

A9 :: Information Extraction

Date : 11-13-2013

Team

Name	Email
Mosharaf Chowdhury	mosharaf@berkeley.edu
Morgan Wallace	morgan@ischool.berkeley.edu
Shreyas	shreyas@ischool.berkeley.edu

Description

We did **suggestion 1** - our project uses a NLTK NE classifier to find **one of the team names** in the match.

- We take soccer match reports from ESPN that describe in flowery language what took place in the match.
- From this we parse for people and organizations to identify the teams and the players that participated in the match.
- We use a downloaded roster to create labels and then use that to cross-reference with our classifier predictions.
- We then go one step further and pass this to a Naïve Bayes Classifier for a 10-fold cross-validation using the extracted information.

Results

Our code results in up to 20% accuracy where the expected value of pure guessing is 5.5%. We are almost 4 times better than random guessing.

Motivation

Our motivation behind using this approach was to generate team names in addition to just the extracted player names.

contributions

Member Name	Contribution
Morgan	Crawled and scraped ESPN for report text and rosters
Mosharaf	NLTK NE parsing to extract player and team names
Shreyas	Added classifier for guessing team names
All	watched the code run for hours

Output

Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	10.3 : 1.0	
Everton = 'F'	Everto : Arsena =	10.3 : 1.0	
Arsenal = None	Newcas : Arsena =	10.1 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.6 : 1.0	
Swansea = 'F'	Swanse : Arsena =	9.3 : 1.0	
Spurs = 'F'	Totten : Arsena =	8.9 : 1.0	
Baggies = 'F'	West B : Arsena =	8.9 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	8.9 : 1.0	
Phil Dowd = 'F'	Sunder : Arsena =	8.9 : 1.0	
Brom = 'F'	West B : Arsena =	8.9 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	9.6 : 1.0	
Arsenal = None	Fulham : Arsena =	9.5 : 1.0	
Fulham = 'F'	Fulham : Arsena =	9.5 : 1.0	
Everton = 'F'	Everto : Arsena =	9.5 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.2 : 1.0	
Liverpool = 'F'	Liverp : Aston =	8.9 : 1.0	
Cardiff = 'F'	Cardif : Aston =	8.9 : 1.0	
Hull = 'F'	Hull C : Aston =	8.7 : 1.0	
Magpies = 'F'	Newcas : Arsena =	8.3 : 1.0	
Baggies = 'F'	West B : Arsena =	8.3 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	10.3 : 1.0	
Everton = 'F'	Everto : Arsena =	10.3 : 1.0	
Magpies = 'F'	Newcas : Arsena =	10.0 : 1.0	
Arsenal = None	Newcas : Arsena =	10.0 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.6 : 1.0	
Liverpool = 'F'	Liverp : Aston =	9.6 : 1.0	
Cardiff = 'F'	Cardif : Aston =	9.5 : 1.0	
Swansea = 'F'	Swanse : Arsena =	9.3 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	9.3 : 1.0	
Hull = 'F'	Hull C : Aston =	9.3 : 1.0	
None			
Most Informative Features			
Hull = 'F'	Hull C : Chelse =	9.3 : 1.0	
Chelsea = 'F'	Chelse : Arsena =	9.0 : 1.0	
Wojciech Szczesny = None	Chelse : Arsena =	9.0 : 1.0	
Everton = 'F'	Everto : Arsena =	8.9 : 1.0	
Michu = 'F'	Swanse : Chelse =	8.8 : 1.0	
Swansea = 'F'	Swanse : Chelse =	8.8 : 1.0	
Arsenal = None	Newcas : Arsena =	8.7 : 1.0	
Fulham = 'F'	Fulham : Chelse =	8.6 : 1.0	
Stoke = 'F'	Stoke : Arsena =	8.6 : 1.0	
Christian Benteke = None	Chelse : Aston =	8.4 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	10.9 : 1.0	
Everton = 'F'	Everto : Arsena =	10.9 : 1.0	
Arsenal = None	Hull C : Arsena =	10.6 : 1.0	
Stoke = 'F'	Stoke : Arsena =	10.4 : 1.0	
Swansea = 'F'	Swanse : Arsena =	9.9 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	9.9 : 1.0	
Spurs = 'F'	Totten : Arsena =	9.4 : 1.0	
Baggies = 'F'	West B : Arsena =	9.4 : 1.0	
Brom = 'F'	West B : Arsena =	9.4 : 1.0	
Toffees = 'F'	Everto : Arsena =	9.0 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	10.3 : 1.0	
Everton = 'F'	Everto : Arsena =	10.2 : 1.0	
Arsenal = None	Cardif : Arsena =	10.1 : 1.0	
Hull = 'F'	Hull C : Chelse =	9.4 : 1.0	
Swansea = 'F'	Swanse : Arsena =	9.3 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.3 : 1.0	
Liverpool = 'F'	Liverp : Aston =	8.9 : 1.0	
Paolo Di Canio = 'F'	Sunder : Arsena =	8.9 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	8.9 : 1.0	
Cardiff = 'F'	Cardif : Aston =	8.8 : 1.0	
None			
Most Informative Features			
Everton = 'F'	Everto : Arsena =	10.1 : 1.0	
Arsenal = None	Newcas : Arsena =	10.1 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.6 : 1.0	
Swansea = 'F'	Swanse : Arsena =	9.3 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	9.3 : 1.0	
Magpies = 'F'	Newcas : Arsena =	8.9 : 1.0	
Baggies = 'F'	West B : Arsena =	8.9 : 1.0	
Brom = 'F'	West B : Arsena =	8.9 : 1.0	
Liverpool = 'F'	Liverp : Aston =	8.3 : 1.0	
Chelsea = 'F'	Chelse : Aston =	8.3 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	9.6 : 1.0	
Liverpool = 'F'	Liverp : Aston =	9.6 : 1.0	
Everton = 'F'	Everto : Arsena =	9.6 : 1.0	
Cardiff = 'F'	Cardif : Aston =	9.5 : 1.0	
Arsenal = None	Newcas : Arsena =	9.4 : 1.0	
Hull = 'F'	Hull C : Aston =	9.3 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.2 : 1.0	
Fulham = 'F'	Fulham : Arsena =	8.6 : 1.0	
Spurs = 'F'	Totten : Arsena =	8.3 : 1.0	
Baggies = 'F'	West B : Arsena =	8.3 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	10.3 : 1.0	
Everton = 'F'	Everto : Arsena =	10.2 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.8 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	9.3 : 1.0	
Cardiff = 'F'	Cardif : Aston =	8.9 : 1.0	
Spurs = 'F'	Totten : Arsena =	8.9 : 1.0	
Baggies = 'F'	West B : Arsena =	8.9 : 1.0	
Swansea = 'F'	Swanse : Arsena =	8.9 : 1.0	
Magpies = 'F'	Newcas : Arsena =	8.9 : 1.0	
Brom = 'F'	West B : Arsena =	8.9 : 1.0	
None			
Most Informative Features			
Chelsea = 'F'	Chelse : Arsena =	9.7 : 1.0	
Arsenal = None	Newcas : Arsena =	9.4 : 1.0	
Hull = 'F'	Hull C : Chelse =	9.3 : 1.0	
Stoke = 'F'	Stoke : Arsena =	9.2 : 1.0	
Norwich = 'F'	Norwic : Chelse =	9.2 : 1.0	
Cardiff = 'F'	Cardif : Aston =	8.9 : 1.0	
Liverpool = 'F'	Liverp : Aston =	8.9 : 1.0	
Michu = 'F'	Swanse : Chelse =	8.8 : 1.0	
Swansea = 'F'	Swanse : Arsena =	8.8 : 1.0	
Sunderland = 'F'	Sunder : Arsena =	8.8 : 1.0	
None			
Accuracy Values: [0.13333333333333333, 0.06666666666666667, 0.2, 0.06666666666666667, 0.2, 0.13333333333333333, 0.13333333333333333, 0.13333333333333333, 0.06666666666666667, 0.06666666666666667]			