UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

# STA4010W Project

Generating Phases of Play in Rugby Union using Recurrent Neural Networks


Thomas Edley
EDLTHO001

Graham Davies
DVSGRA012

**Supervisor:**
Neil Watson


Department of Statistical Sciences
University of Cape Town
South Africa

July 6, 2021

# 1 Background

The sport of Rugby Union is based around two teams competing for the possession of the ball and trying to move closer to the oppositions goal-line. Points are achieved by placing the ball in these goal-line areas, known as scoring a try, or by taking penalty kicks and drop goals, where a player will attempt to kick the ball through the posts situated in the middle of the oppositions goal-line. Teams with possession move the ball by means of phases. These phases are sequences of play that involve actions performed by the players on each team in an attempt to move closer towards the oppositions goal-line.

Statistical analysis has become a huge part of all sports. Whilst most sporting analysis is based around aggregate performance measures over a fixed time interval, the sequential nature of sports is often overlooked (Watson et al., 2020). There have been studies that have used the sequential nature of sport to predict the outcome of games in various sports, as well as other aspects of the game. Recurrent neural networks have been found to be useful in predicting outcomes of soccer games (Goddijn et al., 2018), American football games (Bosch  Bhulai, 2018) and many other sports. The use of RNN's have also been used to assess tactical decision-making in Rugby Union (Watson et al., 2020).

# 2 Aim of Project

The primary aim of this project is to determine whether one can use recurrent neural networks (RNN), in particular long short-term memory (LSTM) networks, to simulate a game of rugby union. This will be achieved by training a RNN that is able to generate realistic phases of play using the sequences of actions by players as input data. Due to the spatio-temporal nature of the data, we aim to use the field locations of these phases, as well as the team and players involved as additional input data. If we are able to achieve our primary objective before the date proposed, we aim to investigate the sequences generated by winning and losing teams to determine if there are any significant differences in the output achieved.

# 3 Methodology

As mentioned previously, we aim to achieve our goal using recurrent neural networks, more specifically, long short-term memory networks. Recurrent neural networks are a type of neural network that one can use to model sequential data. Our understanding of these networks comes from Graves (2014), a study about generating sequences using recurring neural networks. RNN's are similar to standard feed-forward neural networks, except that their output distribution is additionally influenced by internal memory. Sequences are generated from a trained network by iteratively sampling from the output distribution. These samples are then fed into the network as input at the next step, such that the network is taking in new input as well as input from previous iterations. RNN's often fall short as a result of not being able to store information from past inputs for very long (Hochreiter et al., 2001).

Long short-term memory networks are extensions of RNN's that have a longer memory. It allows RNN's to remember inputs for a longer period of time. Figure 1 below displays the structure of a single LSTM cell.

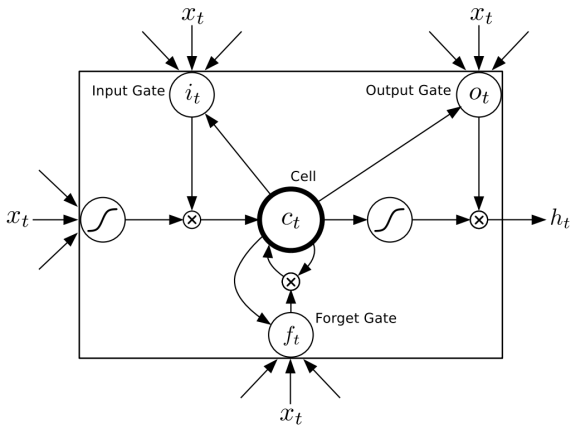


Figure 1: Long short-term memory cell

From Figure 1 above we see that the LSTM cell consists of three gates. The input gate determines whether new input should be let in or not, the output gate determines impact on the output distribution at the current step and the forget gate will delete information if it is not important. The cell state $c$ has information added to it by the various operations in the network.

# 4 Data

The data consists of five different rugby competitions, namely, The Heineken cup, the European Rugby Championship, Super Rugby, The Six Nations and The Rugby Championship. The years/seasons vary between competitions but range from 2013 to 2015. The data in extensive and contains twenty three variables for each observations. The data for each competition is stored on a separate sheet, all of which are in the same formet. The first six lines for the Super Rugby datasheet are shown below:

| id | fx_id | prd | pl_id | tm_id | time | act | act_type | act_res | q3 | q4 | q5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8175178 | 515011 | 1 | 96 | 96 | 0 | 17 | 407 | 0 | 0 | 0 | 0 |
| 8175179 | 515011 | 1 | 21457 | 12 | 3 | 14 | 281 | 286 | 0 | 0 | 0 |
| 8175180 | 515011 | 1 | 10028 | 96 | 3 | 18 | 275 | 277 | 0 | 0 | 0 |
| 8175181 | 515011 | 1 | 10028 | 96 | 5 | 1 | 105 | 120 | 0 | 0 | 0 |
| 8175182 | 515011 | 1 | 10028 | 96 | 9 | 4 | 190 | 193 | 186 | 0 | 0 |
| 8175183 | 515011 | 1 | 96 | 96 | 9 | 15 | 307 | 315 | 0 | 0 | 0 |

| m | x_crd | y_crd | x_end | y_end | score_adv | play_num | set_num | seq_id | ps_times | ps_endst |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 50 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 37 | 37 |
| 0 | 50 | 34 | 86 | 22 | 0 | 0 | 0 | 1 | 40 | 40 |
| 0 | 13 | 47 | 0 | 0 | 0 | 1 | 1 | 2 | 40 | 40 |
| 2 | 13 | 47 | 0 | 0 | 0 | 1 | 1 | 2 | 42 | 42 |
| 0 | 16 | 47 | 50 | 68 | 0 | 1 | 1 | 2 | 46 | 46 |
| 37 | 13 | 47 | 50 | 68 | 0 | 1 | 1 | 2 | 40 | 48 |

Table 1: The first six line from the Super Rugby data sheet

A brief description of the variables are shown below:

| Name | Description |
|---|---|
| id | Identification number of observation |
| fx_id | Identification number of fixture |
| prd | Indicates first or second half. |
| pl_id | Identification number of player |
| tm_id | Identification number of team |
| time | Time in minutes and second(mmss) |
| act | Action of game event recorded |
| act_type | Decribes action using qualifiers |
| act_res | Result of game event |
| q3, q4, q5 | Additional qualifiers |
| m | Metres gained for x direction |
| x_crd, y_crd | X and Y coordinate where the event begins |
| x_end, y_end | X and Y coordinate where the event ends |
| score_adv | Represents the score for the home team |
| play_num | Represents the phase of play in current possesion |
| set_num | ? |
| seq_id | ? |
| ps_times | ? |
| ps_endst | ? |

Table 2: Description of varaibles in the dataset

Due to the size, sequential nature, and extensiveness of the dataset, this dataset will suffice for the purpose of the project.

# 5 Research Plan

## Timeline

The schedule below outlines the proposed dates at which we aim to meet various deadlines by:

| **Task** | **Timeline** |
|---|---|
| Cover relevant literature, have a complete understanding of the data set, and perform data wrangling using R. | July, 31 |
| Skeleton outline write-up complete, as well as Literature Review. | August, 1-6 |
| Submit Progress Report | August, 6 |
| Completed final RNN algorithm | August, 31 |
| Majority of first draft done and begin to evaluate secondary objective | September, 31 |
| First draft and secondary objective code completed | October, 19 |
| Presentations | October, 20 - 22 |
| Final write up and nessacary additions | October, 22 - November, 9 |
| Final hand-in | November, 9 |

Table 3

## Division of Labour

Living in close proximity to one another, we feel we do not need to have a specific division of labour. However, each small task along the way will be discussed so we can move forward together.

**- END -**

# References


Bosch, P., Bhulai, S., 2018. Predicting the winner of NFL-games using Machine and Deep Learning .

Goddijn, S., Moshkovich, E., Challa, R., 2018. A Sure Bet: Predicting Outcomes of Football Matches .

Graves, A., 2014. Generating sequences with recurent neural networks. arXiv:1308.0850v5.

Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., 2001. Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-term Dependencies. A Field Guide to Dynamical Recurrent Neural Networks .

Watson, N., Hendricks, S., Stewart, T., Durbach, I., 2020. Integrating machine learning and decsision support in tactical decision-making in rugby union. Journal of the Operational Research Society doi:10.1080/01605682.2020.1779624.