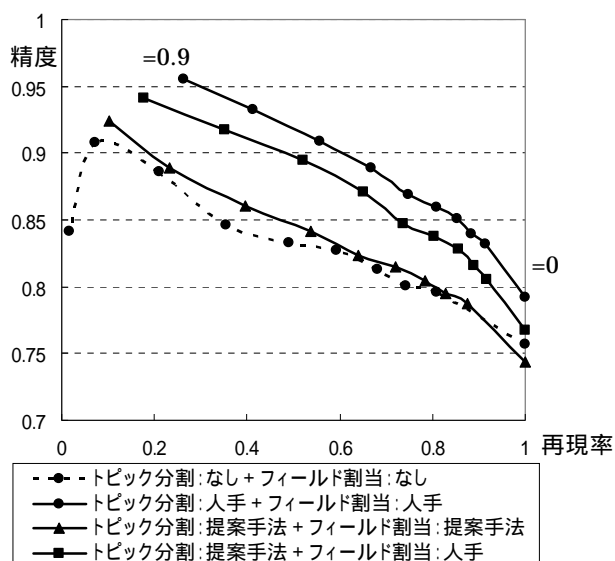


図1 差分の品質評価

フィールドに記載されているかという傾向を考慮する事が有効である事が分かる。また、フィールド割当手法としてSVMが最も性能が悪い事が分かる。SVMによるフィールド割当は、コールメモから生成した分類器を通話に適用する事で実現している。しかし、コールメモで使用する単語の空間と、通話で用いられる単語の空間が異なっていたため、SVMによる手法では、適切にフィールドを識別できなかったと考えられる。

4.2 差分生成に関する実験

4.2.1 実験設定

3章で提案した、トピック解析を行う差分生成法の有効性を実験を通じて確認する。評価データは、4.1節で使用したデータを用いた。その評価データに対して、通話データ D_i とコールメモ M_i を比較しながら、 D_i の中で、 M_i に書かれていない部分を特定した。そして、形態素解析を行い、その部分に含まれる自立語リストを生成し、それを、差分 $D_i - M_i$ の正解データとした。

各差分生成法により、差分と判定された単語に対して、精度と再現率(40事例の平均値)をプロットした結果を図1に示す。縦軸は精度を示し、横軸は再現率を表す。図1において、丸印の点線は、トピック解析を行わない従来の差分生成法の評価結果である。また、丸印の実線は、トピック解析を手で行う、つまり、トピック解析精度が100%で実現できた時の性能である。三角印の実線は、4.1節で実験した結果、自動でのトピック解析性能が最も良かった提案手法の組み合わせによりトピック解析を行った差分生成法の評価結果である。

差分生成の際の各パラメータは、経験的に $\beta=100, N=3$ を使用した。また、 θ は、0から0.9までの範囲で、0.1刻みで値を変えて差分を作成した。図1において、右から順に0から0.9が対応する。

4.2.1 実験結果と考察

図1の丸印の二つの線を比較すると、トピック解析が正しく行われれば、トピック解析後に差分を生成するとい

う提案手法の枠組みにより、差分生成の性能を大幅に改善できる事が分かる。これより、提案手法の枠組みが有効であると結論付ける事ができる。

しかし、トピック解析を自動で行うと、差分生成の性能は、従来手法とあまり変わらない事が分かる。また、フィールド割当を手で行うと、差分生成の性能が向上する。4.1.2節で自動でのトピック解析の精度が57.4%、フィールド割当を手で行った場合が、79.4%である事を考えると、トピック解析精度が低いため、自動でのトピック解析による提案手法の有効性が見られなかったと考えられる。すなわち、フィールド割当に誤りが生じると、本来同一トピックの部分間に絞って差分生成を行っていたものが、誤った部分との間で差分生成が行われることになり、差分生成に悪影響を及ぼしていると考えられる。以上より、差分生成において提案手法の効果が現れない要因として、トピック解析精度の低さが考えられる。トピック分割、及び、フィールド割当手法を改良し、トピック解析精度を向上させる事が今後の課題である。

5. おわりに

本稿では、通話とコールメモとの差分を生成する新たな差分生成方法を提案した。具体的には、通話をトピック分割し、分割された各セグメントに対して、コールメモのどの部分に記述されるかを対応付ける。そして、通話の各セグメントと対応するコールメモのフィールド部分から差分を生成する手法である。また、その要素技術である、トピック分割手法、フィールド割当手法を具体的に示した。特に、フィールド割当手法として、解析対象のセグメント以外に、解析対象と同じようなセグメントがどのフィールドに書かれるか、という傾向を考慮する事で、精度向上をはかる新たな手法を提案した。

コールセンターの実データを用いた実験を通じて、提案した差分生成方式の有効性を示した。また、トピック解析精度が改善できれば、提案した差分生成方式の性能が更に上がり、従来の差分生成方式の性能に比べて、良質な差分を生成できる事を示した。

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] Takeuchi, H., Subramaniam, L. V., Nasukawa, T. and Roy, S., “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] 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.
- [5] Vapnik, V. *Statistical Learning Theory*. Wiley-Interscience, 1998.