

1. domača naloga: Glasovanje za Pesem Evrovizije

Domen Lušina (27172023)

23. oktober 2017

1 Podatki

Podatki so bili predstavljeni v dveh datotekah prva za polfinale in druga za finale Evrovizije. Za analizo sem se odločil uporabiti podatke iz finala. Lahko bi uporabil še podatke iz polfinala, vendar bi moral določiti utež, ki določa kako enakovredni so glasovi polfinala v primerjavi z glasovi finala. Ker se mi je zdel to že sam zase zahteven problem, sem raje uporabil zgolj podatke iz finala.

Podatki so bili predstavljeni tako, da je vsaka vrstica vsebovala pesem. Zadnjih 47 stolpcev pa so predstavljali glasove posamezne države. V tej nalogi smo torej opazovali profile držav (stolpce), kjer so pesmi predstavljale attribute.

2 Računanje razdalj

Naj x označuje profil prve države in y profil druge države, ter naj N označuje dolžino vektorja, ki predstavlja profil države. Razdalja je bila izračunana na podlagi evklidske razdalje (L2 norma):

$$D = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N'}}; \quad x_i, y_i \in \mathbb{R}. \quad (1)$$

Profili držav so vsebovali tudi neznane vrednosti. Če je vsaj en vektor vseboval polje z neznano vrednostjo, smo istoležni polji odstranili iz vektorja. Ker so različni profili vsebovali različno število neznanih vrednosti, smo razdaljo normirali s številom istoležnih polj, ki ne vsebujejo neznane vrednosti v prvem ali v druge vektorju. To število smo označili z N' . V primeru, da je vrednost N' enaka 0 je razdalja D neskončna.

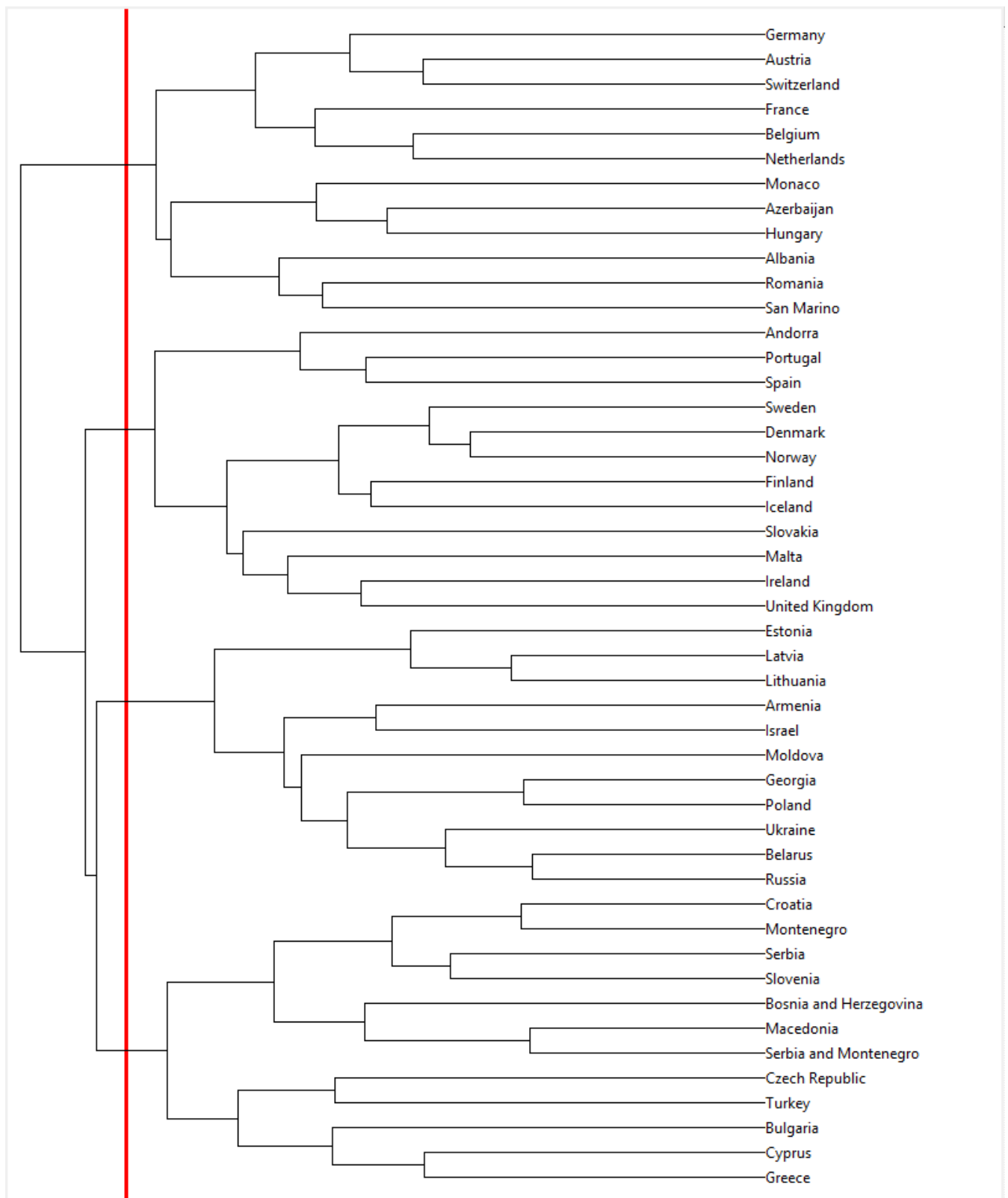
Za razdaljo med skupinami sem uporabil vse metode predstavljene na predavanjih (*angl. average linkage, single linkage, complete linkage*). Ker so razdalje, na podlagi prej določene evklidske razdalje (formula 1), včasih znašale neskončno sem moral določiti način kako te vrednosti obravnavati. Pri povprečni razdalji sem pri seštevanju razdalj izključil razdalje z vrednostjo neskončno in delil s preostalimi povezavami, ki niso znašale neskončno. Podobno sem storil tudi pri minimalni in maksimalni razdalji med dvema skupinama. Množica v kateri smo iskali minimalne oz. maksimalne razdalje ni vsebovala neskončnih razdalj. Na podlagi izrisanih dendrogramov (bolj razvidno je bilo, kako določiti skupine) sem se na koncu odločil uporabiti maksimalno razdaljo med skupinami (*angl. maximum linkage*).

3 Dendrogram

```

-----Germany
----|
      ----Austria
      ----|
            ----Switzerland
----|
      ----France
      ----|
            ----Belgium
            ----|
                  ----Netherlands
----|
      ----Monaco
      ----|
            ----Azerbaijan
            ----|
                  ----Hungary
----|
      ----Albania
      ----|
            ----Romania
            ----|
                  ----San Marino
----|
      ----Andorra
      ----|
            ----Portugal
            ----|
                  ----Spain
----|
      ----Sweden
      ----|
            ----Denmark
            ----|
                  ----Norway
      ----|
            ----Finland
            ----|
                  ----Iceland
----|
      ----Slovakia
      ----|
            ----Malta
            ----|
                  ----Ireland
            ----|
                  ----United Kingdom
----|
      ----Estonia
      ----|
            ----Latvia
            ----|
                  ----Lithuania
----|
      ----Armenia
      ----|
            ----Israel
      ----|
            ----Moldova
            ----|
                  ----Georgia
                  ----|
                        ----Poland
                        ----|
                              ----Ukraine
                              ----|
                                    ----Belarus
                                    ----|
                                          ----Russia
----|
      ----Croatia
      ----|
            ----Montenegro
      ----|
            ----Serbia
            ----|
                  ----Slovenia
      ----|
            ----Bosnia and Herzegovina
            ----|
                  ----Macedonia
            ----|
                  ----Serbia and Montenegro
----|
      ----Czech Republic
      ----|
            ----Turkey
      ----|
            ----Bulgaria
            ----|
                  ----Cyprus
            ----|
                  ----Greece
```

Slika 1: Tekstovni dendrogram.



Slika 2: Grafični dendrogram. Rdeča črta določi skupine in število skupin.

Skupina 1	Max.	Armenia 3.96	Greece 5.65	Serbia 6.55	Serbia and Montenegro 6.6	Turkey 6.79
	Min.	Slovakia 0	Finland 0.66	Belarus 0.78	Slovenia 0.78	Georgia 0.81
Skupina 2	Max.	Norway 4.47	Denmark 4.67	Sweden 4.76	Iceland 4.88	Estonia 4.96
	Min.	Slovakia 0	Macedonia 0.33	Georgia 0.63	Poland 0.65	Albania 0.73
Skupina 3	Max.	Georgia 5.9	Azerbaijan 5.23	Ukraine 7.12	Russia 8.62	Belarus 9.3
	Min.	Slovakia 0	Albania 0.05	Portugal 0.26	Spain 0.53	Macedonia 0.59
Skupina 4	Max.	Macedonia 5.75	Bosnia and Herzegovina 5.81	Bulgaria 7.3	Serbia 7.85	Serbia and Montenegro 8.63
	Min.	Poland 0.34	Switzerland 0.58	Portugal 0.69	France 0.71	Germany 0.72

Tabela 1: Preferenčne izbire posameznih skupin. Z Max. smo označili države z največ povprečnimi glasovi. Z Min. smo označili države z najmanj prejetih glasov skupine.

4 Skupine in njihove preferenčne izbire

Skupine sem določil glede na razliko razdalj med skupinami. Tam, kjer je bila ta razdalja največja sem določil skupine (rdeča črta na sliki 2). Tako sem dobil štiri skupine:

- Skupina 1 (Germany, Austria, Switzerland, France, Belgium, Netherlands, Monaco, Azerbaijan, Hungary, Albania, Romania, San Marino)
- Skupina 2 (Andorra, Portugal, Spain, Sweden, Denmark, Norway, Finland, Iceland, Slovakia, Malta, Ireland, United Kingdom)
- Skupina 3 (Estonia, Latvia, Lithuania, Armenia, Israel, Moldova, Georgia, Poland, Ukraine, Belarus, Russia)
- Skupina 4 (Croatia, Montenegro, Serbia, Slovenia, Bosnia and Herzegovina, Macedonia, Serbia and Montenegro, Czech Republic, Turkey, Bulgaria, Cyprus, Greece)

V tabeli 1 so predstavljene preferirane države skupin (Max.) in tiste, za katere ne glasujejo (Min.). Pod vsako državo v tabeli je številka, ki predstavlja povprečen glas skupine za to državo. Opazi se, da v večini največ glasov prejmejo države znotraj te skupine. Pojavijo se izjeme kot je npr. Slovaška, ki ima najmanj glasov v svoji skupini. Razlog za anomalije je pomanjkanje nastopov v finalu, kar lahko pomeni, da so bile pesmi tista leta slabe in zato

niso prejeli glasov oziroma smo npr. Slovaški določili napačno skupino, ker smo imeli premalo vzorcev (oz. nastopov).

Poudaril bi tudi države katere preferira Skupina 1. Te države niso znotraj držav, ki sestavljajo Skupino 1. To bi lahko utemeljili z velikim deležem imigrantov iz držav za katere so glasovale (npr. Nemčija ima veliko imigrantov iz Srbije in Turčije). Vendar ne poznam načina, s katerim bi lahko objektivno dokazal, da je to razlog za pridobljene rezultate.