

1R-4

コールセンターにおける重要なクレームの絞り込み手法について

梶川泰成[†] 渡邊倫[‡] 大園忠親[‡] 伊藤孝行[‡] 新谷虎松[‡][†]名古屋工業大学 知能情報システム学科 [‡]名古屋工業大学 大学院工学研究科 情報工学専攻

e-mail: {kajikawa, watanabe, ozono, itota, tora}@ics.nitech.ac.jp

1 はじめに

近年、組織内の知識は電子化されたテキストとして、年々増加し、膨大な量が蓄積されている。知的資産が蓄積された膨大な情報源から、必要な情報を選択し、知的資産を再利用することは有益である。また、知的資産のような膨大な情報源から情報を再利用する際、ユーザは情報を得るため情報源の検索を行う。しかし、検索結果がユーザの必要な情報のみであるとは限らない。情報源が膨大である程、不必要な情報も検索結果に含まれる。これを解決するため、絞り込みを行い、必要な情報のみを検索結果に出力するための技術は有益であると考えられる。本稿では、情報源としてコールセンターに蓄積された電話相談のクレームに着目し、クレームを再利用する際、ユーザにとって必要なクレームに絞り込むことで、オペレータの検索効率の向上を図る。

2 コールセンターにおけるクレーム

本稿では、クレームとして、株式会社 INAX お客様相談センター（以降、コールセンターと呼ぶ）に蓄積されたクレームを対象とする。本コールセンターにおいて、オペレータは電話相談の応答のため、過去の電話応対事例参照する。電話応対を蓄積するために、応対内容はデータベース化されている。記録項目は、相談者による問い合わせ内容である質問文書、オペレータによる応対内容である応答文書などを含んでいる。オペレータが過去の事例を参照しようという場合に、そのオペレータにとって、自明な内容を減らすことで、検索効率の向上が見込まれる。本研究では、オペレータにとって自明な内容を取り除いたものを重要なクレームとする。重要なクレームに絞り込むためには、内容によって分類しなくてはならない。しかし、クレームの内容は自由記述形式で記録されており、分類は難しい。さらに文書長が短い [1] ため従来手法による分類の精度が低い。本稿では、コールセンターのクレームの特徴を用いて分類を行う手法を提案する。コールセンターのオペレータに求められるのは顧客の問い合わせに迅速かつ適切に回答することである。そのため、

やり取りの内容を丁寧に記録すること自体がオペレータの評価対象とならない [2]。電話相談記録の応答文書が短い場合、具体的な応答内容を記載する必要がなく、オペレータにとって応答が容易だったクレームと考える。逆に、応答文書が詳細に記述されている場合、詳細に記述する必要があり、応答が難しい場合である可能性が高いと考える [1]。また、出現頻度の高い応答内容は、オペレータであればその応答内容を答える相談内容を知っている可能性が高く、過去のクレームを参照せずに応答可能であると考えられる。以上から、本稿では、短い応答文書で出現頻度の高い応答内容の応答文書のクレームを絞り込みによって検索結果から除き、オペレータの検索効率の向上が可能と考える。

3 絞り込み機構

絞り込みの前処理としてクレームのスコアを計算する。スコアの低いクレームを重要なクレームとする。また、スコア計算を行うため、短い応答文書を選別し、使用される回数が多い応答内容を得るため、選別した応答文書を係り受け解析を行う。解析の結果得られた係り受けが、使用された回数を求めるため、選別された応答文書に出現する確率を求める。

スコア計算の前処理として、短い応答文書を選別する。しかし、コールセンターでは商品名など文字数が多い言葉が多く存在する。そのため、文字数だけで選別を行うのではなく、自立語の数によって応答文書の選別を行う。全てのクレームの応答文書の集合を D とすると、選別された応答文書の集合 D' は次の式で表せられる。

$$D' = \{d | \text{length}(d) < \alpha, \alpha = \text{定数}, d \in D\} \quad (1)$$

ここで、 d は任意の応答文書を示す。関数 $\text{length}(d)$ は応答文書 d の自立語数を返す。定数 α は応答文書の選別を行う閾値を表す。閾値 α 以下の自立語数の応答文書を選別する。本稿では閾値 α を 4 とする。

次に、選別した応答文書に出現する回数の多い係り受けを抽出する必要がある。そのために、選別した全ての応答文書に係り受け解析を行う。解析結果から得られた係り受けを用いて頻出する内容を発見する。そのため、得られた係り受けの集合内で、係り受けのマッチングを行い、係り受けの出現確率を求める。係り受けは二つの自立語で表されている。マッチングの際、言

The Discovery of Important Claims in the Call Center
Yasunari KAJIKAWA, Satoshi WATANABE, Tadachika OZONO, Takayuki ITO, Toramatsu SHINTANI

Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya 466-8555 JAPAN

い換え辞書を使用して表記のゆれを吸収する。係り受けの出現確率は次の式で表せられる。

$$p(a) = \frac{\text{count}(a)}{n} \{a \in A\} \quad (2)$$

ここで、係り受け a の出現確率を $p(a)$ とする。関数 $\text{count}(a)$ は係り受け a が選別された応答文書に出現した回数である。 n は選別された応答文書の数である。

スコア計算は、対象となるクレームの応答文書の係り受けの出現確率の合計によって示される。クレーム c のスコア $\text{score}(c)$ は以下の式で表すことが表せられる。

$$\text{score}(c) = \sum p(a) \{a \in A, c \in C\} \quad (3)$$

ここで、 A は解析結果から得られた応答文書の全ての係り受けの集合であり、 a は集合 A の任意の係り受けである。関数 $p(a)$ は式 (3) で示された係り受け a の出現確率を返す。上記のスコア計算を用いて、クレームを検索した結果のクレームの集合のスコアを計算し、スコアが閾値以上のクレームを重要ではないクレームと分類することで絞り込みを行う。

4 評価

本絞り込み手法を評価するために、実験を行った。あらかじめ重要なクレームと重要ではないクレームに分類した 429 件のクレームのスコアを計算する。重要ではないクレームは、商品の故障に対する修理、商品に対する質問とする。絞り込みの目標は、重要なクレームのみを出力し、重要ではないクレームを最小限に抑えることである。よって、重要なクレームが 98%以上絞り込みの結果に出力することを前提とし、重要ではないクレームが出力されない割合が高い程、評価が高いとする。

評価実験の結果である任意の閾値での絞り込みの結果をグラフ化し、図 1 に示す。図 1 より、グラフの横軸が 6 の値、スコアの閾値が 30 の場合に着目する。絞り込みを閾値 30 で行った際の、結果を表 1 に記す。A は、対象が重要なクレームと分類された割合であり、B は、対象が重要ではないクレームと分類された割合である。閾値が 30 の場合、重要なクレームを 98.05%出力し、重要ではないクレームを 70.74%出力結果から除くという結果が得られた。本実験の結果から、本手法が有用であると考えられる。

5 考察

絞り込みの結果、重要ではないと分類された重要なクレームは、応答文章内に、一般的な応答と一般的ではない応答が記述されていたため、スコアが高くなった。また、重要と分類された重要ではないクレームは

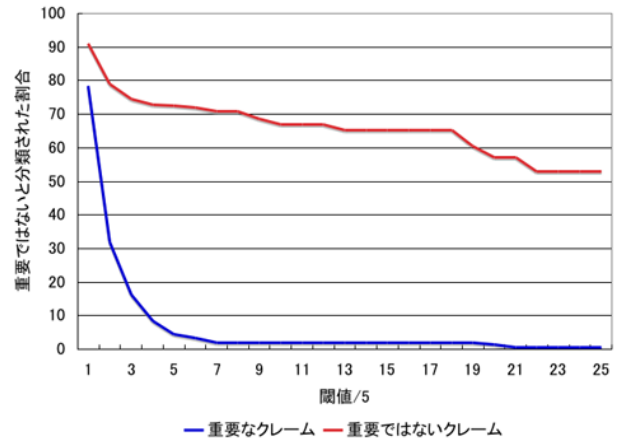


図 1: 評価実験の結果

	A	B
重要ではないクレーム	70.74%	29.26%
重要なクレーム	1.95%	98.05%

表 1: 閾値 30 での絞り込みの結果

以下の二つの特徴が見られた。1) 誤字によって自然言語解析されなかった文章がある。2) 出現確率の高い係り受けと同意であるが、言い換え辞書に含まれていなかった語が応答文書に含まれていた。本手法は、自由記述形式の応答文書に着目し、絞り込みを行うが、自由記述形式であるために、表記のゆれが発生してしまう。これを解消するには、クレームを記録する際に入力を補完する機構などを用いる必要があると考える。

6 おわりに

本稿では、コールセンターにおける重要なクレームを絞り込むことの有用性を述べ、絞り込み手法の説明と、評価実験について述べた。本絞り込み手法により、重要ではないクレームを検索結果から除くことで、オペレータの作業効率の向上が可能である。

参考文献

- [1] 長谷川友治, 大冨忠親, 伊藤孝行, 新谷虎松, ”コールセンターにおける質問応答データの FAQ 作成支援システムの試作”, 日本ソフトウェア科学会第 20 回記念大会論文集, 2003.
- [2] 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”, 人工知能学会誌, vol.16 No.2, pp.219-225, 2001.