DBSJ Japanese Journal
Vol. 16-J, Article No. 4, March 2018

# ニューラル翻訳モデルを用いた回答文書検索のための拡張クエリ生成手法

# Neural Translation Model based Query Generation Method for Answer Document Retrieval

大塚 淳史♡ 別所 克人♢ 西田 京介♡
浅野 久子♢ 松尾 義博♢

**Atsushi OTSUKA　Katsuji BESSHO**
**Kyosuke NISHIDA　Hisako ASANO**
**Yoshihiro MATSUO**

マニュアルや **FAQ** 等の文書を検索する際，ユーザは解決したい事柄を質問として検索システムに入力する．しかしながら，文書に記載されている解決方法は，必ずしも質問に対してキーワードの一致や意味的な類似性があるとは限らず，質問に対して満足のいく検索結果をシステムが提示できないという課題があった．本論文では，質問と回答の内容の関係性をあらかじめ学習し，質問に対する回答内容を先読みすることで，質問を解決するための最適な文書を検索する手法を提案する．質問に対する回答に含まれる可能性の高いキーワードを生成する **Encoder-Decoder** モデルを学習することで，質問に対して，回答内容を予測するモデルを作成する．学習した **Encoder-Decoder** モデルが生成したキーワードを追加した拡張クエリによって回答を検索することで，答えを先読みした検索を実現する．実験の結果，回答内容を学習した **Encoder-Decoder** モデルから生成した拡張クエリを用いることで，**FAQ** 検索の検索精度が有意に向上することを確認した．

**We propose a novel Frequently Asked Question (FAQ) retrieval technique with a neural query expansion model. With the growth in Artificial Intelligence (AI) based systems and mobile communications, FAQ retrieval systems have become widely used in site searches and call center support. However, FAQ retrieval often has lexical gaps between queries and answer documents. To bridge these gaps, we design a query expansion model on the basis of an Encoder-Decoder model as a type of deep neural network. The model learns the words that appear in answers for questions using Q&A pair documents and generates the expanded queries from inputted queries to retrieve answer documents. We evaluate our proposed technique in a multi-domain FAQ retrieval task. Experimental results show that our technique retrieves FAQs more accurately than the previous methods.**

♡ 正会員　日本電信電話株式会社 NTT メディアインテリジェンス研究所
otsuka.atsushi@lab.ntt.co.jp

♢ 非会員　日本電信電話株式会社 NTT メディアインテリジェンス研究所

## 1. はじめに

自然文やキーワードによって所望の情報を入手する情報検索システムは，Web 検索を始め，スマートフォンによる音声検索，音声対話ロボット等様々な場面で活用されている．特に近年では，人工知能（AI) による業務支援の一つとして，FAQ 検索によるコンタクトセンタのオペレータ支援が注目を集めている [23]．コンタクトセンタでは，電話での問い合わせ内容（クエリ）と，応対マニュアルである FAQ 集とを照合し，応対内容を決定している．コンタクトセンタにおける FAQ は数百～数千の QA 文書から構成されることが多く，膨大な QA 文書集合の中から，問い合わせ内容に対応する QA 文書を発見することは，オペレータにとって非常にコストが大きい作業となる．そのため，問い合わせ内容に近い QA 文書を自動で発見する FAQ 検索の重要性はますます高まっている．

FAQ 検索では，QA 文書の質問（Q) 部分に着目し，クエリと Q に記述されている内容が意味的に同一あるいは類似しているかどうかを見ることで高い精度で検索できることが明らかとなっている [22, 5]．しかしながら，Q との意味の類似性だけでは，関連性を判定できない問い合わせや検索クエリも多く存在する．例えば，インターネット接続サービスに関する「動画が急にカクカクするようになった」という問い合わせがあったときに,「動画を見すぎて帯域制限になってしまった可能性があります」という応対をしたい場合を考える．FAQ に「動画がカクカクになったら」のような個別の事象についての QA 文書が存在していることは稀である．オペレータは，動画が急にカクカクしだした原因が，帯域制限による速度低下であると推測して,「帯域制限になった場合の対応方法」という記載の Q の応対方法である回答（A）となる文書を参照して応対することになる．ここで，最終的に閲覧する「帯域制限になった場合の対応方法」と，問い合わせ内容の「動画が急にカクカクするようになった」は意味的な類似性がほとんど存在していないため，クエリと Q の内容の類似性のみを考慮した FAQ 検索では対応できない．

クエリと Q との類似性だけでは関連する QA 文書を発見できない場合，クエリと回答文書に関連性を計算する必要がある．しかしながら，クエリと回答文書との間には，使用される語彙に乖離があり，単純なキーワード一致などの手法では関連する回答を発見出来ないという問題がある [3]．従来の研究では回答文書を検索するために，QA 間の単語共起 [21, 7] や統計的翻訳モデル [20, 14] により回答文書で使用されるキーワードを推定する手法が提案されてきた．しかしながら，これらの手法は単語レベルでの比較であり，複数単語の組み合わせなどは考慮できていない．またモデル作成のために巨大なアライメントデータが必要とした．

本論文では，入力クエリ対して回答文書に出現する単語を予測し生成するモデルを作成する．生成された単語で回答文書を検索することで，適切な QA 文書を提示する情報検索手法を提案する．回答文書に出現する単語の予測には Encoder-Decoder モデルをベースとしたキーワード生成モデルを用いる．Encoder-Decoder モデルは機械翻訳などで多く利用されるニューラルネットワークモデル [16] であり，アライメント等のデータを必要とせず，End-to-End で翻訳モデルを作成できる．Encoder-Decoder モデルの入力を質問，出力を回答とすることで，大量の QA 文書から質問から回答への生成モデルを学習する．また，翻訳モデルである Encoder-Decoder モデルに対して，検索のクエリ生成に特化した改良を加え，モデル学習の高速化や検索に有用なキーワードを生成させる手法も提案する．

評価実験では，応対ログとして Web の質問回答サイトから収集した質問と回答をペアとした QA 文書コーパスによって，Encoder-Decoder モデルを学習する．評価用に作成した質問のタイプが異なる 3 つのドメイン (iphone，コスメ，恋愛相談) の FAQ に対する検索精度を測定する実験を行う．

**本論文の貢献**：本論文は，入力クエリに対して，クエリの答え

となる内容を予測することによる文書検索手法について以下の貢献を果たした．

- Encoder-Decoder モデルをベースとした，高速に学習可能な回答文書検索用の拡張クエリ生成モデルおよび，生成した拡張クエリを用いた FAQ 検索手法を提案．
- 質問のタイプが異なる 3 つのドメインに対して FAQ 検索実験を行い，情報検索型の質問が多い FAQ に対して提案手法が特に有効であることを明らかにした．

本論文の構成は以下の通りである．まず，2 節で本論文での定義とタスク設定を示す．次に 3 節で提案手法について説明する．4 節では評価実験について，実験方法と結果について説明し，5 節で考察を行う．6 節では本論文に関連する先行研究について述べる．最後に 7 節で本論文のまとめを述べる．

## 2. 問題設定

本節では，本論文で対象とする FAQ ならびに学習コーパス，入力クエリの形式について定義する．また，提案手法である回答生成モデル，FAQ 検索のタスクについての定義も行う．

### 2.1 FAQ・学習用 QA・入力クエリ

本研究で扱うテキストである FAQ，学習用 QA，入力クエリについての定義を以下に示す．

**[定義 1]** FAQ は，自然文で記述された QA 文書の集合 $FAQ = \{QA_0, QA_1, ..., QA_i, ..., QA_{|M|-1}\}$ である．本論文において FAQ は，FAQ 検索タスクにおける検索対象文書集合を指すものとする．ここで，$|M|$ は FAQ の全 QA 文書数である．

**[定義 2]** QA 文書は，質問が記載されている文書 $q_i$ と回答が記載されている文書 $a_i$ の組となる文書 $QA_i = (q_i, a_i)$ である．一つの質問（Q) には，必ず一つの回答（A) が存在するものとする．

**[定義 3]** 学習用 QA は，FAQ に関連する話題に関して収集したモデル学習用のコーパスである．回答生成モデルの学習は，学習用 QA 用いて行うものとする．学習用 QA は [定義 2] と同様の QA 文書の形式であるとする．

**[定義 4]** 入力クエリは，自然文または $m$ 個 の単語の集合で表現される．入力が自然文の場合は入力を形態素解析した後，内容語のみを抽出したものを使用する．

### 2.2 回答生成モデル

本論文では，Encoder-Decoder モデルにより，入力クエリに追加するキーワードを生成する．キーワードを生成する回答生成モデルは以下の通り定義する．

**[定義 5]** 回答生成モデルは，入力クエリを入力した時，0〜$N$ 個の単語の列を出力する変換器である．学習では，入力クエリに対して最適な追加キーワードを得るための変換関数 $f$ を学習する．

### 2.3 FAQ 検索タスク

本論文で提案手法を評価する FAQ 検索タスクは以下の通り設定する．

**[問題 1]** 自然文の入力クエリを入力した時，$FAQ = \{QA_0, QA_1, ..., QA_i, ..., QA_{|M|-1}\}$ を，入力クエリに関連する順番に並び替え，上位 $k$ 件の QA 文書を出力する．

## 3. 提案手法

本節では，QA 文書を対象に，回答文書に記載されている内容を予測することによる FAQ 検索手法について述べる．本論文での FAQ 検索のスコア算出までの流れを図 1 に示す．QA 文書に対して，Q のみを使った質問スコアを算出と，Q と A どちらも使用した回答スコアを別々に算出し，最後にスコアを計算する．ここで質問スコアは，先行研究である入力クエリと Q の意味同一に基づく手法 [22] を使用し，統合スコアは質問スコアと回答スコアの線形和によって算出する．このとき，線形和をとるための重みはランキング学習 [10] により決定する．本論文では，回答スコアの算出とそのための回答検索用クエリ拡張について提案する．

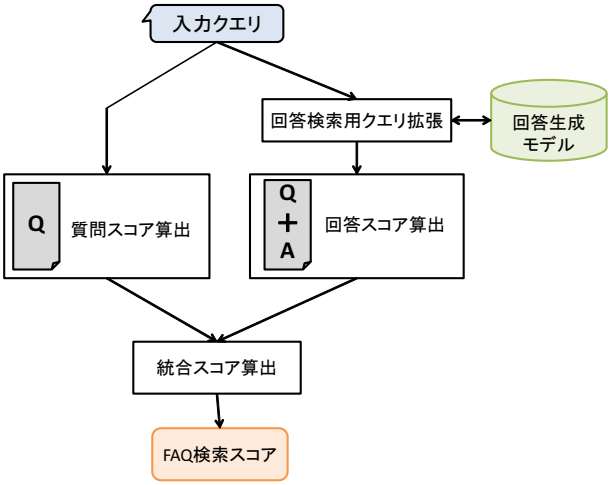



図 1: 回答を考慮した FAQ 検索スコアの算出

まず 3.1 で，回答生成モデルのベースとなる Encoder-Decoder モデルについて述べる．次に，3.2 で質問・回答の関係性を学習する回答生成モデルについて詳述し，最後に 3.3 で拡張クエリを用いた回答スコアの算出について説明する．

### 3.1 Encoder-Decoder モデル

Encoder-Decoder モデルとは，ニューラルネットワークを用いて，系列データから異なる系列データを得るためのモデルである．入力の系列データをエンコード用のニューラルネットワークに順次入力し，1 つの入力ベクトルを作成する．そして，エンコードによって作成したベクトルをデコード用のニューラルネットワークに入力し，出力結果を得ることで，系列データの変換を行っている．

#### 3.1.1 Seq2Seq Model

自然言語処理における，Encoder-Decoder モデルとしては機械翻訳 [16] や対話処理 [17] で用いられている Seq2SeqMoldel がある．Seq2SeqModel では，単語などの言語表現を全てベクトルに変換することによって，Encode-Decode モデルへの適用を可能としている．Seq2SeqModel の構成を図 2 に示す．ここで，$w_0, w_1, w_2$ は入力の単語列，$w'_0, w'_1$ は出力の単語列を示している．以降は Seq2SeqModel のそれぞれの層について説明する．

単語を表現する最も基本的なベクトルは one-hot ベクトルである．one-hot ベクトルとは，ベクトル中で任意の 1 要素の値が 1, その他の要素が 0 となるベクトルである．Seq2SeqModel の embedded 層では，語彙数 $|V|$ の次元数を持つ one-hot ベクトルを，$n$ 次元の連続値ベクトルに変換する．ある単語を表現した one-hot ベクトル $x_t$ の連続値ベクトル $e_t$ への変換は，変換行列 $W_e \in \mathbb{R}^{n \times |V|}$ を用いた以下の式により計算される．

$$e_t = W_e x_t \tag{1}$$

$n$ 次元で表現された単語ベクトルの順列 $\boldsymbol{e} = \{e_0, e_1, ..., e_m\}$ をエンコード用のニューラルネットワークである Long Short-Term Memory (LSTM) に入力する．LSTM とは，系列データを扱うためのリカレントニューラルネットワークの一種であり，内部にメモリセルや忘却ゲートといった内部状態を設定することによって，系列長の長いデータにも対応できるようにしたものである．

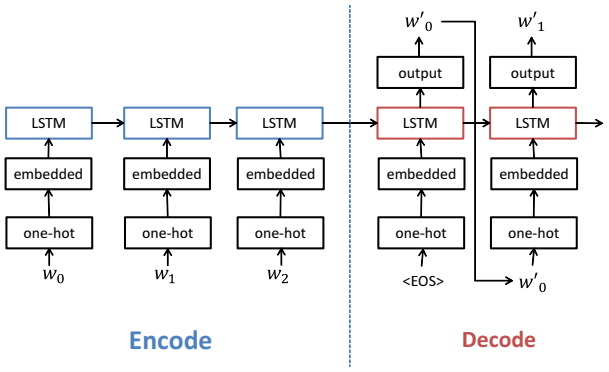



図 2: Seq2SeqModel の構成

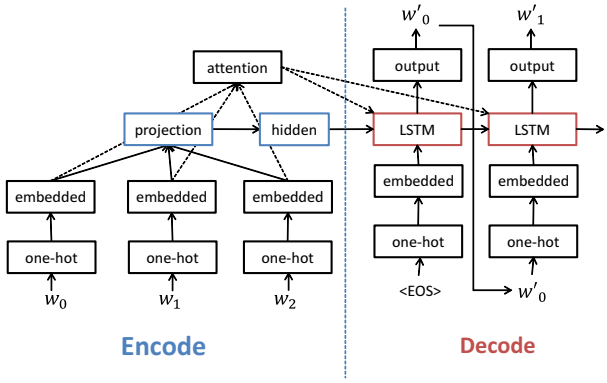



図 3: クエリ拡張用 Encoder-Decoder モデル

LSTM において，ユニットの隠れ層 $h_t$ は以下の式で定義される．

$$\begin{aligned} f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ c_t &= f_t \odot c_{t-1} + i_t \odot tanh(W_c x + U_c h_{t-1} + b_c) \\ h_t &= o_t \odot tanh(c_t) \end{aligned} \quad (2)$$

ここで，$f_t$，$i_t$，$o_t$ は忘却ゲート，入力ゲート，出力ゲートであり，$c_t$ はメモリセルを示している．パラメータは重み行列 $W_f$，$W_i$，$W_o$，$U_f$，$U_i$，$U_o \in \mathbb{R}^{n \times n}$，およびバイアス項 $b_f$，$b_i$，$b_o$，$b_c \in \mathbb{R}^{n \times 1}$ である．$\sigma$ はシグモイド関数，$tanh$ はハイパボリックタンジェントを表している．$\odot$ は行列の要素積である．

Seq2SeqMoldel のデコードにおいても，LSTM を用いるがこのとき，LSTM の隠れ層 $h_{t-1}$ と内部セル $c_{t-1}$ をエンコード用の LSTM から引き継ぐことでエンコードの情報を利用する．デコード時に $j$ 番目に生成される単語 $y_j$ の出現確率 $P(y_j|y_0, y_1, ...y_{j-1}, \boldsymbol{x})$ は，出力層において LSTM の隠れ層 $h_j$ より，以下の式で表わされる．

$$P(y_j|y_0, y_1, ...y_{j-1}, \boldsymbol{x}) = g(W_d h_j + b_d) \quad (3)$$

ここで，重み行列 $W_d \in \mathbb{R}^{|V| \times n}$ およびバイアス項 $b_d \in \mathbb{R}^{|V| \times 1}$ は学習パラメータである．$g$ は活性関数を示しており，ここにソフトマックス関数を用いることで，出力ベクトルを確率分布で表現する．$y_j$ で最も値が大きい要素のインデックスに対応する単語がデコーダーの出力単語となる．

### 3.1.2 Attention 機構

Seq2SeqMoldel の性能を向上可能な手法として attention 機構がある [2]．Seq2SeqModel のエンコーダーでは，全ての単語ベクトルが集約されて 1 つの文脈ベクトルとして表現される．入力単語長が長くなると，文脈ベクトルにおいて入力した単語それぞれの影響が弱まってしまうという問題があった．

Attention 機構では，エンコード時に LSTM の隠れ層の情報を別途保持しておき，デコード時に文脈ベクトルと同様に使用する．この方法により，入力の各単語の情報を影響を弱めることなく使用できるようになる．機械翻訳では，attention 機構は翻訳元と翻訳先の単語の対応関係（アライメント）を学習する機能として導入され，通常の Seq2SeqModel よりも高精度の翻訳が実現できることが明らかになっている．

## 3.2 回答生成モデル

本節では，回答生成モデルについて述べる．回答生成モデルは，入力クエリから，入力クエリの回答となる文書に含まれるキーワードを予測，生成するモデルである．回答生成モデルの学習には，質問と回答文書がペアになった QA 文書の集合を用いる．ベースとなるのは 3.1 で説明した Seq2SeqModel である．本論文では，Seq2SeqModel を文書検索でのクエリ拡張として使用することを目的とした変更を加える．

クエリ拡張モデルでは，ニューラルネットワークの構成と学習で用いる入出力コーパスの点から Seq2SeqModel に変更を加える．以降はそれぞれについて詳述する．

### 3.2.1 クエリ拡張用ニューラルネットワーク構成

機械翻訳や対話処理で用いられる Seq2SeqModel は，自然文で記述された入力に対して，自然文での出力を得る End-to-End の学習モデルである．このモデルでは，入出力の関係を単語の並びまで含めて学習を行うため，文法規則を別途学習する必要がないという利点がある．

しかしながら，検索においては，入力時に自然文形式で入力する場合もあれば，キーワードの集合を入力する場合もある．自然文形式の入力においても，形態素解析等の処理を行い，クエリからキーワードを抽出した上で検索を行うことになるが，このとき，キーワードの並びは自然文形式で入力されたときと，初めからキーワードで入力された場合では異なることが多い．そのため，Encoder-Decoder モデルによって，クエリの内容をエンコードする際，入力する単語の順番によって，デコード時の出力結果が変わってしまうのはあまり望ましいとはいえない．

本論文が提案するモデルでは，Encoder-Decoder モデルのエンコード部分を，入力の順番が影響する LSTM から，入力の順番が影響しない MLP（MultiLayer Perceptron）に変更する．本論文が提案する Encoder-Decoder モデルのネットワーク構成を図 3 に示す．エンコード部分は単語ベクトルを出力する embedded 層，単語ベクトルを集約して文脈ベクトルとして射影する projection 層と，デコーダーへ情報を受け渡すための hidden 層，attention 機構から構成される．embedded 層での one-hot ベクトルから連続値ベクトルへの変換は式 (1) と同様に計算する．projection 層の出力 $p$ と，hidden 層の出力 $h_E$ は次式によって計算する．

$$p = \sum_i e_i \quad (4)$$

$$h_E = tanh(W_e p + b_e) \quad (5)$$

ここで，重み行列 $W_e \in \mathbb{R}^{n \times n}$ およびバイアス項 $b_d \in \mathbb{R}^{n \times 1}$ は学習パ

ラメータである．

Attention 機構は，エンコーダー，デコーダーの hidden 層から作成する．デコード $t$ 回目の LSTM の中間層 $h_{D_t}$ との attetntion$s_t$ は次式によって計算する．

$$u_{it} = tanh(W_{ae}h_{E_i} + W_{ad}h_{D_t} + b_a) \quad (6)$$

$$\alpha_{it} = \frac{exp(u_{it})}{\sum_k exp(u_{ik})} \quad (7)$$

$$s_t = \sum_i \alpha_{it} h_{D_t} \quad (8)$$

ここで，重み行列 $W_{ae} \in \mathbb{R}^{n\times n}$，$W_{ad} \in \mathbb{R}^{n\times n}$ およびバイアス項 $b_a \in \mathbb{R}^{n\times 1}$ は学習パラメータである．

デコードは 3.1.1 で説明した Seq2SeqModel の構成を使用する．1 回目のデコードでは式 (2) において，$h_{t-1} = h_E$ とすることで，エンコード時に計算した hidden 層の出力をデコーダーの LSTM に引き渡している．

### 3.2.2 クエリ拡張用学習コーパス

本論文では，QA 文書の集合である学習用 QA を学習用のコーパスとして使用する．Seq2SeqModel では，変換元と変換先のペアとなる文コーパスをそのまま入力することに特徴がある．日本語では，各文を形態素解析した結果をそのまま入力することになる．例えば，変換前が「動画がカクカクして見られない」，変換後が「私も前に同じ症状になりましたが帯域制限が原因でした」という文を学習するとき，“動画-カクカク-見る-ない”，“私-も-前-同じ-症状-に-なる-帯域制限-が-原因-で-した” のような単語集合のペアをモデルに入力することになる．しかし，Seq2SeqModel の用途を検索でのクエリ拡張としたとき，これら全ての情報を学習させるのは適切ではないといえる．そこで本論文では，並び替えと削除という操作をコーパスに行うことによって，検索にとって有用なキーワードが生成されるようにする．以降はこれらそれぞれの操作について説明する．

**重要単語の並べ替え**　本論文では，Encoder-Decoder モデルが出力した結果を FAQ を始めとする文書検索の拡張クエリとして用いることを目的としている．拡張クエリとして利用することを想定したとき，Encoder-Decoder モデルから出力される単語は，検索にとって有用な単語であることが望ましい．先に述べた例では，“帯域制限” といった単語はできるだけ早く生成されて欲しい語ということになる．しかしながら，Seq2SeqModel の学習では，出力単語は学習コーパスと同様の順番で学習される．先に述べた例では,Seq2SeqModel が最初に出力する語は “私” となる．所望する “帯域制限” のような語は LSTM によって数回デコードした後にしか出力されない．このような問題を回避するために，単語の重み付けによって，変換後の語順を並び替える操作を行う．単語の重み付けには $tfidf$ を用いる．

$tfidf$ とは，情報検索において用いられる単語の重み付け手法であり，検索において，文書の特定性を高める語については高い重みが付与される．ある文書 $d$ 中の単語 $w$ の $tfidf$ は以下の式によって計算される．

$$tf(d,w) = freq(d,w)$$

$$idf(w) = log\frac{|D|}{df(w)} + 1$$

$$tfidf(d,w) = tf(d,w) \times idf(w) \quad (9)$$

ここで，$freq(d,w)$ は文書 $d$ 中に，単語 $w$ が出現頻度であり，$df(w)$ はコーパス $D$ 中の単語 $w$ が出現する文書頻度である．$|D|$ はコーパス $D$ の全文書数を示している．

Encoder-Decoder モデルの学習コーパスに対して，$tfidf$ 単語の重み付けを行う．変換先の文書集合に対しては，各文書で $tfidf$ スコアの降順に単語を並び替える．変換元の文書集合については，並び替えは行わないが，入力単語数の制限のために $tfidf$ のスコア上位 $l$ 個の単語のみ抽出するという操作を行う．

表 1: 実験等データセット文書数

| カテゴリ | 実験用 FAQ | テストクエリ | 学習用 QA |
|---|---|---|---|
| iphone | 150 | 843 | 442,181 |
| コスメ | 100 | 100 | 67,737 |
| 恋愛相談 | 100 | 100 | 71,591 |

**不要単語の削除**　機械翻訳等の自然文を生成するタスクでは，機能語等の全ての単語を生成できるように Encoder-Decoder モデルの学習を行う．しかしながら，検索のクエリ拡張用途での使用を前提とすると機能語等の単語は不要である．そのため，学習コーパスの事前処理として形態素解析後の各単語について，内容語以外の単語を全て削除する操作を行う．

内容語についても，検索精度や学習速度の向上を目的に，不要な単語の削除を行う．ここで不要な語とは，検索対象として想定している文書群で出現しない単語である．回答生成モデルによって新たな単語が追加キーワードとしてクエリに追加されたとしても，FAQ などの検索対象文書群に出現しない語は検索結果に寄与しない．そのため，回答生成を行う Encoder-Decoder モデルの学習時点で単語を削除する．Encoder-Decoder モデルでは，単語の語彙数は embedded 層や出力層でのパラメータ数に影響するので，単語数を制限することでパラメータ数を削減でき，モデルの軽量化，学習の高速化をすることが可能になる．

検索対象文書群を用いた不要単語の削除を行うのは，学習用 QA の回答文書（Decode 側）に対してのみである．質問（Encode 側）は，Encoder-Decoder モデルにおいてユーザが入力するクエリに対応する．入力ではユーザはあらゆる単語を使用することが想定される．そのため, 入力はなるべく多くの語彙に対応していることが望ましい．以上の理由により，Encoder-Decoder モデルの学習において，Q の学習は内容語抽出のみを行い，検索対象文書群の単語による削除は行わない．

### 3.3 回答スコア算出

回答スコアの算出には，単語重みを付与したベクトル空間モデルを用いる．入力クエリから検索で使用するキーワードを抽出し，そこに回答生成モデルから生成したキーワードを追加した拡張クエリを作成する．

QA 文書については，質問と回答文書に出現する単語について，Okapi-BM25 によって重み付けした文書ベクトルを作成する．QA 文書の文書ベクトルと拡張クエリのベクトルとの類似度を計算し回答スコアとする．

## 4. 評価実験

本節では，評価実験について説明する．まず，4.1 で実験に使用するデータセットについて説明する．次に，4.2 で実験の内容について述べ，4.3 で実験結果を示す．

### 4.1 データセット

本論文では，コールセンターの応対支援のための文書検索を目的としている．応対マニュアルは，質問とその回答のペアによって 1 つの文書が構成される FAQ 形式となっているものが多い．そのため本論文においても FAQ を対象に検索の評価実験を行う．実験で使用する検索対象 FAQ と Encoder-Decoder モデルを学習するためのコーパスを表 1 に示す．以降は，実験用 FAQ と学習用 QA について詳述する．

**実験用 FAQ**

実験用 FAQ として,「iphone」,「コスメ」,「恋愛相談」に関する FAQ を人手で作成した．これら 3 つの話題に関する FAQ で実

表 2: iphone データセットでの nDCG

| | ベースライン | Word2vec | 提案手法 |
|---|---|---|---|
| *nDCG*@5 | 0.377 | 0.408 | 0.449 |

験を行う理由として，Kim ら [6]，栗山ら [24] の先行研究がある．先行研究では，Web の質問回答サイトを対象に，質問と回答のタイプについて調査している．「iphone」のようなパソコン・情報機器に関する話題では，事実や手段を聞く情報検索型の質問が多くなされ，回答もその解決方法などが求められることが多い．一方で「恋愛相談」では意見や経験を聞く社会調査型の質問が多く，回答内容も感情的な側面からのものが多いとされている．そして，「コスメ」を始めとする健康・美容関連の話題では，上記で示した質問と回答のタイプが混在しているとしている．つまり，「iphone」，「コスメ」，「恋愛相談」の 3 つの話題の FAQ で実験することで，様々なタイプの質問，回答についても対応できるかを評価することができる．

**Encoder-Decoder モデル学習**

Encoder-Decoder モデルの学習用 QA として，Web から収集した質問回答サイトの QA 文書を用いる．質問回答サイトでは，質問と回答が自然文で記述されている．また，質問に対して回答は複数投稿されるが，その中から最も質問に役に立った回答がベストアンサーとして選ばれる．本論文では，ベストアンサーを質問の回答として QA のペア文書としている．収集には，各話題について，検索対象 FAQ で使用されている単語が含まれる QA を質問回答サイトから収集している．

### 4.2 実験設定

評価実験では，4.1 で説明したデータセットを用いて，FAQ の検索精度を評価する．実験で使用する Encoder-Decoder モデルのユニット数は，embedded 層を 200，LSTM は各ユニットで 200 としている．また，モデル学習と実験には Intel Xeon E5-2640 CPU，メモリ 64GB，NVIDIA Tesla K40 GPU を搭載した計算機を使用する．

評価用の問い合わせ（クエリ）として，実験用 FAQ とは別に作成した自然文のテストクエリを作成する．テストクエリは，実験用 FAQ の各 QA 文書に対して，回答文書を作業者に提示し，回答文書が答えとなるような質問を記述することで作成した．このとき，作成したクエリが QA 文書の質問と同じまたは単純な言い換えになっているものは除外し，クエリを再作成している．「コスメ」，「恋愛相談」の FAQ に関しては各 QA 文書に対して 1 個，「iphone」の FAQ では 1 個以上のクエリを作成した．

FAQ 検索は，クエリと QA 文書の質問（Q）との意味の類似性を測ることで精度よく検索できることが明らかになっている [22]．本論文では，クエリと質問の意味の類似性による検索精度をベースラインとして，回答を用いたことによる検索結果の精度精向上を評価する．回答部分を含めた検索での比較手法は，Word2vec[11] の類似度を用いたクエリ拡張手法 [9] を用いる．

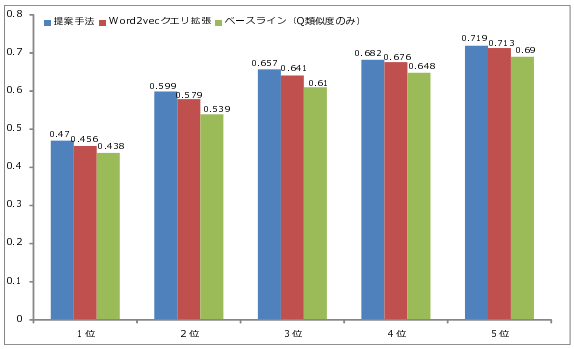
検索実験の評価尺度については，正解率を用いる．正解率とは，検索結果の上位 $k$ 位までに正解となる QA 文書が一つでも含まれていたテストクエリの割合を示すものである．また，「iphone」のデータについては，1 つのクエリに複数の QA 文書が正解となるため，*nDCG* による評価も行う．

### 4.3 実験結果

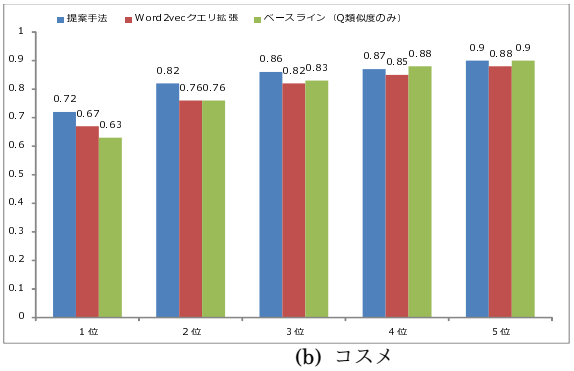
本節では，実験結果を示す．まず，4.2 で説明した FAQ 検索実験の結果を示し，次に，本論文で提案した回答生成の Encoder-Decoder モデルの出力結果と学習時間を示す．

**FAQ 検索精度実験結果**

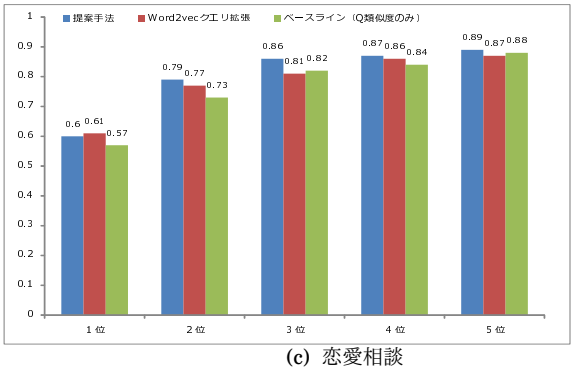
実験結果を図 4a，4b，4c に示す．各図はそれぞれ検索結果上位 1 位～5 位までの正解率を示している．iphone の FAQ は，本論文で最もサイズの大きい学習・テストセットを用いた結果，1 位～5 位いずれにおいても，提案手法が最も良い精度となっている．コスメに関しても同様の傾向を示しており，いずれの順位においても提案手法が優位となっている．恋愛相談に関する FAQ 検索は，ベースラインからの改善が他の話題よりも小さくなっている．特に 1 位の結果では，提案手法よりも比較手法である Word2vec の類似度に基づくクエリ拡張の方が僅かながら優位となっている．




(a) iphone




(b) コスメ




(c) 恋愛相談

図 4: 複数ドメイン FAQ 検索実験における上位 $k$ 件までの正解率

正解含意率の 1 位の結果に対して，“正解”，“不正解” に関する McNemar 検定を実施した．ベースラインの結果に対して，提案手法の結果の $p$ 値 は，「iphone」では $p = 0.0384(< 0.05)$，「コスメ」では $p = 0.0538(> 0.05)$，「恋愛相談」では $p = 0.355(> 0.05)$ となり，「iphone」データにおいて提案手法による精度改善が有意に示された．

「iphone」に関して，検索結果の *nDCG*@5 を計算した結果を表 2 に示す．*nDCG*@5 の評価においても，提案手法によるクエリ拡張を適用した検索結果がベースライン，Word2vec によるクエ

リ拡張手法よりも優位な結果となっている．

回答生成モデル出力結果

本実験によって，拡張されたクエリの例を表 3 に示す．表中の “Word2vec” は Word2vec によって生成された拡張クエリである．入力クエリの単語の合成ベクトルに類似する上位 5 個の単語を抽出している．“提案手法” は本論文で提案したクエリ拡張手法によって生成された拡張クエリである．提案手法の拡張クエリは Encoder-Decoder モデルによって生成されるため，出力されるキーワード数が可変であるという特徴がある．また，クエリを入力したときに Encoder-Decoder モデルが出力した追加キーワード数の比較を表 4 に示す.「iphone」では，拡張の際に平均 4.49 個のキーワードが追加されているが,「恋愛相談」での追加キーワード数は 2.56 個となっており，話題によって追加キーワード数に差が出ていることが分かる．

Encoder-Decoder モデルの学習について，最も学習データ多かった「iphone」データでの計算時間を表 5 に示す．ここでは，3.2.2 で説明した，不要単語の除去作業を行ったときと，行っていないときの学習時間を比較している．出力語彙数は，Encoder-Decoder モデルが出力する単語の語彙数であり，計算時間は学習データ全てを 1 回学習した時にかかった時間の平均値である．FAQ に出現しない不要単語を除去することで，モデルが学習する語彙数が約 80% 削減され，学習時間についても約 65% 高速になっていることがわかる．

## 5. 考察

本節では，実験結果に関する考察を行う．まず，5.1 で，FAQ 検索実験の結果について考察し，5.2 では，学習した Encoder-Decoder モデルについての考察を行う．

### 5.1 FAQ 検索での精度向上に関する考察

正解率による実験結果である図 4a，4b，4c に示すとおり，入力クエリの内容から，回答に出てくる内容を予測することによる検索は有効に機能しているといえる.「iphone」データを用いた nDCG の評価においても，提案手法が最も高い値となっているため，提案手法によって，クエリに対して正解である可能性の高い QA 文書が上位に出現しやすくなっているといえる．

検索結果 1 位について,「iphone」のデータセットに関しては McNemar 検定において有意差が認められたが,「コスメ」,「恋愛相談」では有意差が認められなかった.「コスメ」に関しては $p$ 値が 0.0538 であり有意傾向ありであるといえる.「iphone」と同様に，学習データやテストクエリを拡充することで有意差が認められる可能性が高いのではないかと考えられる．

「恋愛相談」に関しては,「コスメ」よりも学習データが多いにも関わらず，提案手法の優位性が示されなかった．これは,「恋愛相談」では質問に対する妥当な回答が学習できなかったためであると考えられる．表 4 に示した Encoder-Decoder モデルの追加キーワード数では,「恋愛相談」は他の話題と比べて追加キーワード数が少ない．また追加されたキーワードも “告白”, “失恋” といった「恋愛相談」においては一般的な語が多く出力されている．このように,「恋愛相談」においては，回答を検索するための追加キーワードが十分に生成出来ていないことが分かる．先行研究 [6] でも述べられいる通り,「恋愛相談」は社会調査的な内容の質問が多い．このような質問は，特定の正解が存在しているとは限らず，同じ質問であっても回答の内容が異なることが多い．Encoder-Decoder モデルの学習では，質問と回答が 1 対 1 で対応していることが前提となるため，回答内容にバリエーションを持つ「恋愛相談」のような話題に関しては，検索精度を向上するための十分なキーワード生成が行えなかったのではないかと考えられる．

今回の実験では，作成した FAQ データとは無関係に作成された，質問回答サイトの QA を Encoder-Decoder モデルの学習に用いていたが，FAQ 検索結果を向上させる有用な知識を学習することができた．コールセンタ等の利用では，応対ログといった，より検索対象 FAQ と関連性の高い学習データを使用することでより最適な回答内容を生成出来るのではないかと考えている．

### 5.2 回答生成 Encoder-Decoder モデルに関する考察

表 3 に示した提案手法による拡張クエリのキーワードは，単純な単語同士の意味の類似性では生成できないものも含まれている．例えば，“音量の上げ方は” というクエリを入力したときに，Encoder-Decoder モデルが出力する結果として “横” というキーワードがあるが，これは，“音量”, “上げる” といった単語とは一見無関係である．しかしながら，iphone では音量は端末 “横” のボタンを押すことによって調節することが一般的である．これは，提案手法がこのような個別の具体的な事象に対しても，回答内容を正確に学習できることを示唆しているといえる．

提案手法が生成する拡張クエリの特徴的な点として，入力の単語ベクトルの意味の学習がある．表 3 のコスメに関するクエリ “夏のスキンケアのポイントは何？” と，“冬のスキンケアのポイントは何？” は，単語 “夏” と “冬” 以外は同じ文である．しかしながら，提案手法の拡張クエリは夏は日焼け関連，冬は保湿関連と全く違った内容を出力している．一般的にトピックモデルや Word2vec に代表される共起ベースで生成された単語ベクトルは，近い文脈で使用されやすい語の分離が難しいとされる．実際に，Word2vec ベースのクエリ拡張では，“ 用品” や “クレンジングクリーム” など，同じ単語が出力されている．これは，ほとんどの単語が同じと言うことに加えて，“夏” と “冬” が Word2vec 上では近いベクトルであることを意味している．一方で，提案手法の Encoder-Decoder モデルでは，出力が大きく異ることから，モデル内部で保持されている “夏” ベクトルと “冬” ベクトルは大きく異るベクトルであることを表しているといえる．

本モデルがこれらのベクトルを区別できている理由は，質問と回答という，全く性質の異なる文書の共起に基づいてベクトルの学習を行っているからであると考えられる．同一文や文書内の共起では，中で書かている単語が 1 つ変化しても，内容全体は見ればあまり変化しない場合が多い．しかし，質問と回答では，質問中の単語が少しでも違えば，回答内容は大きく変化する．そのため，Word2vec のような手法では，近い意味となってしまう単語も違う意味として学習出来たのではないかと考えられる．

Encoder-Decoder モデルの学習では，モデルが出力する単語を検索対象の FAQ に出現する単語に限定することで，学習の効率化を図った．表 5 に示した通り，不要な単語を除去することで学習速度が約 65% 向上することが確認できた．不要単語を除去した場合においても図 4a や表 2 に示した通り，検索精度が有意に向上することが確認できるため，不要単語を除去は，クエリ拡張のための Encoder-Decoder モデルの学習において非常に効果的であるといえる．

## 6. 関連研究

FAQ 検索は，Q と A という 2 つの文書ペアを検索する情報検索タスクである．特に，FAQ 検索は問い合わせ対応などの場面で多く活用されることから，単純なキーワード組のクエリではなく，Short Messaging Service (SMS) のような自然文を入力を想定したタスクとなっている [8, 15]. 入力クエリが自然文の場合，検索精度を低下させる単語や表現も含まれている．Agarwal[1] らは，接尾辞の一致や N-gram モデルなど自然言語処理を用いて，検索時にノイズとなる表現を解消した FAQ 検索手法を提案している．Yu[21] らは，“長さ” や “頻度” などの質問のタイプによって，回答部分に出現しやすい語が有ることを特定し，それを用いることで回答部分をより高精度に検索できることを報告している．

情報検索において入力されたクエリと検索対象文書のスコアを計算するために，ニューラルネットワークを使用した研究が多く報告されている．Mitra[13] らは，クエリと文書について，直接的な単語一致と分散表現の類似度の一致をニューラルネットワーク

表 3: 評価実験データにおける拡張クエリの例

| カテゴリ | | |
|---|---|---|
| iphone | 入力クエリ: | 音量の上げ方は？ |
| | Word2vec: | 下げる, ボリューム, 調節, 低音, 上げ下げ |
| | 提案手法: | サイレント, 音量, スピーカー, サウンド, 横, 調節 |
| | 入力クエリ: | 画面が割れたら？ |
| | Word2vec: | バキバキ, ひび割れ, 浮き, 割れ, ヒビ |
| | 提案手法: | 交換, 修理, アップル |
| | 入力クエリ: | 海外旅行で通信料金高額にならないようにするには? |
| | Word2vec: | バキバキ, ひび割れ, 浮き, 割れ, ヒビ |
| | 提案手法: | 交換, 修理, アップル |
| コスメ | 入力クエリ: | 化粧品を使い切らないといけない期限ってありますか |
| | Word2vec: | 継続, 年数, 捨てる, 1 か月, 開封 |
| | 提案手法: | 1 年, 製造, 開封, 保管, 期限, 大丈夫 |
| | 入力クエリ: | 夏のスキンケアのポイントは何？ |
| | Word2vec: | 用品, 時短, クレンジングクリーム, お出かけ, 外出 |
| | 提案手法: | 日焼け止め, 紫外線, しみ |
| | 入力クエリ: | 冬のスキンケアのポイントは何？ |
| | Word2vec: | 用品, スクラブ, クレンジングクリーム, ファンケル, シート |
| | 提案手法: | ケア, パック, 保湿, クリーム |
| 恋愛相談 | 入力クエリ: | 社会人のデート代って高いのかな |
| | Word2vec: | 高額, 費用, 超え, 教育費, 貯める |
| | 提案手法: | 表す, スマート, 割り勘 |
| | 入力クエリ: | 振られちゃったんですけど、元気になる方法は？ |
| | Word2vec: | スッパリ, 未練, 忘れる, 勇気, サヤ |
| | 提案手法: | 失恋, 忘れる |
| | 入力クエリ: | 片思いに悩む女性ですが、どう伝えたらいいでしょう？ |
| | Word2vec: | 片想い, 付き合う, 既婚, 思う, 振られる, アプローチ |
| | 提案手法: | 告白 |

表 4: Encoder-Decoder モデルでの追加キーワード数

| | iphone | コスメ | 恋愛相談 |
|---|---|---|---|
| 追加キーワード数 | 4.49 | 3.09 | 2.56 |

表 5: iphone カテゴリデータにおける学習時間比較

| | 単語除去あり | 単語除去なし |
|---|---|---|
| 出力語彙数 | 2,507 | 12,926 |
| 計算時間 (1epoch)[sec] | 1620.2 | 4540.9 |

で統合したモデルを提案している．Jaech ら [4] は，クエリと文書中の単語の分散表現の類似度を Convolutional Neural Network (CNN）で集約することによる検索スコアの計算手法を提案している．Wang ら [18] は，ニューラルネットワークの生成モデルで Generative Adversarial Network（GAN）をクエリから文書を生成するモデルと定義することで検索モデルに適用する手法を提案している．

FAQ のような文書に対して，答えの部分に着目して関連する文書を検索する技術は，質問応答技術の一領域であるといえる．質問応答では，QA を含む大規模テキストを知識源として，ユーザが入力された質問に対して，回答自動で生成して出力する．近年では，質問応答に DeepLearning の技術を用いるものが多く提案されている．Weston ら [19] はニューラルネットワークで構成されたメモリセルにテキストの知識を蓄積する Memory Networks を提案している．Feng ら [12] は，QA 文書を知識源として使用し，CNN によって，Q と A の内容をそれぞれ意味ベクトル化することによる質問応答技術を提案している．

## 7. おわりに

本論文では，質問と回答のペア（QA）の関係性を学習することで，入力クエリを満足させる回答に含まれる単語を予測，生成する回答生成モデルおよび，回答生成モデルを使用した FAQ 検索手法を提案した．Encoder-Decoder モデルにより，End-to-End で入力クエリから，回答文書検索用で使用する追加キーワードを生成する．実験の結果，提案手法は事実や直接的な回答を必要とするような，情報検索型の質問に対応する FAQ 検索において特に有効であることを示した．

今後は，Encoder-Decoder モデルの改良による学習の精度向上および，ニューラルネットワークを用いた End-to-End 型の文書検索についても研究を進めていく予定である．

## [文献]

[1] Amit Agarwal, Bhumika Gupta, Gaurav Bhatt, and Ankush Mittal. Construction of a Semi-Automated Model for FAQ Retrieval via Short Message Service. *Proc of the 7th Forum for Information Retrieval Evaluation (FIRE2015)*, pp. 35–38, 2015.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *Proc of the 5th International Conference on Learning Representations(ICLR2015)*, 2015.

[3] Adam Berger, Rich Caruana, David Cohn, Dayne Freitag, and Vibhu Mittal. Bridging the lexical chasm: Statistical approaches to answer-finding. *Proc of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2000)*, pp. 192–199, 2000.

[4] Aaron Jaech, Hetunandan Kamisetty, Eric K. Ringger, and Charlie Clarke. Match-tensor: a deep relevance

model for search. *CoRR*, Vol. abs/1701.07795, , 2017.

[5] Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. Finding similar questions in large question and answer archives. *Proc of the 14th ACM International Conference on Information and Knowledge Management (CIKM2005)*, pp. 84–90, 2005.

[6] Soojung Kim and Sanghee Oh. Users' relevance criteria for evaluating answers in a social Q&A site. *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 4, pp. 716–727, 2009.

[7] Kanako Komiya, Yuji Abe, Hajime Morita, and Yoshiyuki Kotani. Question answering system using Q & A site corpus Query expansion and answer candidate evaluation. *Springerplus*, Vol. 2, No. 396, pp. 1–11, 2013.

[8] Govind Kothari, Sumit Negi, Tanveer A. Faruquie, Venkatesan T. Chakaravarthy, and L. Venkata Subramaniam. SMS Based Interface for FAQ Retrieval. pp. 852–860, 2009.

[9] Saar Kuzi, Anna Shtok, and Oren Kurland. Query expansion using word embeddings. *Proc of the 25th ACM International on Conference on Information and Knowledge Management(CIKM2016)*, pp. 1929–1932, 2016.

[10] Tie-Yan Liu. Learning to Rank for Information Retrieval. *Found. Trends Inf. Retr.*, Vol. 3, No. 3, pp. 225–331, 2009.

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *CoRR*, Vol. abs/1301.3781, , 2013.

[12] Feng Minwei, Xiang Bing, Glass Michael, Wang Lidan, and Zhou Bowen. Applying Deep Learning to Answer Selection: A Study and An Open Task. *Proc of The 2015 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU2015)*, 2015.

[13] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. Learning to match using local and distributed representations of text for web search. *Proc of the 26th International Conference on World Wide Web (WWW2017)*, pp. 1291–1299, 2017.

[14] Stefan Riezler, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu Mittal, and Yi Liu. Statistical machine translation for query expansion in answer retrieval. *Proc of the 45th Annual Meeting of the Association for Computational Linguistics (ACL2007)*, pp. 464–471, 2007.

[15] Anwar D. Shaikh, Mukul Jain, Mukul Rawat, Rajiv Ratn Shah, and Manoj Kumar. Improving accuracy of SMS based FAQ retrieval system. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 7536 LNCS, pp. 142–156, 2013.

[16] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. *Proc of the 27th International Conference on Neural Information Processing Systems(NIPS2014)*, pp. 3104–3112, 2014.

[17] Oriol Vinyals and Quoc V. Le. A neural conversational model. *Proc of the ICML Deep Learning Workshop 2015*, 2015.

[18] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proc of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2017)*, pp. 515–524, 2017.

[19] Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *Proc of the 5th International Conference on Learning Representations(ICLR2015)*, 2015.

[20] Xiaobing Xue, Jiwoon Jeon, and W. Bruce Croft. Retrieval models for question and answer archives. *Proc of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2008)*, pp. 475–482, 2008.

[21] Zheng-Tao Yu, Zhi-Yun Zheng, and Shi-Ping Tang ; Jian-Yi Guo. Query expansion for answer document retrieval in chinese question answering system. *Proc of 2005 International Conference on Machine Learning and Cybernetics*, pp. 72–77, 2005.

[22] 大塚淳史, 別所克人, 平野徹, 東中竜一郎, 浅野久子, 松尾義博. 文構造を考慮した発話理解に基づく自然文検索. 第 30 回人工知能学会全国大会論文集（JSAI2016), 2016.

[23] 菊地勝由（編）. BUSINESS COMMUNICATION 2016 年 2 月号. ビジネスコミュニケーション社, 2016.

[24] 栗山和子, 神門典子. Q&A サイトにおける質問と回答の分析. 研究報告データベースシステム（DBS）, Vol. 2009, No. 19, pp. 1–8, 2009.

**大塚 淳史 Atsushi OTSUKA**
日本電信電話株式会社 NTT メディアインテリジェンス研究所 研究員. 2013 年筑波大学大学院図書館情報メディア研究科博士前期課程修了．情報検索，対話システムの研究開発に従事．情報処理学会，人工知能学会，日本データベース学会各会員.

**別所 克人 Katsuji BESSHO**
日本電信電話株式会社 NTT メディアインテリジェンス研究所 主任研究員．1994 年大阪大学大学院理学研究科数学専攻修士課程修了．自然言語処理の研究に従事．2009 年，新潟大学博士（工学）．電子情報通信学会，情報処理学会，言語処理学会各会員.

**西田 京介 Kyosuke NISHIDA**
日本電信電話株式会社 NTT メディアインテリジェンス研究所 主任研究員．2008 年北海道大学大学院情報科学研究科博士後期課程修了．博士（情報科学）．人工知能，データマイニングの研究開発に従事．2017 年日本データベース学会上林奨励賞．情報処理学会シニア会員，電子情報通信学会会員，日本データベース学会正会員.

**浅野 久子 Hisako ASANO**
日本電信電話株式会社 NTT メディアインテリジェンス研究所 主幹研究員．1991 年横浜国立大学工学部卒業．自然言語処理に関する研究開発に従事．情報処理学会，言語処理学会各会員.

**松尾 義博 Yoshihiro MATSUO**
日本電信電話株式会社 NTT メディアインテリジェンス研究所 主幹研究員．現在，NTT アドバンステクノロジ株式会社．1990 年大阪大学大学院理学研究科博士前期課程修了．自然言語処理の研究開発に従事．情報処理学会，言語処理学会各会員.