

Úskalí využití geolokačních dat mobilních operátorů

Geolokační data ve veřejné správě

17.-18. dubna 2024, Uměleckoprůmyslové muzeum v Praze

Vysoká škola ekonomická v Praze

Data mobilních operátorů přináší ohromné možnosti a výzvy.

Ale jen v případě, že si uvědomíme jejich slabiny.

Rozšíření oficiální statistiky?

- Oficiální statistika – jasně definované ukazatele s jasně definovaným detailem
- Geolokační data mobilních operátorů
 - možnosti většího detailu (prostorového, časového, ...)
 - možnosti odlišného pojetí (mobility, přítomného obyvatelstva, ...)
- Podobnost s ukazateli oficiální statistiky spíše zdánlivá
- Nutnost nejprve data zcela pochopit
- Nutnost definovat celou řadu parametrů úplně znovu
- Bez absolutní důvěry a spolupráce mezi zadavatelem a zhotovitelem to nejde
- Zhotovitel musí vysvětlit všechna omezení
- Zadavatel musí dobře vystihnout otázku, na kterou hledá v datech odpověď

Kolik má operátor k dispozici SIM?

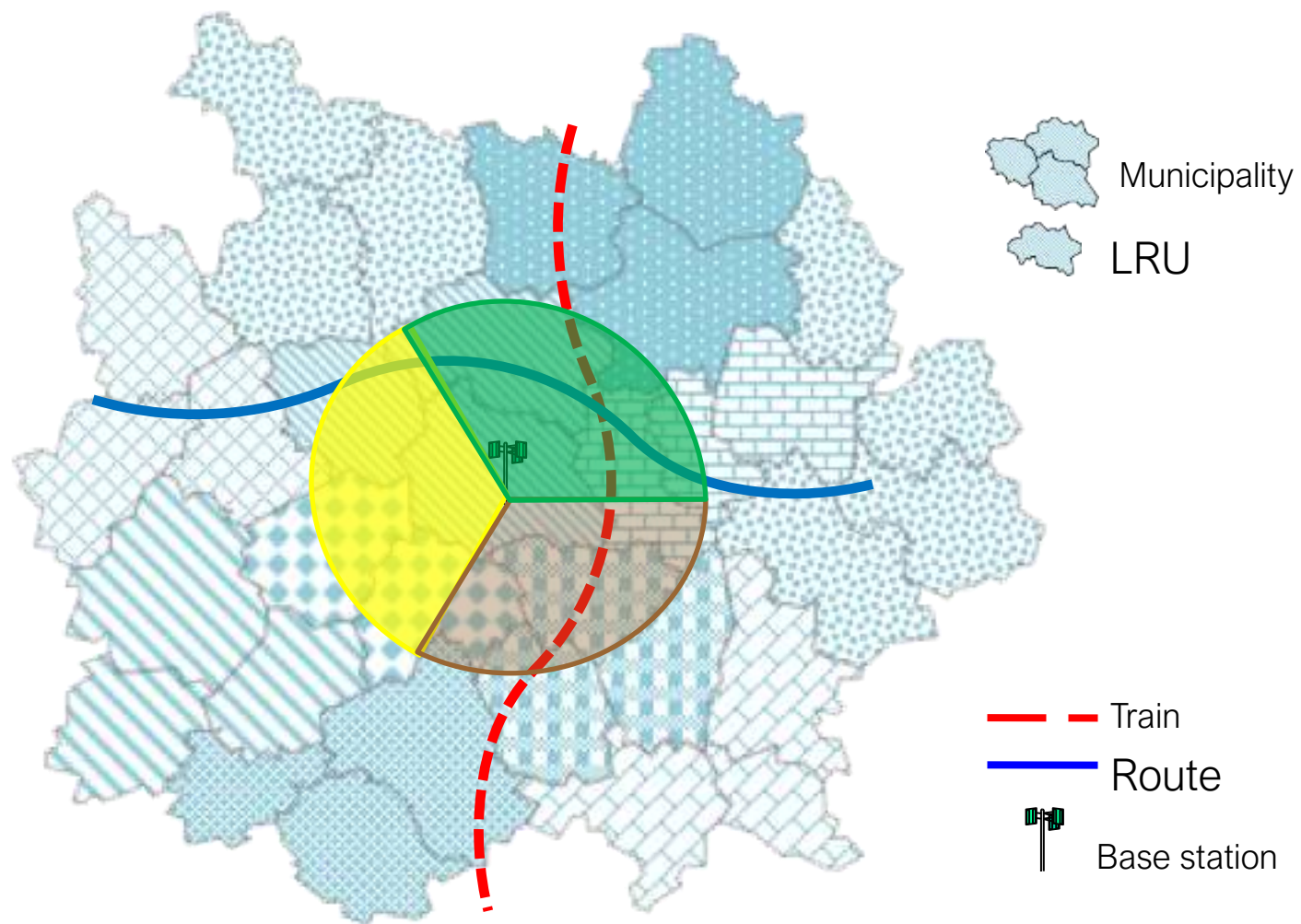
Vysoká škola ekonomická v Praze

Kolik má operátor k dispozici SIM?

- Vyčlenění specifických skupin
 - Technické SIM
 - Virtuální operátoři
 - Cizinci
 - Neaktivní SIM
 - Problémy v regionálních odlišnostech specifických skupin
- Řešení multiSIM
- Řešení osob bez SIM
- **Důležitý parametr pro dopočet na celou populaci**

Jak propojit data ze sítě s územím?

Vysoká škola ekonomická v Praze



Jak propojit data ze sítě s územím?

- Jinak řečeno – jak rozdělit uživatele z jednoho vysílače mezi dvě (či více) obcí v území?
- Využití údajů o hustotě zalidnění obcí
- Využití dat o způsobu využití konkrétního území (v jakém detailu?)
- Pokrytí mobilním signálem je jen matematický model (navíc proměnlivý v čase)
- Problémy s příliš velkým detailem (malé území)

Čas – věčný nepřítel

Vysoká škola ekonomická v Praze

Čas – věčný nepřítel

- Snaha o zachycení spojitě věci diskrétně (nespojité)
- Neznalost výskytu uživatele mezi dvěma intervaly
 - V rámci jednoho vysílače
 - V rámci změny polohy z jednoho vysílače na jiný
- Předpoklad o nutnosti přiřadit čas k některému území

A na scénu přichází *ZADAVATEL....* Aneb KDY, KDE, KDO...

Vysoká škola ekonomická v Praze

KDY (období) a KDE (územní detail)

- Většinou je nutné sledovat delší časové období – úskalí prázdnin, svátků, specifických událostí (např. meteorologických)
- Chování uživatelů je značně sezónní
- Územní detail až na ZSJ
- S růstem detailu se zvyšuje riziko nekvalitního modelu síť-území
- Potíže např. v území sousedícím s hranicí státu

KDO nás vlastně zajímá? – rezidenti

- Uživatel, který během sledovaného období přenocoval v území alespoň X nocí
 - Co znamená přenocoval?
 - Stačí jeden výskyt v nějakém intervalu? Či více výskytů?
 - Nocuje vrátný nebo zdravotní sestra na noční směně? Či chlapec, co se vrací z divadla prvním ranním spojem?
 - Je „X“ zvoleno správně? I když je „X“ větší než 50 %? Studenti na kolejích?
 - Co když někdo nemá více než 50 % přenocování na jednom místě?
 - Stále máme informaci jen o záznamu signálu z přístroje v síti...
 - Přesto má (většinou) v modelech každý uživatel svoji „domovskou“ územní jednotku

Kdo nás vlastně zajímá? – cestující

- Analyzujeme směry mobility (souvislost s volbou detailu území)
- Analyzujeme intenzitu mobility (jak často jedinci takovou cestu realizují) – souvislost s volbou „často/výjimečně“
- Díky přiřazení domovského území analyzujeme skladbu uživatelů v území z hlediska jejich domovské příslušnosti či intenzity mobility
- Výrazně složitější je odhad způsobu dopravy (hromadná, individuální)
- Rizika spojená se zobecněním – jedinec trávící na území (např.) 30min denně = každodenní dojíždka – projíždí, přesedá, vyzvedává oběd, ...?

S čím výsledky srovnat?

Vysoká škola ekonomická v Praze

S čím výsledky srovnat?

- Snaha ověřit si výsledky je pochopitelná
- Jsou data ze SLDB nebo evidence obyvatel srovnatelná?
- Využití zdrojů oficiální statistiky pro přepoččet na celou populaci je krokem v postupu u většiny projektů
- Hledání odlišností mezi výsledky modelu a daty oficiální statistiky je krokem v postupu u většiny projektů
- Výsledky nesmí být ani příliš podobné, ani příliš odlišné
 - Při podobné definici ukazatele a rozdílných výsledcích musí zhotovitel vyjasnit důvod rozdílů
- Pro některé ukazatele neexistuje vhodné srovnání

Na co se ještě nedostalo...

Vysoká škola ekonomická v Praze

Na co se ještě nedostalo...

- Operátoři spolu mluví, spolupracují, ale nesdílí...
 - Prakticky nemožnost řešit multiSIM
 - Nemožnost (či omezená možnost) řešit cizince
- Cestujeme i když stojíme/ležíme/sedíme
- Výpadky sítě
 - Nemusí se najít ideální časový interval
 - Mohou mít vliv na propojení sítě a území
- Bitva s anonymizací údajů

Děkuji za možnost spolupracovat na
projektu „**Zlepšení podmínek pro
decentralizaci a dostupnost
veřejné správy v území**“

Vysoká škola ekonomická v Praze