

Szavazókörzetek újraosztása

András Réka

June 6, 2017

Chapter 1

Bevezetés

A redistricting kifejezés a választói körzetek határainak újra rajzolását jelenti. Erre a folyamatra időnként szükség van, mert egy terület lakosságának a száma nagy mértékben változhat. Választói szemszögből, ez azt jelenti, hogy a sűrűbben lakott körzetek polgárainak szavazata kevesebbet ér, mint a ritkábban lakott kerületekben élő polgároké. Ez ellentmond a demokrácia 'egy ember egy szavát' elvével. A szavazókörzetek újraosztása mégsem az igazságos választásokkal fonódott össze, hanem a gerrymandering fogalmával. A gerrymandering Elbridge Gerry szenátorról kapta a nevét, aki jóváhagyott egy olyan körzet felosztási tervet, amely jelentős előnyhöz jutatta a demokratákat. A különös alakú választókerület egy szalamandrához hasonlított. A Gerry névhez illesztve az angol 'salamander' szót megkapjuk a gerrymandering kifejezést. A gerrymandering vagy választókerület-manipuláció a választói körzeteknek egy olyan felosztását jelenti, amely kedvező/kedvezőtlen helyzetbe hozhat valamely pártot, etnikai, nyelvi, vallási csoportot, stb.. A választási eredmények ilyen manipulációjá kiváltképp a többségi elven működő demokratikus rendszerekben fordulhat elő. Ezek a rendszerek nem arányosan választják meg képviselőiket, az a képviselő nyer aki a szavazatok többségét kapta, így mindig lesznek kárba vesztett szavazatok. Tegyük fel, hogy egy 100 szavazóból álló körzetben, A párt 90 szavazatot kap, B párt 10-et. Ebben az esetben B párt vesztesége 10, A párté 39 (90 mínusz a győzelemhez szükséges 51 szavazat). A választókerület-manipulációt végző párt arra törekszik, hogy az ellenfél kárba vesztett szavazatait maximalizálja, míg az ő esetében minimalizálja a veszteséget. Annak érdekében, hogy az újraosztás folyamata igazságos legyen, az érintett országok különböző törvényeket hoztak egy felosztás elfogadására vonatkozólag. Legtöbb államban független megbízottak végzik az újraosztást, de ha automatizálni lehetne az egész folyamatot, biztosabb lenne, hogy a választópolgárok jogai nem sérülnek.

Chapter 2

GIS, QGIS, Python

A GIS(Geographic Information System) egy számítógépes rendszer, amelyet földrajzi helyhez kapcsolódó adatok megjelenítésére, elemzésére és tárolására dolgoztak ki. Az elképzelés a következő:

1. földrajzi koordinátákkal rendelkező adatok gyűjtése
2. adatok modellezése
3. adatok elemzése
4. eredmények megjelenítése térképen

Adatgyűjtés

A GPS-el felszerelt telefonok, autók folyamatosan képesek térbeli adatokat szolgáltatni, és a szatellitek már mindenki számára elérhető felvételeket készítenek a Föld felszínéről.

Adatmodellezés

A GIS kétféle modellt használ a térbeli adatok tárolására. Az egyik a raszteres(grid) adatmodell. Ez egy pixelekből felépülő adatmodell, ahol minden pixelhez azonos méretű területi egységet rendelünk. A pixelek sorokban és oszlopokban helyezkednek el, ez adja az adatok geometria jellemzőit, ugyanakkor minden mezőhöz értékek sorát rendelhetjük, mindegyiket megfeleltetve egy-egy tulajdonságnak. Gyakori raszter adat fájlformátumok: JPEG2000, IMG, GeoTIFF. A második adatmodell a vektoros adatmodell, ami pontokat (x,y koordináták) tartalmaz, és arra vonatkozó szabályokat, hogy a pontok miként alkotnak vonalakat és poligonokat (zárt vonalak). Ezen adatok tárolása gyakran shapefájlokban történik. Ez igazából legkevesebb három másik különböző formátumú fájl foglal magába. A DBF (dBase file) egy adatbázis tábla, ez tartalmazza az adataink attribútumait. Az shp kiterjesztésű fájl tartalmazza a földrajzi adatokat és ehhez még tartozik egy index fájl. Mivel a Föld térképen való megjelenítésének több módja is van, a vetítésre vonatkozó információk tárolása is szükséges, amennyiben a térképeket egymásra akarjuk illeszteni. A használt vetítés tárolása a projection (.prj) fájlokban történik. A vetítés a következőképpen néz ki:

1. Elkészítjük a Föld szabályos modelljét (szferoid, ellipszoid, gömb...)
2. A 3D modellt levetítjük egy 2D-s koordináta-rendszerbe

Adatelemzés

A QGIS számos előre beépített statisztikai és geometria eszközt nyújt az adatok elemzésére. Például: interpolálás, heatmappek generálása, GPS adatok betöltése, SQL lekérdezések végrehajtása (számolhatunk átlagokat, előfordulásokat, összegzéseket) , convex hurkok meghatározása, osztályozás, kontúrok

meghúzása. Mindezek mellett lehetőséget nyújt az adatok elemzésére általunk írt python vagy c++ kóddal. Amikor betöltünk egy layert a QGIS felépíti az annak megfelelő objektumot, amely a QgsInterface osztály iface nevű instanciáján keresztül érhetünk el. Az így kapott layer objektumtól kérhetünk egy iterátort a featurekhez. A featurek tartalmazzák a geometria információkat (koordináták, terület, stb...) és az adatbázishoz tartozó attribútumok értékeit. Ha a scriptet külső forrásból futtatjuk, importolnunk kell a qgis.core nevű könyvtárat, ez tartalmazza az alap GIS funkcionálisokat. Pythonban írhatunk pluginokat a QGIS-hez, vagy készíthetünk egy különálló applikációt.

Felhasználás

Az adatok reprezentálásának és elemzésének a GIS által biztosított módját sok területen használják: az epidemiológusok a vírusok globális terjedését vizsgálják, a kriminológusok bűnözési mintákat keresnek, a geológusok a Föld fizikai elváltozásait megfigyelve előre jelzik a földrengéseket, vagy lehetséges fosszilis erőforrásokat keresnek, a globális felmelegedéssel kapcsolatos kutatásokban is használnak földrajzi információs rendszereket.

Chapter 3

Algoritmus

3.1 A szimulált hűtés módszere

A szimulált hűtés módszere egy metaheurisztikus algoritmus, amelyet kombinatorikus optimalizációs problémák megoldására használnak. A kombinatorikus optimalizálás, olyan optimalizációs problémákat old meg, ahol a megoldástér véges. A szimulált hűtésnek több változata is van: Párhuzamos Szimulált Hűtés, Gyors Szimulált Hűtés, Adaptív Szimulált Hűtés. A továbbiakban a Klasszikus Szimulált Hűtés módszerét tárgyaljuk, amelyet Kirkpatrick, Gelatt, és Vecchi vezetett be 1983-ban [2], optimalizációs problémák megoldására. A cikkben leírják a termodinamikai rendszerek viselkedését, majd megmutatják hogyan használható az, mint optimalizációs technika. Egy termodinamikai rendszert különböző változók összessége ír le (hőmérséklet, energia, nyomás, entrópia, stb.), míg egy metaheurisztikus algoritmusnak csupán két eszközre van szüksége (költség függvény, szomszédos megoldást generáló függvény). A megfeleltetés a következő: az optimalizációhoz használt költség függvény a rendszer energiájának felel meg. Amikor a rendszer eléri az egyensúlyi állapotát az energia minimális, az optimalizáció célja pedig a költség függvény minimumának a megtalálása. Egy szomszédos állapot pedig egy atom mozgásnak felel meg. Bevezetünk egy hőmérséklet változót a termodinamikai rendszerek mintájára, amely szabályozza a keresésünk viselkedését. Nagy hőmérsékleten a molekulák kaotikus mozgást végeznek, ennek megfelelően az algoritmus is nagy ugrásokat tesz a megoldás térben, így elkerülhető az, hogy lokális minimumba ragadunk. Alacsony hőmérsékleten a molekulák mozgása lassul, az optimalizáló algoritmusunk pedig kis valószínűséggel fogad el gyengébb megoldásokat, gyakorlatilag lokális optimalizálást végez. A fémek hőkezelése során az anyagot magas hőmérsékletig melegítik, ott hűt tartják, majd lassan hűteni kezdik, így változtatva az anyag bizonyos tulajdonságait. Minnél lassabban hűtjük az anyagot annál több idejük lesz az atomoknak elhelyezkedni, vagyis stabilabb kristályok alakulnak ki, így lágyabb fémeket kapunk. A keresésünk is hasonlóképpen viselkedik. Akkor kapunk optimum közeli megoldást, ha a hűtés lassú, azaz minden hőmérsékleten nagy számú iterációt engedünk. Az iterációk során akkor fogadunk el egy új megoldást, ha az jobbnak bizonyul, mint az aktuális megoldás. Amennyiben nem alacsonyabb az új megoldásunk költség függvényének az értéke, akkor a jelenlegi megoldást a Metropolis kritérium alapján, $\exp(-(E - E_0)/T)$ valószínűséggel cseréljük le. Ahol E az aktuális megoldás objektív értéke, E_0 a szomszédos megoldás objektív értéke, T a rendszer hőmérséklete.

3.2 Több-célfüggvényű algoritmus újraosztáshoz

A több-célfüggvényű algoritmus különböző kritériumokat figyelembe véve optimalizál. A mi célunk olyan (K, U) párosításokat találni, amelyek legnagyobb mértékben megfelelnek az általunk definiált objektív függvényeknek. K a körzeteket jelöli, U a unitokat. A unitok a létező legkisebb közigazgatási egységei

a felosztandó országnak. Az ország egy lehetséges felosztása, vagyis egy megoldás, olyan (K,U) párok összessége, ahol minden körzet előre meghatározott számú unitot foglal magába, és egy unit csak egy körzethez tartozhat. Amikor összehasonlítunk két megoldást, ahol V az aktuális megoldás, V' a szomszédos megoldás, három lehetséges eset van:

1. V' minden objektív függvény szerint jobb megoldásnak bizonyul
2. V' csak bizonyos objektív függvényekre mutat jobb eredményt
3. V' minden objektív függvény szerint rosszabb megoldásnak bizonyul

A megoldásainkat egy olyan halmazba gyűjtjük, amely a fenti eseteknek megfelelő szabályok szerint bővül. Ha V' dominál más megoldásokat (1. eset), átveszi a leggyengébb dominált megoldás helyét. Ha az új megoldás nem dominál (nem igaz az 1. eset) és nem dominált (nem igaz a 3. eset), akkor a halmazhoz adódik. Ha V' dominált valamely megoldás által nem kerül a halmazba, viszont a szimulált hűtésnél leírt valószínűség szerint felválthatja a jelenlegi megoldást. Amikor egy új megoldás kerül a halmazba, mindig hozzá rendelődik egy random választott súlyvektor az előre meghatározott súlyok listájából. Az algoritmus végére minden súlyvektorhoz fog legalább egy nem dominált megoldás tartozni. A fent leírt algoritmus pseudokódban, úgy hogy k darab objektív függvénnyel dolgozunk.

Algorithm 3.1 Szimulált hűtés több célfüggvénnyel

Require: $k \geq 1$

Chapter 4

Redistricting

4.1 Általánosan

A megírt program a szimulált hűtés módszerével generálja egy ország megyéinek/államainak/tartományainak lehetséges körzetekre való felosztását. A célfüggvények és a hozzájuk tartozó súlyok tetszőlegesen megadhatók. A pythonban írt plug-in a QGIS-ből futtatható. A programnak szüksége van egy shapefájltra, amely tartalmazza a felosztandó megyéket, és egy másik fájltra amely a felosztandó ország közigazgatási egységeit (unit) tartalmazza. Ehhez a két fájlhoz tartozó adatbázis tábláknak összekapcsolhatóknak kell lenniük a megye azonosítók szerint, tehát tudnunk kell, hogy egy közigazgatási egység melyik megyéhez tartozik. A későbbiekben egy unit szomszédainak a bejárására gyakran kerül sor, ezért a shapefájlok első használata alkalmával egy scriptet is kell futtatni, amely bővíti egy attribútummal a unitokat tartalmazó adatbázis táblát. Az új attribútum az adott unit szomszédos unitjainak az azonosítóját tárolja. Ez később rengeteg futási időt nyer majd a körzetek változtatásához szükséges műveletek során. A plug-in könyvtár-struktúrája tartalmaz egy `config.yaml` elnevezésű fájlt, ezt változtatva állíthatjuk be a paramétereinket. Elsősorban meg kell adni a shapefájlainkban szereplő attribútumok elnevezéseit, ezek helyes ismerete nélkül a program nem lesz futtatható. Meg kell mondanunk összesen hány körzetet akarunk látni a végső felosztásban. A további paraméterek az optimalizáló algoritmus viselkedését befolyásolják, ezek változtatása nélkül is használható a program. A szimulált hűtés módszerének olyan paramétereit változtathatjuk, mint a kezdeti és végső hőmérséklet, a hűtés mértéke és a hőmérsékletenkénti iteráció. Beállítható a megoldás halmaz maximális mérete és a költségfüggvényekhez rendelhető lehetséges súlyok listája. Az is megmondható, hogy egy körzethez minimum hány unitnak kell tartoznia. A használt költségfüggvényeket az `objectives.py` fájl tartalmazza, ezek bővítése értelem szerűen nem adható meg a konfigurációs fájlban. Amennyiben újabb függvényt szeretnénk egy megoldás kiértékeléséhez adni azt implementálnunk kell az `objectives.py` fájlban. A program indítás után iterációnként loggolni fogja a megoldáshalmaz viselkedését. A hőmérséklet lehűlése után a legjobb megoldás megjelenik a QGIS grafikus felületén, úgy hogy az ugyanazon körzethez tartozó unitok azonos színűek lesznek.

4.2 Románia

Chapter 5

Összefoglalás, eredmények

Bibliography

- [1] E. A. Ringc3n-Garc3a, M. A. Guterrez-Andrade et al., A Multiobjective Algorithm for Redistricting, Journal of Applied Research and Technology Volume 11, Issue 3, June 2013, Pages 324–330.
- [2] S. Kirkpatrick; C. D. Gelatt; M. P. Vecchi , Optimization by Simulated Annealing , Science, New Series, Vol. 220, No. 4598. (May 13, 1983), pp. 671-680.