

Twitterの実況ツイートを利用した タイムライン上のネタバレ情報検知

情報・通信工学専攻 コンピュータサイエンスコース 寺田研究室
1431050 齊藤 令

概要

- 興味があるが今必要のない情報
 - 例:ネタバレ情報:楽しみにしていたコンテンツの結果を知ってしまう
- Twitterではネタバレ情報が目に入りやすいので除去したい
- スポーツ等の試合では実況等で積極的にネタバレ情報が発信される
- 実況ツイートを利用して、スポーツ等の試合に関するネタバレ情報遮断が目的

提案システム

- 図1の赤丸をフィルタしたい
- ハッシュタグ等で検索し、実況ツイートを収集
- Bag of Wordsモデルでツイート, ベクトル化
- LSIで次元削減(1~128次元に削減)
- 3つの追加要素をベクトルに付加
- 機械学習で分類器作成

追加要素

- URLの有無
ツイートにURLがあるかないか
- 興奮度
連続した記述を単一文字に正規化
このとき減った文字数をカウント
「お兄様すげええええええええ」
「お兄様すげえ」 興奮度6
- ツイートの勢い
一定時間毎のツイート投稿頻度

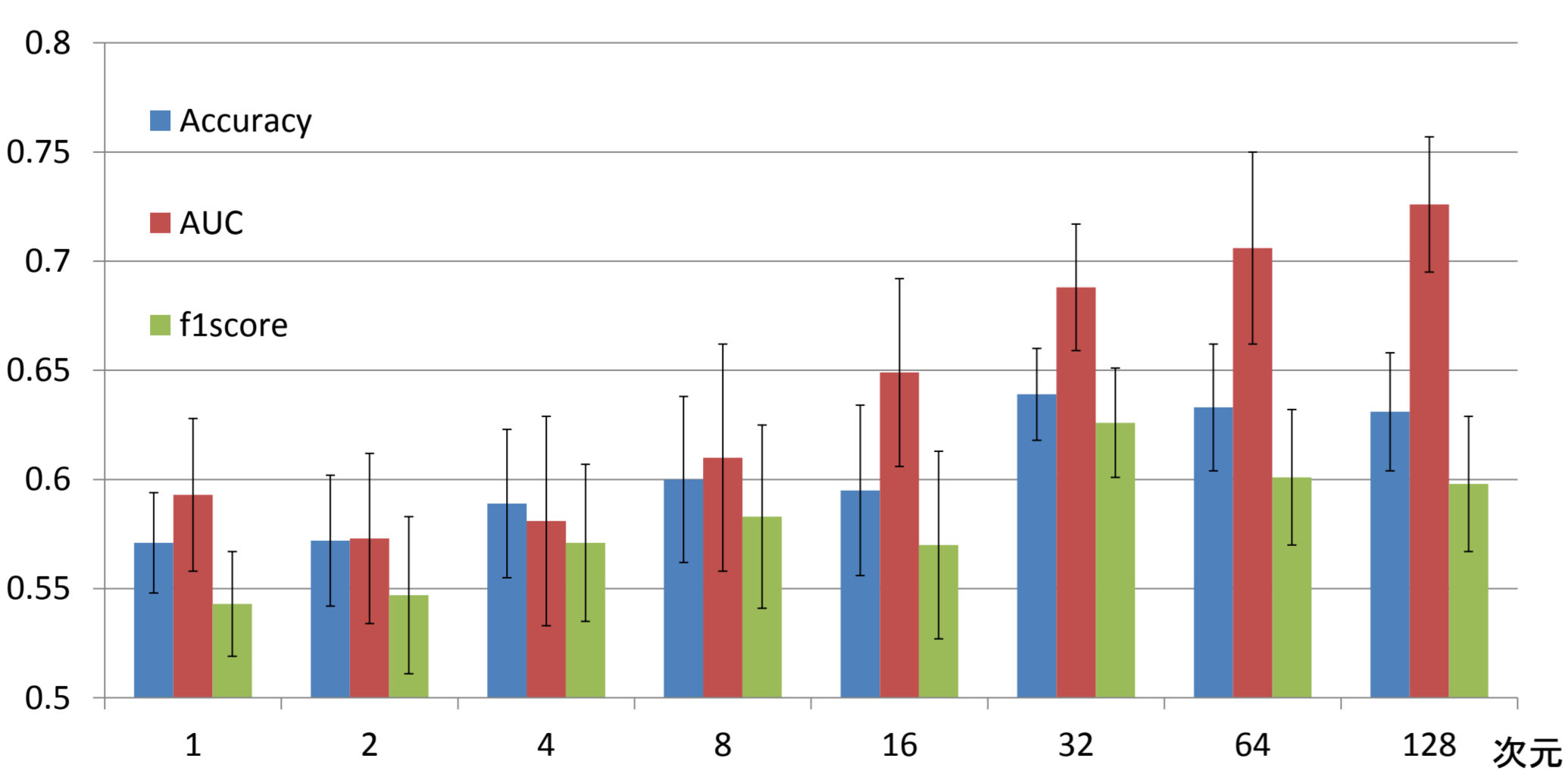


図4 単語ベクトルのみの実験結果

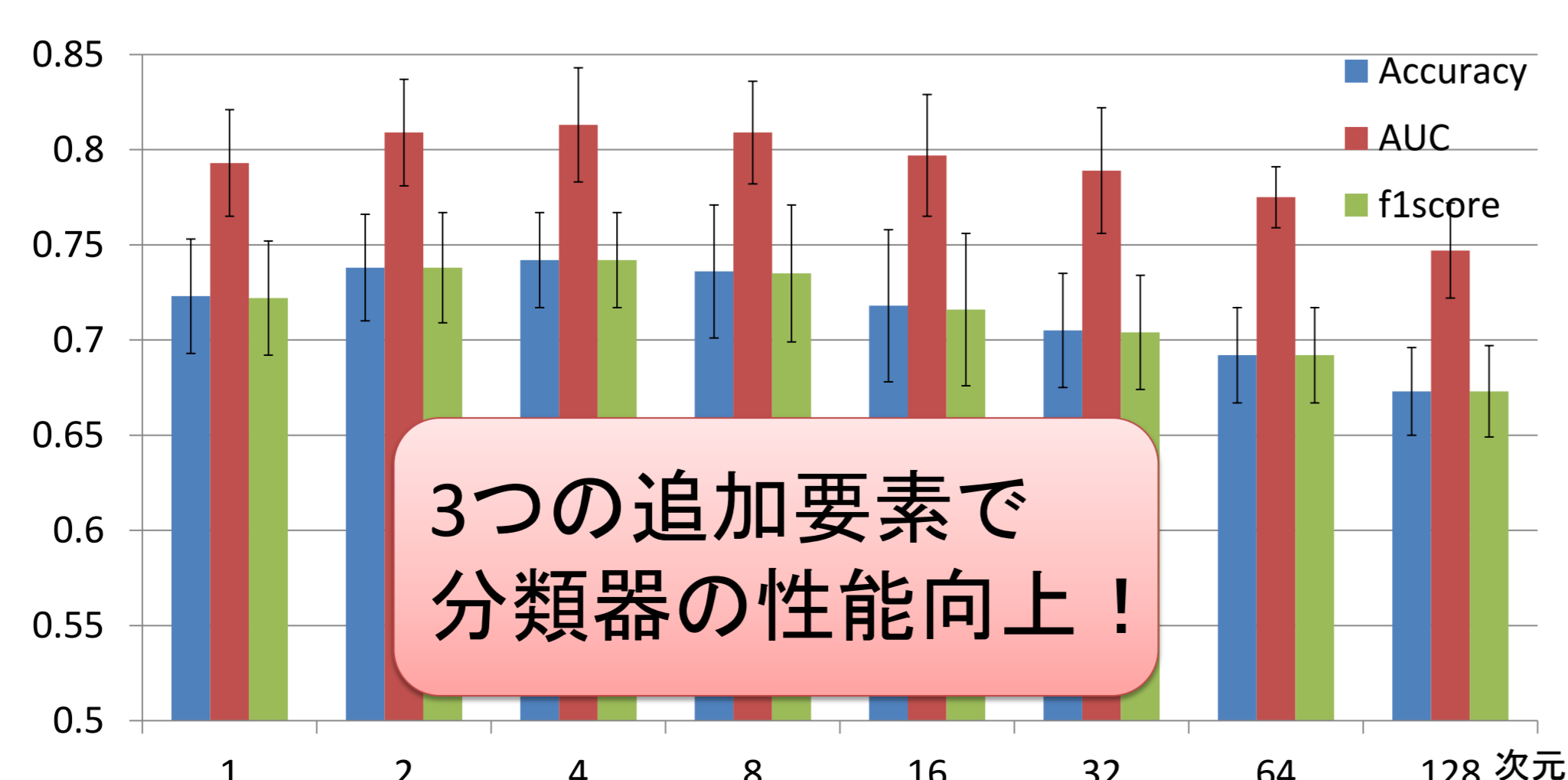


図5 3つの追加要素を追加したときの実験結果



図1 2014年日米野球の実況ツイート

赤丸の文では試合の結果がわかってしまう。これがネタバレ情報でフィルタリングの対象

黄線の部分では試合の展開がある程度読める

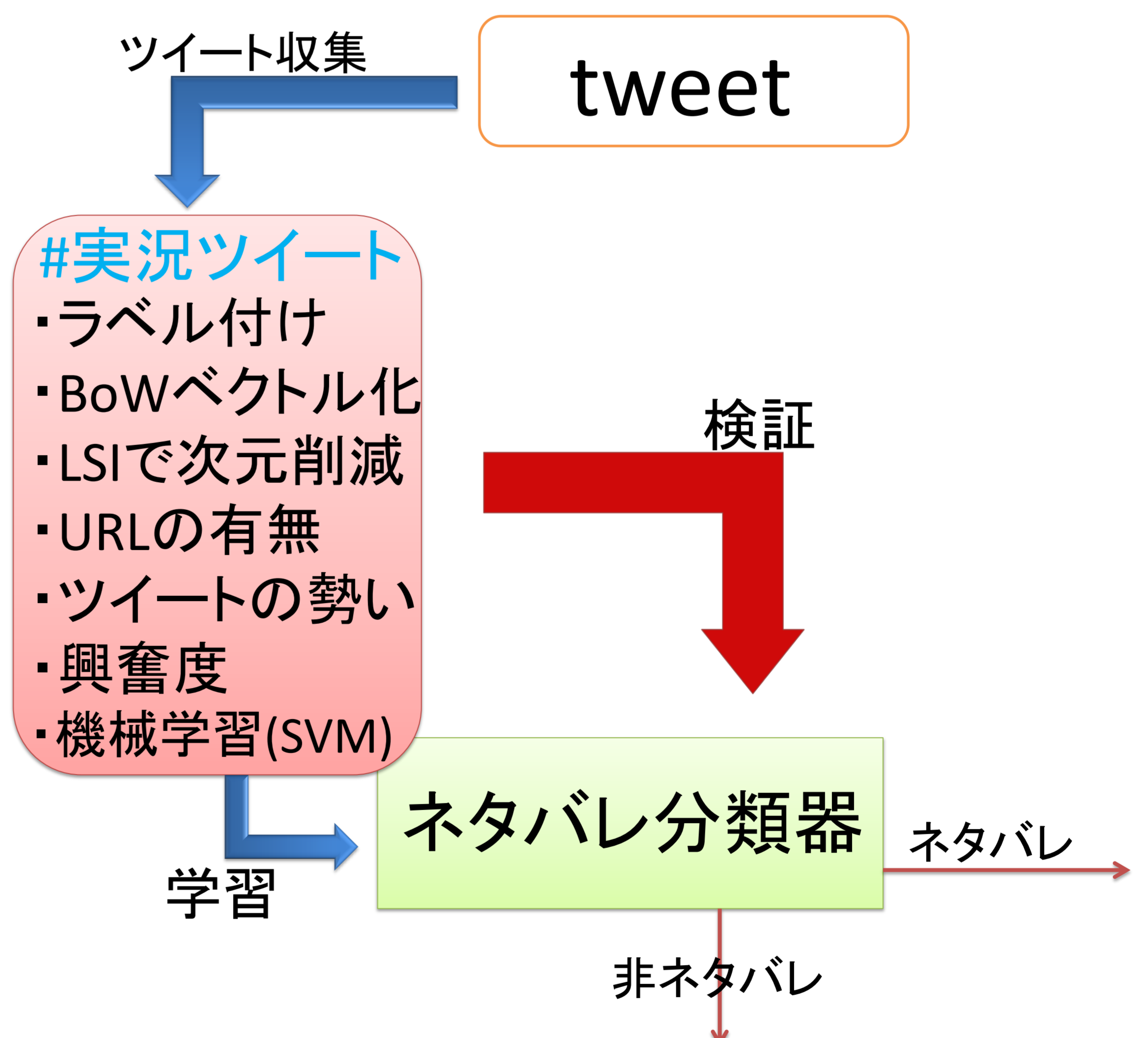


図2 システムの概略図

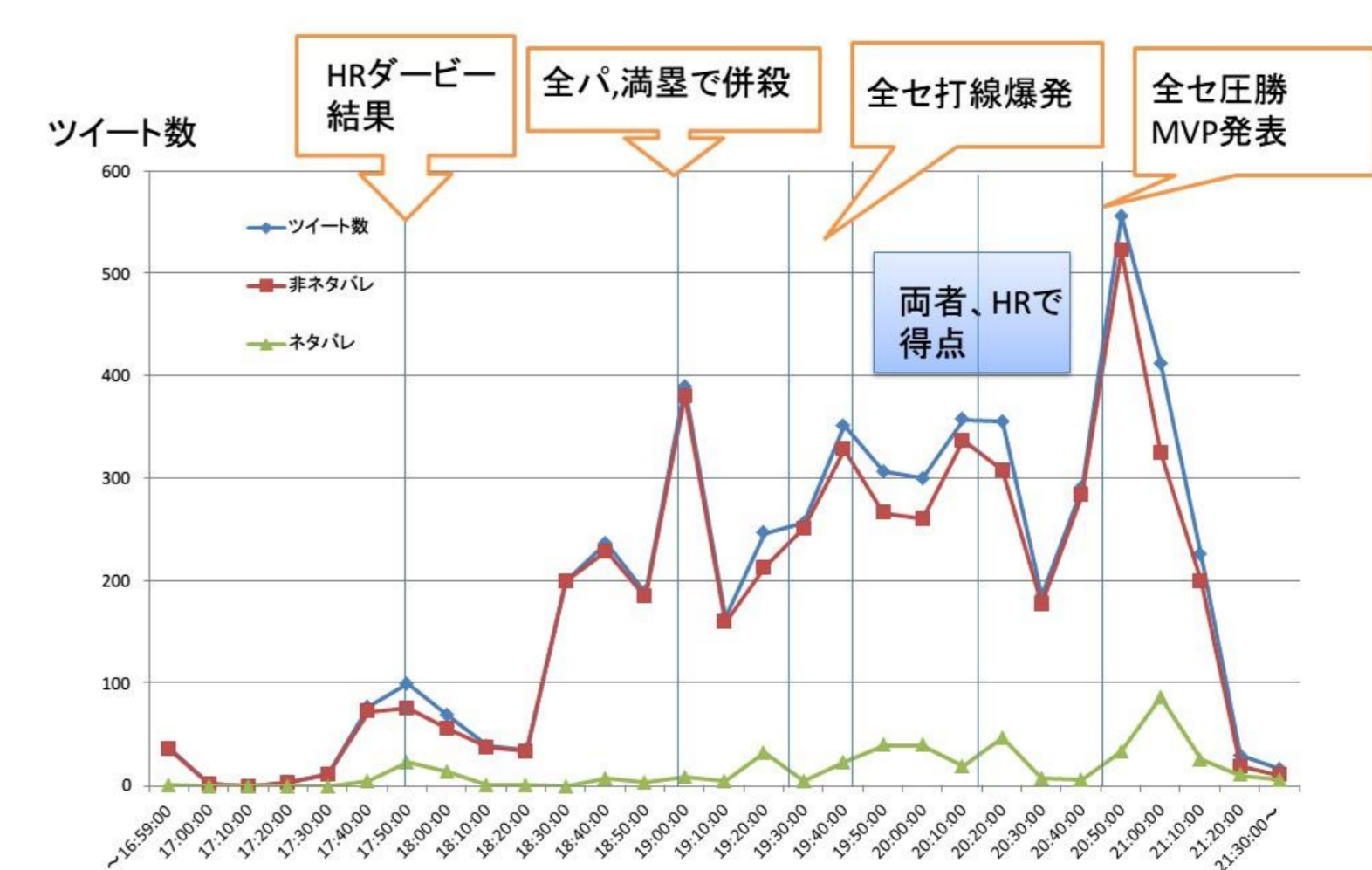


図3 2015プロ野球オールスターゲーム第2戦の実況ツイート推移