

トピックモデルを用いたサッカーの攻撃パターンの分類による 類似プレーの抽出

神谷 啓太（東京大学大学院 工学系研究科 社会基盤学専攻）

中西 航（東京工業大学 環境・社会理工学院 土木・環境工学系）

泉 裕一郎（東京大学大学院 工学系研究科 社会基盤学専攻）

〒113-8656 東京都文京区本郷 7-3-1 TEL:03-5841-6118

E-mail:kamiya@trip.t.u-tokyo.ac.jp

1. はじめに

数多くのスポーツで、センサやデータ計測員を駆使して多量のデータを取得し、統計的手法によってそれらを分析し、得られた示唆や知見をチーム戦略や選手評価に活用することが一般的となっている。サッカーにおいても、ボール支配率やパス・シュート本数などの従来のスタッツを始め、ここ数年では選手の走った経路を逐次記録したトラッキングデータなどが計測されている [1]。

豊富なデータの中から、似たような「一連の攻撃」を探したいというタスクは、スタッフ、選手、観客などがプレーを振り返る際に重要である。現在はデータ計測スタッフやマネージャー部員が逐一、「シュート」や「FK」というアクションタグや、「カウンター」や「クロス」などのプレー内容のフラグを付加している。類似プレーを検索する際には、これらのタグ情報をもとに目的のプレーを抽出できることもあるが、それだけでは限界がある。アクション以外にも「ピッチのどこで」「どんな選手配置の中」など一連の攻撃を説明するような検索条件はたくさん考えられるからである。ただしそれらをひとつひとつ条件立てて検索するのは困難であり、類似プレー抽出の自動化が望まれる。

一連の攻撃を説明するような変数として、「起こったアクション」「ボール位置」「選手配置」など多様な変数が考えられる。これらを変数としたクラスタリングを実施すれば類似プレーの抽出は可能であるが、同じ位置で、同じ順番に、同じ所要時間で展開するプレーはほぼ一つだけであり、複雑な時系列データすべてを見ても類似度は定義しにくい。すなわち、スパース性の問題が生じる可能性がある。また、時空間をある程度離散化してスパース性を解消しても、ひとつの攻撃の長さがそれぞれ異なるという問題が残る、異なる長さのベクトル間の類似度を測定することは難しい。

本研究では、トピックモデルを使用することで、長さの異なる文書でも、一つの共通したベクトルであるトピック分布として再表現できることに着目する。すなわち、そのトピック分布を類似度に使うことで上記の問題を解決することができると思われる。本研究の目的は、トピック

クモデルを用いてサッカーの攻撃パターンの分類を行い、類似プレーの自動抽出を行うことである。第2章でトピックモデルの中でも LDA (Latent Dirichlet Allocation) について概説した後、第3章で分析に用いる変数を選定する。第4章で実際のプレーデータに対して LDA を適用し、推定されたトピックに関する解釈を行う。最後に、本研究のまとめと今後の課題を第5章でまとめる。

2. トピックモデルを用いた類似プレーの抽出手法

本章ではトピックモデルを用いた類似プレーの抽出手法について説明する。まず、提案手法を構築するにあたり必要となるトピックモデル、LDA (Latent Dirichlet Allocation) について概説した後、サッカーにおける一連の攻撃をいかにトピックモデル的な観点から解釈可能かを議論する。その後、トピックモデルを用いたサッカーのプレー内容の分類およびそれに基づく類似プレーの抽出手法について説明する。なお、LDA についての解説は参考文献 [2][3][4] をもとに記す。

2.1. Latent Dirichlet Allocation

LDA は、もともと文書の確率的生成モデルとして提案されたモデルである。ただし、文章の順序は無視し、Bag of Words(BoW) 表現と呼ばれる単語と出現頻度のペアの集合をモデル化する。ここで、BoW 表現は単語が共起している現象を表しており、文書集合として考えると統計的に共起しやすい単語の集合がいくつか存在するはずである。LDA は、この単語の共起性を統計モデルとして数理的に扱うために提案された。

2.1.1 LDA の生成過程

LDA では、文書中の単語がどの潜在トピックによって生成されたかを示す潜在変数を導入する。具体的には、文書 d の i 番目の単語を $w_{d,i}$ として、対応する潜在変数を

$z_{d,i}$ と定義する。潜在トピックの添字集合を $\{1, 2, \dots, K\}$ とする ($z_{d,i} \in \{1, 2, \dots, K\}$)。各潜在変数の値は、それぞれ単語の出現分布 $\phi_k (k = 1, 2, \dots, K)$ に対応する。つまり、文書中の各単語は離散値をとる潜在変数を背後に保持しており、その潜在変数の値が同じ単語はトピックに属する（同じ単語の出現分布に従う）というモデル化を行う。

文書数を M 、文書 d の文章長（総単語数）を n_d とする。LDA では、文章は複数のトピックから構成され、その構成比を離散分布としてもつ。 $\theta_{d,k}$ を、文章 d でトピック k が出現する確率（文章 d でのトピック k の構成比率）とし、トピック分布を $\theta_d = (\theta_{d,1}, \dots, \theta_{d,K})$ とする。 $\phi_{d,v}$ をトピック k における単語 v の出現確率とし、単語の出現分布を $\phi_k = (\phi_{k,1}, \dots, \phi_{k,V})$ とする。 θ_d や $bvecphi_k$ は確率ベクトルであるので、確率ベクトル上の確率分布である Dirichlet 分布 (Dir と表記) による生成を仮定する。すなわち、

$$\theta_d \sim Dir(\alpha), d = 1, \dots, M \quad (1)$$

$$\phi_k \sim Dir(\beta), k = 1, \dots, K \quad (2)$$

ここで、 $\alpha = (\alpha_1, \dots, \alpha_K)$ は K 次元ベクトル、 $\beta = (\beta_1, \dots, \beta_V)$ は V 次元ベクトルで、いずれも Dirichlet 分布のパラメータである。

単語 $w_{d,i}$ や潜在トピック $z_{d,i}$ は離散値なので、多項分布 (Multi と表記) を生成分布として仮定する。すなわち、各文書 $d (= 1, \dots, M)$ において、各単語は以下の生成過程を仮定する。

$$z_{d,i} \sim Multi(\theta_d) \quad (3)$$

$$w_{d,i} \sim Multi(\phi_{z_{d,i}}), i = 1, \dots, n_d \quad (4)$$

2.2. トピックモデルとしてみるサッカーのプレー

自然言語処理では、以下の仮定に基づいてトピックモデルを適用する。

- 文はいくつかのトピック (話題) の重み付き和で構成される
- トピックごとに使われやすい単語の分布がある
- 文は、トピック→単語の順で構成する単語数をサンプリングしたものである
- 文の中での単語の順番は不問とする、すなわち、順不同の単語から文の意味が一意に定まるとする

一方、本研究ではサッカーのプレーに対してトピックモデルを適用するにあたり、上記に対応するような以下の仮定を置く。

- 一連の攻撃は、ボール奪取時の状況やチームの特性に応じたトピック (攻撃パターン) の重み付き和で表現される

- 攻撃パターンごとに選択されやすいアクションや起こりやすい選手配置がある
- 一連の攻撃は、攻撃パターン→アクション/配置の順番に攻撃を構成する要素数をサンプリングしたものである
- 一連の攻撃内でのアクション/配置の順番は不問とする、すなわち、順不同のアクション/配置からプレー状況が定まるとする

最後の仮定について補足する。直感的には、サッカーの一連のプレーにおいてプレーが生起する順番は意味を持ちそうである。しかしながら、本研究では、仮に単語を適切に設定できれば、ある単語の集合を考えたときに起こりうるプレーの順番は実質的に定まると想定する。このことにより、時系列データが格段に扱いやすくなるという大きなメリットがあるためである。ただし、データから分析に有用な単語を生成する方法自体の検討が必要となるため、これを次章で説明する。なお、この仮定には議論の余地が残るが、後述の適用において一定程度の妥当性は示されたと考えている。

2.3. 類似手法の抽出手法

トピックモデルを用いて攻撃パターンを分類し、その分類結果に基づく類似プレーの抽出手法を以下の通り提案する。

1. ボールタッチ／トラッキングデータより文書を作成する
2. 検索対象の文書 d_{req} のトピック分布 $\theta_{d_{req}}$ を推定する
3. データベース中の攻撃データ d_{db} とコサイン類似度 $\cos(\theta_{d_{req}}, \theta_{d_{db}})$ を計算
4. その他検索条件 (攻撃長、特定プレーの存在など) に基づき、類似度の高いものから抽出

3. 入力変数の検討

3.1. 使用したデータの説明

本研究で使用したデータは、2016 明治安田生命 J1 リーグ 1st ステージ第 1 節の計 9 試合に関して、1/30 秒毎にパスやタックルなどボール周辺で発生したイベントおよびその発生時刻と位置を取得したボールタッチデータと、1/25 秒毎に選手及び審判のピッチ上での位置を取得したトラッキングデータの 2 種類である。なお、これらのデータはデータスタジアム株式会社から提供を受けたものである。

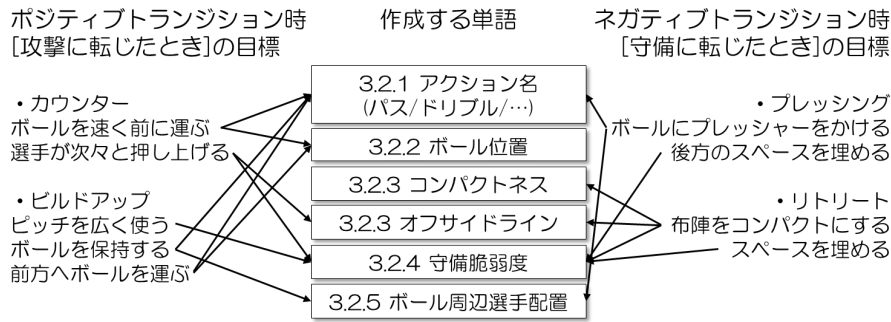


図 1: サッカーの攻防の要素とそれを表す単語の作成

3.2. 単語の作成

LDA の適用のために、ボールタッチデータから本研究で用いる単語を作成する必要がある。ここでの目標は、サッカーの一連の攻撃で発生している複雑な動きを、なるべく端的な単語の羅列で表現することである。そして、作成された単語を並べた文をみれば、人やボールの時系列の移動やその順序が表現される状態を目指す。サッカーの一連の攻撃において実際に生起している事象をもとに検討し (図 1)、以下の 5 種類を作成することとした。

3.2.1 アクション名

プレー内容を表現する単語として、ボールタッチデータに整理されている「アクション名」を選定した。もともとはシュート、パス、GK、フィードなどの単語から構成されている。ただし、「ホームパス」、「アウェイパス」および「トラップ」の出現回数が他の単語と比較してかなり多く、特徴的な単語となりにくい可能性がある。そこで、パスまたはトラップの向きとそのアクションが成功か否かを組み合わせたものをひとつの単語とする。具体的には、提供されたデータ中の「展開力」と「成功 F」を利用した。例えば、「パス」のアクションで、「展開力」が 1 方向、かつ成功の場合、生成される単語は「Pass_1_1」となる。なお、一連の攻撃中に行われる守備側のアクションはすべて「defenseAction」という単語として統一した。

3.2.2 ボール位置

ピッチ上を 54 分割した上で、ボール位置を離散的に表現した単語を作成した。これは、提供データ中の「hotzone6-9」と同義である。図 2 にその定義を示す。

3.2.3 コンパクトネスとオフサイドライン

ピッチ上の選手配置として考えられる指標として、「コンパクトネス」および「オフサイドライン」を統合した単語を作成する。コンパクトネスについて本研究では、「一番前方に存在する選手の X 座標」と「後方より 2 番目に

存在する選手の X 座標」の距離をコンパクトネスと定義する。各チームの選手がその時刻において、どれだけピッチ上で展開できているかを示す指標である。また、オフサイドラインについては「後ろから 2 人目の選手位置」と定義し、ルール上のオフサイドラインとは異なる点に注意されたい。この 2 つの指標は相関をもち、かつ選手がどれくらいの位置・幅でプレーをしているかを表すという点では重複した意味になる可能性がある。そこで、「コンパクトネス×オフサイドライン」の組み合わせとして単語化することとする。本研究では、10m ずつ離散化して単語化かし、守備側の値のみを利用する。

3.2.4 守備脆弱度

ピッチ上の選手配置のもうひとつの指標として、チームの守備力の脆弱性を表す単語を採用する。ここでは自陣の最終ライン付近において相手選手が侵入している程度を守備脆弱度と定義する。具体的には図 2 に示すように、「自軍のオフサイドラインより前方 10m、後方 5m の長方形のうち、最寄りの味方選手から 5m 以上離れており、最近傍選手が相手選手であるような地点の合計面積の割合」である。こちらも守備側の配置のみを文章中の単語として利用する。

3.2.5 ボール周辺選手配置

ボール周辺に味方選手もしくは敵選手がいるかを表す単語として採用する。ボールを中心に半径 5m の円を想定し、4 分割した各扇形セルそれぞれに、ボールに関与しているプレイヤーを除き、誰もいない場合は「0」、味方選手のみの場合は「1」、敵選手のみの場合は「2」、両方の選手がいる場合は「3」の文字を付加していく。前・右・左・後の順に一つの単語を生成する。図 2 に具体例を示す。なお、(2016, 成塚ら) の分析により、「対戦相手との距離が 5m 以上近づくと、そこから次第に向きをそろえる傾向が強くなる」とされている [5]。そこで、本研究でも近傍選手として考慮すべき距離を 5m と設定した。

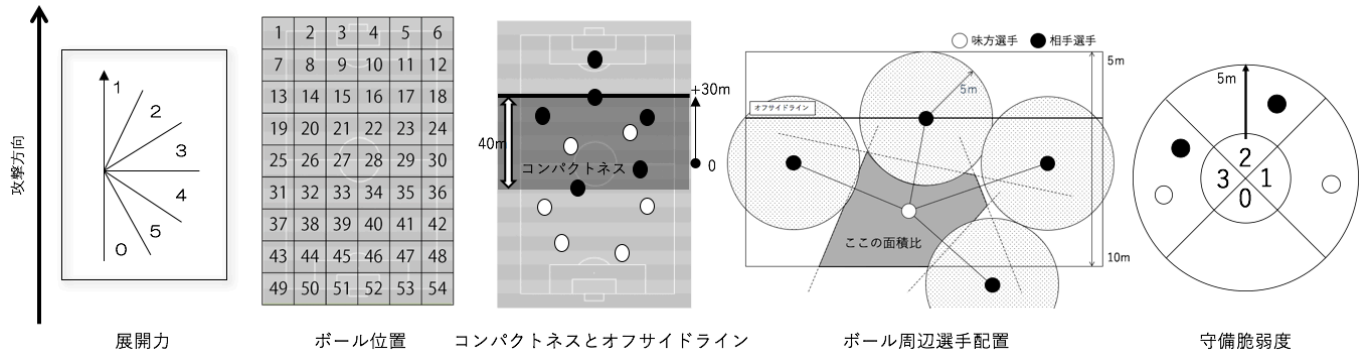


図 2: 作成した単語の例

4. 類似プレーの抽出実験

4.1. 適用条件

第 3 章で作成した全文書データに対して、LDA を適用し、サッカーの攻撃パターンの分類を行う。LDA パラメータとして、トピック数 $K = 5$ 、ハイパーパラメータ α, β の事前分布には無情報一様分布を設定した。高頻度・低頻度単語の削除は行わず、生成したすべての単語を分析対象とした。なお、パラメータ更新が収束した場合、もしくは 50 回繰り返し計算を実施した場合に計算終了とする。

4.2. 推定されたトピックの解釈

以上の設定のもとで推定された 5 つのトピックにおいて特徴的な単語を示し、各トピックにおける攻撃パターンの意味合いを示す (表 1)。作成に当たっては以下の二つを参考にした。一つ目はトピックごとの単語分布 $p(w|z_i), i = \{1, 2, 3, 4, 5\}$ と、デフォルト単語分布 $p(w)$ (すなわち、データ全体における正規化単語ヒストグラム: トピックが 1 つの場合の単語分布) との比を単語ごとに算出し、降順に並べたものである。二つ目は、単語条件付きトピック分布 $p(z_i|w), i = \{1, 2, 3, 4, 5\}$ である。

作成した単語を用いてサッカーの攻撃プレーを表現した結果、おおむねサッカーらしいトピック 5 つに自動分類できたと考えられる。たとえば、トピック 1 はプレー位置が敵陣深くで、守備側のラインも深い。さらに、ボール周囲に相手選手が多く、シュートやクロスも特徴的な単語となった。また、トピック 5 ではボール位置が自陣深く、相手の守備ラインが高く脆弱度も高い一方で、スルーパスやシュート、被ファウルのような単語も特徴となった。これらの結果は、攻撃をシュートで終えるためには、敵陣深くからのスルーパスやクロスまたはカウンターが有効であることを示唆している。これは、一般に想定されるサッカーの知識と類似している。また、サッカーの多くの時間はトピック 4 に表れる自陣でのビルドアップやその阻止、またトピック 3 に表れるアタッキングガード進入の攻防であることや、それ以外の時間が展

開のない自陣でのボールポゼッション (トピック 2) であることもうかがえる。

4.3. 実際の例

次に、トピックの分類結果を実際のプレー内容と照らし合わせた上で検証する。文書条件付きトピック分布 $p(z_i|d), i = \{1, 2, 3, 4, 5\}$ を参照しながら、ほとんど一つのトピックのみで構成されている文書と、複数のトピックで構成されている文書の 2 パターンに関して、実際のプレー内容とトピック内容を比較する。

例えば、トピック 1 が 90% 程度の割合で占めている例では、浦和がボールを前進させシュートまで繋げている様子が確認できた。また、トピック 5 が 90% 程度の割合で占めている例では、鹿島がカウンターで 40m 進み被ファウルを受けた様子が確認できた。さらに、構成トピックが混合しているものの例として (トピック 1:3:5 = 3:4:2)、川崎の一連の攻撃を挙げられる。この例では、川崎が自陣中ほどで相手のパスをカットし、比較的短時間のうちにバイタルエリア付近、さらには敵陣深くまでパスを繋いでいる様子が伺えた。

これらの例から、単語条件付きトピック分布 $p(z_i|w), i = \{1, 2, 3, 4, 5\}$ から推定したトピックごとの攻撃パターンは、文書条件付きトピック分布 $p(z_i|d), (i = \{1, 2, 3, 4, 5\})$ をもとに抽出した実際の攻撃内容から推察される攻撃パターンとある程度一致していることが分かった。特に、複数のトピックの混合で文書条件付きトピック分布が構成されている場合においても、構成されているトピック (攻撃パターン) のそれぞれの特徴がプレー内容に反映されていることが伺えた。

4.4. 類似プレーの抽出結果

最後に、文書条件付きトピック分布 $p(z_i|d)$ の類似度に基づく類似プレーの抽出結果を示す。ただし、全プレーに対して類似プレーの自動抽出を行い、その結果が「似ている」か「似ていない」の判断を下すのは不可能であるため、ここではいくつかの例で検証する。

表 1: 推定されたトピックの解釈

Topic No.	トピックの意味合い	特徴的な単語
1	敵陣深くまで攻めきる	シュート / クロス / スルーパス / ドリブル / CK ボール位置敵陣深く (hotzone1-12 付近) 相手選手多い (neighborplayer に 2) / 守備ライン低い (50m 付近)
2	その他の自陣での展開と短時間で終了した攻撃	ボール位置自陣 (hotzone31-42 付近) / その他特徴的な単語は少ない
3	アタッキングサード 進入の攻防	オフサイド / ボール位置バイタル (hotzone7-24) 守備ライン中盤 (30m 付近)
4	相手守備ブロック形成時のビルドアップ	クリア / GK / フィード / 前方へのパス (失敗含む) ボール自陣 (hotzone31-42 付近) / 脆弱度低い (2% 以下)
5	ポジティブトランジションとカウンター	ブロック / タックル / インターセプト / シュート / スルーパス ドリブル / オフサイド / フィード / 被ファウル ボール自陣深く (hotzone43-54) / 相手選手多い (neighborplayer に 2) 守備ライン高い (0m 付近) / 脆弱度高い (10% 近く)

具体的には、前節で紹介した表 2 における No.7 9 の攻撃について類似した攻撃の検索を試みる。検索対象の攻撃 d_{req} との文書条件付きトピック分布 $\theta_{d_{req}}$ と、他のすべての攻撃 d_{db} の文書条件付きトピック分布 $\theta_{d_{db}}$ のコサイン類似度 $\cos(\theta_{d_{req}}, \theta_{d_{db}})$ を計算し、高いスコアを算出した攻撃を抽出する。ただし、一連の攻撃が少数のアクションのみによって構成されている場合、効果的な攻撃でなく検索結果として好ましくない可能性や、攻撃内容が複雑ではないため、構成されるアクション (CK、シュート) だけで検索できてしまうと思われる。そこで、検索条件として、「15 回以上のアクションによって構成されている攻撃」と設定することとした。

類似プレーの抽出結果を表 3 に示す。例えば攻撃番号 No.9 の検索結果に対してしてみると、フィードやスローインによってバイタルエリア付近までボールを運び、その後パスを繋いでいるという攻撃の様子に共通点が見られた。ただし、検索結果の中には似ているか似ていないかの判断が難しかったり、そもそも攻撃パターンの解釈自体は難しかったりするものも多く存在した。抽出された攻撃パターンがユーザにとって好ましいものであったのかなど、別途議論すべきであると考えられる。

5. おわりに

本研究では、トピックモデルを用いてサッカーの攻撃パターンを分類し、類似した一連の攻撃を自動で抽出する方法の提案を行った。トピックモデルである LDA を適用するにあたり、サッカーのトラッキングデータおよびボールタッチデータから一連の攻撃を説明するような単語の検討を行い、アクション名、ボール位置、選手配置 (コンパクトネス × オフサイドライン)、守備脆弱度、ボール周辺選手配置の 5 種類の単語を作成し、攻撃内容を離散的な文書として表現した。作成した文書に対して

LDA を適用し、トピックごとの単語分布とデフォルト単語分布の比較、および単語条件付きトピック分布の参照を通じて各トピックの意味合いを解釈した結果、おおむねサッカーらしいトピック 5 つに自動分類できたと考えられる。また、実際のプレー内容とトピック分布を比較・検証してみたところ、プレー内容から推察される攻撃パターンとトピックの解釈に基づく攻撃パターンにある程度一致性があることが確認された。さらに、文書条件付きトピック分布の類似度に基づいて類似プレーを抽出した結果、共通した攻撃パターンを持つ攻撃が複数抽出されたことが確認できた。

今後の課題として、はじめに、本研究で設けた仮定の妥当性について、さらなる検証を行いたい。単語の羅列が実際に攻撃プレーを一意に定めうるのか、トピック混合状態が類似するプレーが実用上知りたい類似プレーになっているかなど、多くの適用により知見を深めたい。そのうえで、モデルの数理的拡張として、以下の観点を今後の展望としたい。まず、トピック数自体の推定も含めたモデルの検討が挙げられる。現在、トピック数 K を 5 つに固定して分析を回しているが、チーム戦術や対戦カードによりトピック数自体も異なると想定される。そのため、AIC や BIC を指標としたモデル選択を行う、もしくはディリクレ過程を利用したトピック数の自動推定への応用が望まれる。また、「試合終盤ではゲームが動きやすい」などのサッカーの特徴を反映するため、Dynamic topic model[6] や Topics over time[7] など、トピックの時間変化を含めたモデルに拡張することも課題としてあげられる。さらに、オフENS時の選手配置がファーストディフェンスのやり方を規定すると考えられるため、攻撃に入る直前の守備配置を文書中の単語として採用するなど、文書作成における工夫も必要である。最終的には、変数のスケールや離散化方法など調整を加えたうえで、今後蓄えられてくるであろうデータベースの有効利用のための推薦システムの構築を目指したい。

表 2: 文書条件付きトピック分布と実際のプレーとの関係

No.	文書条件つきトピック分布 (各文書におけるトピック構成比)	実際のプレー内容の概要
1	0.9 : 0.1 : 0.0 : 0.0 : 0.0	浦和がボールを前進させシュートまで繋げる
2	1.0 : 0.0 : 0.0 : 0.0 : 0.0	神戸が自陣深くでクリアしたボールをパスで前線まで運び、サイドをドリブルで切り込む
3	0.0 : 1.0 : 0.0 : 0.0 : 0.0	川崎がハーフライン奥でボールを取得 MF を中心にペナルティーエリア手前でパスを回す展開
4	0.0 : 0.0 : 1.0 : 0.0 : 0.0	鳥栖が敵陣中ほどでボールを取得し、 MF を中心にペナルティーエリア手前でサイドチェンジを展開
5	0.0 : 0.0 : 0.0 : 1.0 : 0.0	湘南が自陣の浅い位置でスローインを取得し、 DF までパスを下てからパスを繋いでいき、 敵陣へと進入したところで被ファウル
6	0.1 : 0.0 : 0.0 : 0.0 : 0.9	鹿島がカウンターで 40m 進み、ファウルを受ける
7	0.4 : 0.0 : 0.6 : 0.0 : 0.0	広島がキーパーからのフィードをパスを繋ぎ、 敵陣深くでスルーパスで終える
8	0.6 : 0.0 : 0.0 : 0.0 : 0.4	川崎が自陣深くで相手のパスをブロック、ポストプレイ キーパーを経由して短い手数で前線に運び、最後はクロスで終わる
9	0.3 : 0.0 : 0.4 : 0.0 : 0.2	川崎が自陣中ほどで相手のパスをカットし、 比較的短時間のうちにバイタルエリア付近、 さらには敵陣深くまでパスを繋ぐ

表 3: 文書条件付きトピック分布と実際のプレーとの関係

No.	コサイン類似度	実際のプレー内容の概要
検索対象 : No.7 in Table2		
7-1	0.99	川崎がキーパーからのフィードをパスを繋ぎ、敵陣深くでスルーパス
7-2	0.99	川崎がキーパー発の FK からパスを繋ぎ、敵陣深くでスルーパス
検索対象 : No.8 in Table2		
8-1	0.99	湘南が自陣の浅い位置でインターセプトし、敵陣深くまで攻める 守備側のクリアなどのアクションあり
検索対象 : No.9 in Table2		
9-1	0.99	名古屋がキーパーからのフィードを起点にボールを前線へと繋げる
9-2	0.95	鹿島がピッチ中ほどからのスローインを起点にパスを繋ぎ、最後にはクロスを上げる

謝辞

本研究で利用したデータはデータスタジアム株式会社に提供いただきました。また、本研究の貸与データは情報・システム研究機構の新領域融合研究プロジェクト『社会コミュニケーション』データ中心科学リサーチコモンズ事業『人間・社会データ』の支援を受けたものです。ここに、感謝の意を表します。

参考文献

- [1] Football LAB (フットボールラボ) とは、Football LAB ～サッカーをデータで楽しむ～、<http://www.football-lab.jp/pages/about/>、2017 年 1 月 31 日閲覧
- [2] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3.Jan (2003): 993-1022.
- [3] 岩田具治：トピックモデル、講談社、2015.
- [4] 奥村学：トピックモデルによる統計的潜在未解釈、コロナ者、2015.
- [5] 成塚拓真、卯田純平、山崎義弘：サッカーの対戦的特徴に現れる普遍的な統計性の探求、統計数理研究所共同リポート 363、スポーツデータ解析における理論と事例に関する研究集会、第 3 巻、pp. 83-90、2016.
- [6] Blei, David M., and John D. Lafferty. "Dynamic topic models." Proceedings of the 23rd international conference on Machine learning. ACM, 2006.
- [7] Wang, Xuerui, and Andrew McCallum. "Topics over time: a non-Markov continuous-time model of topical trends." Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006.