

ベイジアンネットを用いたサッカーの有効戦術推定

岡本健太郎

明治大学総合数理学部先端メディアサイエンス学科

概要: サッカーのような攻守が目まぐるしく変化するスポーツは、戦術のモデルの構成が困難である。そこで、本稿では日本代表のワールドカップに関する試合のデータをベイジアンネットによって学習させ、オフENSEの戦術を決定するモデルを提案する。学習データを増加させるために、シュートの入ったシーンではなく、シュートの是非を用いて戦術の有効無効を判定しデータを作成した。構築したモデルを用いて、エリアごとに分類した推定結果を報告する。

キーワード: データマイニング, ベイジアンネット

Tactical estimation of soccer using Bayesian networks

Kentaro Okamoto

Department of Frontier Media Science, School of Interdisciplinary
Mathematical Sciences, Meiji University

Abstract: Sports in which phase of offense and defense changes frequently. It is difficult to make model structure of tactics of sports. Then, we suggest a tactical decision offensive model by learning offensive elements from Japanese representative match of the world cup data using Bayesian networks. We obtain data of event including shooting, invalid. Moreover, we consider accuracy estimate tactics in categorized areas.

Keywords: Datamining, Bayesian network

1. はじめに

スポーツを強くするためには、得られたデータを客観的、かつ科学的に分析するアプローチが重要である。例えば、バレーボール女子日本代表では、体に心拍数を把握できるウェアラブルデバイスを装着し、その心拍数を監督が見ながら選手の調子を把握して選手交代を行っている。そこで本研究では、サッカー日本代表の戦術を一般的に用いられるベイジアンネットを用いて推定し、有効だった戦術を報告する。

2. 提案手法

2.1 ベイジアンネット

問題領域の確率変数と変数間の関係をモデルで表したものである。事象の確率変数をノードとし、ノード間の相関や依存関係を有向辺で結んだ非循環有向グラフ (Directed Acyclic Graph) で表す。リンク元のノードを親ノード、リンク先を子ノードと呼ぶ。結ばれたリンクの強さを各ノードの条件付き確率表 (Conditional Probability Table) で表す。CPT の計算は DAG の構造を決定することで計算される。従って、ベイジアンネットを戦術推定に用いるためにはノードの決定、及び DAG の構築を行う必要がある。

2.2 ノードの決定, 定義

表1 モデルのノード名とその状態

変数	ノード名	状態	値域
X_1	エリア	ボールを奪取した位置	高, 中, 少
X_2	連携数	ラストプレーまでに經由した人数	少少, 少, 多, 多多
X_3	進行度	エリアの横移動を1, 縦移動を攻撃方向によって ± 10 した値	大大, 大, 小, 小小
X_4	ラストプレー	相手がボールに触れる直前のプレー	※後述
X_5	結果	シュートの是非	1,0

ノードはオフENSEの途中で変化する量である。エリア X_1 はフィールドを縦に5分割, 横に3分割したもので、攻撃の起点となる部分を示す、攻撃方向において右から1~5, 6~10, 11~15の値を割り当て、その値に応じて高, 中, 低の値を格納した。連携数 X_2 はパスの回数を示す。進行度 X_3 はボールの移動を示しており、エリアによって移動の度合いを変化させた。ラストプレー X_4 は相手がボ

ールを触る直前のプレーを示す。図1にエリア、表2にラストプレーの定義を示す。※フリーキックにはキックオフ、スローイン、コーナーキック、ゴールキックも含む。

		1	6	11	
		2	7	12	
相手	ゴール	3	8	13	味方
		4	9	14	ゴール
		5	10	15	

図1 エリアの値の定義

表2 ラストプレーの定義

X ₄	プレー	定義
P ₁	ショートパス	横に2, 縦に1エリア以内の範囲のパス
P ₂	ドリブル	一人で3回以上触れるプレー
P ₃	フリーキック	相手の反則から開始したプレー
P ₄	ロングパス	ショートパス以上の距離のパス

変数 X₁, X₂, X₃は大きな値域の連続量となるので、それぞれのデータを非連続量の値に格納した。サッカーでは相手ゴールに近い場所を高い位置と表現するため、1~5を高く、6~10を中、11~15を低とした。同様に X₂, X₃はそれぞれ {多多(10~), 多(6~9), 少(4~5), 少少(~3)}, {大大(23~), 大(15~22), 小(11~14), 小小(~10)} として値域によって非連続の値を格納した。また、X₅はシュートの是非をブーリアンで示し、1を true とした。

2.3 DAGの構築

DAGの構築には、Rのパッケージの一つである deal を使用した。DAGの構造学習は、ネットワークスコア

$$S(G)=p(G,d)=p(d|G)p(G)$$

を利用して計算されるベイズファクター

$$S(G)/S(G^*)=p(G|d)/p(G^*|d)$$

によって評価される。ただし、G, G*をDAGを構築するベイズネットワークモデル、dを入力されたデータとする。このネットワークスコアの評価にはさまざまな方法が知られているが、dealではよく知られている探索アルゴリズムの一つである欲ばり法が使われている。欲ばり法については参考文献 [2] の pp232 を参照されたい。

5つの確率変数を考えたとき、得られるモデルは無数に存在するため、計算量を減らすために事前にあり得ない有向辺を削除する必要がある。時系列で考えると、ボールを

奪取した際にエリアが決定し、相手選手がボールに触れた時点で連携数と進行度が決定する。同様にして確率変数間の時系列を考慮し、図2のような階層を定め、X₅からすべてのノード、X₄から X₁, X₂, X₃への有向辺などを削除した。

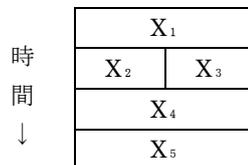


図2 確率変数の階層構造

2.4 学習結果

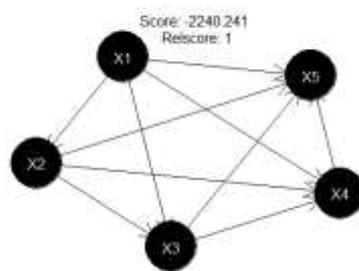


図3 学習結果のDAG

図3に日本代表のデータについて求めたベイズモデルを示す。シュートできる確率を最大化する条件は P (X₁=低, X₂=多多, X₃=大, X₄=ショートパス|X₅=1) で 3.04%であった。次点以降、及びエリアごとの上位10確率は表3、表4で示す。表3の総合的な順位で見ると、ラストプレーの値はまちまちだが、エリアについては高い位置でボールを奪取したとき、ほとんどシュートできていないことが分かる。表4より最も確率が高い事象でも 0.33%と、非常に低い。

サッカー日本代表は細かくパスを回して相手をかく乱しているイメージが強かったが、2位、4位はラストプレーがドリブルであった。

エリアごとの特徴としては、エリア=高の時、3事象すべて連携数が非常に少なく、進行度が小さい。すなわち、高い位置でボールを奪取したらなるべくボールを回さずに少ない人数で攻撃したほうがシュートできる確率が高い。この時取られている戦術はカウンターが考えられる。カウンターとは相手がディフェンスの陣形を組み立てる前に攻め崩してしまおうという戦術で、俗には速攻と呼ばれているものである。

エリア=中の時、上位2事象は連携数が多く、進行度が大きい。すなわち、中盤でボールを奪取したら、多めの人を経由し、シュートのために高い位置に移動するため、横移動は5回以上12回以内でラストプレーに繋げるとシュート

できる確率が高い。この時取られている戦術はクロスが考えられる。クロスとは図4に示すようなプレーであり、図に分かる通り、クロス을上げてシュートするだけで横移動が2回ある。また、クロスをする前にはディフェンスのプレーヤーが横パスでどのサイドから攻めるか選択する段階がしばしば見られるため進行度が増えやすい戦術であるといえる。また、エリア=中のあるときのラストプレーの1位がロングパスであるのもクロスを上げている可能性が高いことを示唆する要因となっている。

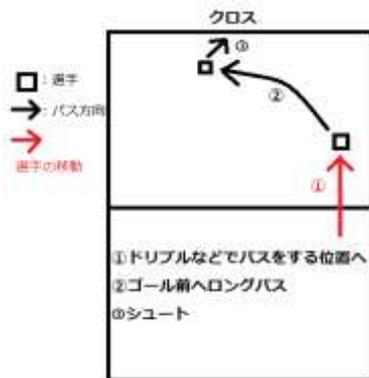


図4 クロスの図解

エリア=低のとき、3事象ともに連携数がとても多く、進行度も大きい、またはとても大きい。すなわち、低い位置でボールを奪取したらじっくりボールを回しながら横移動が4回以内だとシュートできる確率が高い。このとき取られている戦術はカウンターが考えられる。先述した通り、カウンターとは相手の守備の陣形が組み立てられる前に攻め崩そうとする戦術であるが、それゆえになるべく最短距離の縦移動で攻めることが多い。ラストプレーがショートパス、ドリブルであるのもカウンターで攻めている可能性が高いことを示唆している。

表3 シュートできる確率が高い10事象

順位	確率	エリア	連携数	進行度	ラストプレー
1位	0.0304	低	多多	大	ショートパス
2位	0.0284	低	多多	大	ドリブル
3位	0.0282	中	多	大	ロングパス
4位	0.0269	中	多	大大	ドリブル
5位	0.0256	中	多	大大	ショートパス
6位	0.0255	中	多	大	ショートパス
7位	0.0221	低	多多	大大	ロングパス
8位	0.0212	低	多多	大大	ドリブル
9位	0.0206	低	多多	大大	フリーキック

10位	0.0179	低	多多	大	フリーキック
-----	--------	---	----	---	--------

表4 エリアごとのP(結果=YES)を大きくする要因上位3事象

順位	確率	連携数	進行度	ラストプレー
高1位	0.0033	少少	小	フリーキック
高2位	0.0031	少少	小小	ショートパス
高3位	0.0023	少少	小	ドリブル
中1位	0.0282	多	大	ロングパス
中2位	0.0269	多	大	ドリブル
中3位	0.0256	多	大大	ショートパス
低1位	0.0304	多多	大	ショートパス
低2位	0.0282	多多	大	ドリブル
低3位	0.0221	多多	大大	ロングパス
高平均	0.0011	—	—	—
中平均	0.0078	—	—	—
低平均	0.0068	—	—	—

3. おわりに

本結果では、ボールを奪取するエリアが低いほどシュートできる確率が高く、その中でも多くの人数を経由することでシュートをする確率が高いことを明らかにした。しかし、本研究はラストプレーにのみ着目したので、その間の決定的な攻撃の構造の発見には至らなかった。今後は、その対策として、(1) 1プレーごとにショートパス、ロングパス、ドリブル、フリーキックの回数を記録する、(2) ボールの動きに着目した確率変数を設定する、(3) 日本代表だけではデータ数が少ないのでJリーグなどリーグ単位でデータを入力する、(4) 1プレーを評価できる値を設定する、といったことを検討する。

また、データ入力にあたり、TV中継の映像では視点が移動するためエリアやラストプレーの判断が困難であった場面がしばしばあったため、視点を固定させた映像データも求められる。

参考文献

- [1] 上原司, 荒井秀一, MLB 詳細スコアデータから学習した試合構成群間における確率的因果構造に基づく野球選手の投球戦術推定, 電子情報通信学会, p.p.1-3, p.p.5, 2012
- [2] 豊田秀樹, データマイニング入門, 東京図書, p.p.209-240, 2014