

対話における話題の抽出を目的とした発話クラスタリング

福永 隼也[†] 西川 仁[†] 徳永 健伸[†] 横野 光[‡] 高橋 哲朗[‡]

[†]東京工業大学 情報理工学院 [‡]株式会社富士通研究所

fukunaga.s.ab@m.titech.ac.jp {hitoshi,take}@c.titech.ac.jp
 {yokono.hikaru,takahashi.tet}@jp.fujitsu.com

1 はじめに

近年、人間と自然言語によってコミュニケーションを行う対話システムが注目を集めている。対話システムは、ユーザが明確な目的を持つか否かによって「目的指向対話」と「雑談対話」に分類されることが多い。これらの中間的な対話として、その目的が漠然としている対話に焦点を当てた研究もある (Stede and Schlangen, 2004; 市川, 2007)。Stede and Schlangen (2004) はユーザの目標が漠然とした対話を “information-seeking chat” と名付け、そのような対話には雑談のように様々な話題が出現し、次々に遷移していくと述べている。

このような対話では、各時点での話題を判断し、話題に基づいて対話を進めることで、対話システムがより自然な対話を行える。しかし、対話中に出現する話題はドメインによって大きく異なるため、ドメインの知識なしに話題を判断することは難しい。そこで、予め対話システムの想定するドメインに出現する話題を抽出しておくことが考えられる。

話題に基づいて対話を進める従来の対話システムでは、予め話題を手手で定義している (谷津ら, 2013; Lane and Kawahara, 2005)。我々は、対話システムの想定する実際の対話例中の発話をクラスタリングすることでより効率的に話題の抽出を行うことができると考えている。なぜなら、対話例を用いることによって、実際に対話を行わなければ思い付かないような話題を発見しやすくなり、またクラスタリングによって類似した発話を整理することで、話題が抽出しやすくなるためである。

本稿では、対話コーパス中の発話をクラスタリングすることによって話題を抽出するという目標を達成するための第一歩として、発話クラスタリングの精度向上を図る。

2 関連研究

テキストをクラスタリングすることにより、有用な知識を発見する研究はこれまで数多くなされている。東中ら (2011) は、対話中の発話を無限 HMM によってクラスタリングすることにより、対象ドメインの対話にどのような対話行為が現れるかを発見しようとしている。対話中の発話をクラスタリングし、何らかの知識を発見するという点で本研究と類似しているが、本研究では話題を対象としている。

菊池ら (2008) はニュース記事などのテキストから、階層型クラスタリングを用いることによって話題を抽出する手法を提案している。クラスタリングによって話題を抽出するという点で本研究と類似しているが、本研究では対話中の発話を対象としている。

佐藤ら (2016) は Twitter 上の雑談対話をクラスタリングすることにより、対話データをサブドメインに分割し、サブドメインごとに対話モデルを学習している。対話中の発話をクラスタリングし、対話データを分割するという点で本研究と類似しているが、佐藤らのサブドメインは、“!!!” を含む”など発話の表層的な特徴のみによってまとまっているクラスタも含んでおり、必ずしも話題には対応しない。また、発話クラスタリングの結果を評価していない。

本稿では、ガイドと旅行者が観光プランについて対話するコーパスについて、特にユーザの質問発話から話題を抽出することを試みる。

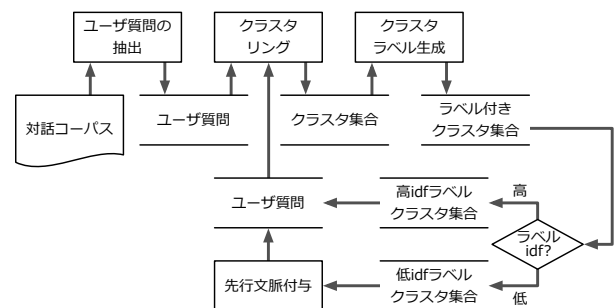


図 1: 提案手法の概要

3 提案手法

3.1 概要

提案手法では、話題を導入することが多いと考えられるコーパス中のユーザの質問をクラスタリングし、各クラスタの特徴的な語をラベルとして付与する。そして、各クラスタのラベルとなる単語の idf を計算し、ラベルの idf が小さいクラスタは、一般語が原因でまとまっていると判断する。そのようなクラスタ中の発話の特徴を明確にするために、クラスタ中の発話に文脈を追加し、再度クラスタリングを行うことで、これらの発話がより適切なクラスタに含まれることを狙う。

図 1 に我々が提案するクラスタリング手法の概要を示す。以降では、各処理の詳細について述べる。

3.2 ユーザの質問の抽出

対話コーパス中の発話から、「～ですか」、「～ですか」など予め設定した 16 のキーワードを含む発話をユーザの質問発話として抽出する。ユーザの質問のみを用いる理由は、ユーザの質問が新たな話題を導入すると考えたからである。また、ユーザの質問は対話システムが対処すべき発話であるため、ユーザの質問によって導入される話題を抽出することは優先度が高いと考える。

3.3 発話クラスタリング

抽出したユーザの質問を整理し、話題を抽出するために群平均法によってクラスタリングする。意味内容の近さで発話をクラスタリングをするために、単語として 2 回以上出現する内容語(名詞、動詞、形容詞、副詞)のみからなる bag-of-words モデルで発話を表現する。固有名詞は NUMBER や PERSON などの意味カテゴリに変換し、その他の単語は正規形を用いる。Bag-of-words のベクトルの各要素は、1 発話を 1 文書として tf-idf により重み付けを行う。また、発話同士の距離尺度には、ベクトル間のコサイン距離を用いる。

3.4 クラスタラベルの生成と idf の計算

各クラスタを形成するのに特徴的な語を 1 語選び、それをクラスタのラベルとする。具体的には、各クラスタに含まれる発話中の各単語の相互情報量を計算し、これが最大となる語をラベルとする。

各クラスタのラベルについて事前に分析したところ、ラベルが「する」「ある」のような一般語になる場合、そのクラスタは話題というよりはそのような一般語を含むためにクラスタにまとまっているという傾向が見られた。たとえば、「ありますか?」という発話だけでは何の話題なのかかわからないが、このような発話が、話題として適切なクラスタに含まれるようにすることで、クラスタリング精度の向上が期待できる。そのために、まず一般語がラベルとなるクラスタをラベルの idf を基準として抽出する。一般語は、どのクラスタにも出現しやすいので、idf が小さくなるはずである。各クラスタについてクラスタを文書としてラベルの idf を計算し、ラベル idf が小さいクラスタを抽出する。

3.5 先行文脈の追加と再クラスタリング

ラベルの idf が小さいクラスタに含まれる発話が、話題としてより適切なクラスタに含まれるようにするために、その発話に先行文脈を追加し、再クラスタリングを行う。先行文脈を追加する理由は、その発話に関する情報が増え、一般的な単語以外でまとまるクラスタに含まれやすくなると思ったからである。ここでは、文脈として、その発話の直前のユーザの一発話を加える。たとえば、前述の「ありますか?」という発話の直前のユーザの発話「おいしいお店って。」を追加することによって食事に関する話題であることがわか

る。再クラスタリングを行う際は、文脈を加えた発話だけでなく、全ユーザ質問を再びクラスタリングする。

4 実験

4.1 実験設定

コーパスとして、京都観光案内対話コーパス(大竹ら, 2008)を使用する。このコーパスは、京都観光案内のエキスパートガイドと模擬旅行者(ユーザ)が、京都市内の一日観光の計画立案を行う対面対話を書き起こしたものである。100 対話が収録されており、1 対話 30 分、総発話数は 62,835 発話である。設定としては、京都観光について十分な知識を持っておらず、明確なプランも持っていないユーザがガイドと相談をしながら一日のプランを決定する際の対話である。

4.2 評価方法

正解データを人手で作成し、クラスタリング結果から正規化相互情報量(NMI:Normalized Mutual Information)、精度、再現率を計算する。

NMI は相互情報量を正規化した尺度であり、計算式は以下の式で表される。

$$NMI(\Omega, \mathbf{C}) = \frac{I(\Omega; \mathbf{C})}{(H(\Omega) + H(\mathbf{C}))/2} \quad (1)$$

$$I(\Omega; \mathbf{C}) = \sum_k \sum_j \frac{|\omega_k \cap c_j|}{N} \log_2 \frac{N \cdot |\omega_k \cap c_j|}{|\omega_k| \cdot |c_j|} \quad (2)$$

$$H(\Omega) = - \sum_k \frac{|\omega_k|}{N} \log_2 \frac{|\omega_k|}{N} \quad (3)$$

$$H(\mathbf{C}) = - \sum_j \frac{|c_j|}{N} \log_2 \frac{|c_j|}{N} \quad (4)$$

ここで、 $\Omega = \{\omega_1, \dots, \omega_K\}$ 、 $\mathbf{C} = \{c_1, \dots, c_J\}$ であり、 ω_k はクラスタリング結果の k 番目のクラスタ、 c_j は正解データの j 番目のクラスタ、 K はクラスタリング結果のクラスタ数、 J は正解データのクラスタ数、 N はユーザ質問の数である。

また、精度、再現率は、発話のペアに着目し計算する。精度はクラスタリングの結果で同じクラスタに含まれる任意の発話の組が、正解データでも同じクラスタに含まれるか割合を表す。一方、再現率は、正解データで同じクラスタに含まれる任意の発話の組が、クラスタリングの結果でも同じクラスタに含まれる割合を表す。計算式は以下の式で表される。

$$\text{精度} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{再現率} = \frac{TP}{TP + FN} \quad (6)$$

ここで、TP は正解データでも、クラスタリングの結果でも同じクラスタに含まれる発話の組の数、FP は正解データでは異なるクラスタに属するにも関わらず、クラスタリングの結果では同じクラスタに含まれる発話の組の数、FN は正解データでは同じクラスタに含

まれるにも関わらず、クラスタリングの結果では異なるクラスタに属する発話の組の数である。

評価に用いる正解データの作成は次のように行った。まず、予め対話に現れる話題をいくつか人手で抽出する。各話題について、その話題を表すと判断したユーザの質問を人手で集めてクラスタを作成し、これを正解データとする。正解データでは、1つのクラスタが1つの話題に対応する。

正解データを作成する際は、「何を尋ねているか」が同じだと判断したユーザの質問を一つのクラスタにまとめる方針をとった。このとき、「何を尋ねているか」はできるだけ汎化する。例えば、「南禅寺の拝観料は、いくらぐらいですか?」と「で、東寺は拝観料とかがいくらぐらいですか?」という発話は、それぞれ南禅寺と東寺の拝観料を尋ねているが、これを汎化して「観光地の料金」という話題に含める。また、その発話だけではなく、その前後の文脈も見て話題を判断する。例えば、同じ「で、おいくらぐらいなんですかね?」という発話が文脈によってはバスなどの交通手段の料金を聞いている場合と、観光地の料金を聞いている場合がありうるが、「観光地の料金」という話題には後者の発話のみを含める。この方法で11個のクラスタを作成した。正解データに含まれる発話数は合計287発話であった。

4.3 実験手順

実験は、正解データに含まれていない発話を含めてユーザの質問全体を用いて行った。クラスタリングアルゴリズムとして階層型クラスタリングを用いているため、閾値を定めてクラスタリング結果を分割する必要がある。分割するクラスタ数を1ずつ変化させ、全ての場合について評価を行う。

ラベルの生成と先行文脈の追加を行うためには、ある一つのクラスタ数でのクラスタリング結果を使用する必要がある。クラスタ数を決定する基準として、1回目のクラスタリング結果の評価を用いる。1回目のクラスタリング結果のNMIが最大となるクラスタ数を、以降のラベルの生成とコンテキストの追加を行うために用いる。

4.4 実験結果

ユーザの質問を抽出した結果、全62,835発話中2,796発話がユーザの質問として抽出された。そのうち100発話をランダムサンプリングし、各発話について質問か否かを人手で判断してユーザの質問抽出の精度を計算した。その結果、100発話中75発話をユーザの質問であった。誤った発話の多くは、相手の言葉を繰り返して相槌を行う「○○ですか」や、「本当ですか」のような相槌、相手の発言を補う「○○ですよ」など、質問というよりもコミュニケーション上の機能を持つ発話であった。

再クラスタリングを行った際の評価結果の変化を示す。ここで、ラベルの生成と先行文脈の追加のために用いるクラスタ数は468であった。各回のNMIが最大

表1: 各回でNMIが最大となる時の評価結果

回数	クラスタ数	NMI	精度	再現率
1回目	468	0.550	0.916	0.090
2回目	431	0.553	0.938	0.092
3回目	420	0.556	0.935	0.092
4回目	421	0.556	0.935	0.092

となる時の評価結果を表1に示す。また、精度、再現率の変化を表す再現率-精度曲線を図2に示す。再現率-精度曲線はクラスタリングの回数ごとに描いている。

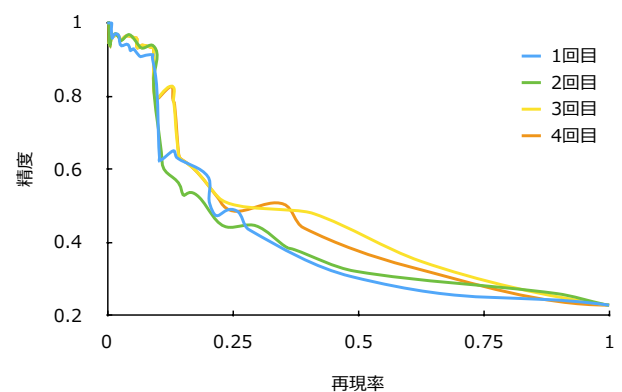


図2: クラスタリング回数ごとの再現率-精度曲線

5 考察

ユーザの質問の抽出精度は75%であり十分な精度とは言えない。質問と同じ形で相槌の機能を持つような発話をキーワードによる抽出では区別できなかったことが原因である。このような発話は、直前の発話などの文脈を考慮することで区別ができると考えられる。

表1を見ると、各回でNMIが最大となる時の評価結果は、精度が高く、再現率が低くなっていることがわかる。これは正解クラスタが11個であるのに対し、クラスタ数が400個以上生成され、まとまるべきクラスタが分割されてしまったのが原因だと考えられる。回数ごとの評価の変化を見ると、1回目と比較して2回目以降は精度が上昇している。また、NMIが最大となるクラスタ数は減少している。一般にクラスタ数を減らすと精度は低下するため、クラスタ数が減少し精度が上昇したという結果から、再クラスタリングによってクラスタの純度が向上したと考えられる。一方、再現率は2回目以降もほとんど変化していない。

図2に示した再現率-精度曲線を見ると、3回目、4回目のクラスタリング結果では曲線が1回目よりも上に位置し、全体的に精度が向上したことがわかる。特に、1回目と3回目を比較すると再現率が0.3以上の場合に、精度が大きく向上していることがわかる。

次に、実際のクラスタリング結果を見て、一般的な

単語でまとまっているクラスタ中の発話に先行文脈を追加し、再クラスタリングを行う効果を確かめた。たとえば、「五百円ですか。」という発話は、1回目のクラスタリングでは、idfの小さいNUMBERという意味カテゴリをラベルとするクラスタに含まれており、一つの話題でまとまっていなかった。この発話に直前のユーザ発話「ここは、入館料は、はい。」を追加し再クラスタリングすることで、ラベル「料」を持つ「観光地の料金」という話題を表すと考えられるクラスタに含まれるようになった。これは、もともと「五百円ですか。」という発話には数を表すNUMBERというシンボルと「円」という名詞のみが内容語として含まれていたが、直前の発話を追加することによって「料」という話題を表しやすい単語が含まれるようになったのが要因だと考えられる。

一方、この方法で効果がない、あるいはむしろ精度が悪化した場合もある。たとえば「六百円ですか。」という発話は、1回目のクラスタリングでは、先程の例と同様のNUMBERをラベルとするクラスタに含まれていた。この発話に直前のユーザ発話「はい。」を追加し再クラスタリングを行っても効果はなく、再びNUMBERをラベルとするクラスタに含まれた。これは、直前のユーザ発話に有益な情報を含んでいないことが原因である。この問題は単純に直前のユーザ発話を用いるのではなく、相槌など内容語を含まない発話を除くことによって解決できると考えられる。

また、「何かお勧めはありますか、近くに。」という発話は、1回目のクラスタリングでは、「ある」というidfの小さい単語をラベルとするクラスタに含まれていた。しかし、このクラスタは、「そういうお勧めのスポットはありますか?」や「なにかお勧めの所ありますか?」のような何かについてガイドのお勧めを尋ねる発話でまとまっていた。これらの発話に直前のユーザ発話を追加し再クラスタリングを行うことで、何についてのお勧めなのかという点で分かれることが期待されたが、実際は全く異なるクラスタに分かれてしまった。特に「何かお勧めはありますか、近くに。」という発話は、直前の「二時、二時十五分。」という発話を追加して再クラスタリングした結果、前述のNUMBERというシンボルをラベルとするクラスタに含まれてしまった。これは、直前のユーザの発話が、ガイドのお勧めという話題に関係のない発話だったことが原因である。「何かお勧めはありますか、近くに。」という発話は直前のガイドの発話、「あと、こっからどこへ向かって行きましょう。まだ、四条戻るには早い、もう一ヶ所ぐらい行けそうですね。」という発話を受けての発話であるため、文脈として追加するにはガイドの発話のほうが適切であった。

またガイドのお勧めを尋ねる発話でまとまっていたクラスタを、ラベルのidfが小さかったことから話題としてまとまっていなと判断してしまったことも原因だと考えられる。提案手法では、クラスタ内の発話中の語のうち相互情報量が最大の語をラベルとしているが、この例では相互情報量が2番目に大きい単語は

「勧め」であった。このことから、相互情報量が最大となる単語だけではなく、2番目以降の単語もそのクラスタの意味的なまとまりを表すと考えられるため、この情報を利用することを検討する必要がある。

6 おわりに

本研究では、話題抽出のための発話クラスタリングの精度向上を図った。発話クラスタリング後に各クラスタにラベルを付与し、そのidfを用いることで一般語でまとまっているクラスタを抽出した。そして、そのようなクラスタに先行文脈を追加し、再びクラスタリングを行った。以上の方法によって、発話クラスタリングの精度が向上することが確認できた。

現在は素性として表層の情報だけを用いているが、より意味内容を表すような素性を使用することで、話題がクラスタとしてまとまりやすくなると考えられる。また、今回ユーザの質問のみを使用したため、利用可能な発話数が大幅に減ってしまった。したがって、今後は、ユーザの質問以外の発話も話題抽出のために使用したいと考えている。また考察から、先行文脈としてユーザの直前の発話だけでは不十分であることと、クラスタの意味的なまとまりを表すためには相互情報量が最大となる単語だけでは不十分であることが示唆された。この点について、今後より深く検討していきたい。さらに、現在は大量のクラスタを人が見て、その中から話題を発見する必要があるため、人の負担を減らす提示方法を検討する予定である。

参考文献

- Ian R. Lane and Tatsuya Kawahara. Incorporating dialogue context and topic clustering in out-of-domain detection. In *Proceedings of IEEE-ICASSP*, pp. 1045–1048, 2005.
- Manfred Stede and David Schlangen. Information-seeking chat:dialogue management by topic structure. In *Proceedings of 8th Workshop on the Semantics and Pragmatics of Dialogue (CATALOG '04)*, pp. 117–124, 2004.
- 菊池 匡晃, 岡本 昌之, 山崎 智弘. 階層型クラスタリングを用いた時系列テキスト集合からの話題推移抽出. *日本データベース学会論文誌*, 7(1):85–90, 2008.
- 佐藤 翔悦, 石渡 祥之佑, 吉永 直樹, 豊田 正史, 喜連川 優. 発話状況を意識したオンライン上の対話における応答選択. 2016年度人工知能学会全国大会論文集 (JSAI 2016), 205-OS-23b, 2016.
- 市川 宙. 情報探索雑談における自然なトピック遷移の実現. 修士論文, 東京工業大学, 2007.
- 大竹 清敬, 堀 智織, 柏岡 秀紀, 中村 哲. 京都観光案内対話コーパスにおける対話行為の分析. *言語処理学会第18回年次大会発表論文集*, pp. 159–162, 2008.
- 谷津 元樹, ジェブカ ラファウ, 荒木 健治. トピック推定を用いたタスクドメインを選択するための発話生成. *言語処理学会第19回年次大会発表論文集*, pp. 142–145, 2013.
- 東中 竜一郎, 川前 徳章, 貞光 九月, 南 泰浩, 目黒 豊美, 堂坂 浩二, 稲垣 博人. 対話行為設計のための発話クラスタリング. *人工知能学会研究会*, 第63回 SLUD 研究会, pp. 37–42, 2011.