

代表文生成技術とFAQ作成の効率向上

Representative Sentence Generation Technology and Application to Highly Efficient FAQ Generation

倉田 早織

小川 哲男

加納 敏行

■ KURATA Saori

■ OGAWA Akio

■ KANO Toshiyuki

企業内外に存在する様々な情報から役だつ知識を得て活用するには、文書を全て読み内容を分析して整理する必要がある。しかし、この作業を人手で行うと多大な時間がかかる。

東芝ソリューション(株)は、蓄積された文書群の情報活用を目的として、文書群から代表文を生成する技術を開発した。生成された代表文を見ることで、文書群の概要が把握でき、コールトラッキング^(注1)分析に活用できるようになる。この代表文生成技術をコンタクトセンターのFAQ (Frequently Asked Question) 作成支援に適用した場合の有効性を評価し、コールトラッキング分析にかかる作業時間を大幅に短縮できるという結論を得た。

Document classification and analysis technologies are required in order to extract useful knowledge from large amounts of information inside and outside an enterprise, and to utilize it. However, these processes involve the expenditure of considerable time and effort.

Toshiba Solutions Corporation has developed a representative sentence generation technology for call tracking analysis, which makes it possible to automatically and efficiently create representative sentences from a group of documents. We have confirmed the effectiveness of this technology by applying it to the generation of frequently asked questions (FAQs) in a contact center.

1 まえがき

企業内外には、多種多様な情報があふれている。これらの情報から役だつ知識を抽出し、活用することが企業活動で重要である⁽¹⁾。例えば、コンタクトセンターには顧客からの意見が蓄積されており、新商品の企画に役だつ知見が含まれている。また、年間30万件以上発行されている特許公報には、幅広い分野の技術情報が含まれており、技術動向の把握や技術戦略の立案に役だつ知識が埋もれている。

文書化された情報から知識を得るには、文書を全て読み、その内容を分析し整理する必要がある。しかし、この作業を行うには、多大な時間を要する。

そこで、多くの文書に共通して記載されている要点が自動的に抽出され、要点とその件数が提示されれば、文書群の概要が把握できる。そして、詳しく読むべき文書の選択ができ、情報の活用が効率的に行えるようになる。文書に共通する要点を表現する文を代表文と呼ぶ。

東芝ソリューション(株)は、代表文を生成する技術を開発した。ここでは、代表文生成技術と、コンタクトセンターでのFAQ (Frequently Asked Question) の作成に対する代表文生成技術の有効性について述べる。

(注1) 顧客との対応内容を集積したもの。

2 代表文生成技術

2.1 関連技術

文書群の分析と整理を目的として、文書中の製品名や年月など、属性値やキーワードに基づく文書の自動分類が行われている。しかし、自動分類の結果として付与されるキーワードだけから文書群の内容を類推するのでは、具体的な内容の的確な把握が難しい。

2.2 代表文生成の処理の流れ

代表文の生成は、まず候補となる文の生成を行い、生成された候補から代表文を選択するという流れで処理される^{(2), (3)}(図1)。以下、処理の内容について述べる。

2.2.1 候補となる文の生成 文書群に含まれる全ての文について構文解析を行う。この解析結果に抽出パターンを適用し、木構造形式で表現された内容として抽出する(図2)。

抽出パターンは、主語や述語という、文の骨格として重要な単語と、単語間の構文的関係を表したものである。抽出パターンにマッチした部分を、木構造形式で表現された内容として抽出する。

そして、木構造形式で表現された内容から、その内容を表した文を生成し候補とする。

2.2.2 候補となる文からの代表文の選択 候補となる文のそれぞれに対して、生成元の文書数を算出し、文書数が高い順に文を選択し、代表文として出力する。

この二つの処理により、文書群から、“メールを送信しない”

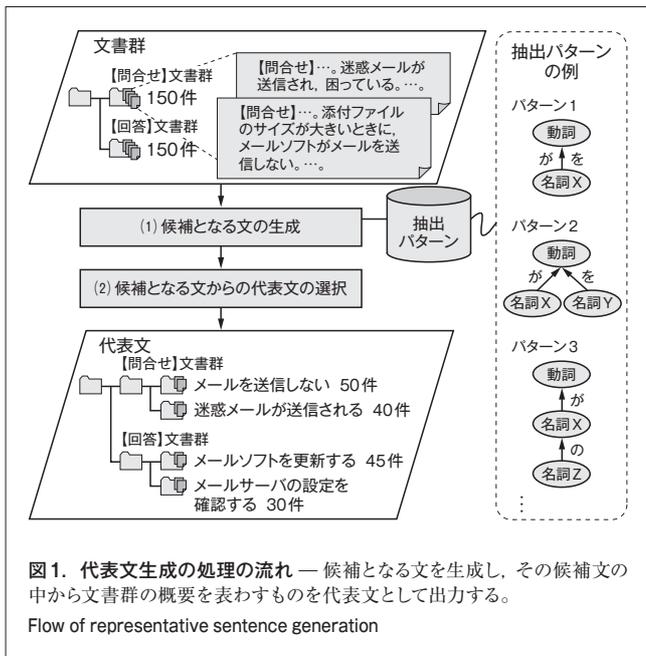
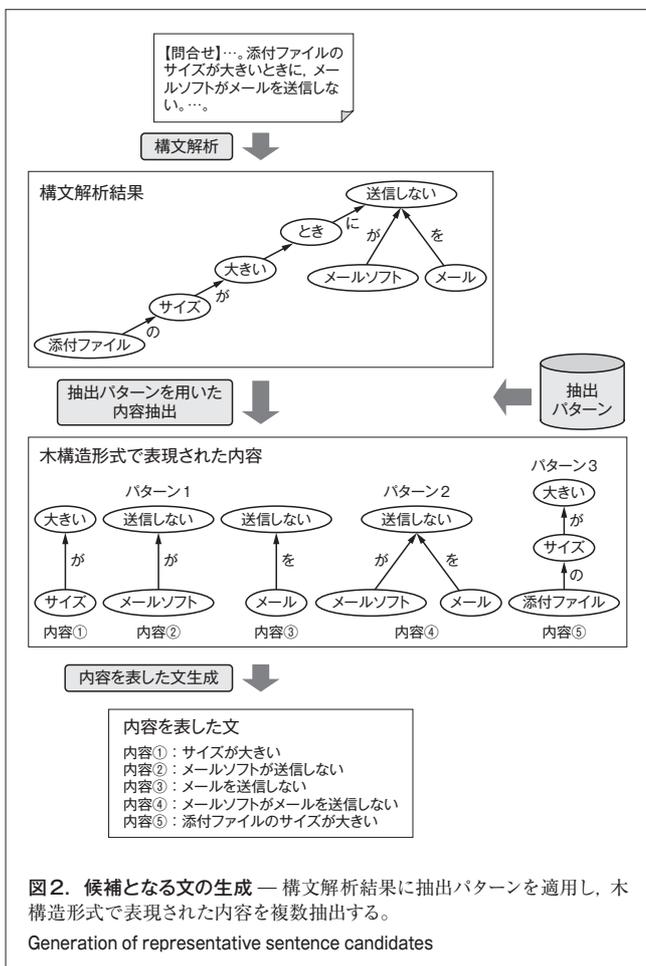


表1. 生成した代表文と文書数
Number of generated representative sentences

No.	代表文	文書数
1	プリンタを購入する	102
2	写真を印刷する	70
3	年賀状を作成する	66
4	パソコンに取り込む	24
5	インクを交換する	16
6	文字を印刷する	11
7	反応がない	11
8	ノズルが詰まる	11
9	インクが出ない	10
10	プリンタを共有する	9
11	インクがなくなる	8
12	プリンタが壊れる	7
13	線が入る	7
14	スキャナを使う	7



や“迷惑メールが送信される”のような、代表文が生成される。このように、文書群の概要が“何が何する”という文の形式で得られる。

Web上のQ&Aサイトに投稿されたプリンタに関する問合せ文書2,000件から生成した代表文と、その代表文と同じ内容が含まれる文書数を集計した。その結果の例を表1に示す。代表文“プリンタを購入する”や“写真を印刷する”，“年賀状を作成する”の文書数が多く、プリンタの購入や写真の印刷、年賀状の作成に関する問合せが多いことがわかる。

このように、代表文生成技術を用いると、文書群の概要が具体的に把握できるようになる。

3 コンタクトセンターソリューションへの適用

コンタクトセンターに集積されたコールトラッキングからFAQを作成する際に、代表文生成技術を適用することで、FAQ作成の効率化が期待できる。

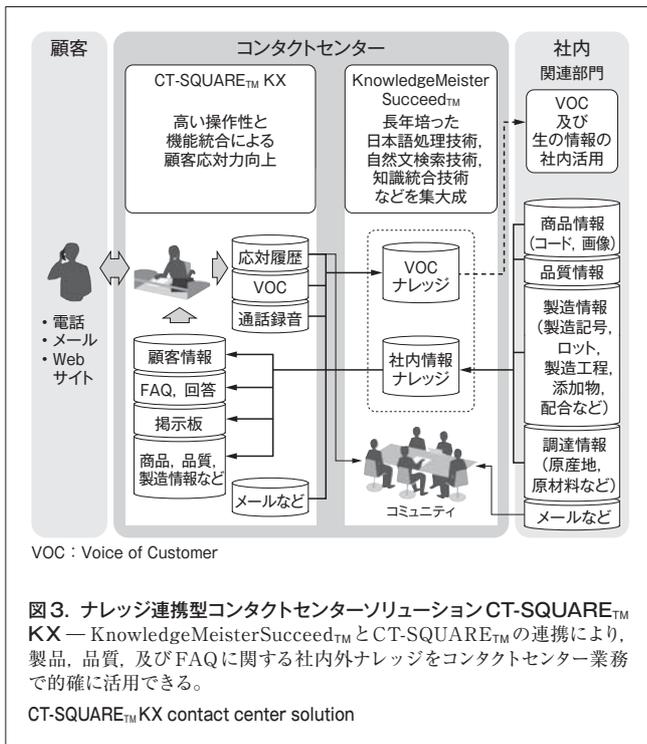
今回、当社のナレッジ連携型コンタクトセンターソリューションCT-SQUARE™ KXにおけるFAQ作成に適用した。

3.1 CT-SQUARE™ KXの概要

CT-SQUARE™ KXは、業務知識の継承を支援するKnowledgeMeisterSucceed™のエンジンを組み込んだ、ナレッジ連携型コンタクトセンターソリューションである(図3)。コンタクトセンターには、電話や電子メールなどで寄せられるコールトラッキングに加えて、FAQや商品情報など膨大な関連情報が蓄積されている。

CT-SQUARE™ KXは、日本語処理技術、自然文検索技術、及び知識統合技術などを集大成したナレッジエンジンを用いて、蓄積された情報を分析することで、顧客への対応時にオペレーターが分析結果を簡単にすばやく利用できるという特長を持つ。これにより、次の三つの課題を解決している。

- (1) 問題発生の予兆を早期に発見する。
- (2) 過去の対応や問合せを今後の対応に生かす。



このFAQのメンテナンスは、スーパーバイザー (SV) と呼ばれるコンタクトセンターの管理者が行う。SVは問合せと回答の内容を把握し、件数の多いものからFAQを整備するのが一般的である。対応業務中若しくは対応業務終了後といった時間的制約の大きいなかで、SVがコールトラッキングを参照するのは負担が大きい。また、人手で行うため、抜け漏れの発生する懸念もある。更に、FAQのメンテナンスは、緊急度の高い案件や通常業務と比較してコンタクトセンターの運営にとって必ずしも必要ではないため、対応が後手となり、やがてはFAQがオペレーターに使われなくなるというケースが散見される。

3.4 FAQ作成支援への代表文生成技術の適用

FAQのメンテナンスにおいて、FAQ作成に代表文生成技術が適用されると、SVの課題を解決できる。コールトラッキングの代表文が得られると、概要の把握が容易になる。更に、問合せ件数の定量的な分類結果が得られれば、FAQ作成の優先度管理強化や効率化が期待できる。

代表文生成技術を適用した際の、FAQメンテナンスの流れを図4に示す。コールトラッキングから生成された代表文を見ることで、出現頻度の高い問合せを把握でき、FAQとして作

(3) 社内の関連情報を効果的に活用して対応スピードと対応品質を上げる。

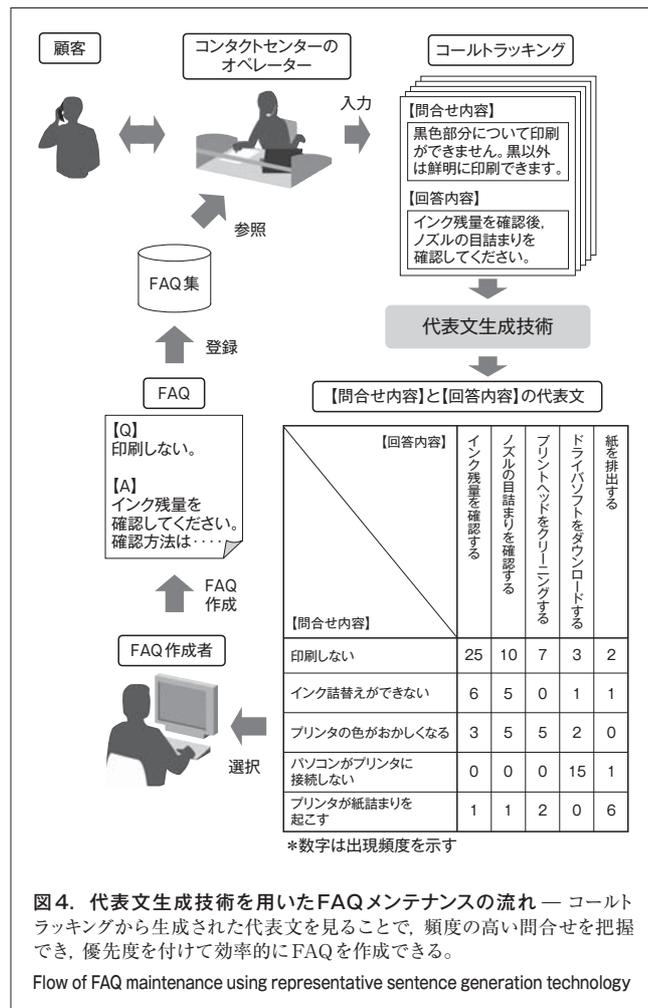
3.2 FAQを活用するための機能

関連情報を活用するための機能の一つに、自然文検索がある。オペレーターはFAQを検索し、検索して得られたFAQに基づいて対応を行う場合がある。この検索には自然言語解析を用いており、顧客からの問合せ内容をそのまま検索要求として使用できる。したがって、キーワードによる検索と異なり、商品知識の浅いオペレーターでも、スキルの高いオペレーターと遜色のない情報検索ができ、関連情報を活用した対応ができる。また、この検索でのオペレーターの操作は、問合せ内容の入力と検索ボタンのクリックだけであり、使い勝手の向上と業務の効率化を実現している。

検索や分析の対象となるナレッジとして、FAQだけでなく、品質情報や、製造工程の製造情報、原材料の調達情報、取扱説明書なども管理している。幅広い情報を管理対象とすることで、コンタクトセンターにナレッジツールが導入されているにもかかわらず、ナレッジの活用が期待どおりに進んでいないという声に対しても有益なソリューションである。

3.3 FAQ作成における課題

コンタクトセンターにおいて、最新のFAQが利用できることが顧客満足度の向上に寄与する。適切なFAQを参照することで、経験や商品知識に個人差のあるオペレーター誰もが、すばやく正しく、同じ内容を回答できるようになり、コンタクトセンターの品質の底上げと均質化を促し、更に顧客満足度の向上につながるからである。



成すべき内容の判断材料となる。出現頻度が高い問合せをFAQの候補として選択し、標準回答として作成した後、FAQ集に登録する。代表文生成技術を用いてコールドトラッキングの概要を把握できれば、SVがなんの支援もなしにFAQを作成する場合と比較して、FAQのメンテナンスにかかる時間を短縮できる。

4 代表文生成技術の有効性の評価

図4に示したように代表文生成技術を適用すると、FAQ作成者が、コールドトラッキングの代表文を参照して問合せの代表文からFAQの候補を選択し、原文を参照しながらFAQを作ることができる。その際選択した代表文が問合せの要点を正しく表現していると、その後の作業がスムーズになる。

そこで、FAQ候補としての適切さを評価指標とし、FAQ作成における代表文生成技術の有効性を評価した。更に、コールドトラッキングの分析にかかる時間も評価指標とした。コールドトラッキングのデータとして、Web上のQ&Aサイトに投稿された問合せを使用した。

以下、それぞれの評価の方法と結果を述べ、評価結果を考察する。

4.1 FAQ候補としての適切さ

4.1.1 評価方法 プリンタに関する問合せ文書500件から代表文を生成した。生成した代表文のうち、生成元の文書数が3件以上であるものは18個であった。この18個の代表文と、生成元の文書数を表2に示す。18個の代表文に関する

生成元の文書数の合計は150件であった。

問合せ文書から生成された代表文が、元の文書の問合せの要点を正しく表しているかどうかを評価するため、問合せの要点と比較した。問合せの要点の抽出は、FAQ作成の未経験者が問合せ文書の原文を読み、問合せ文書からFAQとして欲しい内容を抽出することで行った。

例えば、「年賀状を作ろうと一年ぶりにプリンタを出し印刷すると2色の色が出ません」という問合せについて、「年賀状を作る」、「一年ぶりにプリンタを出す」、及び「印刷する」はユーザーの操作で、「2色の色が出ない」は問題点であると判断し、問合せの要点として問題点である「2色の色が出ない」を抽出した。

4.1.2 評価結果と考察 問合せ文書において、代表文と問合せの要点が内容的に一致した文書数の合計は表2に示したように95で、生成元文書数合計150の63%を占める。

問合せの要点と一致しなかった代表文の事例から、要点に適切に表現できなかった原因を分析した。原因のうちいちばん多かったものは、問合せの文が複文のとき、主節から代表文が生成されないことであった。例えば、代表文「年賀状を作成する」の生成元文書の一つに、「年賀状を作ろうと一年ぶりにプリンタを出し印刷すると2色の色が出ません」という問合せがあった。この場合、望ましい代表文は「2色の色が出ない」であるが、生成元の文書数が1件だったため、3件以上という条件で選択したときに、代表文として得られなかった。

FAQを作成する際、選択した代表文が問合せの要点と異なると、問合せ文書を参照しながら、FAQを作成する作業がスムーズにできない。したがって、問合せの要点と一致しない代表文を減らすことがFAQ作成作業のトータルな効率化につながる。

この事例のように、複文の場合、主節と従属節の両方から候補となる文を生成しているため、問合せの要点と代表文が一致しないという現象が発生すると考えられる。この課題を解決するには、条件を表す複文の場合は、主節だけから候補となる文を生成することが考えられる。

4.2 データの分析及び問合せの選択にかかる時間

4.2.1 評価方法 Q&Aサイトの500件のデータを用いて、人手により問合せの内容を分析し、FAQとして登録すべき問合せを選択する作業を行い、その作業時間を計測した。作業は、FAQ作成の未経験者がひとりで行った。

4.2.2 評価結果と考察 人手による作業時間は、12時間であった。一方、代表文生成エンジンの処理時間は10分であった。実際には、代表文の全てがFAQとならないため、更に、FAQとして登録すべき代表文を選択する作業が発生する。代表文の出力数が18個の場合、選択時間は1分程度であった。したがって、代表文生成エンジンを用いた場合、作業時間は、エンジン処理時間と選択時間の合計であり、10分程度と考えられる。

表2. FAQ候補としての適切さ評価結果

Results of evaluation of appropriateness of FAQs for representative sentences

No.	代表文	生成元文書数	代表文と問合せの要点が一致した文書数
1	プリンタを購入する	35	22
2	年賀状を作成する	31	19
3	プリンタを接続する	15	12
4	写真を印刷する	13	12
5	インクがなくなる	6	3
6	ドライバをインストールする	5	4
7	電源を入れる	5	0
8	ヘッドクリーニングをする	5	1
9	インクを購入する	4	0
10	プリンタが壊れる	4	1
11	スキャナを使う	4	3
12	デジタルカメラで撮影する	4	3
13	反応がない	4	4
14	パソコンに取り込む	3	3
15	ワープロソフトで作成する	3	3
16	用紙をセットする	3	2
17	目詰まりを起こす	3	3
18	パソコンで描く	3	0
合計		150	95

4.3 FAQ作成の効率化

代表文生成技術の有効性を、生成された代表文の適切さと作業時間という二つの観点で評価した。代表文生成エンジンを用いると、FAQに載せるべき問合せを見出す作業が、このエンジンの出力する代表文を見て選択するだけでできるようになる。作業時間も、人手で行うと日単位の時間がかかった作業が分単位に短縮される。代表文生成技術をコンタクトセンターソリューションに適用することで、コンタクトセンターでのFAQ作成の効率化が期待できる。

5 あとがき

代表文生成技術をコンタクトセンターソリューションへ適用したときの有効性を評価した。代表文生成技術は、文書群の概要を文の形式で表すことができ、内容が具体的に把握できるという特長を持つ。この技術を活用することで、FAQの作成が効率化できるという評価結果を得た。

今後更に、SVによる実践的な評価も行っていくとともに、CT-SQUARE™ KXのFAQ作成機能への代表文生成技術を利用した支援機能追加を検討していく。

また、問合せや回答といった分析対象の文書の性質に応じて、適切な代表文を生成する機能の開発など、FAQ候補としての適切さの向上に取り組んでいく。

更に、技術動向の把握や技術戦略の立案に役だつ知識を得るために、特許分析への適用にも取り組む。特許分析では、発明が解決しようとする課題や解決手段に関する代表文を明細書から生成することで、特許マップの作成を支援する活用を考えている。

文献

- (1) 早川ルミ 他. 日本語解析技術を活用した業務支援ソリューション開発への取り組み. 東芝レビュー. 64, 2, 2009, p.30-34.
- (2) 倉田早織. 大量文書の内容が一目でわかる代表文生成技術. 東芝レビュー. 65, 2, 2010, p.68-69.
- (3) 倉田早織 他. “文書クラスターの階層構造を利用した代表文の生成”. 言語処理学会第16回年次大会. 東京, 2010-03, 言語処理学会. 2010, p.43-46.



倉田 早織 KURATA Saori, Ph.D.

東芝ソリューション(株) IT技術研究所 研究開発部主任、理博。言語処理分野の研究・開発に従事。情報処理学会、言語処理学会会員。

Toshiba Solutions Corp.



小川 哲男 OGAWA Akio

東芝ソリューション(株) 業務ソリューション事業部 クロスインダストリーソリューション部。パッケージ型CRMソリューションの設計・開発に従事。

Toshiba Solutions Corp.



加納 敏行 KANO Toshiyuki

東芝ソリューション(株) IT技術研究所 研究開発部主任 研究員。情報知識利活用技術の研究・開発に従事。日本オペレーションズ・リサーチ学会、言語処理学会会員。

Toshiba Solutions Corp.