



Aalto-yliopisto
Perustieteiden
korkeakoulu

On cluster structures of NHL players (Valmiin työn esittely)

Elmeri Lähevirta

20.02.2018

Ohjaaja: TkT *Lauri Viitasaari*

Valvoja: Prof. *Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta

- NHL-pelaajien tilastoja on aloitettu keräämään tarkemmin ja tilastanalytiikkaan on alettu panostaa enemmän 2010-luvulla.
 - NHL lisäsi vuonna 2015 omille sivuilleen kaikkien saataville 45 kehittyneempää ja monipuolisempaa joukkue- ja pelaajatilastoa.
- Elokuva Moneyball (2011) mainittu monessa lähteessä urheilun tilastanalytiikan vallankumouksen käännekohtana.

Tavoitteet

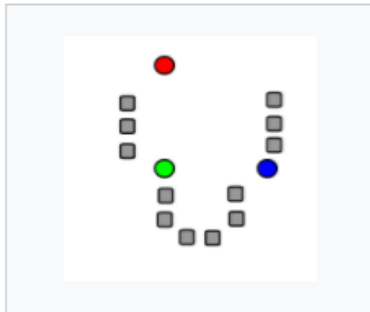
- Tehdä klusterointianalyysi k-means-algoritmilla NHL:n runkosarjan 2016-17 pelaajatilastoista.
 - Tutkia, miten hyvin k-means-algoritmi pystyy erottamaan puolustajat hyökkääjistä ja pystyykö se mahdollisesti erottamaan jopa eri pelaajatyyppejä toisistaan.

Rajaukset

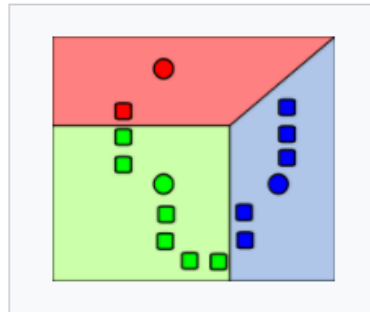
- Maalivahdit jätettiin klusterointianalyysin ulkopuolelle ja analyysissä keskityttiin vain kenttäpelaajiin (hyökkääjät ja puolustajat).

Menetelmät (1/2)

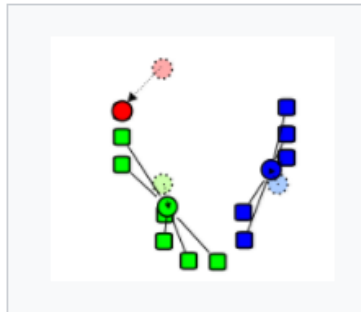
- K-means-algoritmi (kuva Wikipediasta "K-means clustering")



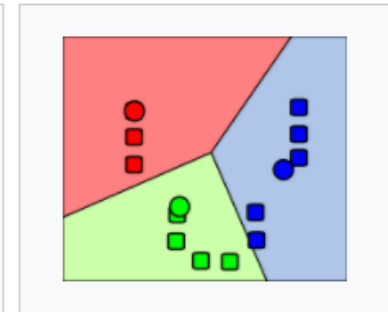
1. k initial "means" (in this case $k=3$) are randomly generated within the data domain (shown in color).



2. k clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.



3. The centroid of each of the k clusters becomes the new mean.



4. Steps 2 and 3 are repeated until convergence has been reached.

– K-means++-alustus

Menetelmät (2/2)

Työvaiheet

1. NHL:n runkosarjan 2016-17 pelaajatilastoiden hankinta ja niihin tutustuminen.
2. Datan rajaaminen ja muokkaaminen sekä sopivien muuttujien valinta.
3. Klusterointimenetelmien teoriaan perehtyminen.
4. Klusterointianalyysin toteuttaminen ja tulosten tulkinta.

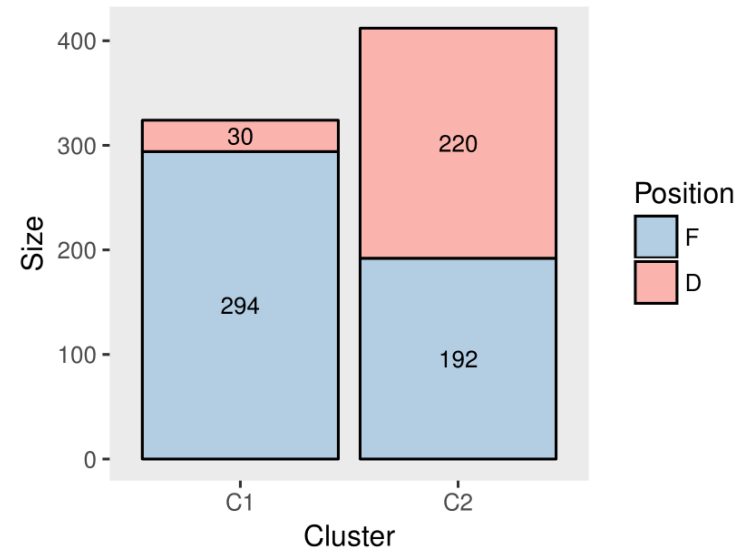
Työkalut

- RStudio
 - Tilastolliseen analyysiin, kuvaajien piirtämiseen ja klusterointianalyysin toteuttamiseen.

Tulokset

$k = 2$

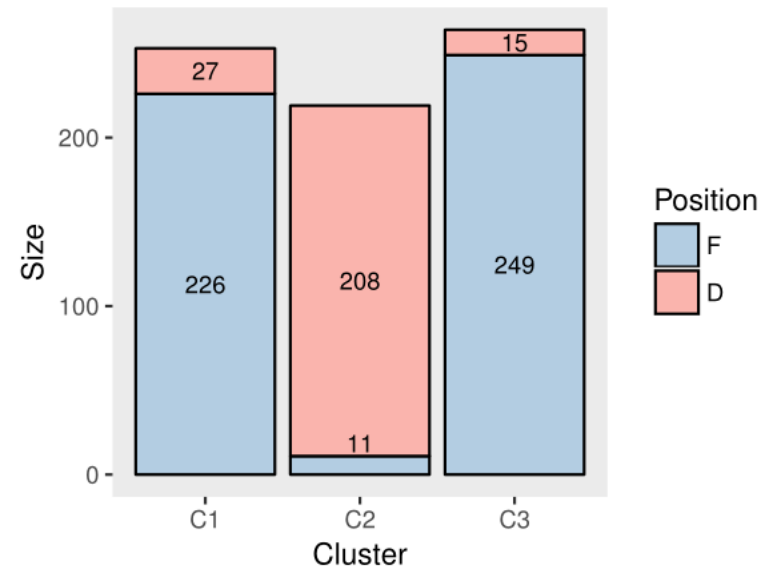
- C1 – hyökkäysorientoituneet pelaajat
- C2 – puolustavat pelaajat



Tulokset

$k = 3$

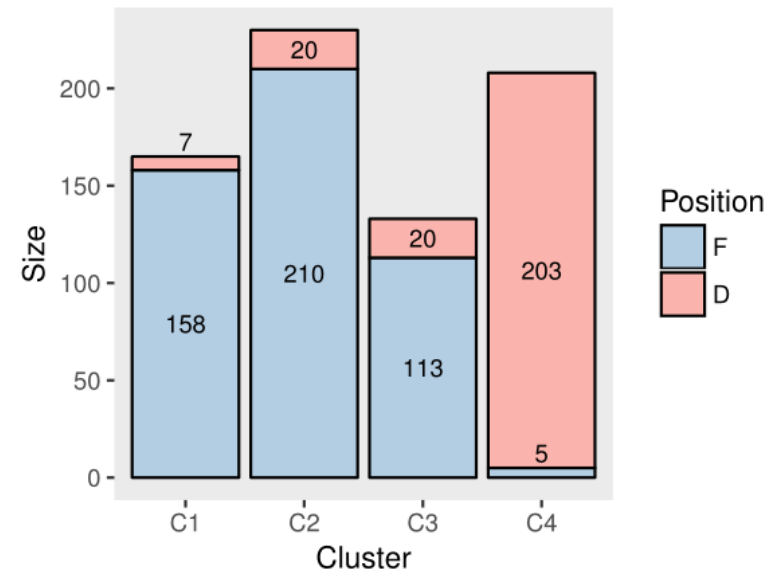
- C1 – voimahyökkääjät
- C2 – puolustajat
- C3 – tehopelaajat



Tulokset

$k = 4$

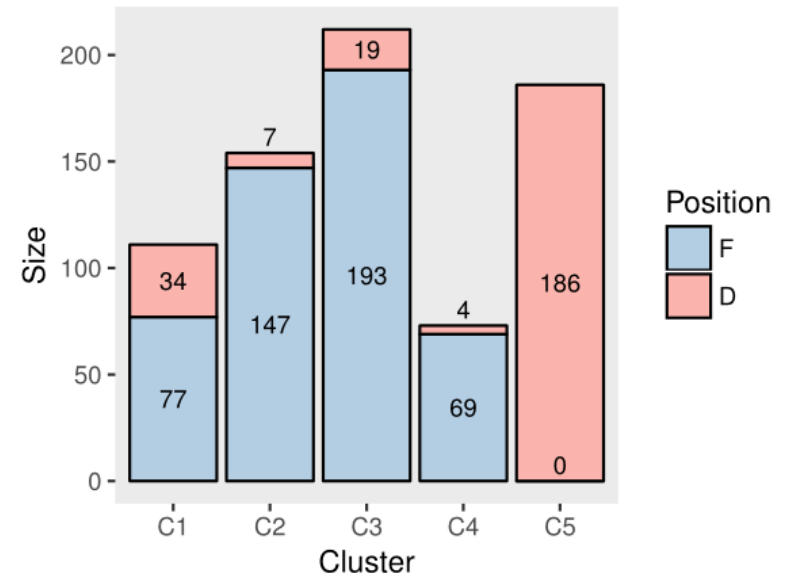
- C1 – tehopelaajat
- C2 – “perus”hyökkääjät
- C3 – voimahyökkääjät
- C4 – puolustajat



Tulokset

$k = 5$

- C1 –
- C2 – tehopelaajat
- C3 – “perus”hyökkääjät
- C4 – voimahyökkääjät
- C5 – puolustajat



Yhteenveto

- K-means-klusterointialgoritmi pystyi erottamaan puolustajat hyökkääjistä eri ryhmiinsä.
- Tämän lisäksi klusteroinnin tuloksena syntyneille ryhmille pystyttiin antamaan pelaajatyypin perustuvia nimityksiä.
- Jatkotutkimuksen aihe k-means-algoritmin soveltuvuus maalivahtien erottelussa.

Tietolähteet ja aineistot

Tietolähteet

- Maimon, Oded Z., and Lior Rokach, *Data Mining and Knowledge Discovery Handbook. 2nd ed.*, New York; London: Springer, 2010.
- Han J, Kamber M, Pei J., *Data mining (third edition)*. Boston: Morgan Kaufmann, (2012).

Aineistot

- NHL:n runkosarjan 2016-2017 pelaajakohtaiset tilastot