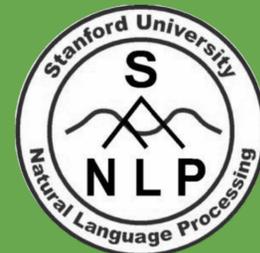




# Generating Football Summaries with NMT



Miguel Ayala  
Mentor: Michael Hahn

## Goal

to produce football (soccer) match summaries from game event vectors with an NMT framework

## Data

### source vectors:

- Football Events Dataset [1]

*7 8 na 1 lyon st etienne yoann gourcuff na na na na na 0 2 na 0 na 0*

### target sentences:

- SkySports Match Summaries

*a second half yoann gourcuff strike cancelled out kurt zoumas opener for les verts as lyon missed out on a third successive league win but stayed on course for champions league qualification*

### examples collected:

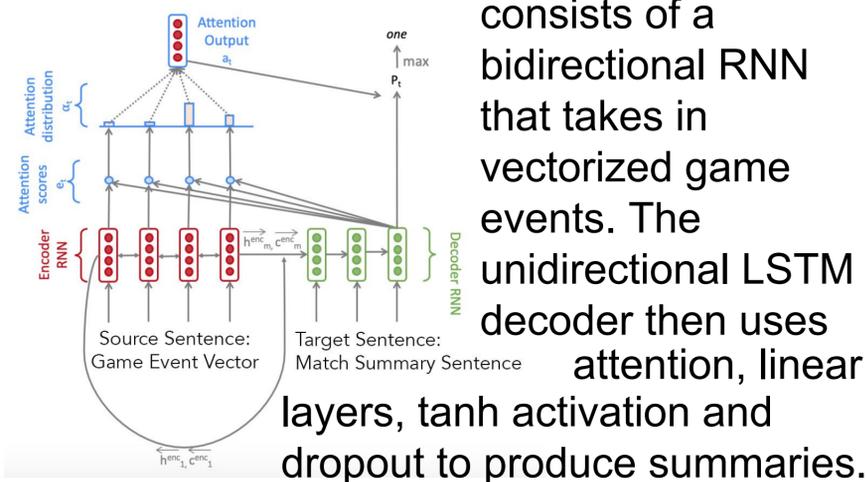
- Train: 8871, Dev: 2535, Test: 1267

## References

[1] Alin Secareanu. Football Events. <https://www.kaggle.com/secareanualin/football292events>. 2017.  
 [2] Nadjet Bouayad-Agha et al. "Perspective-oriented generation of football match summaries: Old tasks, new challenges". 2012.  
 [3] Nal Kalchbrenner and Phil Blunsom. "Recurrent continuous translation models". 2013.

## Approach

Previous approaches were rule-based. These approaches do not sound as natural as human summaries [2]. This uses an NMT [3] system to capture the right style. Encoding consists of a



## Results

comparison of source vector layouts

	Compound	Alpha	Alpha Numeric
<b>BLEU</b>	0.163	0.321	0.678
<b>Coherence</b>	3.4	3.8	7.8
<b>Specificity</b>	0.4	0.9	0.5
<b>Saliency</b>	3.3	3.2	4.9

## Sample Output

input

*4 1 na 1 elche getafe hinestroza przemyslaw tyton na na na na 1 4 2 0 1 0*

output

*the spaniard picked up the ball from the edge of the half hour mark but his shot was blocked by lorient*

reference

*at the other end on loan midfielder hinestroza cut in from the left but his strike was straight at elche custodian przemyslaw tyton*

## Analysis

The Alpha Numeric approach had the best performance in almost every category. The sample output illustrates its strengths and weaknesses: we have the minor non-logical the 'half-hour mark' as a physical location. Nevertheless, it details a shot that is saved by a goalie. This matches both the source vector and the reference. However, specificity is non-existent. Names are omitted or replaced completely.