

自然言語対話システムにおける発話話題の自動生成*

1M-01

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

本研究の目的は、現在の対話話題が尽きたときに次の話題を展開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ことである。図 1 に示すように、ユーザの入力文より文脈情報を抽出する。辞書より意味ネットワークを造り出し、マーカパッシングで次の発話話題を決定する。

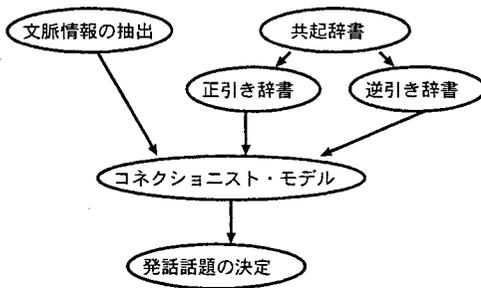


図 1: モデル図

2 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

3 発話話題の決定手順

3.1 文脈情報の抽出

本システムでは、文脈情報として対話文中に表れる全自立語を使用する。文脈として考慮する文は、前 5 文と仮定した。

3.2 共起辞書の作成

EDR 電子化辞書を用いて、共起辞書を以下の手順で作成する。

1. EDR 電子化辞書の見出し語と語義文を抽出する。
2. 抽出された語義文を LimaTK version 0.1.4 を使って形態素解析する。
3. 解析結果より自立語を抽出する。
4. 辞書の見出し語と語義文を表す自立語を 1 レコードにまとめる。

[共起辞書の一部]

相かまえて 決して

相構えて 決して

靉黓 年 とる せい 視力 低下 とき 使う 眼鏡

アイデンティティ 自己 認識

アイデンティティー 自己 認識

3.3 逆引き共起辞書の作成

共起辞書の語義からも活性値を伝搬できるようにするため、共起辞書から逆引き共起辞書を以下の手順で作成する。

1. 共起辞書を一行ずつ読み取る。
2. 見出し語以降の自立語を他行に存在するか探索する。
3. 存在した場合、その行の見出し語をファイルへ書き出す。
4. 1~3 を繰り返す。
5. 同一行を削除する。

[逆引き共起辞書の一部]

決して

構えて 構いて 相かまえて 夢更 相構えて 二度と 年

仮想年 春夏秋冬 春夏秋冬 年記 年紀 例年 例年

数え 世紀の 鬚鬚 老眼鏡 老醜を晒す 老醜を晒す 老女

湯殿始 湯殿始め 湯殿始 湯殿始め 老骨に鞭打つ 春秋 歳霜 春秋 星霜 日月 風霜 仮想年 例年 例年

3.4 話題の生成

抽出された概念と作成された共起辞書を用いてマーカパッシングというコネクションスト・モデルを行う。マーカパッシングとは、ネットワークに活性値を伝播させて単語ノードを発火させる手法である [2]。

3.4.1 共起辞書のネットワーク化

共起辞書の見出し語を表す自立語ノード、その語義数分の語義ノードを作成し、活性リンクで結ぶ。語義ノードはその語義を表す自立語ノードと活性リンクで結ぶ。また、語義ノード同士は抑制リンクで結ぶ。

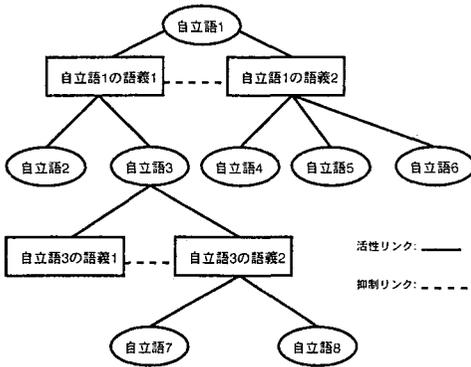


図 2: 意味ネットワーク

3.4.2 活性値の伝播

文脈から得られた全自立語を入力とする。対応する見出し語ノードに活性値 (= 1) を与え、活性リンクの張られた語義ノードへ 50% 伝播する。さらに、この語義ノードから自立語ノードへの 50% の伝播を繰り返していく。また、抑制リンクでは、活性値の 50% のマイナスの値を伝播する。活性値の分布が定常的に落ち着いたら (伝搬値が、0.001 以下)、伝播を終了する。

3.4.3 発話話題の決定

伝搬が終了したら、意味ネットワーク全体の自立語ノードのうち最も活性値の大きい自立語を次の発話話

題とする。ここで新たな話題を抽出したいため、マーカパッシングへの入力単語は最大活性値の選択からは、削除する。また、一般的な単語 (例: こと, もの) 等も選択から削除する。マーカパッシングの入力に前の文の情報を使うことにより、文脈を考慮した発話話題となる。

また、本研究の目的である話題の展開とは、意味ネットワークにおいて現在の話題である単語ノードとは、違ったノードが抽出されることと定義する。

3.4.4 伝搬の性質

共起辞書を引いて活性値の伝搬を行っていくと次第に値が一般語 (例: こと, もの) へ集まっていく傾向が見られる。これは国語辞書の性質によるもので、語義文には一般語が多く含まれているためである。逆に、逆引き共起辞書を引いて活性値の伝搬を行うと、値が特異な語へ集まっていく傾向が見られる。

3.4.5 1ステップごとの辞書切り換え

共起辞書 (正引き), 逆引き共起辞書のそれぞれの性質が反映されるように、まず入力単語に対して正引きを行い活性値を伝播して、正引き結果に対して逆引きを行い、また逆引き結果に対し正引きを行うといった具合に伝搬終了まで繰り返すアルゴリズムでマーカパッシングを行った。

しかし、下記の実行結果は一般語が活性値上位を占めたため、一般語のノードに強い活性制限を与える必要性が見い出された。

[実行結果]

- 値 = 986.945313 @ キー = こと [コト]POS:JLN1,JRN1
- 値 = 277.703125 @ キー = ある [アル]POS:JLR1,JRR1
- 値 = 191.460937 @ キー = もの [モノ]POS:JLN1,JRN1
- 値 = 122.894531 @ キー = 人 [ニン]POS:JLN1,JRN1
- 値 = 102.984375 @ キー = する [スル]POS:JLV1,JEE1
- 値 = 95.828125 @ キー = さま [サマ]POS:JLN5,JRN7
- 値 = 90.433594 @ キー = できる [デキル]POS:JLJ6,JRV1
- 値 = 83.949219 @ キー = 物事 [モノゴト]POS:JLN1,JRN1
- 値 = 59.933594 @ キー = 相手 [アイデ]POS:JLN1,JRN1
- 値 = 59.261719 @ キー = なる [ナール]POS:JLV1,JRV1

4 課題

- 一般語についての問題点
 - 正引きを行うと必ず活性値上位を占める一般語に活性制限を与える。

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] 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.