

# コールメモを利用したコールセンタ向け音声対話要約方式の提案

田村 晃裕† 石川 開† 西光 雅弘†  
 †日本電気株式会社 情報・メディアプロセッシング研究所

## 1. はじめに

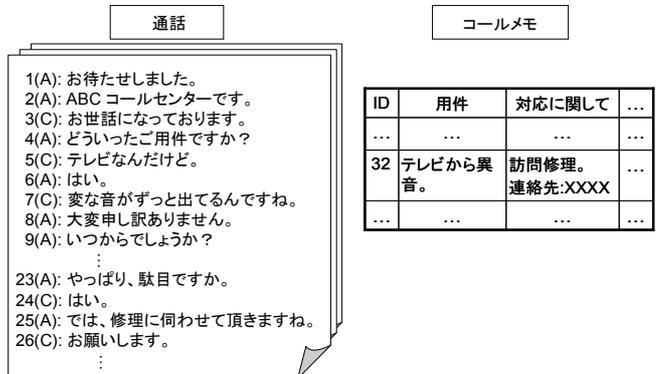
コールセンタは、企業にとって顧客と直接コンタクトする窓口であるため、近年、様々な企業において、顧客の声を獲得する場として重要視されている。実際に、多くの企業が、コールセンタでの顧客とのやりとりから、製品やサービスの問題、潜在リスク、顧客ニーズなどを把握し、製品やサービスの改善、顧客満足度の向上を進めている。コールセンタでは、対応内容を管理、分析する為に、対応者が対応内容をまとめたコールメモを作成する(図1に通話とそのコールメモの例を示す)。しかし、多くの場合、対応者が記憶や通話中の簡単なメモに基づいて作成するため、作成時間がかかると共に、対応者のスキルによって品質がばらつくといった問題がある。また、コールセンタ管理者は、問題のある対応を発見した場合、その対応の通話音声、或いは通話音声を音声認識したテキストから内容を把握し、問題の解決や再発防止策を講じる。しかし、通話は内容と無関係な部分も多く含むため、通話音声全体を聞くコストは膨大となる。また、音声認識したテキストも閲覧性が悪いため、内容把握に時間がかかるという問題がある。

そのような背景で、通話から対応内容の要点をまとめた要約を作成するニーズが高まっている。岩崎らは、対応者と顧客の対話から重要文を抽出する事で営業日報を自動生成する手法を提案している[1][2]。このように、対応内容をまとめた要約は、コールメモの代替となりうる。あるいは、対応者がコールメモを作成する際のドラフトに利用できる。また、管理者は、要約から対応内容を把握する事で、内容把握のコストも削減する事ができる。

要約の研究は従来から盛んに行われている[3]。本研究では、要約対象の通話から重要文(発話)を特定、抽出し、抽出した重要文の集合を要約とする、重要文抽出による要約手法を提案する。従来の重要文抽出手法には、重要な文にタグが付与された学習データを用いる教師ありの手法[1][2][4]と用いない教師なしの手法[5][6]がある。教師ありの手法は、学習データの作成にコストがかかるという問題がある。一方で、従来の教師なしの手法は、対応報告、管理、分析の観点で文の重要度を決定しないため、コールセンタのニーズにあった要約を必ずしも生成できるとは限らない。

そこで我々は、対応者がコールメモによって報告するような内容を、対応の管理や分析に重要な内容として要約に含める事で、ニーズにあった要約を生成する手法を提案する。なお、提案手法は、コールセンタに蓄積された過去のコールメモを利用する事で、学習データの作成を必要としない。具体的には、高頻度語に着目して内容とは無関係な発話を要約から除くと共に、過去の通話とコールメモのペアから、要約対象の通話の中から報告されるべき重要な内容の発話を重要文として抽出する。

以降、2節で従来の要約手法とその課題を説明し、3節で提案手法を提案する。そして、4節で提案手法をコー



※左の数字は発話ID,括弧内はA:対応者,C:顧客

図1 通話とコールメモの例

ルセンタの実データに適用して有効性を検証する。最後に、5節で本稿をまとめる。

## 2. 従来手法

岩崎らは、コールセンタの対応者と顧客の対話から重要文を抽出する事で営業日報を自動生成する手法を提案している[1][2]。まず、対話中の各文を6タイプ(「営業日報を構成するタイプ」「質問」「応答」「結論」、営業日報に不要なタイプ「挨拶」「返事」「その他」)に分類する。分類は、学習データとSVMを用いて実現する。そして、「質問」「応答」「結論」の文に対し、文タイプごとに重要文抽出を行う。その後、抽出した重要文に情報の不足がある場合、情報を補完する文を追加し、最後に、テンプレートを用いて、対話の表現から営業日報の表現へと変換する。岩崎らの手法は、対話中の各文を前記6種類の文タイプに分類する際、学習データが必要で、学習データの作成にコストがかかる。

教師なしの重要文抽出手法の代表的な方法としてLead法[5]やtf・idf法[6]がある。Lead法は、重要箇所はテキストの最初の方に位置するとの仮定の下で、テキストの最初の文から順に重要文として抽出する方法である。しかし、コールセンタの対話は、最初に、要約としては不適切なオープニングの挨拶等が存在する。そして、要約に含めるべき顧客の用件の要点や対応者の対応は、テキストの途中に位置する。そのため、Lead法等の位置情報に基づく要約手法はコールセンタの対話には適さない。

一方、tf・idf法は単語の出現頻度を利用した方法である。文の各単語 $w_i$ に対して次の式で定義するtf・idfで重み付けをし、文の重要度を文中に出現する単語の重みの総和とする。そして、重要度の高い文から順に要約として抽出する。

$$tf \cdot idf_i = tf_i * idf_i$$

$$tf_i = f(w_i) / \sum_k f(w_k), idf = \log\left(\frac{|D_{all}|}{|D_{w_i \in d}|}\right)$$

$f(w_i)$  は単語  $w_i$  の出現数、 $|D_{all}|$  は総テキスト数、 $|D_{w_i \in d}|$  は単語  $w_i$  を含むテキスト数である。tf・idf法は、

要約対象に特徴的な単語（要約対象に頻出し、その他での出現頻度が少ない単語）を含む文を重要視し抽出する。

しかし、要約対象の通話に特徴的な単語と、コールセンタの業務（応答報告、管理、分析）で重要な単語は必ずしも一致しないため、コールセンタの業務上、不要な文が要約に含まれたり、必要な文が含まれない可能性がある。例えば、一部の顧客に特徴的で内容とは無関係な表現（顧客の癖で発話の開始時に「ちょっと」をつけるなど）は、コールセンタの業務上、不要であるが、要約に含まれる可能性がある。また、用件把握はコールセンタの業務で重要であるが、頻出する用件は、idf 値が小さくなるため、要約に含まれない可能性がある。

### 3. 提案手法

本節では、コールセンタのニーズにあった要約を生成する手法を提案する。我々は、応答者がコールメモによって報告している内容は、応答の管理や分析において重要であると考え、要約に含めるべきであると考えた。ただし、要約時には、要約対象の通話に対応するコールメモがない状況を考える。そこで、提案手法は、コールセンタに蓄積された過去のコールメモを利用し、要約対象の通話の各発話に対して、応答者により報告される度合い（「報告度」と呼ぶ）を推定し、報告度の高い発話を抽出し、抽出した発話集合を要約とする。

以降では、まず 3.1 節で、過去のコールメモに含まれる単語を重要視して報告度を推定する、ベースラインの手法とその問題を説明する。次に 3.2 節で、ベースライン手法の問題を解決した、過去の通話とコールメモのペアから報告度を推定する手法を提案する。

#### 3.1 コールメモに基づく手法（MF 法）

本手法は、過去のコールメモにおいて使用頻度が高い単語ほど報告されやすい単語と考え、単語  $w$  の報告度  $R\_Score(w)$  を次式のコールメモ頻度(Memo Frequency; MF)で推定する手法である。ここで、コールメモ頻度は、過去のコールメモの集合を母集団とした文書頻度である。

$$R\_Score(w) = \frac{|M_{w \in m}|}{|M_{all}|}$$

$|M_{w \in m}|$  は単語  $w$  を含むコールメモ数、 $|M_{all}|$  は総コールメモ数である。そして、本手法は、要約対象の各文の報告度を文中の単語の報告度の平均で計算し、報告度の高い文から順に抽出し、要約とする。

##### 3.1.1 MF 法の問題

コールメモは、応答者が関係者に内容を報告するという目的で作成するため、簡潔な表現で記載される傾向がある。また、関係者間でのみ通じる用語もしばしば使われる。一方、通話中では、顧客の理解や状況に応じて様々な表現が使われる傾向がある。したがって、コールメモと通話では、同一内容であっても、表現の不一致が生じる。図 1 の例では、通話中では、用件が「変な音が出る」と表現されている。一方、コールメモ中には、応答者により「異音」という簡潔な表現で記載されている。

MF 法では、過去のコールメモでの使用頻度に基づいて各単語の報告度が推定されるため、コールメモでは使われない単語は重要視されない。したがって、通話中で報

告すべき箇所であっても、コールメモと異なる表現で出現すると、重要な発話として抽出されないという問題がある。例えば、図 1 においては、発話 ID7 の発話が重要な発話として抽出されない可能性がある。

### 3.2 通話とコールメモ対に基づく手法（AS 法）

3.1.1 節で述べた問題点を考え、本節では、過去の通話とコールメモのペアから報告度を推定する要約手法を提案する。

#### 3.2.1 単語の相関による報告度推定

我々は、通話に対応するコールメモがある場合、通話の中からコールメモとの共通部分を特定する手法を提案している[7]。この手法[7]は、通話中の単語  $w$  が、対応するコールメモ  $M$  の内容になる可能性を、次式(1)の通り、コールメモ中の単語  $v \in M$  との相関の強さ（Association Strength; AS）で計算する。

$$AS(w, M) = \max_{v_i \in M} \sum_{i=1}^N AM(w, v_i) \quad \dots (1)$$

ここで、 $AM(w, v_i)$  は、単語  $w$  と  $v_i$  の相関係数である。例えば、相互情報量、 $\chi^2$  乗、Z 値を用いる事ができる。また、式(1)では  $M$  中の単語に対し、相関係数  $AM(w, v_i)$  の値の大きい上位  $N$  個に限って総和をとる。

$AS(w, M)$  は、表層の一致を考慮しないため、通話とコールメモ間の表現の違いに影響されずに、コールメモの内容になる可能性を計算できる。そこで、本節では、この  $AS(w, M)$  を通話中の単語  $w$  の報告度として採用する。

しかし、要約対象の通話には対応するコールメモが存在しない。そこで、要約対象に対応するコールメモとして、過去のコールメモに出現した単語が、過去のコールメモでの出現確率  $p_{\tilde{M}}$  ( $= w$  を含む過去のコールメモ数 / 過去のコールメモ総数) で存在する、仮想的なコールメモ  $\tilde{M}$  を考える。そして、単語  $v$  の報告度を、仮想的なコールメモ  $\tilde{M}$  との  $AS(w, \tilde{M})$  で計算する（次式(2)参照）。式(2)は、式(1)においてコールメモ中の単語に、過去のコールメモでの出現確率を重み付けした式である。

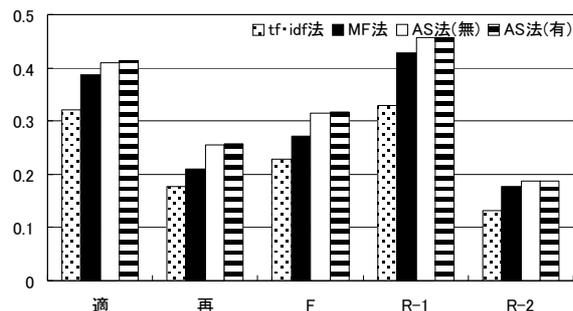
$$AS(w, \tilde{M}) = \max_{v_i \in \tilde{M}} \sum_{i=1}^N AM(w, v_i) p_{\tilde{M}}(w) \quad \dots (2)$$

提案手法 AS 法は、要約対象の各文の報告度を、文中の各単語に対して式(2)で計算した報告度の平均とする。そして、報告度の高い文から順に抽出し、要約とする。ここで、 $AM(w, v_i)$  は、過去の通話とコールメモのペアから計算し、 $p_{\tilde{M}}(w)$  は、過去のコールメモから計算する。

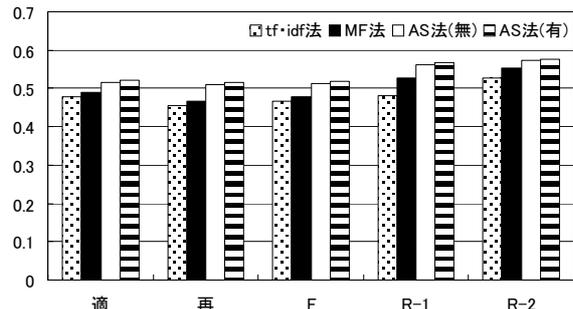
#### 3.2.2 不要発話削除の導入

通話は、「はい」「えー」などの相槌や「申し訳ありません」などの内容とは無関係な発話を多く含む。これらの発話は、要約には含めるべきではない。そこで、内容とは無関係な発話（以降、「不要発話」と呼ぶ）は、あらゆるトピックの応答で出現するとの仮定の下に、「発話が出現した通話数 / 全通話数」で定義される文書頻度が閾値  $\theta$  より高い発話を不要発話として特定する。そして、特定した不要発話は、要約には含めないという制約を付け加える事で、要約の質の改善をはかる。

(a) 音声認識テキストに対する性能



(b) 書き起こしテキストに対する性能



※適：適合率、再：再現率、F：F値、R-1(2)：ROUGE-1(2)

図2 各手法の indicative 要約の性能比較

#### 4. 評価実験

本節では、ある国内コールセンタにおける通話音声とコールメモを用いて、種々の要約手法で生成した要約の品質と提案手法の有効性を検証する。

##### 4.1 実験設定

###### 実験データ

実験データは、コールセンタでなされた顧客と応対者の通話音声 4,596 通話とそれに対応するコールメモを用いる。通話音声に対しては、下記の 2 種類のテキストを作成し用いる。

1. **音声認識テキスト**：通話音声の内容を読み上げた音声を音声認識したテキスト。認識結果は、単語正解率(PC)が、応対者 91.5%、顧客 89.9%、単語正解精度(WA)が、応対者 87.7%、顧客 84.8%である。正解の単語数を N、置換誤りの単語数を S、挿入誤り単語数を I、脱落誤り単語数を D とすると、 $PC=(N-S-D)/N$ 、 $WA=(N-S-D-I)/N$  である。
2. **書き起こしテキスト**：通話を人手で書き起こしたテキスト。PC、WA が共に 100%である時に相当する。

###### 評価セット

評価セットは、実験データより無作為に抽出した 40 事例を用いる。評価セットの要約の正解は、書き起こしに対して、人手により以下の 2 種類を作成した。

1. **indicative 要約**(一目で応対内容が分かる要約)：コールメモの代替やコールメモ作成時のドラフト、或いは、管理者が内容を一目で把握する為に用いる事を想定。要約率 (= 要約の文字数 / 要約元の文字数) が 30%となる要約を作成する際に抽出すべき文を人手で特定し、その文集合を正解とする。

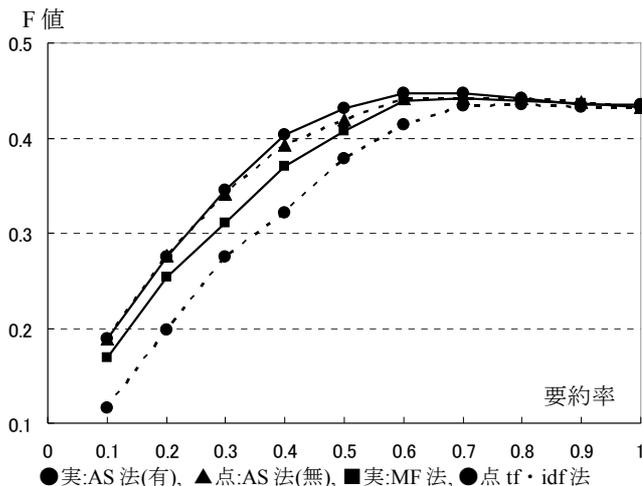


図3 各手法の informative 要約の性能評価

2. **informative 要約**(通話内容を漏れなく含む要約)：管理者が応対を詳細に分析したり、検索やマイニング等の分析元として使う事を想定。要約率を定めず、通話内容を漏れなく含む要約を作成する際に抽出すべき文を人手で特定し、その文集合を正解とする。

###### 評価対象手法

評価対象の要約手法は下記 4 種類である。要約時の抽出単位は、書き起こしに対しては文単位、音声認識テキストに対しては、音声認識エンジンが検出した発話単位である。

1. **tf-idf 法**：2 節で説明した手法。
2. **MF 法**：3.1 節で説明した手法。
3. **AS 法 (無)**：3.2.1 節で説明した手法。不要発話削除は行わない。報告度を求める際に用いる相関係数  $AM(w, v_i)$  は Z 値を採用し、N は 16 とする。
4. **AS 法 (有)**：AS 法 (無) に 3.2.2 節で説明した不要発話削除を導入した手法。手法の設定は AS 法 (無) と同じである。不要発話削除で、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 (2-gram) に変換して、統計量を計算する。また、不要発話削除の際に用いる閾値  $\theta$  は 0.5 とする。

###### 評価尺度

評価は、文献[8]より、文抽出の F 値、ROUGE-N[9]などの客観評価尺度と主観評価尺度との相関が高い事が示されている事から、客観評価により行う。用いた評価尺度は次の尺度である。また、本実験の評価では、単語として自立語のみを使用する。

###### • 文抽出の適合率、再現率、F 値

適合率は、手法が抽出した文の内、正しい割合。再現率は、正解の文の内、手法が抽出した割合。F 値は、「 $2 \times \text{適合率} \times \text{再現率} / (\text{適合率} + \text{再現率})$ 」である。また、音声認識テキストが要約対象の場合、正解と手法が抽出する文の文境界が一致せず、文が完全に対応しない。そこで文献[8]にならい、手法が抽出した各文に対して、その文中の単語が 50%以上占めている文が正解に含まれる場合、正しく文抽出が行われたものとみなした。

## ● ROUGE-N

正解と手法による要約との類似度を、次式の通り、正解の要約の N-gram に対する再現率で求める方式である。

$$ROUGE - N(C, R) = \frac{C_m(N - gram)(C, R)}{C(N - gram \in R)}$$

$C_m(N - gram)(C, R)$  は、手法の要約 (C) と正解の要約 (R) 間で一致する N-gram の数、 $C(N - gram \in R)$  は、正解の要約中の N-gram の数である。本研究では、1-gram (ROUGE-1)、2-gram(ROUGE-2)を用いた。

## 4.2 実験結果

4.1 節で説明した各手法の indicative 要約に対する性能を調べるため、各手法で要約率 30%の要約を作成し、indicative 要約の評価セットで性能評価を行った。結果を図 2 に示す。同様に、informative 要約に対する性能を調べるために、各手法で要約率を 0.1 から 1 まで 0.1 刻みに変化させて要約を作成し、informative 要約の評価セットで性能評価を行った。なお、本評価は、音声認識テキストを用いた場合で代表させる。また、評価指標は、F 値に代表させる。結果を図 3 に示す。

## 4.3 考察

### 4.3.1 提案手法の有効性検討

図 2、3 より、従来の tf・idf 法より、MF 法や AS 法 (無/有) の方が、要約生成性能が良い事が分かる。これより、コールメモを利用し、応答者が報告するような内容を要約に含める事が有効である事が分かる。また、MF 法より AS 法 (無/有) の方が、性能が良い事が分かる。この結果より、コールメモと通話のペアを利用して、単語間の相関の強さを基に報告度を求める事で、コールメモと通話間の表現の揺れを吸収し、品質のよい要約を生成できる事が確認できた。

### 4.3.2 音声認識誤りへの頑健性検討

図 2 より、全ての手法で、書き起しテキストよりも音声認識テキストの方が生成される要約の質が下がる事が分かる。これは、置換誤りや脱落誤りにより、本来要約として含まれるべき単語がテキスト中に現れなくなっている事が理由である。また、MF 法と AS 法 (無) の F 値を比較すると、音声認識テキストの場合、AS 法(無)の MF 法に対する誤り改善率<sup>1</sup>は 5.8%、書き起しテキストの場合、6.9%であり、書き起しテキストの方が、性能改善が大きい事が分かる。これは、MF 法が AS 法 (無/有) より、認識誤りに頑健である事を意味する。これは、AS 法 (無/有) は、認識誤りがある状況では、単語が湧き出したり、脱落、置換されるため、過去の通話とコールメモのペアから単語間の正しい相関を捉えづらく、報告度の推定が誤りやすくなるのに対し、MF 法は、過去のコールメモから報告度を求めるため、各単語の報告度の推定には認識誤りは悪影響を及ぼさない事が理由と考えられる。しかし、その欠点のもとでも、音声認識率 (精度) が今回程度あれば、AS 法 (無/有) の効果が発揮できる事が示せた。

### 4.3.3 不要発話削除の有効性検討

図 1 や図 2 の要約率が低い部分(30%以下)では、AS 法 (無) と AS 法 (有) で生成される要約に質の差は、ほとんど見られない。これは、AS 法 (無) で報告度が上位 30%以内に入るような文には、不要発話が含まれにくい事を意味する。この事から、AS 法 (無) は、要約率 30%に含まれる重要な文と不要発話の識別が可能といえる。一方、要約率が大きい場合には、AS 法(有)の方が AS 法(無)よりも性能がよく、不要発話削除の効果が確認できる。これは、要約率が大きい場合、報告度だけで重要か否かを判定し難しい文が出現し、その際、不要発話から削除する事で、結果として要約に適した文が要約に残り、要約の品質が保たれているからと考えられる。

## 5. おわりに

本稿では、コールセンタのニーズにあった要約を生成する手法として、過去の通話とコールメモのペアから通話の各発話の報告度を求め、報告度の高い発話を要約に含める手法を提案した。そして、コールセンタの実データを用いた実験を通じて有効性を確認した。また、要約率が大きい場合、高頻度な発話を不要発話として要約から除く事で要約の質を改善できる事を確認した。

今後は、主観評価により、生成した要約が、コールセンタにおける応対報告や管理、分析にどの程度役立つかを確認する予定である。また、本稿では、重要文抽出手法による要約手法を提案したが、今後は、さらに重要文の集合から、人間が読みやすい要約を生成する手法を開発する予定である。

## 参考文献

- [1] 岩崎 礼次郎, 荒木 健治. “コールセンターの対話データを対象とした営業日報自動生成のための重要文抽出手法”, 第 19 回人工知能学会全国大会論文集 1E1-02, 2005.
- [2] 岩崎 礼次郎, 荒木 健治. “コールセンターにおける対話データを用いた営業日報の自動生成”, ことば工学研究会資料, SIG-LSE-A503-14, 2006.
- [3] 奥村 学, 難波 英嗣. “テキスト自動要約に関する最近の話題”, 自然言語処理, Vol.9, No.4, 97-116, 2002.
- [4] 藤井 康寿, 山本 一公, 北岡 教英, 中川 聖一. “重要文抽出に基づく講義音声の自動要約”, 情報処理学会論文誌, Vol.51, No.3, 1094-1106, 2010.
- 山本 国会会議録を対象とする話言葉要約
- [5] H.P. Edmundson. “New methods in automatic abstracting”, journal of ACM, 74-81, 2004.
- [6] K. Zechner. “Fast Generation of Abstracts from General Domain Text Corpora by Extracting Relevant Sentences”, In Proc. of the 16<sup>th</sup> International Conference on Computational Linguistics, 986-989, 1996.
- [7] 石川 開, 田村 晃裕, 安藤 真一, 西光 雅弘. “音声対話と報告テキストからの共通箇所抽出手法”, 第 5 回音声ドキュメント処理ワークショップ, 2011 (発表予定).
- [8] 岩野 公司, 広畑 誠, 新中 庸介, 古井 貞照. “重要文抽出による音声自動要約手法とその客観評価法についての検討”, 電子情報通信学会技術研究報告, SP2005-20, 2005.
- [9] C-Y. Lin. “ROUGE: A Package for Automatic Evaluation of Summaries”, In Proc. of Workshop on Text Summarization Branches Out, 74-81, 2004.

<sup>1</sup> (AS法(無)の F 値-MS 法の F 値)/(1-MS 法の F 値)×100