

Projektbericht zum Modul Information Retrieval und
Visualisierung Sommersemester 2021

Visualisierung von Daten des Videospiels Fifa 21

Johannes Lange

7. September 2021

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
1 Einleitung	1
1.1 Anwendungshintergrund	1
1.2 Zielgruppen	1
1.3 Überblick und Beiträge	2
2 Daten	3
2.1 Technische Bereitstellung der Daten	3
2.2 Datenvorverarbeitung	3
3 Visualisierungen	4
3.1 Analyse der Anwendungsaufgaben	4
3.2 Anforderungen an die Visualisierungen	4
3.3 Präsentation der Visualisierungen	4
3.3.1 Visualisierung Eins	4
3.3.2 Visualisierung Zwei	5
3.3.3 Visualisierung Drei	5
3.4 Interaktion	5
4 Implementierung	5
5 Anwendungsfälle	5
5.1 Anwendung Visualisierung Eins	6
5.2 Anwendung Visualisierung Zwei	6
5.3 Anwendung Visualisierung Drei	6
6 Verwandte Arbeiten	6
7 Zusammenfassung und Ausblick	6

Abbildungsverzeichnis

1	Darstellung des Scatterplots Quelle: eigene Darstellung	4
2	Darstellung der Parallelen Koordinaten Quelle: Eigene Darstellung	5
3	Darstellung einer FUT Karte inklusive relevanter Werte am Beispiel Cristiano Ronaldo Quelle: https://www.ea.com/de-de/games/fifa/fifa-21/news/fifa-21-player-ratings-best-strikers-st-cf	6
4	Ausschnitt aus Baumdarstellung Quelle: Eigene Darstellung	6

Abkürzungsverzeichnis

FUT Fifa Ultimate Team

1 Einleitung

Die Visualisierung von Daten nimmt mit Hinblick auf Big Data und damit immer unübersichtlicheren Datengrundlagen an Bedeutung zu. Um aus großen Mengen von Daten neue Informationen zu gewinnen reicht es nicht aus die Daten direkt zu analysieren. Durch die Anwendung der richtigen Visualisierungstechniken können neue Informationen gewonnen werden. In diesem Bericht geht es um die Visualisierung von Daten aus dem Computerspiel FIFA 21. Da sich diese Daten allerdings auf tatsächliche Fussballer beziehen besteht die Hoffnung mittels Visualisierung dieser Daten Rückschlüsse auf die Sportler selbst schließen zu können. So könnten Erkenntnisse aus den Daten deutlich wertvoller sein als zuerst vermutet.

Der gegebene Datensatz ist äußerst umfangreich, zu über 16000 Sportlern sind jeweils 107 Variablen festgehalten. Diese reichen von *Name* über *Verein* bis hin zu *Schusskraft* eines Spielers. Da nicht alle Eigenschaften eines Fussballers konkret festgehalten werden können, sind einige der Werte geschätzt. Da aus diesem Grund wie oben beschrieben nicht sicher auf die tatsächlichen Spieler zurückgeschlossen werden kann, bietet sich eine Analyse der Daten bezüglich der Nähe zur Realität an. Doch selbst wenn die Daten nicht zu neuen Erkenntnissen bezüglich der realen Sportler führen können sich für Spieler des Videospiels neue Erkenntnisse ergeben.

1.1 Anwendungshintergrund

Was sind die Daten, sind die Visualisierungslösungen angebracht, möglicherweise auf einzelne Variablen eingehen[Munzner2008]

Die Frage, welche sich bei diesen Daten stellt ist, welchen Nutzen sie außerhalb eines Videospiels haben. Deswegen ist mein Ansatz zu vergleichen wie genau diese Daten die Wirklichkeit widerspiegeln. Ein Beispiel dafür könnte sein ob sich das Rating eines Spielers mit der körperlichen Entwicklung eines Spielers bewegt. Sollte dies zutreffen müssten Spieler im Alter ihres physischen Peaks das höchste Rating haben.

1.2 Zielgruppen

Da es sich bei den Daten um die eines Videospiels handelt ist davon auszugehen, dass eine der Zielgruppen der Daten Spieler des Videospiels sind. Diese würde ich weiterhin in die des Spielmodus *Karriere* und die des Spielmodus *Fifa Ultimate Team FUT* unterteilen.

Spieler des Karrieremodus spielen das Spiel als Manager eines Vereins indem die Entwicklung von jungen Spielern wichtig ist und vor allem der Vergleich von Variablen wie Gehalt, Alter, Rating oder Potential relevant ist. In diesem Fall bietet sich eine Darstellung per Scatterplot an, da sich so schnell die aussichtsreichsten Fussballer erkennen lassen. Da der Karrieremodus nicht Online gespielt wird und die Schwierigkeit manuell festgelegt werden kann, ist davon auszugehen, dass das Vorwissen der Gruppe stark schwankt. Während einige Spieler bereits viel Wissen und

sehr genaue Informationen benötigen, brauchen andere erst einmal Grundwissen bezüglich der Spieler.

Spieler des FUT Modus sammeln die Fussballer wie in einer Art Sammelkarten Spiel. Mit den Karten der Fussballer kann dann Online gegen andere Spieler angetreten werden. Dabei sind vor die aktuellen Werte der Spieler von Relevanz, da sich diese nicht durch Training oder ähnliches ändern lassen. Weil dieser Modus sehr kompetitiv ist kann ein Fussballer bereits durch einen schlechten Wert in einer Kategorie für Spieler unbrauchbar werden. Dadurch, dass *FUT* zu den Onlinespielen gehört in denen viele Spieler sich durch *Min-Maxing*[1] einen Vorteil verschaffen wollen ist von einem hohen Vorwissen der Anwender der Visualisierung auszugehen.

Fifa 19 Spieler -> Viel Vorwissen, Fussball Fan -> recht viel Vorwissen, Sportinteressent -> Kaum Vorwissen

Zielgruppen für Scatterplot:

Parallele Koordinaten:

Baumdiagramm: Hier lässt sich erkennen welche Sportliga Europas im Durchschnitt die besten Spieler hat. Deswegen ist dies für X interessant.

Mögliche Zielgruppen: Videospieler, Teilnehmer einer Fantasy Fußball Liga

1.3 Überblick und Beiträge

Bei den Daten handelt es sich um einen Datensatz des Fussball Videospiels Fifa 19 dieser enthält Daten zu über 18000 Fussballern. In diesen enthalten sind neben Variablen wie Name, Verein und Alter auch ein ungefähr nach der Qualität des Fussballers festgelegtes Rating sowie Bewertungen seiner fussballerischen Fähigkeiten (Schießen, Passen, Dribbling, Verteidigen, Geschwindigkeit, Physis, etc.)

Die erste verwendete Visualisierungstechnik ist der Scatterplot, dieser ist gut dazu geeignet zwei Variablen der Fussballer zu vergleichen. Bspw. könnte untersucht werden ob ein Fussballer entsprechend seines Könnens (Rating) verdient (Wage).

Bei der zweiten verwendeten Visualisierungstechnik handelt es sich um Parallele Koordinaten. Diese, im Vergleich zum Scatterplot, etwas fortgeschrittene Technik eignet sich um mehr als zwei Werte miteinander zu vergleichen. Ihr Einsatz bietet sich daher an um Werte eines Fussballers in den Attributen: Schießen, Passen, Dribbling, Verteidigen, Geschwindigkeit oder Physis miteinander zu vergleichen und dies vorher nach der Position des Fussballers zu Filtern um zu untersuchen ob sich ein Spieler anhand seiner Attribute dafür eignet auf dieser Position zu spielen. (Bsp. ein Stürmer benötigt vor allem Geschwindigkeit, Schießen und Dribbling, ein Verteidiger hingegen Verteidigen, Physis und Geschwindigkeit).

Die dritte Visualisierungstechnik ist die Darstellung als Baumdiagramm. durch sie werden Spieler ihren Teams und diese ihren jeweiligen Ligen zugeordnet.

Als dritte und letzte Technik habe ich die Baumdarstellung ausgewählt um sichtbar zu machen welche Fussballliga Europas im Durchschnitt die besten Spieler hat. Deswegen ist dies für X interessant.

2 Daten

Beschreibung der gegebenen Daten, Sind sie zur Beantwortung der Fragestellungen geeignet? Welche zusätzlichen Daten wurden genutzt

Grundsätzlich eignen sich die Daten gut um die gewünschten Fragestellungen beantworten zu können, jedoch enthält der Grunddatensatz einige Felder, die für die Visualisierung nicht nötig sind, deswegen wurde der Datensatz in der Vorverarbeitung noch verkleinert (Siehe 2.2 auf Seite 3). Außerdem ist der Datensatz sehr groß, was gerade bei den parallelen Koordinaten zu Problemen führen kann wenn der ganze Datensatz angezeigt wird, deswegen wurde sich bei den parallelen Koordinaten dazu entschieden nach zusätzlichen Dimensionen wie Nationalität zu Filtern um dies so übersichtlicher zu gestalten. Da Datenwerte wie Größe, Alter und Rating der Spieler diskret sind wurde sich dazu entschieden im Scatterplot die Anzahl an Spielern welche in diesem Punkt enthalten sind auszugeben. Weiterhin wurde die Opazität der Punkte verringert, da so zu sehen ist an welchen Stellen sich mehrere Spieler überlagern.

2.1 Technische Bereitstellung der Daten

Die Daten sind in einem privaten Github gehostet¹. Dort liegen die Daten der drei Visualisierungen jeweils als *.csv* vor. Die verwendeten Variablen sowie die Anzahl an Fussballern wurden jeweils der Visualisierungstechnik entsprechend angepasst um den Datensatz so klein wie möglich zu halten.

Auch die Daten der Visualisierung als Baumdiagramm liegen als *.csv* vor und werden erst im Programm zu einer *.json* encoded und dann wieder zu einem Baumdiagramm decoded. Da in den Daten der Spieler keine Daten zu den Fussballligen in denen die Vereine spielen, enthalten sind, sind diese zusätzlich hinzugefügt worden.

2.2 Datenvorverarbeitung

Die Daten der Fussballer bestehen aus ca. 19000 einzelnen Fussballern, jeder dieser hat 93 einzelne Variablen, diese sind für die Visualisierung generell nicht alle relevant. Weiterhin unterscheiden sich die relevanten Variablen auch zwischen den verschiedenen Visualisierungstechniken.

Außerdem wurde der vorhandene Datensatz aus zwei verschiedenen Gründen gekürzt, zum einen ist die Darstellung durch die Größe des Datensatzes erheblich verlangsamt worden und zum anderen enthält der Datensatz Fussballer für die nicht alle Werte enthalten sind.

¹https://github.com/JohannesLange/Visualisierung_FIFA19/tree/master

Für die Darstellung des Baumdiagramms wurde sich auf die fünf größten europäischen Fussballligen konzentriert. Deswegen enthält der Datensatz für diese Visualisierungstechnik nur Daten von Spielern, deren Team Teil einer dieser Ligen ist.

3 Visualisierungen

3.1 Analyse der Anwendungsaufgaben

Wie hilft Scatterplot/Parallele Koordinaten/Baumdiagramm die genannten Problemstellungen zu beantworten?

3.2 Anforderungen an die Visualisierungen

Wie muss die Visualisierung designed werden um das Zielproblem gut beantworten zu können?

3.3 Präsentation der Visualisierungen

Visualisierungstechniken vorstellen, Interaktivität zeigen, Designentscheidungen begründen(Erfüllen diese die Anforderungen?), Diskutieren wieso nicht andere Techniken verwendet wurden(Expressivität und Effektivität).

3.3.1 Visualisierung Eins

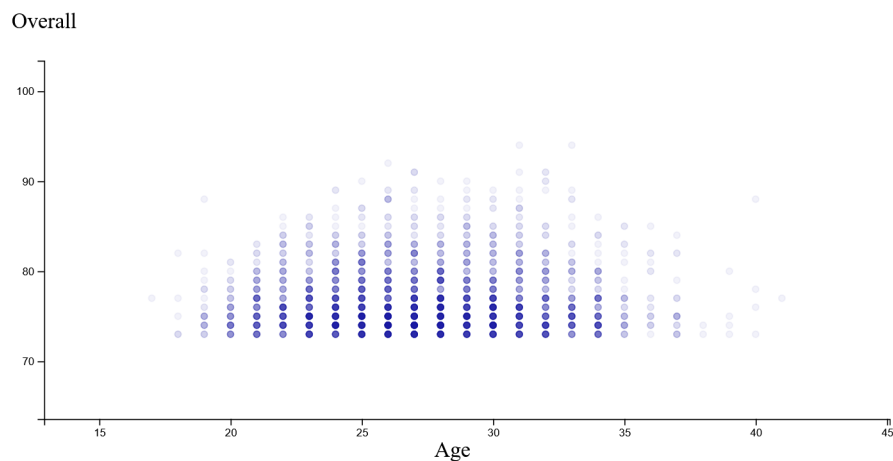


Abbildung 1: Darstellung des Scatterplots
Quelle: eigene Darstellung

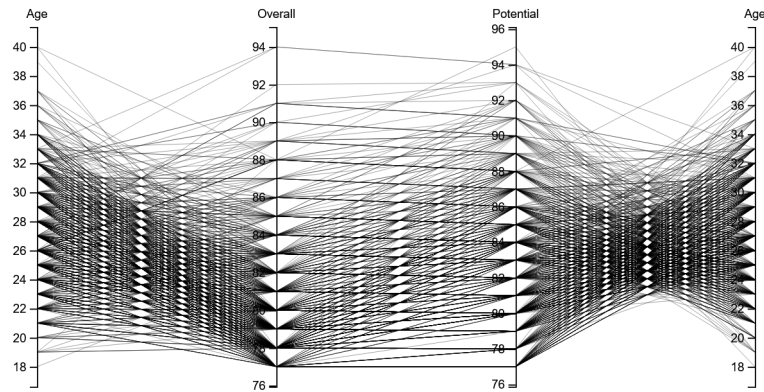


Abbildung 2: Darstellung der Parallelen Koordinaten
Quelle: Eigene Darstellung

3.3.2 Visualisierung Zwei

3.3.3 Visualisierung Drei

3.4 Interaktion

Interaktionen in den Visualisierungen(möglicherweise Interaktion zwischen den Techniken), Warum genau diese Techniken, welche Zwecke erfüllen sie für die Anwender, Warum wurden andere nicht umgesetzt

4 Implementierung

Wie ist der Quellcode gegliedert, was lies sich aus den Übungen übernehmen, Wie sieht die Datenstruktur des Modells aus -> in dem verschiedene Zustände der Interaktion gespeichert wurden (Success record)

5 Anwendungsfälle

Spezifischen Anwendungsfall für Nutzergruppen vorstellen, der an Hand der Visualisierungstechniken visuell erkennbar ist, wäre dies auch mit anderen Techniken möglich gewesen? Aufwand mit anderen Techniken vergleichen

Anhang: Git-Historie

Literatur

- [1] *Min-Maxing*. de-DE. Feb. 2014. URL: <https://mein-mmo.de/lexikon/min-maxing/>
(besucht am 04.09.2021).