# Andjela Todorović

# November 24, 2018

Segmentacija tkiva korišćenjem MCL algoritma

Klasterovanje (clustering) je tehnika istrazivanja podataka koja otkriva objekte (koji se opisuju odredjenim atributima) i deli ih u grupe (klastere) cineci ih preglednijim i korisnijim.

Ljudsko oko lako može prepoznati prirodne klastere, ali je ovo i više nego zahtevan zadatak za računar. Postoji dosta razlicitih algoritama za klasterovanje (k-means, GMM..) svaki sa razlicitom slozenoscu i razlicitom primenom u zavisnosti od prirode problema.

Problem koji ovde razmatramo je segmentacija respiratornog tkiva. Ulazni set podataka se sastoji od 800 slika tkiva plucne maramice. Zanima nas kako na zadatom setu podataka možemo odvojiti jezgro, citoplazmu i ostale celijske organele u klastere, da bismo dalje posmatrali promene koje se odvijaju na ovim organelama. Ukoliko prevedemo ovaj problem u formu koja je čitljiva računaru, cilj nam je da obojimo svaku od slika tako da njene komponente budu obojene razlicitim bojama, npr. citoplazma zutom, jezgro i organele braon, a pozadina svetlo plavom bojom.

Svaku sliku možemo predstaviti kao funkciju (vektorski) sa odredjenim ponasanjem ili kao graf (rasterski). Konkretno, procesuiranje nad slikama predstavljenim kao grafovima funkcionise na principu razlike u pikselima, odnosno svaki piksel (zadat npr pomocu RGB skale kao vektor ili tacka u trodimenzionalnom prostoru) predstavljamo cvorom grafa, a kazemo da izmedju dva cvora u grafu postoji ivica ako su oni na rastojanju manjem od unapred zadanog ρ . Ovo rastojanje se naziva Euklidsko rastojanje razlike.

Kod CCL (conected component labeling) algoritma, pikseli se grupisu u komponente na osnovu slicnosti u njihovim osobinama (boji). Svi pikseli u povezanoj komponenti dele slican intenzitet i na neki nacin su medjusobno povezani. Jednom kada se svi pikseli povezu, svakoj se komponenti pridruzi odredjen graylevel ili boja koja je predvidjena za tu grupu piksela.

Ovaj algoritam radi na principu 4-connectivity ili 8-connectivity, odnosno poredi svaki piksel sa susednih 4 (gore, dole, levo, desno) ili susednih 8 piksela. U praksi, odnosno na nasem datasetu, CCL se nije pokazao kao adekvatan za resavanje problema, jer 4-connectivity nije davao precizne rezultate a 8-connectivity je imao preveliki variance.

Zato se, umesto da se posmatra rastojanje izmedju cvorova za ispitivanje egzistencije ivice u grafu, pretpostavi unapred da su svi cvorovi u grafu povezani na neki nacin (nema izolovanih cvorova). U ovom slucaju nije moguce primeniti CCL, te se koristi MCL algoritam (Markov cluster algorithm).

Markov cluster algorithm krece od grafa sa ogromnim brojem cvorova sa ciljem da se taj broj cvorova smanji. Funkcionise na principu matematickog bootstrapovanja, odnosno od postojeceg grafa kreira novi iskljucivo koriscenjem informacija iz ulaznog seta podataka.

Ovaj algoritam deterministicki izracunava verovatnoce proizvoljne setnje (random walk) kroz graf, koristeci dva operatora (ekspanzije i inflacije). Ekspanzija se matematicki modeluje kao normalan proizvod stohastickih matrica, a u praksi zeli da pocevsi od prvog cvora prosiri skup cvorova koji posmatramo. Inflacija je matematicki modelovana kao Hadamardov proizvod, odnosno gde se svaki element mnozi sa njemu odgovarajucem u drugoj matrici (obe matrice moraju biti kvadratne i istih dimenzija), a ovo se radi sa ciljem da se odrede verovatnoce da se iz jednog cvora grafa (jednog piksela) stigne u drugi. Svakoj grani je dodeljena verovatnoca da se iz jednog cvora grane stigne u drugi. Uzimaju se samo verovatnoce od najveceg znacaja a postojece grane sa malim verovatnocama se brisu. Na ovaj nacin se ispostavlja da se graf jako precizno podeli u komponente nakon odredjenog broja koraka, a u praksi ovo obuhvati sve piksele slicnih boja, odnosno sa malom razlikom u vrednosti.