

# プレミアリーグ過去10年データと 試合前テキストを用いた LSTM確率予測分析

27番 廣瀬林太郎

# プレミアリーグとは

- The Football Association Premier League Limited
- イングランドのプロサッカー1部リーグのこと
- 世界のプロサッカーリーグの中で最も市場規模が大きい
- 1年（8/15~6/26）、380試合で1シーズン
- チームの強さが均衡しており、予測難易度が高い

# 実験理由・背景

- 実用に耐える**妥当な確率**を作るうえで**試合前テキスト**が効どこまで効くか確かめるため
- Home／Draw／Awayは不均衡で、とりわけ引分の予測が難しい

# 実験

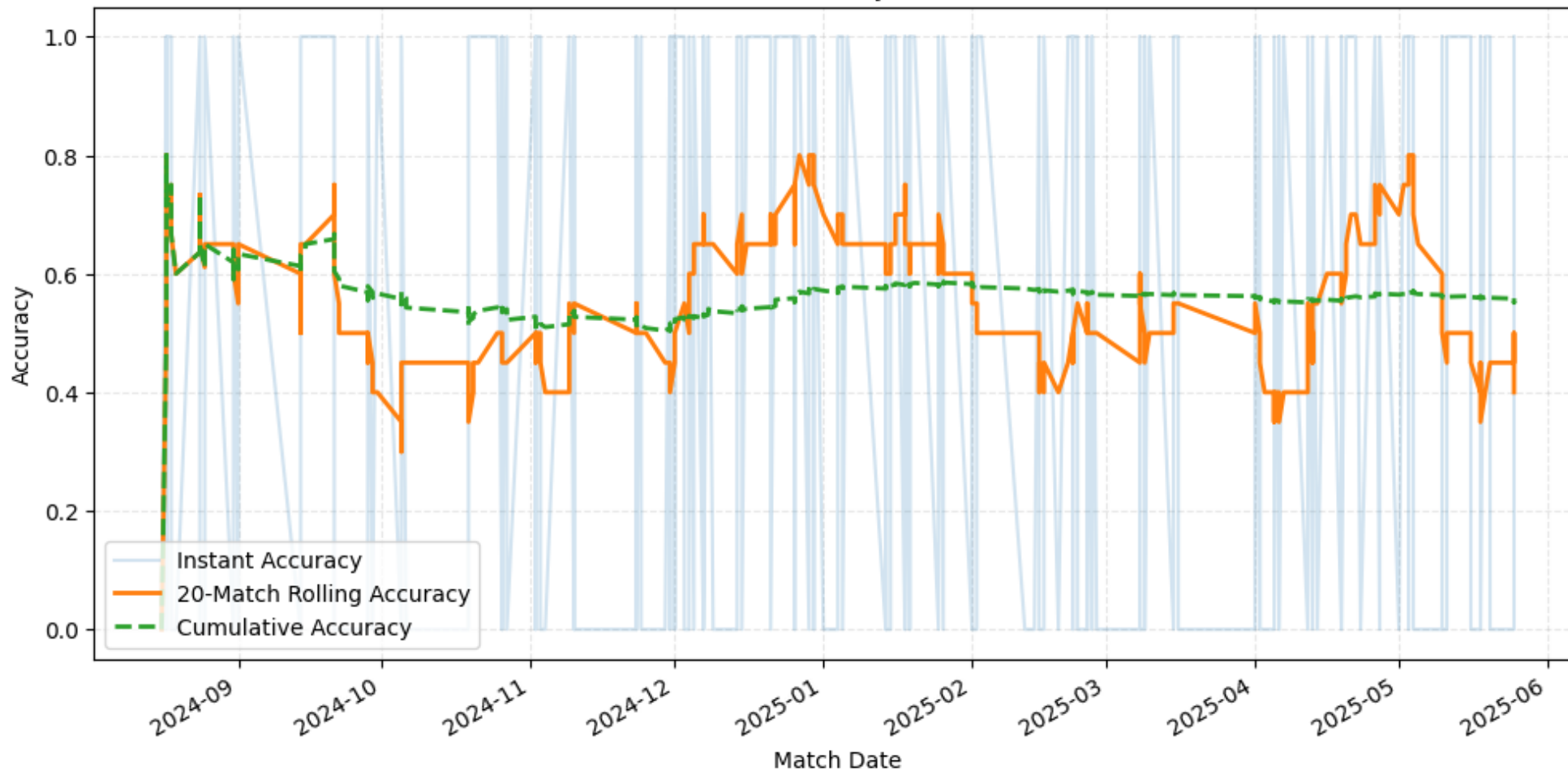
- プレミアリーグ各試合の勝敗（Home／Draw／Away）確率を目的変数とする **LSTM**を構築
- 説明変数は、試合時点より前に算出した数値特徴（Elo差、直近フォーム、ホームアウェイ別フォーム、休養日差、過去対戦の指数減衰相性、昇格フラグ）を入力
- **試合前7日間**のSNS投稿（Bluesky）をクラブ別キーワードで収集し、前処理後に **Doc2Vec**で埋め込み、PCAで圧縮したテキスト特徴を入力
- 正答率の目標としては **50%**

# チーム別正答率

チーム	試合数	正解数	正答率	テキスト数
Liverpool	8	6	0.750	9,187
Wolves	7	5	0.714	10,704
Newcastle	9	6	0.667	11,458
Leicester	8	5	0.625	10,181
Ipswich	7	4	0.571	4,900
Aston Villa	7	4	0.571	6,520
Man City	8	4	0.500	5,602
Tottenham	8	4	0.500	8,547
Crystal Palace	8	4	0.500	6,701
West Ham	7	3	0.429	9,761
Bournemouth	7	3	0.429	11,938
Brighton	7	3	0.429	13,080
Southampton	8	3	0.375	13,200
Brentford	8	3	0.375	5,898
Chelsea	8	3	0.375	10,639
Fulham	8	3	0.375	14,748
Man United	8	3	0.375	6,254
Arsenal	7	2	0.286	9,641
Everton	7	2	0.286	13,688
Nott'm Forest	7	0	0.000	9,836

# 瞬間精度と20試合移動平均精度

Multimodal Model Accuracy Over 24/25 Season



# 考察

- 小標本の高精度は過大評価の可能性
- Nが増えるほど正答率は**全体水準** ( $\approx 0.48$ ) に回帰
- **Draw予想**が難しいと予想していたが**Away予想**のほうが精度が低かった
- **テキスト量が多い** = **高精度**ではない  
(例: Man United 13,688語 でも 0.375)

# まとめ

- 含意：**確率の妥当性**が出なかった

## 今後

- テキストの**質的特徴**
- **直前の選手情報**
- **オッズ、Poisson**統合