

パス経路の類型化に基づくサッカー試合記録データからの知識獲得

Grouping of Pass Sequences in Soccer Game Records

平野章二*¹
Shoji Hirano

津本周作*¹
Shusaku Tsumoto

*¹ 島根医科大学医学部医療情報学講座

Department of Medical Informatics, Shimane Medical University

This paper presents a method for finding interesting pass patterns from soccer game records. Given a set of temporal sequences recording positions of a ball and actions of players, it generates groups of sequences containing similar pass routes by performing multiscale matching and hierarchical clustering.

1. はじめに

サッカー、野球等の試合記録データは、チームの戦いを科学的に検証し、今後強化すべき弱点や課題を浮かび上がらせるための重要な素材である。しかしながら、従来の解析は打率や盗塁数、シュート数などを中心とした基礎的ものが中心であり、データのもつ情報を十分に活用できていない現状にある。

本稿では、サッカーの試合記録データを対象とし、戦略上重要な要素となる「パス回し」、特に、ゴールに結びつくパス回しの特徴について、パス経路の類型化を通じて知識を獲得する手法を提案する。本研究で用いる試合記録データにおいて、パスの経路はパス行為の発生時間を媒介変数とする2次元の位置座標系列として表現されるため、その類型化は基本的に時系列の類型化と解釈することができる。しかし、パス経路の類型化は通常の観測データの類型化に比べ以下の二点において注意を要する。まず第一は、パス系列(経路)の長さや系列の構成点数、点間の間隔が全ての系列で不均一であることである。パス経路は連続したパス行為を1つの系列として連結したものであるが、パスを中継する人数や個々のパスの距離は当然異なるため、不等間隔サンプリングの時系列の類型化問題として取り扱わなければならない。第二は、観察スケールを変化させる必要性である。これは、パス回しが戦略的な要素、すなわち、サイドから切り込む形で攻める等の広範囲な行動パターンを反映すると同時に、ワンツーパスで相手を交わす方法を多用しているなど、局所的な行動パターンの特徴も反映することから、それらの類似性をスケールを変えて比較する必要性による。我々の提案する方法は、図形認識の立場からこの問題に取り組むもので、多重スケールマッチング [1, 2] による系列比較を基礎としている。これにより、不等間隔のパス系列において、広範な行動パターンの類似性と局所的な行動パターンの類似性を同時に比較することを試みるものである。

2. データ構成と前処理

本研究では、2002年5月から6月にわたって開催されたFIFAワールドカップの予選及び決勝トーナメント全64試合の試合記録データを用いた。データは株式会社アソボウズからご提供いただいたもので、試合中に発生したパス、トラップ、シュート等のイベントそれぞれについて、その行為者、位置、時間等の情報が時系列的に記録されている。表1にデータの概略を示す。同表において、Serはシリーズ番号であり、連続

表 1: 試合記録データの概略

| Ser | Time | Action | T_1 | P_1 | T_2 | P_2 | X_1 | Y_1 | X_2 | Y_2 |
|-----|----------|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 20:28:12 | KICK OFF | セネガ | 10 | | | 0 | -33 | | |
| 1 | 20:28:12 | PASS | セネガ | 10 | セネガ | 19 | 0 | -50 | -175 | 50 |
| 1 | 20:28:12 | TRAP | セネガ | 19 | | | -175 | 50 | | |
| 1 | 20:28:12 | PASS | セネガ | 19 | セネガ | 14 | -122 | 117 | 3004 | 451 |
| 1 | 20:28:14 | TRAP | セネガ | 14 | | | 3004 | 451 | | |
| ⋮ | | | | | | | | | | |
| 169 | 22:18:42 | P END | フラン | 15 | | | 1440 | -685 | | |

したイベントに同一番号が振られている。以下、イベント発生時刻 (Time)、イベント種類 (Action)、行為者1のチーム (T_1) と背番号 (P_1)、行為者2のチーム (T_2) と背番号 (P_2)、行為者1の位置座標 (X_1, Y_1)、行為者2の位置座標 (X_2, Y_2) と続く。行為者1および2は行為の主体者と対象者を示しており、例えばパスの場合は出し手が行為者1、受け手が行為者2に相当する。また、X軸およびY軸はコート中央を原点として、長辺および短辺方向に対応し、原点はコート中央である。表1の2行目の例は、(0, -50) に位置するセネガルの10番の選手から (-175, 50) に位置するセネガルの19番の選手に対してパスが出された、と解釈される。

このような試合記録データから、以下の手続きによりパス経路をしめる系列群を抽出する。

1. IN GOAL を含むシリーズを抽出する。
2. 同一シリーズの中で、連続した PASS イベントを切り出す。ここでは、系列の過分割をさけるため、パスと連続するトラップ、ドリブル、センタリング、クリア、ブロックなどのイベントがある場合はそれらを含む形で切り出すこととしている。また、インターセプトは出し手と受け手のチームが異なるパスとして記述されているが、ここでは区別することなく一つのパス行為として含めている。
3. 切り出したイベントから、行為者1の位置に相当する X_1 及び Y_1 を取り出して連結し、 $p(t) = \{(X_1(t), Y_1(t)) | 1 \leq t \leq T\}$ として時系列化する。(簡単のため、以後は $X_1(t)$ および $Y_1(t)$ をそれぞれ $x(t)$ 、 $y(t)$ と表記する。)

図1に、上記手続きにより切り出されたパス経路の一例を示す。また、表2にこの系列と対応する区間のデータを示す。図1において、縦軸はゴールを結ぶ軸であり、上端の+5500前後がフランスのゴール位置に相当し、下端の-5500前後がセネガルのゴール位置に相当する。この例は、座標 (-333, 3877) のフランス18番の選手からパスが始まり、コート中央でセネガルがパスカットをした後、コート左側から攻めあがりセンタリン

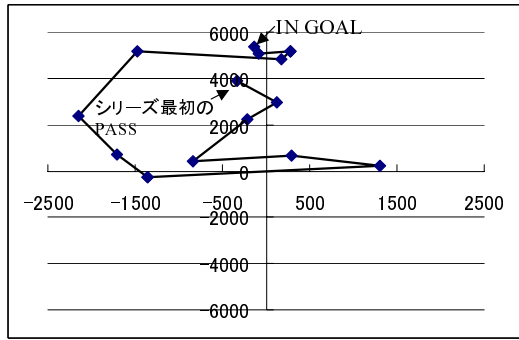


図 1: パス系列の例.

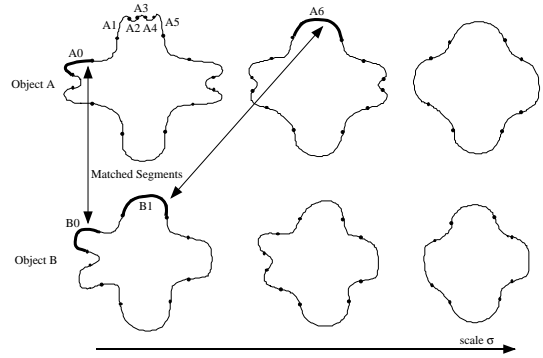


図 2: 多重スケール比較の例.

表 2: 図 1 に対応する記録データ

| Ser | Time | Action | T_1 | P_1 | T_2 | P_2 | X_1 | Y_1 | X_2 | Y_2 |
|-----|----------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|
| 47 | 20:57:07 | PASS | フラン | 16 | フラン | 18 | -333 | 3877 | 122 | -2958 |
| 47 | 20:57:08 | PASS | フラン | 18 | フラン | 17 | 122 | 2958 | -210 | -2223 |
| 47 | 20:57:10 | DRIBBLE | フラン | 17 | | | -210 | 2223 | -843 | -434 |
| 47 | 20:57:14 | PASS | フラン | 17 | フラン | 4 | -843 | 434 | 298 | -685 |
| 47 | 20:57:16 | PASS | フラン | 4 | フラン | 6 | 298 | 685 | 1300 | 217 |
| 47 | 20:57:17 | TRAP | フラン | 6 | | | 1300 | 217 | | |
| 47 | 20:57:19 | CUT | セネガ | 6 | | | -1352 | -267 | | |
| 47 | 20:57:19 | TRAP | セネガ | 6 | | | -1352 | -267 | | |
| 47 | 20:57:20 | PASS | セネガ | 6 | セネガ | 11 | -1704 | 702 | -2143 | 2390 |
| 47 | 20:57:21 | DRIBBLE | セネガ | 11 | | | -2143 | 2390 | -1475 | 5164 |
| 47 | 20:57:26 | CENTERING | セネガ | 11 | | | -1475 | 5164 | | |
| 47 | 20:57:27 | CLEAR | フラン | 17 | | | 175 | 4830 | | |
| 47 | 20:57:27 | BLOCK | フラン | 16 | | | 281 | 5181 | | |
| 47 | 20:57:27 | CLEAR | フラン | 16 | | | 281 | 5181 | | |
| 47 | 20:57:28 | SHOT | セネガ | 19 | | | -87 | 5081 | | |
| 47 | 20:57:28 | IN GOAL | セネガ | 19 | | | -140 | 5365 | | |

グを上げ、ゴール前でフランスのブロックを受けた後にシュートが決まった場面を示している。

同様の処理を全試合のゴールシリーズに適用し、 N 個のゴールシリーズに対応する N 個のパス経路群 $P = \{p_i | 1 \leq i \leq N\}$ を得る。ここで、 p_i の添え字 i は i 番目のゴールシリーズに対応するパス系列を示す。

3. 系列の比較と類型化

前節において切り出したパス系列群 P に含まれる系列のペア $\{(p_i, p_j) \in P | 1 \leq i < j \leq N\}$ に対し、多重スケールマッチングを適用して類似度を算出し、類似度行列を作成する。構築された類似度行列を入力として階層的クラスタリング法を適用し、系列を類型化する。

多重スケールマッチングは、2 つの系列を A, B を入力とし、それぞれの部分系列の対応関係を異なる視野スケールに渡って探索し決定する方法である。以下に概略を示すが、詳細については文献 [2] を参照されたい。

簡単のため、 p_i 及び p_j をそれぞれ A, B に置き換える。まず、系列 A の X_1 の系列 $x(t)$ を考える。このとき、スケール σ における系列 $X(t, \sigma)$ を、 $x(t)$ とスケールファクター σ をもつガウス関数 $g(t, \sigma)$ との畳み込みとして以下のように表現する。

$$X(t, \sigma) = \int_{-\infty}^{+\infty} x(u) \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-u)^2/2\sigma^2} du \quad (1)$$

ここで、ガウス関数は畳み込みにおける加算重みの分布を規定する。その特性から明らかなように、 σ が大きいほど遠方の加算重みが増し、より広範にわたる平滑化が行なわれる。すなわち、スケールの増加に従い近傍値との平滑化が進み、より変曲点の少ない滑らかな系列が得られる。 Y_1 の系列 $y(t)$ について

も同様に $Y(t, \sigma)$ を得る。系列上の各点における曲率は次式で与えられる。

$$K(t, \sigma) = \frac{X'Y'' - X''Y'}{(X'^2 + Y'^2)^{3/2}}, \quad (2)$$

ここで、 X', X'' は $X(t, \sigma)$ の t による 1 次および 2 次微分である。次に、曲率の符号の変化から系列上の変曲点の位置を求め、隣接する変曲点を両端とする部分輪郭、すなわち凹凸セグメントを構築する。スケール $\sigma^{(k)}$ におけるパス系列 $A^{(k)}$ を R 個のセグメントの集合とすると、 $A^{(k)} = \{a_i^{(k)} | i = 1, 2, \dots, R^{(k)}\}$

ここで、 $a_i^{(k)}$ はスケール $\sigma^{(k)}$ における i 番目のセグメントを示す。ここまでの処理をもう 1 つの入力系列 B に適用し、 S 個のセグメントからなる系列組 $B^{(h)} = \{b_j^{(h)} | j = 1, 2, \dots, S^{(h)}\}$

を得る。ここで、 $\sigma^{(h)}$ は系列 B の観察スケールである。多重スケールマッチングにおけるマッチング手続きは、図 2 に示す通り、 A, B を構成する全てのセグメント組から、対応漏れの無い、かつこのセグメント組の相違度の総和を最小にする組を探索することに相当する。部分系列 $a_i^{(k)}$ と $b_j^{(h)}$ の相違度 $d(a_i^{(k)}, b_j^{(h)})$ は次式の通りである。

$$d(a_i^{(k)}, b_j^{(h)}) = \frac{|\theta_{a_i}^{(k)} - \theta_{b_j}^{(h)}|}{\theta_{a_i}^{(k)} + \theta_{b_j}^{(h)}} \left| \frac{l_{a_i}^{(k)}}{L_A^{(k)}} - \frac{l_{b_j}^{(h)}}{L_B^{(h)}} \right| \quad (3)$$

ここで、 $\theta_{a_i}^{(k)}$ および $\theta_{b_j}^{(h)}$ は各セグメントに沿った接ベクトルの回転角、 $l_{a_i}^{(k)}$ および $l_{b_j}^{(h)}$ は各セグメントの長さ、 $L_A^{(k)}$ および $L_B^{(h)}$ は対象系列 A, B のスケール $\sigma^{(k)}, \sigma^{(h)}$ における総セグメント長である。

4. 実験結果と結び

提案法を 64 試合のデータに適用した類型化の結果について、講演当日において示す予定である。

参考文献

- [1] F. Mokhtarian and A. K. Mackworth (1986): Scale-based Description and Recognition of planar Curves and Two Dimensional Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(1): 24-43
- [2] N. Ueda and S. Suzuki (1990): A Matching Algorithm of Deformed Planar Curves Using Multiscale Convex/Concave Structures. IEICE Transactions on Information and Systems, J73-D-II(7): 992-1000.