

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Jakub Arnold
Název práce Bayesovská optimalizace hyperparametrů pomocí Gaussovských procesů
Rok odevzdání 2019
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Milan Straka **Role** Vedoucí
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Cílem diplomové práce je vytvoření praktického nástroje pro optimalizaci hodnot hyperparametrů modelů strojového učení, především neuronových sítí. Většina algoritmů strojového učení využívá hyperparametrů, jejichž hodnota není v průběhu trénování upravována, ať už proto, že to není možné (například neexistuje derivace ztrátové funkce), nebo proto, že pro jejich nastavení není k dispozici dost dat. Nastavování hyperparametrů pak probíhá ve velkém případě ručně, což vyžaduje značné úsilí a nikterak nezaručuje optimalitu výsledků. V některých případech lze použít úplnou enumeraci vybraných hodnot hyperparametrů, což je ovšem výpočetně náročné a v případě komplexnějších modelů to může být neproveditelné. Přitom ovšem korektní nastavení hodnot hyperparametrů je ve většině případů kritické pro dosažení světově nejlepších výsledků.

Diplomová práce přistupuje k optimalizaci hyperparametrů jako k tzv. black-box optimalizaci, a je tedy aplikovatelná k nalezení optimálního řešení libovolného systému, jehož úspěšnost je měřitelná funkce hledaných hyperparametrů. Řešitel konkrétně využívá bezparametrickou bayesovskou optimalizaci Gaussovských procesů, která poskytuje jedny z nejlepších možných výsledků v situaci, kdy je hyperparametrů menší množství (do deseti či dvaceti) a evaluací desítky či stovky, což je nejběžnější scénář při použití hlubokých neuronových sítí.

První dvě kapitoly seznamují čtenáře s úlohou a popisují obecnou variantu bayesovské optimalizace. Třetí kapitola buduje teorii Gaussovských procesů. Následující kapitola se věnuje konkrétním detailům použití bayesovské optimalizace (paralelní evaluace, celočíselné hyperparametry, volba kernelu, implementační detaily). Samotná implementace vytvářeného nástroje je popsána v kapitole páté a kapitola šest vyhodnocuje optimalizaci hyperparametrů v několika úlohách, od triviálních po komplikované hluboké sítě pro značkování a rozpoznávání řeči.

Práci považuji za velmi kvalitní. Teoretická část popisuje bayesovskou optimalizaci a teorii Gaussovských procesů včetně kompletních důkazů a tvoří tak jedinečný studijní materiál na toto téma, který podle mého nejlepšího mínění nemá v existujících publikacích obdoby. Samotná implementace je velmi flexibilní a nabízí širokou funkcionalitu včetně automatického paralelního a distribuovaného vyhodnocování, což je funkcionalita srovnatelná se službami na komerční bázi (např. Google Vizier). Současně poskytuje velké množství vizualizací vlivu hyperparametrů (jak individuálních tak interakcí všech dvojic) na úspěšnost modelu, a to v libovolném čase optimalizace, což nabízí originální způsob interpretace chování hyperparametrů a jejich vzájemných působení.

V neposlední řadě obsahuje práce sadu experimentů, která porovnává kvalitu nalezených hyperparametrů popisovanou metodou s náhodnou a ruční volbou hyperparametrů. Oceňuji, že kromě jednodušších úloh jsou využity hluboké sítě dosahující nejlepších známých výsledků v oblasti zpracování textu. Přestože v tomto případě, kdy již byly hyperparametry určeny ručně na základě rozsáhlé manuální evaluace, je jejich vylepšení velmi obtížné, popisovaná metoda toho dosáhla ve třech ze čtyřech případů.

Celkově hodnotím diplomovou práci jako velmi povedenou a prokazující schopnost samostatné výzkumné činnosti, vzhledem k tomu, že spojuje obtížnou teoretickou stránku, robustní implementaci schopnou využít libovolného množství paralelních či distribuovaných CPU a GPU jednotek, nové vizualizační techniky, a v neposlední řadě experimenty vyhodnocující výsledky na modelech dosahujících nejlepších známých výsledků.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Datum 2. června 2019

Podpis