

Predstavitev primera »big data«

Daniel Gruden

18.3.2021

Zbiranje podatkov s statičnimi obcestnimi senzorji

Zbiranje podatkov s statičnimi obcestnimi senzorji je metoda, ki omogoča stalno monitoriranje prometa na znani lokaciji. Opazujemo število vozil in povprečno hitrost vsako minuto in podatke hranimo v podatkovni bazi. Podatke rabimo za napoved zastojev in minimiziranje časa potovanja. Senzorji niso enakomerno razporejeni po cestni mreži, ampak le na določenih lokacijah: to omogoča bolj ekonomično uporabo senzorjev, ampak tudi otežuje analizo podatkov. Senzorji tudi niso povsem zanesljivi in nekaterim manjkajo podatki za več minut na dan. Težko je tudi določiti povezavo med podatki zbranimi na enem senzorju in na drugemu, tudi če sta zelo blizu, ker vozila med merjenjimi spreminjajo hitrost in senzorji naredijo samo 1 meritev na minuto.

Podatki niso točni, za njihovo uporabo je treba odstraniti šum in preveriti ali podatki manjkajo. Šum odstranimo z algoritmom tako, da ga jemljemo kot stokastični proces, medtem ko jemljemo signal kot deterministični. Problem manjkajočih podatkov na moremo rešiti, tako, da gledamo podatke pred in po »luknji«, ker vrednosti podatkov močno nihajo, tako da najdemo donor z Markovskimi verigami.

Zbiranje podatkov s senzorji na vozilih

Na vozilih delimo dva različna tipa senzorjev:

- tehnični senzorji: zaznavajo delovanje vozila, monitorirajo okvare in omogočajo napovedovanje vzdrževanja
- okoljski senzorji: zaznavajo cesto, vreme, druga vozila.

Informacije ob vožnji so potrebne takoj, tako, da je potrebno zmogljiv računalnik za hitro procesiranje podatkov. Tako pridobljene informacije lahko dopolnimo z navigacijo GPS.

Problem pri okoljskih senzorjih je, da lahko merimo samo neposredno bližino avta. Če pa avtomobil povežemo z internetom, s pomočjo 4G/5G lahko izkoristimo še fiksne obcestne senzorje in senzorje na drugih vozilih, lahko vozniku podamo veliko več zanimivih funkcij: lahko napovedujemo prometne konice v realnem času, izbiramo najbolj optimalno pot, obvestimo voznika glede na morebitne prometne nesreče, poplave na cestah, o velikih dogodkih, kot so koncerti športne tekme, ki pritegnejo veliko ljudi in mašijo ceste.

Podatke je možno tudi deliti z zavarovalnici in omogočiti nove zavarovalne pogoje, npr. plačevanje glede na čas uporabe vozila, glede na prevožene kilometre ali glede na nevarnost območja, po katerem ponavadi vozimo. Največji izziv je torej deljenje podatkov iz vozila v omrežje, ki postavlja vprašanja o varstvu podatkov in prejetju podatkov iz omrežja, kar predstavlja vprašanja o kibernetski varnosti.

Senzorji morajo biti zanesljivi in dajati točne podatke, saj želimo izogniti uporabnikom težave z dodatnim vzdrževanjem, ki bi bilo potrebno, če bi bili senzorji nezanesljivi. Morajo biti tudi točni, ker bojo uporabniki tem podatkom zaupali in se ne bojo spraševali, ali so podatki točni ali ne.

Zbiranje podatkov za meterološke namene

V Sloveniji imamo različne meteorološke postaje:

- meteorološka postaja 1. reda: meri samodejno zračni tlak (z elektronskimi barometri), meterološko vidnost, temperaturo zraka (s termometrom), vlažnost zraka, količino padavin (Hellmanov dežemeter), vrsto, čas trajanja in jakost padavin (s pluviografom), višino snežne odeje (s snegomerom), smer in hitrost vetra (z anemometrom), gostoto toka globalnega in difuznega sončnega sevanja, trajanje sončnega sevanja, višino baznih oblakov in oblačnost. Operator meri še druge spremenljivke.

-letalska meteorološka postaja: samodejno meri enake količine kot meteorološka postaja 1. reda ampak je urink opazovanjprilagojen delovanju letališč.

-podnea postaja: meri temperaturo zraka, vlažnost zraka, višino padavin., vrsto, čas in trajanje padavin, višino snežne odeje in ponekod še druge količine.

-padavinska postaja: enkrat dnevno izmeri višino padavin, snežne odeje in novo zapadlega snega.

- samodejna postaja: meri podobno, kot druge postaje, le da je povsem avtomatizirana.

Dobre lastnosti avtomatične postaje so neodvisnost od opazovalca, subjektivnost opazovalca, slabosti avtomatične postaje so odvisnost od elektrike, višje cene instrumentov in to, da imajo klasične postaje dolg niz reprezentativnih meritev.

Vir ne podaja informacije, ali bi bili podatki z avtomatičnimi senzorji bolj točni od ročnih meritev, podaja pa, da je zaradi dobrih in slabih lastnosti enega in drugega načina merjenja bolje uporabiti oba tipa postaj. Vsekakor je pa treba, da bodo meritve točne, instrumente redno umerjati.

Big data

Big data je izraz, ki označuje zbiranje velikih količin podatkov. Čeprav so ljudje zbirali podatke od začetka zgodovine, se je v zadnjih dveh desetletjih močno povečala hitrost in obseg zbiranja, kar pomeni, da tako velike količine podatkov ni mogoče več obravnavati tradicionalno. Izraz ne označuje samo zbiranje podatkov, ampak tudi procesiranje le-teh.

Mislim, da zbiranje podatkov z obcestnimi senzorji lahko označimo kot big data proces, ker je število senzorjev zelo veliko, velika je tudi frekvenca merjenja in posledično je proizvodnja podatkov ogromna.

Mislim da zbiranje podatkov za vozila spada v big data kategorijo, ker zbiramo mnogo podatkov iz veliko različnih senzorjev bodisi na vozilu bodisi z oddaljenimi, brezžično povezanimi, senzorji. Težava je počistiti te podatke, jih analizirati in jih predstaviti v neko skladno celoto.

Mislim da zbiranje podatke za meteorološka opazovanja ne spada v big data, vsaj če gledamo samodejne postaje predstavljene v viru. Samodejne postaje ne merijo več količin kot klasične vremenske postaje, niti z večjo frekvenco. Torej ni veliko več podatkov v primerjavi s klasičnimi vremenskimi postajami in, z izjemo učinkovitosti ni razloga, da bi podatke obdelovali računalniško in ne analogno kot v preteklosti.

Izjema je uporaba metapodatkov za pridobivanje visoko kvalitetnih dolgih časovnih nizov podnebnih podatkov, kjer računalniki omogočajo avtomatično vrednotenje izmerjenega podatka in ocena vpliva dejavnikov na merjene spremenljivke v času in prostoru, ampak ne mislim, da samo analiza podatkov z računalnikom zadostuje, da primer označimo kot big data.

Aha momenti

Članek o statičnih senzorjih podaja, da je na Nizozemskem približno 60 tisoč senzorjev na 24 tisoč lokacijah, kar se mi zdi ogromno. Nisem vedel, da je promet tako močno opazovan, kot to podaja članek.

Preden sem prebral članek, sem si mislil, da je povezava vozil v omrežje samo neka marketinška ideja, preden mi je članek dal pomisliti, na vse možnosti, ki jih ta infrastruktura odpira.

Nisem vedel, da so metapodatki pomembni pri ocenjevanju meteroloških spremenljivk. Metapodatki npr. lega, nadmorska višina, so pomembni da merimo le sprembo podnebja in vremena, ne pa tudi lega meterološke postaje, vrsta in stanje merilnega sistema, način merjenja in opazovanja meteroloških spremenljivk in datumi sprememb naštetega.

Viri

- Puts, M., Daas, P.J.H., Tennekes, M., & de Blois, C. (2018). Using huge amounts of road sensor data for official statistics, AIMS Mathematics, 4(1): 12–25, doi:10.3934/Math.2019.1.12(fokus na točko 1 in 2)
- McKinsey & Co. (2016). Monetizing car data. New service business opportunities to create new customer benefits. Najdeno na <https://www.mckinsey.com/~media/mckinsey/industries/automotive%20and%20assembly/our%20insights/monetizing%20car%20data/monetizing-car-data.ashx>
- Nadbath, M. et al. (2015). Podnebna spremenljivost Slovenije v obdobju 1961 –2011: Meteorološka opazovanja I. Ljubljana: Ministrstvo za okolje in prostor, Agencija RS za okolje. Najdeno na <http://meteo.arso.gov.si/uploads/probase/www/climate/text/sl/publications/Meteoroloska%20opazovanja%20I%20splet.pdf>(fokus na težave pri merjenju)