

トピックモデルとタスクの知識を用いた言語モデルによる 野球実況中継の構造化

佐古 淳[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院自然科学研究科 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学工学部 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: tsakoats@me.cs.scitec.kobe-u.ac.jp, {takigu,ariki}@kobe-u.ac.jp

あらまし 本研究では、スポーツ中継、特に野球実況中継に対して構造化を行うことを目的としている。構造化を行うことで検索や要約、番組推薦のための特徴量として用いることが可能である。本研究では、ラジオ実況中継音声の音声認識結果から構造化を行うアプローチを採用する。音声認識の機構に試合状況の推定を取り入れることにより、野球の知識と協調可能な音声認識の提案を行う。これに加え、トピックモデルに基づく言語モデルとの統合を行うことにより、試合状況だけでなく、実況中継における話題の遷移も考慮した認識を行う。実験により、通常の音声認識器を用いる場合に比べ、単語正解精度、キーワードのF値、構造化正解率ともに向上することを確認した。

キーワード 構造化, 言語モデル, 情報検索, スポーツ中継

Structuring Baseball Live Games Based on Topic Model and Task Dependent Knowledge

Atsushi SAKO[†], Tetsuya TAKIGUCHI^{††}, and Yasuo ARIKI^{††}

[†] Graduate School of Science and Technology, Kobe University Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

^{††} Faculty of Engineering, Kobe University Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

E-mail: tsakoats@me.cs.scitec.kobe-u.ac.jp, {takigu,ariki}@kobe-u.ac.jp

Abstract The purpose of this study is to automatically structure sports live speech, especially baseball live speech using Large Vocabulary Continuous Speech Recognition system. Structures are useful for information retrieval, summarization and contents recommendation. In this paper, we propose the situation dependent speech recognition which is the speech recognition method of incorporating the baseball dependent knowledge. Moreover, a language model based on topic model is integrated into proposed speech recognition method. This method enables to seek a word sequence as well as a situation sequence and a topic sequence. The experimental results showed that the proposed approach improved the word accuracy, F-measure of keyword and structuring correct rate.

Key words Structuring, Language Model, Information retrieval, Sports Live Games

1. はじめに

近年、デジタルテレビやWWWなどの発展により、映像や音声など、多くのマルチメディア・コンテンツを所有することが可能となってきた。しかし、このような大

量のコンテンツの中から欲しい情報を探し出すことは容易ではない。潜在的には興味のあるコンテンツであっても、目に触れることがないために知らないまま視聴機会を失う場合もある。コンテンツに対してインデックス情報を付与し、データベースを構築しておくことで、検索

表 1 野球中継のシーケンスが持つ情報

Table 1 Baseball information of each sequence.

構造情報	イニング, アウトカウント, ストライクカウント, ボールカウント
------	--------------------------------------

を行ったり, 要点だけ, または好みのシーンだけの視聴を行うことが可能となる。また, 番組推薦のための情報としても用いることができる。しかし, コンテンツが大量に存在することから, 人手によりインデックス情報を付与することは困難であり, コンピュータが自動的にインデックス情報を付与できることが望まれる。本研究では, マルチメディア・コンテンツとしてスポーツ実況中継, 特に野球実況中継を対象とし, 構造化を行うことを目的としている。

野球中継を構成する要素のイメージを図 1 に示す。野球中継は映像と音声のシーケンスによって構成され, 各シーケンスは表 1 のような情報を保持している。本研究では, 図 1 の情報のうち, 「イニング, アウトカウント, ボールカウント, ストライクカウント」といった情報を, シーケンスに対して連続的に付与していくことを構造化と定義している。

従来, 野球実況中継の構造化手法として, 音声から単語系列とともに試合状況の系列を同時に推定する状況推定音声認識について提案を行ってきた [1]。音声認識結果から状況を推定し, 推定された状況によって言語モデルを切り替えることにより, 野球のルールや試合進行に照らしてあり得ない誤認識を修正することが可能であった。これにより, 構造化を行う上で重要なキーワード正解精度の向上が得られた。

しかし, 一方で, ルールや試合進行に直接関係しない単語については改善が得られていなかった。そのため, 全体としての単語正解精度はほとんど改善されていなかった。本研究では, この点について改善を得るため, トピックモデルを用いることにより, 試合状況だけでなく, 細かな話題の遷移に対する依存まで考慮した音声認識手法の提案を行う。

以下, 次章で提案手法について述べる。この中で, 単語系列と共に状況の系列を推定する音声認識を定式化し, トピックモデルを用いた言語モデル構築手法, 及び状況推定音声認識との統合方法について述べる。その後, 3. 章で実験について述べ, 4. 章でまとめる。

2. 状況推定音声認識

本章では, 従来提案を行ってきた状況推定音声認識について述べ, その後, トピックモデルとの統合を行う。まず, 状況推定音声認識の定式化について述べる。

2.1 状況推定音声認識の定式化

一般的な音声認識は, 観測される音声の特徴系列を

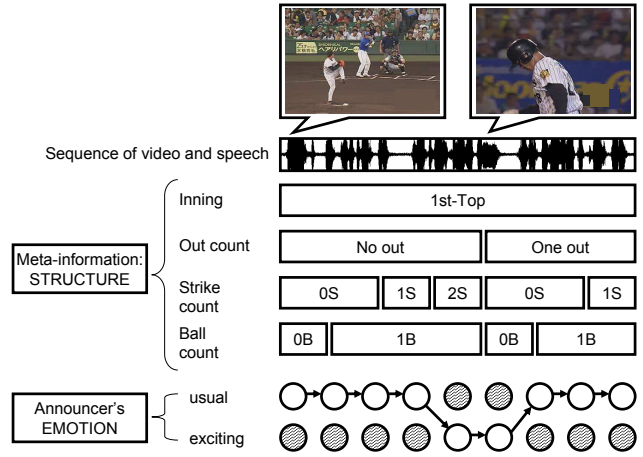


図 1 野球中継の構成要素

Fig. 1 Components of baseball game.

$\mathbf{O} = \{o_1 \dots o_t\}$, 単語系列を $\mathbf{W} = \{w_1 \dots w_N\}$ とすると, 特徴系列 \mathbf{O} に対して, 尤もらしい単語系列 \mathbf{W} を求めることに相当し, 以下のように定式化される。

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} P(\mathbf{W}|\mathbf{O}) \quad (1)$$

$$= \underset{\mathbf{W}}{\operatorname{argmax}} P(\mathbf{O}|\mathbf{W})P(\mathbf{W}). \quad (2)$$

ベイズの定理により式 2 が導かれる。 $P(\mathbf{O}|\mathbf{W})$ は音響モデル, $P(\mathbf{W})$ は言語モデルである。言語モデルは通常, N-gram によって表されるため, 一般的な音声認識は最終的に以下の式で表される。

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} P(\mathbf{O}|\mathbf{W}) \prod_{i=1}^N P(w_i|w_{i-N+1}^{i-1}). \quad (3)$$

次に, 音声認識に知識を用いる枠組みとして本研究で提案する状況推定音声認識の定式化について述べる。状況推定音声認識では, 音声認識に利用したい知識を「状況の系列」として表し, 観測される音声の特徴系列から, 単語系列と状況系列を同時に推定する。本研究では, 「状況」として「ストライクカウント, ボールカウント, アウトカウント, アナウンサーの興奮状況」を用いた。状況系列を $\mathbf{S} = \{s_1 \dots s_N\}$ とすると, 状況推定音声認識は以下のように定式化できる。

$$(\hat{\mathbf{S}}, \hat{\mathbf{W}}) = \underset{\mathbf{S}, \mathbf{W}}{\operatorname{argmax}} P(\mathbf{S}, \mathbf{W}|\mathbf{O}). \quad (4)$$

式 4 は, ベイズの定理により次のように展開できる。

$$(\hat{\mathbf{S}}, \hat{\mathbf{W}}) = \underset{\mathbf{S}, \mathbf{W}}{\operatorname{argmax}} P(\mathbf{O}|\mathbf{W}, \mathbf{S})P(\mathbf{S}, \mathbf{W}). \quad (5)$$

ここで, $P(\mathbf{O}|\mathbf{W}, \mathbf{S})$ は状況に依存した音響モデルと考えられる。特に, 状況に含まれる「アナウンサーの感情」に依存する。また, $P(\mathbf{S}, \mathbf{W})$ は, さらに以下のように展開できる。

$$\begin{aligned}
P(\mathbf{S}, \mathbf{W}) &= P(s_1, \dots, s_N, w_1, \dots, w_N) \\
&= P(w_1)P(s_1|w_1) \\
&\quad \times \prod_{i=2}^N P(w_i|w_1^{i-1}, s_1^{i-1})P(s_i|s_1^{i-1}, w_1^i). \quad (6)
\end{aligned}$$

ここで、近似によって簡単化を行う。まず、 $P(w_i|w_1^{i-1}, s_1^{i-1})$ について、

- 単語は直前の $N-1$ 個の単語と直前の状況にのみ依存

と近似することにより、以下の式を得る。

$$P(w_i|w_1^{i-1}, s_1^{i-1}) \approx P(w_i|w_{i-N+1}^{i-1}, s_{i-1}) \quad (7)$$

これは、状況依存 N-gram と考えることができる。これによって、「フォアボール」という単語はボールカウントが 3 のときに発話されやすい、といったような野球の知識を表現できると考えられる。次に、 $P(s_i|s_1^{i-1}, w_1^i)$ について、

- 状況は直前の状況と現在の単語、及び状況遷移確率を最も高くするような、M 単語以内の現在の単語とのペアの単語

という近似を行う。これにより以下の式を得る。

$$P(s_i|s_1^{i-1}, w_1^i) \approx \max_{y=(i-1 \dots i-M)} P(s_i|s_{i-1}, w_i, w_y) \quad (8)$$

これは、単語のペアに依存した状況遷移モデルと考えることができる。「三振」という発話だけによってアウトカウントの遷移をとらえるよりも、「投げた、三振」のように、単語のペアによって遷移をとらえた方がより精度を高められる。ただし、単語のペアは、隣接するとは限らず、「投げた、直球ストライク、三振」のように、離れて現れることが多いため、M 単語以内の単語とのペアのうち、最も遷移確率が高いものを選択するようにした。

最終的に、状況推定音声認識は以下のように定式化できる。

$$\begin{aligned}
(\hat{\mathbf{S}}, \hat{\mathbf{W}}) &= \underset{\mathbf{S}, \mathbf{W}}{\operatorname{argmax}} P(\mathbf{O}|\mathbf{W}, \mathbf{S})P(w_1)P(s_1|w_1) \\
&\quad \times \prod_{i=2}^N P(w_i|w_{i-1}, s_{i-1}) \\
&\quad \times \max_{y=(i-1 \dots i-M)} P(s_i|s_{i-1}, w_i, w_y) \quad (9)
\end{aligned}$$

一般的な音声認識の定式化と比較して異なる点は以下の通りである。

- 音響モデルが状況に依存する。
- 言語モデルが状況に依存する。
- 発話された単語に依存して状況遷移する状況遷移モデルが新たに存在する。

状況系列とは、音声認識に利用したい「知識」を表現する系列である。状況に依存した音響モデル・言語モデルは、すなわち、知識に依存した音響モデル・言語モデルであると言える。このように、単語と同時に状況を推定することで、知識を用いた音声認識が可能となる。

2.2 トピックモデルとの統合

状況推定音声認識は、試合進行に基づく状況が変化しない限り、音声認識に関するモデルが変化しない。しかし、アナウンサーの発話は、試合進行と無関係の部分でも特徴が変化する。例えば、投球に関する発話の後には、球種や球速についての話題が多く現れる、などである。一方で、このような話題全てにラベルを付与することは困難である。このため、細かい話題の遷移をラベルから学習することはできない。

話題の遷移を自動的にラベルなしで学習する手法として、文書に含まれる単語を特徴量として類似した特徴量を持つ文書をクラスタリングし、クラスター間の遷移確率を求める Stochastic Switching N-gram (SS N-gram) が提案されている [2]。しかし、SS N-gram では、特徴の次元数が大きいことから多くの学習データを必要とし、また、標準的な N-gram との重み付き結合が必要であるなど言語モデルの調整が困難な側面があった。

これに対し、我々は PLSA [3] により文書が持つトピックの割合を推定し、このトピックの割合を特徴量として Ergodic HMM によりトピックのクラスタリングを行い、そのクラスター (HMM 状態) 間の遷移確率を求める手法について提案を行ってきた [5]。これにより、話題の遷移を考慮することができ、細かな言語モデルの調整を行うことなく認識率の向上を実現することができる。本研究では、前節の状況推定音声認識と、トピック HMM を用いた言語モデルとの統合を行う。

まず、次節で PLSA について簡単に説明し、その後 2.2.2 節で状況推定音声認識とトピックモデルとの統合について述べる。

2.2.1 PLSA による言語モデル適応

PLSA [3] はテキストモデリングの一手法であり、文書を潜在トピックの混合によって表現する。すなわち、文書 d における単語 w の発生確率を、潜在トピック z を用いて、

$$P(w|d) = \sum_z P(w|z)P(z|d) \quad (10)$$

のように表現する。PLSA では、EM アルゴリズムによってパラメータを学習するため、bi-gram または tri-gram 確率を学習するには莫大な計算量を必要とする。そのため、近似的に N-gram 確率を推定する手法として、unigram スケーリングによる言語モデル適応が提案されている [4]。 $P(w_i|w_{i-1}w_{i-2})$ の適応後の確率は以下の式によって求められる。

$$P(w_i|w_{i-1}w_{i-2}) \propto \frac{P(w_i|d)}{P(w_i)} P(w_i|w_{i-1}w_{i-2}) \quad (11)$$

2.2.2 トピック HMM

トピックモデルを用いた言語モデル構築手法の概要を図 2 に示す。まず、野球実況中継コーパスに対して PLSA

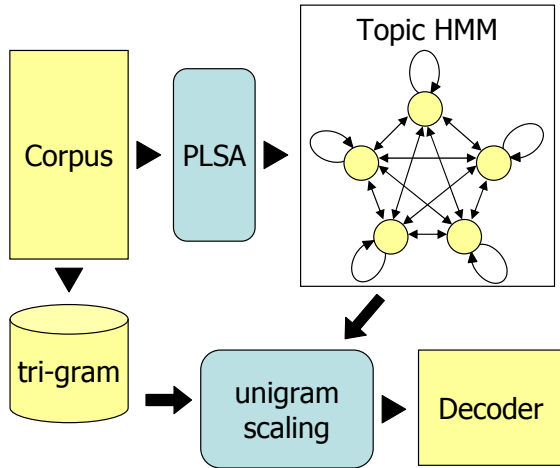


図2 トピック HMM を用いた言語モデル適応の概要.
Fig.2 Language model adaptation using topic HMM.

を行い、各潜在トピックにおける単語の unigram 確率 $P(w|z)$ 及び、発話毎の潜在トピック重み $P(z|d)$ を求める。1 発話は 1 つの句点で区切られた範囲を基本とした 10 ~ 20 単語程度で構成されている。学習コーパス中の全発話数は約 8 千発話であった。

次に、各発話の潜在トピック重み $P(z_1|d) \cdots P(z_C|d)$ をベクトル化したものを特徴量として HMM を学習する。HMM は ergodic HMM とする。ただし、 C は潜在トピックの数である。野球中継において、状況・話題はインニング毎に初期化されるものと考えられる。そこで、発話をインニング毎に分割し、1 インニング 1 サンプルとして HMM 学習に用いた。初期段階では、HMM のすべての状態に同じ分布を与えておく。学習により、似た状況・話題についての発話が HMM の状態に偏って現れる。すなわち、HMM の各状態が、類似トピックのクラスターとなる。各状態は、クラスター内に含まれる文書のトピック分布の平均を持つ(図 3)。状態 k の平均トピック分布を $P(z_1|k) \cdots P(z_C|k)$ で表す。ただし、学習後、 $\sum_{i=1}^C P(z_i|k) = 1$ となるように正規化する。HMM の状態遷移確率は、話題の遷移確率と捉えることができる。

以上のようなトピック HMM と状況推定音声認識の統合を行う。統合の目的は、試合状況に関する発話については状況推定音声認識の枠組みを用い、試合状況に変化がない場合にはトピック HMM による言語モデル適応の枠組みで認識を行うことである。認識のイメージとしては、第 1 パスとして通常の音声認識を行い仮説を出力する。次に、第 2 パスとして、トピック HMM による言語モデルを用いてリスコアリングを行う。さらに、第 3 パスとして、状況推定音声認識の枠組みを用いてリスコアリング、及び状況推定モデルを用いて状況の推定を行う。これらの目的を満たすよう、状況依存音声認識において、

話題系列 $\mathbf{K} = (k_1, \dots, k_t)$ として、以下のように定式化を行う。

$$\begin{aligned}
 P(\mathbf{S}, \mathbf{W}) &= \sum_{\mathbf{K}} P(\mathbf{S}, \mathbf{W}, \mathbf{K}) \\
 &= \sum_{\mathbf{K}} P(s_1, \dots, s_N, w_1, \dots, w_N, k_1, \dots, k_N) \\
 &\approx \prod_{i=1}^N P(k_i | k_{i-1}, s_{i-1}) P(w_i | w_{i-N+1}^{i-1}, k_i, s_{i-1}) \\
 &\quad \times P(s_i | s_{i-1}, w_{i-M+1}^i). \tag{12}
 \end{aligned}$$

ここで、 $P(k_i | k_{i-1}, s_{i-1})$ は、状況 s_{i-1} の下でのトピック遷移確率であり、トピック HMM における状態遷移確率を用いる。 $P(w_i | w_{i-N+1}^{i-1}, k_i, s_{i-1})$ は状況と話題に依存した言語モデルである。これは、状況毎にトピック HMM を学習し、そのトピック HMM から unigram rescaling による言語モデル適応によって推定する。式 11 より、

$$P(w_i | w_{i-1}^{i-2}, k_i, s_{i-1}) \propto \frac{P(w_i | k_i, s_{i-1})}{P(w_i)} P(w_i | w_{i-1}^{i-2})$$

と表せる。状況と話題の両方に依存することから、状況・話題依存言語モデルと呼ぶものとする。次節で学習方法について述べる。

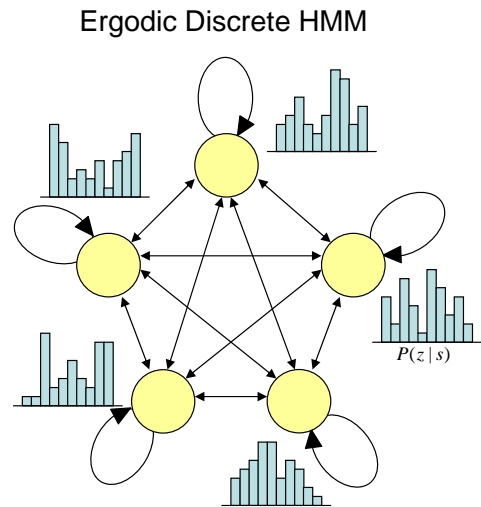


図3 トピック HMM.
Fig.3 Topic HMM.

2.3 確率モデルの学習

本節では、状況推定音声認識で用いる各確率モデルの学習方法について述べる。従来、状況依存音響モデルについても研究を行ってきたが、本稿では簡単のため省略し、状況に依存しない通常の音響モデルを用いるものとする。以下、状況依存言語モデル、及び、状況遷移モデルについて述べる。

2.3.1 状況・話題依存言語モデル

本節では、状況・話題依存言語モデルの学習法について述べる。従来、状況依存言語モデルは、状況への依存が強い単語のみ確率モデルを学習し、そうでない単語については通常の言語モデルにおける N-gram 確率をそのまま用いていた。これは、状況依存のコーパスで学習を行うと学習サンプル数が小さくなってしまいうため、状況に依存しない単語については、多くのサンプルを用いて学習した通常の N-gram 確率の方が推定精度が高いためであった。

本研究では、上記のように単語毎に特別にモデルを切り替えたりせず、単純に、状況毎にトピック HMM による言語モデルを構築する。unigram rescaling によって通常の N-gram に対して適応を行うことから状況毎の学習サンプルが少なくなっても、極端に偏ったモデルにはならないと考えられる。

まず、全ての状況を含む学習コーパスに対して PLSA を行い、その後、状況毎にトピック HMM の構築を行う。最終的には、このトピック HMM から unigram rescaling によって状況と話題に依存した言語モデルが得られる。

2.3.2 単語依存状況遷移モデル

本節では、単語に依存した状況遷移モデルについて述べる。単語依存の状況遷移モデルは $\max_{y=(i-1 \dots i-M)} P(s_i | s_{i-1}, w_i, w_y)$ という式で表される。状況遷移のイメージを図 4 に示す。まず、野球のルール上の制約により、ストライクカウントとボールカウントが同時に増えたり、一度に 2 つ増えたりすることはない。また、アウトになったり、フォアボールになったりすると、次の打者となり、カウントは 0-0 に戻る。状況遷移確率のうち、野球のルールが許さない遷移については、単語の依存を考慮せずに確率値ゼロを与えている。

次に、野球のルールが許す遷移について、その遷移が起こるときにアナウンサーがどのような発話を行っているかを考察する。例えば、ボールカウントが増加するような遷移においては、確かに「ボール」という発話が行われるが、「ボール」という単語は、「ボールが高く上がった」のような発話の中でも現れる。このことから、単純に単語ひとつを見て状況を遷移させることは出来ないと考えられる。本来は、 $P(s_i | s_{i-1}, w_i \dots w_{i-M})$ のように過去 M 単語の系列によって状況を遷移させることが望ましい。しかし、これではデータスパースネスの問題が生じてしまう。そこで、過去 M 単語の中から、現在の単語との組み合わせで、最も状況遷移確率が大きくなる単語を選択する。これにより、「投げた、アウトコース、ボール」のような発話と「ボールが高く上がった」のような発話を区別し、状況遷移の精度を高めることができる。

状況遷移確率は以下の手順で計算する。学習データとして、実況中継の書き起こしテキストに、状況が遷移す

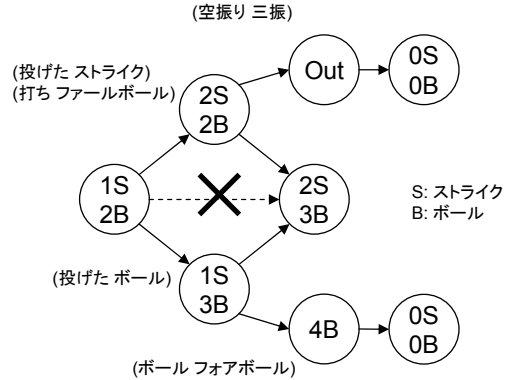


図 4 状況遷移モデル

Fig. 4 State transition model.

表 2 状況遷移を引き起こしやすい単語ペア

Table 2 Word pairs tend to raise state transition.

状況遷移	w_y	w_i
アウト	空振り	三振
	取り	アウト
ストライク	投げ	ストライク
	直球	ストライク
ボール	アウトコース	ボール
	第一球	ボール
フォアボール	投球	フォアボール
	ボール	フォアボール
ファールボール	投げ	ファール
	打ち	ファールボール

ると考えられる場所にタグを付与したものをを用いる。単語ペアに依存した状況遷移確率は以下の式で求められる。

$$P(s_i | s_{i-1}, w_i, w_y) = \frac{N(w_i, w_y, s_i | s_{i-1})}{N(w_i, w_y)} \quad (13)$$

ここで、 $N(x, y)$ は単語 x と y が M 単語の距離内に出現する回数である。表 2 に、状況遷移を引き起こす確率が高いと計算された単語のペアを示す。

3. 実験

提案手法を用いて音声認識、及び構造化実験を行った。まず、実験条件について述べる。

3.1 実験条件

ベースラインの音響モデルは、日本語話し言葉コーパス (CSJ: Corpus of Spontaneous Japanese) モニター版 [6] のうち、男性話者 200 名の講演音声を用いて作成した。音響分析条件と HMM の仕様を表 3 に示す。これらの条件で音響モデルを作成し、さらに、MLLR+MAP [7] により音響モデル適応を行った。音響モデル適応は、テストセットと同一話者の音声データを用いて行った。適応データの分量は、約 1 時間半であった。

言語モデルは、野球実況中継音声の書き起こしテキス

表 3 音響分析条件と HMM の仕様

Table 3 Condition of acoustic analysis and HMM specification.

音響分析	サンプリング周波数	16kHz
	特徴パラメータ	MFCC(25 次元)
	フレーム長	20ms
	フレーム周期	10ms
	窓タイプ	ハミング窓
H	タイプ	244 音節
M	混合数	32 混合
M	母音 (V)	5 状態 3 ループ
M	子音+母音 (CV)	7 状態 5 ループ

表 4 キーワード一覧

Table 4 Keyword list.

ストライク, ボール, ファールボール, フォアボール, アウト, 空振り, 三振
--

トから tri-gram モデルを作成した。書き起こしテキストは、話し言葉を書き起こしたものである。異なり単語数は約 3,000, コーパスサイズは約 8 万形態素であった。

認識の手順は、2.2.2 節で述べた通り、第 1 パスとして通常の音声認識を行い仮説を出力し、次に、第 2 パスとして、トピック HMM による言語モデルを用いてリスコアリングを行った。その後、第 3 パスとして、状況推定音声認識の枠組みを用いてリスコアリング、及び状況推定モデルを用いて状況の推定を行った。第 1 パスのデコーダーには Julius [8] を用いた。

3.2 実験結果

単語正解精度、キーワードの F 値、構造化正解率を表 5 に示す。キーワードには、構造化にとって重要であると考えられる表 4 のような単語を用いた。構造化正解率は、ピッチャーの投球毎に構造情報が正解している割合である。通常の音声認識、及びトピック HMM に基づく言語モデルを用いた音声認識では、状況の推定機構が存在しないため、構造化正解率は求めている。

表より、提案手法による精度の向上が確認できる。ただし、単語正解精度についてはトピック HMM と提案手法の差は見られない。状況推定音声認識では、試合状況に関係のある単語についてしか改善が得られないため、全体の単語数に対して、少ない単語の改善しか得られないためである。ただし、キーワードの F 値については、改善を得ることができた。これに伴い、構造化正解率についても改善が得られた。

4. まとめ

本研究では、野球実況中継の構造化を目的として、状況と話題を推定音声認識を定式化し、実験を行った。実験結果より、従来の音声認識手法よりも単語正解精度・キーワードの F 値が向上し、また、構造化正解率も向上

表 5 実験結果

Table 5 Experimental results.

	単語 ACC	KEY F 値	構造化
通常音声認識	65.0%	74.4 %	-
状況音声認識	65.1%	74.9 %	70.3%
トピック HMM	65.7%	75.2%	-
提案手法	65.7%	75.5%	72.8%

することが確認できた。今後の課題として、状況推定を高精度化することがあげられる。

文献

- [1] 佐古 淳, 有木 康雄, “知識を用いた音声認識による野球実況中継の構造化,” 第 6 回音声言語シンポジウム, SP2004-136, pp.85-90, 2004-12.
- [2] T. Nagano, M. Suzuki, A. Ito and S. Makino, “Language Modeling using Stochastic Switching N-gram”, in Proceedings of the 18th International Congress on Acoustics, V, pp.3697-3700, 2004.
- [3] T. Hofmann, “Probabilistic Latent Semantic Analysis”, in Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99), 1999.
- [4] D. Gildea and T. Hofmann, “Topic-Based Language Models Using EM”, in Eurospeech'99, pp.2167-2170, 1999.
- [5] 佐古 淳, 滝口 哲也, 有木 康雄, “PLSA に基づくトピック HMM を用いた言語モデル構築の検討,” 日本音響学会平成 18 年度春季研究発表会, 1-P-23, pp.189-190, 2006-03.
- [6] 古井 貞熙, 前川 喜久雄, 伊佐原 均, “『話し言葉工学』プロジェクトのこれまでの成果と展望,” 第 2 回話し言葉の科学と工学ワークショップ, pp.1-6, 2002.
- [7] 緒方 淳, 有木 康雄, “音素事後確率に基づく信頼度を用いた音響モデルの教師なし適応,” 信学技報, SP2001-105, 2001.
- [8] A. Lee, T. Kawahara, K. Takeda, M. Mimura, A. Yamada, A. Ito, K. Ito and K. Shikano, “Continuous Speech Recognition Consortium — An Open Repository for CSR Tools and Models —”, in Proc. IEEE Int'l Conf. on Language Resources and Evaluation, 2002.