

文字対話における対話関係成立の可否判別

三ノ宮 和俊 山本 和英

長岡技術科学大学 電気系

{sannomiya, yamamoto}@jnlp.org

表1 文字対話の例

	ツイート	返信
例1	おはようございます	おはようございます。
例2	9時わりぽよ。今起きたなう。	私もいま起きた。

1 はじめに

自然言語処理の文生成分野において、人間同士で行われるような自然な対話を自動的に生成することは困難とされている。この要因として、話題の変動が激しいため、対話の応答として適切な処理が行えていない点などが考えられる。また、実際の人間と対話を行う対話(雑談)システムについての研究も既に行われているが、実用的なレベルにはまだ達しておらず、対話の精度を高めていくことが課題である。

本研究では対話の精度を向上させるために、別々の人物によって作成された二つの文が対話として成立しているかどうかを自動的に判定することを目的とする。これは、不特定多数の二文間において、対話関係が成立しているかどうかの判別が高い精度で行うことができれば、対話システムなどにおける対話の精度が向上すると考えたからである。実際に人間同士で行われた対話内容を大量に収集するのは困難であるため、Web上のテキストで行われた人間同士のやりとりを対話として扱うことにした。このテキストでの人間同士のやりとりを文字対話として扱い、文字対話に対して対話関係が成立しているかどうかの判別を行った結果を示す。また、判別の際に使用した素性の有効性についても検証を行う。

2 関連研究

今井ら[1]はユーザの入力に対して複数の既存対話システムにより生成された応答候補の中から最適な応答文を選択することで、自然な対話を行うことを目指した。最適な応答文を選択するために、実際に人対人で行われたチャットログとの類似度から入力文に対する応答文を選択する方法や、入力文と各応答候補中の単語の共起頻度を求めて入力文との関連性がある応答文を選択する方法などを行っていた。

稲葉ら[2]は対話において意味的な正しさを持つ発話間には語の共起が存在すると仮定し、共起語対を発話間に含むかどうかで対話エージェントと人間との対話の評価を行っている。結果より共起語対による評価と人手による評価に相関関係が確認されている。

我々の目的は対話関係の可否判別を行うことであり、対話関係が成立していることに加えて不成立であることも判定する必要がある。この点が本研究と関連研究との違いである。

3 文字対話について

本研究では、Web上のテキストで行われた人間同士のやりとりを文字対話として扱う。実際に行われた対話は大量に収集するのが困難な点、対話の内容に偏りが生じる可能性がある点、不特定多数の人物からの収集が困難な点などが問題である。Web上のテキストであれば、これらの問題点を気にする必要がなく、容易に収集することが可能である。ただし、「人間同士のやりとり」が必ず行われていることを保証する必要がある。そこで、本研究ではTwitter^{*1}に着目した。

Twitterとは、140文字の文字制限の中で「ツイート」と称

する短文を発信できるWebサービスのことである。ツイート以外にも、あるツイートに対してツイートを行う「返信」、他のユーザーが投稿したツイート内容を再投稿する「リツイート」などを行うことが可能である。

会話とは基本的に「応答ペア」という会話の中の対をなす二つの発話によって成り立っていると考えられている[3]。この二つの発話を「第1の発話」と「第2の発話」という順序で、二人の話者が交互に発していくことで対話が形成されており、「第一の発話」が「第二の発話」を導くことになる。本研究では、実際に発信されたツイート内容に対して、ツイート送信者とは別の人間が返信を行っている点から、ツイートを「第一の発話」、返信を「第二の発話」と考え、ツイート・返信の関係は人間同士の対話が成立している関係にあると仮定して、文字対話として用いることにした。表1に文字対話の例を示す。なお、本研究では1ツイート単位で扱っていくため、単文や重文などは考慮していない。以降、ツイートと返信の1つの組み合わせを「発話対」として表記する。

4 可否判別までの流れ

文字対話について対話関係の可否判別を行うことは、二値分類問題として扱うことができる。本研究ではSVMを用いて対話関係の可否判別を行うことにした。対話関係の可否判別の際に用いる素性として、以下の三つの要素を使用する。

- (A) 語の共起情報
- (B) 対象発話対の疑問判定
- (C) 内容語・未知語クラスタリング

4.1 語の共起情報

語の共起とは、任意の文書内においてある語と別の語が同時に出現することである。もし、発話対内に共起しやすい単語が存在した場合、その発話対は対話として関係があると期待できる。そこで本研究では、素性の一つとして語の共起情報を用いることにした。語の共起情報の獲得には、Twitterから収集した約20万の発話対を使用し、獲得する品詞は名詞、動詞、形容詞、未知語とした。品詞情報の獲得には形態素解析器 MeCab^{*2}を使用した。

共起情報については、共起の強さを表す統計的指標の一つである対数尤度比[4]として用いることにした。これは稲葉ら[2]の研究で、対話の意味的な正しさに観点を置いた際の評価結果において、人間による評価と対数尤度比を用いた場合の評価に強い相関があると報告されており、対数尤度比が本研究の目的に適していると考えたからである。対数尤度比

^{*1} Twitter, <http://twitter.com>

^{*2} 形態素解析器 MeCab, Version0.98, 工藤拓,
<http://mecab.sourceforge.net/>

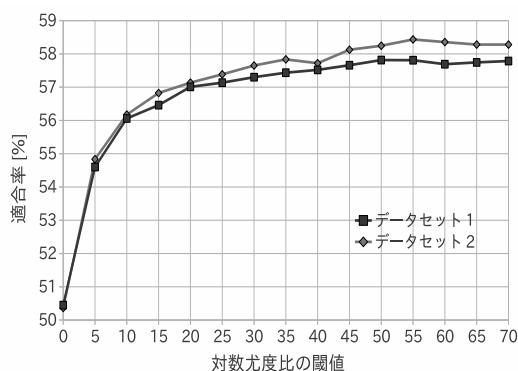


図1 対数尤度比の閾値と対話関係の判別結果

の算出方法を式(1)に示す。

$$LLR = a \log \frac{aN}{(a+b)(a+c)} + b \log \frac{bN}{(a+b)(b+d)} + c \log \frac{cN}{(a+c)(c+d)} + d \log \frac{dN}{(b+d)(c+d)} \quad (1)$$

本研究では語 A と語 B についての対数尤度比を算出する際、式(1)の各変数は以下の意味を表す。

- a : 語 A がツイート文、語 B が返信文に存在する場合の発話対の数
- b : 語 A がツイート文に存在し、語 B が返信文に存在しない場合の発話対の数
- c : 語 B が返信文に存在し、語 A がツイート文に存在しない場合の発話対の数
- d : 語 A も語 B も存在しない場合の発話対の数
- N : 対数尤度比算出の際に使用した全発話対の数

式1によって算出された二語間における対数尤度比の値が $LLR > 0$ であれば、その二語は従属の関係にあると言える。しかし、対数尤度比の値がどの程度あれば従属の関係に信頼性が持てるのかわからないため、以下の手順で行った予備実験により対数尤度比に関する閾値を求めることにした。

- 手順1 対話関係が成立したデータと成立していないデータを合わせたデータセットを作成する(データセット1)
- 手順2 データセットに依存した閾値にならないように、データセット1とは内容が異なるデータセットを作成する(データセット2)
- 手順3 作成したデータセット1・2に対して、語の共起情報のみを素性として使用し、対話関係が成立しているか否かの判定を行う
- 手順4 判定結果において、素性の特徴量を付与するための対数尤度比の閾値を0から5刻みで変化させ(0, 5, 10, ...), 閾値ごとの精度を比較し、最も良い精度が得られたときの閾値を求める

予備実験の結果は図1のようになり、データセット1・2に対して最も高い精度となった55を対数尤度比の閾値として設定した。

4.2 対象発話対の疑問判定

対話において、疑問や質問を相手に問いかけた場合、その問いに対する応答が相手から返ってくることが期待される[3]。もし、問いかけに対して適切な応答をせずに疑問や質問を返した場合、対話として成立していないと考える。平井ら[5]の研究においても、文脈情報として質問であるか否かを考慮した場合、発話選択の精度が向上したと報告されており、本研究の目的である対話関係の判別においても効果的な素性だと

表2 疑問判定に関する素性の付与方法

	ツイートが疑問ではない場合	ツイートが疑問である場合
返信が疑問ではない場合	素性として1を付与	素性として1を付与
返信が疑問である場合	素性として1を付与	素性として0を付与

考える。文字対話における疑問判定の素性は表2に基づいて仮定した。なお、ツイートまたは返信が疑問であるか否かについては、疑問符が文内で使用されているかどうかで行った。

この素性の有効性を確認するために、発話対において疑問符がどの程度の割合で使用されているのか、またツイートと返信のどちらにも疑問符が使用されている発話対はどの程度存在するのかについて調査を行った。なお、調査では会話関係が成立している発話対を1万対使用した。調査の結果、疑問符が使用されている発話対は全体の約25%(2650対)あり、素性の対象となる発話対がある程度存在することがわかった。また、疑問符が使用されている発話対のうち、ツイートと返信のどちらにも疑問符が使用されている発話対は約10%(242対)あり、これは全発話対の約2%に当たる。調査で使用した発話対は会話関係が成立していることから、対話関係が成立している大部分の発話対は、疑問の応答に対して疑問を用いていないことがわかり、表2の付与方法の正当性を確認することができた。

4.3 内容語・未知語クラスタリング

対話関係の判別に語の関連性が有効であると考え、語のクラスタリングを行うことにした。また、本研究では文字対話としてTwitterのデータを用いているため、未知語の出現が多いと考えた。Twitterから獲得した1万の発話対について、MeCabを用いて形態素解析を行い、未知語がどの程度存在するかを調査したところ、約2万語の未知語が存在した。この結果から、一対中に約2語の未知語が存在することになり、Twitterにおける未知語は重要な要素であると考えた。以上のことから、本研究では名詞、動詞、形容詞の内容語と未知語に関するクラスタリングをデータクラスタリングツール「bayon^{*3}」を用いて行った。形態素解析器としてMeCab、コーパスにはWeb日本語Nグラム^{*4}の7-gramデータを使用した。クラスタリング時に用いる素性の獲得方法を図2に示す。図2の対象語がクラスタリングの対象となる内容語又は未知語を現し、その対象語の特徴となる要素として素性を付与する。本研究では同じ7-gramデータ中に出現した対象語以外の内容語又は未知語を素性とし、その素性の値として7gramデータの出現頻度を使用した。なお、対話関係の可否判別の際に用いる素性については、作成された内容語・未知語クラスタを用いて以下の手順で付与を行った。

- 手順1 MeCabを用いてツイート文と返信文の形態素解析を行う
- 手順2 ツイート文、返信文中に出現する内容語・未知語の全組み合わせに対して、同じクラスタに所属するかどうかの判定を行う
- 手順3 同じクラスタに所属する内容語・未知語の組み合わせが存在すれば、クラスタリングの素性として1を付与、存在しなければ0を付与する

^{*3} データクラスタリングツール bayon, Ver.0.0.11, Mizuki Fujisawa, http://code.google.com/p/bayon/wiki/Tutorial_ja

^{*4} 工藤拓, 賀沢秀人. Web 日本語 N グラム 第 1 版. 言語資源協会発行

(GIVE, NIL, 過去)
(NIL, GIVE, 中心)

4. 意味構造の抽出

意味検索を実装するにあたり、意味構造を計算する仕組みが必要である。本研究では、日英翻訳エンジン ATLAS[6]の翻訳過程から意味構造を取り出すことにした。ATLASは中間言語方式の翻訳方式を採用しており、原文を辞書と文法規則に基づき解析して意味構造を計算する。意味構造においては、ノードは概念記号と呼ばれる意味を表す記号で識別される。

5. 検索エンジンの設計

クエリー、検索対象文書を意味最小単位の総和としたときに、そのマッチング方法および各文書の評価値計算方法が重要である。

ATLASの概念記号は形態素よりも抽象化されている。そのため、意味最小単位のマッチは完全一致とする。

評価値計算に関しては、意味最小単位の検索は文単位で行うため、文書(Dk)の評価値は文の評価値(Si)の総和とすることにした。すなわち、

$$Dk = \sum_i Si$$

文の評価値は、以下のようにして行う。データベース作成にあたり、検索対象文書を意味解析して意味最小単位を抽出し、それぞれの意味最小単位の逆文書頻度を計算しておく。検索時では、入力キーを意味解析して得られた意味最小単位 Mn につき、文中の Mn の出現回数に Mn の逆文書頻度をかけあわせる(=En とする)。さらに、En を合計したものに、文中に出現した M の数の 2 乗をかけあわせる。すなわち、文 i の評価値(Si)は以下の式で表される：

$$Si = (\sum_n \text{Idf}(i,n) \times \text{Freq}(i,n)) \times (\sum_n C(i,n))^2$$

ただし、

$\text{Idf}(i,n)$ = 文 i に出現する Mn の idf 値

$\text{Freq}(i,n)$ = 文 i における Mn の出現回数

$C(i,n) = 0$ if 文 i に Mn が出現せず

$C(i,n) = 1$ if 文 i に Mn が出現する

文中に出現した M の数の 2 乗をかけあわせる理由は、入力キー内の意味最小単位が同時に

1 文中に現れれば現れるほど、その文が入力キーに類似した有向グラフを含むと考えられるからである。例えば、意味最小単位の同時出現数が 2 より 3 のほうがはるかに意味が類似する可能性が高い。これにより、解析誤りにより有向グラフが一致しなくても、意味最小単位が同時にマッチすれば評価値が高くなる。

自然言語文で入力したとき、意図しない意味最小単位で検索対象文書を検索してしまう場合がある。この問題に対しては、ユーザに、意味最小単位に相当するクエリー中の表現を提示して選択させることで解決することにした。

6. プロトタイプシステムの開発

以上で説明した検索方法により、意味検索のプロトタイプシステムを構築した。

検索結果は文書ごとにランキングされて、入力キーと類似しているものから順に表示される(図 2)。

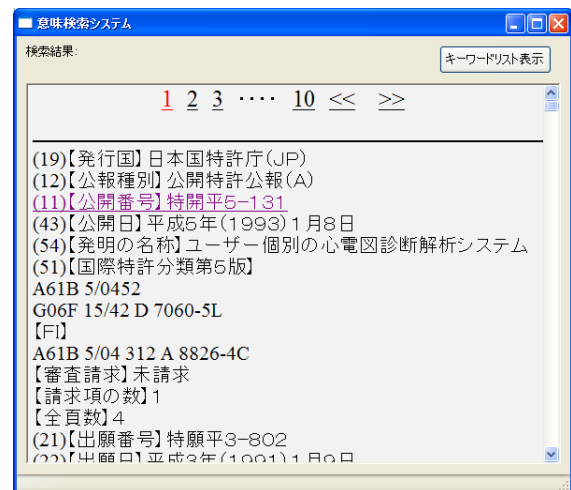


図 2. 検索結果画面

検索された文書の 1 つをクリックすると、明細書が表示されるが、評価値が高い文が強調表示される(図 3)。

表 5 対数尤度比の算出を行った語の組に関する情報

	算出が行えた組の総数	閾値 55 を越えた組の数
異なり数	約 50 万組	約 1,200 組
延べ数	約 70 万組	約 24,000 組

表 6 疑問判定において対応できない対話例

発話対の原文 (／は発話対の区切り文字を表す)
ラーメンライスすか!? /
梅塩ラーメン+小ライス 何でわかったん!?
一人映画いいっしょ? なんか純粋に映画楽しめるしよ /
最近ひとり映画はまってきてん てか会社帰れんとかありえんの!?

表 7 追加実験の際に用いた負例データの例

発話対の原文 (／は発話対の区切り文字を表す)
おはよ! 職場って職場? 現場に! /
おはようございます。寒さも増して、職場でもマスク姿が増えてきたので気を付けなきゃですね!
了解一。気を付けてお帰りくださいー。 /
不具合ですかね?? これからは気を付けます

している。

しかし、予備実験により算出した対数尤度比の閾値に関しては、表 5 の判別実験の評価データに対して行った結果より、多くの共起情報を使っていない状態であると言え、改善の余地があると言える。この対策として、対数尤度比を算出する際に使用する語の頻度情報をより大規模なコーパスから求めることで、語のスパースネスを解消することができ、対数尤度比の閾値を下げて信頼性のある共起情報を獲得することができると考える。

表 4 の例番号 4・5・6 は、対象発話対の疑問判定についての素性で対話関係が成立していないと判別された負例データの例である。どの対話内容とも疑問への応答をせずに疑問を問いかけているため、対話として成立していないと言える。しかし、表 6 のような疑問への応答を行った後に疑問を問いかける場合を考慮できておらず、判別の精度を悪くした一つの要因だと考える。

表 4 の例番号 7・8・9 は、内容語・未知語クラスタリングの素性でのみ対話関係が成立していると判別された正例データの例である。素性として付与される際に用いられた語の対は、例番号 7 では「わかる ⇔ 言える」、例番号 8 では「来る ⇔ ごろ」、例番号 9 では「ネット ⇔ ドコモ」「ネット ⇔ Android」「ネット ⇔ 使える」「の ⇔ する」であった。特に例番号 9 の例文に関しては未知語である「Android」と「ネット」が同じクラスタに属しているという特徴を用いている点から、期待した結果が得られた例と言える。しかし、実際にこのような例は少数であり、未知語の大半はクラスタの精度を下げる要因となっていた。また、未知語に限らず精度を下げる要因となっている語が存在しており(「の ⇔ する」など)、これらのデータを除外することで、クラスタリングの精度が上がり、対話判別の精度が向上すると考える。

5.5 追加実験

表 3 より、ベースラインの結果は負例データに関する適合率が高くなっており、これは負例データ内に形態素が照合しない発話対が多く存在したことを意味する。追加実験では、図 3 の負例データの作成方法に「ツイート文内の名詞、動詞、形容詞、未知語のいずれか 1 語と同一の語が返信文にも含まれている」という条件を加え、形態素の照合がある負例データに対する各素性の精度を評価した。表 7 に使用した負例データの例を示す。

表 7 より、発話対間に同一語が存在しているが対話として成り立っていないことがわかる。このようなデータに対して、形態素の照合のみの素性(ベースライン手法)では対応す

表 8 追加実験についての判別結果

使用した素性情報	正例データに関する適合率	負例データに関する適合率	全体の適合率
ベースライン	0.380	0.0	0.190
(A)	0.442	0.290	0.366
(B)	0.978	0.024	0.501
(C)	0.099	0.915	0.507

ることができない。追加実験では、表 7 のような負例データ 1 万件と正例データ 1 万件に対して、3 種類の素性を用いて再度判別実験を行った。なお、各素性を単体で用いた場合の適合率のみを算出しており、素性の有無で対話関係の判別を行った。結果を表 8 に示す。表 8 内の素性情報を表す記号は表 3 に対応している。

表 8 より、形態素の照合では対応できないことが確認できる。また、対象発話対の疑問判定(B)の正例データに関する適合率、内容語・未知語クラスタリング(C)の負例データに関する適合率が極端に高くなっているのは、素性の付与が行われていないことが要因となっているだけで、素性の効果によるものではなかった。その結果、語の共起情報(A)を素性とした場合のみ、ベースラインを上回っていることになるが、発話対間に同一語が存在する発話に対してはほぼ対応できていない結果となった。

6 おわりに

本研究では、別々の人物によって作成された二つの発話対によって構成される文字対話について、対話として成立しているか否かの可否判別を自動的に行う方法を提案した。判別を行う際に使用する素性として語の共起情報、対象発話対の疑問判定、内容語・未知語クラスタリングを用いた。判別実験の結果、各素性のみを用いた場合の適合率、素性を組み合わせた場合の適合率共にベースラインの結果を上回ることができなかった。しかし、各素性によって対話関係の可否判別が適切に行われている発話対に違いがあったことから、今回使用した 3 種類の素性は対話関係の可否判別の素性として効果はあると考える。

また追加実験により、形態素の照合では対応できない発話対に対して、提案した素性が対応できるかを検証した。結果より、語の共起情報を素性とした場合にベースラインを上回ることができたが、対応できているとは言えない結果であった。

今後の課題としては、前節で述べた各素性の問題点を改善し、対話関係の可否判別を高精度で行えるようにすることが挙げられる。また、実際に対話システムへの実装を行い、対話の精度が向上するかどうかについての調査を行う。

参考文献

- [1] 今井健太, ジェブカ ラファウ, 荒木健治. 複数の対話システムからの応答候補文を用いた最適応答文選択手法の性能評価. 情報処理学会研究報告, Vol.2010-FI-97 No.10, 2010.
- [2] 稲葉通将, 鳥海不二夫, 石井健一郎. 語の共起情報を用いた非タスク指向型対話エージェントの評価. 情報処理学会研究報告, Vol.2009-MPS-75 No.18, 2009.
- [3] 寺村秀夫, 佐久間まゆみ, 杉戸清樹, 半澤幹一. ケーススタディ 日本語の文章・談話 pp.140-151, 1990.
- [4] Ted Dunning. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, Vol.19, No.1, pp.61-74, 1993.
- [5] 平井尚樹, 稲葉通将, 鳥海不二夫, 石井健一郎. 対話エージェントのための文脈を考慮した統計的発話選択. 合同エージェントワークショップ&シンポジウム 2010 論文集, 2010.