The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

製品マニュアル文からの質問自動生成

Automatic generation of questions from product manual sentences

佐藤 紗都[*1] 伍井 啓恭[*2] 奥村 学[*3]
Sato Sato Hiroyasu Itsui Manabu Okumura

[*1] 東京工業大学工学院 School of Engineering Tokyo Institute of Technology
[*2] 三菱電機株式会社 Mitsubishi Electric Corporation
[*3] 東京工業大学科学技術創成研究院 Institute of Innovative Research Tokyo Institute of Technology

In this research, as the first step to develop a system that automatically generates related question answer pairs from sentences in a product manual, we present a method to automatically generate question sentences from manual sentences. As the result of experiments with a data set of about 1400 sentences, we obtained BLEU score of 62.11 by comparing generated sentences with manually prepared question sentences.

1. はじめに

電化製品の高度化の 1 つの方向性に，製品に対するユーザの質問に音声などで電化製品自身が応答することが挙げられる．例えば，炊飯器に向かって「白米でお粥を作るとき、どのくらい時間がかかりますか？」と聞いたときに，炊飯器が「約 41～47 分かかります」と答えることができれば有用と考えられる．しかし，このような質問応答システムの構築には製品に関する質問応答対が必要であり，これを人手で作成することには大きなコストがかかる．一方で，製品に必ず付属している製品マニュアルは製品の使用方法などが網羅されているため，これを用いて自動で質問生成することができればそのコストを削減することが可能である．そこで，本研究では製品の使用方法などが網羅されている製品マニュアルを用いて，関連する質問応答対を獲得する第一歩として，マニュアル文から関連する質問文を自動生成する手法を示す．

具体例を図 1 に示す．例 1 に示すように入力のマニュアル文「炊飯中の「チリチリ」等の音はIH特有の通電音によるもので、故障ではありません」に対し質問文，「炊飯中に変な音がするのは故障ですか？」と出力する．また，例 2 のようなマニュアル独特の入力に対しても対応できるようにする．

質問を自動生成する試みは古くから行われており，規則に基づく手法，テンプレートを用いる手法，ニューラルネットワークを用いる手法の 3 つに大きく分類することができる．

規則に基づく手法の1つとして，平叙文を質問文に変換する汎用規則を用いて質問文候補を複数生成し，分類器を用いてランキングする手法[Heilman 2009]が提案されているが，英文の質問生成特有のルールが多く，日本語に直ちに適応することは難しい．

テンプレートを用いる手法として，質問応答サイトに対する検索履歴ログを用いて，検索クエリを入力として，自動的にテンプレートを生成し，質問を自動生成する手法[Zhao 2011]が提案されており，本研究にもっとも類似するが，大量の検索履歴が利用可能であることを前提としたものであり，大量に入手できない場合には不向きである．

ニューラルネットワークを用いる手法の1つとして，教師付き学習と強化学習を組み合わせてリカレントニューラルネットワークモデルを学習させる手法[Yuan 2017]が挙げられるが，この手法は質問応答データが大量に入手できない場合には不向きである．

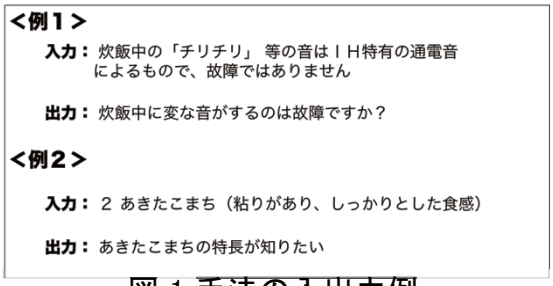

図 1 手法の入出力例

一方で打者成績とイニング速報からテンプレートを生成し，それを用いて，打者成績からイニング速報を生成する手法[村上 2016]が提案されている．この手法ではデータ数が少ない場合でも対応可能であるため，本研究ではこのテンプレート生成手法を用いて質問生成を試みる．

以下その具体的な手法を説明し，それに対する評価実験の結果と考察を述べる．

2. 提案手法

2.1 システム概要

図 2 に質問生成システムの概要を示す．提案手法は大きく学習フェーズと質問文生成フェーズの 2 つに分けられる．

学習フェーズでは，人手で作成されたマニュアル文と質問文の対を訓練データとして，マニュアル文と質問文の自動対応付けを行い，対応した単語個所をタグに抽象化した後，クラスタリングを行い，マニュアル文-質問文の対の代表テンプレートリストを生成する(テンプレート生成)．

質問生成フェーズでは，入力されたマニュアル文に対して，自動対応付けにより，生成した代表テンプレートリストから候補テンプレートを抽出し，タグを単語で置換することで質問文を生成する(質問文生成)．

次節からテンプレート生成と質問文生成について詳しく説明する．

連絡先：佐藤紗都，東京工業大学工学院，横浜市緑区長津田町 4259 R2-7，satosato@lr.pi.titech.ac.jp

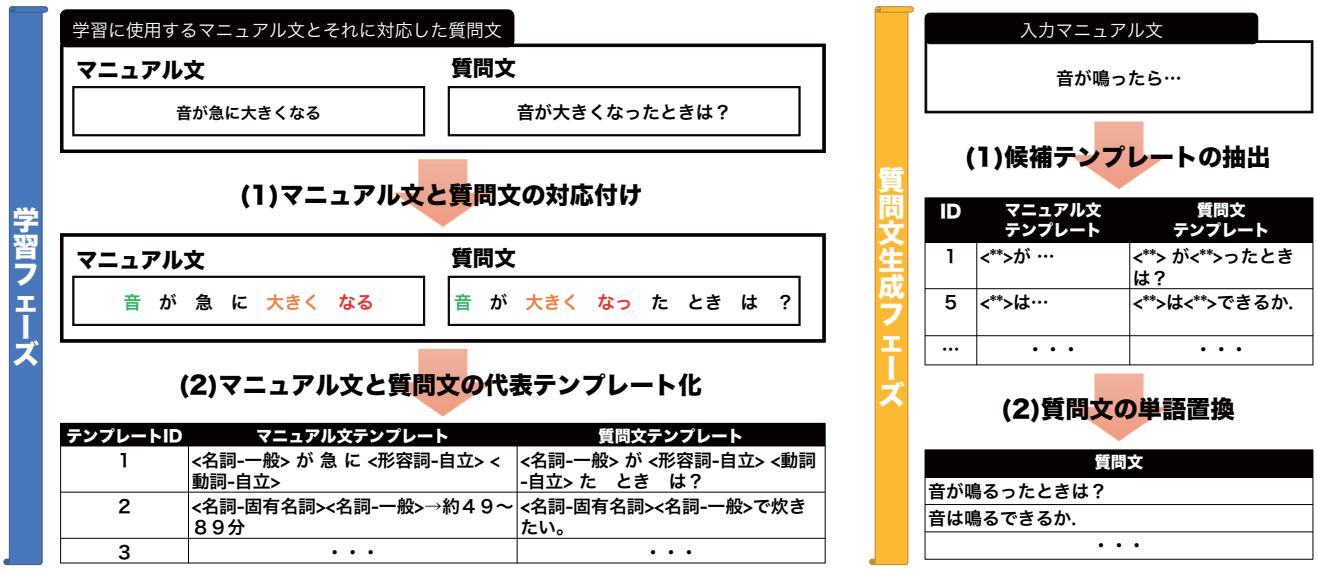

図 2 質問文生成システム概要

2.2 テンプレート生成

テンプレート生成では大きく分けて，(1) マニュアル文と質問文の対応付け，(2) マニュアル文と質問文の代表テンプレート化という 2 つの処理を行う．

(1) マニュアル文と質問文の対応付け

MeCab(http://mecab.sourceforge.jp/)を用いてマニュアル文と質問文をそれぞれ形態素解析する．次に，その結果に対し，Jing ら[Jing 1999]の隠れマルコフモデルに基づいたアライメント手法を適用し，マニュアル文と質問文の原形単語間の対応付けを行う．

(2) マニュアル文と質問文の代表テンプレート化

(1)の結果を用い，まずマニュアル文と質問文の汎化を行う．具体的には，(1)で対応付いた単語のうち，自立語であるものは変数化して良いと考え，マニュアル文と質問文のその単語個所を「<品詞-品詞細目 1>」という品詞タグに置き換えることで汎化を行う．例えば，単語が一般名詞であった場合は「<名詞-一般>」，形容詞自立であった場合は「<形容詞-自立>」と置き換えるため，図1では「音が急に大きくなる」というマニュアル文が「 <名詞-一般> が 急 に <形容詞-自立> <動詞-自立>」という風に汎化される．ここで対応づいた自立語の単語をタグに置き換えた文をテンプレート文と呼ぶ．

次に，マニュアルテンプレート文-質問テンプレート文の対をクラスタリングする．具体的には，質問文に含まれる内容語及びタグの TF-IDF 値で構成されるベクトルを特徴量として k-means 法によるクラスタリングを行い，生成されたクラスタごとに固有のテンプレート ID を付与する．

最後に，生成されたクラスタの中心に最も近いテンプレート対をそれぞれのクラスタの代表テンプレート対として，代表テンプレートリストを生成する．

2.3 質問文生成

質問文生成では大きく分けて，(1) 候補テンプレートの抽出，(2) 質問文中のタグの単語置換という 2 つの処理を行う．説明のために，あるクラスタ k の代表マニュアルテンプレートを rep_k，入力されたマニュアル文を m とする．

(1) 候補テンプレートの抽出

2.2 節(1)と同じアライメント手法を用いて，入力マニュアル文と代表マニュアルテンプレートの対応付けを行い，一定以上の割合対応付けが取れたものを候補テンプレートとする．図 3 に入力マニュアル文と代表マニュアルテンプレートの対応づけの流れを示す．

まず，m を MeCab で形態素解析し，rep_k のタグになっている部分と同じ位置の m の単語個所を品詞タグに置き換える．この置き換えを行なった m を m'とする．さらに置き換えた m の単語の品詞をインデックス，品詞ごとの m の単語リスト(単語の出現順にソート)を値とした品詞単語辞書を保持しておく．

次に，2.2 節(1)と同じアライメント手法を用いて，rep_k と m'の単語間の対応付けを行う．また，rep_k と m'の単語間の対応付けがされた割合(対応割合)を求める．対応割合Rは rep_k の単語数をl_r，m'の単語数を$l_{m'}$，対応が取れた単語数をl_aとすると次の式で表すことができる：

$$R = l_a / max(l_r, l_{m'}).$$

以上の操作を各代表テンプレートについて行い，パラメタとして与えられた対応割合下限R_{limit}以上の代表マニュアルテンプ

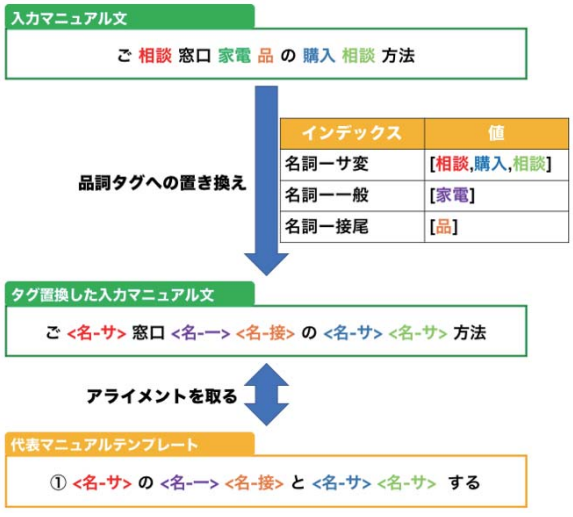

図 3 入力文と代表テンプレートの 対応づけ

レートと代表質問テンプレートの対を候補テンプレート対として出力する.

(2) 質問文中のタグの単語置換

図 4 に示すように品詞単語辞書を用いて, 出力された候補テンプレート対の代表質問テンプレートの品詞タグを単語に置換する.

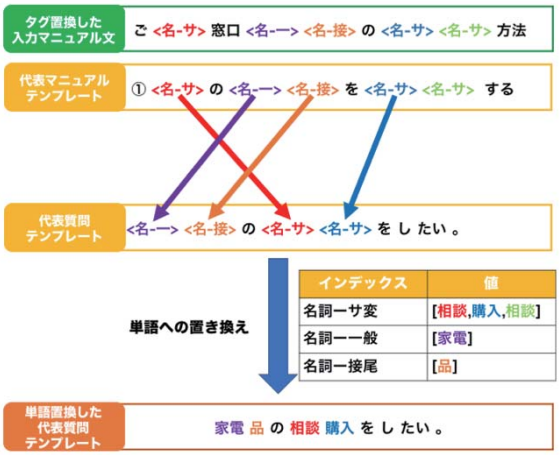

図 4 質問文の単語置換

具体的には代表質問テンプレートの先頭から 1 単語ずつ見ていき, もし品詞タグであった場合, 品詞単語辞書のそのインデックスを持つ単語リストの先頭の単語に置換し, その単語を単語リストから除外する処理を文末まで繰り返す.

3. 実験

実験には公開されている冷蔵庫と炊飯器のマニュアルの各文および質問文を人手で対応づけした 1358 文対を利用した.

3.1 テンプレート生成

まず, 代表テンプレートリストを生成するため, テンプレート生成実験を行なった. MeCab で形態素解析を行う際に, 新語を含むテキストの形態素解析用辞書として, mecab-ipadic-NEologd (https://github.com/neologd/mecab-ipadic-neologd) を用いた.

変化させたパラメタはクラスタ数・テンプレート変数化割合・クラスタリング特徴量である. クラスタ数は 10 および 50 から 300 まで 50 刻みの値で変化させた. テンプレート変数化割合とは質問文の汎化を行なう際に質問文中の単語数のうち品詞タグに変わった割合である. このテンプレート変数化割合が高ければ高いほど, 汎化していると言える. テンプレート変数化割合の閾値とその閾値以上の質問文数を表 1 に示す. また, 1358 文のテンプレート変数化割合の平均は 29%で, 1 単語も変数化しなかった文は 409 文であった. テンプレート変数化割合の閾値は 5~15% の間で変化させた.

クラスタリング特徴量は Unigram, Bigram, Trigram, Unigram + Bigram, Bigram + Trigram, Unigram + Bigram + Trigram, それぞれの TF-IDF を取った特徴量をそれぞれ試した.

評価はシルエット係数[Rousseeuw 1987]を用いた. シルエット係数とは各サンプルについて算出され, k-means などによってクラスタリングされた各サンプルがその割り当てられたクラスタ内にどの程度収まっているかを示す値(1 が最高値, -1 が最低値)である.

表 1 テンプレート変数化割合と質問文数

テンプレート変数化割合の閾値[%]	質問文数[文]
1	949
5	949
10	895
15	811
20	680
25	543

3 つのパラメタを変化させた結果, 全サンプルに対するシルエット係数の平均が大きくなった場合の上位 5 位を表 2 に示す.

表 2 クラスタリング結果

テンプレート変数化割合閾値[%]	クラスタリング特徴量	クラスタ数	平均シルエット係数
15	Unigram + TF-IDF	300	0.23
15	Unigram + TF-IDF	250	0.22
10	Unigram + TF-IDF	300	0.21
15	Unigram + TF-IDF	200	0.2
10	Unigram + TF-IDF	250	0.2

表 2 からテンプレート変数化割合閾値 15%, クラスタリング特徴量 Unigram + TF-IDF, クラスタ数 300 の場合がもっとも平均シルエット係数が高かったため, このクラスタで代表テンプレートリストの生成を行った. またこのリストを用いた場合, 代表文とアライメントが取れる単語数およびその品詞が一致し, 質問文生成可能であると考えられる質問数は 375 文であった. 以下にテンプレート例を 2 つ示す.

(マニュアルテンプレート例 1)
マニュアルテンプレート:<名詞-一般>(<名詞-固有名詞>) 約 8 4 ～ 88分
質問テンプレート: <名詞-一般>で<名詞-固有名詞>を作るとき、どのくらい時間がかかる？
(マニュアルテンプレート例 2)
マニュアルテンプレート: <名詞-一般> の <名詞-一般>
質問テンプレート:<名詞-一般> の <名詞-一般> について

また, クラスタに属するサンプルのシルエット係数をクラスタごとに平均したクラスタシルエット係数平均の分布は図 5 のようになった. このように全体の 96%のクラスタシルエット係数平均が 0 か正の値であった. 以下, このクラスタシルエット係数平均の下限を指定したクラスタを用いて質問生成を行う.

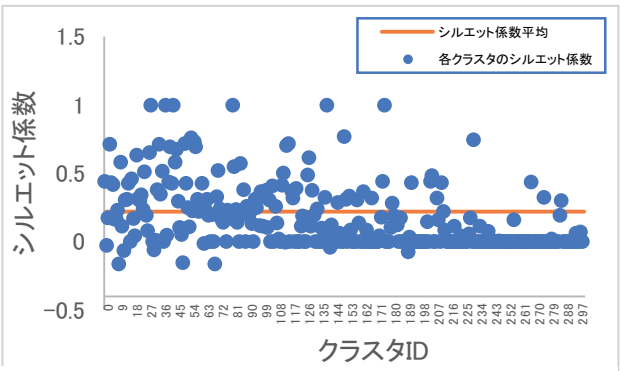

図 5 クラスタシルエット係数平均の分布

3.2 質問文生成

前節で得たテンプレートリストを用いて，質問文生成実験を行なった．人手で作成された正解質問文と生成された質問文の差異を機械翻訳の評価として用いられる BLEU[Papineni 2002]で評価した．また，生成可能である文のうち質問文を生成できた文の割合も評価として用いた．対応割合下限R_{limit}を 60%から 95%まで 5%刻みで変化させ，クラスタシルエット係数平均下限S_{limit}を-0.2(クラスタシルエット係数平均の最低値)から 0.2 まで 0.1 刻みで変化させた結果の BLEU 上位 5 位を表 3 に示す．

表 3 質問文生成結果

対応割合下限[%]	クラスタシルエット係数平均下限	BLEU	平均生成質問数[文]	生成できた割合[%]
70	−0.1	62.11	1.75	93.07
75	−0.1	62.11	1.76	92.53
80	−0.1	62.03	1.77	90.93
75	−0.2	62.02	1.75	93.33
70	−0.2	62.01	1.75	93.87

表 3 で BLEU と質問文を生成できた割合が最も高い対応割合下限 70%，クラスタシルエット係数平均下限-0.1 の場合について生成例を 2 つ示す．

（生成例 1）
入力マニュアル文:「うま早」「お急ぎ」はできません
正解質問文:寿司用のご飯を「お急ぎ」で炊きたい 。
生成質問文 1:寿司用のご飯を「お急ぎ」で炊きたい 。
生成質問文 2:中華粥を「お急ぎ」で炊きたい 。
[ほか 9 文，同様の炊飯器でできる調理質問が出力された]
（生成例 2）
入力マニュアル文:①音が急に大きくなる。音色が変わる→お使い始め、暑いとき、ドアの開け閉めが多いなどのときに高速運転に切り替わり、強い力で冷やします
正解質問文:音が大きくなった
生成質問文:音が大きくなった

生成例から，正解とは異なるが自然な質問文が得られていることおよび正解文と同様の質問文が得られていることがわかる．

また，生成結果が人手作成と同一でない生成文について，被験者による 5 段階の評価尺度(MOS 値)による主観評価を実施した．表 4 に実験設定，図 6 にその結果を示す．

表 4 実験設定

項目	詳細
被験者	12 人(男性 7 人，女性 5 人)
人手作成文	ランダム抽出した(生成不可文を除く)50 文中，人手生成文と異なる結果であった 26 文
自動生成質問文	上記に対応する自動生成質問文 38 文（N ベストを含む）
評価尺度(MOS 値)	5:非常に良い，4:良い，3:どちらでもない，2:悪い，1:非常に悪い，の 5 段階評価

生成文のうち 48%が人手作成文と同じ文となったが，図 6 からもわかるように残りの 52%の生成文の約 6 割(生成文全体の 30.1%)が MOS 値 4 以上となった．同一文を除く生成文の MOS 値平均は 3.50 で人手作成文は 4.25 であった．これらから，生成文全体のうち 78.1%が人手作成文と同一か主観評価 4 以上であることを確認できた．

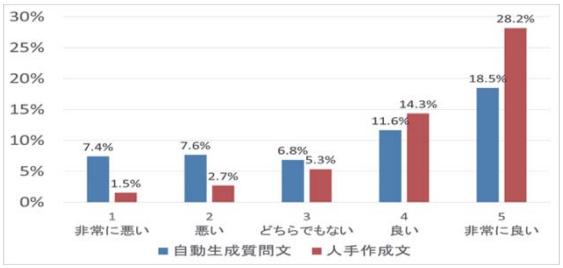

図 6 同一文を除く文の MOS 値比較

4. おわりに

本研究ではマニュアル文から関連する質問文を自動生成する手法を示した．実験の結果，人手作成文を正解として，BLEU スコア 62.11 が得られるとともに，生成文全体のうち 78.1%が人手作成文と同一か主観評価 4 以上であることが確認できた．

今後の課題は大きく分けて 2 点考えられる．第一にクラスタリングの精度の向上による代表テンプレートリストの改善と穴埋めアルゴリズムの改良を行い，よりよい質問文を増やすことである．第二に解答文生成システムも同様に構築し，製品マニュアルに書かれている記述から，関連する質問応答対を自動生成するシステムを構築する予定である．

参考文献

[Heilman 2009] Michael Heilman et al.: Question Generation via Overgenerating Transformations and Ranking , *CARNEGIE-MELLON UNIV PITTSBURGH PA LANGUAGE TECHNOLOGIES INST*, 2009.

[Zhao 2011] Shiqi Zhao et al.: Automatically generating questions from queries for community based question answering , *Proceedings of the 5th International Joint Conference on Natural Language Processing* , pp.929-937, 2011.

[Yuan 2017] Xingdi Yuan et al.: Machine Comprehension by Text-to-Text Neural Question Generation, *Proceedings of the 2nd Workshop on Representation Learning for NLP* , Association for Computational Linguistics , pp.15-25, 2017.

[村上 2016] 村上 聡一朗, 他.: 打者成績からのイニング速報の自動生成, *言語処理学会 第 22 回年次大会 発表論文集*, pp.338-341, 2016.

[Jing 1999] Hongyan Jing et al.: The decomposition of human-written summary sentences, *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp.129-136, 1999.

[Rousseeuw 1987] Peter J.Rousseeuw: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics 20* , pp.53-65, 1987.

[Papineni 2002] Papineni et al.: BLEU: A Method for Automatic Evaluation of Machine Translation, *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp.311-318, 2002.