

Napovedovanje prihodov avtobusov LPP

Karmen Gostiša (63130057)

25. november 2017

1 Uvod

Cilj naloge je zgraditi čim boljši model linearne regresije, ki bo znal iz časa odhoda avtobusa z začetne postaje napovedati čas prihoda na končno postajo.

2 Ocenjevanje točnosti

Interno sem na učnih podatkih točnost ocenjevala na naslednji način:

1. Učne podatke sem razdelila na dva dela - učni in testni del. Učni del je vseboval vse vrstice, kjer je bil mesec časa odhoda avtobusa (`DepartureTime`) različen od novembra, testni del pa je vseboval samo podatke iz meseca novembra. S tem sem uporabila logiko ocenjevanja na tekmovalnem strežniku, ki je za testne podatke vseboval mesec december.
2. Za vsak testni primer sem napovedala čas potovanja. Pri tem sem uporabljala različne napovedne modele.
3. Izračunala sem čas prihoda avtobusa (`ArrivalTime`) in ga odštela od resničnega časa prihoda. Rezultat sem prištevala spremenljivki `MAE` in jo na koncu delila s številom testnih primerov. Tako sem dobila srednjo absolutno napako.

3 Napovedni modeli

Pri vseh napovednih modelih sem uporabila naslednjo logiko: za vsako smer linije (`RouteDirection`) sem izdelala svoj model linearne regresije s pomočjo priložene kode v navodilih naloge. Za testne primere, ki vsebujejo takšno smer linije, ki je v učnih podatkih ni, sem izdelala še modele s številko linije (`Route`). V kolikor pa je testni primer vseboval tako smer linije kot tudi številko linije, ki ju ni bilo v učnih, pa sem za ta primer za čas potovanja (`TravelTime`) privzela kar povprečje vseh učnih primerov.

DVUP Pričela sem z metodo, ki je posamezno vrstico boljše časovno opisala. Dodala sem ji naslednje izračunane attribute: 7 atributov za dan v tednu (`boolean`), 1 atribut za vikend (`boolean`), 23 atributov za uro v dnevu (`boolean`) in 1 atribut za prometno špico

(`boolean`), kjer sem na podlagi skoraj desetletja izkušenj s prevozi LPP določila prometno najbolj obremenjene ure.

DVUP+M Metodi DVUP sem dodala še 12 atributov za mesec v letu.

DVUP+P Metodi DVUP sem dodala še 15 atributov za praznike oz. dela proste dneve v letu 2012.

4 Rezultati

Ime metode	Oddaja	Ocena s preverjanjem na učnih podatkih	Ocena na tekmovalnem strežniku
DVUP		174.88	187.50
DVUP+M		182.76	187.91
DVUP+P	*	174.38	187.13

Tabela 1: Rezultati napovednih modelov.

Metoda DVUP+P se je izkazala za najboljšo izmed vseh treh in se v povprečju zmotila za 187.13 s. To je za približno minuto bolje od napovedovanja povprečnega časa vsake linije. Metoda DVUP+M deluje slabše kot osnovna različica te metode DVUP in sklepam, da je atributov v tem primeru že preveč in pride do prekomernega prilagajanja modela (angl. *over-fitting*), kar vodi v slabše napovedovanje na novih primerih.

5 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelala sama.