

Topološka analiza podatkov

Detekcija oblik

Špela Čopi and Nika Kastelec

January 15, 2020

Povzetek

V poročilu vam bomo predstavili implementacijo detekcije oblike, ki smo jo izdelali v sklopu predmeta Topološka analiza podatkov. Za vhodne podatke problema smo vzeli množico točk in s pomočjo različnih topoloških prijemov določili ali predstavlja katero izmed vnaprej izbranih oblik. Reševanje problema temelji na določanju homološke grupe vhodnih podatkov s pomočjo obstojne topologije (*persistance topology*), gradnje Vietoris-Rips kompleksa in izračunavo Bettijevih števil.

1 Uvod

Problem prepoznavanja oblik je bil v nalogi skrčen na prepoznavanje določenih oblik. To so disk, črta, krožnica, sfera in torus. Te oblike po večini spadajo v različne homološke grupe, zato smo iskali postopek za določanje homoloških grup. Odločili smo se za izgradnjo Vietoris-Rips kompleksov in računanje Betti števil pri različnih razdaljah med točkami. Na podlagi spreminjanja števila komponent, lukenj in votlin smo nato določili homološko grupo. V drugem koraku smo morali ločiti med črto in diskom, ker oba spadata v isto homološko grupo. Tega smo se lotili z iskanjem preseka med sfero in črto oziroma diskom, ki ga nato odstranimo od sfere.

2 Podatki

Vhodne podatke smo izdelali s pomočjo programa Blender za 3D modeliranje. V programu smo zmodelirali disk, črto, krožnico, sfero in torus ter iz njih pridobili množice točk. Velikost vhodnih podatkov (oziroma število točk v množici) je bilo zaradi slabše učinkovitosti implementacije nekoliko omejeno navzdol.

Imeli smo učno in testno množico. Razlike med učnimi in testnimi modeli so bile v velikosti, poziciji in rotaciji.

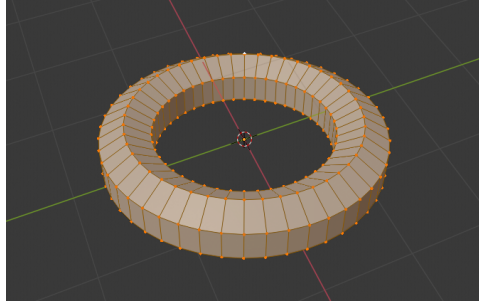


Figure 1: Primer modela v programu Blender.

3 Metode

3.1 Izgradnja Vietoris-Rips kompleksa

Pri izgradnji Vietoris-Rips kompleksa je pomembno kako izberemo interval ϵ is katerega vzemamo ϵ -e (razdaljo za grajenje kompleksa). V teoriji bi interval izbrali tako, da bi za posamezen objekt avtomatično izbrali interval, kjer se Bettijeva števila spreminjajo. Ker implementiran algoritem ni dovolj učinkovit in je prepočasn za velike ϵ -e, sva interval izbirali empirično. Za naše učne primere ϵ vzamemo iz intervala $(0, 1)$.

3.2 Bettijeva števila

Bettijeva števila β_i smo za posamezni Vietoris-Rips kompleks izračunali s pomočjo knjižnice `mogutda`. Ker so vsi zadani primeri vloženi v tri demenzionalni prostor, smo za vsak kompleks izračunali le β_0 , β_1 in β_2 .

Tako za vsak oblak točk, za katerega želimo prepoznati obliko, dobimo tabelo s tremi stolpci (za vsako Bettijevo število) in j vrsticami, kjer je j število Vietoris-Rips kompleksov, ki smo jih zgradili.

Potem iz danih stolpcev zgoraj opisane tabele razberemo prevladujoče Bettijevo število. To naredimo tako, da najprej preštejemo število zaporednih ponovitev. Nato za vsako izračunamo kakšen procent vseh izračunanih predstavlja. Na koncu izberemo tisto število, ki zadošča pogojem o dovolj velikem odstotku zaporednih ponovitev in in je hkrati največje od takih. Na tak način izbiramo zato, ker se zelo velika števila pojavijo malokrat, poleg zelenih razultatov pa se velikokrat ponovi 0, ki se pojavi kadar se različno deminzjonalne luknje zapolnijo (v zadnjih vrsticah) in preden se luknje v drugih dimenzijah pokažejo (v začetnih vrsticah, ko ima kompleks še veliko komponent). Ta način ni najbolj učinkovit, ampak deluje na izbranih oblikah.

3.3 Prepoznavanje modelov iz iste homološke grupe

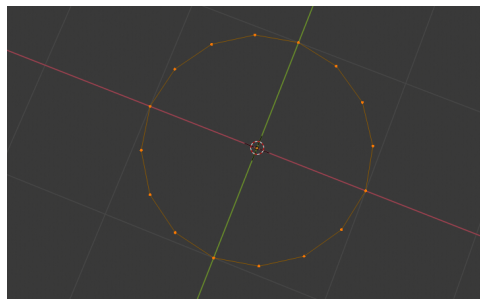
Ker imata črta in disk obe eno komponento, nič lukenj in nič votlin, se njuna Bettijeva števila ujemajo, dbimo vrednosti $\beta_0 = 1$, $\beta_1 = 0$ in $\beta_2 = 0$. Da lahko črto in disk razločimo, konstruiramo oblak točk, ki predstavlja sfero, nato pa od nje izločimo točke, ki so blizu prvotnega oblaka točk (črta ali disk). Tu moramo biti pozorni, da sfera ni prevelika in da je prav postavl-

jena v prostor. Potem na sferi brez izbranih točk ponovno kličemo algoritem za prepoznavo oblik in v primeru, da algoritem prepozna 2 komponenti (torej je disk razdelil sfero na dva dela) lahko sklepamo, da je prvotni oblak točk predstavljal disk, če pa je prepoznal eno komponento z dvema 1-dimenzijskima luknjama (torej črta prebode sfero na dveh mestih) lahko sklepamo, da je prvotni oblak točk predstavljal črto. Pri tem moramo biti pozorni na to, da je gostota točk na modelu (črta, disk) dovolj velika, da je razdalja med modelom in sfero pravilno izračunana.

4 Rešitve

Metodo testiramo na učnih in nekaj testnih primerih. Z nekaj ročne pomoči pri določanju intervalov za gradnjo Vietoris-Rips kompleksov je program za vse modele izpisal pravilno rešitev.

Oglejmo si nekaj vhodnih in ishodnih podatkov:



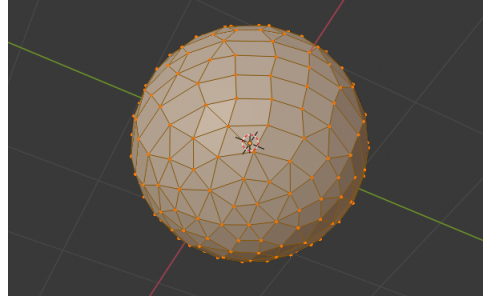
(a) Model krožnice v Blanderju

```

1  [32, 0, 0]
2  [32, 0, 0]
3  [32, 0, 0]
4  [32, 0, 0]
5  [32, 0, 0]
6  [1, 1, 0]
7  [1, 1, 0]
8  [1, 1, 0]
9  [1, 1, 0]
10 [1, 1, 0]
11 [1, 1, 0]
12 [1, 1, 0]
13 [1, 1, 0]
14 [1, 1, 0]
15 [1, 1, 0]

```

(b) Bettijeva števila
pri različnih ϵ -ih,
iz katerih algoritem
izračuna 1, 1, 0



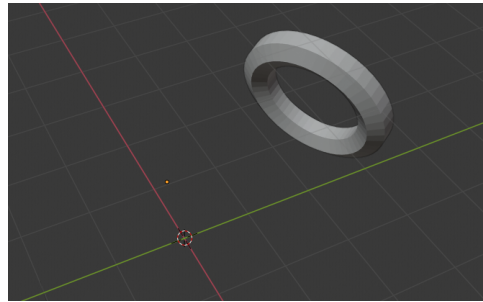
(a) Model krožnice v Blanderju

```

20 [6, 0, 0]
21 [4, 1, 0]
22 [2, 1, 0]
23 [1, 3, 0]
24 [1, 7, 0]
25 [1, 6, 0]
26 [1, 6, 0]
27 [1, 4, 0]
28 [1, 2, 0]
29 [1, 0, 0]
30 [1, 0, 0]
31 [1, 0, 1]
32 [1, 0, 1]
33 [1, 0, 1]
34 [1, 0, 1]
35 [1, 0, 1]

```

(b) Bettijeva števila pri različnih ϵ -ih, iz katerih algoritem izračuna 1, 0, 1



(a) Model krožnice v Blanderju

```

3 [288, 0, 0]
4 [194, 2, 0]
5 [100, 4, 0]
6 [100, 4, 0]
7 [6, 6, 0]
8 [6, 6, 0]
9 [6, 6, 0]
10 [6, 6, 0]
11 [1, 289, 0]
12 [1, 145, 0]
13 [1, 2, 1]
14 [1, 2, 1]
15 [1, 2, 1]
16 [1, 2, 1]
17 [1, 2, 1]
18 [1, 2, 1]

```

(b) Bettijeva števila pri različnih ϵ -ih, iz katerih algoritem izračuna 1, 2, 1

5 Zaključek

Program se da izboljšati na več načinov. Lahko bi avtomatizirali izbiro intervala za ϵ za gradnjo Vietoris-Rips kompleksa tako, da bi brez ročnega nastavljanja dobro prepoznal objekt ne glede na njegove relativne razdalje.

Knjižnica mogutda Bettijevih števil morda ni najbolj optimalna, saj je časovno zelo potrana, še posebej za $betta_i, i > 2$, kar tudi preprečuje prepoznavanje oblik višjih dimenzij.

Za boljše določanje Bettijevih števil vhodnih podatkov, bi morali najprej izločiti interval, kjer se število komponent, lukenj in votlin spreminja. Nato pa bi glede na dolžino intervala ter karakteristike sprememb avtomatično določili prag za izbiro Bettijevih števil.

6 Viri

<https://www.inegi.org.mx/eventos/2014/big-data/doc/P-MichaelLesnick.pdf>

<https://pypi.org/project/mogutda/>