

Generiranje naključnih omrežij z metodo strojnega učenja

Mitja Kuščer

Diplomsko delo

Mentor: izr. prof. dr. Marko Bajec



Motivacija

- Želimo generirati naključna omrežja, ki v svojih lastnostih čimbolj posnemajo realna omrežja
- Generiranje naj poteka avtomatsko, brez potrebe po prisotnosti domenskega eksperta
- Pristop naj bo uporaben za vrsto različnih omrežij

Omrežja in njihove lastnosti

■ Socialna omrežja

- Omrežja prijateljev
- Omrežja poslovnih sodelovanj
- Omrežja telefonskih klicev

■ Informacijska omrežja

- Omrežje spletnih strani in povezav med njimi
- Omrežja citiranj med znanstvenimi članki

■ Tehnološka omrežja

- Električno omrežje
- Fizično internetno omrežje

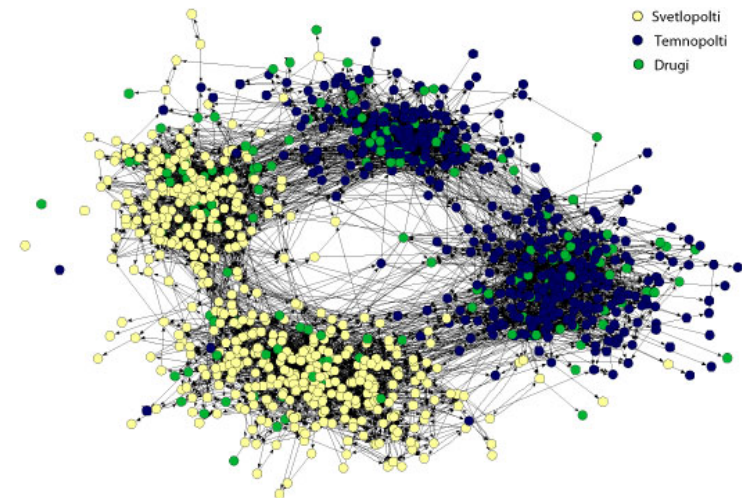
■ Biološka omrežja

- Omrežja prehranjevanja
- Omrežja metaboličnih poti
- Gensko urejevalno omrežje

■ Majhen premer

■ Potenčna porazdelitev stopenj

■ Skupnosti



Modeli naključnih omrežij

■ Poissonov model omrežij

- Majhen premer
- Binomska (Poissonova) porazdelitev stopenj
- Brez skupnosti

■ Model BA (Barabási in Albert)

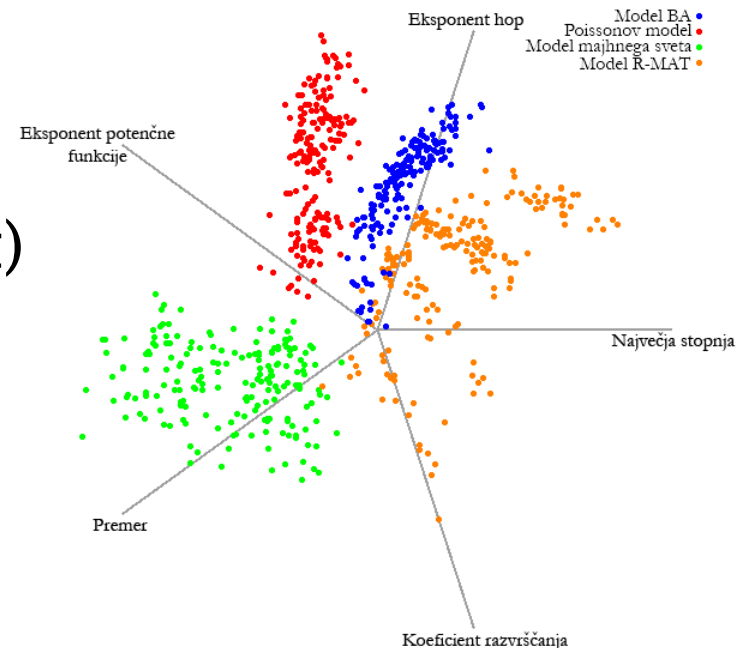
- Majhen premer
- Potenčna porazdelitev stopenj
- Ne generira skupnosti

■ Model majhnega sveta

- Majhen premer
- Eksponentna porazdelitev stopenj (z vrhom v k)
- Generira skupnosti znotraj omrežja

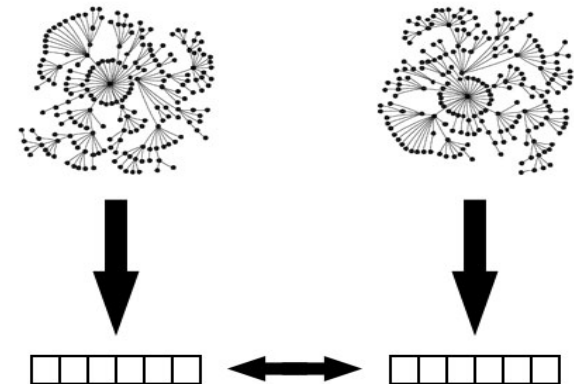
■ Model R-MAT

- Obljublja majhen premer
- Potenčno porazdelitev stopenj
- Skupnosti v omrežju



Omejitve pri strojnem učenju

- Omrežja neobvladljiva kot celota
Preslikamo v n -dimenzionalni vektor lastnosti
- Pomanjkanje podatkov
Težko je zbrati zadostno količino podatkov, iz katerih bi lahko sestavili množico omrežij, ki bi bila uporabna v različnih primerih
- Neoznačena učna množica
Za razpoložljiva realna omrežja je potrebno s pomočjo domenskega eksperta poiskati najprimernejši model naključnih omrežij



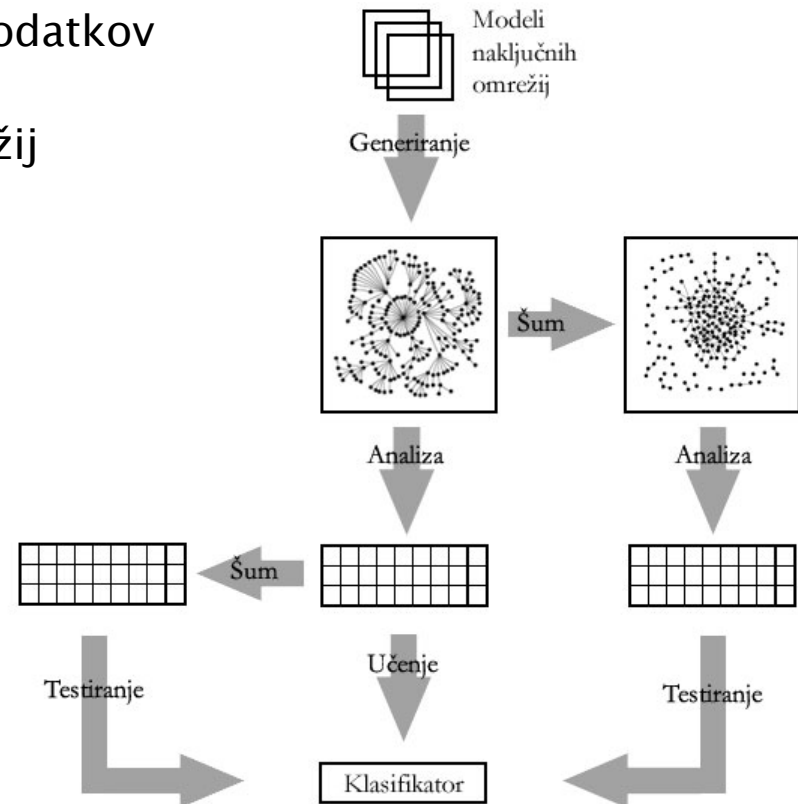
Model M-generator

■ Gradnja in testiranje klasifikatorja

- Klasifikator raje konstruiramo iz podatkov naključno generiranih omrežij
- Izberemo modele naključnih omrežij
- Generiramo čim večjo množico naključnih omrežij
- Analiziramo generirana omrežja
- Konstruiramo klasifikator

■ Prednosti

- Preprosto dodajanje novih modelov naključnih omrežij
- Ne potrebujemo velike množice realnih podatkov



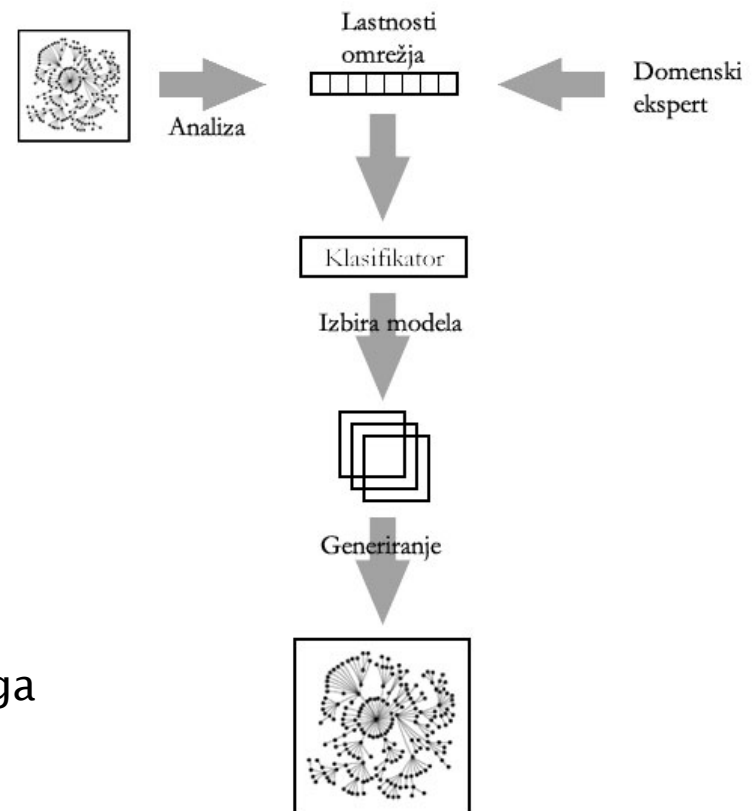
Model M-generator

■ Generiranje omrežja

- Analiziramo realno omrežje ali pa prepustimo nastavitve parametrov domenskemu ekspertu
- S pomočjo klasifikatorja izberemo primeren model
- Z izbranim modelom generiramo naključno omrežje

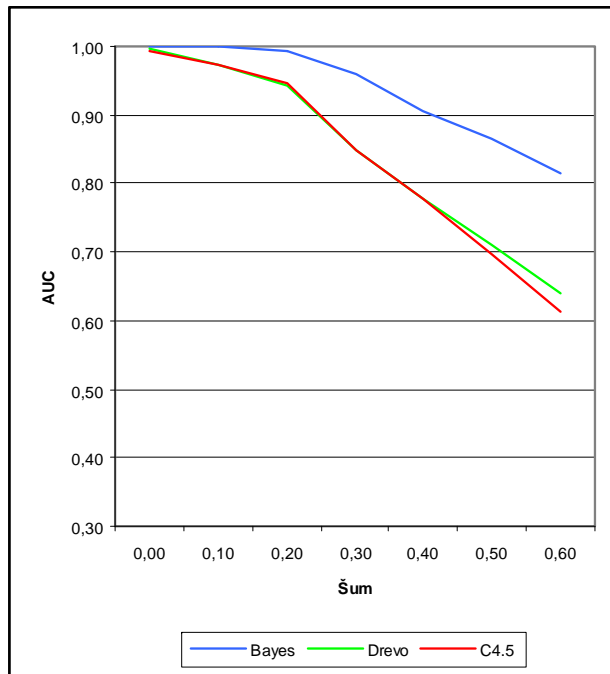
■ Prednosti

- Samodejna izbira modela iz množice razpoložljivih
- Ni potrebno posredovanje domenskega eksperta

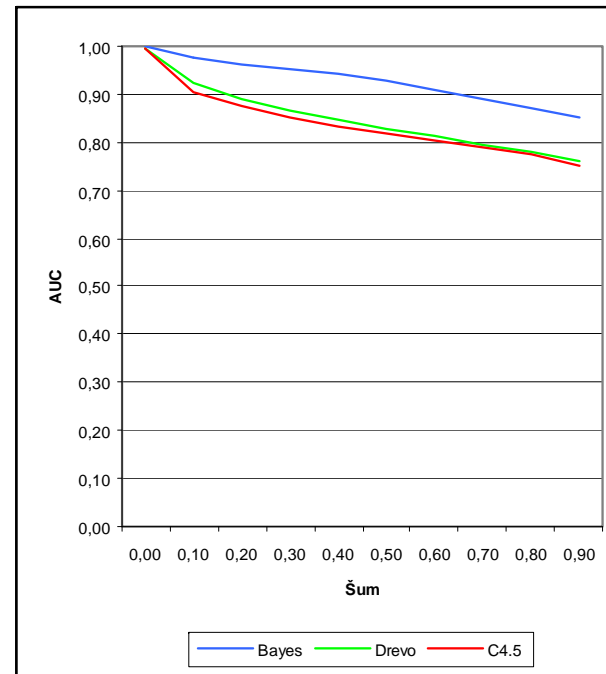


Rezultati testiranja

- Testiranje klasifikatorja na podatkih s šumom



Šum dodan omrežjem



Šum dodan vektorjem lastnosti

Rezultati testiranja

■ Testiranje nad realnimi omrežji

Analizirano omrežje	Izbran model	Verjetnost
Omrežje 500-tih največjih komercialnih letališč v ZDA	Model R-MAT	0.785
Omrežje sodelovanja med znanstveniki	Model majhnega sveta	0.606
Omrežje ženskega kluba	Poissonov model	0.995
Električno omrežje zahodnega dela ZDA	Model majhnega sveta	0.829
Socialno omrežje delfinov	Model majhnega sveta	0.722
Nevronsko omrežje	Model R-MAT	0.930

■ Primerjava realnega in naključno generiranega omrežja

	Vozlišča	Povezave	Alfa	Napaka R^2	Največja stopnja	Povpr. stopnja	Premer	Eksp. hop	Koeficient razvrščanja
Omrežje letališč:	500	2,980	-0.761	0.6581	145	11	4	2.610	0.6175
Model R-MAT:	500	2,874	-0.656	0.7414	206	11	3	2.257	0.1759
Model R-MAT:	10,000	49,862	-1.175	0.7664	539	9	4	4.422	0.01397

Sklep

- Z rezultati smo zadovoljni
- Pazljivost pri interpretaciji rezultatov
Zaradi pomanjkanja označenih podatkov ne moremo sklepati o uspešnosti pristopa v splošnem
- Možne številne izboljšave
 - Razširitev z dodatnimi klasifikatorji in atributi
 - Konstruiranje klasifikatorja nad realnimi omrežji za manjši, specifični nabor omrežij