Primjena Blumovog filtera u algoritmima detekcije noviteta

Damir Delijic 1/21 C

Contents

[Definicija problema 3](#_Toc138071562)

[Predstavljanje tekstualnih dokumenata 3](#_Toc138071563)

# Definicija problema

Zadatak se sastoji u dizajniranju i implementaciji algoritma za detekciju noviteta u skupu tekstualnih dokumenata. Algoritam treba da bude zasnovan na odgovarajucoj modifikaciji ideje na kojoj je zasnovan Blumovog filtera. Detekcija noviteta predstavlja sposobnost prepoznavanja novih obzervacija, odnosno onih koje se u odredjenoj mjeri ne uklapaju u poznatu distribuciju. Smatra se polu-nadgledanom metodom detekcije anomalija. Trening podaci se sastoje iz iskljucivo cistih obzervacija koje ne sadrze anomalije.

# Predstavljanje tekstualnih dokumenata

Dokumenti su predstavljeni pomocu karakteristicne matrice. Kolone matrice odgovaraju dokmentima, a redovi odgovaraju šinglovima koji su izvuceni iz poznatog skupa dokumenata. U -tom redu i -toj koloni stoji jedinica ako je taj šinlg prisutan u tom dokumentu, u suprotnom se u toj celiji nalazi nula.

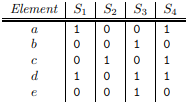


Figure : Matricna reprezentacija skupova