

アノテーション付きサッカー軌跡データからの チーム戦術パターン抽出

Extraction of Frequent Patterns in Trajectories of Football Players

鈴木 湧人¹ 尾崎 知伸^{2*}
Yuto Suzuki¹ Tomonobu Ozaki²

¹ 日本大学大学院 総合基礎科学研究科

¹ Graduate School of Integrated Basic Sciences, Nihon University

² 日本大学 文理学部

² College of Humanities and Sciences, Nihon University

Abstract: In this paper, we propose a framework to extract frequent moving patterns from trajectory databases on multiple football players. At the first stage in the framework, clustering techniques are applied to extract meaningful partial trajectories as motifs and convert the trajectory databases into a set of motifs associated with the original partial trajectories. Then, in the second stage, we apply a sequential pattern miner developed for treating motifs with partial trajectories and enumerate all the frequent patterns. Experimental results using the trajectory databases on nine matches in J. League (Japanese professional football league) show the effectiveness of the proposed framework from the aspects of the computation time. We also succeeded in extracting meaningful patterns representing some offensive and defensive strategies formed by multiple football players.

1 はじめに

近年、バレーボールやバスケットボール、サッカーに代表される様々なスポーツ分野において、経験や直観ではなく、実際に記録されたプレーデータに基づき、選手の能力やチームの戦術を分析することが重要視され、様々な研究が行われている。例えば、サッカー分野では、選手のパフォーマンスを定量的に評価する研究 [1] やホームとアウェイでの戦術の違いに関する研究 [2]、プレーの可視化に関する研究 [3] など、主に個人的なプレーと戦術の可視化に関する研究が多数報告されている。

本研究では、サッカーを対象とし、複数人が運動する戦術的な動きを表すパターンを抽出するための枠組みを提案する。具体的には、選手とボールの軌跡データを対象に、

1. クラスタリング技術を用い、意味のある部分軌跡をモチーフとして抽出する
2. 時間と位置情報を考慮しながらモチーフの系列をパターンとして抽出する

という、2段階からなる枠組みを考える。また特に、後段における系列パターンの抽出を実現するため、既存の単一区間イベント系列を対象としたパターン発見技術 [4] を拡張し、位置情報を考慮したパターンを列挙するアルゴリズムを開発する。

本論文の構成は、以下の通りである。2章では、軌跡データやそこから得られる区間イベント、区間イベント系列など、入力に関する種々の定義を導入する。3章では、今回の抽出対象であるパターンに関する定義を行い、実際の列挙アルゴリズムを示す。4章で、Jリーグ9試合のデータを用いて行った評価実験の結果を示し、考察を行う。最後に5章でまとめを行い、今後の課題を述べる。

2 準備

本章では、準備として、分析の対象である軌跡データや区間イベント、区間イベント系列に関して形式的な定義を与える。その後、軌跡データ集合を区間イベント系列へと変換する手法について概説する。

*連絡先：日本大学文理学部情報科学科
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: tozaki@chs.nihon-u.ac.jp

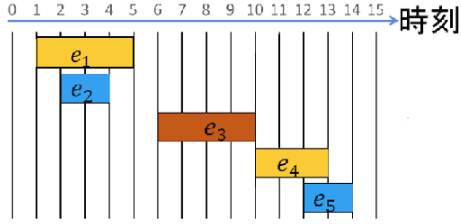


図 1: 区間イベント系列の例

2.1 区間イベントと区間イベント系列

区間イベントとは，レンタルビデオの貸し出しや病院への入退院，毎日の睡眠など，開始時刻と終了時刻を持つ事象を指す．本論文では，区間イベント $e = \langle l_e, e^+, e^-, a_e \rangle$ を，イベント内容を表すラベル l_e とイベントの開始・終了時刻 $e^+, e^- \in \mathbb{N}^+$ ，イベントに付随した属性 a_e の 4 項組で表現する．また， k 個の区間イベントの集合（系列）

$$E = \left\{ \begin{array}{l} e_1 = \langle l_{e_1}, e_1^+, e_1^-, a_{e_1} \rangle, \\ \vdots \\ e_k = \langle l_{e_k}, e_k^+, e_k^-, a_{e_k} \rangle \end{array} \right\}$$

を，区間イベント系列と呼ぶ．図 1 に， $e_1 \sim e_5$ の 5 つの区間イベントで構成される区間イベント系列の例を示す．図からわかる通り，各区間イベントには，時間的な前後関係や重なりが存在することとなる．

2.2 軌跡データと部分軌跡

オブジェクト o に関する軌跡データ tr とは，時刻 $t \in \mathbb{N}^+$ における 2 次元平面上の o の位置座標 (x_t, y_t) とそれに対する属性 a_t の 3 項組 (x_t, y_t, a_t) を要素とし，それらを時刻に沿って系列化したものであり，

$$tr = \{(x_1, y_1, a_1), \dots, (x_t, y_t, a_t), \dots, (x_T, y_T, a_T)\}$$

$(1 \leq t \leq T)$ と表現される．なお， T は tr の要素数，すなわち長さを表す．また本論文では，各要素の時刻は 1 から始まり T まで連続していることを仮定する．

軌跡データ tr における時刻 i から j ($1 \leq i \leq j \leq T$) の（連続する）部分軌跡を

$$tr[i, j] = \{(x_i, y_i, a_i), \dots, (x_j, y_j, a_j)\}$$

と表記する．部分軌跡 $p = tr[i, j]$ の重心を

$$(\bar{x}^p, \bar{y}^p) = \left(\frac{1}{j-i+1} \sum_{k=i}^j x_k, \frac{1}{j-i+1} \sum_{k=i}^j y_k \right)$$

と表記する．これを拡張し， n 個の部分軌跡の集合 $P = \{p_1, \dots, p_n\}$ の重心を，各部分軌跡 p_i の重心の重心

$$(\overline{X^P}, \overline{Y^P}) = \left(\frac{1}{n} \sum_{k=1}^n \bar{x}^{p_k}, \frac{1}{n} \sum_{k=1}^n \bar{y}^{p_k} \right)$$

と定義する．以上を用い，部分軌跡の集合 P の広がりを

$$d_1(P) = \max_{p \in P} \sqrt{(\bar{x}^p - \overline{X^P})^2 + (\bar{y}^p - \overline{Y^P})^2}$$

と定義する．同様に，2 つの部分軌跡 $p_1 = tr[i_1, j_1]$ と $p_2 = tr[i_2, j_2]$ の最短距離を

$$md(p_1, p_2) = \min_{i_1 \leq k_1 \leq j_1, i_2 \leq k_2 \leq j_2} \sqrt{(x_{k_1} - x_{k_2})^2 + (y_{k_1} - y_{k_2})^2}$$

と定義し，部分軌跡の集合 P の広がりを

$$d_2(P) = \max_{p_i, p_j \in P, i \neq j} md(p_i, p_j)$$

と定義する．式が表すように， d_1 は，全体の重心に対して各部分軌跡が最大でどの程度離れているのかを基準としている．また d_2 では，それぞれの軌跡が最大でどの程度離れているかを基準としている．

2.3 軌跡データの区間イベント系列への変換

N 個のオブジェクトに関する長さ T の軌跡データの集合

$$TR = \left\{ \begin{array}{l} tr^1 = \{(x_1^1, y_1^1, a_1^1), \dots, (x_T^1, y_T^1, a_T^1)\}, \\ \vdots \\ tr^N = \{(x_1^N, y_1^N, a_1^N), \dots, (x_T^N, y_T^N, a_T^N)\} \end{array} \right\}$$

とする．ここで，部分軌跡 $tr^k[i, j]$, ($1 \leq k \leq N, 1 \leq i \leq j \leq T$) が形状の面などで強い意味性 \mathcal{M} を持つとき，それをモチーフ [10] と呼び $\mathcal{M}(m, tr^k[i, j])$ と表記する．なお， m はモチーフの種類を区別するラベルとする．

TR に含まれる各モチーフを区間イベントとみなすことで得られる区間イベント系列を

$$M(TR) = \left\{ \langle m, i, j, a \rangle \mid \begin{array}{l} 1 \leq i \leq j \leq T, \\ tr \in TR, \mathcal{M}(m, tr[i, j]) \end{array} \right\}$$

と表記する．また，各区間イベント $\langle m, i, j, a \rangle \in M(TR)$ から，抽出元となった部分軌跡を得る操作を

$$Em(\langle m, i, j, a \rangle) = tr[i, j] \text{ と表記する．}$$

本論文では，部分軌跡の形状に着目してモチーフ \mathcal{M} の発見を行う．具体的には，各軌跡データを固定長の部分軌跡へと分割した上で，クラスタリング技術 BIRCH[5] を適用する．その後，各クラスタに対してラベル m を付与することで，モチーフを区別する．また，座標から得られる情報（相対的な位置関係や進行方向など）に従い，属性 a を決定する．

3 区間イベント系列パターン

本章では、軌跡データの集合 TR から得られる区間イベント系列 $M(TR)$ におけるパターンを定義する。また、時間、空間を考慮した出現を導入するとともに、パターンの支持度（頻度）を定義する。

3.1 端点系列とパターン

文献 [6, 4] に従い、区間イベントの端点に関する順序関係を導入する。またこの順序関係を用い、区間イベント系列を、具体的な時刻を抽象化し順序だけを考慮した端点の系列へと変換する。

区間イベント $e = \langle l_e, e^+, e^-, a_e \rangle$ に対し、イベントの端点、すなわち e の開始または終了を記号 $ep \in \{ep^+, ep^-\}$ で表す。また ep に対し、それが開始・終了のどちらかであるか返す関数を

$$m(ep) = \begin{cases} + & (ep = e^+) \\ - & (ep = e^-) \end{cases}$$

とし、各端点の時刻を

$$t(ep) = \begin{cases} e^+ & m(ep) = + \\ e^- & m(ep) = - \end{cases}$$

と表す。記号 $l(ep) = l_e$ と $a(ep) = a_e$ を、それぞれ端点が属する区間イベントのラベル l_e と属性 a_e とする。加えて、ラベルや属性間の順序関係を \prec_L と表記する。以上の準備の下、2 端点 x, y 間の全順序関係 \prec を以下のように定義する。すなわち x と y に対して、時刻、開始・終了、イベントラベル、属性の順に比較を行い、その順序関係を定める。

$$x \prec y \text{ if } \begin{cases} t(x) < t(y) \vee \\ t(x) = t(y) \wedge m(x) = + \wedge m(y) = - \vee \\ t(x) = t(y) \wedge m(x) = m(y) \wedge l(x) \prec_L l(y) \vee \\ t(x) = t(y) \wedge m(x) = m(y) \wedge l(x) = l(y) \\ \wedge a(x) \prec_L a(y) \end{cases}$$

一方、2 つの端点の時刻に関する 2 項関係を、

$$x \oplus y = \begin{cases} < & (t(x) < t(y)) \\ = & (\text{otherwise}) \end{cases}$$

と表記する。

上記の順序関係と記法を利用することで、具体的な時間を排除し、区間イベント系列を各区間イベントの開始・終了の順序関係に着目した系列に変換することが可能である。すなわち、区間イベント系列 $E = \{e_1, \dots, e_{|E|}\}$

を $2|E|$ 個の端点 $ep \in \bigcup_{1 \leq i \leq |E|} \{ep_i^+, ep_i^-\}$ からなる端点系列

$$\langle ep^1 \oplus_1 ep^2 \oplus_2 \dots \oplus_{(2|E|-1)} ep^{(2|E|)} \rangle \\ (\forall i, j : 1 \leq i < j \leq 2|E| \rightarrow ep^i \prec ep^j)$$

へと変換する。なお（具体的な時刻情報を除いて）区間イベント系列と端点系列は一対一に対応し、端点系列から各区間イベントの前後関係や重なりが復元できることから、今後両者を同一視する。また、データセット中に現れる端点の系列を、区間イベント系列パターンと呼び、本論文での抽出対象とする。

例として、図 1 に示す 5 つの区間イベントからなる区間イベント系列を考える。この区間イベント系列は、10 個の端点を用いて

$$\langle ep_1^+ < ep_2^+ < ep_2^- < ep_1^- < \\ ep_3^+ < ep_4^+ = ep_3^- < ep_5^+ < ep_4^- < ep_5^- \rangle$$

と表現される。

3.2 パターンの一般特殊関係と出現

それぞれ I 個と J 個の区間イベントから構成される区間イベント系列パターン（端点系列）

$$\alpha = \langle ep_1^\alpha \oplus_1^\alpha ep_2^\alpha \dots \oplus_{2I-1}^\alpha ep_{2I}^\alpha \rangle \\ \beta = \langle ep_1^\beta \oplus_1^\beta ep_2^\beta \dots \oplus_{2J-1}^\beta ep_{2J}^\beta \rangle$$

を考える。このとき、 $I \leq J$ かつ α が順序関係を保ちつつ β に現れる、すなわち条件

$$\forall i (1 \leq i \leq 2I) : ep_i^\alpha \equiv ep_{j_{2I}}^\beta \wedge \\ \forall i (2 \leq i \leq 2I) : ep_{i-1}^\alpha \oplus ep_i^\alpha = ep_{j_{i-1}}^\beta \oplus ep_{j_i}^\beta$$

を満たす $j_x (1 \leq x \leq 2I)$ が存在するとき、 α は β 中に出現するといいい、 $\alpha \sqsubseteq \beta$ と表記する。なお、 $ep^\alpha \equiv ep^\beta$ は、端点 ep^α と ep^β を有する区間イベント e^α, e^β において、イベントラベルと属性が一致すること ($l(ep^\alpha) = l(ep^\beta) \wedge a(ep^\alpha) = a(ep^\beta)$) を表す。また、 $\alpha \sqsubseteq \beta$ のとき、 α は β よりも一般的である (β は α より特殊である) と言う。

軌跡データ集合 TR から得られる区間イベント系列 $E = M(TR)$ に対し、区間イベント系列パターン α の出現を、集合

$$O(E, \alpha) = \{o \subseteq E \mid o \sqsubseteq \alpha \wedge |o| = |\alpha|\}$$

と表記する。また、区間イベントの最大時間差に関するパラメタを k としたとき、時間制約 k を満たす、すなわち末尾の端点の（終了）時刻と先頭の端点の（開始）時刻の差が k 以内である α の出現の集合を

$$O_k(E, \alpha) = \left\{ o \in O(E, \alpha) \mid \begin{array}{l} o = \langle ep_1^o \oplus_1^o \dots \oplus_{2|o|}^o \rangle \\ \wedge t(ep_{2|o|}^o) - t(ep_1^o) \leq k \end{array} \right\}$$

と定義する．一方，出現の元となる部分軌跡集合の広がり $d_x(x = 1, 2)$ の上限を s としたとき，空間制約 s を満たす α の出現の集合を

$$O^s(E, \alpha) = \{o \in O(E, \alpha) \mid d_x(\{Em(p) \mid p \in o\}) \leq s\}$$

と定義する．また，時間制約 k ，空間制約 s の双方を満たす α の出現の集合を

$$O_k^s(E, \alpha) = O_k(E, \alpha) \cap O^s(E, \alpha)$$

と表記する．

3.3 パターンの支持度

これまでに，単一の点系列を対象とした系列パターンの評価値がいくつか提案されている [7, 8, 9]．本論文ではこのうち，[7] で提案された系列先頭頻度及び系列全体頻度の考え方を拡張し，軌跡データにおける部分区間イベント系列パターンへと応用する．

軌跡データ集合 TR から得られる区間イベント系列 $E = M(TR)$ に対し，区間イベント系列パターン α の条件なし先頭支持度を

$$H_Freq(E, \alpha) = |\{t(ep_1^o) \mid \langle ep_1, \dots \rangle \in O(E, \alpha)\}|$$

と定義する．これは，各出現における最初の区間イベント開始時間を集計することに相当する．

また同様に，時間・空間に関する制約を考慮した上で区間イベント開始時間を集計することで，時間制約 k のもとでの α の先頭頻度

$$H_Freq(E, \alpha, k) = |\{t(ep_1^o) \mid \langle ep_1, \dots \rangle \in O_k(E, \alpha)\}|$$

と，時間制約 k と空間制約 s のもとでの α の先頭頻度

$$H_Freq(E, \alpha, k, s) = |\{t(ep_1^o) \mid \langle ep_1, \dots \rangle \in O_k^s(E, \alpha)\}|$$

をそれぞれ定義する．なおこれらの定義は，時間幅 k の滑走窓を考え，各窓の先頭において出現が現れた回数を数えていることに相当する．

各系列先頭頻度を用い，パターン α の系列全体頻度を， α の任意の部分パターンの先頭頻度の最小値

$$T_Freq(E, \alpha, k) = \min_{\gamma \sqsubseteq \alpha} (H_Freq(E, \gamma, k))$$

$$T_Freq(E, \alpha, k, s) = \min_{\gamma \sqsubseteq \alpha} (H_Freq(E, \gamma, k, s))$$

と定義する．

詳細な証明は省略するが， $T_Freq(E, \alpha, k)$ に関しては，パターンの特殊化に関する逆単調性

$$\forall \alpha, \beta : \alpha \sqsubseteq \beta \rightarrow T_Freq(E, \alpha, k) \geq T_Freq(E, \beta, k)$$

が成り立つが， $T_Freq(E, \alpha, k, s)$ に関しては逆単調性が成り立つとは限らない点に注意が必要である．

3.4 区間イベント系列パターンの列挙

本論文でのデータマイニング問題を，以下のように定義する．

軌跡データの集合 TR から得られる区間イベント系列 $E = M(TR)$ と，時間に関する制約 k ，空間に関する制約 s ，支持度に関する制約 σ ，パターンを構成する区間イベントの最大数 N を入力とし，条件 $T_Freq(E, \alpha, k, s) \geq \sigma$ を満たすイベント数 N 以下の区間イベント系列パターン α をすべて列挙する．

この問題に対し，本論文では，系列パターン列挙アルゴリズム $TPrefixSpan$ [6] のパターン拡張と，時間幅を考慮した系列全体頻度による枝刈り，時間・空間を考慮した頻度によるチェックを組み合わせた幅優先探索に基づく列挙アルゴリズム TP_LS^{TR} を開発した．図 2 にアルゴリズム TP_LS^{TR} を示す．

$TP_LS^{TR}(E, \sigma, k, s, N)$	
Input	
	E : 区間イベント系列
	σ : 最小支持度
	k : 時間制約
	s : 距離制約
	N : 区間イベント数制約
Output	
	P : 区間イベント系列パターンの集合
1:	$P \leftarrow \{\}$
2:	$P_1 \leftarrow$ 頻出なイベントの集合
3:	$P \leftarrow P \cup P_1$
4:	$TP_LS^{TR}(E, \sigma, k, s, 1, N, P_1, P)$
5:	return P
$TP_LS^{TR}(E, \sigma, k, s, L, N, C, P)$	
1:	$C' \leftarrow \{\}$
2:	for each $p \in C$
3:	for each $p' \in extension(p)$
4:	if $T_Freq(E, p', k) \geq \sigma$
5:	$C' \leftarrow C' \cup \{p'\}$
6:	if $T_Freq(E, p', k, s) \geq \sigma$
7:	$P \leftarrow P \cup \{p'\}$
8:	if $C' \neq \{\} \wedge L < N$
9:	$TP_LS^{TR}(E, \sigma, k, s, L + 1, N, C', P)$

図 2: アルゴリズム TP_LS^{TR}

区間イベント系列パターン α の系列全体頻度 (T_Freq) の計算には， α の部分パターンが必要とされる．このことに着目し，アルゴリズム TP_LS^{TR} では，長さ $($

間イベント数)の短いパターンから順に区間イベント系列パターンを列挙している。

また各パターン p の拡張は、関数 $extension$ によって行われる。関数 $extension(p)$ は、サイズ n のパターン p に、区間イベント $e = \langle l_e, e^+, e^-, a_e \rangle$ を追加することで新たなパターン p' (の集合) を生成する。この関数は、TPrefixSpan[6]における拡張関数と同様、 p の端点系列において最後に現れる開始点を表す端点 ep_{last}^+ より後ろの任意の位置 i に e の開始点 ep^+ を追加・挿入する。また e の終了点 ep^- を、 ep^+ を追加・挿入した位置 i より後ろの任意の位置 j に追加・挿入する。追加・挿入位置 i, j と、端点間の時間に関する2項関係 $\oplus \in \{<, =\}$ のすべての組み合わせに対してパターン p を拡張することで、 p より特殊なパターンの集合を生成する。

生成された各パターン p' は、時間制約を考慮した系列全体支持度 $T_Freq(E, p', k)$ の値が σ 以上の場合に C' へと追加され、次の拡張に利用される。一方、支持度が σ 未満の場合は C' へ追加されず、結果として枝刈りされることになる。なお、時間制約を考慮した系列全体支持度は逆単調性を有するため、この枝刈りにより列挙の完全性が損なわれることはない。また、 p' に対する時間・空間を考慮した系列全体支持度 $T_Freq(E, p', k, s)$ の値が σ 以上であれば、 p' は求めるべき頻出パターンであるので、それを p へと追加する。

4 評価実験と考察

提案手法の有効性を評価するため、Java 言語を用いて BIRCH[5] 及び TP_LS^{TR} を実装し、サッカー軌跡データを用いて評価を行った。

4.1 データセット

実験には、2015年から2016年に行われたJリーグのリーグ戦9試合に関するデータ¹を利用した。具体的には、ボールと選手の軌跡データ(0.2秒間隔)とそれに付随するプレイ内容(パスやシュートなど)を利用する。軌跡データ数は(11人×2チーム+ボール)×9試合の合計207である。

今回の実験では、時間幅を{3, 4, 5}秒、スライド幅を2秒として各軌跡データを部分軌跡へと分割し、それらをBIRCH[5]を用いてクラスタリングすることで区間イベント(モチーフ)の抽出を行った。また各区間イベントの属性 a には、

1. フィールド上のエリア

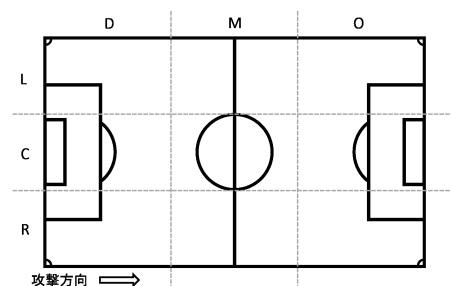


図3: フィールド上のエリア

表1: 得られた区間イベント数

時間幅	3	4	5
区間イベント数	429,169	387,513	218,485
区間イベントの種類数	13	15	23

2. 選手と対応したプレータグ

の2つを採用した。ここで「フィールド上のエリア」とは、プレイが行われたエリアを表す。図3に示す通り、サッカーフィールドを縦横それぞれ等間隔に3分割、計9つのエリアに区切る。縦方向には、自陣のゴール側を守備エリア、センターサークル付近を中盤エリア、敵陣のゴール側のエリアを攻撃エリアとして、それぞれのラベルをD, M, Fとする。また、横方向は左側のエリアからラベルをL, C, Rとする。プレイエリアは縦と横のラベルを連結したものであり、例えば、図3中の一番左かつ一番上のエリアのラベルは「DL」となる。一方、「選手と対応したプレータグ」とは選手の部分軌跡の開始時間と終了時間内に現れるプレータグとする。

表1に、これらの変換を通じて得られた区間イベント系列の大きさ(区間イベント数)と区間イベントの種類数を示す。

4.2 評価実験と考察

実験では、最小支持度とウィンドウ幅、パターンサイズの最大値をそれぞれ $\sigma \in \{100, 150, 200\}$ と $k \in \{10, 15, 20\}$ 秒および $N = 5$ に設定し、得られるパターン数と実行時間を集計した。実験には、Linuxマシン(CPU: Intel Xeon 3.20GHz, メインメモリ: 24GB)を使用した。実験結果を表2および表3に示す。なお、表2は区間イベント間距離 d_1 を採用し、 $s = 20m$ とした場合の結果である。また表3は距離 d_2 を用いて、 $s = 10m$ とした場合の結果である。

表2, 表3より、区間イベント集合間の距離の定義に関わらず、現実的な計算時間でパターンの抽出が達成できていることが分かる。また、パラメータ σ, k に対

¹データスタジアム株式会社
(<https://www.datastadium.co.jp>) より

する抽出パターン数の変化に関し、 d_1 と d_2 とで類似する傾向が見られた。一方、計算時間に関しては、すべての場合において d_1 を採用した方が速いことが分かる。

次に、図 4、図 5 に区間イベント集合間の距離定義にそれぞれ d_1, d_2 を採用し、ウィンドウ幅 $k = 10$ 、最小支持度 $\sigma \in \{100, 150, 200\}$ の場合に得られたパターン数のサイズ（イベント数）毎の分布を示す。図 4、図 5 より、区間イベント集合間の距離定義、最小支持度 $\sigma \in \{100, 150, 200\}$ に関わらずサイズ 3、4 の場合にパターン数が多い傾向があることが分かる。

表 2: 距離 $d_1, s = 20m$ の実験結果

$k \setminus \sigma$	パターン数			実行時間(分)		
	100	150	200	100	150	200
10	4,713	2,842	896	41	35	11
15	89,248	54,698	31,513	169	127	96
20	796,149	625,594	283,395	579	386	273

表 3: 距離 $d_2, s = 10m$ の実験結果

$k \setminus \sigma$	パターン数			実行時間(分)		
	100	150	200	100	150	200
10	5,817	3,782	1,059	61	41	27
15	94,732	71,701	38,315	472	289	163
20	923,317	722,619	346,521	725	561	428

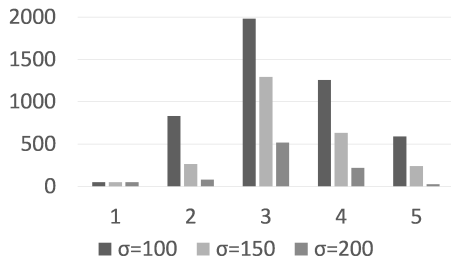


図 4: 距離 $d_1, k = 10$ におけるパターン数のパターンサイズごとの分布

次に、 $\sigma = 100, k = 10$ の場合に得られたパターンの例として、以下に示す 2 つのパターン P_D, P_O について簡単に考察する。なお、パターン中における各イベントラベル（数字とアルファベット）は、モチーフの種類（数字）とフィールドの位置（アルファベット）の組み合わせである。

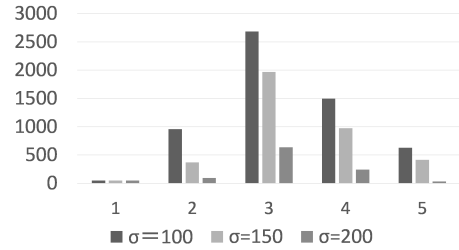


図 5: 距離 $d_2, k = 10$ におけるパターン数のパターンサイズごとの分布

$$p_D = \langle 12LM^+ < 9LM^{1+} < 9LM^{2+} < 12LM^- < 9LM^{1-} < 9LM^{2-} \rangle$$

$$p_O = \langle 1CD^+ = 3CD^+ < 1LD^+ < 1CD^- = 3CD^- < 1LD^- \rangle$$

パターンに対応する部分軌跡の例を図 6 と図 7 に示す。 p_D （図 6）では、自陣において 3 名のディフェンス選手が連動し、ラインを上げるような動きが確認できる。また p_O （図 7）では、自陣からオーバーラップのような動き（ラベルが 12:LM）とスペースを空ける動き（ラベルが 9:LM）が確認できる。以上より、両パターンが守備または攻撃に関する戦略に対応していることが示唆される。

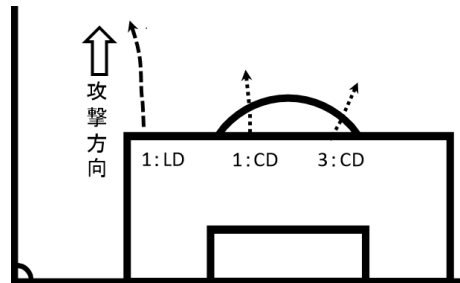


図 6: DF のラインを上げる動き

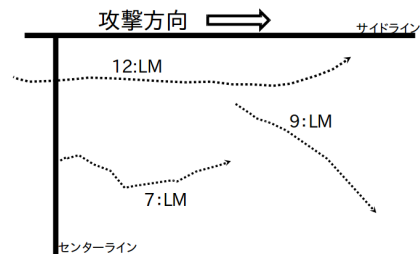


図 7: サイドを駆け上がる動き

5 まとめ

本論文では、サッカーにおける選手同士の意味のある動き、すなわち守備・攻撃に関する戦術を抽出することを目的に、軌跡データを対象としたパターン発見技術を提案した。具体的には、選手の軌跡データから意味のある部分軌跡をモチーフとして抽出し、それらを区間イベントへと変換した上で、頻出系列パターン発見技術を適用している。

今回開発したパターン発見アルゴリズムでは、時間と空間に関してパラメタが必要となる。しかし、サッカーにおける戦術的な動作は時間的空間的スケールが多様であり、意味のあるパターンの抽出には適切なパラメタの設定が不可欠となる。従って、今後の課題の一つとして、パラメタ値の変更に伴う結果の違いを確認するとともに、パラメタの最適化について検討を行う必要がある。また、頻出パターン発見の特性上、多くの類似パターンが抽出されてしまう。得られたパターンを定量的に評価するとともに、統計的有意性によるフィルタリングを行うことも重要な課題の一つである。

参考文献

- [1] Kang, C.-H., Hwang, J.-R. and Li, K.-J.: Trajectory analysis for soccer players, *Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on*, IEEE, pp. 377–381 (2006).
- [2] Lucey, P., Oliver, D., Carr, P., Roth, J. and Matthews, I.: Assessing team strategy using spatiotemporal data, *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 1366–1374 (2013).
- [3] Bialkowski, A., Lucey, P., Carr, P., Matthews, I., Sridharan, S. and Fookes, C.: Discovering Team Structures in Soccer from Spatiotemporal Data, *IEEE Transactions on Knowledge Data Engineering*, Vol. 28, No. 10, pp. 2596–2605 (2016).
- [4] 鈴木 湧人, 尾崎 知伸: 単一区間イベント系列からの頻出系列パターンマイニング, 人工知能学会第 108 回知識ベースシステム研究会, SIG-KBS-B504-05, pp. 24–29 (2016).
- [5] Zhang, T., Ramakrishnan, R. and Livny, M.: BIRCH: A New Data Clustering Algorithm and Its Applications, *Data Mining and Knowledge Discovery*, Vol. 1, No. 2, pp. 141–182 (1997).
- [6] Wu, S.-Y. and Chen, Y.-L.: Mining nonambiguous temporal patterns for interval-based events, *IEEE transactions on knowledge and data engineering*, Vol. 19, No. 6, pp. 742–758 (2007).
- [7] 宏治岩沼: テキスト系列マイニングにおける有用性尺度について, 人工知能学会誌, Vol. 27, No. 2, pp. 136–145 (2012).
- [8] Achar, A., Laxman, S. and Sastry, P. S.: A unified view of the apriori-based algorithms for frequent episode discovery, *Knowledge and Information Systems*, Vol. 31, No. 2, pp. 223–250 (2012).
- [9] Yang, J., Wang, W. and Yu, P. S.: Mining Surprising Periodic Patterns, *Data Mining and Knowledge Discovery*, Vol. 9, No. 2, pp. 189–216 (2004).
- [10] Agarwal, P., Shroff, G., Saikia, S. and Khan, Z.: Efficiently Discovering Frequent Motifs in Large-scale Sensor Data, *Proceedings of the Second ACM IKDD Conference on Data Sciences*, pp. 98–103 (2015).