

隠れマルコフモデルによるトピックの遷移を捉えた談話構造解析

柴田 知秀 黒橋 禎夫
東京大学大学院情報理工学系研究科

{shibata,kuro}@kc.t.u-tokyo.ac.jp

1 はじめに

我々は料理教示発話を対象とし、映像の要約・検索を高度化するために、発話の談話構造を解析し、料理の作業構造を自動抽出している [5]。その談話構造解析において、一文ずつ逐次的に談話構造を求めているため、局所的な関係は捉えられるが、全体の作業の流れを考慮していないという問題がある。全体の流れとは、料理では例えば、「下ごしらえ」-「炒める」-「盛り付け」といったトピックの遷移のことである。そこで、一度談話構造を解析し、局所的な関係を捉えた後、全体のトピックの遷移を推定し、推定されたトピック列に基づいて談話構造解析の高度化を行なう。

トピックの遷移は隠れマルコフモデル (HMM) を用いてモデル化を行なう。HMM の隠れ状態がトピック (下ごしらえ、炒める、盛り付けなど) にあたり、トピックが遷移しながら作業に関する発話 (皮をむく、火をつける) がシンボルとして出力されるとする。本研究で扱う作業教示発話では、作業やコツ、雑談など、様々なタイプの発話が含まれるが、作業の説明が中心であり、それらがトピックを特徴付けると考える。

2 関連研究

Hearst らは、同一語の繰り返しによる語の結束性に基づいたテキストセグメンテーションの手法 Text Tiling を提案している [2]。この手法ではトピックの遷移を直接扱うのではなく、語の分布の偏りを利用してセグメンテーションを行なっている。

Blei らは、Aspect HMM を用いてテキストセグメンテーションを行なっている [3]。Barzilay らは、地震などの 5 つのドメインを対象として、隠れマルコフモデルを用いて生コーパスからトピックの遷移モデルを構築し、文の並び順の決定と要約の生成の 2 つのタスクに利用している [1]。トピック (震度の情報、被害の情報など) が遷移しながら、そのトピックが出力する言語モデルを元に文が生成されるというモデルをとっている。

表 1: 用言格フレームの例

用言	格	用例
切る (1)	ガ	【主体】
	ヲ ニ	豚肉、大根、こんにゃく、… 正方形、楕形、三角形、…
切る (2)	ガ	【主体】
	ヲ ノ	水気、水分、汁気、… なす、豆腐、肉、…
入れる (1)	ガ	【主体】
	ヲ ニ	塩、油、野菜、… 鍋、ボール、容器、フライパン、…
入れる (2)	ガ	【主体】
	ヲ ニ	包丁… 魚、腹、付け根、…

3 現在の談話構造解析の概要

まず、現在の談話構造解析の手法について述べる。前処理として、以下の処理を行なう。

1. 入力文を JUMAN/KNP で形態素・構文・格解析する。
2. 自動構築した用言・名詞の格フレーム (表 1) を用いて省略解析を行なう。
3. 節末の表層パターンを用いて発話タイプを認識する (表 2)。

上記の処理の後に、省略解析結果・発話タイプ・語連鎖・表層ルールを統合し談話構造解析する。

談話構造のモデルとして、節を一つのノードとし、関係するノードがリンクされたグラフ構造を考える。談話構造の解析の様子を図 2 に示す。

節間の関係として以下に示すもの考える。

1. 一文内における節の係り受け関係
2. 任意の 2 つの節における同格の関係 (同一文内に限らない)
3. 主節 (「～と思う」などといった節を除いた、一文で最後の節) 間の関係

まず、一文内で係り受け関係にある節間の結束関係を決定する。付与する関係は、順接 (～て、(連用形))、

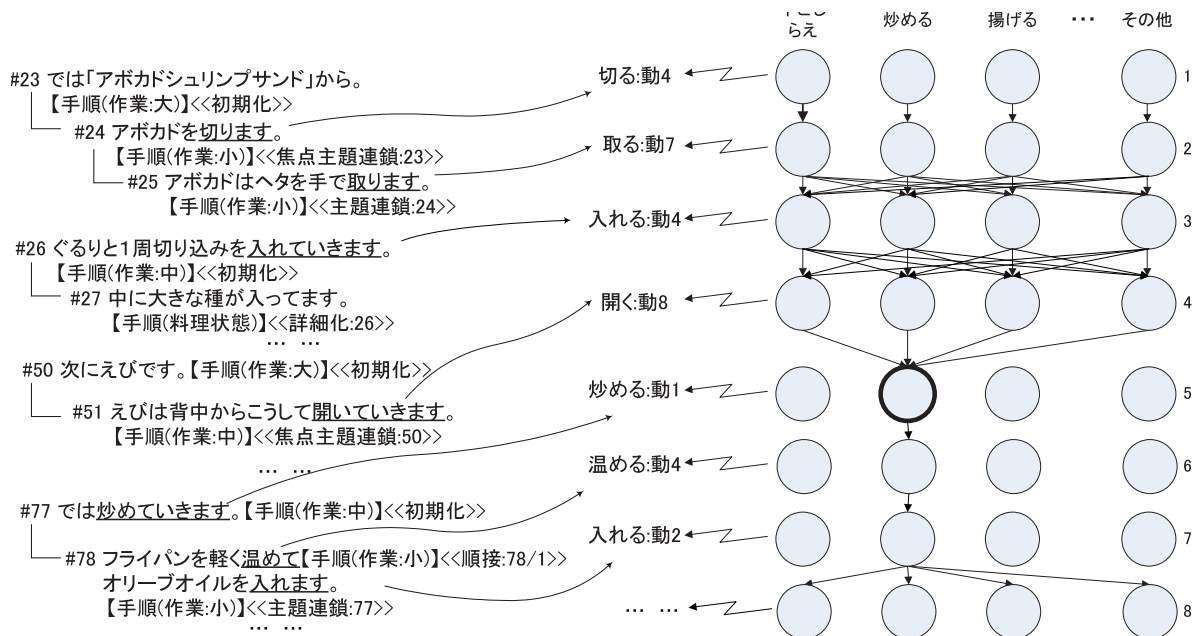


図 1: 談話構造と隠れマルコフモデルによるトピックの推定

表 2: 発話タイプの分類

手順	その他
[作業:大] ・さ、では、ステーキの材料にかかります。 [作業:中] ・強火で油を温めましょう。 ・じゃあ炒めていきましょう。 [作業:小] ・お鍋にお水を入れます。 [料理状態] ・ニンジンの水分がなくなりました。	[留意事項] ・最初に肉をパラパラに炒める事がポイントです。 [代替可] ・もし半個ぐらいでしたら、手で搾って頂いても結構です。 [食品・道具提示] ・材料は、牛ひき肉、百五十グラムです。 [雑談] ・暑くなってきましたね。 [効果] ・そうすると最初の方はアクが出てきます。

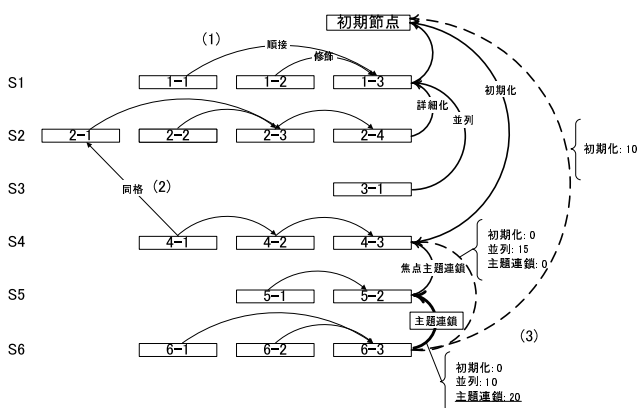


図 2: 談話構造のモデル

並列 (KNP で並列構造とされたもの)、理由 (~から、~ので)、条件 (~と、~たら)、修飾 (~ずに) などであり、括弧内に示すような表層パターンにより決定される。

次に、任意の 2 つの節において、用言の原形とその格要素が一致するかどうかをチェックし、同格の関係

にあるものを検出する。

最後に、主節間の関係を求める。談話構造の初期状態として初期節点を考え、初期節点に接続することは、その発話から新しいトピックが始まることを意味し、この時の関係を“初期化”とする。初期化以外の主節間の結束関係としては、並列、対比、主題連鎖、焦点主題連鎖、詳細化、例示、質問応答、同格などの関係を考える。主節間の関係は、種々の表層の手がかりをもとに、各入力文に対して、関係をもつ以前の発話(接続文)とその間の結束関係を逐次的に求める [6]。新しいトピックが導入された後に古くなったトピックに接続することはないという仮定をおき、入力文は談話構造の一番最後の子供の発話にのみ接続可能と考える。そして、さまざまな接続可能文との間のさまざまな結束関係を考慮し、最終的に最も高い確信度を得た関係を採用する。

確信度は表 3 に示すようなルールにより決定される。表 3 において、適用範囲とはどれだけ離れた発話

表 3: 談話構造解析のルール

結束関係	スコア	適用範囲	接続可能文パターン	入力文パターン
初期化	50	*	~	[手がかり表現]
並列	5	1	~	そして~
並列	40	*	[並列]	~さらに~
対比	30	1	~	むしろ~
詳細化	30	1	~	すなわち~
詳細化	15	1	~	<留意事項>
焦点主題連鎖	25	1	<食品提示>	<作業:小>
焦点主題連鎖/ 主題連鎖	X	*		[語連鎖]
同格	20 × (類似度)	*	[主節の 2 用言の類似度]	
主題連鎖	500 × (共起度)	*	[主節の 2 用言の共起度]	

との関係まで考えるか、接続可能文パターン、入力文パターンは、それぞれに対する表層表現、節間の結束関係 ([] で括られたもの)、発話タイプ (<>で括られたもの) などのパターンである。手がかり表現は「では」「次は」「そうしましたら」といったものを考え、用言の類義/共起関係はコーパスから自動獲得した [5]。ルールが一致した場合には、結束関係欄の関係に対して、スコア欄の点数が与えられる。

このような解析の結果、図 1 の左のような談話構造が得られる。

しかし、現在の自然言語処理技術では、話し言葉を正確に解析するのは難しく、省略の解析が正しく行なえずに語連鎖を捉えられないことや、用言の類義/共起関係を捉えられないことのため、話のまとまりを捉えられずに、「初期化」されてしまうことがある。例えば、図 1 の例の 26 文目では、入れるの二格にアボカドを補えないことにより、語連鎖を捉えることができず、26 文目で初期化されてしまっている。

そこで、図 1 に示すように、まず、発話タイプが作業である節の用言の格フレームを抽出し、それぞれの格フレームがどのトピックから出力され、トピックが遷移するかを隠れマルコフモデルで推定する。全体のトピックの遷移を推定した結果、例えば、23 文目から 25 文目までのトピックが下ごしらえであり、26 文目から 27 文目のトピックも下ごしらえであることが推定できれば、26 文目では初期化せずに、25 文目に接続するように談話構造を改善するといった処理を行なう。

4 節で、隠れマルコフモデルを用いてトピックを推定する方法について述べ、5 節で推定されたトピックを用いて、談話構造解析を高度化する方法について述べる。

4 隠れマルコフモデルによるトピックの推定

4.1 隠れマルコフモデルの定式化

隠れマルコフモデルを用いてトピックの遷移のモデル化を行なう。HMM は以下のパラメータを持つ。

- 隠れ状態 s_i : トピックにあたる。本研究では以下の 8 種類 ($N = 8$) を考える。

下ごしらえ、蒸す、ゆでる、揚げる、煮る、炒める、盛り付け、その他

- 初期状態確率 π_i
- 状態遷移確率 a_{ij} : 状態 i から状態 j への遷移確率であり、トピックの遷移確率にあたる。
- 出力シンボル確率 $b_j(k)$: 状態 s_j から格フレーム cf_k が出力される確率。

4.2 パラメータの推定

3 節で述べた談話構造解析を行ない、談話構造木から、発話タイプが、<作業:大>、<作業:中>、<作業:小>である節の格フレームを抽出する。ただし、以下のものは除外する。

- 同格であると判定された格フレーム
繰り返しの発話であるので抽出しない。
- 条件節
条件節は、前のトピックを表すことが多いので抽出しない。
例) 切りましたら 炒めていきます。

Baum-Welch アルゴリズムを用い、初期状態確率 π_i 、状態遷移確率 a_{ij} 、出力シンボル確率 $b_j(k)$ を推定する。ただし、以下の制約を設ける。

- 同じ談話構造木内の格フレームは同一トピックをとる。例えば、図 1 の例では、時刻 $T = 1$ と $T = 2$ 、 $T = 5, 6, 7$ は同じトピックを遷移するようにする。
- 談話構造木の最初の発話のタイプが<作業:大/中>で、用言の原形がトピックと一致する場合、その木のトピックを決定する。例えば、図 1 の時刻 $T = 5$ では「炒めていきましょう」という<作業:中>の発話があるので、トピックを「炒める」とする。その他に、「~の下ごしらえです」「~の準備です」にマッチする場合はトピックを「下ごしらえ」とするなどのルールを用意した。

表 4: トピックの推定の実験結果

提案手法	ベースライン
72 / 110 (65.5%)	43 / 110 (39.1%)

また、推定されたトピックは物体モデルを学習する際に利用する [4]。物体モデルを学習するには、その物体が大写しになっている画像を大量に収集する必要がある。その際に、トピックが「炒める」などの部分からは収集せず、「下ごしらえ」であるところから画像を収集することにより、画像の収集精度を向上させることができる。

5 談話構造解析の高度化

得られたモデルを用いて格フレームのトピックを推定し、その結果をもとに談話構造解析の高度化を行なう。

隣接する格フレームのトピックが連続する場合は、それらが接続するような談話構造解析のルールを追加することにより、談話構造を修正する。この処理により、直接関係を捉えられなかったものが、トピックが同一という観点を介して関係性を捉えることができる。

6 実験

NHKの「きょうの料理」のクローズドキャプションを用いて実験を行なった。約2年分のデータ202日分を用いた。1番組は約200発話からなり、1発話の平均文字数は約20文字である。

5番組を対象としてトピックの推定と談話構造解析を用いたセグメンテーションの実験を行なった。トピックの推定は談話構造木単位で評価し、ベースラインは、トピックが決定されているところから、 $p(s_i|cf_k)$ を計算し、それを基に格フレームのトピックを推定する手法である。また談話構造解析の評価では、各文がどの文とどの関係で結束しているかを評価するのではなく、関係が「初期化」である発話でトピックセグメンテーションし、その結果を評価した。

実験の結果、ベースラインよりも精度が向上していることが分かる。これはトピックの遷移を学習できていることを示している。また、誤りの多くはトピックが「炒める」「焼く」「煮る」などといった境界があいまいな所であり、4節で述べた、映像との対応付けの際のトピックが「下ごしらえ」の部分抽出するというタスクでは十分な精度を達成している。また、トピックセグメンテーションの実験では、従来の談話構造解析よりもF値が向上している。これは、全体のトピックの遷移を考慮しながら談話構造を解析することの有

表 5: セグメンテーションの実験結果

	Precision	Recall	F
提案手法	67.5%	77.1%	0.720
旧手法	43.1%	88.5%	0.579

効性を示している。

7 結論

本稿では、隠れマルコフモデルを用いてトピックの遷移をモデル化し、談話構造解析を行なう手法を提案した。今後は、高度化された解析結果から再び、用言の共起関係などを学習し、談話構造解析を向上させる予定である。

参考文献

- [1] Regina Barzilay and Lillian Lee. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proceedings of the NAACL/HLT*, pp. 113–120, 2004.
- [2] Hearst.M. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, Vol. 23, No. 1, pp. 33–64, March 1997.
- [3] David M.Blei and Pedro J.Moreno. Topic segmentation with an aspect hidden markov model. In *Proceedings of the 24th annual international ACM SIGIR conference on Reserch and development in information retrieval*, pp. 343–348. ACM Press, 2001.
- [4] 加藤紀雄, 柴田知秀, 黒橋禎夫. 言語情報と映像情報の統合による物体のモデル学習と認識. 言語処理学会 第11回年次大会, 3 2005.
- [5] 柴田知秀, 黒橋禎夫. 料理教示発話の理解と作業構造の自動抽出. 情報処理学会 自然言語処理研究会, No. 2004-NL-164, pp. 117–122, 11 2004.
- [6] 黒橋禎夫, 長尾眞. 表層表現中の情報に基づく文章構造の自動抽出. 自然言語処理, Vol. 1, No. 1, pp. 3–20, 10 1994.