

# Fuzzy кластеровање

Маја Црнкомарковић 21/2017

Марко Бабић 77/2017

14. мај 2021.

Семинарски рад у оквиру курса  
Рачунарска интелигенција



# Садржај

<b>1</b>	<b>Увод</b>	<b>3</b>
1.1	Дефиниција кластеровања . . . . .	3
1.2	Врсте кластеровања . . . . .	3
1.3	Различити типови кластеровања . . . . .	4
1.4	Fuzzy кластеровање . . . . .	4
<b>2</b>	<b>Сегментација слика</b>	<b>4</b>
2.1	Fuzzy c-means . . . . .	5
2.2	Наша имплементација . . . . .	6
<b>3</b>	<b>Закључак</b>	<b>7</b>
<b>4</b>	<b>Литература</b>	<b>9</b>

## 1 Увод

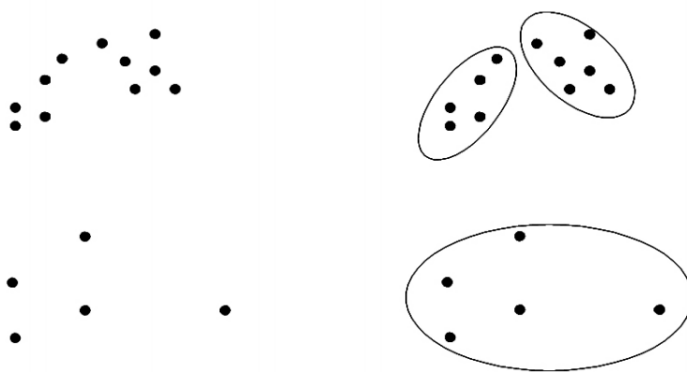
Сегментација слика је један од најпрострањенијих начина за коректно класификовање пиксела у апликацијама заснованим на одлучивању. Сегментација је техника која партиционира слику на униформне и непреклапајуће делове, засновано на некој мери сличности. Ова техника има велики број примена у анализи слика, медицинском процесирању слика, географском информационом систему, и др. Последњих година, исказано је велико интересовање за анализу слика, и с временом се ова област све више развија. Наш рад се бави проблемом сегментације слика коришћењем технике fuzzy кластеровања. Највише пажње ће бити посвећено проблему детекције тумора на мозгу сегментовањем рендгенских снимака.

### 1.1 Дефиниција кластеровања

Не постоји формална дефиниција кластеровања. **Кластер анализа** је проналажење група објеката таквих да су објекти у једној групи међусобно сличнији у односу на објекте у различитим групама. **Кластеровање** се односи на поступак издвајања кластера. Проблем кластеровања се може дефинисати на следећи начин: Дат је жељени број кластера  $K$ , скуп података од  $N$  тачака и функција за мерење растојања. Потребно је пронаћи партиције скупа података тако да се минимизује вредност функције за мерење.

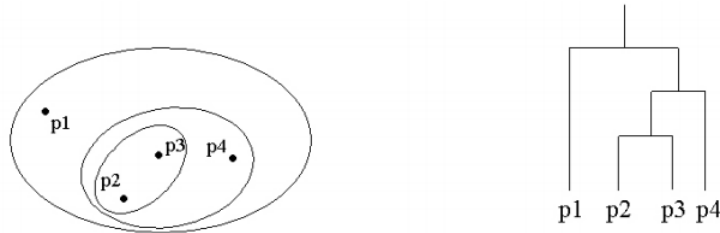
### 1.2 Врсте кластеровања

- **Партиционо кластеровање:** Подела скупа података у непреклапајуће подскупове (кластере) такве да је сваки податак тачно у једном подскупу.



Слика 1: Лево - почетни подаци. Десно - Партиционо кластеровање.

- **Хијерархијско кластеровање:** Скуп угњеждених кластера организован у облику хијерархијског стабла.



Слика 2: Лево - хијерархијско кластеровање. Десно - денгограм.

### 1.3 Различити типови кластеровања

- Ексклузивно/неексклузивно кластеровање
- Fuzzy (расплинуто/нерасплинуто) кластеровање
- Делимично/комплентно кластеровање
- Хетерогено/хомогено кластеровање

### 1.4 Fuzzy кластеровање

Код класичног кластеровања, подаци су подељени у одређен број дисјунктних кластера, где један елемент може припадати само једном кластеру. У fuzzy кластеровању тачка може припадати већем броју кластера са неком тежином између 0 и 1. Збир свих тежина је једнак 1. Сличне карактеристике има вероватносно кластеровање. Fuzzy кластеровање се још назива и расплинуто кластеровање.

## 2 Сегментација слика

Постоје многобројне методе и разноврсна литература за издвајање информација са слике и њену поделу на различите регионе. Свака од тих метода сусреће се са одређеним ограничењима која се огледају у временској сложености или тачности. Разлог за то је што не постоје јасне границе између објеката на слици. Fuzzy кластеровање показало се као веома добар начин за превазилажење овог проблема.

Сегментација слика коришћењем fuzzy кластеровања била је предмет великог интересовања кроз године. Неки од алгоритама који се баве овом темом су Fuzzy C-Means (FCM), Gustafson-Kessel (GK), Gaussian Mixture Decomposition (GMD), Fuzzy C-Varieties (FCV), Adaptive Fuzzy-C varieties

(AFC), Fuzzy C-Shell (FCS), Fuzzy C-Spherical Shells (FCSS), Fuzzy C-Rings, Fuzzy C-Quadric Shells (FCQS), Fuzzy C-Rectangular Shells (FCRS) и други. Међу свим горе наведеним алгоритмима, метод FCM је најприхваћенији начин за сегментацију слика јер омогућава мањи губитак информација у односу на алгоритме класичног кластеровања.

## 2.1 Fuzzy c-means

Fuzzy c-means је први пут представио Dunn а потом га је модификовао Bezdek.

Следи детаљан опис алгоритма:

- Улазни параметри: Подаци које желимо да кластерујемо (скуп елемената  $x$  димензије  $n$ ) и број који означава колико кластера желимо да добијемо ( $k$ ).
- Излазни параметри: Матрица припадности кластерима (димензије  $n \times k$ ) и матрица центроида кластера (димензије  $k \times d$  где је  $d$  димензија сваког појединачног елемента из скупа  $x$ ).

Кораци алгоритма:

1. Насумично додељујемо вредности за све тежине у ознаци:  $w_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq k$  уз услове:
  - (a)  $\sum_{j=1}^k w_{ij} = 1$ ,  $\forall i \in 1, 2, \dots, n$
  - (б)  $0 < \sum_{i=1}^n w_{ij} < n$ ,  $\forall j \in 1, 2, \dots, k$
2. Рачунање центроида за све кластере у ознаци  $c_j$  помоћу формуле:

$$c_j = \frac{\sum_{i=1}^n w_{ij}^p x_i}{\sum_{i=1}^n w_{ij}^p} \quad (1)$$

Напомена: ако је  $p = 0$ , имаћемо понашање класичног k-means алгоритма.

3. Ажурирање вредности матрице припадности користећи формулу:

$$w_{ij} = \frac{\frac{1}{\text{dist}(x_i, c_j)^{\frac{1}{p-1}}}}{\sum_{q=1}^k \frac{1}{\text{dist}(x_i, c_q)^{\frac{1}{p-1}}}} \quad (2)$$

4. Понављати кораке 2. и 3. док центроиди не остану исти у две итерације за редом.

Мера квалитета кластеровања је сума квадратне грешке (енг. Sum of Squared Error):

$$SSE = \sum_{j=1}^k \sum_{i=1}^n w_{ij}^p \text{dist}(x_i, c_j), \quad p \in 1, \dots, \infty \quad (3)$$

## 2.2 Наша имплементација

Решење проблема сегментације слика уз помоћ алгоритма fuzzy c-means смо имплементирали у програмском језику Python. Алгоритам је имплементиран у функцији `fuzzy_c_means`. Она као аргументе прима:

- *data* - низ података (то је уств. матрица која је димензије  $n \times d$ );
- *n* - цео број који означава број података које желимо да кластерујемо;
- *k* - цео број који означава број кластера;
- *d* - цео број који означава димензију појединачног податка из скупа података које желимо да кластерујемо;
- *p* - цео број који означава парамеар за fuzzy формулу којом одредјујемо вероватноћу припадности неког податка за сваки од кластера;
- *max\_iter* - цео број који означава максимални број итерација због безбедности,

док као повратну вредност враћа:

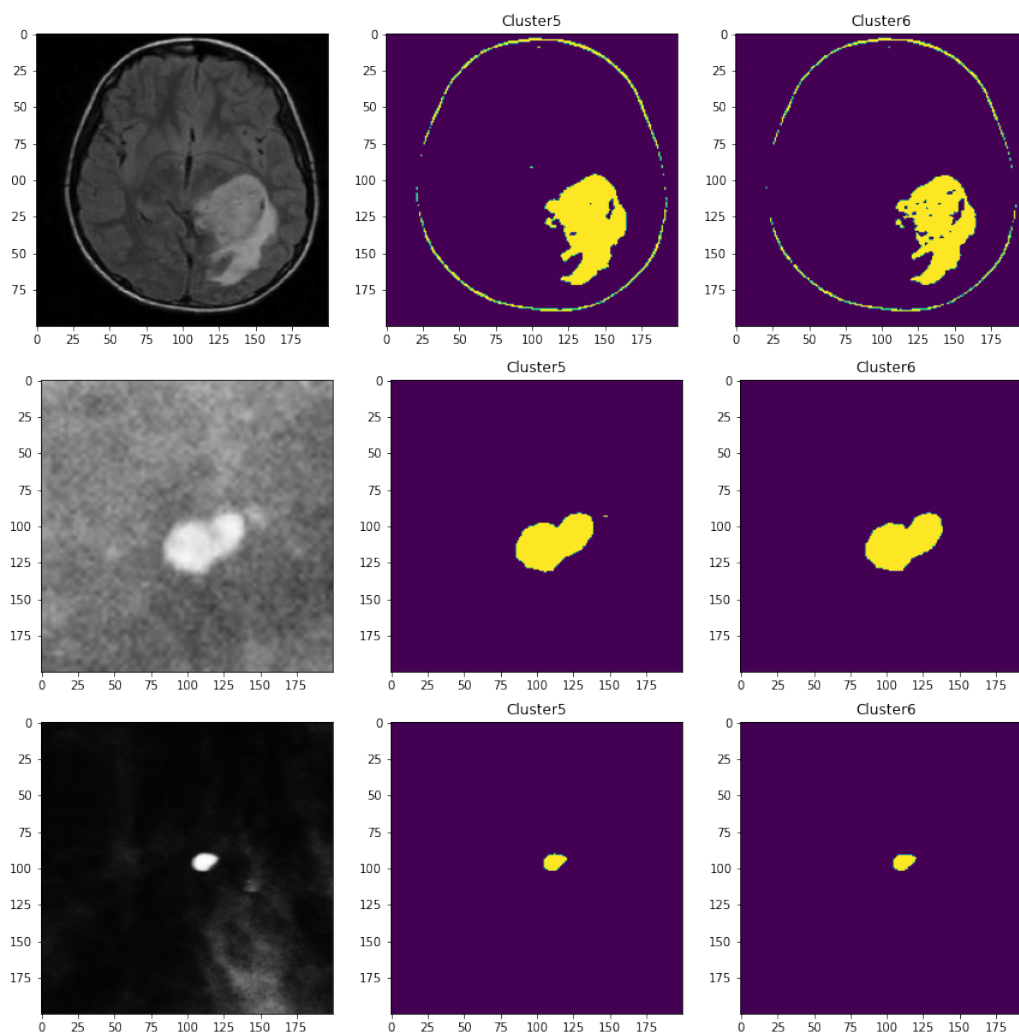
- матрицу  $n \times k$  у којој ће елемент на позицији бити  $w_{ij}$ , тј. тежина са којом *i*-ти елемент припада *j*-том кластеру;
- матрицу  $k \times d$  у којој ће се чувати центроиди свих кластера.

Дакле, како је једна слика представљена као матрица пиксела где је сваки пиксел димензије 3, њу парсирамо у низ како бисмо је могли, као аргумент, проследити нашој функцији. Након што наша функција као повратну вредност врати матрицу припадности података (у нашем случају пиксела димензије 3) кластерима сваком од пиксела додељујемо вредност центроида кластера за који је вредност највећа у матрици припадности. Затим низ пиксела враћамо у матрицу полазних димензија и посматрамо је као слику. У нашем случају врло је битно да се јасно сегментује одређени детаљ на слици и због тога бинаризујемо слику, односно представљамо је са само две боје.

Python библиотеке коришћене у нашем решењу су:

- `numpy`
- `matplotlib.pyplot`
- `os`
- `cv2`
- `time`
- `math`

Резултати нашег алгоритма на неким рендгенским снимцима тумора:



### 3 Закључак

Кластеровање је један од најкоришћенијих начина за сегментацију слика. Огромна предност fuzzy кластеровања у односу на остале технике сегментације слика је то што је врло отпоран на неправилне границе. Током рада користили смо више различитих слика које смо сегментовали уз испробавање различитих вредности одређених параметара и различитог броја кластера и у већини случајева најбоље се показало коришћење 5 кластера уз параметар fuzzy формуле ( $p$ ) који је једнак 2. Сегментација слика уз помоћ алгоритама fuzzy кластеровања је своју примену између осталог нашла у медицини где је веома битна прецизност резултата. Наравно, fuzzy кла-

стеровање има широк спектар примене поред сегментације слика у разним другим областима неке од њих су економија, географија и друге.



## 4 Литература

- M.S. YANG - A Survey of Fuzzy Clustering, October 1993.
- XL Xie, G Beni - A validity measure for fuzzy clustering, 1991.
- Donald E. Gustafson, William C. Kessel - Fuzzy clustering with a fuzzy covariance matrix, 1979.
- Мирјана Маљковић - Скалабилни кластер алгоритми, 2008.
- Ненад Митић - Предавање о кластер анализи на Математичком факултету, 2020.