

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

Important Sentence Extraction Method for Automatic Generation of Business Days Report for Conversation Data of Call Center

岩崎 礼次郎^{*1}
Reijirou IWASAKI

荒木 健治^{*1}
Kenji ARAKI

^{*1} 北海道大学情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

Since it is very heavy labor to make the business days report of every day at a call center, it is hoped that the business days report is generated automatically. Therefore, we have proposed an important sentence extraction method for automatic generation of the business days report for conversation data of a call center.

1. はじめに

近年、音声認識技術やテキスト要約技術の発展に伴い、その技術を利用したシステムの実用化が望まれている。現在では、ニュース音声の自動字幕化システム [安藤 00] や、講演音声自動要約 [2] などの研究が行われている。ところが、これらの文章要約技術といったものは、必ずしも万能ではないため目的によって異なった手法を用いる必要がある。

そこで、本稿では、コールセンターにおいて毎日作成される営業日報に着目し、これを自動的に作成することにより人にかかる負担を軽減することを目的とし、コールセンターの対話データの中から営業日報に記載すべき重要文を抽出する手法について提案を行った。

2. コールセンター対話データと営業日報

対話データからの重要文抽出や要約技術といったものが存在 [畑山 01] [安藤 00] するが、それらの手法の有効性は対象となる対話データにより異なってくる。そこで、まずはコールセンターの営業日報における特徴がどのようなものかについて述べる。

2.1 コールセンター対話データと営業日報の特徴

証券会社のコールセンターにおける株式に関する話題を中心とする全 12 対話 320 文の対話データと営業日報を元に調査を行った結果いくつかの特徴が発見された。1 つ目の特徴は、コールセンターの対話データには挨拶、返事といった営業日報に不必要な部分が多く見つかった。これは電話のため必ず最初と最後に挨拶が行われ、相手の会話に頷くなどといったことが多々行われるためだと考えられる。2 つ目の特徴としては、営業日報は記述される構成が定まっている。これは、次のような 3 つの内容で構成されている。

- お客様からの用件に関する記述
- オペレーターの対応に関する記述

- どのような結論に至ったかに関する記述

また、この構成のうち結論部分が省略されることや、営業日報を書く人にとって周知の事実は省略されることもある。3 つ目の特徴は、営業日報の語尾が「～の問い合わせ」「～の説明」という対話データには出現しないが営業日報としては高い頻度で記述されている。(図 1)

Sample 1

資金運用についてのご相談。

お客様が高齢のこともあり、MMF、国際分散型の投資をご紹介します。

ご検討いただけるとのこと。

Sample 2

CP, CD についての問い合わせ。

個人のお客様への販売はお勧めしていない旨説明。

図 1: 営業日報の例

2.2 従来手法による問題点

対話を要約する手法として代表的な 3 つの手法と比較を行い、それぞれの手法における問題点を明らかにする。1 つ目は、もっとも一般的な tf-idf 法などを用いた重要文抽出手法 [奥村 02] がある。この手法においては重要な文は確かに抽出されるが、今回のような営業日報のように構成が定まっているものに対しては必ずしも欲しい情報が得られないという問題がある。2 つ目は、一文単位で要約を行う手法 [堀 99] である。この手法はそれぞれの文を短くするという手法であるため、営業日報に不要な挨拶などの文はそれだけで一文であることが多くこの手法ではこれらが残ってしまうという問題がある。3 つ目は、Lead 法 [Edmundson 69] という方法である。この手法は文章の先頭に重要な文が書かれているとし、先頭を重要文として抽出する手法である。この方法も、コールセンターは対話という状況のため挨拶などの不必要な語が先頭に来ることが多いという問題がある。

連絡先: 岩崎 礼次郎, 北海道大学情報科学研究科, 札幌市北区北 14 条西 9 丁目, TEL:+81-11-706-7389, AX:+81-11-709-6277, iwasaki@media.eng.hokudai.ac.jp

3. 営業日報のための重要文抽出手法

このようなコールセンター対話と営業日報の特徴から、営業日報を自動生成するための手法として、対話データを一文単位で営業日報を構成する3つの要素と営業日報に不必要な挨拶、返事といった文になるよう学習による分類を行い(分類されたものを文タイプとする)、その後、文タイプごとの重要文を抽出しその抽出された重要文を営業日報にするべき文とする手法を提案する。本手法により、営業日報の構成に適した重要文の獲得が可能となるだけでなく、文タイプごとに異なったアプローチが可能となり、さらに明確に不必要な文が判別できるようになる。本稿では前回の発表[岩崎 04]と比べて、分類の際の素性情報の追加や、重要文抽出の際に文タイプごとに重要度の重み付けの変更を行うことにより精度の向上を目指した。

3.1 文タイプの分類の構成

今回、文タイプを分類するにあたり Support Vector Machine[SVM]という2値分類を行う手法を利用した。分類する素性としては、形態素解析ツール chasen[茶筌]による形態素情報として、字面情報と詞情報を利用した。実験データ全12対話、320文とし、11対話を学習データ、残りの1対話を実験データとする交差検定を行った。このとき、学習データの正解は6つの文タイプによって予め正解を与えた。その文類するべき文タイプとして6つの文タイプの設定を以下のように行い実験を行った。

- 営業日報に記載されるべき内容の文
 - 質問:お客様からの用件に関する記述
 - 応答:オペレーターの対応に関する記述
 - 結論:どのような結論に至ったかに関する記述
- 営業日報に不必要な内容の文
 - 挨拶:挨拶や呼びかけ
 - 返事:はい、了解しました。等の簡潔な返事
 - その他

また、SVM は1つの文タイプとそれ以外という風に分類が行われるため、1つの文が複数の文タイプに割り当てられる可能性がある。

3.2 文タイプ分類結果

営業日報の特徴を調べた対話データと同様の全12対話、320文の対話データを元に各文タイプごとに分類した結果を図2に示す。

	F-measure	Precision	Recall
1. 挨拶	77.31	97.87	63.89
2. 質問	56.98	46.23	74.24
3. 応答	55.17	38.94	94.62
4. 返事	83.67	91.11	77.36
5. 結論	31.58	30.00	33.33
6. その他	7.06	4.48	16.67

図2:SVMによる分類精度

図2の結果から、挨拶、返事における Precision が共に90%以上の精度が出ていることが分かる。このことから、挨拶、返事といった文タイプでは分類がほぼ間違いないものであるため、この文タイプに分類された文は他の文タイプの候補からは除外することが可能となる。

3.3 文タイプごとの重要文抽出手法

文タイプごとに分類された文の中から tf-idf 法による各文の重要度を元に営業日報に必要な文タイプ「質問、応答、結論」のそれぞれにおいて最も重要な文の抽出を行った。また、重要文は候補の中から上位1割の文とした。

さらに、このとき文タイプに分類されていることを利用し、以下のように重要度の変更を行った。はじめに、先ほどの分類結果より、「挨拶、返事」に分類される文を他の文タイプの候補から除外した。これらを除外することにより、「質問」タイプがコールセンター対話の先頭にくるため Lead 法を利用できる。このため、「質問」タイプは文タイプの中で先頭から4文を候補とした。同様に「結論」タイプは末尾に来るため文タイプの後ろから4文を候補とした。次に、「質問」タイプにおいては日本語の特徴より「～ですか?」「～なの?」の例に見られるように「か」「の」といった語が語尾に来る場合質問文である可能性が非常に高いという特徴を持っているため、語尾がこのような語のときは重みを増加させた。最後に「質問」「応答」「結論」という文脈情報を利用することによりそれぞれの文タイプにおける重要文の抽出を行った。

3.4 文タイプごとの重要文抽出の評価結果

文タイプごとに抽出された重要文は、それぞれの重要文を人手により営業日報とした際、正解の営業日報と同意のものを作成することができるものとして適切かどうかにより以下の4種で評価を行った。実験結果を表1に、実際の例を図3に示す。

- 1: 同意の営業日報が作成可能
 - 2: 文の内容が不十分である(不足)
 - 3: 異なった内容になる
 - 4: 日報には記述されていないが、日報として書いてもなんら問題が無いもの
- 以上のことから、選択された重要文が適切であると思われるのは1+2+4の場合であるといえる。

表1:文タイプごとの重要文抽出結果

	1	2	3	4	1+2+4
質問全12文	41.7	8.3	50.0	0.0	50.0
応答全18文	55.6	16.7	27.7	0.0	72.3
結論全12文	66.7	0.0	25.0	8.3	75.0

<p>抽出された重要文 質問:田中ともうしますが,CPやCDは,私個人でも購入できるかどうかをおしえていただけますか. 応答:CPやCDSは,基本的に単騎金融市場で取り扱われる商品で発行額面が大きいので,私共としましてはお勧めしておりません. 結論:じゃあやめておきます.</p>
<p>営業日報 CD,CPについての問い合わせ. 個人のお客様への販売はお勧めしていない旨説明.</p>

図3:抽出された重要文と日報の比較例

以上の結果より,営業日報の3つの構成を元にした重要文の獲得が行われた.要約の精度はまだ不十分であるといえるが,本手法は営業日報においても要約すべき重要文が全体では66.7%の精度で獲得することが出来たことから,この点においては本手法が有効であると考えられる.

4. 文要約

最後に,抽出された重要文を元に堀ら[堀 99]の提案した手法を用いることにより一文単位の要約を行った.この手法は,単語がどの程度自然に接続しているかを表す言語スコアと,その単語がどの程度文の中で話題語らしいかを表す単語重要度スコアを利用することにより要約する手法である.

この結果,営業日報と完全に営業日報と同様の要約文は生成されず,重要な情報が欠如してしまったり,不自然な語の繋がりとといったものが出来てしまったが,意味的に近いものや話し言葉から営業日報に記載する形の言葉に変化させればよいと思われる20%程度見受けられた(図4).このため,営業日報にする際には今回利用した要約手法の改良や,語尾を営業日報に適した形に変える方法などをとる必要があると考えられる.

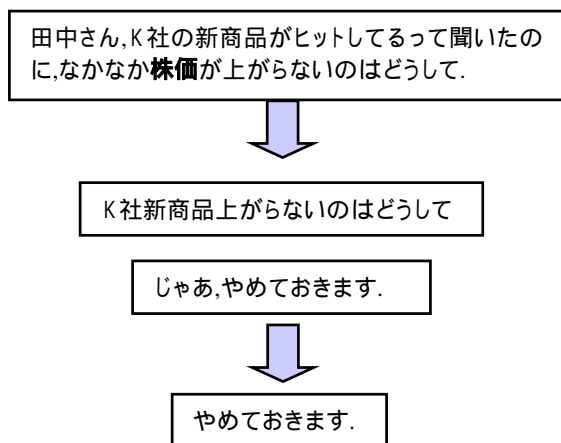


図4:文要約結果の例

5. まとめ

本稿では,コールセンターの対話データの中から営業日報に記載すべき重要文を抽出する手法について提案を行った.その結果,今回の提案した分類により営業日報の構成に適した重要文の獲得に有効であることが確認された.しかしながら,その重要文から営業日報にすることは従来の手法では不十分であった.今後は,さらなる分類精度の向上と抽出された重要文の語尾の変換による営業日報の生成を予定している.

参考文献

[岩崎 04] 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からの営業日報の自動生成のための文タイプ分類,平成 16 年度電気・情報関係学会北海道支部連合大会講演論文集,281,pp.345(2004-10)

[畑山 01] 畑山 満美子,松尾 義博,白井 諭:重要語句抽出による新聞記事自動要約,自然言語処理,2001

[安藤 00] 安藤 彰男: ニュース音声自動字幕化システム,音声言語情報処理学会研究会,SP2000-101,pp.31-36.

[堀 01] 堀 智織,古井 貞熙: 講演録作成を目的とした講演音声自動要約,日本音響学会,2001 年秋季講演論文集 ,2-1-11,pp.69-70.(2001-10)

[堀 99] 堀 智織,古井 貞熙: 話題語と言語モデルを用いた音声自動要約手法の検討,電子情報通信学会技術研究報告,Vol99,No525,SP99-110,pp.103-108

[Furui 04] Sadaoki Furui, Tomonori Kikuchi, Yousuke Shinnaka, Chiori Hori:Speech-to-Text and Speech-to-Speech Summarization of Spontaneous Speech, IEEE transactions on speech and audio processing, vol.12, NO. 5 JULY 2004

[奥村 02] 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, pp.97-116, 2002

[遠山 00] 遠山義洋,西田豊明: 話題構造の抽出と変形による対話録の自動要約, 2000 年度人工知能学会全国大会(第14回)論文集, pp.157-160, 2000

[Youngjoong 02] Youngjoong Ko, Jinwoo Park, Jungyon Seo:Automatic Text Categorization using the Importance of Sentences, COOLING2002

[Edmundson 69] Edmundson,H.P.:New methods in automatic abstractiong, journal of ACM 1969

[SVM] <http://chasen.org/~taku/software/TinySVM/>

[茶筌] <http://chasen.aist-nara.ac.jp/hiki/ChaSen/>