LATVIJAS UNIVERSITĀTE
DATORIKAS FAKULTĀTE

Mašīnmācīšanās metožu lietojumi gēnu ekspresijas analīzei

Autors: Georgijs Pliska
Darba vadītājs: profesors Dr. sc.comp. Juris Vīksna
Latvijas Universitātes Matemātikas un informātikas institūts

PROBLĒMA

RNS Ekspresijas dati - tehnoloģijas labi attīstītas, salīdzinoši lētas, atkārtojami eksperimentu rezultāti, pieejamas daudzas datu kopas.

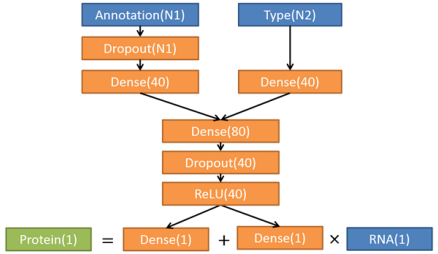
Proteīnu koncentrāciju dati - salīdzinoši jaunas, diezgan dārgas, pagaidām vēl nedaudz datu kopu, atkārtojamība un precizitāte nav skaidra.

Problēma - cik cieši ir saistītas proteīnu koncentrācijas ar gēnu ekspresiju, kas šo saistību ietekmē, kādas ir iespējas proteīnu koncentrācijas prognozēt no ekspresijas datiem?

MĒRĶIS

Par darba pamatu paņemts ~3 g sens pētījums un tā ietvaros izstrādāts DL algoritms, kas pārbaudīts uz tobrīd pieejamajām datu kopām. Mērķi ir pārbaudīt DL algoritma darbību uz jaunākām datu kopām un tūvak izpetīt darba principus.

IZPĒTE UN ANALĪZE



Korelācijas salīdzinājums

	MT1	MT3	MT5	NCI5	NCI60	T13	T29	T43
LR (RNS)	0.62	0.52	0.61	0.54	0.53	0.43	0.49	0.49
DL (RNS)	0.64	0.47	0.62	0.60	0.61	0.52	0.54	0.51
DL (RNS + GO)	0.64	0.47	0.62	0.59	0.60	0.52	0.55	0.50

REZULTĀTI

DL Modelis paradīja sagaidāmu rezultātu ar jauniem un veciem datiem bet ietekme uz rezultāti bija mazāk par prognozējamiem, balstoties uz iepriekšējiem pētījumiem.

SECINĀJUMI

- DL modelim ir potenciāls priekš turpmākas attīstības un izmantošanas nākotnē priekš jauniem pētījumiem.
- Ir nepieciešams dzīļak iepazīties un izpētīt DL modeļi.

TĀLĀKAIS VIRZIENS

Paplašināt testu kopās, uzlabot algoritmu un, ideālā gadījuma, adaptēt programmatūru jaunākai un šobrīd lietotai DL izstrādes videi.