

Rozhodovacie džungle

Mgr. Simon Horvát | vedúci práce, RNDr. Ľubomír Antoni, PhD. | Ústav informatiky PF UPJŠ

Motivácia/Cieľ

Zníženie pamäťových nárokov stromových klasifikačných modelov využitím štruktúry DAG (directed acyclic graphs) a tak zvýšiť použiteľnosť modelu v mobilných alebo zabudovaných procesoroch, kde je pamäť obmedzeným zdrojom.



Úvod do problematiky

V práci rozoberáme rozhodovacie džungle, ktoré boli navrhnuté Shollonom (Microsoft Research Ltd) na konferencii NIPS 2013. Idea rozhodovacej džungle je úzko prepojená so známymi klasifikačnými algoritmi, rozhodovací strom a rozhodovací les, ktoré v práci analyzujeme tiež. Počet uzlov v rozhodovacích stromoch rastie exponenciálne s hĺbkou stromu. Pre niektoré aplikácie, napríklad na mobilných alebo zabudovaných procesoroch, je pamäť obmedzeným zdrojom, a tak exponenciálny rast stromov obmedzuje ich hĺbku, teda aj ich potenciálnu presnosť.

Na rozdiel od rozhodovacích stromov, DAG v rozhodovacej džungli umožňuje viacero ciest od koreňa ku každému uzlu, a teda sa zdá byť sľubnou alternatívou k spomínaným algoritmom, s výhodou ušetrenia pamäte.

Pridaná hodnota

Vlastný model

- Zovšeobecňili sme binárnu džunglu na všeobecný graf

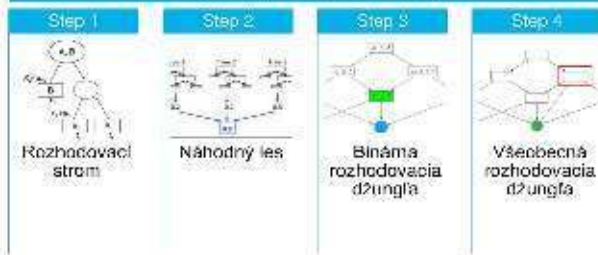
Nový optimalizačný algoritmus

- V práci navrhujeme vlastný optim. algoritmus, ktorý je menej greedy, ako pôvodný

Zdrojový kód

- Súčasnou prácou je zdrojový kód k spomenutým algoritmom i implementácia nášho modelu

Evolúcia acyklických algoritmov

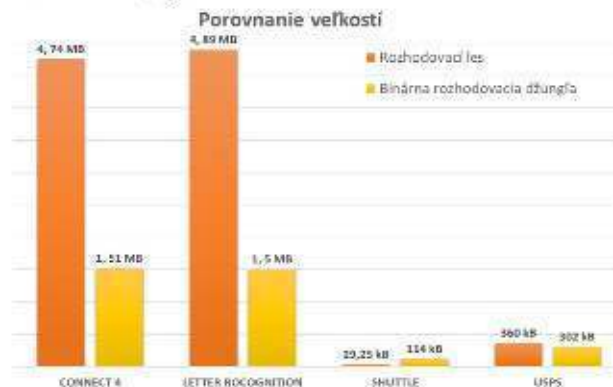


Dáta / Pozorovania

Pôvodný model binárnej rozhodovacej džungle bol testovaný Pohlomom [2] na týchto datasetoch:

Data set	Veľkosť	Dimenzia	Atribúty
CONNECT 4	67 552~	42(126)	kategorické
I FTTPR RECOGNITION	20 000~	16	numerické
SHUTTLE	43 500/14 500	9	numerické
USPS	3 873/1 797	64	numerické

Pohlom experimentálne overil hypotézu, že binárne džungle vedú ušetriť okolo 50% pamäte.



Výsledky

My sme tento experiment simulovali na našej, všeobecnej rozhodovacej džungli.



Na grafe môžeme vidieť ako si v teste vedú rozhodovacie lesy (baseline), binárne džungle od microsoftu a nakoniec náš model všeobecných džunglí.

Zhrnutie

V práci sme sa venovali mimoriadne zaujímavému klasifikačnému modelu rozhodovacích džunglí, ktoré sa zdajú byť výhodnou alternatívou známých klasifikačných acyklických modelov (rozhodovací strom a les).

Navrhli sme vlastný zovšeobecnovaný model, ktorý v priemere, na testovacích dátach, ušetril až 91% pamäte (oproti binárnym džungliam je to 88%).

Works Cited

- [1] Jamie Shollon, Toby Sharp, Pushmeet Kohli, Sebastian Nowozin, John Winn, and Antonio Criminisi. Decision jungles: Compact and rich models for classification. In C.J.C. Burges, L. Bottou, M. Welling, Z. Chahramani, and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems 26, pages 234–242. Curran Associates, Inc., 2013.
- [2] J. Pohlom. Decision jungles. Seminar Report, Fakultät für Mathematik, Informatik und Naturwissenschaften Lehr- und Forschungsgebiet Informatik, 2014.
- [3] K. Komardová. Rozhodovací stromy a lesy, 2012, Orma, ISBN 978 80 7204 785-7.