Basketball play-by-play information extraction

Project Proposal and Work Plan

WRITTEN BY

Name: Xavier Rubiés Cullell
Date: 12/03/2021

REVIEWED AND APPROVED BY

Name:
Date:

Project overview and goals

The project is carried out at Computer Science Department and is supervised by Pere-Pau Vázquez Alcocer.

Traditionally, in the basketball world, one information source was generated from a match development: the play-by-play. It consists in a document describing the game action by action and indicating the time every action occurred. However, since season 2013-2014, the NBA (the most important basketball league) is using motion cameras that record the players and ball position 25 times every second. Each team has access to this data and uses machine and deep learning to exploit it.
The purpose of this project is to program the automatic extraction of the maximum amount of available information (relating to team/individual player performance) from play-by-play match data, in order to show the possibilities of this data source and to prove its value.

The project main goals are:

1. Programming an automatic obtention of a box score (the main statistics source) of a match from play-by-play data
2. Programming of automatic match features extraction. They can be more direct, such as, the longest time without scoring by a team, or more abstract, such as trying to determine whether a timeout made a change in a team performance
3. Design and implementation of visualisations describing the performance of a team, player or top X players in a match.
4. Generation of a dynamic visualisation of the evolution of a match (a visual play-by-play)
5. Making the tool able to read play-by-play from other/any sources (specially to be able to deal with alternative languages to English)

Project background

Although there exist previous projects involving automatised treatment of basketball match data, the project starts from scratch. It does not belong to any organisation or company. The student is the only person completely working on it and he decides its direction and programs it all by himself. As stated before, the project is supervised by Pere-Pau Vázquez Alcocer, who might contribute with suggestions about what to do or how to approach it. The main project initial ideas were also provided by the author.

Work Plan

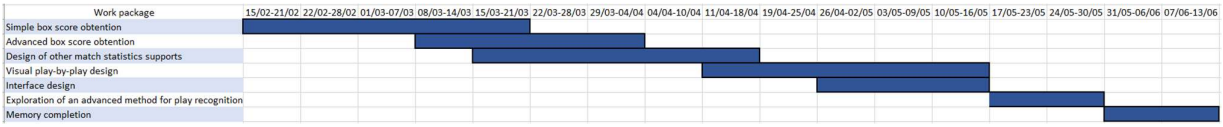
Tasks and Milestones. Gantt Diagram

Describe the main envisaged tasks and milestones of the project providing their estimated periods of time and deadlines, respectively. If informative, include a Gantt diagram showing the dependencies among tasks.

*Note that one of the milestones has to be the delivering of a **Project Critical Review**; in this semester, due to April 30th 2021. In it, the project evolution is to be discussed and, if necessary, the work plan redefined.*

1. Simple box score obtention
2. Advanced box score obtention
3. Design of visualisations and tables showing statistics, either of the match or of the evolution during a season
4. Design of a visual play-by-play
5. Interface design
6. Exploration of a method (such as transformer or clustering) that adapts to other play-by-play syntactic formats and to other languages
7. Memory completion

The project is started by using *Basketball Reference* site as the data source. The intention of the author is to be able to deal with play-by-play in different formats and different languages, in order to be able to treat play-by-play data from "any" league. However, as different data sources show data in very different formats, finding a general approximation for understanding different play-by-play sources might make little sense. That is why the author will probably program the extraction of a few more sources independently, not by using a general method. If other packages need more time, this work package might be discarded.



Current state

At this time, the student has developed the following packages and subpackages:

1. Simple box score obtention **– Finished**
   1.1. Definition of the "standard" play-by-play language and format
   1.2. From "standard" play-by-play to box score
   1.3. From *Basketball Reference* play-by-play language to "standard" play-by-play
   1.4. Adaption to format in *Basketball Reference*
   1.5. Creation of the full process and detail correction

2. Advanced box score obtention **– On process**
   2.1. Interval-restricted box score obtention **– Finished**
   2.2. Study of additional variables to be added (either devised by the author or found on sources) and its programming **– On process**
3. Design of visualisations and tables showing statistics, either of the match or multiple matches during a season **– On process**
   3.1. Design of match statistics extraction methods **– On process**
   3.2. Design of a graphic showing the shooting distance statistics of a match
   3.3. Design of a method allowing to see a visualisation of the evolution of statistics during a season

Meeting and communication plan

Describe the meeting and communication plan established between you and your advisor(s) in order to developed the proposed project.

The meetings are being done at convenience, being set approximately once every week and a half.

Generic skills

The following generic skills will be promoted and assessed during the development of the project. Note that an initial set of generic skills has been defined by your advisor when creating the project proposal. Check whether this initial proposal is still the most suitable one for the project and make, if necessary, the required changes both in this document and in the academic platform (Racó).

Be aware that if you have some of the third level generic skills not scored yet with A or B, you can work them in your TFG in order to obtain your Bachelor degree with the set of generic skills completely acquired.

#	Generic Skill	Assessed
GS1	Innovation and entrepreneurship	
GS2	Societal and environmental context	
GS3	Oral and written communication	X
GS4	Teamwork	
GS5	Survey of information resources	X
GS6	Autonomous learning	X

GS7	Communication in a foreign language	X
GS8	Gender perspective	