

## 発話を出力単位とする HMM による対話に対する話題分割

但馬 康宏<sup>†</sup> 北出 大蔵<sup>††</sup> 中野未知子<sup>††</sup> 藤本 浩司<sup>†††</sup> 中林 智<sup>††††</sup>  
小谷 善行<sup>†</sup>

<sup>†</sup> 東京農工大学 工学部 情報工学科

東京都小金井市中町 2-24-16

<sup>††</sup> トランスコスモス (株)

<sup>†††</sup> (株)NIWS-FEG

<sup>††††</sup> テンソルコンサルティング (株)

E-mail: †{ytajima,kotani}@cc.tuat.ac.jp

あらまし 本研究において、比較的長い対話に対する話題分割を行う手法を提案する。隠れマルコフモデル (HMM) による話題分割は、これまでも盛んに研究されており、音声認識の分野で特に成果をあげている。しかし、一般的に対話を単語の列として取り扱うため、長さが数百語程度以上の対話の場合にその対話の発生確率が著しく低くなり、有効精度不足となる。本手法では、対話における発話を HMM の出力単位として話題分割を行う。対話における 1 発話ごとにベイズ推定によりあらかじめ話題のラベルを付けた後、そのラベル列を出力する HMM を構成することにより話題の切り替わりを特定する。ここで、HMM はすべての状態間の遷移を許したモデルとした。68 名の被験者で対話実験を行い、62 対話を作成し、本手法の有効性を検証した。この結果、1500 単語程度の長さの対話にたいして良好な分割精度を出せたことを報告する。

キーワード 対話の構造, 話題分割, 隠れマルコフモデル, ナイーブベイズ

## A dialogue segmentation method via utterance based HMM

Yasuhiro TAJIMA<sup>†</sup>, Daizo KITADE<sup>††</sup>, Michiko NAKANO<sup>††</sup>, Koji FUJIMOTO<sup>†††</sup>, Tomo  
NAKABAYASHI<sup>††††</sup>, and Yoshiyuki KOTANI<sup>†</sup>

<sup>†</sup> Tokyo University of Agriculture & Technology

<sup>††</sup> Transcosmos Inc.

<sup>†††</sup> NIWS-FEG Inc.

<sup>††††</sup> TensorConsulting Co. Ltd.

E-mail: †{ytajima,kotani}@cc.tuat.ac.jp

**Abstract** We propose a dialogue segmentation and topic structure finding method via Hidden Markov Model (HMM). HMM has been applied for this problem in previous studies and its advantages have been shown. Nevertheless, the length of the dialogue must be restricted about a hundred words because of computational errors, i.e. the occurrence probability of a dialogue which has a thousand words tends to be less than  $10^{-1000}$  and we fail to construct HMM because of lack of computational precision. In this paper, we propose a new approach for this problem by HMM whose state outputs a symbol of an utterance. Every utterance is classified into some symbols of a segment by a Bayesian classifying method, then we construct an HMM for the target dialogue. The HMM in our method can handle a long dialogue whose length is about 1500 words for 1000 kinds of words. We used 62 dialogues by 68 testee and evaluate our method.

**Key words** dialogue structure, text segmentation, Hidden Markov Model, naive Bayes

## 1. はじめに

対話の分析に対する研究は過去に多くなされており、対話の大局的な理解に不可欠な研究分野となっている。特に近年、ウェブにおけるチャットや音声認識技術の進歩などによりテキスト化された大量の対話記録が手軽に入手できるため、その重要性はますます高まっている。その中でも、対話をその話題などによる段落に分割するテキストセグメンテーションは、対話の構造理解においても重要な課題である。

テキストセグメンテーションにおける重要な研究はいくつかあるが、隠れマルコフモデル (HMM) を用いた方法は、機械学習の応用手法として特に音声認識の分野で強力な手法であることが示されている。越仲らによる変分ベイズを用いた HMM の学習アルゴリズムによる話題分割 [5] や、今井らによる 1 つの段落が複数の話題を含む場合の話題抽出 [4] などでその有効性が確かめられている。

テキストセグメンテーションの代表的手法としては、HMM によるもの以外にも、Hearst による教師無しデータからの分割手法が TextTiling アルゴリズムとして広く知られている [3]。また、この手法の改良に関する研究 [2] も多くなされ、他の確率モデルからの分割手法も提案されている [1]。

一般的な HMM による話題分割手法では、隠れ状態が話題を表し、それぞれの状態において、その話題における単語の出力分布を学習により獲得する。したがって、単語を出力記号とする HMM である。ここで、およそ 1000 種類の単語からなる対話を考える。対話の長さが 1500 語であるとする、各単語の出現率が均等ならば、対話の発生確率はおよそ  $(10^{-3})^{1500} = 10^{-4500}$  となる。これは、一般的な倍精度では表現できず、このような対話に HMM を用いる際に大きな困難となる。

本研究において、発話を出力単位とした HMM を用いたテキストセグメンテーション手法を提案する。本手法では、まずはじめに各発話をベイズ推定によりその発話に属する話題を推定し、各発話に対して話題番号を付与する。その後、話題番号の列を HMM を用いてモデル化することにより、最終的な話題の切れ目を決定する。したがって、本手法における HMM の隠れ状態は、話題を表す点では従来の手法と変わらないが、各状態の出力記号が話題番号である点に特徴がある。本研究では、HMM の構成には、Baum-Welch アルゴリズムを用い、各状態から任意の状態に遷移できる全遷移型として評価実験を行った。その結果、約 2000 種類の単語から成る 1500 語程度の長さの対話に対して、7.8 割の精度で正しい話題分割ができた。

## 2. 本研究における対話の話題分割

本研究では、以下の視点から話題分割を行った。

- 対話は 2 名で行われる。
- 対話者はそれぞれ電話口など、口頭のみでコミュニケーションできる。
- 一方の対話者は対話に目的を持っており、その目的を達成すべく対話を進める (active speaker)。
- もう一方の対話者は対話の目的を知らされていない

(passive speaker)。

- 対話の目的および進め方は、図 1 に示すようなフローチャートで記述することができる。

図 1 におけるチャートでは、それぞれの箱が話題を表し、対話がこのチャートに沿って進んだか否かを HMM を用いて判定することとなる。これらのチャートをシナリオチャートと呼ぶ。

## 3. 提案手法

本提案手法では、HMM の学習に用いる教師データが必要である。これらは、あらかじめ各発話について、シナリオチャートのどの話題に分類されるべきかをラベル付けされた対話である。それらを用いて、以下のステップにて対話の話題分割を行う。

- (1) ベイズ推定により、発話に対して話題番号を付与する。
- (2) 前ステップで付与された話題番号の列を生成する HMM

を用いて、対話に対する最適な状態遷移を求める。

ここで利用する HMM には、以下のような特徴がある。

- 隠れ状態は、シナリオチャートの話題に対応している。
- 隠れ状態から出力されるのは、各発話に対する話題番号であり、その話題番号もシナリオチャートの中の話題と 1 対 1 に対応している。

図 2 に本手法でのデータの流れおよび処理の概要を示す。

過去の研究において、HMM は話題分割の有力なツールであったが、隠れ状態が直接単語を出力することにより対話を表現することが一般的である。その場合、対話全体の生成確率を表現するためには、対話の長さを適切に制限しなければ計算精度不足となる。本手法でも表現できる対話の長さに限界はあるが、一般的な手法より長い対話を同じ計算精度で表現できる。

さらに、本手法の特徴として、HMM の学習に要する計算時間を比較的少なくできる点が挙げられる。

### 3.1 ベイズ推定による発話に対する話題番号付け

第 1 ステップのベイズ推定による話題番号付けについて述べる。  $T$  を  $n$  発話から成る対話とし、各発話を  $u_1, u_2, u_3, \dots, u_n$  と表す。ここで  $u_i$  は  $T$  における  $i$  番目の発話である。各発話  $u_i$  は、  $m$  種類の単語  $w_1^{(i)}, w_2^{(i)}, \dots, w_m^{(i)}$  から成るとする。

ベイズ推定は、発話内の単語に対するナイーブベイズにより、話題番号を決定する。すなわち、発話  $u_i$  が話題  $D_k$  での発話である確率  $P(D_k|u_i)$  を

$$P(D_k|u_i) = \frac{P(u_i|D_k)P(D_k)}{P(u_i)}$$

として求める。ここでさらに、

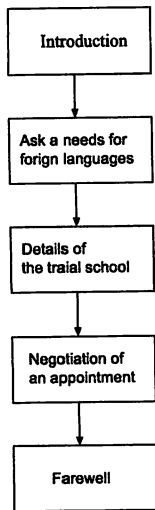
$$P(u_i|D_k) = P(w_1^{(i)}|D_k)P(w_2^{(i)}|D_k)\dots P(w_m^{(i)}|D_k)$$

と近似し、単語  $w$  について、

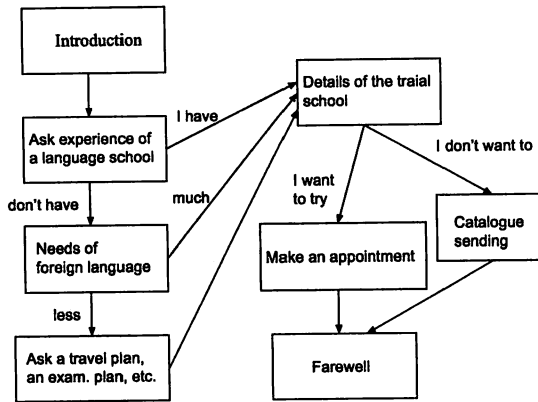
$$P(w|D_k) = \frac{\sum_{u \in H_k} c(w, u)}{\sum_{u \in H_k} \sum_{v \in W_u} c(v, u)}$$

とした。ここで  $H_k$  は、訓練データにおいて話題  $k$  に分類されるすべての発話の集合であり、 $W_u$  は発話  $U$  に表れるすべての単語の集合、 $c(w, u)$  は発話  $u$  における  $w$  の出現個数である。

上記の方法で発話  $u_i$  に付けられた発話番号を  $o(u_i)$  と表す。



Scenario chart 1



Scenario chart 2

図 1 シナリオチャート

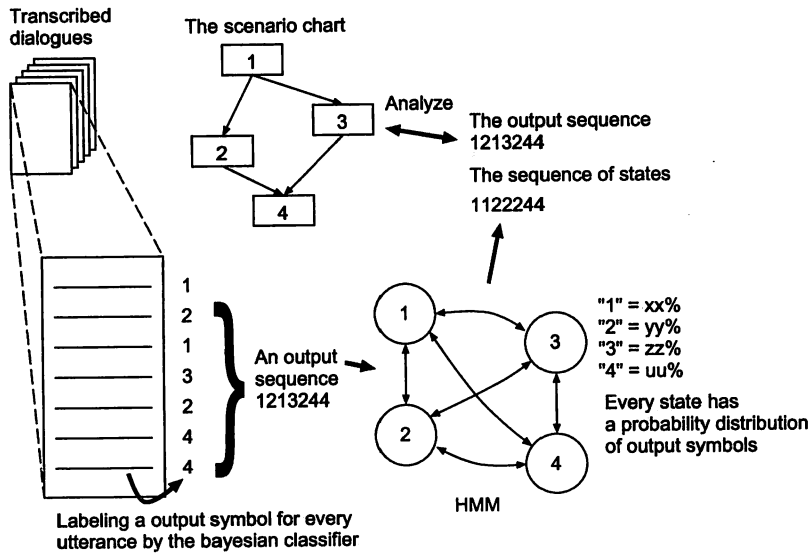


図 2 提案手法の概要

次のステップの HMM の構成では、対話  $T = u_1 u_2 \dots u_n$  に対する発話番号の系列  $o(u_1) o(u_2) \dots o(u_n)$  を出力する HMM を構成する。

### 3.2 HMM の構成

学習データから Baum-Welch アルゴリズムにより話題番号の系列  $o(u_1) o(u_2) \dots o(u_n)$  を出力する HMM を構成する。ここで、HMM の隠れ状態はそれぞれ、シナリオチャートにおける話題に対応し、すべての状態から任意の状態へ遷移可能な全遷移型の HMM とした。全遷移型の HMM とすることで、シナリオチャートに沿っていない対話の流れについても明らかにすることができる。本研究におけるシナリオチャートは、解析時に定めるものであり、対話がシナリオチャートに基づいている必然性はない。したがって、シナリオチャートとまったく違う状態遷移が学習データを適切に表している場合もあるため、全遷移型の HMM を用いている。

次に  $p$  を発話番号の列  $o(T)$  の HMM における生成確率とし、 $n$  を対話  $T$  における発話の総数とする。以下を満たす  $q$  を考える。

$$\log(p) = \log(q^n) \quad (1)$$

このとき  $q$  は、対話  $T$  における 1 発話あたりの平均出現確率を表している。すなわち、この  $q$  の値を求めることで、 $T$  がシナリオチャートに合致しているかどうかを他の対話と比べることができる。

## 4. 評価実験

評価実験として、ある架空の英会話教室への勧誘を目的とした対話を設定し、以下の条件で実際の対話を行い、その対話記録に対して本手法を適用し評価を行った。

● 被験者は、4 名の active speaker と 62 名の passive speaker である。

passive speaker はそれぞれ、1 名の active speaker に割り当てられ、active speaker から電話をかけることにより対話を行った。

● 実験対話の総数は 62 である。  
● すべての対話は、人手によりテキストへの書き起こされ、利用される。

● 形態素解析に用いたプログラム：mecab 0.91  
● 解析に用いるシナリオチャートは、2 種類。図 1 にそれぞれを示した。

- ひとつの対話あたりの平均発話数：167.11
- 実験全体での発話数：10358
- ひとつの発話に含まれる単語数の平均：10.58
- 全ての対話における単語数の合計：109573

シナリオチャート 1 については、全ての対話の発話に対して話題番号が付けられ、シナリオチャート 2 については、28 の対話について話題番号が付けられた。したがって、シナリオチャート 2 に関しては、シナリオチャート 1 よりも少ない学習データで実験を行った。

以下は、この実験対話の一部分である。ここで“A”は active

speaker を表し、“B”は passive speaker を表す。発話の最後の“( )”内の数字が話題番号を表す。

Talk1:

- B: もしもし。(0)  
A: あ、XXX 様のお宅でしょうか？(0)  
B: あ、はいそうです。(0)  
A: 私、トランスコスモス英会話スクールの YYY と申します。(0)  
B: あ、はいどうも。(0)  
A: えー先日は東京農工大学と共催させて頂きました、英語学習に関するセミナーにご参加いただきまして、まことにありがとうございました。(0)  
B: はい。(0)  
A: 本日は失礼ながら、セミナーの際におうかがいしましたご連絡先をもとにお電話をおかけしております。(0)  
B: あ、はい。(0)  
A: 英語学習にご興味をお持ちの、XXX 様にお伝えしたいことがございますので、(0)  
B: はい。(0)  
A: えー本日、ただ今少々お時間を頂戴してもよろしいでしょうか？(0)  
B: あ、はい。(0)  
A: ありがとうございます。えー、トランスコスモス英会話スクールは、来年で開校 30 周年を迎えることになりまして、(0)  
B: はい。(0)  
A: え、これを機に無料の体験スクールの開催を、えー計画しております。(0)  
B: はい。(0)  
A: XXX 様は先日、英語学習に関するセミナーに、ご参加いただいたということなのですが、(0)  
B: あ、はい。(0)  
A: これまでにですね、英会話スクールなどへの、えー、ご入学などを、お考えになったことがございますでしょうか？(1)  
B: 今までは特になかったですけど。(1)  
A: ええ、これから先、では、ぐ、(1)  
B: 今回、そうですねー(1)  
A: 具体的にどこか、英語をお使いになる機会はございます... かね？(2)  
B: 機会、ん、あ、まー、特に機会とかはないですけど、英語一今後のために、(2)  
A: まあ、お仕事とかご旅行とかですかね。(2)  
B: そうですね、はい。(2)

.....

### 4.1 話題分割の結果

表 1 に、シナリオチャート 1 を用いて話題分割のした場合の正答率を、表 2 に、シナリオチャート 2 を用いて話題分割のした場合の正答率を示す。ここで正答率は (正しい話題番号が付けられた発話数) / (対話における全発話数) で算出した。

これらの結果より、本手法は対話に対する話題分割に十分な

表 1 話題分割の正答率 (シナリオチャート 1)

dialogue name	correct utterances	total / utterances	rate	dialogue name	correct utterances	total / utterances	rate
talk1	69	/ 69	1.0000	talk38	27	/ 48	0.5625
talk3	62	/ 63	0.9841	talk40	32	/ 34	0.9411
talk4	42	/ 51	0.8235	talk41	80	/ 82	0.9756
talk5	46	/ 56	0.8214	talk42	53	/ 57	0.9298
talk6	66	/ 70	0.9428	talk43	68	/ 74	0.9189
talk7	94	/ 97	0.9690	talk44	68	/ 73	0.9315
talk8	57	/ 63	0.9047	talk45	68	/ 128	0.5312
talk9	48	/ 53	0.9056	talk46	41	/ 55	0.7454
talk11	105	/ 129	0.8139	talk47	71	/ 74	0.9594
talk12	68	/ 71	0.9577	talk48	91	/ 103	0.8834
talk13	35	/ 36	0.9722	talk49	86	/ 90	0.9555
talk14	65	/ 72	0.9027	talk50	93	/ 97	0.9587
talk15	113	/ 124	0.9112	talk51	42	/ 195	0.2153
talk16	154	/ 156	0.9871	talk52	94	/ 99	0.9494
talk17	103	/ 107	0.9626	talk53	85	/ 96	0.8854
talk18	67	/ 69	0.9710	talk54	63	/ 81	0.7777
talk20	69	/ 71	0.9718	talk55	19	/ 34	0.5588
talk21	56	/ 61	0.9180	talk56	108	/ 138	0.7826
talk23	54	/ 58	0.9310	talk57	78	/ 80	0.9750
talk24	30	/ 37	0.8108	talk58	79	/ 99	0.7979
talk25	23	/ 30	0.7666	talk59	98	/ 129	0.7596
talk26	59	/ 62	0.9516	talk62	205	/ 211	0.9715
talk27	55	/ 58	0.9482	talk63	93	/ 95	0.9789
talk28	34	/ 41	0.8292	talk64	98	/ 101	0.9702
talk29	51	/ 54	0.9444	talk66	56	/ 74	0.7567
talk31	70	/ 78	0.8974	talk67	80	/ 93	0.8602
talk32	63	/ 64	0.9843	talk68	162	/ 164	0.9878
talk33	60	/ 61	0.9836	talk69	132	/ 146	0.9041
talk34	63	/ 64	0.9843	talk70	71	/ 94	0.7553
talk35	42	/ 47	0.8936	talk71	62	/ 91	0.6813
talk36	62	/ 67	0.9253	average			0.8775
talk37	43	/ 44	0.9772	mean dev.			0.1768

表 2 話題分割の正答率 (シナリオチャート 2)

dialogue number	correct utterances	total utterances	rate
talk1	59 / 70		0.8429
talk3	50 / 64		0.7813
talk6	61 / 71		0.8592
talk7	87 / 98		0.8878
talk8	57 / 64		0.8906
talk9	45 / 54		0.8333
talk11	52 / 130		0.4000
talk12	53 / 72		0.7361
talk13	27 / 37		0.7297
talk15	61 / 125		0.4880
talk16	103 / 157		0.6561
talk17	71 / 108		0.6574
talk18	56 / 69		0.8116
talk21	53 / 62		0.8548
talk27	41 / 65		0.6308
talk32	57 / 65		0.8769
talk33	59 / 61		0.9672
talk41	77 / 83		0.9277
talk44	68 / 74		0.9189
talk50	91 / 98		0.9286
talk52	57 / 100		0.5700
talk57	71 / 82		0.8659
talk59	93 / 130		0.7154
talk62	80 / 212		0.3774
talk63	78 / 96		0.8125
talk64	84 / 102		0.8235
talk68	87 / 165		0.5273
talk69	93 / 147		0.6327
average			0.7501
mean dev.			0.1603

精度を持っていると言える。

表 3 本手法と一般的な HMM モデルとの学習計算時間の比較

time(s)	talk1	talk11	talk12	talk13	talk14	average
our method	441	398	431	417	445	427
standard	73904	72558	74109	75007	82781	75671

#### 4.2 一般的な手法との比較

本研究における評価実験で利用した対話に対して、一般的な HMM を構成して、比較検討する。前章までで述べたように、一般的な隠れ状態が単語を出力する HMM では、今回の実験で得られた、長さ 1500 語程度、1 対話内での語彙数およそ 2000 の対話に対しては、有効な計算精度を確保できず適切な結果を得ることができなかった。

表 3 に本手法による HMM を構成する学習時間と、一般的な HMM を構成する際の学習時間との比較を示す。その結果、格段に高速であることが示された。これらの結果から、本手法の有効性が示された。

#### 4.3 本手法で得られた HMM の応用

次に、式 1 において定義した、対話における 1 発話あたりの

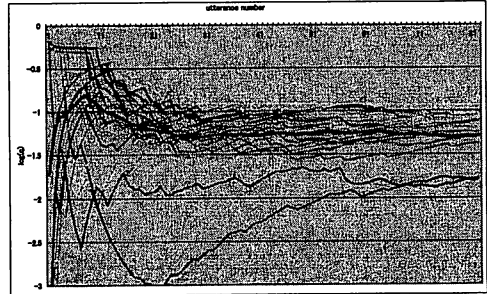


図 3 1 発話あたりの平均出現率 (シナリオチャート 2)

平均出現率  $q$  を考える。図 3 から、シナリオチャート 2 で分析した結果、各対話ごとの  $\log(q)$  の値が読み取れる。図において、黒太曲線が各対話の平均値である。グラフより、実験対話では、対話の前半にモデルとかけ離れた展開をすることが多く、対話の後半になるほどモデルに沿った展開が行われていることがわかる。

#### 5. おわりに

本稿において、HMM を用いた対話の話題分割について新しい手法を提案し、その特徴を評価実験により明らかにした。本手法における HMM は一般的な手法に比べ長い対話 (1 つの対話はおおよそ 1500 語) を扱うことができ、話題分割の精度も平均 75% - 87% と良好であった。さらに本研究では、対話がモデルに沿って進められているか否かを評価する方法を示した。その結果、評価実験における対話では、前半にモデルから離れる傾向があり、対話の後半になるほどモデルに沿った進め方となっていることが示された。今後の課題として、複数人数による対話に対する分割手法の考案、対話の目的がない自由対話における話題抽出方法の考案などが挙げられる。

#### 文 献

- [1] D. Beeferman, A. Berger, and J. Lafferty, Statistical models for text segmentation, *Machine Learning*, 34(1-3), pp.177-210, 1999.
- [2] 別所 克人, クラスタ内変動最小基準に基づくテキストセグメンテーション, *情報処理学会論文誌*, 47(3), pp.957-967, 2006.
- [3] M. A. Hearst, Texttiling: segmenting text into multi-paragraph subtopic passages, *Computational Linguistics*, 23, pp.33-64, 1997.
- [4] 今井 亨, R. Schwartz, 小林 彰夫, 安藤 彰男, 話題混合モデルによる放送ニュースからの話題抽出, *信学論 J81-D-II-9*, pp.1955-1964, 1998.
- [5] 越仲 孝文, 奥村 明俊, 磯谷 亮輔, HMM の変分ベイズ学習によるテキストセグメンテーション及びその映像インデキシングへの応用, *信学論 J89-D-9*, pp.2113-2122, 2006.