Meta stabilna stanja Markovljevih lanaca

Tvrtko Glunčić, Pavao Turić, Matej Vojvodić

Studeni 2025

Sadržaj

1 Opis problema, motivacija

Brojni složeni dinamički sustavi mogu se matematički opisati pomoću homogenih Markovljevih lanaca s konačnim skupom stanja. Pojedina stanja Markovljevog lanca predstavljaju konfiguracije sustava, dok prijelazne vjerojatnosti opisuju vremensko ponašanje dinamičkog sustava. U takvim se sustavima često pojavljuje fenomen metastabilnosti – situacija u kojoj se sustav dugo zadržava unutar određenih skupova stanja, dok su prijelazi između tih skupova rijetki. Identifikacija metastabilnih stanja od iznimne je važnosti jer omogućuje bolje razumijevanje dugoročne dinamike sustava, pojednostavljenje modela te učinkovitiju analizu i predikciju ponašanja.

Povijesno gledano, problem je posebno istraživan u području dinamike molekularnih konformacija, gdje metastabilna stanja odgovaraju različitim stabilnim oblicima molekule što je iznimno važno u proizvodnji lijekova. No isti matematički problem pojavljuje se i u mnogim drugim područjima: u analizi društvenih i komunikacijskih mreža metastabilna stanja predstavljaju grupe korisnika koje često komuniciraju, a algoritam koji ćemo mi implementirati u takvoj primjeni detektira zajdnice bez prethodnog znanja o njihovom broju i veličini samo iz velikog broja raznih oblika interakcija i komuniciranja. Ako promatramo diskretni graf kojim se toplina širi, metastabilna stanja su ekvivalent gotovo toplinski izoliranih područja u tom grafu, ako ona postoje naravno, dok se prijelazna matrica interpretira kao operator raspodjele topline na grafu čijom uzastopnom primjenom na čvorove dobijemo kako se toplina dugoročno raspodjeli. U neuroznanosti se koristi za dijagnoze jer se naša moždana aktivnost može izjednačiti sa skokovima između raznih stanja poput stanja mirovanja, mentalne aktivnosti i slično. To može biti korisno kod analize epileptičkih obrazaca i reakcija na podražaje što je važno za dijagnostiku. Još jedna važna primjena je analiza ponašanja kupaca, metastabilna stanja su uobičajene faze koje kupac prolazi: informiranje, razmatranje kupnje, kupnja, neaktivnost pa dugotrajnim promatranjem i analizom podataka trgovine mogu optimizirati vrijeme i iznos popusta kako bi zaradile što više novaca. Također, stabilni klimatski režimi su još jedna prirodna pojava koja se može modelirati metastabilnim stanjima Markovljevog lanca.

2 Uvod

Formalno, problem se svodi na pronalaženje takve permutacije stanja Markovljevog lanca kojom se njegova prijelazna matrica može transformirati u oblik gotovo blok-dijagonalne matrice. Dijagonalni blokovi predstavljaju metastabilna stanja obilježena čestim unutarnjim prijelazima, dok su elementi koji se ne nalaze unutar blok-dijagonalne strukture mali i odgovaraju rijetkim prijelazima između različitih skupova stanja.

Klasični pristupi rješavanju ovog problema temelje se na analizi svojstvenih vrijednosti i svojstvenih vektora prijelazne matrice, odnosno na pronalaženju tzv. Perronovog klastera – skupa svojstvenih vrijednosti bliskih jedinici. Međutim, takve metode nisu numerički stabilne, osobito u slučajevima kada ne postoji jasna razlika među svojstvenim vrijednostima ili kada je izračun stacionarne distribucije osjetljiv na numeričke pogreške što je česta situacija. Dodatno, ove metode često zahtijevaju pretpostavku reverzibilnosti Markovljevog lanca što je velika pretpostavka koja ne mora vrijediti uvijek u stvarnom svijetu i izračun više svojstvenih parova, što povećava složenost algoritma.

Zbog navedenih ograničenja razvijen je alternativni pristup temeljen na dekompoziciji po singularnim vrijednostima. Ova metoda koristi činjenicu da je dovoljno izračunati samo singularni vektor koji odgovara drugoj najvećoj singularnoj vrijednosti prijelazne matrice kako bi se razdvojila metastabilna stanja. Analizom predznaka tog vektora moguće je razvrstati stanja u odvojene skupine i rekurzivno provoditi postupak dok se ne otkriju sve metastabilne cjeline. Prednost ovoga pristupa jest u tome što ne zahtijeva poznavanje stacionarne distribucije, ne oslanja se na jasan razmak u spektru svojstvenih vrijednosti te omogućuje znatnu uštedu u računalnom vremenu jer se u svakom koraku izračunavaju samo dvije najveće singularne vrijednosti.

Cilj ovog seminara je, uz male preinake, implementirati SVD-algoritam opisan u [1] koji identificira metastabilna stanja Markovljevih lanaca te ispitati njegovu učinkovitost na odabranim primjerima.

3 Opis metode

U prvom koraku naše metode primjenjuje se dekompozicija na singularne vrijednosti (SVD) na matricu prijelaza P. Promatra se drugi lijevi singularni vektor u_2, koji odgovara drugoj najvećoj singularnoj vrijednosti. Njegove komponente služe za inicijalnu biparticiju skupa stanja: stanja s pozitivnom komponentom $u_2(i)$ dodjeljuju se prvom klasteru, dok stanja s negativnom komponentom pripadaju drugom klasteru. Stanja s komponentom jednakom nuli mogu se proizvoljno pridružiti jednom od klastera bez značajnog utjecaja na rezultat.

Ova spektralna biparticija temelji se na ideji srodnoj Fiedlerovom vektoru iz teorije grafova, gdje drugi svojstveni vektor Laplaceove matrice reflektira prirodno particioniranje grafa. Intuitivno, SVD identificira dominantne smjerove varijacije matrice prijelaza, dok u_2 otkriva podjelu stanja na dvije slabo povezane skupine koje predstavljaju potencijalne metastabilne klastere.

3.1 Heuristika gustoće prijelaza

Dobivena podjela ocjenjuje se heuristikom gustoće prijelaza. Za podjelu stanja na dva klastera C_1 i C_2, defniraju se:

- Prosječna unutar-klasterska povezanost

$$D_{\text{in}} = \frac{\sum_{i,j\in C_1} P_{ij} + \sum_{i,j\in C_2} P_{ij}}{|C_1| + |C_2|}$$

- Prosječna međuklasterska povezanost

$$D_{\text{out}} = \frac{\sum_{i\in C_1,\, j\in C_2} P_{ij} + \sum_{i\in C_2,\, j\in C_1} P_{ij}}{|C_1| + |C_2|}$$

Na temelju toga definira se omjer

$$r = \frac{D_{\text{out}}}{D_{\text{in}}}.$$

Ako je omjer manji od praga δ (tj. $r < \delta$), klasteri se smatraju metastabilno odvojenima. U suprotnom, podjela se odbacuje jer ne predstavlja jasnu metastabilnu strukturu. Parametar δ predstavlja razinu tolerancije: manja vrijednost zahtijeva izraženije slabe međuklasterske veze kako bi se podjela prihvatila.

3.2 Rekurzivna biparticija

U slučaju uspješne biparticije, matrica se permutira tako da stanja klastera C_1 i C_2 postanu uzastopna u indeksiranju. Isti postupak zatim se rekurzivno primjenjuje na svaki klaster zasebno. Svaki klaster se ponovno dijeli primjenom SVD-a i heuristike gustoće sve dok daljnja podjela ne bude moguća bez narušavanja kriterija metastabilnosti ili dok se ne dosegne minimalna dopuštena veličina klastera.

Proces generira hijerarhijsko stablo podjela, pri čemu konačan skup klastera ne ovisi o redoslijedu grananja. Rezultat je razdioba skupa stanja na disjunktne metastabilne klastere s jakim unutar-klasterskim tranzicijama i slabim međuklasterskim prijelazima.

3.3 Rafiniranje metodom *k*-sredina

Nakon inicijalnog određivanja klastera provodi se rafiniranje pomoću algoritma k-sredina (k-means). Broj klastera K jednak je broju dobivenom rekurzivnom biparticijom. Svaki klaster C_k predstavlja se centroidom

$$\mu_k = \frac{1}{|C_k|} \sum_{i \in C_k} P_{i:},$$

gdje je $P_{i:}$ redak matrice prijelaza koji opisuje distribuciju prijelaza iz stanja i.

Svako stanje i ponovno se pridružuje klasteru čiji je centroid najbliži u Euklidskoj metriki:

$$d(i, k) = \| P_{i:} - \mu_k \|_2.$$

Nakon prerazdiobe stanja izračunavaju se novi centroidi te se postupak ponavlja sve dok ne dođe do stabilizacije članstva klastera.

3.4 Određivanje redoslijeda klastera

Za vizualno uočljivu blokovnu strukturu matrice potrebno je odrediti optimalan redoslijed klastera. Svaki klaster predstavlja se centroidom μ_k, dok se udaljenost između klastera k i l definira kao

$$d(k, l) = \| \mu_k - \mu_l \|_2.$$

Problem određivanja poretka reducira se na problem putujućeg trgovca (TSP). Zbog računske složenosti koristi se pohlepna heuristika najbližeg susjeda: započinje se proizvoljnim klasterom, a svaki sljedeći bira se kao najbliži neobiđeni klaster prema definiranoj metriji udaljenosti.

3.5 Konačna permutacija matrice

Na temelju dobivenog poretka konstruira se permutacijski vektor π kojim se indeksiranje stanja preuređuje. Permutacijom redaka i stupaca matrice prijelaza

$$\tilde{P} = \Pi P \Pi^T,$$

gdje je Π pripadna permutacijska matrica, dobiva se nova matrica s izraženom blokovnom strukturom uz glavnu dijagonalu. Blokovi odgovaraju metastabilnim klasterima, dok su izvanblokovski elementi znatno manji.

4 Evaluacija

Na kraju se provodi analiza pomoću standardnih metrika kvalitete klasteriranja te se uspoređuju:

- rezultati dobiveni isključivo SVD biparticijom,
- rezultati nakon dodatnog rafiniranja metodom k-sredina,
- utjecaj različitih vrijednosti praga δ na konačnu strukturu klasteriranja.

U svrhu kvantitativne procjene kvalitete dobivenog klasteriranja metastabilnih stanja, korištene su dvije metrike: **metastability index** i **omjer gustoća unutar i između klastera**. Ove metrike komplementarno ocjenjuju koliko su identificirani klasteri koherentni iznutra i koliko su međusobno razdvojeni, što su ključne karakteristike metastabilne strukture Markovljevih lanaca.

Metastability index (MSI) definira se kao očekivana vjerojatnost da Markovljev lanac, nakon što uđe u određeni klaster, u njemu ostane barem jedan korak. Neka je $M \in \mathbb{R}^{n \times n}$ matrica prijelaza, $\pi \in \mathbb{R}^n$ stacionarna raspodjela (ili, u našem slučaju, njena aproksimacija iz stupčanih zbrojeva), i neka su stanja podijeljena u K disjunktnih klastera $C_1, C_2, \ldots, C_K$.

Tada je metastability index definiran izrazom:

$$\mathrm{MSI} = \sum_{k=1}^{K} \pi(C_k) \cdot P_{kk}$$

gdje je $\pi(C_k) = \sum_{i \in C_k} \pi_i$ ukupna stacionarna masa klastera C_k, a $P_{kk} = \frac{1}{\pi(C_k)} \sum_{i \in C_k} \sum_{j \in C_k} \pi_i M_{ij}$ vjerojatnost da lanac ostane unutar istog klastera. Visoka vrijednost MSI-a (bliska 1) ukazuje na to da se gotovo svi prijelazi u jednom koraku odvijaju unutar istog klastera, što potvrđuje metastabilnost i ispravnost strukture klastera.

Omjer gustoće između i unutar klastera dodatno mjeri koliko su klasteri odvojeni u smislu prijelazne dinamike. Za svaki klaster C_k definiraju se:

- prosječna unutar-klasterska gustoća:

$$\mathrm{D}_{\mathrm{intra}}^{(k)} = \frac{1}{|C_k|^2} \sum_{i,j \in C_k} M_{ij}$$

- prosječna međuklasterska gustoća:

$$\mathrm{D}_{\mathrm{inter}}^{(k)} = \frac{1}{2|C_k|(n - |C_k|)} \sum_{\substack{i \in C_k \\ j \notin C_k}} (M_{ij} + M_{ji})$$

Zatim se za svaki klaster izračuna omjer $r_k = \mathrm{D}_{\mathrm{inter}}^{(k)} / \mathrm{D}_{\mathrm{intra}}^{(k)}$, a ukupna metrika je njihova srednja vrijednost:

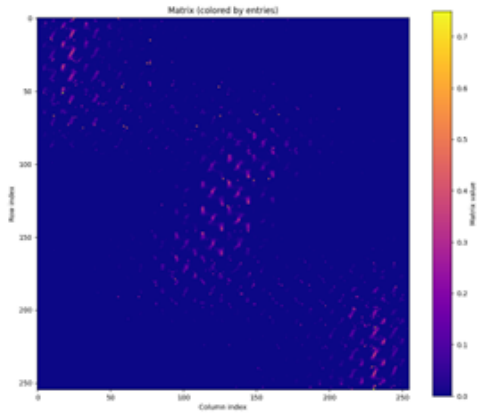
$$\text{Inter/Intra Ratio} = \frac{1}{K} \sum_{k=1}^{K} r_k$$

Što je ovaj omjer manji, to je gustoća prijelaza između klastera manja u odnosu na gustoću unutar klastera, što implicira da je podjela stanja strukturirana i dinamički utemeljena.

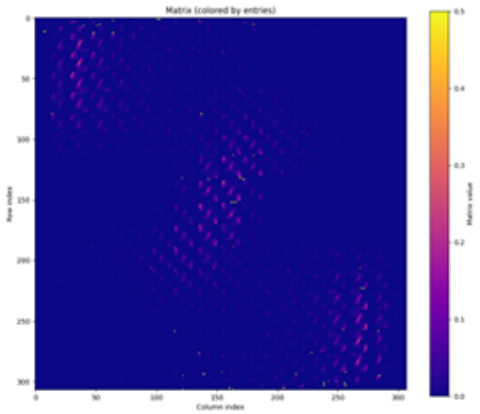
Ove dvije metrike zajedno osiguravaju sveobuhvatan uvid u kvalitetu klasteriranja: dok MSI mjeri stabilnost ponašanja unutar klastera, omjer gustoća testira koliko su ti klasteri međusobno izolirani. Kombiniranom primjenom obje metrike moguće je objektivno procijeniti je li struktura dobivena algoritmom vjerna metastabilnim domenama sustava.

5 Rezultati

Početni podaci koje smo dobili izgledali su ovako:

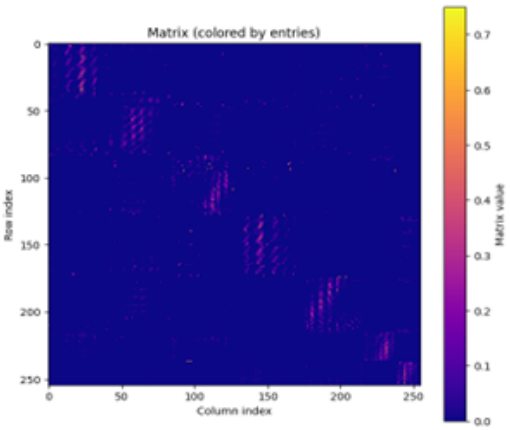


Slika 1: Početno stanje matrice za dataset PH300

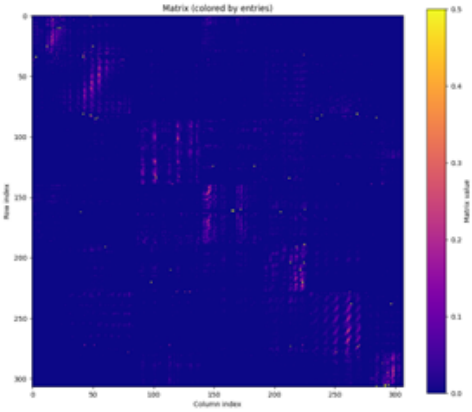


Slika 2: Početno stanje matrice za dataset PH500

Nakon provedenog algoritma, dobili smo sljedeće rezultate:



Slika 3: Konačno stanje matrice za dataset PH300



Slika 4: Konačno stanje matrice za dataset PH500

Za oba dataseta se vidi promjena matrice iz uglavnom neuređene na početku na matricu koja izgleda puno više kao blok dijagonalna matrica.

Dvije gore navede mentrike smo izračunali za vrijednosti δ od 0.5 do 5.5 te nakon svakog koraka postupka; nakon SVD, k-means i na kraju nakon TSP. Rezultati su prikazani u dvije tablice za dva dataseta koje smo koristili. U tablicama SVD označava nakon provedbe samo SVD-a, KMEANS nakon provedbe SVD-a i k-means-a te TSP završni rezultat za dan δ.

Δ	MSI_SVD	Ratio_SVD	MSI_KMeans	Ratio_KMeans
0.5	0.9016	0.0528	0.9169	0.0418
1.5	0.8333	0.0867	0.8613	0.0763
2.5	0.8567	0.0779	0.8619	0.0760
3.5	0.3258	0.3036	0.4477	0.3306
4.5	0.3408	0.2778	0.4379	0.3450
5.5	0.2859	0.3413	0.4238	0.3445

Tablica 1: Za dataset PH300

Δ	MSI_SVD	Ratio_SVD	MSI_KMeans	Ratio_KMeans
0.5	0.6880	0.1385	0.8169	0.0765
1.5	0.6668	0.2124	0.7241	0.1394
2.5	0.4154	0.3236	0.4702	0.2827
3.5	0.4485	0.3024	0.4788	0.2640
4.5	0.3967	0.2996	0.5885	0.2478
5.5	0.3630	0.3222	0.5106	0.3240

Tablica 2: Za dataset PH500

Rezultati pokazuju da se MSI u pravilu povećava nakon k-meansa — što ukazuje da profinjenje u prostoru matrice prijelaza uspješno redistribuira neka stanja u klastere gdje imaju stabilnije ponašanje. Također se smanjuje Inter/Intra omjer, što znači da su klasteri postali i međusobno razdvojeniji.

Parametar δ, koji kontrolira koliko su preklapanja između klastera dopuštena tijekom rekurzivne dekompozicije, ima izrazit utjecaj na stabilnost i odvojenost klastera. Za niže vrijednosti δ (npr. 0.5–1.5), funkcija `ima_smisla` razdvaja samo one granične podklastere koji su jasno odvojeni. Rezultat su veći, stabilniji klasteri sa snažnijom unutarnjom povezanošću i manjim međuklasterskim prijelazima. S porastom δ, algoritam dopušta veću razinu difuznosti i dijeli strukturu na sve finije dijelove, čime raste broj klastera, ali istovremeno dolazi do pada MSI-a i rasta Inter/Intra omjera. To se jasno očituje u rezultatima za oba skupa podataka (PH300 i PH500), gdje se nakon $\delta > 2.5$ značajno pogoršava mjera stabilnosti, a međuklasterski prijelazi postaju sve češći.

Zaključno, rezultati sugeriraju da postoji optimalni raspon vrijednosti parametra δ za koji se dobivaju klasteri koji su dovoljno razlučeni, a istovremeno zadržavaju snažnu unutarnju metastabilnost. Taj raspon može ovisiti o karakteristikama ulazne matrice, no za oba analizirana skupa podataka pokazao se kao stabilan u intervalu $[0.5, 1.5]$.

6 Literatura

Metoda opisana u ovom poglavlju temelji se na pristupu za identifikaciju metastabilnih stanja Markovljevih lanaca temeljenom na dekompoziciji po singularnim vrijednostima (SVD), kako je prikazano u radu:

Literatura

[1] D. Fritzsche, V. Mehrmann, D. Szyld and E. Virnik *AN SVD APPROACH TO IDENTIFYING METASTABLE STATES OF MARKOV CHAINS*, 2008.