

Shluková analýza funkčních ploch pro sledování urbánního vývoje olomouckého regionu

Martin Klícha

Katedra geoinformatiky, Přírodovědecká fakulta, Univerzita Palackého v Olomouci, tř.
Svobody 26, 771 46, Olomouc, Česká republika
Kontakt.manas.klicha@gmail.com

Abstrakt. Hlavním cílem bakalářské práce je využití zejména lokálních metod shlukové analýzy pro sledování výskytu statisticky významných shluků vybraných funkčních ploch v rámci města Olomouc i v rámci celého olomouckého regionu. Shlukovou analýzu poté aplikovat na rozmístění funkčních ploch v různých časových obdobích a to v letech 1930, 1955, 1985, 1999 a 2010. Významným výsledkem práce je nejen vizualizace analýz ve formě map, ale také komentář popisující změny rozložení funkčních ploch v čase.

Klíčová slova: shluková analýza, funkční plochy, shluky

Abstract. The main aim of this bachelor thesis is the use local cluster analysis methods or monitoring the occurrence of statistically significant clusters of selected functional areas within the city of Olomouc and also in the framework of the Olomouc region. Cluster analysis then applied to the distribution of functional areas at different times in 1930, 1955, 1985, 1999 and 2010. An important result of the work is not only the visualization of analysis in the form of maps, but also a commentary describing the changes in the location of functional areas in the time.

Keywords: cluster analysis, functional areas, clusters

1 Úvod

Shluková analýza je vícerozměrná statistická metoda, která slouží ke klasifikaci objektů do skupin neboli shluků. Hlavní princip analýzy je shlukovat jednotlivé prvky, které jsou si co nejvíce podobné. Shluková analýza je používána v řadě vědních oborů, jako je biologie, medicína nebo geografie.

V této práci je kladen důraz na lokální metody shlukové analýzy a jejich aplikaci pro sledování změn urbánního vývoje v Olomouci a olomouckého regionu.

Dnešní struktura měst často odráží historickou minulost a její vývoj závisí na mnoha faktorech. Vývojem a strukturou měst se dnes zabývá vědní obor územní plánování, proto výstupy této práce mají sloužit zejména urbanistům a geografům, kteří se zabývají vývojem a uspořádáním prostorové struktury měst.

2 Cíle práce

Hlavním cílem práce je využití zejména lokálních metod shlukové analýzy (např. Moranovo I, G_i^*) pro sledování výskytu statisticky významných shluků vybraných funkčních ploch v rámci města Olomouc i v rámci celého olomouckého regionu.

Před vlastním započítáním analýz bude digitalizací doplněna databáze funkčních ploch. Shluková analýza bude poté aplikována na rozmístění funkčních ploch v různých časových obdobích a to v letech 1930, 1955, 1985, 1999 a 2010.

Hlavní výsledky práce lze rozdělit do tří kategorií. První kategorie obsahuje samotnou digitalizaci funkčních ploch pro rok 2010, na které se podílela také ještě jedna studentka. Tato digitalizace je základem pro další bakalářské práce. Do druhé kategorie lze zařadit vizualizace analýz ve formě map. Velice významnou a poslední kategorií výsledků, bude komentář popisující shlukové analýzy funkčních ploch a jejich změny rozložení v čase.

Při zjišťování dostupnosti dat pro různá časová období bylo zjištěno, že data v letech 1930, 1955, 1985, 1999 jsou pouze pro území Olomouce, nikoli pro celý olomoucký region. Z tohoto důvodu, byla práce zaměřena především na analýzy na území Olomouce. Shlukové analýzy pro celý olomoucký region byly vyhotoveny pouze pro rok 2010.

3 Použité metody a postupy zpracování

3.1 Použitá data a programy

V této podkapitole bude shrnuto jaká data a především jaký typ dat byl použit. Použitá data byla vždy ve formátu SHP a lze je rozdělit dle původu do dvou skupin. Data převzatá, která byla zapůjčena Katedrou geoinformatiky Univerzity Palackého v Olomouci. A data vzniklá především za účelem této práce.

Mezi data převzatá patří data funkčních ploch od studentky Zuzany Zapletalové, která se ve své loňské diplomové práci zabývala funkčními plochami pro roky 1930, 1955, 1985 a 1999. Data těchto funkčních ploch byla pořízena digitalizací historických územních plánů a jsou důležitou součástí této práce. Tyto data jsou však pouze pro katastrální území Olomouce v daných letech, nikoli pro olomoucký region. Dále bych do této skupiny také zařadil hranice katastrů pro jednotlivé roky a hranice obcí spadající do olomouckého regionu.

Mezi data vniklá za účelem této práce patří vlastní digitalizace funkčních ploch pro rok 2010. Tato digitalizace byla prováděna pro celý olomoucký region a podílela se na ní také další studentka. Podrobnější popis vlastní digitalizace a případných problémů je uveden v podkapitole 3.2.

V práci byl použit především hlavní produkt od společnosti esri ArcGIS a konkrétně jeho starší verze 9.3. K analýzám byl využit nástroj Hot spot Analysis (Getis-Ord G_i^*), který je součástí toolboxu Spatial statistic. Ke grafickým úpravám a editacím byl zvolen volně dostupný grafický program GIMP. Pro tvorbu textové části Microsoft Office Word 2003.

3.2 Digitalizace funkčních ploch

Digitalizace funkčních ploch byla nezbytnou součástí této práce a to zejména z důvodu absence dat funkčních ploch pro olomoucký region pro rok 2010. Probíhala ve spolupráci ještě další studentky. Celý olomoucký region byl rozdělen na dvě rovnocenné části 29 a 25 obcí. Obě tyto části byly později spojeny do jedné výsledné vrstvy funkčních ploch.

Při vlastní digitalizaci bylo vycházeno z nejnovějších dostupných územních plánů olomouckého regionu, současně s tím proběhla kontrola nad katastrální mapou a ortofotem. Vše bylo digitalizováno při měřítku územního plánu 1 : 5000.

Počáteční vrstva byla poskytnuta z Magistrátu města Olomouce, která již obsahovala základní rozložení funkčních ploch. Avšak nebyla zde dodržována základní topologická pravidla. Polygony se překrývaly, byly nedotažené či neúplné. Z tohoto důvodu bylo lepší polygony smazat a začít digitalizovat znovu.

V průběhu digitalizace se vyskytla řada problémů, které bylo potřeba operativně vyřešit. Jednalo se především o neaktuálnost územních plánů. V mnohých případech nastal případ, kdy v územním plánu byla dotyčná funkční plocha evidována jako návrh či výhled, ale ve skutečnosti byla tato funkční plocha již zrealizována. Typickým příkladem takových ploch byl například jižní obchvat Olomouce nebo část rychlostní silnice do Hranic, která v územních plánech byla zaevidována jako návrh. Tyto typy problémů byly obecně řešeny s ohledem na využití nově zdigitalizovaných funkčních ploch v této práci. Kde bylo potřeba zajistit co nejaktuálnější stav funkčních ploch. V jednom případě nebylo území obce zcela pokryto územními plány. Jedná se o katastrální území obce Hlubočky, které se nachází severně od Olomouce. Z tohoto důvodu část katastrálního území této obce nebyla digitalizována.

Před spojením obou částí digitalizace byla provedena řada úprav. Mezi tyto úpravy patří zejména spojení jednotlivých kategorií pomocí nástroje merge a poté použití nástroje Explode Multi-part Feature, pomocí kterého byly rozděleny jevy ze stejných kategorií, které spolu přímo nesoudily, nebo na sebe nenavazovaly.

Výsledná vrstva funkčních ploch olomouckého regionu má rozlohu 608 km², je dostupná ve formátu SHP a obsahuje 13 454 polygonů, kterým bylo přidáno 17 atributů. Mezi ně patří i klasifikace, zda se jedná o stav, návrh nebo výhled dané funkční plochy. Tato digitalizace je důležitým výstupem této práce. Vrstva funkčních ploch olomouckého regionu pro rok 2010 je základem dalších bakalářských prací.

3.3 Předzpracování dat

Tato podkapitola je věnována velmi stručnému shrnutí úpravě dat před samotným spuštěním shlukové analýzy.

Úplně počátečním vstupem byly tedy vrstvy funkčních ploch pro jednotlivé roky. Tyto data funkčních ploch bylo třeba nejprve agregovat. Po konzultaci s vedoucím práce byla vybrána agregace dat, která je založena na principu „rozsekání“ funkčních ploch pomocí pravidelné hexagonální sítě a spočítání procentuálního zastoupení rozlohy jednotlivých kategorií funkčních ploch v daném hexagonu. V tomto případě bylo nutné zaručit stále stejnou velikost hexagonů, proto velikost sítě byla nastavena

tak, aby byla totožná s velikostí maximálního možného rozsahu území, tedy s velikostí olomouckého regionu. Hexagonová síť byla vytvořena pomocí extenze Repeating Shapes for ArcGIS. Rozsekat funkční plochy se podařilo pomocí nástroje Identity. Poté už pomocí dalších několika jednoduchých úprav, které jsou dostupné v programu ArcGIS, byly získány jednotlivé vrstvy vybraných funkčních ploch, obsahující atribut procentuálního zastoupení rozlohy v daném hexagonu. Toto byl hlavní vstupní atribut pro shlukovou analýzu.

3.4 Typy lokálních shlukových analýz

Obecně prostorově založené statistické metody, na rozdíl od těch neprostorových, obsahují prostorovou složku, jako je například blízkost daných prvků, velikost plochy a další prostorové vztahy. Jsou tedy určeny speciálně pro práci s geografickými daty. Jsou navrženy především pro souhrn hlavních charakteristik prostorového rozložení, identifikaci statisticky významných prostorových shluků (hot/cold spots) či prostorových outlierů, posouzení celkového modelu shlukování či disperze a v neposlední řadě i modelováním prostorových vztahů.

Pro mapování shluků a jejich analýzu, především pak k lokalizaci statisticky významných shluků vysokých hodnot, nízkých hodnot či prostorových outlierů, slouží sada nástrojů v programu ArcGIS. Tyto nástroje jsou dostupné v toolboxu Spatial Statistic. Používají dvě základní metody analýz.

První metoda LISA identifikuje shluky (bodů) s podobnými hodnotami a shluky s rozdílnými hodnotami, tj. nezabývá se velikostí hodnot. Určuje, zda se vyskytuje shluk prvků, nebo prostorový outlier.

Druhá metoda Getis – Ord GI* identifikuje shluky vysokých hodnot (hot spots) a shluky s nízkými hodnotami (cold spots). Tato metoda je naprosto vyhovující pro účel této práce, proto byla po konzultaci s vedoucím práce vybrána jako hlavní metoda pro tvorbu analýz.

3.5 Analýza Getis-Ord GI*

Tato metoda nazývána také Hot spot analýza, slouží k identifikaci rozmístění prostorových shluků vysokých hodnot (hot spots) a prostorových shluků nízkých hodnot (cold spot).



Obr. 1. Schéma průběhu analýzy Getis-Ord GI*.

Analýza Getis – Ord GI* je dána vztahem:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1}}} \quad (1)$$

kde x_j je hodnota atributu prvku j vstupující do analýzy, $w_{i,j}$ prostorová váha mezi prvky i a j , n je hodnota celkového počtu prvků a:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (1)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \left(\bar{X} \right)^2} \quad (3)$$

G_i^* index je přímo vyjádřen hodnotou z-score, proto nejsou potřeba již žádné další výpočty pro testování statistické významnosti.

Analýza je založena na principu náhledu na jednotlivé prvky v rámci souvislosti s prvky sousedními pomocí prostorových vztahů. Výsledný GI* index je vypočítáván pro jednotlivé prvky. Jeho hodnota je vyjádřena přímo hodnotou z-score. Nejsou tedy nutné žádné další výpočty pro testování statistické významnosti. Prvek s vysokou hodnotou, však ještě nemusí být nutně statisticky významný prostorový shluk vysokých hodnot. Pro splnění funkce statisticky významného shluku vysokých hodnot, musí prvek nejen obsahovat vysokou hodnotu, ale musí být také obkloповán vysokými hodnotami sousedních prvků. Lokální součet hodnoty prvku a hodnot jeho okolí je proporcionálně srovnáván se součtem hodnot všech prvků v území. Pokud je lokální součet značně odlišný od očekávaného lokálního součtu, znamená to, že tento rozdíl nemůže vzniknout náhodně a jedná se tedy o statisticky významný výsledek (Z-score)

Výstupem jsou hodnoty Z-score a p-value pro každý prvek. Tyto hodnoty určují statistickou významnost prostorového shlukování a jsou zapsány v atributové tabulce analýzy.

Z-score je test statistické významnosti, který nám pomáhá rozhodnout, zda přijmout nebo odmítnout nulovou hypotézu. Nulová hypotéza v případě hot spot analýz je vyslovena takto: „Hodnoty (prvky) jsou v území rozmístěny náhodně (neexistuje zde

prostorové shlukování). P-value je pravděpodobnost, se kterou zavrhneme nulovou hypotézu.

Vysoká hodnota Z score a malá hodnota p-value (statisticky významná, tj. $< 0,05$) pro daný prvek znamená existenci prostorového shluku vysokých hodnot v okolí (hot spot). Nízká záporná hodnota Z score a malá hodnota p-value pro daný prvek znamená existenci prostorového shluku nízkých hodnot v okolí (cold spot). Čím je Z score větší nebo menší, tím je shlukování intenzivnější. Z score blízké nule znamená, že se zde nevyskytuje žádný zjevný shluk.

Velmi důležitý je výběr koncepce prostorových vztahů používaných pro analýzy. Měl by být založen na pochopení interakce prostorových vztahů analyzovaných prvků. U hot spot analýz je obecně doporučováno, používat metodu konstantní vzdálenosti. Lze ale použít i metody založené na inverzní vzdálenosti.

Vstupní data musejí mít definován souřadnicový systém a to především z důvodu práce s daty, v nichž je nutné definovat jednotky.

Analýza pracuje s pouze s body (centroidy) polygonů. Vyžaduje variabilitu vstupních hodnot, pracuje pouze s kladnými hodnotami. Vstupní data nesmí obsahovat nulové hodnoty.

3.6 Postup zpracování

Nejprve bylo třeba řádně nastudovat dostupnou literaturu, poté přišla na řadu digitalizace a zjišťování dostupnosti dat. Dalším důležitým krokem a fází mnohých rozhodování nastala při předzpracování dat. Hlavní a zároveň nejdůležitější částí byla samostatná tvorba shlukových analýz. Poté musely být analýzy vhodně vizualizovány. Tomuto kroku bylo také věnováno hodně pozornosti. Bylo třeba najít co nejlepší řešení z hlediska čitelnosti a přehlednosti mapových výstupů. Poslední část této bakalářské práce byla tvorba textové části, kde byly shrnuty všechny důležité poznatky dané problematiky.

4 Závěr

Pro shlukové analýzy byly vybrány 4 typy funkčních ploch. Funkční plochy bydlení, veřejného vybavení, výrobních ploch a zeleně. Celkem bylo vyhotoveno 9 map ve formátu A3.

Z důvodu nedostupnosti dat byly analýzy pro olomoucký celý region vytvořeny pouze pro rok 2010. V ostatních letech pak pouze na katastrálním území Olomouce. Protože při analýzách velmi záleží na velikosti analyzovaného území, byly ještě vyhotoveny analýzy na katastrálním územím Olomouce z roku 1930. Ty slouží především pro lepší porovnávání shlukování funkčních ploch v čase.

Reference

1. ANSELIN L. Local indicators of spatial association – LISA. *Geographical Analysis*, Ohio State University Press, 1995
2. Esri, <http://webhelp.esri.com>, ArcGIS Desktop 9.3 Help
3. GETIS, A., ORD J. K. The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*. no 3. 1992