

文書カテゴリを利用した文書クラスタリングの コールセンターFAQ改善への適用

木村 英志[†], 高島 俊哉[†], 重岡 知昭[†], 森田 豊久[†]

株式会社日立システムズ[†]

1. はじめに

企業のコールセンターは、利用者からの商品注文の受付や、問い合わせに対する情報の提供の役割を担う。

問い合わせを受け付けると、まず、対応手順や回答が定められているか否かを判断する。定められている場合は、手順書やよくある問い合わせ集(FAQ)に記載の対応方法を用いる。手順書やFAQに対策が存在しない場合は、過去履歴の検索や専門家に対する2次問い合わせを行い、解決を試みる。受付業務効率化や対応品質向上のためには、現時点の状況に適合するよう、手順書やFAQをメンテナンスし、拡充していく必要がある。

本稿では、FAQの品質向上を目的として、文書クラスタリングを用いてFAQを改善する候補を抽出する手法を提案する。ここでは、FAQを改善する候補を抽出するために、文書集合を自動的に分ける技術である文書クラスタリングに加えて、“FAQ”や“問い合わせ履歴”といった文書カテゴリによる区分を利用する。

また、コールセンターで記録された問い合わせ受付履歴から、改善の候補を抽出し、実際のFAQに適用した評価実験の結果を報告する。

2. コールセンターFAQ改善支援

FAQの改善には、有益な情報を持つ過去の問い合わせ受付履歴の把握や、陳腐化したFAQの発見が必要である。しかし、類似する過去の問い合わせの収集やFAQ活用状況の判定、登録や更新の判断など作業者の負担は大きい。

FAQの作成を支援する方法として、文章間の関連度を適用した手法[1]がある。これは、ある問い合わせ受付情報と、問い合わせ受付履歴群を抽出して、FAQ作成支援の参考とする手法である。しかし、“FAQの作成”に着目した手法であるため、FAQの陳腐化や更新の必要性といったFAQ運用に必要な評価はできないという課題がある。

また、文書の類似性に着目した文書クラスタリングでは、文書群の内容を類推するのは難しいという問題について、文書群の代表文を自動生成することで支援する手法[2]がある。文書クラスタリングの結果に代表文を付加し、文書群の具体的な記述内容理解を支援しているが、新しいFAQの候補抽出支援であり、既存のFAQに対しての評価には至っていない。

3. 文書カテゴリを用いた文書クラスタリング

FAQ作成候補の抽出に加えて、FAQの陳腐化を抽出するために、文書カテゴリを用いた文書クラスタリングを提案する。本手法では、問い合わせ受付履歴とFAQに文書カテゴリを付与し、統合したデータをその対象とする。ここで、文書カテゴリとは、文書が問い合わせ履歴情報であるかFAQであるかといった区分で、予め定められているものとする。

文書カテゴリを用いた文書クラスタリングは次の手順を踏む。

(1) 対象データへの文書カテゴリの付与

対象データに文書の種類を特定するカテゴリを付与する。「問い合わせ受付情報」または「FAQ」をカテゴリ値とする。

(2) 対象データの統合

各文書データから分類する文書と文書の識別子を抽出して、統合する。

(3) 統合データの文書クラスタリング

統合データに対して、文書のクラスタリングを実行する。まず各文書の特徴量を算出する。その上で、精度が高いとされる凝集型階層的クラスタリング[3]を用いる。中でも、各クラスタ内すべての対象間の距離の平均をクラスタ間の距離とする、群平均法を用いることとする。

(4) クラスタ要素の文書カテゴリによる区分

文書クラスタリングの結果を、クラスタを構成する文書の文書カテゴリに基づいて、次の通り区分する。

Document Clustering using Document Categories, and Its Application to Improvement of FAQs in Call-Center

[†]Eiji Kimura, Toshiya Takashima and Tomoaki Shigeoka, Toyohisa Morita, Hitachi Systems, Ltd.

表 1: クラスタ判定区分

判定区分	文書カテゴリ	
	問い合わせ	FAQ
FAQ 活用クラスタ	含む	含む
FAQ 無しクラスタ	含む	含まない
活用無しクラスタ	含まない	含む

4. 評価実験

(1) 本手法の実行方法

本実験においては、各文書の特徴量として、TF-IDF 値を正規化したベクトルを用いた。TF-IDF 値取得時の形態素解析には mecab[4] を利用し、文書間の距離の算出にはユークリッド距離を用いた。

更に、文書クラスタリング精度向上のために、一定程度の距離に類似文書が見つからない問い合わせや FAQ に関しては、予めクラスタリング対象から除外する処置を行った。

(2) 評価対象データ

特定顧客を対象とする全ての問い合わせ受付情報について、本手法を用いた FAQ 改善前後での受付対応時間の比較を行った。FAQ 改善前期間に収集した問い合わせ受付履歴と FAQ データに、本手法を適用した。結果クラスタの判定区分から”FAQ 無しクラスタ”と”活用無しクラスタ”を抽出し、コールセンター担当者により FAQ を更新した。評価対象データを表 2 に示す。

表 2: 評価対象データ

項目	FAQ 改善前	FAQ 改善後
評価期間	2013 年 8 月	2013 年 9 月
問い合わせ件数	1,325 件	1,145 件
FAQ 件数	1,006 件	1,009 件

(3) 評価結果

本実験では、クラスタ判定区分の 1 つである”FAQ 無しクラスタ”に対応する FAQ を新規追加し、”活用無しクラスタ”の FAQ を見直すという FAQ 改善を実施した。

改善した FAQ に関する問い合わせ受付を抽出し、FAQ 改善前後で受付対応時間を比較した。ここで、改善した FAQ に関する問い合わせ受付情報は、文書類似度を元に FAQ 改善前後からそれぞれ問い合わせ受付を抽出することとし、本研究で用いたクラスタリングツールを流用した。

表 3 に、新規登録した FAQ と修正した FAQ と

関係する問い合わせ受付対応時間の削減率と、これらの改善した FAQ と関係のない問い合わせの受付対応時間の削減率を示す。結果として、受付対応時間は従来の約 40~約 70%減となった。

表 3: 受付対応時間比較

計測対象 受付履歴	FAQ 改善前	FAQ 改善後	対応時間 削減率
新規 FAQ を活用	37.3 分 (12 件)	11.4 分 (18 件)	69.5%
修正 FAQ を活用	12.7 分 (13 件)	7.6 分 (5 件)	40.2%
改善 FAQ 非活用	41.1 分 (840 件)	36.5 分 (844 件)	11.1%

(小数点第 2 位以下四捨五入)

5. おわりに

本研究では、FAQ のメンテナンスという、従来は実施自体が困難な作業に着目した。問い合わせ受付履歴と FAQ に文書カテゴリを付与し、その統合データに文書クラスタリングを実行することで、FAQ 活用有無を判定する手法を提案した。

この判定結果を FAQ 改善に活用した結果、改善した FAQ に関連する問い合わせの受付対応時間を、FAQ 改善前後で比して約 40~約 70%削減できた。

今回は、1 か月間のデータを対象とした。今後は、長期間データを対象とした分析を進めていく。また、FAQ を既に活用している問い合わせ群の FAQ 評価を対象としていきたい。

6. 参考文献

[1]VextContact, ベクスト株式会社, <<http://www.vext.co.jp/product/vextcontact>> (2013 年 12 月 24 日閲覧)

[2]倉田 沙織, 小川 哲男, 可能 敏. 代表文生成技術と FAQ 作成の効率向上. 東芝レビュー Vol.66 No.9, 2011

[3]F. Murtaghf. A survey of recent advances in hierarchical clustering algorithms. THE COMPUTER JOURNAL VOL.26 NO.4 pp.354-359, 1983.

[4] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto. Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.