

非タスク指向型対話システムにおける 話題展開へのゼロ照応解析導入の提案

初田 玲音[†]
奈良工業高等専門学校[†]

市川 嘉裕[‡]
奈良工業高等専門学校[‡]

1 はじめに

近年、非タスク指向型対話システムの研究が盛んである。非タスク指向型の対話システムでは、対話すること自体が目的となることが多いため、タスク指向型対話システム以上に自然な応答をすることが求められる。しかし、非タスク指向の対話では一般的にドメインが制限されないため、システムがユーザの発話を正確に理解することや、その上で対話の流れに沿った発話をするなど様々な点で課題がある。特に、人同士の雑談では自然に話題が展開されていくことを考えると、非タスク指向型対話システムにおいても自然に話題展開する能力が重要であるといえる。

話題展開を対象とした従来の研究では、例えば語の意外度に基づくもの [1] やランダム性を考慮したもの [2] が提案されているが、いずれも直前のユーザ発話から話題語を推定して発話を生成している。しかし、一般的に日本語、特に雑談等に用いられる口語文では主語や目的語等の主要な情報が省略されることが多く、対話中の話題語が省略されることが考えられる。従って直前のユーザ発話のみを考慮した発話生成は、連続する対話の流れを考慮した自然な発話になりにくいと考えられる。

そこで本研究では、今村ら [3] の最大エントロピーモデルを用いたゼロ代名詞照応解析付き述語項構造解析を利用して、二つ以上連続する会話文を考慮してユーザの発話から欠落した情報を補完してから話題展開させる手法を提案する。本稿では、対話実験を実施する準備段階の予備実験について報告する。

2 対象となるタスク

本研究では、入出力モダリティがテキストのみのテキスト対話システムにおける、二つ以上連続する対話文の後に続けて話題展開するタスクを対象とする。図 1 にタスクが生じる場面(対話シーン)の例を示す。

図 1 は、上からユーザとシステムが交互に発話する場面を表している。システムは、この続きとして、対話の流れに沿い、かつ自然に話題展開する発話生成が求められる。例えば、図 1 では次にシステムが“銀閣寺は銀色ではありませんからね”と発話すると、最後のユーザ発話中の“金色”だけでなく、対話中の“金閣寺”の話題に沿った自然な発話をしたと言える。しかし、例えばシステムが“銀色も綺麗ですね”と発話すると、最後のユーザ発話中の“金色”、“綺麗”は考慮されているが、対話中の“金閣寺”の要素が考慮されていない。このような発話は、間違いではないが、自然な発話ではない。

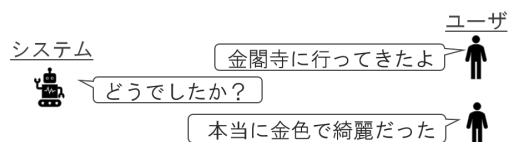


図 1 対話シーンの例

本稿では、最後のユーザ発話より前にある話題を考慮した発話生成を対象とする。

3 提案手法

図 2 に本研究で提案するシステムの概略図を示す。

提案システムでは、まず直前のユーザ発話の述語項に対してゼロ代名詞照応解析を行う。これにより、対話の流れを考慮しつつ直前のユーザの発話で省略された語を補完する。次に、情報が補完された発話を基に、別の手法を利用して話題展開を行う。以下で、提案システムを構成する 2 つの機構について詳述する。

3.1 ゼロ代名詞照応解析部

ゼロ代名詞照応解析部のベースとなる述語項構造解析には、今村ら [3] の手法を採用した。この手法は文内に存在する項、文間の項、外界照応を同時に解析できる特徴に加え、特に対話を対象としている。

この手法は、述語文節、項の候補名詞句、両者の関係を素性化し、必須格のガ、ヲ、二格独立に、候補から最も各格にふさわしい名詞句を選択器で選択する。選択器のモデルは、最大エントロピー分類に基づく。選択器は、発話内の述語ごとに、候補名詞句集合から、尤もらしい名詞句を選択する。

提案システムのゼロ照応解析部では、ベースとなる手法に対して以下の点で異なる。

- 素性が異なる。ベースの手法では Speaker 素性と Syn 素性が存在するが、本稿では存在しない。これは、NAIST テキストコーパス [4] のみを用いたからである。従って、オリジナルと比べて補完の精度が低下する可能性がある。
- 必須格辞書(述語に対して、格が必須なら 1 を返す辞書)から、述語に対してヲ格補完が必須でない場合は補完を行わない。これは、今村らの結果より、述語に項が必要な場合と述語に項が必要でない場合を合わせて 90% 以上の精度があるので、ヲ格は必須格辞書から補完すべきか否か判断出来ると考えたためである。

3.2 話題展開部

話題展開部には、伊藤ら [1] の手法を採用した。これは、All About(<http://allabout.co.jp/>)の各記事とそれぞれが属するカテゴリ情報を利用して、対話中の話題に対して意外性を持った記事を正規化した類似度で判定し、意外性を持った発話を記事から

Introduction of Zero-Anaphora Resolution to Topic Extension Method for Non-task-oriented Dialogue System

[†] Leon Hatsuda, National Institute Of Technology, Nara College

[‡] Yoshihiro Ichikawa, National Institute Of Technology, Nara College

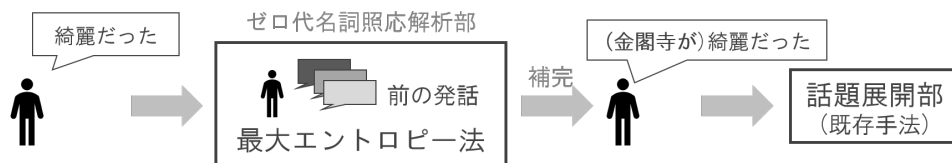


図2 提案システムの概略図

表1 各必須格での補完成功率

格	ガ	ヲ	ニ
補完成功率	43.5%	95.7%	86.4%

表2 補完が成功した例

ユーザ	金閣寺に行ってきたよ		
システム	どうでしたか？		
ユーザ	本当に綺麗だった。		
述語	ガ格	ヲ格	ニ格
綺麗	金閣寺	補完なし	Null

生成する手法である。

対話システムが発話する文の生成を次の5段階で行う。

1. ユーザ発話からの特徴語抽出
2. ユーザ発話と記事との関連度算出
3. ユーザ発話と記事カテゴリとの関連度算出
4. 記事の意外性スコア算出
5. ユーザ発話と関連度が高く、関連度が小さいカテゴリに属する記事を選び発話を生成する。

4 予備実験

4.1 実験内容

提案手法の対話実験による評価に先駆けて、予備実験として、実験者が用意した対話シーンに対してゼロ代名詞照応解析を行い、補完の成功率を調べた。対話シーンとは、図1に示したような人とシステムが交互に発話した文章の組のことを指す。

予備実験では図1のような対話シーンを8つ用意した。本研究では表層格のガ格、ヲ格、ニ格に対して解析を行うため、それぞれの補完が期待されるシーンを2つずつ、それに加えて補完に意図のない対話シーンを2つ用意している。

なお、NAIST テキストコーパス [4] を用いて各必須格のモデルを学習させる際、環境の都合によりコーパス量を無作為に減らし、全体の約28.6%のコーパス量で実験を行った。

4.2 実験結果

各必須格での補完成功率を表1に示す。補完成功率とは、対話中で補完対象となった全ての述語に対し、補完に成功した述語の割合である。結果より、ヲ格、ニ格の補完成功率が高く、ガ格の補完に成功した割合が他の格に比べて低いことが得られた。しかし、ヲ格は必須格辞書から述語に対して補完するか判断した結果であり、必須格辞書を使わずに解析した場合、補完成功率は13.6% (22述語中3個成功) であった。同様に、ニ格に対しても、解析された結果、全ての述語に特殊名詞句 Null が出力されており、ニ格補完が必要であった述語のみを対象に分析を行うと補完成功率は0% (3述語中0個成功) であった。

補完が成功した例を表2に示す。表の例では、最後のユーザ発話の述語「綺麗だ」に対して解析が行われた。ガ格に金閣寺が補完され、ヲ格は補完なし、ニ格はNullであった。ここで、ヲ格の補完なしとは、ヲ格必須格辞書を引いた結果、述語に対してヲ格が必要ないと判断されたため補完されなかったことを意味する。

一方、出力がNullの場合、候補名詞句から確率を計算した結果、尤もらしい名詞句がNullであったことを意味する。解析結果を文章に反映すると「本当に [金閣寺が] 綺麗だった」となり、対話の流れに沿った補完に成功したといえる。

5 まとめ

本稿では、ゼロ代名詞照応解析付き述語項構造解析を利用し、二つ以上連続する会話文を考慮してユーザの発話から欠落した情報を補完してから話題展開させる手法を提案した。補完の成功率を調べる予備実験の結果より、ヲ格、ニ格の補完成功率が高く、ガ格の補完に成功した割合が他の格に比べて低いことを確認した。

今後の課題として、コーパスの改善がある。本稿でゼロ代名詞照応解析部のモデルの学習に使用したコーパスは本来のコーパスから無作為に減らしたものであり、完全ではない。また、実験結果より、ニ格の補完結果は本稿で用意した対話タスクでは高い成功率であったが、補完が必要な述語における成功率が低いため、素性テンプレート等を再検討する必要があると考える。加えて、用意した対話シーンの対話をより長くした場合の補完成功率の変化についても検討したい。

参考文献

- [1] 伊藤直之, 西川侑吾, 大野和久: 語の意外度に基づき話題展開する非タスク指向型対話システム. 人工知能学会全国大会論文集, Vol. 29, pp.1-4.
- [2] 三上佳孝, 萩原将文: 対話におけるランダム性を考慮した話題展開手法. 日本感性工学会論文誌, Vol.17, No.3, pp.365-373.
- [3] 今村賢治, 東中竜一郎, 泉朋子: 対話解析のためのゼロ代名詞照応解析付き述語項構造解析. 自然言語処理, Vol.22, No.1, pp.3-26.
- [4] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治: 述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から. 自然言語処理, Vol.17, No.2, pp.25-50, April 2010.