

トピックモデルを用いた 攻撃パターンの分類による 類似プレーの自動抽出

第 6 回スポーツデータ解析コンペティション
サッカー部門

神谷啓太*, 中西航**, 泉裕一郎*

* 東京大学 大学院工学系研究科 社会基盤学専攻

** 東京工業大学 環境・社会理工学院 土木・環境工学系

分析の背景／目的および方針

- 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
 - スタッフ、選手、観客などがプレーを振り返る際に、目的のプレーと似たようなプレー(「類似プレー」)を参考にできたら便利
- 「シュート」や「FK」というアクションタグだけの検索は限界がある
 - 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
 - それらをひとつひとつ条件立てて検索するのは困難



目的および方針

- 一連の攻撃を説明するような変数：
 - 「起こったアクション」「ボール位置」「選手配置」 など…
- 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だが…
 - 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
 - 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く、ほとんどのものは違って見える(スパース)
 - 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
- そこで、トピックモデルを適用し、そのトピック分布を類似度に使う

目的

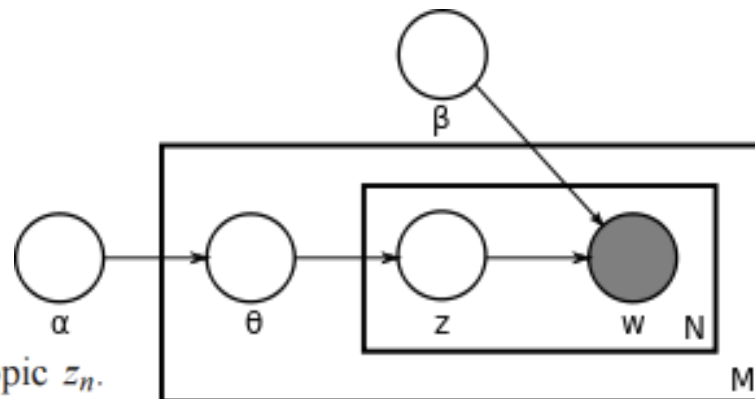
- ボールタッチデータ/トラッキングデータから
類似した一連の攻撃を自動で抽出する方法の提案
- トピックモデルを用いた攻撃プレーのクラスタリングを実施、
および得られたトピックの解釈
- 類似プレーを推薦するシステムの提案

手法



- トピックモデル(LDA; Latent Dirichlet Allocation)
 - 文書が複数の潜在的なトピックから確率的に生成されると仮定したモデル

- Choose $N \sim \text{Poisson}(\xi)$.
- Choose $\theta \sim \text{Dir}(\alpha)$.
- For each of the N words w_n :
 - Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
 - Choose a word w_n from $p(w_n | z_n, \beta)$, a multinomial probability conditioned on the topic z_n .



変数	自然言語における考え方	サッカーにおける考え方
d	文書	一連の攻撃
w	単語	「起こったアクション」 「ボール位置」「選手配置」 etc
z	トピック	「攻撃パターン」 攻撃がどのような傾向かの指標

- 単語分布 $p(w|z)$ よりトピックの意味解釈が可能
- 文書 d ごとに付加されるトピック分布 $p(z|d)$ よりrecommendationが可能

自然言語処理における論理

文はいくつかのトピック(話題)の重み付き和で構成されている

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→単語の順番に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
単語があれば意味のある文の復元は可能

サッカーにおける論理

ボール奪取時の状況やチームの特性に応じて、「一連の攻撃」はそれらから判断される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

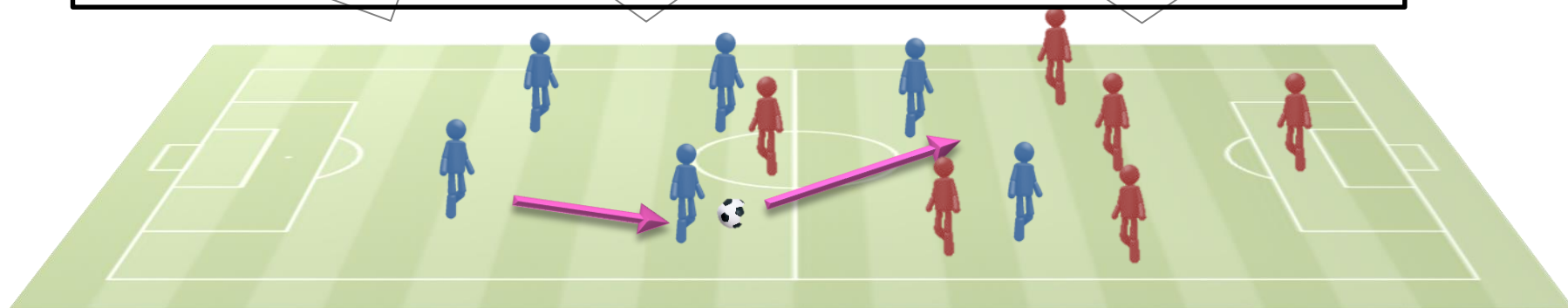
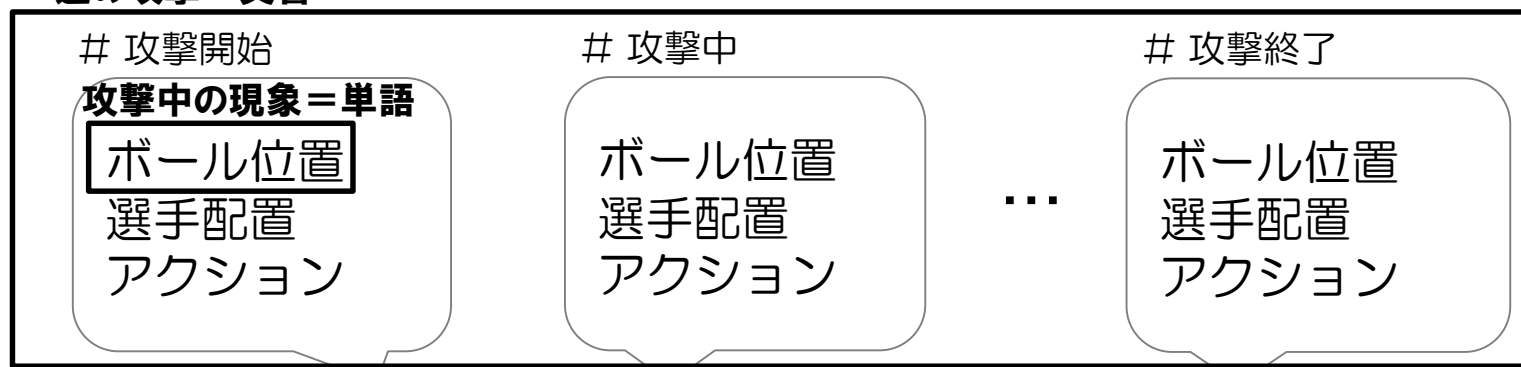
一連のプレーは攻撃パターン→アクション/配置の順番にサンプリングしたものである

- 一連のプレー内部の順番を問わないことの意味
 - 時系列データが格段に扱いやすくなる
 - 単語を適切に設定すれば、実質的にプレー展開は定まる
- ただしそのための単語生成方法の検討が必要(自然言語処理との違い)

文と単語の作り方のイメージ

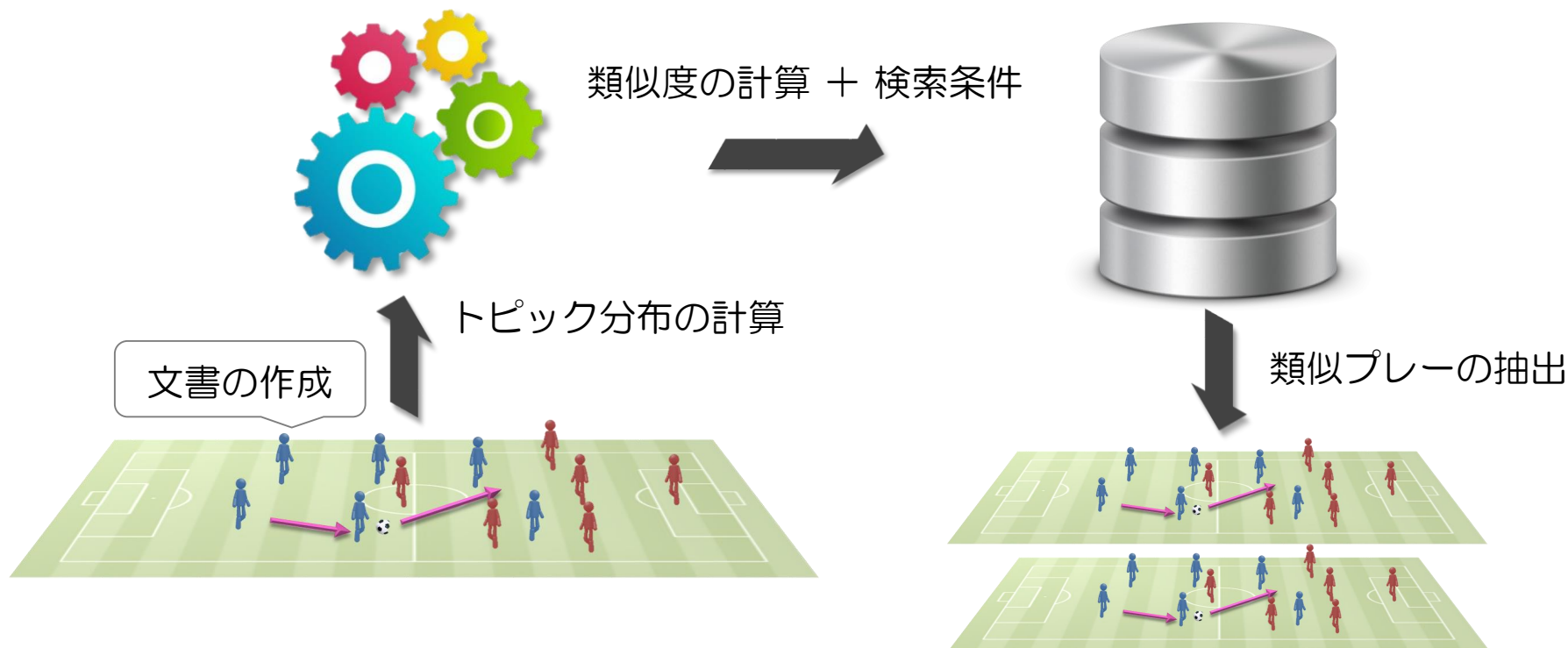
- サッカーの攻撃におけるトピックモデルの考え方
 - ボールタッチデータにおける攻撃番号ごとにユニークな「文書」とみなす
 - 攻撃中の現象を説明するような値を「単語」とみなす
 - 攻撃中のボール位置や選手配置、起こったアクション など
- 同じ現象を同じ単語に、違う現象を違う単語に割り当てる
 - 変数の取捨選択と、離散化の粗さ・細かさを検討する

一連の攻撃＝文書



LDAを用いた類似プレーの抽出

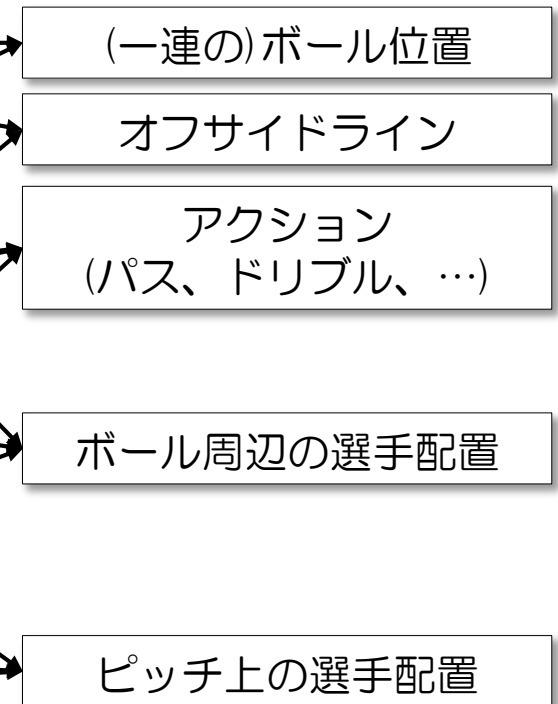
- 文書ごとに付加されるトピック分布より類似プレーを抽出する
ある「一連の攻撃」 d_{req} に関して、
 - ボールタッチ/トラッキングデータより文書を作成する
 - トピック分布 $z_{req} = p(z|d_{req})$ を推定
 - データベース中の攻撃データとコサイン類似度 $\cos(z_{req}, z_{db})$ を計算
 - その他検索条件に基づき、類似度の高いものを抽出



分析に使う文と単語の生成

単語の作成方針

- 文に含まれる単語作成のため、サッカーの「一連の攻撃」で実際に起きていることをもとに検討する
- 例) ボールを奪取して攻撃に転じたとき
 - カウンター
 - ボールを速く前に運ぶ
 - 後ろから選手が次々と押し上げる
 - ビルドアップ
 - ピッチを広く使う
 - 前方へボールを運ぶ
 - ボールを保持する
- 例) ボールを奪われて守備に転じたとき
 - プレッシング・リトリート
 - ボールにプレッシャーをかける
 - コンパクトにする
 - スペースを埋める

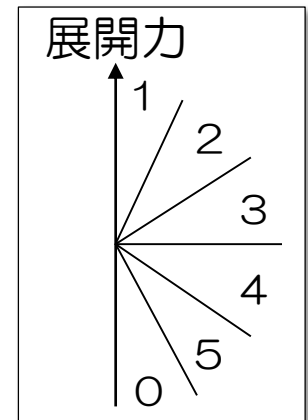


複雑な動きをなるべく端的に単語の羅列で表すことを目指す
 これにより「人やボールの移動」は文として結果的に表現できる

作成した単語 ～アクション名～

12

- 配布データの「アクション名」を用いる
 - “シュート”, “クロス”, “GK” ...
- 「(ホーム/アウェイ)パス」と「トラップ」の出現回数がかかなり多く、特徴的な単語となりにくい
 - パスor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
 - 配布データ中の「展開力」と「成功F」を利用
 - ex) パス×展開力1×成功 : “Pass_1_1”
 - ex) トラップ×展開力3×失敗: “Trap_3_0”
- 一連の攻撃中に行われる守備側のアクションは“defenseAction”に統一



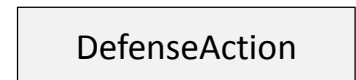
攻撃側のアクション

	A	B	C	D	E	F
1	アクションID	アクション名	アクション名	アクション名	アクション名	アクション名
16	15	シュート	シュート	シュート	Shoot	Shoot
17	16	GK	GK	GK	GK	GK
18	17	直接FK	直接FK	FK	FK	FK
19	18	キャッチ	キャッチ	キャッチ	Catch	Catch
20	19	警告(イエロ)	警告(イエロ)	警告(イエロ)	YellowCard	Yellow
21	20	PK	PK	PK	PK	PK
22	21	CK	CK	CK	CK	CK
23	22	間接FK	間接FK	FK	FK	FK

パスorトラップの場合



守備側のアクション



作成した単語 ～ボール位置～

13

- 配布データの「HOTZONE6-9」を用いる

- ピッチ上を54分割して単語生成

– ex

hotzone_1

hotzone_2

⋮

hotzone_54

- ※「一連の攻撃」中に行われる
守備側アクション時にはhotzoneを
反転し、向きをそろえる

攻撃方向

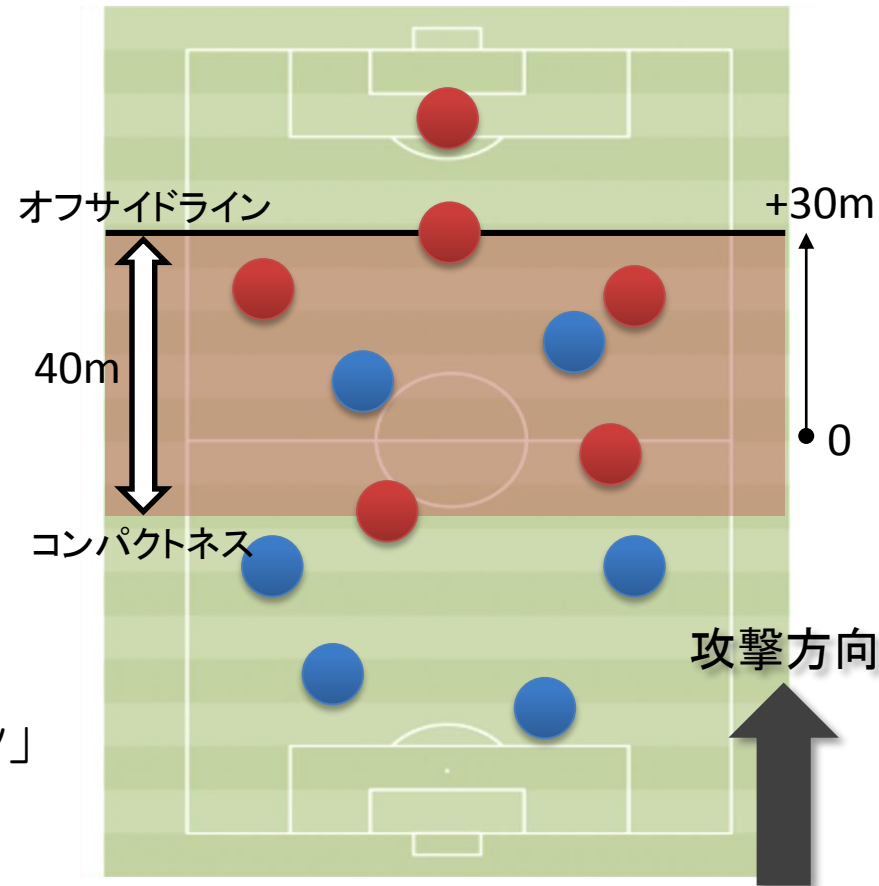


作成した単語 ～選手の配置～

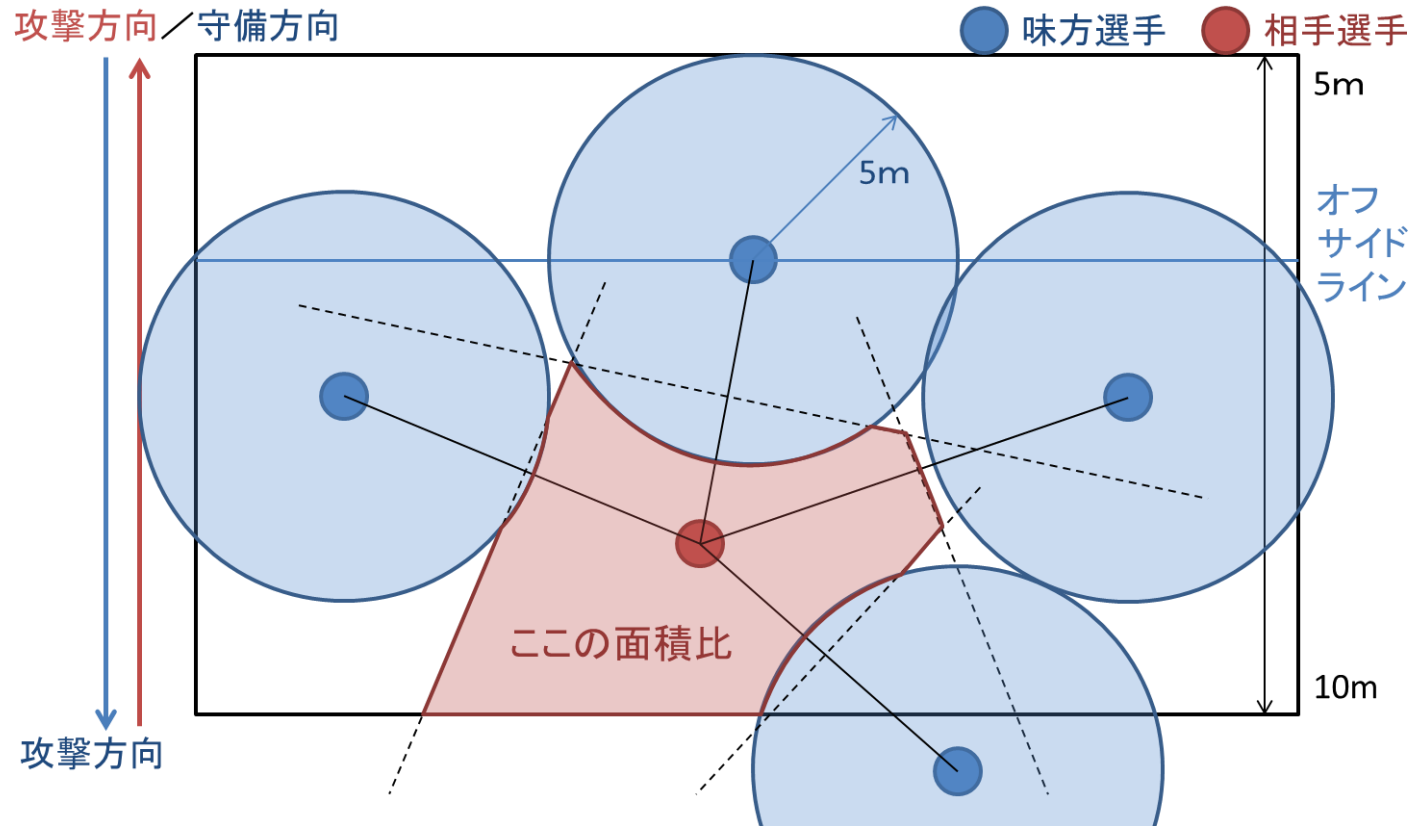
14

- コンパクトネス
 - 「後ろから2人目の選手と最前選手との幅」
- オフサイドライン
 - 「後ろから2人目の選手位置」
 - ルール上のオフサイドラインとは異なる
- ただし、これらの値は相関をもつ
 - 相関係数 -0.2 ほど
 - 単語として重複する意味となる恐れがある
 - 「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
 - 10mずつ離散化して単語化
 - 守備側の値のみを単語化
 - ex)

Compactness_40m__Offsideline_30m



作成した単語 ～守備脆弱度～



- 「オフサイドラインの後ろ5m前10mの長方形のうち、最寄りが相手で、かつ、最寄りの味方が5m以上離れてる面積比」
- この指標を1%刻みで離散化したものを単語として設定
 - ex) Volnerability_0% ... Volnerability_15%
- 守備側のみの値を単語として登録

作成した単語 ～ボール周辺選手配置～

16

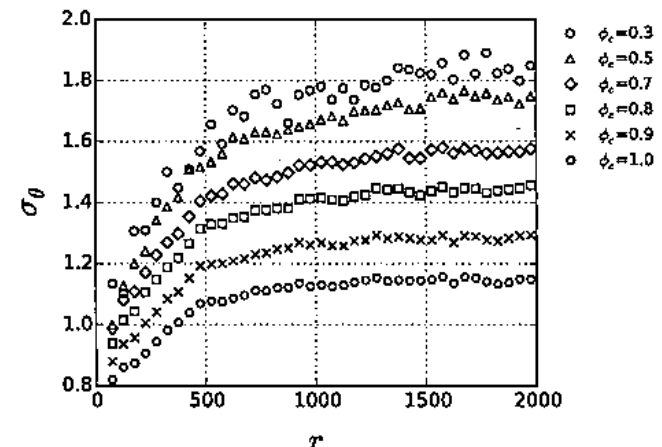
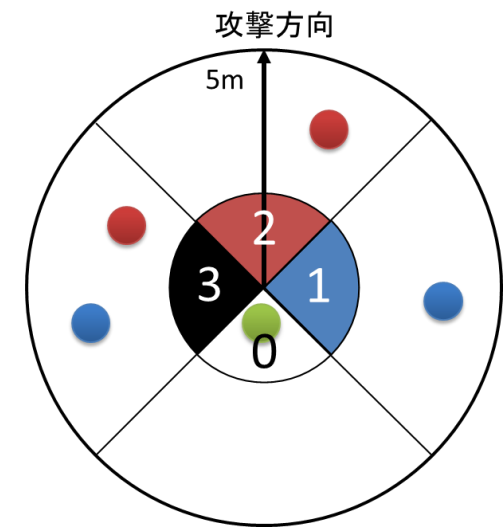
- ボール周辺に味方選手／敵選手がいるかを表す単語
- ボールを中心に半径5mの円を想定、4分割した各扇形セルに

- 誰もいない：0
- 味方選手のみいる：1
- 敵選手のみいる：2
- 両方の選手がいる：3

それぞれ上記の値を記入
(ボールに関与しているプレイヤーを除く)

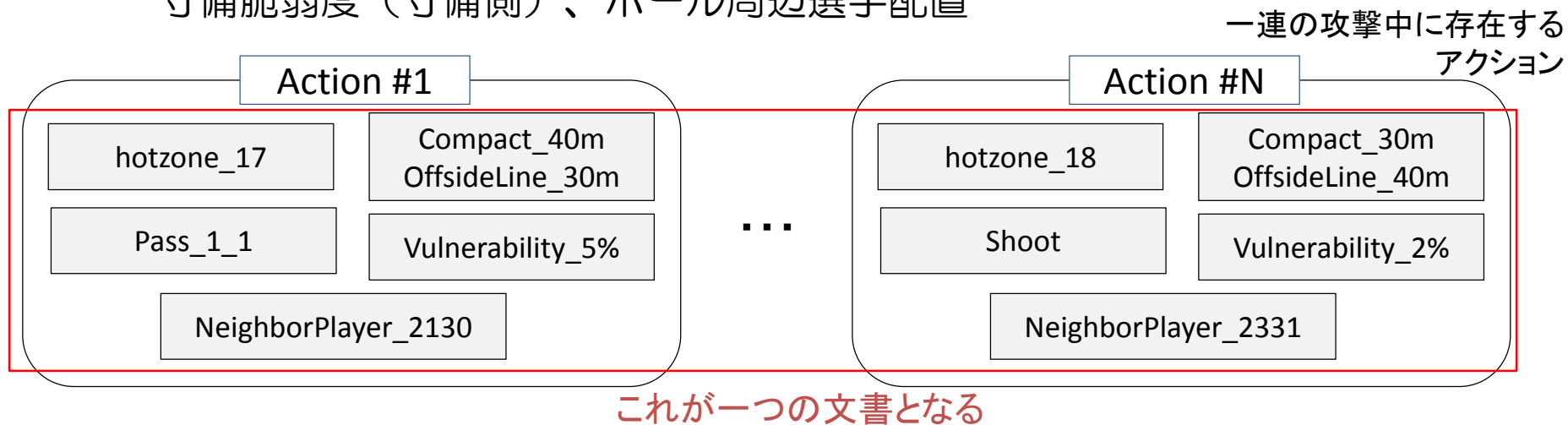
- 前・右・左・後の順に一つの単語を生成
- ex) 右の選手配置の場合：

NeighborPlayer_**2****1****3****0**



(2016, 成塚ら)「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」
→ 近傍選手として考慮すべき距離を5mとした

- 攻撃番号 = 1つの文書
→ 一連の攻撃期間における位置・アクション等の情報 = 各単語
- 構成する単語
 - 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）（守備側）
守備脆弱度（守備側）、ボール周辺選手配置



- ボール位置および選手位置を介した攻撃-守備の関係を捨象して表現
 - 複雑な時系列データを、特徴を残しながらも解析しやすい形へ
 - これらがどのように分類されるか？ → トピックの推定

トピックモデルの適用と結果の解釈

- LDAパラメータ設定
 - トピック数 : 5
 - ハイパーパラメーター事前分布 : 無情報一様分布
 - 計算終了条件 : パラメータ更新の収束 or 50回繰り返し計算実施
- 実装詳細
 - Python用ライブラリ genism
 - 文書イメージ

	A	B	C	F	K	L
	attackID	hotzone	actionWord	VulnerabilityAttackWord	Compactness_Offside	NeighborPlayerWord
1						
2	文書1	hotzone_27	KickOff	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_20,30	NeighborPlayer_1111
3		hotzone_27	Trap_0,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_20,30	NeighborPlayer_1111
4		hotzone_27	Pass_0,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_20,30	NeighborPlayer_1111
5		hotzone_34	Trap_4,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_20,30	NeighborPlayer_1111
6		hotzone_33	Pass_2,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_20,30	NeighborPlayer_1111
7		hotzone_13	Pass_2,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_20,30	NeighborPlayer_3111
8		hotzone_14	Trap_1,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_30,40	NeighborPlayer_3113
9		hotzone_8	Shoot	VulnerabilityAttack_3,4	CompactnessDefense_30,40_OffsideLineDefense_30,40	NeighborPlayer_1331
10		hotzone_3	BallOut	VulnerabilityAttack_3,4	CompactnessDefense_30,40_OffsideLineDefense_30,40	NeighborPlayer_1311
11	文書2	hotzone_51	GK	VulnerabilityAttack_1,2	CompactnessDefense_10,20_OffsideLineDefense_10,20	NeighborPlayer_1111
12		hotzone_26	defenseAction	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_10,20	NeighborPlayer_3222
13		hotzone_26	Pass_1,0	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_10,20	NeighborPlayer_3131
14		hotzone_35	Pass_2,1	VulnerabilityAttack_6,7	CompactnessDefense_20,30_OffsideLineDefense_0,10	NeighborPlayer_3111
15		hotzone_30	Trap_0,1	VulnerabilityAttack_6,7	CompactnessDefense_20,30_OffsideLineDefense_0,10	NeighborPlayer_1111
16		hotzone_36	Pass_4,1	VulnerabilityAttack_5,6	CompactnessDefense_20,30_OffsideLineDefense_10,20	NeighborPlayer_1111
17		hotzone_35	Trap_3,0	VulnerabilityAttack_4,5	CompactnessDefense_20,30_OffsideLineDefense_10,20	NeighborPlayer_3113
18		hotzone_20	Pass_3,1	VulnerabilityAttack_0,1	CompactnessDefense_20,30_OffsideLineDefense_10,20	NeighborPlayer_3111

- 推定されるもの
 - トピック条件付き単語 $p(w|z)$
 - どの単語(アクション/位置/配置…)がどのトピックで起こりやすいか
 - 文書条件付きトピック $p(z|d)$
 - 一連の攻撃はどのトピックのどのような重み付けで生成されているか

推定されたトピック($p(w|z)$)の例

20

- トピック $z=1$ の上位単語の抜粋

rank	word	probability
1	neighborplayer_1111	0.064
2	vulnerabilitydefense_4_5	0.032
3	vulnerabilitydefense_5_6	0.031
4	defenseaction	0.027
	...	
7	neighborplayer_3111	0.024
	...	
11	compactnessdefense_40_50__offsidelinedefense_40_50	0.019
12	neighborplayer_1311	0.019
13	compactnessdefense_10_20__offsidelinedefense_40_50	0.017
	...	
16	trap_1_1	0.016
17	pass_2_1	0.014
	...	
21	hotzone_4	0.013
22	pass_1_1	0.013
23	hotzone_1	0.012
	...	
30	hotzone_2	0.009
	...	

<単語のおおよその意味>

守備脆弱度少々高い

周囲のどこかには相手選手

深いディフェンスライン

前方へのPassとその阻害

前方でのプレー



<トピックの意味合い>

敵陣深くでの攻撃プレー

- $p(w|z)$ は文章全体に出現頻度の高い w ほど大きな確率となりやすい
 - 例：neighborhood 1 1 1 1 はどのトピックでも最上位
- そこで $p(z|w)$ での解釈(トピック頻度)を併用する
 - 「ある単語が出現するとすれば、どのトピックから出てきやすいか？」
- 各トピックの意味合いと出現頻度/トピック頻度の高い特徴的な単語

トピック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)
トピック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)/その他特徴的単語は少ない
トピック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)
トピック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%)
トピック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)

推定されたトピックの解釈

22

- おおむねサッカーらしいトピックに分かれたといえる
 - 攻撃をシュートで終えるためには、敵陣深くからクロス/敵陣深くへのスループス(トピック1)や、カウンター(トピック5)が有効
 - サッカーの多くの時間は、ビルドアップとその阻止(トピック4)、さらにアタッキングサードへの進入とその阻止(トピック3)である
 - それらが手詰まりになると、展開が見込めないボール保持(トピック2)
- 実際の例で確認してもおおむね良い結果である
- トピック1が90%程度：浦和がボールを前進させシュート
 - 遠藤(パス)森脇(トラップ・パス) … (パス)阿部(パス)宇賀神(スループス)武藤(ドリブル・シュート) …
- トピック5が90%程度：鹿島がカウンターで40m進み被ファール
 - 西(パス)遠藤(トラップ・パス)小笠原(トラップ・パス)金崎(トラップ・パス)小笠原(トラップ・被ファール)
- 混合しているもの(トピック1:3:5 = 3:4:2)：川崎の一連の攻撃
 - 大島(パス)奈良(トラップ・パス)中村(トラップ・パス)狩野(トラップ・パス)森谷(トラップ・パス)中村(トラップ・パス)小林(トラップ・パス)狩野(パス)大久保(トラップ・パス) …

(5)カウンター

(3)バイタル

(1)最前線

- 川崎の一連の攻撃
 - 大島(パス)奈良(トラップ・パス)中村(トラップ・パス)狩野
(トラップ・パス)森谷(トラップ・パス)中村(トラップ・パス)小林
(トラップ・パス)狩野(パス)大久保(トラップ・パス)...



類似度0.99

檜崎 正剛	hotzone_52	フィード
安田 理大	hotzone_37	トラップ
安田 理大	hotzone_31	アウェイパス
イ スンヒ	hotzone_32	トラップ
イ スンヒ	hotzone_38	アウェイパス
オーマン	hotzone_40	アウェイパス
竹内 彬	hotzone_39	アウェイパス
シモビッチ	hotzone_21	アウェイパス

類似度0.98

塩田 仁史	hotzone_52	ハンドクリア
ネイツ ペチュニク	hotzone_46	トラップ
ネイツ ペチュニク	hotzone_46	クリア

類似度0.95

西 大伍	hotzone_18	スローイン
鈴木 優磨	hotzone_12	アウェイパス
宇佐美 貴史	hotzone_12	ホームパス
柴崎 岳	hotzone_16	アウェイパス
金崎 夢生	hotzone_16	トラップ
金崎 夢生	hotzone_17	アウェイパス
カイオ	hotzone_11	トラップ
カイオ	hotzone_12	アウェイパス
金崎 夢生	hotzone_11	トラップ
金崎 夢生	hotzone_12	アウェイパス
カイオ	hotzone_12	トラップ
カイオ	hotzone_5	クロス
東口 順昭	hotzone_4	キャッチ

- 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、その後パスをつないでいる攻撃の様子に共通点があるような攻撃が抽出された
- ただしトピック分布だけでは抽出しきれない場合も多くある
 - アクションタグやアクション数など確定的な条件付けによる検索も必要

まとめと今後の展望

- ボールタッチデータ/トラッキングデータから類似した一連の攻撃を自動で抽出する方法の提案
- 一連の攻撃を説明するような単語の生成
 - 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）（守備側）
守備脆弱度（守備側）、ボール周辺選手配置
- トピックモデルを用いた攻撃プレーのクラスタリングを実施、および得られたトピックの解釈
 - おおむねサッカーらしいトピックに分類することができた
- 類似プレーを抽出する手法の提案
 - 共通した動きを持つような攻撃が抽出された（プレーの中にあった）
 - ただしトピック分布以外の項も検索条件として導入する必要あり

- トピック数自体の推定も含めたモデルも検討
 - チーム戦術や対戦カードにより異なると想定される
 - AIC / BIC でモデル選択 or Dirichlet Process
- 12月の発表会では、トピックの時間変化を含めたモデルに拡張？
 - 「試合終盤ではゲームが動きやすい」などの表現
 - Dynamic topic model / Topic over time
 - オフェンス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から、時系列性を考慮？
 - 攻撃に入る直前の守備配置
- 変数の調整・修正
 - 連続変数をそのまま導入する
 - 各変数のセルサイズ・セル数の変化による感度分析
- 推薦システムの構築

- (2011) 浅井「MPEG-1 映像における類似プレー推定での複数基準動作例を用いた場合の性能評価」
http://ci.nii.ac.jp/els/110009622805.pdf?id=ART0010089894&type=pdf&lang=en&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1477896041&cp=
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, Jan (2003): 993-1022.
- Albert トピックモデル
http://www.albert2005.co.jp/technology/machine_learning/topic_model.html
- データ解析: LDAの実装(gensim)
<https://openbook4.me/projects/193/sections/1154>