

プロ野球の実況ツイートを 対象とした マルチラベル分類

日本大学文理学部
情報システム解析学科尾崎研究室
山田慎也

研究の背景

Twitterでは「実況書き込み」が行われている

実況書き込みは
野球中継などでも行われている

これらをまとめるサイトが欲しい・・・

Twitter の実況書き込みを利用したスポーツ映像の要約
(小林尊志 ,野田雅文, 出口大輔, 高橋友和 ,井手一郎 ,
村瀬洋/2011)

ツイートをを用いた野球のLiveシステムの研究

(Hogun Park, Sun-Bum Youn, Geun Young Lee, Heedong Ko/2011)



第一歩として・・・

ツイートの分類が最低限必要

分類ラベル

実際のツイートを見てみると。。

例 1

澤村投手続投か…ボールのキしも落ちてきているが
リリース陣が疲れてるから仕方ないか。
傷を広げないでほしい。頑張れ！

・澤村投手続投
・リリース陣が
疲れてる

状況

・ボールの
キしも落ちて
きている

解説

・頑張れ！

応援

例2

バカかてめえ。ゲッツーとか・・・腹立つわ。

バカかてめえ

ゲッツーとか

腹立つわ

野次

状況

感想

一つのツイートには、複数の内容が含まれる。

一つのツイートに、複数のラベルを割り当てる(タグ付け)

今回使用するラベル

状況 解説 感想 応援 野次 その他

研究の背景

試合の実況ツイートを集め、分類すると
ユーザーの意見、考え、感想が分かる

というわけで・・・

実際に自分の手で振り分けてみました

データについて

Twitter4Jを利用

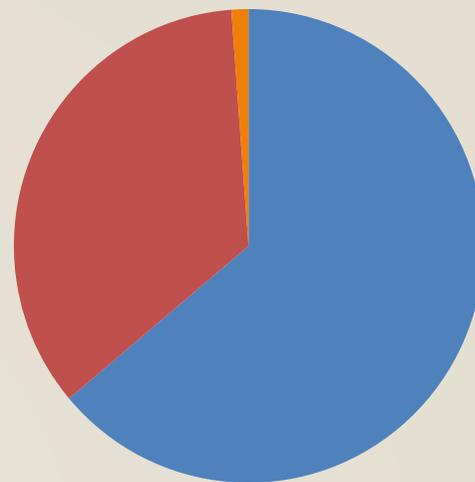
- ・データセットは野球のデータを使用
- ・収集期間は2013年5月14日から6月20日の間に
行われた2013年度日本生命セ・パ交流戦の
読売ジャイアンツの全24試合を対象
- ・24試合 13勝10敗1分け .565 交流戦順位**3位**

●巨人 3-5 ロッテ○
○巨人 5-4 ロッテ●
○巨人 7-2 西武●
○巨人 2-1 西武●
●巨人 3-7 日本ハム○
●巨人 1-2 日本ハム○
●巨人 1-2 楽天○
○巨人 10-4 楽天●
○巨人 3-2 オリックス●
△巨人 5-5 オリックス△
●巨人 0-5 ソフトバンク○
●巨人 2-3 ソフトバンク○

●巨人 3-4 ロッテ○
●巨人 2-3 ロッテ○
●巨人 1-5 西武○
○巨人 7-3 西武●
○巨人 4-1 日本ハム●
○巨人 4-2 日本ハム●
○巨人 5-3 楽天●
●巨人 3-5 楽天○
○巨人 7-1 オリックス●
○巨人 3-0 オリックス●
○巨人 6-1 ソフトバンク●
○巨人 11-3 ソフトバンク●

ラベル数の内訳

総ツイート数
46,912ツイート



ラベル数が1: **63.9%**

ラベル数が2: **35.0%**

ラベル数が3: **1.1%**

ラベル数が1: **解説**、**状況**など

ラベル数が2: **解説**と**状況**、**状況**と**感想**など

ラベル数が3: **解説**と**状況**と**感想**、**解説**と**感想**と**野次**など

解説が一番多い

機械学習を使うには、データを属性ベクトルで表す

文字数はそのツイートの文字数
頻出用語は収集した
ツイートから作成
選手名の有無
野球用語の有無
感情語の有無
応援用語の有無
野次用語の有無

今回使用する属性

- ・文字数
- ・選手名
- ・感情語
- ・野次用語
- ・頻出用語
- ・野球用語
- ・応援用語

判定にはそれぞれ辞書を作成し、用いた

-使用した辞書

- ・選手名辞書・野球用語辞書・感情語辞書
- ・応援用語辞書・野次用語辞書
- ・頻出用語辞書・標準のIPA 辞書

例

・ ロペス ナイス ヒット ! ! !

選手名、応援用語、野球用語が判定される

・ クソ!! 誉めるところ全くない。

ムカつくわ

野次用語、感情語の属性が判定される

分類について

分類- ある事例に対し, その事例が
属するラベルを決定する問題

単一ラベル分類

各事例を一つの
ラベルに分類

まだいける! がんばれ!

応援のみ

マルチラベル分類

各事例を複数の
ラベルに
同時に分類

澤村投手続投か…
ボールのキレも落ちてきているが
リリース陣が疲れてるから
仕方ないか。
傷を広げないでほしい。頑張れ!

状況 解説 応援

Gjorgji Madjarov ら (2012)

An extensive experimental comparison of methods
for multi-label learning

Pattern Recognition 45 3084-3104

マルチラベル分類

Problem

Transformation

問題を複数の単一ラベル
問題へと変換する
問題変換に基づく手法

Rakel

PPT BR RPC CLR

Algorithm

Adaptation

既存アルゴリズムの
拡張を伴うアルゴリズム
適応に基づく手法

ML-KNN

BP-MLL MMP PCT

評価基準

Gjorgji Madjarov ら (2012)

An extensive experimental comparison of methods
for multi-label learning

Pattern Recognition 45 3084-3104

- マルチラベル分類での、分類精度の評価に関して大きく2つのカテゴリが提案されている

①事例に基づく評価基準

事例ごとに評価値を算出し、その平均を求める

②ラベルに基づく評価基準

ラベル毎に評価値を算出してから平均を求める
ラベル毎の集計後に評価値を算出する

①事例に基づく評価基準

左:実際 / 右:予測

tweet	解説	状況	感想	...	その他
1	1 / 1	1 / 1	0 / 1	...	0 / 0
2	0 / 1	1 / 0	1 / 0	...	0 / 0
3	0 / 0	0 / 0	0 / 0	...	1 / 1
⋮					



事例(今回はtweet)にラベルが存在しているかないか実際に予測

- F値
- 再現率
- 適合率
- ハミングロス
- 精度

	予測:1	予測:0
実際:1	2	1
実際:0	2	1

再現率: $\frac{2}{3}$

適合率: $\frac{1}{2}$ → F値

調和平均 ↓

F値

②ラベルに基づく評価基準

左:実際 / 右:予測

tweet	解説	状況	感想	...	その他
1	1/0	1/1	0/1	...	0/0
2	0/1	1/0	1/0	...	0/0
3	0/0	0/0	0/0	...	1/1
⋮					
⋮					

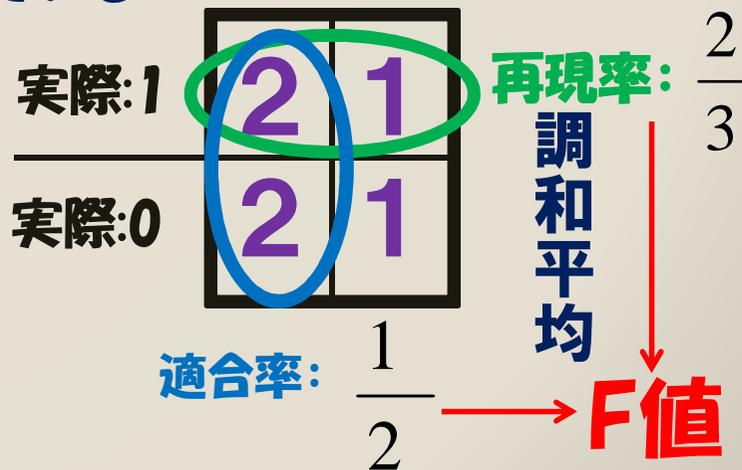
Micro (circled around the 'その他' column)

Macro (circled around the bottom row of the table)



それぞれ平均を出している 予測:1 予測:0

- Macro F値
- Macro 再現率
- Macro 適合率
- Micro F値
- Micro 再現率
- Micro 適合率



実験について

実験の目的

①手法間(アルゴリズム間)の比較
- ライブラリMulanを使用

②どのくらいの学習データが必要・適切なのか
- 直前N試合による比較と累積による比較をする

例 ①直前N試合による比較 N=3の時

直前3試合(1, 2, 3試合目)を使い4試合目を予測

1	2	3	4	5	...	24
---	---	---	---	---	-----	----



直前3試合(2, 3, 4試合目)を使い5試合目を予測

1	2	3	4	5	...	24
---	---	---	---	---	-----	----



直前3試合(21, 22, 23試合目)を使い24試合目を予測

1	2	...	21	22	23	24
---	---	-----	----	----	----	----

例 ② 累積(1~N試合)による比較

累積3試合(1, 2, 3試合目)を使い14試合目を予測

1	2	3	4	5	...	24
---	---	---	---	---	-----	----



累積4試合(1, 2, 3, 4試合目)を使い15試合目を予測

1	2	3	4	5	...	24
---	---	---	---	---	-----	----



累積23試合(1, 2, ..., 23試合目)を使い24試合目を予測

1	2	...	21	22	23	24
---	---	-----	----	----	----	----

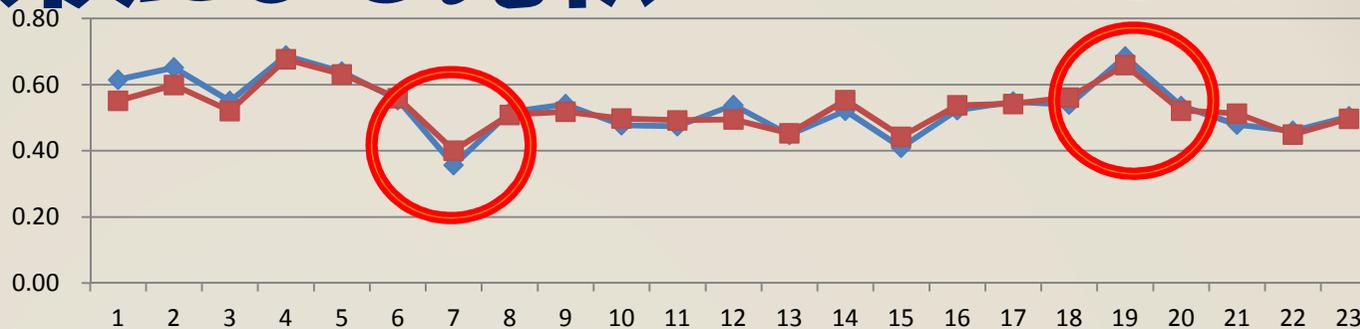
**これら2つの手法(アルゴリズム)と
11個の評価値を使って分類を行う**

では結果を見てみましょう

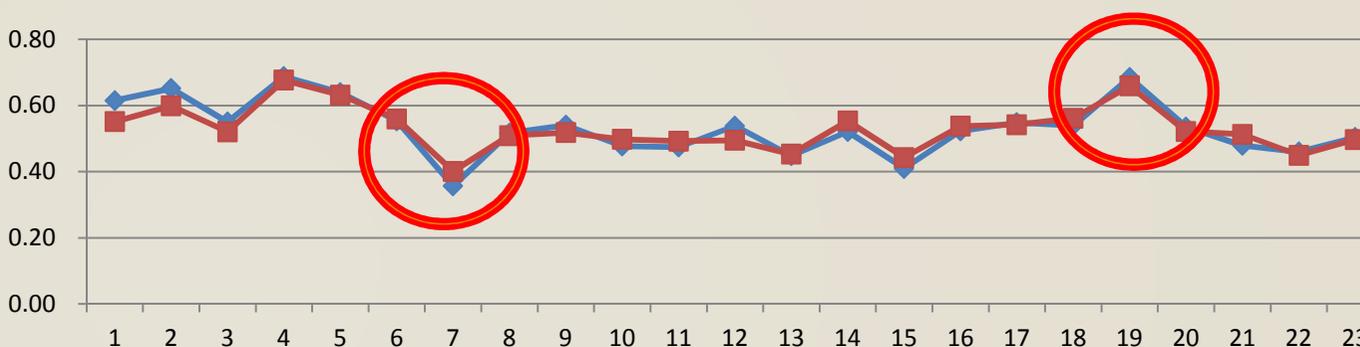
累積による比較

— Raker
— ML-KNN

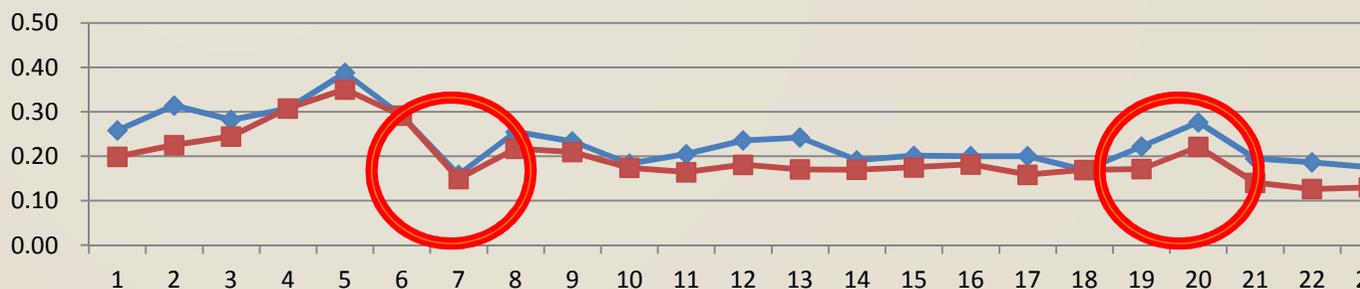
F値



Micro
F値



Macro
F値



F値、MicroF値はあまり差がないが
MacroF値はRakerの方が大きい

直前N試合による比較

	F値	microF値	macroF値
直前1試合	52.8%	52.4%	19.9%
直前2試合	52.4%	51.4%	19.3%
直前3試合	52.2%	51.0%	19.0%
直前4試合	52.6%	51.3%	18.6%
直前5試合	52.3%	51.1%	18.5%
直前6試合	51.5%	50.1%	17.4%
直前7試合	51.3%	50.0%	16.7%
直前8試合	51.6%	50.5%	17.3%
直前9試合	51.4%	50.4%	17.1%
直前10試合	51.6%	50.4%	17.2%

全体的にあまり数値が高くない。
特にmacroF値が低いという結果になった

ラベル別の結果

N	評価値	解説	実況	感想	応援	野次	その他
1	再現率	91.0%	30.3%	5.1%	0.9%	6.7%	15.2%
1	適合率	68.7%	30.7%	16.8%	3.8%	17.8%	22.5%
1	F値	77.6%	22.8%	5.4%	1.3%	6.4%	11.9%
2	再現率	91.3%	26.5%	3.4%	0.4%	6.1%	14.9%
2	適合率	67.8%	35.6%	19.6%	15.6%	21.2%	23.1%
2	F値	77.4%	20.4%	4.7%	0.8%	5.5%	14.3%
3	再現率	91.4%	24.8%	3.3%	0.2%	4.2%	15.9%
3	適合率	68.1%	38.0%	16.1%	5.6%	21.9%	20.3%
3	F値	77.5%	20.1%	4.7%	0.3%	5.2%	14.8%
4	再現率	93.2%	22.3%	2.4%	0.7%	5.4%	12.9%
4	適合率	67.8%	31.9%	16.3%	10.3%	23.7%	21.0%
4	F値	78.0%	19.5%	3.6%	0.9%	6.2%	11.9%
5	再現率	93.2%	22.1%	1.8%	0.1%	5.0%	11.3%
5	適合率	66.8%	27.4%	16.7%	2.1%	25.3%	21.9%
5	F値	77.4%	18.9%	2.8%	0.2%	6.2%	12.6%

解説はとてもよく当たるが、応援は全く当たらない

一 結論

生のデータを自分で振り分け、データセットを作った
自動分類を試みた

一 考察

- 辞書作成があまりうまくいかず、よい結果を得られなかった
のではないか
- 手法間では差があまりない
- 評価基準ではMacroF値が極端に低かった

一 今後の課題

- 対戦チームごとに比較する
- まとめサイトにする際視覚化を行う