

## 文クラスタ混合分布 N-gram の検討

清水徹 大野晃生 黒岩眞吾 樋口宜男

(株)KDD 研究所

〒356-8502 埼玉県上福岡市大原 2-1-15

Tel.: 0492-78-7383 e-mail: shimizu@kddlabs.co.jp

あらまし 本論文では、十分な学習データ量が確保できない場合の N-gram モデルの構築法として、文レベルの混合分布 N-gram と類似タスクのデータを加えて学習する方法を組み合わせた手法を提案する。本手法は、ターゲットトピックのデータと類似タスクのデータを足しあわせて文をクラスタリングし、各クラスタの N-gram から混合分布 N-gram を構成するステップと、ターゲットトピックのデータだけを用いて混合分布 N-gram の混合比を学習するステップからなる。本手法のメリットは、類似タスクを用いた学習データ量の確保と、N-gram の混合比の学習による不要な学習データの除去にある。実験の結果、提案法はターゲットトピックのデータだけから学習した trigram よりもクロスエントロピーが低下することが確認された。

キーワード • クラスタリング • 統計的言語モデル • 混合分布 N-gram • 会話音声

## A Study on Sentence-Level Mixture N-gram based on Sentence Clustering.

Tohru Shimizu Teruo Ohno Shingo Kuroiwa  
Norio Higuchi

KDD R&D Laboratories Inc.

2-1-15 Ohara Kamifukuoka-shi, Saitama 356-8502

Tel. 0492-78-7383 e-mail: shimizu@kddlabs.co.jp

### Abstract

This paper proposes a new method for developing statistical N-gram language models which integrate sentence-level mixture N-grams and selective use of similar task data. In this method, component N-gram parameters are estimated using both target topic data and similar task data, then the sentence-level mixture N-gram model is adapted by using only target topic data. This approach has the advantage that it can use more data for training and remove useless clusters, which are far from target topic data. The experiment results show that this method achieves the cross-entropy reduction compared with the standard trigram.

key words • Clustering • Statistical language model • Mixture N-gram • Conversational speech

## 1. はじめに

筆者らは、会話音声の認識に使用する N-gram ベースの統計的言語モデルの構築法について検討を進めている。直前の N-1 単語の履歴に基づいて N 番目の単語を確率的に予測するモデルである N-gram は、単純な言語モデルであることと、大量のテキストデータがあれば容易に構築できることから、新聞記事や放送音声のディクテーションを対象として広く用いられている [1]

N-gram の問題点は、文の構造をより正確に捉えるために N の値を大きくすると、学習データに現れない N 単語連鎖の補完が必要になってくることである。この問題点を解決するため、予測単位である単語をクラス化するクラス N-gram [2] や、予測単位を単語列に拡張するフレーズ N-gram [3, 4]、あるいはクラス N-gram とフレーズ N-gram を組み合わせた方法 [5] が提案されている。クラス N-gram は、複数の単語を 1 つのカテゴリとして取り扱うことにより、パラメータ数を少なくし学習セットに現れなかった単語列に対する遷移確率を推定する効果を狙ったものであるが、単語をクラス化することによる湧き出し誤りが生じやすいという問題点がある。フレーズ N-gram は、単語 N-gram の予測単位を単語列に拡張することにより、単語の予測力を改善する方法である。この方法は、予測単位を長くすることにより湧き出し誤りを軽減できる効果があるが、基本的に長単位 N-gram と同様な効果を狙ったものであるため、trigram に適応した場合には、bigram ほどの改善効果が得られていない [4, 5]。このため、十分な学習データが確保できない会話音声用 N-gram 構築においては、予測単位の変更だけでは十分な推定精度を確保することは難しいと考えられる。

これに対して、1 つの N-gram を学習する母集団を変えるアプローチとして、大量データが確保できる別タスクのデータを加えて学習する方法 [6, 7]、学習データを類似トピックに分割して、各々のトピック毎の N-gram の推定精度を上げる方法 [8, 9, 10, 11, 12] が提案されている。別タスクのデータを加えて学習する方法では、加えたデータがタスクのカバー率向上に寄与する場合に認識性能が向上することが報告されている [6]。しかし、タスクの違いが大きい場合は必要となるデータ量が膨大になる上に、場合によっては加えたデータが逆に N-gram の推定精度を劣化させる副作用も懸念され

る [1]。一方、学習データを分割する方法では、分割に伴いクラスタ当たりの学習データ量が少なくなることから、クラスタ数の設定やクラスタリングの工夫が必要になる。

また、N-gram 一般に言えることであるが、学習データが決まればモデルが自動的に構築できる反面、モデルの制約力を簡便な方法によって変えることは難しい。

本論文では、これらの結果を踏まえて、(1) 別タスクのデータを加えてもターゲットタスクに対する推定精度の劣化が生じる可能性が少ない、(2) クラスタ数の増加に伴うクラスタ当たりの学習データ量の減少がモデルの劣化に直接結びつかない、(3) モデルの制約力を簡便に変えることのできるという 3 つの特長を有する手法を提案する。提案法は、ターゲットトピックのデータと類似タスクのデータを足しあわせて文をクラスタリングし、各クラスタの N-gram から混合分布 N-gram を構成するステップと、ターゲットトピックのデータだけを用いて混合分布 N-gram の混合比を学習するステップからなる。

以下、第 2 節では、ターゲットトピックのデータに類似タスクのデータを加えて N-gram を構築する方法を説明し、第 3 節では、クラスタリング方法を説明する。第 4 節では、会話音声のテキストデータを用いたクロスエントロピーの評価によって、本手法の有効性を示す。

## 2. 類似タスクデータの追加によるターゲットトピック N-gram 構築法

### 2.1 文レベルの混合分布 N-gram

まず、ターゲットトピック N-gram を構築する上で重要な概念となる文レベルの混合分布 N-gram について述べる。

クラスタ数を  $m$ 、クラスタ  $k$  に対するモデルを  $M_k$ 、クラスタ  $k$  における単語  $w_i$  の N-gram 確率を  $P(w_i | w_{i-N+1}^{i-1}, M_k)$ 、クラスタ  $k$  に対する混合比を  $\lambda_k$  とした場合、単語数  $T$  からなる文  $w_1^T$  の文レベル混合分布が与える出現確率  $P(w_1^T)$  は、(1) 式で表される。

$$P(w_1^T) = \sum_{k=1}^m \lambda_k \prod_{i=1}^T P(w_i | w_{i-N+1}^{i-1}, M_k) \quad (1)$$

(1) 式に示すように、文レベルの混合分布 N-gram では、まずクラスタ毎に 1 文の確率を計算し、次に混合比を乗じる [9, 10]. すなわち、いずれかのクラスタ N-gram の文の確率が高くなければ、混合分布 N-gram の確率が高い値にはなりにくい. すなわち、文レベルの混合分布 N-gram は、文空間を分割してそれぞれの部分文空間に対して N-gram でモデル化するものであるから、文断片がいずれかの部分文空間の特徴を有していたとしても 1 文全体がいずれの部分文空間にも属さないような文は生成されにくい特長がある. また、各部分文空間に属する文の語彙数は、文空間全体の語彙数のサブセットとなることから、部分文空間の語彙数が少なければ、単語連鎖のパターンも制約され、低い次数の N-gram であっても間接的に文の構造をモデル化できる可能性がある. 従って、タスク・話題・話者の役割・会話の流れなどに応じて定型的表現が頻出する会話音声に対しては、定型的表現の違いに基づき文空間がうまく分割可能であるならば、文レベルの混合分布 N-gram は有効なモデル化手法を提供すると考えられる.

## 2.2 ターゲットトピック N-gram 構築手順

ターゲットトピックの学習データ量が不足する場合の学習データ量を補う方法として、ターゲットトピックのデータに類似タスクのデータを加えて N-gram を構築する手順について説明する.

ここでは、ターゲットトピックのカバー率を向上させることができること、別タスクのデータによる N-gram 推定精度の低下など副作用が生じにくいことを考慮し、加えるデータはターゲットトピックに類似したタスクのデータであることを想定する. しかし、類似タスクであるとは言っても、中には N-gram の推定精度を劣化させるようなデータが混在している可能性もあり、何らかの方法でこのような副作用を生じるデータを除去する必要がある. ここでは、副作用を生じるデータを予め人手で取り除くのは困難であるという観点から、類似タスクデータをターゲットトピックデータとともにクラスタリングし、ターゲットトピックと関

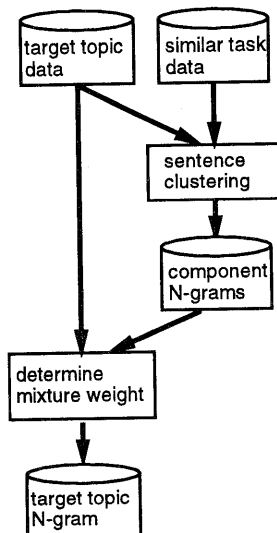


図 1. トピック依存モデル作成の処理の流れ

連性がないクラスタを除去するという方法で副作用を生じるデータを除去するアプローチをとる.

本手法の概略を図 1 に示す.

**手順 1: 学習用セットのクラスタリング** ターゲットトピックデータの学習用セットに類似タスクデータの学習用セットを加えた学習用セット (以下、全学習用セットと呼ぶ) を作成する. 全学習用セットについて、bigram 尤度ベースの距離に基づく文のクラスタリングを行う. 各クラスタ毎にクラスタ bigram を構築し、全学習用セットに含まれる各文を各クラスタ bigram に最も距離が近いクラスタに割り当てる. (bigram 尤度ベースの距離に基づく文のクラスタリング手法は第 3 節で詳述する)

**手順 2: 混合分布 N-gram の混合比の学習** 各クラスタに割り当てられた学習データから N-gram を構築する. 次に、ターゲットトピックデータの学習用セットに対する確率値が最も高くなるような混合比  $\lambda_k$  をターゲットトピックの学習用セットを用いて EM アルゴリズムにより推定する (この段階で、ターゲットトピックに全く関連のないクラスタの混合比は 0、関連性の薄いクラスタは混合比の値が小さくな

る。つまり、副作用を引き起こすようなデータが混合分布 N-gram の確率値に寄与する割合は自動的に小さくなる)。

## 2.3 クラスタの選択に基づき混合分布 N-gram の制約力を変える手法

クラスタリングの過程で、各クラスタに割り当てられた学習データに対する各クラスタ N-gram のエントロピーが求まっている。このエントロピーの値に基づいて N-gram の制約力をコントロールする。具体的には、全クラスタ中からエントロピーの小さいいくつかのクラスタを選択し、選択したクラスタに属する学習データとクラスタ N-gram のみで混合比  $\lambda_k$  を計算する。エントロピーの小さいクラスタのみを選択するので、混合分布 N-gram のエントロピーは、全てのクラスタから作成した混合分布 N-gram のそれより小さくなる<sup>1</sup>。

なお、学習用セットのエントロピーの代わりにテスト用セットのクロスエントロピーを用いれば、クロスエントロピーを下げることも可能になる。

## 3. Bigram 尤度ベースの距離に基づくクラスタ N-gram の自動生成

本節では、未知データに対するクロスエントロピーが小さくなることを目的としたクラスタリング手法について述べる。

学習用セットを分割して各クラスタに含まれるデータのみからクラスタ N-gram を構築する方法では、クラスタの分割に伴い各クラスタに割り当てられる学習データ量が減少する。このため、クラスタの分割による文空間の分割効果よりもクラスタ内に含まれる学習データの量的な減少によるモデル推定精度の劣化が顕著となり、未知データに対する混合分布 N-gram のクロスエントロピーが大きくなってしまふ可能性がある。本論文で提案する方法は、N-gram パラメータ推定用データとは別に未知データを準備し、未知データに対するクロスエントロピーが小さくなるように N-gram を学習するアイデアに基づいている。従来手法と最も大きく異なるのは、クロスエントロピー最小

<sup>1</sup>ここで求めたエントロピーは、選択されたクラスタのみに対するエントロピーで、学習用セット全体のエントロピーの計算とは母集団が異なっていることに注意する必要がある。

規準に基づき、1つのクラスタ N-gram を複数クラスタのデータから学習する点にある。図2に従来法である単一クラスタの学習データからクラスタ N-gram を学習する方法を、図3に提案法である複数クラスタの学習データからクラスタ N-gram を学習する方法を示す。提案法は、クロスエントロピーが最小となるようなクラスタ群を全クラスタの中から選択するので、1つのクラスタ N-gram のクロスエントロピーは全クラスタのデータから学習された N-gram のそれを上回ることは原理上ない(但し、混合分布 N-gram では混合比  $\lambda_k$  が乗じられることから、混合分布 N-gram のクロスエントロピーが、クラスタ数が1の N-gram を上回る可能性はある)。

クラスタリング手法としては、新聞記事を対象とした手法として段落や記事単位から出発し、類似した段落あるいは記事をグループ化してゆくボトムアップ的な方法 [9] (すなわち、1つの段落や記事に含まれる文は同じクラスタに属することになる) と、全ての文を1クラスタとしクラスタを徐々に分割して行くトップダウン的な方法 [11, 12] がある。本手法では、予めタグ付けされていないデータを用いてクラスタリングすること、段落や記事単位ではなく文単位でクラスタリングするのが目的であることからトップダウンクラスタリングを採用した。また、クラスタリングにおける距離尺度としては、bigram 尤度ベースの距離を用いている [13]。

ここでは、クラスタ分割に関する計算量を削減するために、2分割の繰り返しでクラスタを分割した。具体的な手順は以下の通りである。

- (1) 学習用セットの分割 学習用セットを bigram パラメータ推定用データとクラスタ選択用データとにランダムに切り分ける (クラスタ選択データは bigram パラメータ推定用データに対する未知データとして利用する)。
- (2) 初期クラスタ Bigram パラメータ推定用データを1クラスタとする。クラスタに属する文からクラスタ bigram を学習する。
- (3) クラスタの2分割 各クラスタについて以下の処理を行う。  
当該クラスタのクラスタ bigram を用いて当該クラスタに含まれる bigram パラメータ推定用データ各文の単語当たりのエントロピー

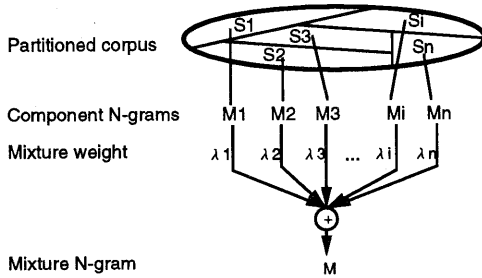


図 2. クラスタ N-gram を単一のクラスタから作成する方法 (従来法)

を求める。1つのクラスタからはエントロピーの低い文を、もう1つのクラスタからはエントロピーの高い文をそれぞれ僅かずつ取り除く。このようにして、1つのクラスタから含まれる文が僅かに異なる2つのクラスタを作成する。各クラスタに属する文からクラスタ bigram を学習する。

(4) クラスタ分割の最適化 上記クラスタを初期値として、準最適なクラスタを求める。

具体的には、bigram パラメータ推定用データ各文に対して、全てのクラスタ bigram に対するエントロピーを求め、エントロピーが最も小さくなるクラスタに割り当てる。クラスタ選択データ各文も同様にクラスタに割り当てる。この段階で、bigram パラメータ推定用データが1文も割り当てられなかったクラスタは消去する。次に、各クラスタに割り当てられたクラスタ選択データを用いて、クラスタ選択データに対する平均クロスエントロピーが最も小さくなるようなクラスタ群を求める(クラスタ群を求めるにあたっては、クラスタ群から作成されるクラスタ bigram のクロスエントロピー減少効果が最も高くなるようなクラスタから順にクラスタ群に加えて行き、この作業をクロスエントロピーの減少が止まるまで繰り返す)。この段階で、クラスタ群が同一となったクラスタは1つのクラスタにまとめられる。この作業を、クラスタ選択データに対する平均クロスエントロピーの値が収束する(すなわち、前回との差分が一定のしきい値以下になる)まで繰り返す。

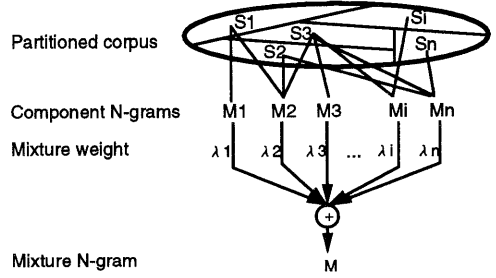


図 3. クラスタ N-gram を複数のクラスタから作成する方法 (提案法)

クラスタ数が予め定められた値に達するまで、(3)、(4)を繰り返す。

## 4. 評価実験

### 4.1 実験条件

実験に使用したターゲットトピックデータは、「通信に関するカスタマケアサービス」タスクの国番号、地域番号の問い合わせ [14](但し、[14]で用いた評価データに比較して表現の自由度は緩和されている) データ 4,738 文、約 8 万単語で、類似タスクデータは、同タスクのターゲットトピック以外のデータ約 25,000 文 (55 万単語) を使用した。

ターゲットトピック・類似タスクともに、80%を学習用データ、残り 20%を評価用セットとした。さらに、学習用セットの 90%を N-gram モデルパラメータ推定用データに、残り 10%をクラスタ選択用データにランダムに切り分けた。クラスタ N-gram の作成には、CMU-Cambridge Toolkit[15]を使用した。

### 4.2 クラスタリング手法の評価

ターゲットトピックと類似タスクの評価データを加えた評価データを用いて、クラスタ分割が進むに従ってモデルのクロスエントロピーの減少が観察されるかどうか評価した。混合分布 N-gram の平均クロスエントロピーの変化を図 4 に示す。提案法 (multi-clusters, 2gram および multi-clusters, 3gram) と比較するために、単一クラスタから bigram を作成する方法 (single cluster, 2gram) とク

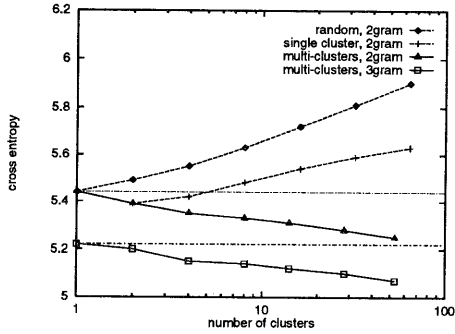


図 4. クラスタ数の増加による平均クロスエントロピーの減少効果

ラスタをランダムに分割し単一クラスタから bigram を作成する方法 (random, 2gram) を併せて示した。単一クラスタから bigram を作成する方法との比較では、クラスタ bigram を複数クラスタの学習データから作成する手法の効果、クラスタをランダムに分割する方法との比較では、クラスタリングによる文空間の分割効果を見ることができる。

図 4 では、提案法の平均クロスエントロピーがクラスタ数の増加に伴い単調に減少することがわかる。クラスタ数が最も大きいクラスタ数 53 の条件では、bigram のエントロピーがクラスタ数が 1 の trigram に近い値を示している。特徴的なのは、trigram でも bigram と同様に大きなクロスエントロピーの減少が観察されることである。この結果は、提案する混合分布 N-gram が文空間を部分文空間に分割することにより、低い次数の N-gram であってもより詳細に文の構造をモデル化可能なことを示している。

これに対して、単一クラスタからクラスタ bigram を作成した場合は、クラスタ数の増加に従って一旦下がったクロスエントロピーが増加に転じている。また、クラスタをランダムに作成した場合は、クラスタ数の増加に伴いクロスエントロピーは単調に増加する。クロスエントロピー増加の原因は、1 つのクラスタ当たりの学習データ量が減少し、クラスタ bigram を十分に推定できなかったのが原因と考えられる。

次に、クラスタ数を最も増やした 53 の場合において、各クラスタにどのようなデータが割り当て

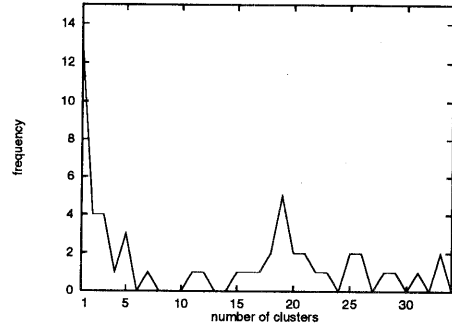


図 5. 1 つのクラスタ N-gram の学習に用いたクラスタ数の頻度分布

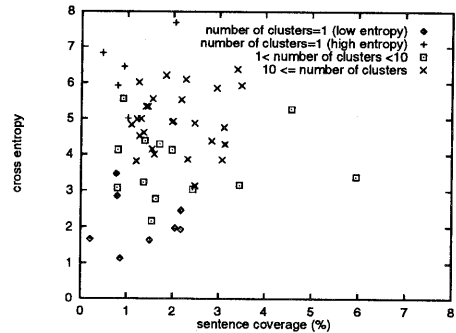


図 6. 各クラスタに割り当てられたテストデータの全テストデータに対する割合と平均クロスエントロピー

られているか、クラスタ trigram のクロスエントロピーと語彙数を評価した。各クラスタ trigram のクロスエントロピーは、評価データを各クラスタに割り当てて求めた。その際、各クラスタ trigram の学習に用いられたクラスタ数と語彙数に基づき 53 のクラスタを以下の 4 つのグループに分類した。

- グループ 1: 単一クラスタから学習されたクラスタ trigram で語彙数の少ないグループ
- グループ 2a: 単一クラスタから学習されたクラスタ trigram で語彙数の多いグループ
- グループ 2b: 10 未満かつ 1 を超えるクラスタから学習されたグループ
- グループ 3: 10 を超えるクラスタから学習されたグループ

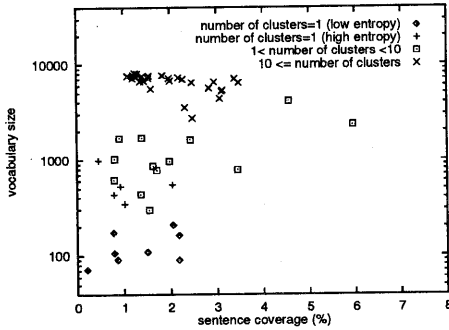


図7. 各クラスタに割り当てられたテストデータの全テストデータに対する割合と N-gram 学習に要した語彙数

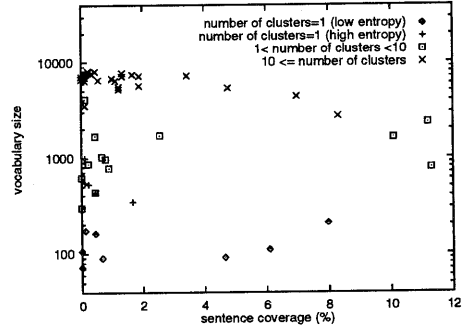


図8. ターゲットトピックにおける各クラスタに割り当てられたテストデータの全テストデータに対する割合と N-gram 学習に要した語彙数

表1. 混合分布 N-gram によるターゲットトピックに対するエントロピー削減効果

model	cross entropy
mixture, 3gram, target	3.75
3gram, target	3.98
3gram, all	3.89

学習に要したクラスタ数を 10 で区切った理由は、N-gram 学習に要したクラスタ数の頻度分布が図 5 に示すようにクラスタ数が少ないグループと多いグループに分かれているため、その境界を 10 と見なしたからである。

各クラスタの平均クロスエントロピーを図 6 に、各クラスタ N-gram の語彙数を図 7 に示す。図 6,7 から、グループ 2a を除く 3 つのグループでは、学習に使用されたクラスタ数が多いほどクロスエントロピーならびに語彙数が大きいこと、割り当てられたテストデータの割合はクロスエントロピーや語彙数にはあまり依存しないことがわかる。

#### 4.3 ターゲットトピックで混合比を学習した混合分布 N-gram の評価

次に、ターゲットトピックの学習用セットで混合比を学習した混合分布 trigram (mixture, 3gram, target) のターゲットトピックの評価用データに対するクロスエントロピーを表 1 に示す、比較のために、ターゲットトピックの学習用セットで作成した

単一クラスタ trigram(3gram, target) と全学習用セットで作成した単一クラスタ trigram(3gram, all) の値も示した。

表 1 より、提案法で作成したモデルが、ターゲットトピックの学習用セットのみを用いて学習した trigram と全学習用セットを用いて学習した trigram のいずれに比較してもクロスエントロピーを低くできることがわかる。この結果は、類似タスクのデータを加えて学習することにより trigram の推定精度が向上可能なことを示している。

次に、各クラスタ trigram のクロスエントロピーと評価データの配分割合に基づき混合分布に使用するクラスタを選択し、制約力を変えたモデルを構築する実験を行った。

クラスタの選択に先立って、ターゲットトピックの学習用データがどのようなクラスタに配分されるか観察した。図 8 に各クラスタに割り当てられた評価データの割合と N-gram 学習に要した語彙数を示す。ターゲットトピックに加えて類似タスクのデータを加えた図 7 の場合と比較すると、ターゲットトピックのデータだけを用いた場合は、特定のクラスタへの集中度が高いことがわかる。この結果は、ターゲットトピックに追加した類似タスクのデータのうち、ターゲットトピックに類似していないデータは、クラスタリングの過程で自動的にターゲットトピックデータが属するクラスタとは別のクラスタに配分されたことを示している。すなわち、提案法は、N-gram の推定精度を劣化させるようなデータを自動的に除去可能なこと

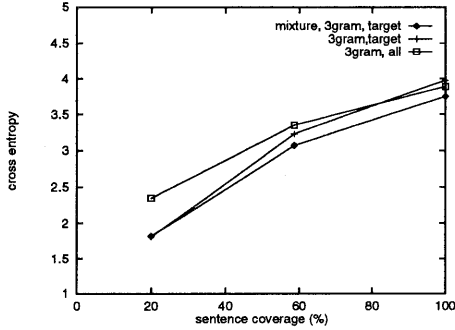


図 9. 使用するクラスタ数を変えた場合の混合分布 N-gram のエントロピー低下効果

を示している。

次に具体的にクラスタを選択し、それらのクラスタに対するエントロピーを求めた。ここで選択したクラスタは、(1)グループ1に属するクラスタのみ、(2)グループ1および2bに属するクラスタ、(3)全クラスタ(表1と同じ条件)の3つの条件である。(1),(2),(3)の順にエントロピーが大きくなる設定である。図9に表1に示した3種類のモデルに対するクロスエントロピーを示す。図9より、いずれの条件においても提案する混合分布 N-gram が最も小さいクロスエントロピーを与えていることがわかる。

## 5. むすび

本論文では、十分な学習データ量が確保できない場合の N-gram モデルの構築法として、文レベルの混合分布 N-gram と類似タスクのデータを加えて学習する方法を組み合わせる手法を提案した。本手法は、bigram 尤度ベースの距離に基づいて混合分布 N-gram を自動生成するため、ターゲットトピックの学習用セットと類似タスクの学習用セットがあれば容易にターゲットトピックに対する N-gram のクロスエントロピーを改善することができる。また、利用するクラスタ数を変えることにより N-gram の制約力を変えることも容易で、クラスタ化しない N-gram にはない特長を持っている。

今後の課題としては、音声認識実験に適用した場合の評価を行うこと、混合分布 N-gram のパラメータ数が多いことから、パラメータ数の削減を

行うことが挙げられる。

## 参考文献

- [1] 古井貞熙, “大語彙連続音声認識の現状と展望,” 日本音響学会公演論文集, 1-6-10, pp.19-22, March 1997.
- [2] P. F. Brown, V.J.D. Pietra, P.V. deSouza, J.C. Lai and R.L. Mercer, “Class-based n-gram Models of Natural Language,” Computational Linguistics, vol.18, no.4, pp.467-479, 1992.
- [3] E.P. Giachin, “Phrase Bigrams for Continuous Speech Recognition,” Proc. ICASSP95, pp.225-228, 1995.
- [4] 森信介, 山地治, 長尾真, “予測単位の変更による n-gram モデルの改善,” 情報処理学会研究報告, SLP19-14, pp.87-94, 1997.
- [5] 政瀧浩和, 松永昭一, 匂坂芳典, “品詞および可変長単語列の複合 N-gram の自動生成,” 信学論 (D-II), vol.J81-D-II, no.9, pp.1929-1936, 1998.
- [6] R. Iyer, M. Ostendorf and M. Meteer, “Analyzing and Predicting Language Model Improvements,” Proc. IEEE Workshop on Automatic Speech Recognition and Understanding, pp.254-261, 1997.
- [7] F. Weng, A. Stolcke and A. Sanker, “Hub4 Language Modeling Using Domain Interpolation and Data Clustering,” Proc. DARPA Speech Recognition Workshop, pp.147-151, 1997.
- [8] R. Kneer and V. Stainbiss, “On the Dynamic Adaptation of Stochastic Language Models,” Proc. ICASSP93, pp.II-586-II-589, 1993.
- [9] R. Iyer, M. Ostendorf and J. Rohlicek, “Language Modeling with Sentence-Level Mixtures,” Proc. ARPA Workshop on Human Language Technology, pp.82-86, 1994.
- [10] R. Iyer, M. Ostendorf and J. Rohlicek, “Modeling Long Distance Dependence in Language: Topic Mixtures vs. Dynamic Cache Models,” Proc. ICSLP96, pp.236-239, 1996.
- [11] P. Clarkson and A. Robinson, “Language Model Adaptation Using Mixtures and an Exponentially Decaying Cache,” Proc. ICASSP97, pp.799-802, 1997.
- [12] Y. Gotoh and S. Renals, “Document Space Models using Latent Semantic Analysis,” Proc. Eurospeech97, pp.1443-1446, 1997.
- [13] 清水徹, 大野見生, 樋口宜男, “文のクラスタリングに基づく統計的言語モデル,” 日本音響学会講演論文集, 1-6-14, pp.31-32, March 1997.
- [14] T. Kato, S. Kuroiwa and N. Higuchi, “Area Code, Country Code, and Time Difference Information System and Its Field Trial,” Proc. IVTTA98, pp.5-9, 1998.
- [15] P. Clarkson and R. Rosenfeld, “Statistical Language Modeling Using the CMU-Cambridge Toolkit,” Proc. Eurospeech97, pp.2707-2710, 1997.