

8B-6

対話文脈における話題の一貫性に関する研究*

上野 正巳 唐澤 博†
山梨大学‡

1 はじめに

我々は、マン-マシン・インタフェースの一環として自然言語処理について研究しているが、マン-マシン間の円滑なコミュニケーションを行なうために、文脈について考える必要がある。本研究は対話を対象とし、対話全体に関する一貫性(Coherence)を対話中の各文から得た話題の展開から捉える。

本研究で対象とする対話は、グライスの会話の公準を満たしているものとする。その上で、対話中の各文において話題となっている物事を捉え、その話題間の意味的、概念的距離を指標として、対話中の話題の展開を定量的に捉え、一貫性の指標を得ることを目標としている。

言い換えると、協調的対話に関するグライスの公準に定量的指標を与える工学的試みであるといえる。

一貫性を捉えることにより、

- 対話理解時に、概念上の多様性による曖昧性を解消できる。
- 発話生成時に、相手が必要としている情報を素早く提供することができる。

といった事が可能になる。

2 話題の一貫性の評価手順

本研究では、対話中に現れる各文から話題を抽出し、話題の変化から一貫性を捉える。このため、図1の手順で処理を行ない、一貫性を評価する。

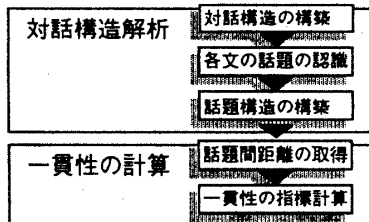


図1: 一貫性の評価手順

対話構造解析

- 対話構造の構築** 既存の方法を利用し対話の係り受けなどから構造を構築する。
- 各文の話題の認識** 対話構造を利用し各文中で話題となっているものを取得する。
- 話題構造の構築** 各文の話題と対話構造を利用し話題間の構造を得る。

一貫性の計算

- 話題間距離の取得** シソーラスを利用して話題間の距離を求める。
- 一貫性の指標計算** 話題間距離を利用し対話の一貫性の指標となる値を計算する。

*Coherence in Dialogue Context.

†Masami UENO, Hiroshi KARASAWA

‡Yamanashi Univ.

3 対話文からの話題の抽出

従来、話題を処理対象とした研究では文の主題と話題を混然して扱っているものが多かった。そこで、それぞれを次のように定義し扱うことにした。

主題 提題の助詞「は」により文中に導入され、話題の基点となるもの。

話題 対話者間で共有される関心の対象。

さらに、対話中に現れる文は、全て話題を持っているものとし、各文から話題を取得する。また、主題を単純に話題として扱わず、次のようなルールで話題を決定した。

| 文のタイプ | 話題となる格 |
|--------------------|--------|
| 他動詞の文 | 対象格 |
| 自動詞文、形容(動)詞文 | 主格 |
| 上記以外の「～が<名詞>だ」形式の文 | <名詞> |

また、対話は「質問」-「応答」のような係り受けをはじめとする構造(対話構造)を持っており、例えば、「応答」文が与えられるまで「質問」文の話題が決定できない場合などがある。このため、本研究では話題取得のために、この対話構造を利用する。対話構造に関しては既存の研究の方法を利用した。

また、本研究では話題間の関係を捉えるために、既に多くの研究者によってある程度研究が進んでいるシソーラスにおける関係の分類を利用して話題間に距離の概念を導入する。そして、話題間の距離を用いて一貫性を捉える。

4 対話中の話題の構造化

対話全体の一貫性を捉えるために次の仮定を行なう。

仮定 ある時点までの対話を見た場合、その時点までの対話全体にも話題が存在する。

その上で、各文から取り出した話題は図2のように階層的構造を成しながら、文全体の話題につながっていくものと定義する。

この対話全体の話題は、文脈全体を支配している話題とも捉えることができる。

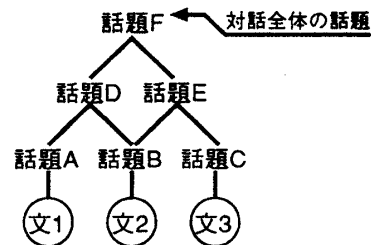


図2: 話題の構造

対話全体の話題をこのように構造化し、最上位の話題から個々の文の話題への距離を考える。この距離は、文脈全体を支配している話題と、個々の文が担っている話題との概念的な距離を表している。そこでこの距離を測るために、話題間に距離を設定する。

5 話題間の距離

話題間に距離を導入するために、まず話題間の関係を捉える。これには、既存のシソーラスの研究を利用する。

5.1 シソーラス上の関係分類

文献[6]によると、シソーラス中の用語間の意味的な関係には、等価、階層、連想という三つの基本的な関係がある。

- 等価関係** 複数の用語が同じ概念を表しているときみなされる場合の、優先語および非優先語の間の関係。
- 階層関係** 上位および下位のレベルを示す語。上位語は、類あるいは全体を表し、下位語は、その一要素、あるいは一部分を表す。
- 連想関係** 階層的ではなく概念的に密接に関連し、かつ等価集合に入らない用語間の関係。

これらの関係を利用するわけであるが、「連想関係」はシソーラス作成者の主観が入りやすく適用分野への依存度が大きいため、本研究では除外した。

5.2 話題間の距離の設定

対話の各文から得られた話題間に概念的距離を設定するために、シソーラスの構造を利用する。今回は、文献[7]で一般的な方法として紹介されている、シソーラス中での2単語間のリンクの数を距離として利用する方法を採用する。2単語間にリンクがない場合は距離は ∞ となる。よって、2つの話題 x, y 間の距離 $D(x, y)$ は次式で定義される。

$$D(x, y) = \begin{cases} x, y \text{ 間のリンク数} & (x, y \text{ 間にリンクが張れる場合}) \\ \infty & (x, y \text{ 間にリンクがない場合}) \end{cases}$$

なお、2単語間に複数のリンクが存在する場合は、最短のものを利用する。

6 話題の一貫性の指標の算出

対話を通じて動的に話題を捉えるために、発話がなされるたびに対話全体としての話題を捉える。具体的には話題ノードをパスカルの三角形に積み上げていく。

話題 T_x, T_y について、 T_x が T_y の上位語である場合を

$$T_x \supset T_y$$

であらわす。このとき、関係 \supset には推移率が成り立つ。また、隣接する2ノードの話題 ${}_{n-1}T_m, {}_{n-1}T_{m+1}$ から上位ノードの話題 ${}_nT_m$ を以下のように定義する。

$${}_nT_m = \begin{cases} {}_{n-1}T_m & ({}_{n-1}T_m \supseteq {}_{n-1}T_{m+1} \text{ または } {}_{n-1}T_{m+1} = \phi) \\ {}_{n-1}T_{m+1} & ({}_{n-1}T_m \subsetneq {}_{n-1}T_{m+1} \text{ または } {}_{n-1}T_m = \phi) \\ T_c & (T_c \supseteq {}_{n-1}T_m, T_c \supseteq {}_{n-1}T_{m+1}) \\ \phi & ({}_{n-1}T_m \text{ から } {}_{n-1}T_{m+1} \text{ へのパスがない}) \end{cases}$$

Step.1 対話構造を構築する。

Step.2 k 番目の発話文が提示する(または継続する)話題を T_k とする。話題提示のない文は対話対をなす文の話題を引き継ぐ。

Step.3 話題間距離 $D({}_nT_m, {}_nT_k)$ ($1 \leq k \leq m$)を計算する。

Step.4 D_k の期待値 E と分散 V を求める。

$$E = \sum_{k=1}^n D_k/n$$

$$V = \sum_{k=1}^n (D_k - E)^2/n$$

E は対話文脈全体を支配している話題からの個々の発話文の話題の平均距離を表している。また V は個々の発話文の話題の変動を表している。 E および V は、話題の一貫性を示す指標として利用できると考えられ、これらを組み合わせることで以下のような対話特徴を捉えることができる。

- (1) E が大で V が小の場合 話題の変化が大きいが関連している話題同士のふくらみが乏しい
- (2) E が小で V が大の場合 話題の変化が小さくそのふくらみが豊かである
- (3) E が小で V が小の場合 話題の変化が小さいがそのふくらみが乏しい
- (4) E が大で V が大の場合 話題の変化が大きく関連している話題同士のふくらみも大きい

ここで「話題のふくらみ」とは、概念シソーラスのリンクを上昇下降する程度に対応する。

対話制御の立場からは(2),(3)が重要である。例えば、話題について指向性の強い対話に制御したい場合には(2)が満たされるような発話内容を優先する。発話すべき内容が複数競合を起こしているような場合に、発話前に予め E, V を計算して優先付けするという方法によってその競合を、円滑な対話を生成するという立場で解消することが期待できる。

7 おわりに

対話中の各文から取り出した話題間の距離を測ることで、一貫性の指標と考えられる値を得る方法を示した。この値は対話制御、発話内容の競合解消などに利用できる。

また、学部学生に協力してもらい、端末間対話により20組のべ369文からなる会話データを収集し本研究手法を適用したところ、実際のデータに対しても有効であることが確認できた。

今後の課題としては、

- 話題間の距離の設定の緻密化
- より多くの対話例によるアルゴリズムの検証
- コンピュータ上でのインプリメンテーションによる検証

などを、考えている。

参考文献

- [1] 工藤育男: 文と文の結束性をとらえるための知識, 自然言語処理 76-3(1990)
- [2] 竹下敦: 対話のインタラクション構造を用いた話題の認識: 情処研究報告書 92-NL-87, pp75-82(1992)
- [3] 山梨正明: 文脈構造に基づく言語理解の研究, 文部省科研資料, 言語情報処理の高度化研究報告 7, pp87-96(1991)
- [4] 垣内隆志, 上原邦昭, 三上理, 豊田順一: 文脈のボトムアップ処理とトップダウン処理の融合について, 第31回情処全大 7N-1, pp1165-1166(1985)
- [5] 内田ユリ子, 石崎俊, 井佐原均: テキストにおける首尾一貫性と文脈表現構造, 第38回情処全大 4D-2, pp245-246(1989)
- [6] Jean Aitchison, Alan Gilchrist, THESAURUS CONSTRUCTION 2nd Edition: (内藤, 中倉, 影浦訳: シソーラス構築法, 丸善(1989))
- [7] 長尾真: 言語工学, 昭晃堂(1982)