

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Ondřej Hubsch
Název práce Reducing Number of Parameters in Convolutional Neural Networks
Rok odevzdání 2021
Studijní program Informatika **Studijní obor** Teoretická informatika

Autor posudku Roman Neruda
Pracoviště ÚI AV ČR

Role Oponent

Text posudku:

Předkládaná práce se zabývá tématem redukce parametrů konvolučních neuronových sítí pro klasifikaci obrazových dat. Jde o velmi studované a aktuální téma. Konvoluční sítě jsou dnes nejúspěšnější modely hlubokého učení pro zpracování obrazových dat. Na druhou stranu jde o velké modely s milióny parametrů, takže jejich redukce je důležitá v praktických aplikacích, například při využití v mobilních zařízeních. Autor ve své práci navrhuje několik optimalizačních metod pro redukci parametrů založených na evolučních algoritmech a zpětnovazebném učení a v experimentální části ukazuje jejich efektivitu.

Text vlastní práce je rozdělen do sedmi kapitol. Úvodní kapitola popisuje motivaci problému, definuje cíle práce a představuje její strukturu.

Druhá a třetí kapitola obsahuje úvod do oblasti použitých architektur hlubokých neuronových sítí, včetně konvolučních a rekurentních sítí typu LSTM, a zavádí i optimalizační algoritmy dvou typů: evoluční algoritmy a zpětnovazebné učení (algoritmus REINFORCE). Následuje podrobný přehled souvisejících prací v oborech hledání architektur neuronových sítí, pruningu a extrakce znalostí.

Návrh původních algoritmů autora je obsahem kapitoly 4. Autor definuje prohledávací prostor pomocí konceptu bloků sítí, navrhne redukci bloků a tím vytvoří optimalizační úlohu. Pro její řešení vytvoří několik algoritmů založených na evoluční optimalizaci (kapitola 4.3.2) a posilovaném učení (kapitola 4.3.3). Celý postup redukce parametrů sítě je doplněn o závěrečnou fázi pruningu a dotrénování parametrů (kapitola 4.4).

Pátá kapitola je rozsáhlým ověřením efektivit navržených algoritmů na experimentech se state-of-the-art architekturami konvolučních sítí ResNet a Wide ResNet a databází CIFAR-10. V závěru autor shrnuje a komentuje dosažené výsledky a navrhuje možnosti pokračování v daném tématu.

Hlavní přínosy práce shrnuji v následujících bodech:

- Autor ve své práci navrhl několik různých algoritmů optimalizujících architekturu sítě s ohledem na zachování efektivit sítě a redukci jejich parametrů. Evoluční řešení pracuje buď jako multi-kriteriální s algoritmem selekce NSGA-II nebo agreguje dvě kritéria podle aktuálních přístupů v literatuře. Navíc jsou evoluční řešení rozlišena reprezentací vah (algoritmy Evo a Evo-W). Další řešení využívá zpětnovazebného učení rekurentní sítě, která řídí proces vybírání sítě. Navržené algoritmy navazují na nejnovější výsledky v oblasti.

- Experimentální ověření efektivit algoritmů je velmi pečlivé a zevrubné. Autor porovnává relativní výsledky jednotlivých variant oproti několika baseline přístupům a tam,

kde to je možné, snaží se o srovnání s existujícími přístupy (kapitola 5.8). Experimenty navíc ukazují, že některé varianty autorových algoritmů dosahují srovnatelných až lepších výsledků oproti literatuře.

Následují komentáře a otázky k obhajobě:

- Redukce parametrů má mnoho společného s problémem NAS – obecným hledáním architektur sítí, jak ukazuje například inspirace použitím REINFORCE dle algoritmu ENAS nebo úvaha o nahrazování bloků sítě jinými. Myslíte si, že by šlo využít a upravit i nějaké další přístupy NAS pro redukci parametrů?
- V diskusi zmiňujete možnost automatizace redukce bloků, případně optimalizaci parametru s nastaveného v práci na hodnotu 10. Zkoušel jste například menší hodnotu s?

Závěrem bych rád konstatoval, že práci považuji za velmi kvalitní a vzhledem k dosaženým výsledkům vhodnou k publikaci na mezinárodním fóru. Rád ji doporučuji k obhajobě.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhuje na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 11. 6. 2021

Podpis