

6M-3 自然言語対話システムにおける発話話題の自動生成*

梶野 博仁, 唐澤 博†

山梨大学 工学部‡

E-mail: {kajino, karasawa}@jewel.yamanashi.ac.jp

1 はじめに

現在の対話の話題を掘り下げたり、転換することをシステム側から行うことができるようにする。ユーザの入力文よりキー概念を抽出し、また文脈より観点到に相応する概念を抽出する。キー概念より概念ベースを造り出し、関連度計算で次の発話話題を決定する。

2 概要

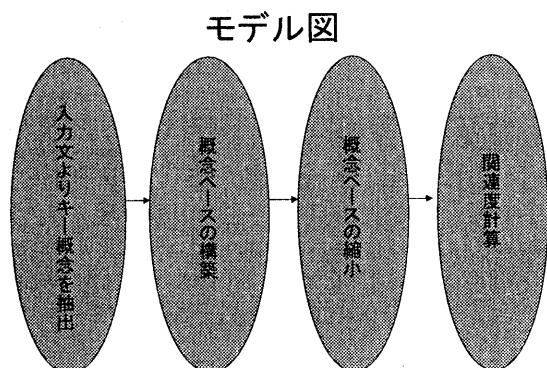


図 1: モデル図

3 概念の定義

本研究での概念の単位は、単語とする。ある概念 A の意味特徴を表す単語の集合をその概念の一次属性と呼ぶ。さらに、一次属性の意味特徴を表す単語の集合を二次属性と呼ぶ。よって概念は、n 次属性まで表すことができる [1].

$$\text{概念} A = a_1^1, a_2^1, \dots, a_i^1, \dots, a_N^1$$

$$a_i^1 = a_1^2, a_2^2, \dots, a_i^2, \dots, a_M^2$$

*Automatic topic generation for natural language dialog system

†Hirohito Kajino, Hiroshi Karasawa

‡Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

4 発話話題の決定手順

4.1 キー概念の抽出

ユーザの入力文からキー概念という「文の意味を解くために最も重要な概念」を抽出する。また、話題とは、「対話者間で共有される関心の対象」と定義する [2]。入力文は、それから取り出される話題について話しているため、話題は文の意味を解くために最も重要な概念であると考えられる。これに従い、キー概念の抽出は話題の抽出方法 [2] により行うことができる。

(例文) 私は、馬を持っている。

表 1: 文のタイプと話題となる格

| 文のタイプ | 話題となる格 |
|-------------------------|---------|
| 他動詞の文 | 述語の直前の格 |
| 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) |

話題は「馬」が取り出される。主題は「私」であるが、キー概念を考えると「馬」が取り出される。入力文は、それから取り出される話題について話しているため、話題は文の意味を解くために最も重要な概念であるといえる。よって、キー概念の抽出方法は、話題の抽出方法によって行うことが可能である。

4.2 概念ベースの構築

概念ベースの構築には、電子化国語辞書を用いる。辞書の見出し部分にある単語を概念とし、その語義文中の自立語を形態素解析で抽出し、その概念の一次属性とする (図 2) [3].

(例) 「馬(うま)」家畜の一。たてがみが長い。

4.3 概念ベースの縮小化

概念ベースには、辞書の語義文の単語すべてが含まれているため、今後概念ベースを計算に用いる場合に無意味な計算が増え、計算量が現実的なものではなくなってしまうことが考えられる。そこで組み合わせ最適化の

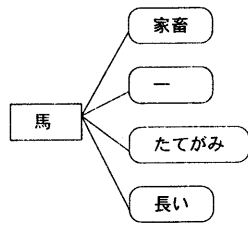


図 2: 概念ベース

手法としてマーカパッシングを用いて、概念ベースのサイズを縮小する。

4.3.1 マーカパッシングの適用

1. 現在の入力文と前5文のキー概念を抽出する。
2. まず、辞書の見出し語を単語ノード、その各語義について語義ノードを作り、単語ノードからはその単語の語義ノードへ、語義ノードからはその語義文中の各単語の単語ノードに活性リンクを張り、一つの見出し語の複数の語義ノードには抑制リンクを張り、辞書を意味ネットワーク化する。
3. 現在のキー概念と前5文のキー概念をこの意味ネットワークへ入力する。
4. 入力されたキー概念の対応する単語ノードにある活性値を与える。
5. 以降はネットワーク上で活性値の伝搬を繰り返す。活性値を活性リンクの張られた語義ノードへ $m\%$ を伝搬する。さらに、語義ノードから活性リンクの張られた単語ノードへ $m\%$ を伝搬する。抑制リンクでは、マイナスの値を伝搬する。活性値の分布が定常的に落ち着いたら、伝搬を終了する。
6. 各入力単語に対し、最も活性値の高い語義ノードがその単語の語義となる。

4.4 関連度計算

現在の入力文のキー概念とその一次属性の関連度を前5文のキー概念を踏まえて計算する。関連度の最も大きい前5文のキー概念を観点とし、その観点と最も一致度の大きい現在の入力文のキー概念の属性を次の発話話題とする。観点とは、「文脈に共通した概念」と定義する。

4.4.1 一致度 Match

二つの m 次属性の一致度は、それぞれの $m+1$ 次属性の一致単語数 p を 0 から 1 の範囲に正規化したものとする [1].

$$a_i^m = a_1^{m+1}, a_2^{m+1}, \dots, a_i^{m+1}, \dots, a_L^{m+1}$$

$$b_j^m = b_1^{m+1}, b_2^{m+1}, \dots, b_j^{m+1}, \dots, b_M^{m+1}$$

$$Match(a_i^m, b_j^m) = \frac{p(\frac{1}{L} + \frac{1}{M})}{2}$$

4.4.2 関連度 Sim

概念 A と概念 B との観点概念 C における類似度は、A と B の一致度と観点概念 C と A の一致度、観点概念 C と B の一致度の三者の相乗平均の平均とする。

$$A = a_1^1, a_2^1, a_3^1, a_4^1, a_5^1$$

$$B = b_1^1, b_2^1, b_3^1, b_4^1$$

$$C = c_1^1, c_2^1, c_3^1, c_4^1, c_5^1, c_6^1$$

$$Sim(A, B|C) =$$

$$\frac{\sum_{i=1}^N \sqrt[3]{Match(c_i^1, a_x i^1) \times Match(c_i^1, b_y i^1) \times Match(a_x i^1, b_y i^1)}}{N}$$

上式の Match の引数で使われている x, y は、A, B の一次属性を対応する観点概念 C の一次属性との一致度の合計が最大になるように並び替えたことを表す。また、N は観点概念 C の一次属性数を表す [1].

5 課題

- マーカパッシングにおいて、活性値の伝播率 $m\%$ を決定する。
- 関連度計算における一致度の計算で、最大になるように並べ替える手法について決定する。
- より多くの例を使ってモデルを評価する。
- アルゴリズムをコンピュータに実装して評価する。

参考文献

- [1] 入江, 渡部, 河岡, 松澤: 知的メカニズムのための概念間の類似度定量化方式, 第 58 回情報処理学会講演論文集, 1999.
- [2] 上野, 唐澤: 対話文脈における話題の一貫性に関する研究, 情報処理学会第 46 回全国大会講演論文集, 8B-6, pp.201-202, 1993.
- [3] 笠原, 松澤, 石川: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌 vol.38, 1997.