共通状態と連結学習を用いたHMMによるコールセンタ対話の要約

東中竜一郎† 南泰浩‡ 西川仁† 堂坂浩二‡ 目黒豊美‡ 小橋川哲† 政瀧浩和†
吉岡理† 高橋敏† 菊井玄一郎†

† 日本電信電話株式会社 NTT サイバースペース研究所
‡ 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

1 はじめに

テキストデータの要約研究は多い [5]．要約手法としては，文書の最初のN文を抽出する方法（LEAD法）や機械学習の手法によって重要な文を特定し，それらを抽出する方法などがある [3, 7]．また，近年では，要約を整数計画問題（ILP）と置いて，重要と考えられる単語を最も多く被覆するような文を選択する手法も考案されている [2]．

本稿では，複数のドメインに分かれたテキストデータの要約を扱う．ここで，「複数のドメインに分かれた」とは，ひとつのテキストデータが複数のドメインの内容を含むということではなく，単一ドメインのテキストデータの集合が複数ドメイン分あるということである．複数のドメインにまたがるテキストデータを扱う場合，従来，個々のドメインについて，隠れマルコフモデル (HMM) などを用い，要約器を学習するアプローチが用いられてきた [1]．しかしながら，ドメインが多くなるにつれ，学習データの作成コストが高くなるという問題があった．

本研究では，学習データ作成のコスト低減のため，要約の正解を作成せずに要約器を学習する手法を提案する．具体的には，ドメインラベルのみが付与されたテキストデータ集合から各ドメインに特徴的な系列をHMMによって学習し，あるドメインのテキストデータの要約を行うとき，このドメインに特徴的な系列に該当する箇所のみを要約として抽出する．ここで，HMMの学習には，状態として，すべてのドメインに共通なシンボルを出力する「共通状態」を追加し，各ドメインに特徴的な系列を特定の状態から出力されやすくする手法である「連結学習」を用いる．なお，本稿において，HMMの学習はEMアルゴリズムによるものを指す．

本稿では，共通状態と連結学習を用いたHMMの作成法と，作成したHMMをコールセンタ対話（お客様センタ）の要約に適用した結果について報告する．コールセンタでは電話の故障受付，契約，設置など，さまざまな種別の対話を扱う．よって，これらは複数ドメインに分かれたテキストデータである．コールセンタでは大量のコールを扱う上，個々の対話は一般に長い．そのため，オペレータや分析者がすべての対話を効率的に振り返ることが難しく，要約技術の適用によって，コールセンタにおける対話の分析が容易になると考えられる．

2 共通状態と連結学習を用いたHMM

われわれが提案するHMMは，二者対話の分析に用いられるSpeaker HMM (SHMM) [6] を拡張し，系列の分類問題に適用できるようにしたものである．SHMMは，話者1（speaker1）と話者2（speaker2）のそれぞれに対応する状態を持ち，各状態は，対応する話者の発話（発話内容を表すシンボル）のみを出力する．各状態はどの状態にも遷移可能である．われわれは，各ドメインの対話データから個別に学習されたSHMMを複数組み合わせて，新たなHMMを構成する．たとえば，図1のように組み合わせる．

図1に示すHMMについて，ある系列の入力があり，そのときの最尤の状態系列（ビタビデコーディングなどで求められる）が $< 1, 3, 4, 2 >$ だったとすると，それぞれの状態がどのドメインに対応しているかを見ることで，$< 1, 2, 2, 1 >$ というドメイン系列を得ることができる．SHMMの組み合わせ方には3種類ある．以下にそれぞれを説明する．

2.1 エルゴディック

「エルゴディック」は独立に学習されたSHMMをエルゴディックに等確率で接続したHMMである．トポロジーとしては図1である．エルゴディックでは，すべての状態が等確率で接続されているため，分類は，各SHMMにおける発話の頻度分布に左右される．例

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

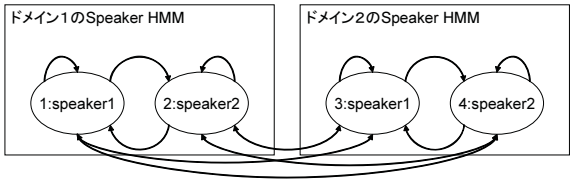

図 1: Speaker HMM を組み合わせた HMM

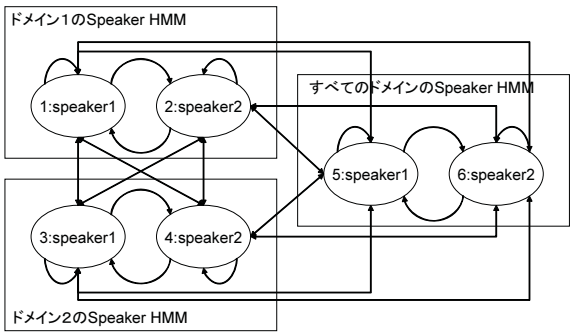

図 2: 共通状態を持つエルゴディック

えば，ある発話がドメイン 2 に比べドメイン 1 に高頻度で出現するのであれば，その発話はドメイン 1 の SHMM から出力され，結果，ドメイン 1 と分類される．

2.2 共通状態を持つエルゴディック

どのドメインにも共通に現れる発話系列というものが存在する．例えば，コールセンタの対話であれば，すべてのドメインの対話に共通して，挨拶のやり取りや個人情報の確認などが現れる．エルゴディックではこういった共通した発話系列を既存のクラスのどれかに割り振ってしまう．つまり，たまたま，挨拶がドメイン 1 に多少多く出現したからという理由で，挨拶はドメイン 1 に分類されてしまう．できればこのような複数のドメインにまたがって出現するものは，どのドメインにも分類されないようにモデル化するのがよい．

そこで，挨拶のような発話はドメイン 1 でもドメイン 2 でもなく，共通ドメインというものを仮定して，そのドメインに割り振ることを考える．これは，図 2 に示す形状を持つ HMM で実現できる．この HMM ではエルゴディックに加えて，すべてのデータから学習された SHMM を持ち，すべての状態がエルゴディックに接続されている．すべてのデータから学習された SHMM は全ドメインの系列をモデル化しているため，すべてのドメインに共通した系列を表すと考えられる．なお，すべてのドメインのデータから学習された SHMM に含まれる状態を共通状態と呼ぶ．

このような HMM を用いることで，ある入力系列に対して，最尤の状態系列が $< 1, 4, 5, 6, 3, 2 >$ である場合，$< 1, 2, 0, 0, 2, 1 >$ のように入力系列をドメイン

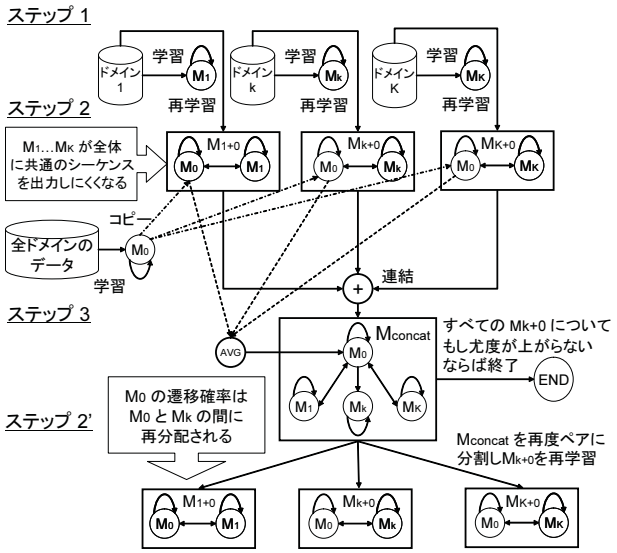

図 3: 連結学習を用いて HMM を学習する手続き

系列に分類することができる．ここで，入力における $< 5, 6 >$ は共通状態であり，これらに対応するドメイン系列はドメイン 1，ドメイン 2 のどちらにも属さず，共通の系列であるというように分類される．共通状態を持つことで，無理矢理どちらかのドメインに入力系列を分類しなくても良いため，ドメイン分類の精度向上が期待できる．

2.3 共通状態と連結学習

共通状態を持つ HMM にも問題があり，それは，全体のデータから学習された SHMM の出力分布がなだらかになってしまうことである．これは，複数のドメインの情報を平均化したようなモデルを学習してしまうことに起因する．この影響で，入力系列に対して最尤の状態系列を求めると，一切共通状態を経由しないことが起こり得る．これを解決する手段は二つある．

一つの解決策は，共通状態の数を増やすことである．そうすることで，尖った分布を持つ共通状態を保持することができ，その結果，入力系列に対する最尤の状態系列が共通状態を通過する可能性が出てくる．

もう一つの解決策は，連結学習 [4] を用いることである．本稿ではこちらの解決策に着目する．この手法により，各ドメインにおける系列と全ドメインに共通して現れるような系列の出力分布を，特定の状態に集中させることができる．具体的には，下記の手続きによって学習される HMM を用いて入力系列をドメインラベルの系列にデコードする．なお，この手続きを図としてまとめたものが図 3 である．

ステップ 1 M_k $(M_k \in M, 1 \leq k \leq K)$ をそれぞれ D_k から学習した SHMM とする．ここで，$D_k =$

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

$\{\forall d_j | c(d_j) = k\}$ であり，M_0 はすべてのドメインのデータから学習した SHMM である．すべてのデータとはすなわち，D である．ここで，K は全体のドメイン数であり，$c(d_j)$ は系列 d_j の属するドメインを表す．

ステップ 2 $M_k \in M$ と M_0 のコピーを同じ初期確率，同じ遷移確率でもって接続する．このモデルを，M_{k+0} と呼ぶ．そして，M_{k+0} を $\forall d_j \in D_k$ の学習データで再学習する．ここで，$c(d_j) = k$ である．

ステップ 3 M_{k+0} $(1 \leq k \leq K)$ をすべて統合して一つの HMM にする．この HMM を M_{concat} と呼ぶ．ここで，統合の際，M_0 のコピーの出力確率は K で平均化される．もし，M_{k+0} のいずれも学習データに対する尤度が改善しないようであれば，この処理を抜ける．そうでない場合は，ステップ 2 に戻る．このとき，すべての k について，M_0 と M_k を接続するが，M_0 から $M_l (l \neq k)$ への遷移確率は一度足され，その後，M_0 の自己遷移と M_k への遷移に均等に分配される．

3 コールセンタ対話の要約

K 個のドメインからなるコールセンタ対話のデータがあるとき，まず，前節で説明した HMM を学習する．そして，ドメイン k の対話データが入力されたとき，各発話がドメイン k の状態から出力された確率を forward-backward アルゴリズムで得る．ここで得られた事後確率を発話の重要度とみなし，これを元に各発話中の単語の重要度を決定する．最後に，単語重要度の総和が要約長内で最大になるように発話を選択し，要約とする．要約処理は，学習フェーズとデコーディングフェーズからなる．それぞれを以下に説明する．

学習フェーズ D $(d_1 \ldots d_N)$ をコールセンタ対話のすべてのデータとし，DM^k $(DM^k \in DM, 1 \leq k \leq K)$ をドメイン k に与えられるドメインラベルだとする．$U_{d_i,1} \ldots U_{d_i,H}$ は対話 d_i 中の発話系列である．ここで，H は d_i 中の発話数を指す．

まず，D から，2 種類のモデルを構築する．一つはトピックモデル (TM) であり，もう一つはわれわれの提案する HMM である．トピックモデルは対話データ中の各発話を一つのトピックラベルに落とし込む処理に必要である．この処理は，HMM の特徴量があまりに高次元になると学習が困難になるため，これを回避するために行う．トピックモデルを作る方法としては probabilistic latent semantic analysis (PLSA) や latent Dirichlet allocation (LDA) などがある．本研究では，LDA を用い，モデルは bag-of-words を特徴量として学習し，この結果，$\mathrm{P}(z|w)$ を得る．ここで，w は単語であり z はトピックである．このトピックモデルを用いることにより，D の各発話について，トピックラベルを付与することができる．すなわち，$\operatorname{argmax}_z \sum_{w \in \mathrm{words}(U_{d_i})} \mathrm{P}(z|w)$ となる z を各発話に割り振る．

D 中のすべての発話にトピックラベルを付与し終えたら，トピックラベルの系列を HMM で学習する．

デコーディングフェーズ d_j を入力された対話とし，$DM(d_j)$ $(\in DM)$ を対話 d_j に対してドメインラベルを得るテーブルとし，$U_{d_j,1} \ldots U_{d_j,H_{d_j}}$ を d_j 中の発話系列とする．ここで，H_{d_j} は対話中の発話数である．まず，学習フェーズで作成した TM を使って発話系列をトピック系列 $T_{d_j,1} \ldots T_{d_j,H_{d_j}}$ にし，そして，われわれの提案する HMM を用いて，forward-backward アルゴリズムにより，$DM(d_j)$ に対する事後確率 $P_{d_j,1} \ldots P_{d_j,H_{d_j}}$ を得る．ここで，発話 $U_{d_j,l}$ 中の単語 w の重要度を $P_{d_j,l} \cdot \mathrm{tf}(d_j, w)$ と定め，この総和を要約長内で最大化するように，対話中の発話を ILP の定式化を用いて選択する．ここで，tf は d_j における w の頻度を返す関数である．要約の冗長性を減らすため，同じ w については一度しか総和の計算に用いない．

4 実験

コールセンタ対話の模擬データを独自に収集した．データ収集には，90 人の実験参加者が参加した．参加者はオペレータとユーザに分かれて，予め準備されたシナリオにしたがって対話を行った．オペレータには実際にコールセンタにおける応対経験者を用いた．

対話のドメインは，金融，インターネットサービスプロバイダ，自治体への問い合わせ，通信販売，PC サポート，電話についての問い合わせの 6 種類である．それぞれのドメインについて 15–20 のシナリオを用意し，これらに基づいて，オペレータとユーザは，別室に分かれ電話を介して音声で通話した．本実験ではこの通話を書き起こしたものをデータとして用いた．

対話データの収集は二度にわたって行われ，それぞれ，391 対話と 307 対話を収録した．以降，初回の 391 対話を学習データ，第二回の 307 対話をテストデータと呼ぶ．一対話にはおおよそ 130–150 発話が含まれ，一発話の平均長は約 11 文字である．要約の正解として，一人の作業者（作業者 A）が，すべての対話について，発話を抽出することにより，250 文字，500 文

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

表 1: 250 文字の要約長における発話抽出の F 値

学習セット	(a) エルゴ	(b)+共通状態	(c)+連結学習
set1	0.211	0.220^{a}	$\mathbf{0.254}^{aabb}$
set1–2	0.219	0.229^{aa}	$\mathbf{0.256}^{aabb}$
set1–3	0.226	0.228	$\mathbf{0.248}^{aabb}$
set1–4	0.225	0.235^{a}	$\mathbf{0.268}^{aabb}$
set1–5	0.226	0.237^{a}	$\mathbf{0.263}^{aabb}$

表 2: 500 文字の要約長における発話抽出の F 値

学習セット	(a) エルゴ	(b)+共通状態	(c)+連結学習
set1	0.395	0.397	$\mathbf{0.432}^{aabb}$
set1–2	0.403	0.406	$\mathbf{0.432}^{aabb}$
set1–3	0.403	0.405	$\mathbf{0.431}^{aabb}$
set1–4	0.406	0.416^{aa}	$\mathbf{0.444}^{aabb}$
set1–5	0.407	0.412	$\mathbf{0.431}^{aabb}$

字要約を作成した．全 698 対話から 120 対話をサンプリングし，もう一人の作業者（作業者 B）との発話抽出の一致率（Cohen's κ）を調べたところ，250 文字要約，500 文字要約について，それぞれ 0.43 と 0.53 であり，中程度の一致であった．本実験では，作業者 A のデータを評価時の正解として用いた．

提案手法の有効性，および，学習データ量の増加による効果を検証するため，本実験では，まず，学習データから，各ドメインの対話を 50 対話ずつ抽出した．残りの 91 対話は本実験では用いない．そして，各ドメインの対話を 10 対話ずつに分けた後，各ドメインの対話を 10 対話ずつ含むセットを 5 つ作成した．これらを，set1 . . . set5 と呼ぶ．6 ドメインあるため，各セットには 60 対話が含まれる．そして，set1（=60 対話），set1–2，set1–3，set1–4，set1–5（=300 対話）をそれぞれ学習データとして，われわれの提案する HMM を学習し，テストデータについて要約を出力させ，発話抽出の精度を F 値で算出した．ここで，set1–N は set1 から setN のすべてのセットを合わせた対話集合を指す．トピックモデルにおけるトピック数は 100 とし，SHMM の状態数は各話者 1 つずつ，共通状態の状態数は各話者 3 つずつとした．

表 1 と表 2 は，要約長を 250 文字，500 文字に制限した場合の要約精度である．ここで，(a) エルゴとは共通状態を持たない HMM (cf. 2.1 節) である．(b) と (c) は，それぞれ，共通状態を付加した HMM (cf. 2.2 節) と，その HMM にさらに連結学習を適用した HMM (cf. 2.3 節) である．F 値の肩にある a，b は，それぞれ (a)，(b) よりも t-test で統計的に有意な差であることを示す．aa のように二つあれば，p <0.01，一つであれば，p <0.05 を表す．ボールドは各行で最も精度の高い数値を示す．表から分かるように，連結学習を用いることで精度が向上することが分かる．また，若干ではあるものの，データ量を増やす（set1 に他のセットを加えていく）ことで，要約精度が向上していることが分かる．(a) や (b) がデータ量の増加とともに，徐々に精度を向上させているのに比べ，(c) は少ないデータからでも精度が比較的高い．これは，連結学習によって各ドメインの特徴が効率的に学習されているからだと考えられる．

5 まとめと今後の課題

共通状態と連結学習を用いた HMM を提案し，コールセンタ対話の要約に適用した結果について報告した．今後，さらなる精度改善と音声認識結果への適用について検討していく予定である．

参考文献

[1] Regina Barzilay and Lillian Lee. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proc. HLT-NAACL*, pp. 113–120, 2004.

[2] Dan Gillick and Benoit Favre. A scalable global model for summarization. In *Proc. the Workshop on Integer Linear Programming for Natural Language Processing*, pp. 10–18, 2009.

[3] Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proc. the 18th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR)*, pp. 68–73, 1995.

[4] Kai-Fu Lee. *Automatic speech recognition: the development of the SPHINX system*. Kluwer Academic Publishers, 1989.

[5] Inderjeet Mani. *Automatic summarization*. John Benjamins Publishing Company, 2001.

[6] Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. Analysis of listening-oriented dialogue for building listening agents. In *Proc. SIGDIAL*, pp. 124–127, 2009.

[7] Miles Osborne. Using maximum entropy for sentence extraction. In *Proc. the ACL-02 Workshop on Automatic Summarization*, pp. 1–8, 2002.

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.