

ニュース情報検索システム「NetTv」における 質問種別の推定*

☆丸本理貴人, 田中克幸, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

情報網・Web 2.0の発展や放送のデジタル化により、情報整理が困難なメディア、映像、画像、音響などの普及が、情報の氾濫を招いている。情報量の爆発とプラットフォームの多様化により、ユーザがほしい情報を入手できない状況にあり、効率的にユーザが欲しい情報だけを入手できる環境が、必要とされている。そこで、我々人同士のコミュニケーション手段である音声を紹介したインターフェースが注目を集めており、音声インターフェースによる乗換案内システム[1]や情報案内システム[2, 3]が研究されてきた。本研究で構築しているNetTv[5]は、ネットニュース動画において、動画インデキシングと音声インターフェースを用いた検索システムである。すなわち、ユーザが快適に動画を視聴でき、視聴中に生じた疑問をその場で解決できるシステムを構築し、ユーザの検索負担軽減を目的としている。

しかし、NetTvでは、“○○って何”や、“○○って誰”といった質問に対して、GUIでWikipediaのページを表示するという形で返答が可能であるが、それ以外の質問に対して返答することができないという問題があった。ユーザの疑問をより柔軟に解決するためには、幅広い質問に対して返答することが求められる。そこで本研究では、NetTvに質問応答モジュールを構築するための1ステップとして、質問の種別を推定する。自然言語の分類問題はいくつか研究[7, 8, 9]が行われているが、本研究では、Bag of WordsとRNNEncoder-Decoderモデル[6]を用いて文書データのベクトル変換を行い、各ベクトル変換手法に対して、Support Vector Machine(SVM)とDeep Neural Network(DNN)の学習手法を用いて質問種別を推定し、その結果を比較する。

2 NetTvの概要

NetTvでは、日々変化するダイナミックな環境において、音声認識により情報検索することを目的としている。そのため、インターネットニュースの映像メディアに対して、以下のような機能を円滑に行えることに重点を置き、システムを構築している。

- ネットニュースにおける動画インデキシング

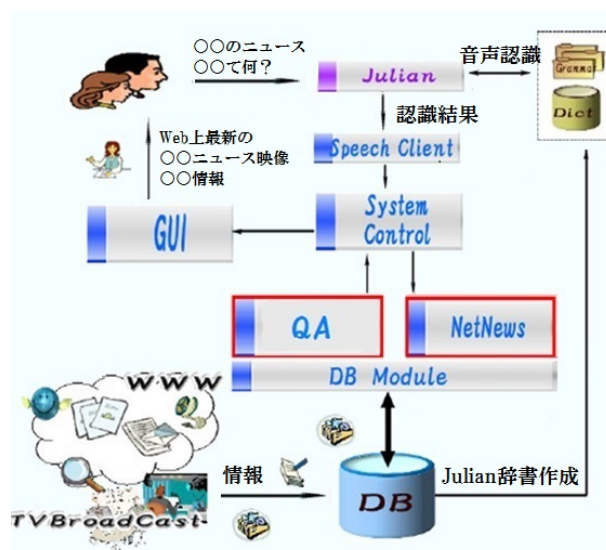


Fig. 1 NetTvの構成

- 音声インターフェースによる検索と質問応答機能

NetTvでは、Fig. 1に示すようにNetNewsというモジュールが中心的役割を果たしている。本研究では、ユーザの幅広い質問に対して、回答できるようなQA(質問応答)モジュールを構築した。

2.1 NetNews

Fig. 2にNetNewsの構成を示す。NetNewsはネット上に流れるニュースに対して、周りのテキスト情報をメタ情報として付与し、ニュースをインデックス化することにより、音声検索を可能にする。ニュースのダイナミックなコンテンツ変化に対応できるように、リンクと詳細記事をニュースサイトから自動的に集めてくる。集めた各詳細記事のページにhtmlパーザをかけ、記事部分と動画部分を切り出し、詳細記事とヘッダーを文単位に分解する。その後、茶筌[4]を用いて、形態素解析を行う。この結果、URL、動画、ヘッダー、詳細記事の単語に関するインデックステーブルが作成され、キーワードによる検索が可能となる。

2.2 音声インターフェース

NetTvでは、NetNewsモジュールで作成したデータベースから音声認識用の辞書を自動生成している

*Estimation of question type in news information dialogue system 「NetTv」. by MARUMOTO, Rikito, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe University)

Table 1 質問タイプ

ファクトイド型質問		短いフレーズで回答できる質問	富士山の高さは？
ノンファクトイド型質問	定義型質問	人物や物事の定義を尋ねる質問	ファクトイド型質問とは何ですか？
	why 型質問	理由・原因を尋ねる質問	地震はなぜ起きるの？
	how 型質問	形容や方法, 手続きを尋ねる質問	どうやって頭がよくなるの？

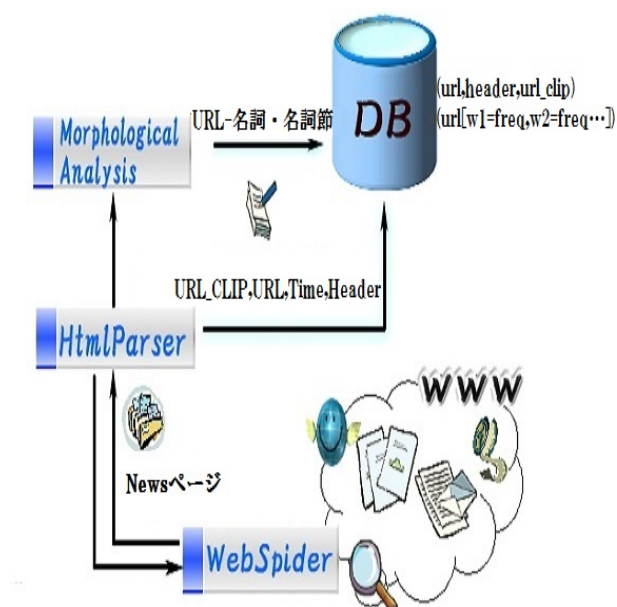


Fig. 2 NetNews モジュールの構成

ので、ネットニュース情報が日々更新されるたびに、音声認識用の辞書をダイナミックに変更することができ、ドメインの最新情報に関する音声認識を実現している。音声認識には Julian を使用しており、文法を設定することで、“〇〇のニュース”といった発話によって、そのキーワードに関連するニュースをユーザーに提示している。

3 質問応答

NetTv システムでは、ユーザーの幅広い質問に対応するため、質問応答モジュールを構築する必要がある。質問応答技術は、自然言語による質問を入力とし、そこから回答部分のドキュメントを抽出することで、ユーザーの質問に応答する技術である。Table 1 で示すように、質問応答で扱われる質問は、短いフレーズで回答でき、単純な事実や事象に関して尋ねるファクトイド型質問と、文章で回答でき、理由や方法、定義などを尋ねるノンファクトイド型質問に分けられる。また、ノンファクトイド型質問は大きく分けると、人物や物事の定義を尋ねる定義型質問と、理由

や原因を尋ねる why 型質問、形容や方法、手続きを尋ねる how 型質問の 3 タイプに分けられる。

質問応答モジュールでは、ユーザーの質問に対して、質問種別を推定した後、各質問種別に適した処理を行い、ユーザーに回答を提示する。本研究では、質問応答モジュールを構築するために、ファクトイド型質問、定義型質問、why 型質問、how 型質問の 4 つの質問種別を推定する。

4 質問分類

4.1 ベクトル変換

文書データは、学習のためにベクトルに変換する必要がある。文書データをベクトル変換するための手法として、Bag-of-Words と RNN Encoder-Decoder モデルを用いる。

4.1.1 Bag-of-Words

Bag-of-Words は、文書データを表現するために最も広く用いられている手法である。BoW では文書中に含まれている単語の出現頻度でベクトル化を行う。しかし、単語の出現頻度でのみベクトル化を行うため、語順の情報が失われる。本研究では、 $TF-IDF$ に基づいて、学習に不必要な単語を除いている。 $TF-IDF$ は、文書中の単語に関する重要度を示す指標であり、 TF と IDF の 2 つの指標の積で計算される。文書 d 内で任意の単語 t の出現回数を n_{td} とし、 d 内すべての単語の出現回数の和を $\sum_s n_{sd}$ としたとき、 $TF = n_{td} / \sum_s n_{sd}$ で計算される。したがって、 TF の値が大きい単語すなわち、出現頻度が高い単語は重要である。また、任意の単語 t を含む文書数を df_t とし、全文書数を D としたとき、 $IDF = \log(D/df_t)$ で計算され、 IDF が高い単語は重要である。これは、多数の文書に出現する単語は重要でないことを意味する。本研究では、 $TF-IDF$ が低い単語、すなわち出現頻度が 3 以下で、全体の 7 割以上の文書で出現するような単語は TF が低く、 IDF が低いいため削除する。そうして得られたベクトルデータに対して、Latent Semantic Indexing(LSI) を用いて次元削減を行い、200 次元の特徴量を分類に使用している。

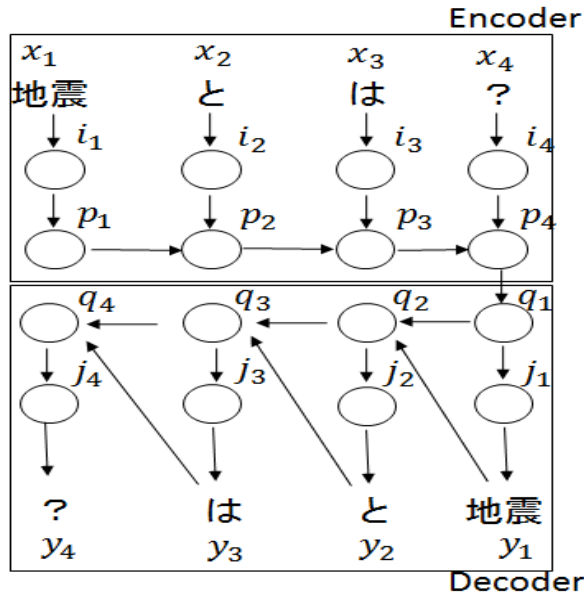


Fig. 3 RNN Encoder-Decoder
「地震とは？」という入力に対する例

4.1.2 RNN Encoder-Decoder

RNN Encoder-Decoder は、機械翻訳の分野で頻繁に用いられており、Encoder 側と Decoder 側の 2 つの RNN を中間層で結合することで構築する。機械翻訳では、入力を翻訳したい文、出力を翻訳結果として学習させることでモデルを構築するが、本研究では出力データに入力データと同じデータを設定して学習を行い、中間層のベクトル値を特徴量として質問種別の分類に使用している。

RNN Encoder-Decoder モデルを Fig. 3 に示す。 i_n と j_m は埋め込み層と呼ばれ、単語情報を多次元実数ベクトルで表す。埋め込み層である i_n と j_m は式 (1), (2) で計算する。

$$i_n = \tanh(W_{xi}x_n) \quad (1)$$

$$j_m = \tanh(W_{qj}q_m) \quad (2)$$

活性化関数としては、双曲線正接関数を用いている。ここで、 W_{xi} , W_{qj} は、ネットワークの重みである。

隠れ層である p_n と q_m は式 (3), (4), (5) で計算される。ここで、 g は活性化関数であり、本研究では Long Short-Term Memory(LSTM) を用いている。 W_{ip} , W_{pp} , W_{pq} , W_{yq} , W_{qq} はネットワークの重みである。

$$p_n = g(W_{ip}i_n + W_{pp}p_{n-1}) \quad (3)$$

$$q_1 = g(W_{pq}p_{|W|}) \quad (4)$$

$$q_m = g(W_{yq}y_{m-1} + W_{qq}q_{m-1}) \quad (5)$$

最終的に y_m は式 (6) のようにソフトマックス関数で計算される。 W_{jy} は、ネットワークの重みである。本研究では、 $q_{|W|}$ (Fig. 3 の q_4) を特徴量ベクトルとして抽出し、分類に使用する。

$$y_m = \text{softmax}(W_{jy}j_m) \quad (6)$$

4.2 分類手法

文書をベクトル変換して得られた特徴量ベクトルを用いて、質問種別のクラス分類を行う。本研究では、分類手法として SVM(Support Vector Machine) と DNN(Deep Neural Network) を用いる。

4.2.1 Support Vector Machine

SVM は、入力データをカーネル関数を用いて高次元へ写像し、写像空間においてマージンが最大となるような識別境界を求める方法である。識別境界は式 (7) で求められる。また、カーネル関数は式 (8) の RBF カーネルを使用する。

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i \quad (7)$$

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (8)$$

4.2.2 Deep Neural Network

DNN は、従来のニューラルネットワークを多層構造にしたものである。入力層、複数の中間層、出力層で構成されており、中間層の出力を式 (9)、出力層の出力を式 (10) とする。

$$y = \text{sigmoid}(Wx + b) \quad (9)$$

$$z = \text{softmax}(\tilde{W}y' + \tilde{b}) \quad (10)$$

ただし、 W, b は中間層における重みとバイアスであり、 $\tilde{W}, \tilde{b}$ は出力層における重みとバイアスである。また、 y' は最後の中間層における出力である。

5 評価実験

5.1 実験条件

データセットとして、Yahoo! 知恵袋データ (第 2 版) を用いた。ファクトイド型、定義型、why 型、how 型の 4 クラスの質問種別を手動でラベル付けし、学習データ数は 1000 データ (各クラス 250 データ)、評価データ数は 200 データ (各クラス 50 データ) を用いた。入力データの次元数は、Bag of Words で文書をベクトル変換し、LSI で次元圧縮した 200 次元と、RNN Encoder-Decoder の中間層のベクトルを抽出した 200 次元を用いた。

Table 2 Precision · Recall · F-measure の平均

	Precision(%)	Recall(%)	F-measure(%)
BoW + SVM	86.4	85.0	85.7
BoW + DNN	87.4	87.5	87.4
RNN ED + SVM	93.7	94.0	93.8
RNN ED + DNN	92.4	92.0	92.2

各分類方法のパラメータとして、SVMのパラメータは、グリッドサーチで得た $C = 8.0, \gamma = 0.125$ (BoW), $C = 32.0, \gamma = 0.125$ (RNN Encoder-Decoder) とした。また、DNNにおいては、中間層数を2層、各中間層のノード数を500ノードとし、学習率0.001、バッチサイズ5として確率的勾配法を用いてモデルを学習した。評価指標として Precision · Recall · F-measure を用いた。

5.2 実験結果

質問種別のクラス分類結果を Table2 に示す。Precision · Recall · F-measure は4クラスの平均である。Table2 より、最もよい結果が得られたのは、RNN ED + SVM の手法であることがわかる。次いで、RNN ED + DNN である。この結果から、BoW と RNN ED の文書のベクトル変換法において、BoW より RNN ED が有効であるといえる。これは、BoW では単語の出現頻度で文書のベクトル変換を行うために、文書の語順情報が失われるのに対して、RNN ED は文書の系列情報を加味したベクトル変換ができていたためであると考えられる。しかし、SVM と DNN の分類手法を比較すると、BoW と RNN ED のどちらのベクトル変換法においても、あまり差異がない。原因として、データ数が少ないことが考えられる。データ数を増やすことで、SVM より DNN が良い結果を得ることができると考えられる。

6 おわりに

本研究では、ニュース情報検索システム「NetTv」において、ユーザーがニュース視聴の際、幅広い質問に対して回答できるように、ファクトイド型質問、定義型質問、why 型質問、how 型質問の4つの質問種別のクラス分類を行う質問応答モジュールを構築した。BoW+SVM, BoW+DNN, RNN ED+SVM, RNN ED+DNN の4つの手法を比較した結果、RNN ED+SVM が最も良い結果を得ることができ、質問種別の分類においては RNN Encoder-Decoder ベクトル変換が有効であることを示した。

今後の課題としては、NetTv に質問応答モジュールを統合し、NetTv を使用してもらい、ユーザー満足

度の調査を行うことを考えている。また、NetTv において、ユーザーのニーズを理解し、必要な機能を調査し、新たな機能を構築することで NetTv のシステム向上につながると考えている。

参考文献

- [1] Lori Lamel *et al.*, “User evaluation of the mask kiosk, ”, Speech Communication, 131-139, 2002.
- [2] Koichiro Yoshino, Tatsuya Kawahara, “News Navigation System based on Proactive Dialogue Strategy, ” Int’l Workshop Spoken Dialogue Systems, 2015.
- [3] 木田 学, 他 “音声情報案内システムにおける質問応答データベース最適化手法の検討,” 情報処理学会, 81-86, 2006.
- [4] 松本 裕治 他, “形態素解析システム茶筌 version 2.2.9 使用説明書,” 2002.
- [5] 田中 克幸, 滝口 哲也, 有木 康雄, “NetTv : NetNews とテレビ放送のクロスプラットフォームにおける音声検索,” 日本音響学会誌, 179-180, 2007.
- [6] Kyunghyun Cho *et al.*, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, ” Empirical Methods in Natural Language Processing, 2014.
- [7] 加藤 玲大, 馬 青, 村田 真樹 “深層学習を用いた QA サイト質問文のカテゴリ分類,” 情報処理学会, 1-6, 2016.
- [8] Yoom Kim, “Convolutional Neural Networks for Sentence Classification, ” Empirical Methods in Natural Language Processing (EMNLP), 1746-1751, 2014.
- [9] 渡邊 直人 他, “QA コミュニティにおける質問者の期待に基づく質問分類に関する一検討,” 第3会データ工学と情報マネジメントに関するフォーラム (DEIM2011), B5-1, 2011.