

## 対話履歴に基づく話題推定方法

中田 貴之 安藤 真一 奥村 明俊

NEC マルチメディア研究所

{t-nakata, ando, okumura}@ccm.cl.nec.co.jp

### 1. はじめに

近年、社会の情報化・ボーダーレス化が進むなかで、異言語コミュニケーションの支援手段として、自動通訳の実現に向けた研究開発が進められている。対象を旅行会話などに限定することでほぼ実用的な性能を発揮することが可能となっており、今後さらに性能向上をはかることでその利用場面も拡大していくと思われる[1]。

自動通訳システムの利用場面の拡大とともに、その会話内容や利用場面に応じた適切な通訳処理が今までにまして重要になってくる。例えば旅行会話において "It's rare." という発話があった場合、これがレストランでの肉の焼き方に関する発話ならば「レアです。」、その他での発話ならば「まれです。」というように話題に応じた適切な通訳結果が得られるのが望ましい。このためには、利用者の対話履歴の内容から現在の話題を推定し、その推定結果によって多義語の曖昧性を解消するといった手法が考えられる。

本稿では、自動通訳システムにおいて多義語の曖昧性解消や音声認識精度の向上を図るため、話者間の対話履歴から現在の話題を推定する方法について、性質や問題点を明らかにするとともにその解決方法について提案する。

### 2. 話題推定

本稿で扱う話題は、ある対話が行なわれている場所、そこでの行為・目的、出現する物、およびそれらに関連する周辺知識によって特徴つけられるものを想定している。旅行会話の日英双方向自動通訳システムを利用する場面の話題としては、「ホテル」、「レストラン」、「観光」、「トラブル」といったものがあげられる。

従来、上記の話題にあたるものを推定する手法として複合語をトリガー情報として用いたもの[2]や辞書・シソーラスなどを参照したもの[3]がある。前者の手法は、翻訳時にある分野辞書のトリガー

となる複合語が使用されると、その分野の話題と判定し、対応する分野辞書の優先順位を上げるという手法である。後者の手法は、物の名前をそれが出てくる場面から引くことができる視覚辞書を利用、ある場面に現われる物の名前を場面知識とし、さらにシソーラスの上位-下位関係や概念を用いて頑健性を高める手法である。

これらの手法は文書には有効であるが、自動通訳システムによる対話のように一つの発話ごとに適用する場合には、以下のような制約により話題の推定が困難と考えられる。

#### (1) 文単位の処理による制約

- ・ 対話文は一文一文が短く、そこに含まれる特徴語は数個程度しかない
- ・ ある一文に出現する単語の頻度はほとんど1であるため、統計的頻度情報に基づいた手法をそのまま適用することが困難である

#### (2) 対話構造による制約

- ・ 対話文では一つの事柄について述べるとき、2文以上に分割して述べるものがしばしばある
- ・ 以前の文で述べたことを省略や照応で表すことがある
- ・ 話題の遷移が頻繁に起こる

本稿では、話題推定アルゴリズムとして、k最近傍法[4]にクラスタリング手法と対話履歴利用を組み込んだ手法を提案する。このk最近傍法を用いた理由は、入力文に対する逐次的な話題推定処理に適し、かつ上記2つの制約に対しても有効な話題推定を可能とするためである。

まず文単位の処理による制約に対する解法として、各話題に関する文集合を、その中の各文の類似度をもとにより小さな部分文集合へとクラスタリングする(この結果できた各クラスタをサブトピックと呼ぶことにする)。そして、入力文とこの

サブトピックとの関連度の計算により話題推定を行なう。これは各話題中の類似文をクラスタ化することで、入力文の関連度の計算対象の特徴語数を増やすことが目的である。

次に対話構造による制約に対する解法として、現在の入力文を対話履歴文と組合せ、対話履歴文集合とする。そして、この対話履歴文集合と各話題との関連度の計算により話題推定を行なう。この目的は先程と同じく計算対象の特徴語数を増やすことと、対話特有の対話履歴文の間の関連性や対話の流れを関連度の計算中で扱えるようにするためである。

### 3. 話題推定アルゴリズム

話題推定の基本アルゴリズムとして用いる k 最近傍法、および TF・IDF を用いたクラスタリング手法について述べる。

#### 3.1. 基本アルゴリズム

学習データ中の各文を表すベクトルを  $D_j$ 、入力文を表すベクトルを  $X$  とする。このときベクトルの各要素はそれぞれ一つの単語に対応し、その単語の重要度に応じた値がベクトルの要素の値として与えられる。ここでは各単語の重みは TF・IDF により求めた。

文と文の類似度を、ベクトル空間法に基づき以下のようなベクトルの内積によって求める。

$$Sim(X, D_j) = \frac{\sum_i x_i \times d_{ij}}{\|X\|_2 \times \|D_j\|_2}$$

次に、学習データ  $D_j$  が各カテゴリ  $C_i$  に属する条件付き確率を以下のように定める。

$$Pr(C_i | D_j) = 1 / (D_j \text{ が属するカテゴリ数})$$

このとき入力文  $X$  と各カテゴリ  $C_i$  との関連度は、類似度の値が大きいものから k 個の学習データの和をとった値

$$Rel(C_i | X) = \sum_{D_j \in \text{類似度上位}k\text{文書}} Sim(X, D_j) \times Pr(C_i | D_j)$$

により得られる。

#### 3.2. 話題のクラスタリング

次に、各話題をより小さな類似文の集まりへクラスタリングする手法を説明する。この手法は、ある文集合中の特徴語を求め、その語を含む文集

合と含まない文集合へ再帰的に 2 分割していくというものである。

手法：TF・IDF を用いたクラスタリング手法

- (1) 文集合の各文に含まれる単語のうち最大の TF・IDF 値を持つ語を求める。
- (2) (1) で得られた単語を含む文集合と含まない文集合に 2 分割する。
- (3) TF・IDF が閾値に到達するまで(1),(2)を再帰的に繰り返す。

この手法で得られるクラスタは、各話題を強く特徴付ける語およびその関連語の単語集合となる。

#### 3.3. 対話履歴の導入

対話履歴を利用した話題推定における類似度は、入力文を  $X$ 、対話履歴を  $X_{r_i}$  として

$$Rel(C_i | X, X_{r_1}, \dots, X_{r_n}) = \lambda Rel(C_i | X) + \lambda_{r_1} Rel(C_i | X_{r_1}) + \dots + \lambda_{r_n} Rel(C_i | X_{r_n})$$

で求める。ここで  $\lambda$ 、 $\lambda_{r_i}$  は入力文および各対話履歴文に対する重みである。

### 4. 評価実験

提案手法の有効性を評価するために、旅行会話を対象として学習データと評価データを用意した。クラスタは閾値を変えて 3 種類作成した。対話履歴の重みは 2 種類用意した。

#### 4.1. 学習データ

評価実験では、独自に構築した旅行会話コーパスの約 25,000 文を学習データとして利用した。学習データは人手により、旅行における主要な話題である「ホテル」、「レストラン」、「ショッピング」および「その他」の正解ラベルを付与してある。ここで「その他」とは旅行場面に出てくる話題のうち最初の 3 つの話題に該当しないもので、「交通」や「病気・けが」などがある。複数の話題に分類される文があった場合は、複数の話題の正解ラベルを付与することにした。また、全ての話題の正解ラベルを付与された文は「基本会話」という一つの話題として扱うことにした。

#### 4.2. 評価データ

評価データは性質が異なるものを 2 種類用意し

た。1つめは旅行における典型的な対話を想定し作成したもので、全部で62対話896文からなる(以後これを典型的評価データと呼ぶ)。もう1つはより実際の対話に近い文を得るために、非典型的な対話も含めて想定し作成したもので、全部で45対話498文からなる(以後これを非典型的評価データと呼ぶ)。

典型的評価データに含まれる対話は旅行において誰もが使い、実際の対話では始めの方に出てくるような典型的な基本対話例であり、上記の各話題に頻出する単語や表現が多く使われている。また、1文1文が短く冗長な表現は少ない。非典型的評価データはより実際の場面で交わされる対話に近いものであり、個別の話題に言及したものが多く、使用される単語や表現は典型的な対話に含まれるものから学習データにはほとんど現れないものまで様々である。表現も冗長でくだけた口語的なものが多い。

#### 4.3. 話題のクラスタリング

「ホテル」、「レストラン」、「ショッピング」のいずれか一つ以上の正解ラベルが付与されている学習データ内のデータ8,457文に対して前記手法を用いたクラスタリングを行なった。3種類の閾値でクラスタをつくったところ、各々のクラスタ数は8,409クラスタ(クラスタ数大)、3,845クラスタ(クラスタ数中)、2,203クラスタ(クラスタ数小)、となった。クラスタリングの分割の際、TF・IDF値が閾値以上の語を含まない文は、一文一クラスタとして展開した。また、話題の正解ラベルとして「その他」のみが付与されている文と、4つの話題すべての正解ラベルが付与されている文(「基本会話」)に関しては今回はクラスタリングを行なわなかった。ラベルの組合せの数から選択すべき話題の総数は13種類となる。

一話題を一クラスタとした場合、クラスタ数は最小となり、13クラスタとなる。一文を一クラスタとした場合、クラスタ数は最大となり8,457クラスタとなる。

#### 4.4. 対話履歴の利用

対話履歴を利用した場合の効果を調べるため、ここでは入力文、前文、前々文の3文を対話履歴とする。またこれら3文の重みとして、均等な重みと時系列な重みの2種類を与える。重みは前者を(入力文、前文、前々文) = (0.33, 0.33, 0.33)、

後者を(0.5, 0.3, 0.2)とした。

### 5. 実験結果

4.3で述べた13種類の話題に関する推定実験を、典型的評価データと非典型的評価データを用いて行なった。

#### 5.1. 典型的評価データに対する結果

クラスタ数を変化させた場合の典型的評価データに対する話題推定結果を図1に示す。これによると、各話題において一文を一クラスタとした場合の推定精度が最も良く、各話題に属する文集合をまとめて一つのクラスタとした場合の推定精度が最も悪くなっている。

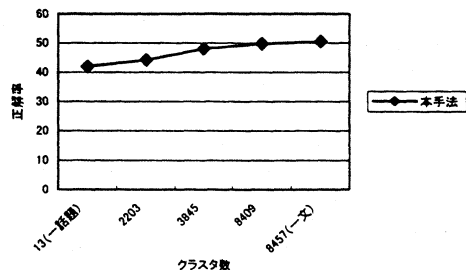


図1 典型的評価データに対する評価結果

#### 5.2. 非典型的評価データに対する結果

クラスタの大きさを変化させた場合の非典型的評価データに対する話題推定結果を図2に示す。これによると、クラスタ数中の場合が一文一クラスタの場合を推定精度で若干上回っている。これは、類似文を文集合として扱うことで、入力文との類似性計算において推定精度が高まることを示している。

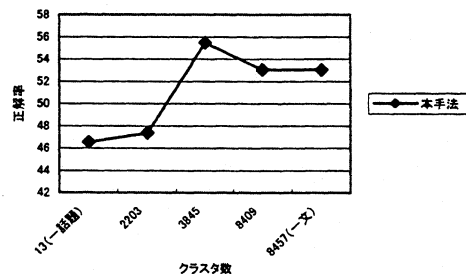


図2 非典型的評価データに対する評価結果

### 5.3.対話履歴を利用した結果

評価には典型的評価データを用い、一文ークラスタの場合とクラスタ数中の場合に関して調べた。入力文 1 文だけから話題を推定した場合の推定精度は一文ークラスタの場合とクラスタ数中の場合においてそれぞれ 59.2%、56.0%、3 文で均等な重み付けを行なった場合で 72.0%、70.0%、時系列な重み付けを行なった場合で 76.7%、77.0%となった。

## 6. 考察

典型的評価データを用いた評価実験において、一文ークラスタとした場合に最も高い推定精度が得られているのは、評価データが典型的な対話であり、かつ一文の長さが短いため、入力文の特徴語と学習データの特徴語が完全に一致する可能性が高いためと思われる。これに対し、一話題ークラスタとした場合に最も低い推定精度が得られているのは、様々なサブピックが含まれる話題を一つのクラスタとして扱うと、その話題を強く特徴づける単語の効果が失われるためと思われる。

例えば、学習データの「手にとって見てもいいですか。」という文は、ショッピングの際に商品を吟味しているところと考えられ、「ショッピング」という話題に属する。ここで入力文として「あれを手にとってもいいですか。」があった場合、一文ークラスタとして話題推定すると、先程の学習データとの類似性から正しい結果が得られる。しかし一話題ークラスタとして話題推定すると、「手が痛い。」や「体調がいい。」といった学習データを含む「その他」へと誤って推定される。

非典型的評価データの推定結果を解析すると、一文ークラスタの場合とクラスタリングした場合において、正解率がほぼ同じでも、個々のどの文を正解したかは大きく異なっていた。一文ークラスタが正解を出しやすいのは文中に話題を比較的強く特徴づける語がある場合で、クラスタが正解を出しやすいのは、文中の個々の単語は話題を強く特徴付ける語でなくても複数個並べて文としてみると、話題が特定できるような場合であった。

つまり、典型的な入力文に対しては一文ークラスタと比較し入力文の特徴を忠実にとらえることができ、非典型的な入力文に対してはクラスタと比較し、典型例からのずれ、内容や表現のゆれに対する頑健性を持たせることができる。今回の評価結果では、一文ークラスタに対してクラスタ数

中の推定精度は、典型的評価データにおいてほぼ同等、非典型的評価データに対しては、若干の改善が得られている。よって大量の旅行会話データを収集して、典型的対話および非典型的対話の双方に対して最適なクラスタ数を定めることによって実用的な話題推定が可能となると考えられる。

## 7. おわりに

本稿では、自動通訳における多義語の曖昧性解消や音声認識率向上に対する、話題の推定問題に関して述べ、その解決方法について提案した。13種類の話題に対し、典型的・非典型的双方の対話を 50%以上の精度で推定できた。典型的な対話に対しては、一文をークラスタとして扱う場合に最も話題推定の精度が上がる事が確認された。また、非典型的な対話に対しては、クラスタリングを導入した場合、一文ークラスタの場合に比べて若干推定精度が向上することがわかった。従って、適切なサイズのクラスタを導入することで典型的・非典型的両方の入力文に対し高い推定結果が得られる。

ここで述べた手法は、話題の遷移など時系列な流れを推定する場合に、より有効に働くものと思われる。今後は、対話履歴とクラスタをより緊密に関連付けることで、さらなる推定精度向上をはかる予定である。

## 参考文献

- [1] 奥村明俊、山端潔、土井伸一、坂井信輔、亀井真一郎、高橋一裕、花沢健、佐藤研治、渡辺隆夫:旅行会話向け自動通訳システム、情報処理学会第 61 回全国大会、2U-3 (1999)
- [2] 羽鳥洋美、神山淑朗:分野判定トリガー情報のフィードバックによる Web 翻訳、情報処理学会第 63 回全国大会、5Y-5 (2000)
- [3] 角田達彦、田中英彦:英語名詞の多義性解消における文脈としての場面情報の評価、自然言語処理、Vol.3 No.1 (1996)
- [4] Y. Yang: Expert Network. Effective and Efficient Learning from Human Decisions in Text Categorization and Retrieval, ACM SIGIR (1994)