

# 平成 27 年度 卒業研究論文

題目 一球速報と実況音声認識を用いた  
野球映像の自動タギングに関する研究

指導教員 服部 峻

提出者 室蘭工業大学 情報電子工学系学科

氏名 荒澤 孔明

学籍番号 12024006

提出年月日 平成 28 年 2 月 12 日

# 目次

第1章	まえがき	1
第2章	提案手法	3
2.1	システム概要	3
2.2	一球速報を用いたイベントの取得	5
2.3	実況音声認識	5
2.4	タギングアルゴリズム	7
2.4.1	イベント開始時間の算出	7
2.4.2	イベント終了時間の算出	8
2.5	野球試合モデルに依る推定イベント開始時間の算出	10
第3章	システム評価と今後の課題	13
3.1	野球試合モデルの精密化に関する評価と課題	15
3.2	各推定方法に依るタギング精度に関する評価と課題	16
3.3	各パラメータに関する評価と課題	20
第4章	むすび	24
	謝辞	25
	参考文献	26

# 目次

2.1	システム概要図 . . . . .	4
2.2	イベント $E_i$ のイベント開始時間 $T_i$ の算出方法 . . . . .	7
2.3	推定方法 (iii) に基づいたイベント $E_i$ のイベント時間 ( $T_i$ から $T'_i$ まで) の算出 . . . . .	9
3.1	$\Delta t_1$ と $\Delta t_2$ の変化に依る F 値 . . . . .	21
3.2	$\Delta t_1$ と $\Delta t_2$ の変化に依る二乗誤差 . . . . .	21
3.3	$\Delta t_1$ の変化に依る二乗誤差 . . . . .	23
3.4	$\Delta t_2$ の変化に依る F 値と二乗誤差 . . . . .	23
3.5	$\Delta t_s$ に依る F 値と二乗誤差の変化 . . . . .	23
3.6	$\Delta t_p$ に依る F 値と二乗誤差の変化 . . . . .	23
3.7	$w_l$ に依る F 値と二乗誤差の変化 . . . . .	23
3.8	$w_r$ に依る F 値と二乗誤差の変化 . . . . .	23

# 表目次

3.1	推定イベント開始時間のみに基づくタギング精度 . . . . .	15
3.2	各推定方法に依るタギング精度 (データ 1) . . . . .	17
3.3	各推定方法に依るタギング精度 (データ 2) . . . . .	17
3.4	各推定方法に依るタギング精度 (データ 3) . . . . .	17
3.5	各推定方法に依るタギング精度 (データ 4) . . . . .	17
3.6	各推定方法に依るタギング精度 (データ 5) . . . . .	17
3.7	各推定方法に依るタギング精度 (5 試合平均) . . . . .	18

# 第 1 章

## まえがき

スポーツ報道番組等ではハイライト映像がよく使用される。その映像の多くは番組側で制作されているため、同じ試合のハイライト映像でも番組によって異なったシーンがハイライト映像として使用される。ハイライトシーンはその試合の見どころでもあることから、試合を全て見られなかった人達に対して、短時間で試合を楽しんでもらうのがハイライト映像を制作する目的とも言えるだろう。すなわち、番組側は多数の視聴者が見たいと思うシーンを推測しながらハイライト映像を制作する。しかしながら、視聴者が見たいと思うシーンは個々人で異なり、それら全てのシーンをハイライト映像に含めるわけにはいかない。例えば、ある視聴者が選手 A を応援していて、その選手 A のプレーシーンをハイライト映像として視聴したいと思っていたとする。しかし、仮にその試合で選手 A にあまり目立った活躍が無かった場合、番組側が制作するハイライト映像で選手 A のプレーシーンが使用されることはおそらく無いだろう。つまり、このハイライト映像はその視聴者のニーズには応えられなかったということになる。従って、番組側が制作するような多数の視聴者の一般的なニーズ予測に基づいたハイライト映像では、個々の視聴者のニーズに全て応えるということは不可能であると言える。

そこで、視聴者自身でハイライト映像を制作することができれば、前述した、番組側が制作するハイライト映像の課題について解決できるのではないかと考えた。視聴者自身がハイライト映像を制作する手段の一つとして、「視聴者が予めスポーツ中継番組を録画し、自分が見たいシーンだけを集めて編集する」等が挙げられる。しかしながら、これは映像全体に対して、見たいシーンが存在した場合そのシーンを編集し、それ以外のシーンは早送りにする等の作業を行うというのが一般的であるため非常に手間がかかる。

では、もし、その試合映像がシーン毎にチャプター分割されていたらどうだろうか。チャプターとは、試合映像の各シーンに見出しを付け、見たいシーンへ直接移動できる機能のことである。前述した録画した映像から視聴者自身が見たいシーンを探すという作業も、既にシーン毎にチャプター分割されているため、映像を全て見る必要が無くなり、視聴者はシーンを選抜するだけで良いことになる、つまり余計な手間をかけずに済む。すなわち、予めチャプター分割されたスポーツ映像を視聴者に提供することができれば、各々が見たいシーンを容易に視聴することが可能になるのではないかと考えた。そこでスポーツ映像を自動でシーン毎にチャプター分割するシステムを提案する。映像をチャプター分割するためには「その映像内でどのよ

うな出来事が起こったか」, 及び, 「その出来事はいつ起こったか」という二つの情報が必要になる. 本研究では野球映像の自動チャプター分割を目指し, 一球速報の Web テキストと実況音声認識を用いて, 「その映像内でどのような出来事が起こったか」, 及び, 「その出来事はいつ起こったか」という二つの情報の取得手法を提案する. また, 本研究で目指すチャプター分割はタギング (タグ付け) [1] と呼ばれる分割手法で, これは, 単に見出しを付加したシーン毎に分割するのではなく, タグ情報と呼ばれる, そのシーンで起こった詳細な情報も付加したシーン毎に分割する手法のことを言う. 本提案システムは, このタグ情報をシステム利用者に提供することで, それらのタグ情報に基づいたシーンの選抜を可能にする.

## 第 2 章

# 提案手法

### 2.1 システム概要

映像をシーン毎にタギングするには基本的に「その映像内でどのような出来事が起こったか」、及び、「その出来事はいつ起こったか」という二つの情報が必要になる。野球映像のタギングでも同様にこれら二つの情報が必要になるが、それらの情報の細かさ、つまり、タギングをする際の分割の細かさも考慮する必要がある。野球映像をタギングする場合、一球毎に「いつからいつまでが〇〇投手が何球目を投げたシーン」等のようにタギングするケースや、イニング毎に「いつからいつまでが先攻チームの攻撃シーン」等のようにタギングするケースといった様々なタギングケースが考えられるが、本稿ではタギングをする際の分割の細かさを打者毎に定め、「いつからいつまでが打者〇〇の打席で、どのような結果になったシーン」のようなタギングを目指す。ここで、映像のタギングに必要な二つの情報のうち、「その映像内でどのような出来事が起こったか」というフィールドをイベント、「その出来事はいつ起こったか」というフィールドをイベント時間と言う。

イベントの取得には、一球速報 [2] と呼ばれる Web テキストを利用する。本研究で目指すタギングは打席シーン毎の分割であるため、イベント  $E_i$  ( $i = 1, 2, \dots, N$ ) も打席シーン毎に取得する。これを図 2.1 の Step 1 に示す。また、あるイベント（打席シーン）について、そのイベントの打者が打席に入った時の状況とその打者名、及び、その打席シーンの結果をそのイベントの内部要素とする。イベントの取得手順としては、タギング対象の試合が終了した後、一球速報を用いて打者が変わる毎、つまり打席シーン毎に取得する。取得するイベント  $E_i$  の例に「2 アウト 1 塁の状況で、第  $\Delta$  打席目に〇〇選手がホームランを打った」を挙げる。但し、同じ選手が複数回打者として登場したシーンであっても違うイベントとして扱う。

次に、これらのイベントにイベント時間を付加していくのだが、ここではイベント時間について、まず、全イベント  $E_i$  に対してイベント開始時間  $T_i$  ( $i = 1, 2, \dots, N$ ) を付加する。これを図 2.1 の Step 2 に示す。次に、イベント  $E_i$  に対して、次のイベント  $E_{i+1}$  のイベント開始時間  $T_{i+1}$  をイベント  $E_i$  のイベント終了時間とすることで、イベント時間が算出される。イベント開始時間の算出方法として、一般的な野球中継番組で行われている実況放送を利用できないかと検討した。野球中継番組では一般的に、実況者、及び、解説者の音声も放送される。

その会話内容には選手の特徴、プレーの実況と解説、次の展開の予測などが含まれる。その中で、ある打者が打席に入っていることを示す実況に着目した。ある打者が打席に入っていることを示す実況とは、例えば「ここでバッター〇〇です」のような実況である。そこでプロ野球中継番組から2試合を参考にして、登場した打者延べ139人に対し、打席シーンの中で、その打者が打席に入っていることを示す実況が少なくとも一つ以上存在したか否かという調査を行った。ある打者の打席シーンとは、その打者が打席に立った瞬間から、出塁した、またはアウトを宣告された瞬間までのことを言う。調査結果より、ある打者の打席シーン中、その打者が打席に入っていることを示す実況が少なくとも一つ以上存在した確率は83%であった。

そこで、このようなある打者が打席に入っていることを示す実況を利用して、各イベントにイベント開始時間、及び、イベント終了時間を付加する、以下の手法を提案する。例えば、1番打者A、2番打者Bに対して、実況者はAが打席に入る瞬間に「1番のAです」、Bが打席に入る瞬間に「次はBです」と実況したとする。このようなケースで「1番のAです」と実況された瞬間をAの打席シーン、つまり、Aのイベントの開始時間、また、「次はBです」と実況された瞬間をBの打席シーン、つまり、Bのイベントの開始時間、及び、Aのイベントの終了時間とすることでイベント時間を算出する。このプロセスでイベント時間を付加した複数のイベントを生成することで、タグ情報を含む打席シーン毎のタギングを実現する。

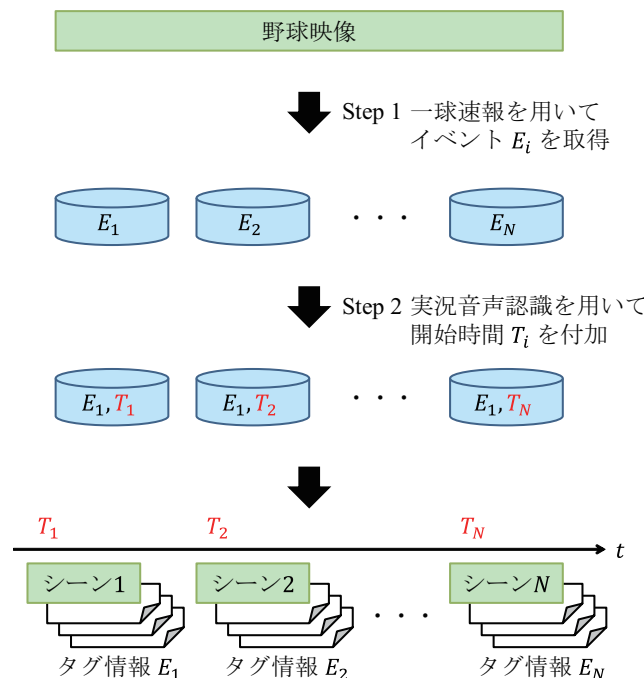


図 2.1 システム概要図



## 2.2 一球速報を用いたイベントの取得

本システムでは、Yahoo! JAPAN スポーツナビが提供する Web サイトである一球速報 [2] を用いてイベントを取得する。一球速報からはプロ野球を中心とした野球の試合のリアルタイム速報、及び、試合後の情報等を取得することができる。本稿では、タギング対象の試合が終了した後に、この一球速報の Web テキストを用いて打者が切り替わる毎にイベントを取得し、これらのイベントをタグ情報とする。

一般的なハイライト映像には含まれないようなシーンを視聴したいと思うユーザに対して、そのシーンを提供することが本システムの目的の一つであるため、シーンを選抜させる際にはより詳細なタグ情報を付加したシーンを提供する必要がある。例えば、あるユーザが選手 A のプレーした全てのシーンを視聴したいと思っている場合、ユーザはタグ情報に選手名「A」が含まれるシーンを選抜することになる。この場合、システムは守備や走塁等の選手 A の打席以外のシーン、また、選手 A の関与が極僅かであったシーンも、そのシーンに選手 A がどのように関与したかを示すタグ情報をユーザに提供しなければならない。以上のことから、一球速報の Web テキストを用いて、イベントとしてある打者が打席に入った時の状況とその打者名、及び、その打席シーンの結果を取得するのに加え、その打者の打席シーン中に、走塁、守備のいずれかに関与した選手、また、関与した内容も取得する。イベントの例として、「ある打者〇〇が 2 アウト 2 塁の状況で第 3 打席に立ち、サードゴロを打ちアウトを宣告された、そのシーンには、投手として選手□□、1 塁走者として盗塁を決めた選手××、打球処理をした三塁手として選手△△が関与した」を挙げる。

## 2.3 実況音声認識

各イベントにイベント開始時間を付加するために、野球中継番組の実況音声認識を利用する。実況音声を認識するために音声認識ソフトとして AmiVoice [3] を使用する。AmiVoice には、マイクまたは音声ファイルから音声認識した結果をテキストとして出力する機能、AmiVoice 搭載の辞書にユーザ指定の単語を新語として追加登録する機能がある。本研究では、音声認識する際のノイズをより少なくするため、一試合の実況音声を音声ファイルに変換して認識する。また、音声認識する際の辞書には AmiVoice 搭載の標準大汎用音響モデルを使用する。

野球中継における、周囲の歓声や場内アナウンスは目立ったノイズとなり、イベント開始時間を付加するための「ここでバッター〇〇です」のような、ある打者が打席に入っていることを示す実況を正確に認識することは困難であった。そこで「ここで〇〇です」や「〇〇が打席に入りました」等のある打者が打席に入っていることを示す実況に含まれる「選手名」に着目した。選手名の認識は、ある打者が打席に入っていることを示す一文の認識よりも認識率は高くなることが期待される。

本来、選手 A の打席シーンについて「ここで打席に A が入ります」のような一文を試合全体の実況音声認識結果から取得し、その実況時間を選手 A のイベント開始時間とする手法を

検討していた。しかし、その一文を正確に認識することができなかつたため、試合全体の実況音声認識結果から選手名「A」のみの取得に変更したところ、「次のバッターはAです」や「Aが守備でファインプレーをしました」等あらゆる場面で実況された選手名「A」が取得された。このことから、試合全体の実況音声認識結果から選手名「A」を取得することで、選手名「A」が複数出現することを想定し、それら複数の「A」の中で、本来認識することを目指していた「ここで打席にAが入ります」のような打者「A」が打席に入っていることを示す実況に含まれる選手名「A」を推測する手法を提案する。

- 実況ポイント：実況音声認識で打者名が出現した瞬間のことを言う。また、 $E_i$ の打者名の実況ポイントを出現した順に $P(i, 1), P(i, 2), \dots, P(i, j)$ とする。但し、その打者名が打者以外のシーンで実況された場合も $P(i, j)$ に含まれる。
- 打席実況ポイント：上記の実況ポイントのうち、打席に入っていることを示す実況に含まれる実況ポイントのことを言う。但し、各イベントの打席シーン中で1回のみ出現とは限らない。

選手名が正確に音声認識されるかどうかは、認識する際の辞書に影響されると考えた。AmiVoice 搭載の辞書には、佐藤、高橋など一般的な苗字は登録されているのに対し、珍しい苗字や、外国人の名前は登録されていないと推測した。ここで推測したと述べたのは、AmiVoice 搭載の辞書における登録単語の詳細については確認できなかったためである。従って、AmiVoice 搭載の辞書へ選手名を新語として追加登録することを検討した。単語登録方法としては、一球速報を用いて取得したイベントに基づき、その試合でプレーした選手のみを抜粋し、それらの選手名を記述文法を用いた新語に変換して辞書へ追加登録する。ここで記述文法とは「実況者が選手名を実況する際によく使う言い回し」のことを言う。例えば、ある選手名「A」を音声認識させる際、単に「A」という選手名を辞書へ追加登録するだけでは、野球映像に含まれる周囲のノイズによって高い認識率は期待されない。そこで、実況者が選手名を実況する際によく使う言い回しである、「○番のA」「ピッチャーのA」等のような複数の単語の集合を一つのフレーズとして辞書へ追加登録する。この記述文法を用いた辞書登録によって選手名の高い認識率が期待される。但し、記述文法で用いる打順、及び、ポジションも一球速報を用いて取得する。以上のように選手毎に記述文法が用いられた新語を生成し、それらの新語をAmiVoice 搭載の標準大汎用音響モデルに追加登録する。本研究では、それらの新語を追加登録した辞書を用いて実況音声を認識する。

## 2.4 タギングアルゴリズム

### 2.4.1 イベント開始時間の算出

イベント開始時間の算出には、野球中継の実況音声認識を用いた。本研究では、実況音声を認識し、「ここで〇〇です」のようなある打者が打席に入っていることを示す実況が存在した時、その実況時間がその打者のイベント開始時間となるのではないかと仮説を立てた。この仮説に基づき、複数の実況ポイントの中から「ここで〇〇です」のようなある打者が打席に入っていることを示す実況を推測し、イベント開始時間を算出する手法を提案する。

まず、イベント  $E_1$  については例外的にイベント開始時間  $T_1$  が算出される。本提案システムはテレビレコーダへの搭載を想定しているため、番組録画した時刻を認識し、また、一球速報を用いて試合開始時刻を取得することで、録画した映像の中での試合開始時間を算出し、その試合開始時間をイベント  $E_1$  のイベント開始時間  $T_1$  とする。

続いて、イベント  $E_2$  以降のイベント開始時間の算出方法を述べる。イベント  $E_i$  について、複数の実況ポイント  $P(i, j)$  が存在する時、その中からイベント  $E_i$  の打者が打席に入っていることを示す打席実況ポイントを推測する必要がある。よって、まず、イベント  $E_i$  が映像内のどの辺りで起こるのかを推定イベント開始時間  $\hat{T}_i$  ( $i = 1, 2, \dots, N$ ) として算出する。また、本研究では、イベント  $E_i$  のイベント開始時間  $T_i$  となる打席実況ポイントは、推定イベント開始時間  $\hat{T}_i$  の付近にあると仮定した。そこでパラメータ  $\Delta t_1$  (分) と  $\Delta t_2$  (分) を用い、 $\hat{T}_i + \Delta t_1$  から  $\hat{T}_i + \Delta t_2$  までの近傍領域を探索し、その領域内で最初に出現した実況ポイント  $P(i, j)$  を打席実況ポイントと推測することで、その打席実況ポイントをイベント開始時間  $T_i$  とする。但し、領域内に実況ポイント  $P(i, j)$  が存在しなかった場合は、推定イベント開始時間  $\hat{T}_i$  をそのままイベント開始時間  $T_i$  とする。図 2.2 にはイベント  $E_i$  の打者の実況ポイントを「●」とした時、また、その試合に出場した他の選手名の実況ポイントを「▲」とした時のイベント開始時間  $T_i$  の算出方法を示す。

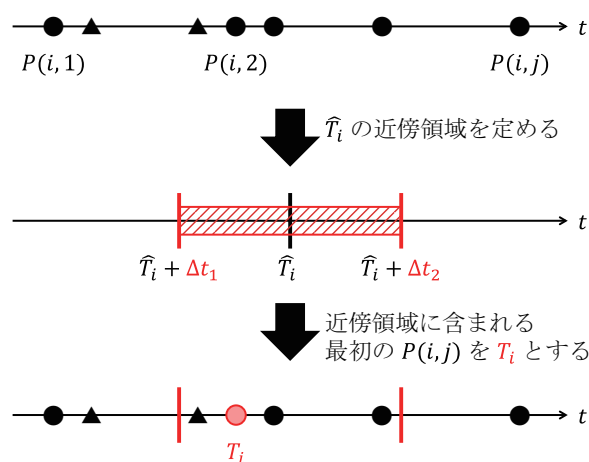


図 2.2 イベント  $E_i$  のイベント開始時間  $T_i$  の算出方法

## 2.4.2 イベント終了時間の算出

全てのイベント  $E_i$  にイベント開始時間  $T_i$  を付加した後、それらのイベント開始時間  $T_i$  に基づき、各イベント  $E_i$  にイベント終了時間  $T'_i$  ( $i = 1, 2, \dots, N$ ) を付加する。イベント  $E_i$  の終了時間  $T'_i$  には次のイベント  $E_{i+1}$  のイベント開始時間  $T_{i+1}$  が付加される。但し、イベント開始時間  $T_i$  を算出する際は、イベントの前後関係を考慮せずに、各イベント  $E_i$  で独立したイベント開始時間  $T_i$  の算出をしたため、 $T_i \geq T_{i+1}$  となる場合がある。このような場合には、イベント  $E_i$  の終了時間  $T'_i$  には次のイベント  $E_{i+1}$  のイベント開始時間  $T_{i+1}$  ではなく、次節の推定方法 (i), (ii), (iii), (iv) においてそれぞれ、 $A$ ,  $B \times \beta_i$ ,  $B' \times \beta_i$ ,  $B_w \times W_1(E_i) \times W_r(E_i)$  をイベント開始時間  $T_i$  に加算した値をイベント終了時間  $T'_i$  とする。また、最後のイベント  $E_N$  には次のイベント  $E_{N+1}$  のイベント開始時間  $T_{N+1}$  が存在しないため前ケースと同様に、次節の推定方法 (i), (ii), (iii), (iv) においてそれぞれ、 $A$ ,  $B \times \beta_N$ ,  $B' \times \beta_N$ ,  $B_w \times W_1(E_N) \times W_r(E_N)$  をイベント開始時間  $T_N$  に加算した値をイベント  $E_N$  の終了時間  $T'_N$  とする。図 2.3 には推定方法 (iii) を用いた時のイベント時間の算出方法を示す。

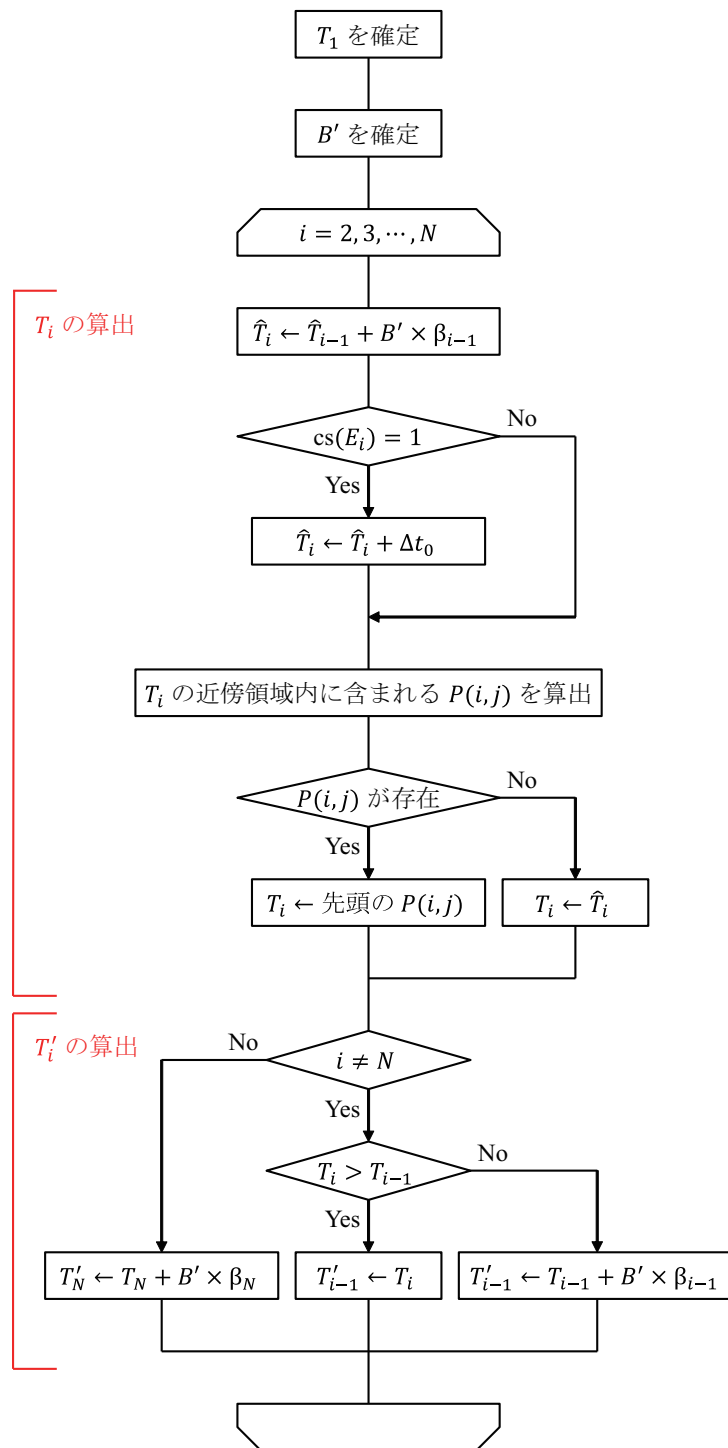


図 2.3 推定方法 (iii) に基づいたイベント  $E_i$  のイベント時間 ( $T_i$  から  $T'_i$  まで) の算出

## 2.5 野球試合モデルに依る推定イベント開始時間の算出

イベント  $E_i$  の推定イベント開始時間  $\hat{T}_i$  の算出する際、野球試合のモデル構造として、イベント  $E_i$  の前に「どのようなシーンがあったか」、「そのシーンにどのくらいの時間を要したか」という2つを考慮して式を導く。筆者の提案した4種類の野球試合モデルに依る推定イベント開始時間の算出方法を述べる。

### (i) 単位イベント当たりの平均時間を用いた推定

この推定方法ではまず、一球速報を用いて取得した試合時間  $T$  に基づき、単位イベント（打席シーン）当たりの平均時間  $A$  を次の式より算出する。

$$A = \frac{\text{(試合時間)}}{\text{(総イベント数)}} = \frac{T}{N} \quad (2.1)$$

次に、この単位イベント（打席シーン）当たりの平均時間  $A$  を用いて、イベント  $E_i$  の推定イベント開始時間  $\hat{T}_i$  を次の式より算出する。

$$\hat{T}_i = \hat{T}_{i-1} + A = T_1 + A \times (i - 1) \quad (2.2)$$

### (ii) 単位一球当たりの平均時間を用いた推定

この推定方法ではまず、一球速報を用いて取得した各イベント  $E_i$  に要した球数  $\beta_i$  ( $i = 1, 2, \dots, N$ ) に基づき、それぞれのイベントに要した球数を合計し、試合全体の総球数を求める。この総球数を用いて単位一球当たりの平均時間  $B$  を次の式より算出する。

$$B = \frac{\text{(試合時間)}}{\text{(総球数)}} = \frac{T}{\sum_{i=1}^N \beta_i} \quad (2.3)$$

次に、この単位一球当たりの平均時間  $B$  を用いて、球数が  $\beta_i$  であるイベント  $E_i$  の推定イベント開始時間  $\hat{T}_i$  を次の式より算出する。

$$\hat{T}_i = \hat{T}_{i-1} + B \times \beta_{i-1} = T_1 + B \times \sum_{k=1}^{i-1} \beta_k \quad (2.4)$$

### (iii) 推定方法 (ii) に攻守交替の時間を考慮した推定

推定方法 (ii) と同様のアプローチで推定イベント開始時間の算出をするが、攻守交替直後のイベントに対してのみ、攻守交替時間をパラメータ  $\Delta t_s$  (秒) として推定イベント開始時間に加算する。但し、イベント  $E_i$  が攻守交替直後のイベントか否かについても一球速報を用いて取得し、次の関数を用いて表す。

$$\text{cs}(E_i) = \begin{cases} 1 & (E_i \text{ が攻守交替直後のイベント}) \\ 0 & (\text{otherwise}) \end{cases} \quad (2.5)$$

この推定方法ではまず、式 (2.3) と式 (2.5) に基づき、攻守交替を考慮した場合の単位一球当たりの平均時間  $B'$  を次の式より算出する。

$$B' = \frac{T - \Delta t_s \times \sum_{i=1}^N \text{cs}(E_i)}{\sum_{i=1}^N \beta_i} \quad (2.6)$$

次に、この攻守交替を考慮した場合の単位一球当たりの平均時間  $B'$  を用いて、イベント  $E_i$  の推定イベント開始時間  $\hat{T}_i$  を次の式より算出する。

$$\hat{T}_i = \hat{T}_{i-1} + B' \times \beta_{i-1} + \Delta t_s \times \text{cs}(E_i) \quad (2.7)$$

(iv) 推定方法 (iii) に投手交代、単位一球当たりの投球時間への重みを考慮した推定

推定方法 (iii) の野球試合モデルをより精密化するために、推定方法 (iii) の野球試合モデルに、投手が交代するシーンに要する時間、及び、イベント  $E_i$  に依って単位一球当たりの投球時間に重みを考慮する以下の推定イベント開始時間  $\hat{T}_i$  の算出方法を提案する。この推定方法は、推定方法 (iii) と同様のアプローチで推定イベント開始時間の算出をするが、直前に投手交代があったイベントに対してのみ投手交代の時間をパラメータ  $\Delta t_p$  (秒) として推定イベント開始時間に加算する。さらに、イベント  $E_i$  が起こったのが試合の終盤であった場合、または、そのイベント  $E_i$  開始時、塁上にランナーが存在した場合、それぞれで投手が一球投球するのに要する時間に重みを付加する。但し、本研究では試合の終盤を7回以降（試合が延長戦に入った場合も同様）と定義する。以降この推定方法による算出手順を述べる。

この推定方法ではまず、イベント  $E_i$  の直前に投手交代が存在したか否かを一球速報を用いて取得し、次の関数を用いて表す。

$$\text{cp}(E_i) = \begin{cases} 1 & (E_i \text{の直前に投手交代}) \\ 0 & (\text{otherwise}) \end{cases} \quad (2.8)$$

次に、一球速報を用いてイベント  $E_i$  が起こるイニングを取得し、この際、投手が一球投球するのに要する時間に加える重みであるパラメータ  $w_l (\geq 1)$  に基づき、次の関数で表す。

$$W_l(E_i) = \begin{cases} w_l & (E_i \text{が7回以降のイベント}) \\ 1 & (\text{otherwise}) \end{cases} \quad (2.9)$$

さらに、一球速報を用いてイベント  $E_i$  開始時のランナーの有無を取得し、この際、投手が一球投球するのに要する時間に加える重みであるパラメータ  $w_r (\geq 1)$  に基づき、次の関数で表す。

$$W_r(E_i) = \begin{cases} w_r & (E_i \text{開始時にランナーが存在}) \\ 1 & (\text{otherwise}) \end{cases} \quad (2.10)$$

式 (2.5), 式 (2.8), 式 (2.9), 及び, 式 (2.10) を用いて, 単位一球当たりの加重平均時間  $B_w$  を次の式より算出する.

$$B_w = \frac{T - \Delta t_s \times \sum_{i=1}^N \text{cs}(E_i) - \Delta t_p \times \sum_{i=1}^N \text{cp}(E_i)}{\sum_{i=1}^N \beta_i \times W_l(E_i) \times W_r(E_i)} \quad (2.11)$$

このイベント  $E_i$  における単位一球当たりの加重平均時間  $B_w$  を用いて, イベント  $E_i$  の推定イベント開始時間  $\hat{T}_i$  を次の式より算出する.

$$\begin{aligned} \hat{T}_i &= \hat{T}_{i-1} + \Delta t_s \times \text{cs}(E_i) + \Delta t_p \times \text{cp}(E_i) \\ &\quad + B_w \times W_l(E_{i-1}) \times W_r(E_{i-1}) \times \beta_{i-1} \end{aligned} \quad (2.12)$$



## 第3章

# システム評価と今後の課題

本研究では、打者毎のイベントを取得した後、これらのイベントにイベント時間を付加するという野球映像のタギングアルゴリズムを提案した。イベントの取得に関しては全て一球速報の Web テキストに依存しているため、各イベントのタグ情報は全て正確であることを前提とし、イベントの取得に関する評価は行わない。従って本章ではイベントに付加されたイベント開始時間とイベント終了時間についてのみ評価を行う。

イベント時間の評価を行うためには、実際のイベントの開始時間と終了時間を定義する必要がある。実際のイベント開始時間は、各打者が打席に入った時に、その打者の名前やシーズン成績等がディスプレイに表示された瞬間と定義する。実際のイベント終了時間は基本的に、次のイベントの実際のイベント開始時間と定義するが、そのイベントが攻守交替する直前のイベントであった場合のみ、実際のイベント終了時間は、攻守交替時にイニング表がディスプレイに表示された瞬間と定義する。以上の実際のイベント開始時間と終了時間の定義に基づき、本タギングアルゴリズムのイベント時間の評価に次の指標を設けた。但し、以下の式の「正しく認識されたイベント時間」とは、あるイベント  $E_i$  の実際のイベント時間とシステムが認識したイベント時間の重なり合った部分の時間のことを言う。

$$(\text{再現率}) = \frac{(\text{正しく認識されたイベント時間の合計})}{(\text{実際のイベント時間の合計})}$$

$$(\text{適合率}) = \frac{(\text{正しく認識されたイベント時間の合計})}{(\text{システムのイベント時間の合計})}$$

評価実験には5試合の野球中継番組の録画映像を用いた。但し、5試合の内4試合が9回まで放送された中継番組で、残り1試合(Data 1)は7回まで放送された中継番組である。さらに5試合全てで映像内にCMが含まれる映像を用いた。また、タギング精度には再現率と適合率のF値、及び、実際のイベント(打席シーン)のイベント開始時間とシステムが認識したイベント開始時間との二乗誤差(秒)を用いて評価し、このF値をシステム全体のタギング精度とする。さらに、本研究で定義した $\Delta t_1$ (分)、 $\Delta t_2$ (分)、 $\Delta t_s$ (秒)、 $\Delta t_p$ (秒)、 $w_l$ 、及び、 $w_r$ はそれぞれ以下の範囲で実験を行う。

- $-15 \leq \Delta t_1 \leq 15$  (1分刻み)
- $\Delta t_1 \leq \Delta t_2 \leq 15$  (1分刻み)
- $0 \leq \Delta t_s \leq 300$  (30秒刻み)
- $0 \leq \Delta t_p \leq 300$  (30秒刻み)
- $1.0 \leq w_l \leq 1.5$  (0.1刻み)
- $1.0 \leq w_r \leq 1.5$  (0.1刻み)

### 3.1 野球試合モデルの精密化に関する評価と課題

本研究のタギングシステムではイベント  $E_i$  について、まず、推定イベント開始時間  $\hat{T}_i$  を算出する。次に、パラメータを用いた近傍領域で実況ポイントを探査し、最初に出現した実況ポイント  $P(i, j)$  をイベント開始時間  $T_i$  とした。この時、この実況ポイント  $P(i, j)$  はあくまでもイベント  $E_i$  の打者が打席に入ったことを示す「ここで〇〇です」のような打席実況ポイントの推測に過ぎないため、場合によっては、その実況ポイント  $P(i, j)$  がイベント  $E_i$  の打者の打席シーン以外のシーンで実況された実況ポイントである可能性も考えられる。探索範囲である近傍領域を大きく設定し過ぎると、探索のターゲットであった打席実況ポイントではなく、イベント  $E_i$  の打者の打席シーンとは異なるシーンで実況された実況ポイント  $P(i, j)$  を探知してしまう可能性が高まるため、近傍領域はできるだけ小さい方が好ましい。そのためには、推定イベント開始時間の算出の時点で正確に算出しておく必要がある。推定イベント開始時間が正確に算出されると、実況ポイントの誤探知の可能性を低くするだけでなく、もし探索範囲内に実況ポイントが存在せず、推定イベント開始時間をそのままイベント開始時間とする例外処理の場合でも正確なイベント開始時間を付加することができるためタギング精度の向上につながることを期待される。そこでまず、各推定方法で算出された推定イベント開始時間の正確性を検証し、本研究で提案した野球試合モデルの精密化の有用性に関する評価を行う。

表 3.1 には推定方法 (i) から (iv) のそれぞれで算出された推定イベント開始時間をそのままイベント開始時間にする、つまり、実況音声認識を用いず、イベント開始時間の推定のみに基づいたタギングの精度を示す。推定方法 (i) から (iv) になるに連れてタギング精度が向上している、つまり、提案した推定イベント開始時間の算出がより正確であったことが確認できた。しかし、実際のイベント（打席シーン）のイベント開始時間とシステムが認識したイベント開始時間との二乗誤差（秒）を見ると、推定方法 (i) に基づいたタギングの二乗誤差がその他の推定方法に基づいたタギングの二乗誤差と比べて、誤差が大きいことが目立ち、その他 3 つの誤差にあまり大きな変化は見られなかった。このことから、モデル構造の精密化が未だ不十分であったのではないかという見方もでき、更に精密化できる可能性も確認できた。

表 3.1 推定イベント開始時間のみに基づくタギング精度

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	—	—	—	—	—	—	0.162	0.131	0.145	306.043
(ii)	—	—	—	—	—	—	0.311	0.253	0.279	218.216
(iii)	—	—	90	—	—	—	<b>0.343</b>	<b>0.279</b>	<b>0.308</b>	<b>214.696</b>
(iv)	—	—	90	0	1.0	1.0	<b>0.343</b>	<b>0.279</b>	<b>0.308</b>	<b>214.696</b>

## 3.2 各推定方法に依るタギング精度に関する評価と課題

前節で示した推定イベント開始時間の正確性の評価を踏まえ、次に、野球試合モデルの精密化に依って推定イベント開始時間が正確に算出されたことがタギングシステム全体にどのように影響したかを評価する。表 3.2 から表 3.6 には本研究で用いた 5 試合毎の各推定方法に依るタギング精度を示す。表 3.2 から表 3.6 を見ると、推定方法 (i) から (iv) にかけて、つまり、野球試合モデルをより精密化するに連れてタギング精度が向上する傾向があることが分かる。また、実際のイベント（打席シーン）のイベント開始時間とシステムが認識したイベント開始時間との二乗誤差も推定方法 (i) から (iv) にかけて小さくなっている。この結果を野球映像のモデリングの観点から考察する。推定方法 (i) と推定方法 (ii) では、それぞれ単位イベント当たりの平均時間、単位一球当たりの平均時間を用いて推定イベント開始時間を算出しているため、この算出方法では、投手が投球していないシーン、例えば、選手の交代シーンや攻守交替のシーンも投球しているのと同じ扱いとしての算出式になる。一方、推定方法 (iii) や (iv) では、野球映像を「打者が打席に入り、投手が投球するシーン」と「選手の交代、及び、攻守交替のシーン」の二つのモデルに構造化したため、より正確に推定イベント開始時間の算出が行われたことが、推定方法 (i) から (iv) にかけて、つまり、野球試合モデルをより精密化するに連れてタギング精度が向上した要因と考察する。

表 3.2 各推定方法に依るタギング精度 (データ 1)

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	-8	3	-	-	-	-	0.680	0.442	0.536	147.840
(ii)	-3	6	-	-	-	-	0.735	0.554	0.632	101.313
(iii)	-3	6	0	-	-	-	0.735	0.554	0.632	101.313
(iv)	-1	4	90	90	1.0	1.5	<b>0.774</b>	<b>0.614</b>	<b>0.685</b>	<b>51.246</b>

表 3.3 各推定方法に依るタギング精度 (データ 2)

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	-9	4	-	-	-	-	0.646	0.343	0.448	226.795
(ii)	-8	6	-	-	-	-	0.589	0.311	0.407	250.039
(iii)	-8	5	270	-	-	-	0.587	0.351	0.439	237.768
(iv)	-3	6	150	270	1.0	1.4	<b>0.712</b>	<b>0.501</b>	<b>0.588</b>	<b>129.531</b>

表 3.4 各推定方法に依るタギング精度 (データ 3)

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	-6	12	-	-	-	-	0.656	0.363	0.467	273.179
(ii)	0	6	-	-	-	-	0.691	0.538	0.605	94.281
(iii)	0	6	0	-	-	-	0.691	0.538	0.605	94.281
(iv)	0	8	60	30	1.0	1.0	<b>0.742</b>	<b>0.567</b>	<b>0.643</b>	<b>93.313</b>

表 3.5 各推定方法に依るタギング精度 (データ 4)

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	-11	4	-	-	-	-	0.470	0.230	0.308	380.260
(ii)	-6	-1	-	-	-	-	0.510	0.331	0.401	204.679
(iii)	-7	-2	180	-	-	-	0.579	0.363	0.446	164.528
(iv)	0	1	30	150	1.0	1.1	<b>0.703</b>	<b>0.578</b>	<b>0.635</b>	<b>79.201</b>

表 3.6 各推定方法に依るタギング精度 (データ 5)

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	-6	7	-	-	-	-	0.477	0.281	0.354	277.716
(ii)	3	5	-	-	-	-	0.489	0.337	0.399	225.825
(iii)	3	5	30	-	-	-	0.504	0.347	0.411	234.103
(iv)	-3	8	30	30	1.0	1.5	<b>0.564</b>	<b>0.347</b>	<b>0.430</b>	<b>201.514</b>

しかし、5試合平均の観点から考察すると、見解の相違が生じる。表 3.7 には推定方法 (i) から (iv) に基づいたタギングに関しての、5試合平均の精度を示す。表 3.7 からは、推定方法が (i) から (iv) になるに連れてタギング精度の向上は見られるものの、その上昇率が低いことや実際のイベント（打席シーン）のイベント開始時間とシステムが認識したイベント開始時間との二乗誤差が推定方法 (ii) から (iv) にかけて大きくなっていることより、精密化に依るタギング精度への影響が期待よりは低いとも考えられる。

この結果について、筆者は、平均精度はパラメータの最適化次第で大きく向上すると考えている。本研究の評価実験では各パラメータに関して最適化への考慮が無かったため、平均精度で期待通りの結果が得られなかった。特に、表 3.7 の 5 試合平均における推定方法 (iv) に基づいたタギング精度が、表 3.2 から表 3.5 における推定方法 (iv) に基づいたタギング精度よりも大きく下回っていることが分かる。表 3.2 における推定方法 (iv) に基づいたタギング精度と表 3.7 の 5 試合平均における推定方法 (iv) に基づいたタギング精度との差が最も大きく、その差は 0.228 であった。この結果より、本来ならば高い精度が得られる手法であるのにも関わらず各パラメータの最適化が不十分であった場合、その手法で得ることができるはずの精度に達しないことも考えられることから、各パラメータの最適化は今後の課題としていく。

表 3.7 各推定方法に依るタギング精度（5 試合平均）

推定方法	$\Delta t_1$	$\Delta t_2$	$\Delta t_s$	$\Delta t_p$	$w_l$	$w_r$	再現率	適合率	F 値	二乗誤差
(i)	-9	6	-	-	-	-	0.559	0.304	0.394	289.940
(ii)	-3	6	-	-	-	-	0.502	0.363	0.421	<b>198.485</b>
(iii)	-4	7	60	-	-	-	0.524	0.360	0.427	199.403
(iv)	-3	10	90	30	1.1	1.2	<b>0.588</b>	<b>0.373</b>	<b>0.457</b>	202.628

表 3.2 から表 3.7 より、いずれの推定方法についても、再現率が適合率を上回っていることが分かる。再現率は実際のイベント（打席シーン）時間の内、正しく認識された時間の割合である。再現率が高い傾向があるということは、システムが実際の打席シーンよりも前の時間をイベント開始時間と認識したイベントが比較的多くあったのではないかと考察ができる。この要因としては2つ挙げられる。1つ目の要因は、本タギングシステムの実況ポイントの探索方法にある。本タギングアルゴリズムでは推定イベント開始時間を算出し、その近傍領域内で「最初に」出現した実況ポイントをイベント開始時間とする処理を行う。従って、この場合、正確に推定イベント開始時間が算出され、その直前に正しい開始時間となり得る打席に入ったことを示す「ここで〇〇です」のような打席実況ポイントがあったにも関わらず、パラメータに依って、打席実況ポイントよりも前の実況ポイントをイベント開始時間としてしまう可能性が出て来る。従って、この実況ポイントの探索方法では再現率が高く、またそれに伴い、適合率が低くなる傾向になり易いと考えられる。この傾向を検証するために今後、近傍領域で最初に出現した実況ポイントをイベント開始時間とするのではなく、近傍領域内で複数の実況ポイントを探し、その中で更に、打席実況ポイントとなる実況ポイントを推定するアルゴリズムを検討し、本アルゴリズムとの比較実験を行っていく。

再現率が適合率を上回っている要因の2つ目として、実況者によって実際のイベント（打席シーン）の開始時間よりも前に打者名が実況されたイベントが多かったからと考えられる。1つ目の要因については、実際のイベント開始時間よりも前の実況ポイントをイベント開始時間とされたイベントが多く存在したのではないかと考察であったため、理論上は再現率が高くなるが、その反面極端に適合率が低くなってしまいうことも考えられる。本評価実験からは、適合率が極端に低いことは確認できなかったため、おそらく、この2つ目の要因が主に影響していることが考えられる。本評価実験に用いた5試合は全て異なった実況者による中継番組であったため、それぞれの実況者は打席シーンの直前にある打者が打席に入ったことを示す実況をする傾向があるのではないかと推測できる。このことに関しては、今後実際に評価実験に用いた試合を調査することで確認を図る。さらに再現率が適合率を上回っていることについて、この2つ目の要因の影響が大きいとすると、近傍領域を定める際の下限と上限を決める  $\Delta t_1$  と  $\Delta t_2$  の最適化に有用なデータになり得ることが考えられる。実況者毎の打席実況ポイントの実況傾向を実況者データベースに登録し、システムに学習させることで、実況者に依って実況ポイントを探索する範囲を任意に定められる可能性があるだろう。

### 3.3 各パラメータに関する評価と課題

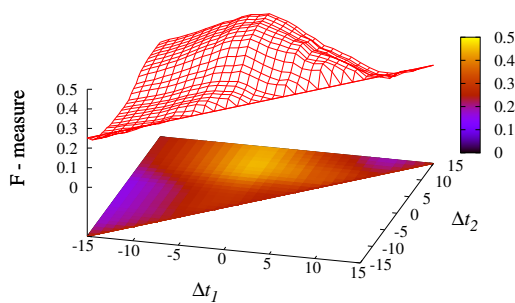
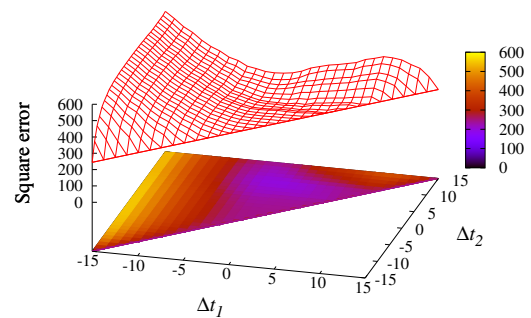
図 3.1 には、表 3.7 の野球映像 5 試合の平均精度に基づき、推定方法 (iv) における最高精度であった時の、 $\Delta t_s = 90$ ,  $\Delta t_p = 30$ ,  $w_l = 1.1$ ,  $w_r = 1.2$  を固定し、 $\Delta t_1$  と  $\Delta t_2$  の変化に伴う F 値を、図 3.2 には、実際のイベント（打席シーン）のイベント開始時間とシステムが認識したイベント開始時間との二乗誤差（秒）を示す。また、図 3.3 には表 3.7 に基づき、推定方法 (iv) における最高精度であった時の、 $\Delta t_2 = 10$ ,  $\Delta t_s = 90$ ,  $\Delta t_p = 30$ ,  $w_l = 1.1$ ,  $w_r = 1.2$  を固定し、 $\Delta t_1$  ( $\leq \Delta t_2 = 10$ ) の変化に伴う F 値、及び、二乗誤差を示す。また、図 3.4 には表 3.7 に基づき、最大精度であった時の  $\Delta t_1 = -3$ ,  $\Delta t_s = 90$ ,  $\Delta t_p = 30$ ,  $w_l = 1.1$ ,  $w_r = 1.2$  を固定し、 $\Delta t_2$  ( $\geq \Delta t_1 = -3$ ) の変化に伴う F 値、及び、二乗誤差を示す。

図 3.1 と図 3.2 より F 値が高くなるに連れて、二乗誤差が小さくなっていることが確認できた。 $\Delta t_1$  と  $\Delta t_2$  は、「ここで〇〇です」のような打者が打席に入っていることを示す打席実況ポイントを探査する範囲を定めるためのパラメータであるため、各試合、特に各実況者による依存性が非常に高いと考えられる。また、図 3.4 からは  $\Delta t_2$  の値が一定値を超えたところで F 値に大きな変化が見られなくなったことが分かる。一方で図 3.3 からは  $\Delta t_1$  の値は F 値に大きく影響していることが分かる。本研究のタギングアルゴリズムでは、イベント開始時間を算出するために推定イベント開始時間を求め、その近傍領域内に存在する最初の実況ポイントをイベント開始時間とする手法を提案した。この近傍領域の下限を定めるパラメータが  $\Delta t_1$  であり、上限を定めるパラメータが  $\Delta t_2$  である。このことから、領域の上限を定めるパラメータ  $\Delta t_1$  に関しては、値をある程度高く設定することで実況ポイントが出現した場合その値より高い値に設定してもタギング精度に影響しないことが分かる。一方、領域の下限を定めるパラメータ  $\Delta t_2$  に関しては、下限を高く設定しすぎると、探索のターゲットとしていた「ここで〇〇です」のような打席実況ポイントが仮にその領域よりも前に存在していた場合見逃してしまうことも考えられ、逆に低く設定しすぎると、探索のターゲットとしていた打席実況ポイントとは異なる、例えば、前の打者の打席シーンで実況された「次は〇〇です」のような実況ポイントを探査してしまうことも考えられる。

また、図 3.1 からも見られるように、 $\Delta t_1 = \Delta t_2$  ( $-15 \sim 15$ ) の時の F 値が全て同じ値 (0.251) になったことについて考察する。この場合、あるイベント  $E_i$  の推定イベント開始時間  $\hat{T}_i$  から実況ポイント  $P(i, j)$  を探索する近傍領域が範囲ではなく一点である状況のことを言う。つまり、その一点のみを探索し  $E_i$  の打席実況ポイント  $P(i, j)$  が存在しない場合、 $E_i$  のイベント開始時間  $T_i$  には推定イベント開始時間  $\hat{T}_i$  がそのまま付加されることになる。 $\Delta t_1 = \Delta t_2$  の場合は、推定イベント開始時間  $\hat{T}_i$  の付近に定めたある一点のみで実況ポイント  $P(i, j)$  を探索するため、実況ポイント  $P(i, j)$  が存在する可能性は非常に低く、全てのイベント  $E_i$  で開始時間  $T_i$  に、推定イベント開始時間  $\hat{T}_i$  が付加されたことが考えられる。



次に、図 3.1 からは  $\Delta t_1$  は 0 に近づくと F 値が高くなる傾向が見られるが、その傾向には反し、 $\Delta t_1 = 15$ ,  $\Delta t_2 = 15$  の時の F 値が  $\Delta t_1 = 14$ ,  $\Delta t_2 = 15$  の時の F 値よりも高くなっていることが分かる。前者では前述したように、近傍領域が領域ではなく一点に定めるため、その一点に実況ポイントが存在する可能性は低い。後者では、近傍領域が範囲として定められているため、その範囲で探索して実況ポイントを見つけられたイベントがいくつかあることが想定される。しかし、F 値が最も高かった時のパラメータが  $\Delta t_1 = -3$ ,  $\Delta t_2 = 10$  であったことを参考にすると、 $\Delta t_1 = 14$ ,  $\Delta t_2 = 15$  における近傍領域で存在した実況ポイントは「ここで〇〇です」のような打席実況ポイントではない可能性が高く、これが  $\Delta t_1 = 14$ ,  $\Delta t_2 = 15$  の時の F 値の減少に影響し、 $\Delta t_1 = 15$ ,  $\Delta t_2 = 15$  の時の F 値が  $\Delta t_1 = 14$ ,  $\Delta t_2 = 15$  の時の F 値よりも高くなったと考えられる。

図 3.1  $\Delta t_1$  と  $\Delta t_2$  の変化に依る F 値図 3.2  $\Delta t_1$  と  $\Delta t_2$  の変化に依る二乗誤差

次に、攻守交替に要する時間として定義したパラメータ  $\Delta t_s$  に関する考察を行う。図 3.5 には表 3.7 に基づき、最大精度であった時の  $\Delta t_1 = -3$ ,  $\Delta t_2 = 10$ ,  $\Delta t_p = 30$ ,  $w_l = 1.1$ ,  $w_r = 1.2$  を固定し、 $\Delta t_s$  の変化に伴う F 値、及び、二乗誤差を示す。 $\Delta t_s$  に関しては、表 3.2 から表 3.6 を見ると、各試合で、不規則な値に最適化されていることが分かるが、図 3.5 から  $\Delta t_s$  の値を変化させても F 値の変化が少ないことから、野球試合モデルの精密化があまり影響していないのではないかという考察ができる。日本プロ野球においては、2 分 15 秒を目安に攻守交替を行うように推奨されているのに対し、図 3.5 から読み取れる 5 試合平均の攻守交替の最適化された平均時間は 90 (秒) と誤差が生じた。これに関しては、今後更に評価試合数を増やし、実験を行うことで検証していく。図 3.6 には表 3.7 に基づき、最大精度であった時の  $\Delta t_1 = -3$ ,  $\Delta t_2 = 10$ ,  $\Delta t_s = 90$ ,  $w_l = 1.1$ ,  $w_r = 1.2$  を固定し、 $\Delta t_p$  の変化に伴う F 値、及び、二乗誤差を示す。表 3.2 から表 3.6 を見ると、 $\Delta t_p$  もそれぞれ不規則に最適化されているような傾向が見られるが、図 3.6 からは、最高精度を得た  $\Delta t_p = 30$  をピークに  $\Delta t_p$  の値が大きくなる連れて、F 値が減少していることが分かる。このことから、今回の評価実験では各試合毎で  $\Delta t_p$  の値に大きな誤差が生じたものの、今後は最適化を図る価値のあるパラメータであるのではないかという考察ができる。

図 3.7 には表 3.7 に基づき、最大精度であった時の  $\Delta t_1 = -3$ ,  $\Delta t_2 = 10$ ,  $\Delta t_s = 90$ ,  $\Delta t_p = 30$ ,  $w_r = 1.2$  を固定し、 $w_l$  の変化に伴う F 値、及び、二乗誤差を示す。表 3.2 から表 3.6 を見ると、各試合で最適化された  $w_l$  の値が全て 1.0 であることや、図 3.7 から  $w_l$  の重みを増加させた時タギング精度が減少していくことから、イベント  $E_i$  が終盤 (7 回以降) のイベントであった場合に単位一球当たりの投球時間に付加する重み  $w_l$  は野球試合の精密化には大きく影響しないことを示した。図 3.8 には表 3.7 に基づき、最大精度であった時の  $\Delta t_1 = -3$ ,  $\Delta t_2 = 10$ ,  $\Delta t_s = 90$ ,  $\Delta t_p = 30$ ,  $w_l = 1.1$  を固定し、 $w_r$  の変化に伴う F 値、及び、二乗誤差を示す。このイベント  $E_i$  開始時に塁上にランナーが存在した場合に単位一球当たりの投球時間に付加する重み  $w_r$  についても、図 3.8 から  $w_r$  を変化させても F 値に大きな影響を与えていないことが分かるが、本研究で提案した  $w_r$  は一試合を通して定数化し、また、イベントにおけるランナーの有無の判断だけで重み付けを行った。今後は、「ノーアウト満塁」または「ツーアウト 1 塁」、「投手が打席に入った」または「4 番が打席に入った」等といったイベントの状況を更に細かく分析し、単位一球当たりの投球時間に重み付けをしていくことで  $w_r$  の有用性を検証していく。

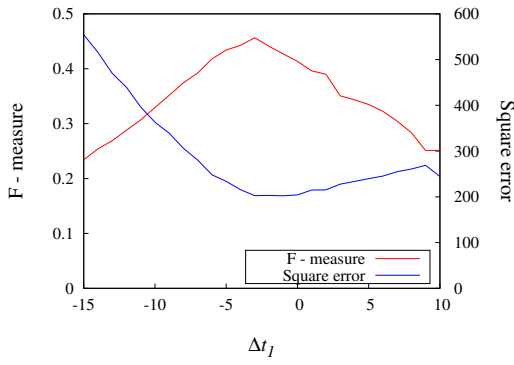


図 3.3  $\Delta t_1$  の変化に依る二乗誤差

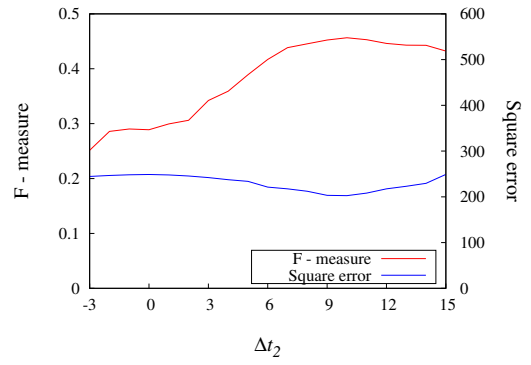


図 3.4  $\Delta t_2$  の変化に依る F 値と二乗誤差

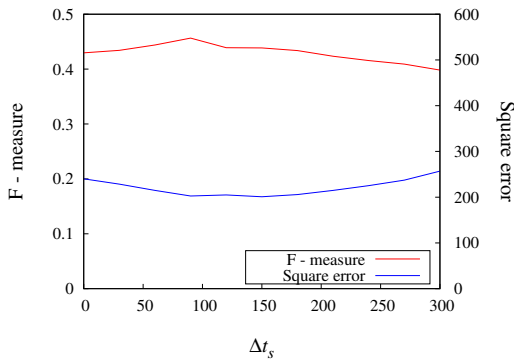


図 3.5  $\Delta t_s$  に依る F 値と二乗誤差の変化

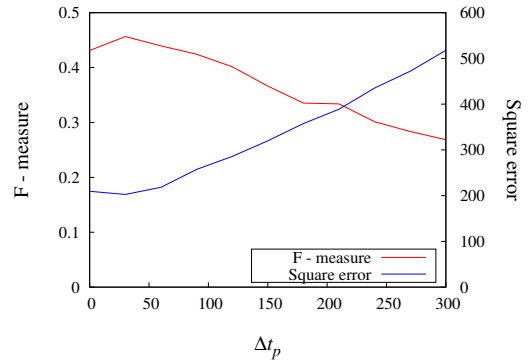


図 3.6  $\Delta t_p$  に依る F 値と二乗誤差の変化

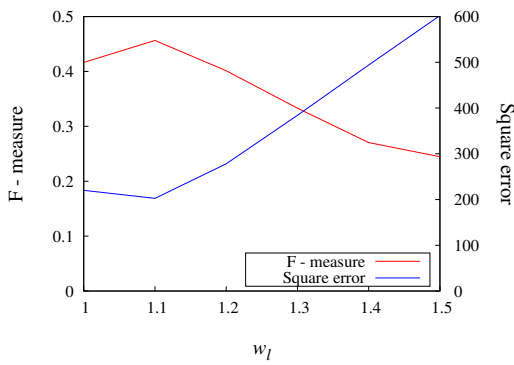


図 3.7  $w_l$  に依る F 値と二乗誤差の変化

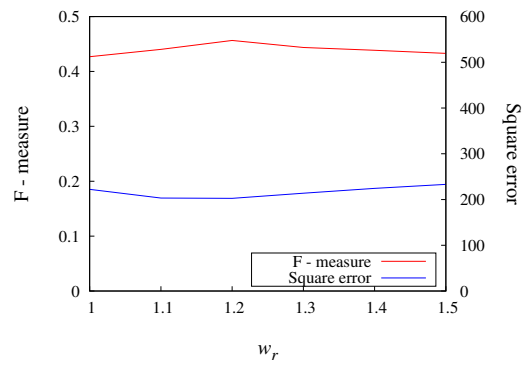


図 3.8  $w_r$  に依る F 値と二乗誤差の変化

## 第4章

# むすび

本研究では，野球映像を自動タギングするために，一球速報の Web テキストを用いてその試合で起こった打席毎のイベントを取得し，それらのイベントに基づき推定したイベント開始時間とその近傍領域，及び，実況音声認識を用いてイベント開始時間となる打席実況ポイントを探査するタギングアルゴリズムを提案した．また，推定イベント開始時間を正確に算出するために，投手交代に要する時間，ランナーの有無等のイベントに依る単位一球当たりの投球時間の重みを取り入れた野球試合モデルを導入し，タギング精度の向上と，実況ポイントを探査する際の推定イベント開始時間の近傍領域を定めるパラメータの最適化に有用なデータを取得できる可能性が確認できた．そして今後は，その他の各パラメータの最適化を検証するためにより多くの試合数で評価実験を行ったり，更に正確な推定イベント開始時間を算出できるように野球試合のモデル構造を検討したりしていく．また，実況音声認識以外の技術を用いたタギング技術の検討も視野に入れている．例えば，映像解析によるタギング [4, 5, 6] であったり，一球速報の速報という特徴を活かしたタギングであったり等といった，複数の技術を組み合わせたタギング手法を提案していくことで，更なる精度の向上を目指す．

# 謝辞

本研究に際して、様々なご指導を頂きました服部峻助教に厚く御礼申し上げます。また、日常の議論を通じて多くの知識や示唆を頂いた服部研究室の皆様にも深く感謝の意を表します。

## 参考文献

- [1] 宮原 正典, 青木 政樹, 滝口 哲也, 有木 康雄, “顔表情からの関心度推定に基づく映像コンテンツへのタギング,” 情報処理学会論文誌, Vol.49, No.10, pp.3694–3702 (2008).
- [2] Yahoo! JAPAN スポーツナビ – 一球速報 –, <http://live.baseball.yahoo.co.jp/npb/game/2015102401/score>.
- [3] アドバンスド・メディア, 音声認識ソフト AmiVoice SP2, <http://sp.advanced-media.co.jp/>.
- [4] 椋木 雅之, 寺尾 元宏, 池田 克夫, “カット構成の規則性を利用したスポーツ映像のプレイ単位への分割,” 電子情報通信学会論文誌 D, Vol.J85-D2, No.6, pp.1016–1024 (2002).
- [5] 阿部 孝司, 年代 光宏, 林 貴宏, 木村 春彦, “放送型野球中継映像における打席シーンの分割に必要なスコアテロップ内の情報領域の抽出と意味推定,” 画像電子学会誌, Vol.36, No.5, pp.701–709 (2007).
- [6] 望月 貴裕, 藤井 真人, 八木 伸行, 篠田 浩一, “投球の次ショットに重きを置いたシーンのパターン化と離散隠れマルコフモデルを用いた野球放送映像の自動イベント分類,” 映像情報メディア学会誌: 映像情報メディア, Vol.61, No.8, pp.1139–1149 (2007).