

Posudek vedoucí diplomové práce

Název práce: Optimalizace tvorby trénovacího a validačního datasetu pro zvýšení přesnosti klasifikace v DPZ

Autorka práce: Barbora Chytilová

V literatuře je popsáno mnoho přístupů, jak navrhnout trénovací/validační dataset a způsobů dělení terénních dat na trénovací a validační část. Jedním z tvrzení bylo, že pro dosažení maximální přesnosti klasifikace je ideální podíl 1/3 trénovacích a 2/3 validačních dat (Foody, 2009). Literatura uvádí i to, že v případě klasifikace pomocí algoritmu SVM je pro dosažení stejné/podobné přesnosti klasifikace potřeba nižší počet trénovacích bodů než v případě klasifikačního algoritmu Maximum Likelihood (Foody, 2004). Cílem práce bylo testovat měnící se podíl/množství trénovacích a validačních dat a vyhodnotit vliv těchto parametrů na přesnost klasifikace multispektrálních dat senzoru Sentinel-2A a porovnat metody Maximum likelihood a Support Vector Machine. Testy by měly odpovědět na otázku, zda má podíl trénovacích a validačních dat, resp. jejich množství vliv na výsledky hodnocení přesnosti klasifikace.

Barbora Chytilová ke zpracování tématu přistupovala zpočátku do určité míry s despektem, který se ale podařilo do značné míry překonat. Práce má standardní členění. Z jednotlivých kapitol se poměrně dobře podařilo zpracovat část literární rešerše týkající se výběru trénovacích a validačních dat (kapitola 2). Jako modelové území byla vybrána lesně-luční krajina v Podkrkonoší. Jak se v průběhu práce ukázalo, v tomto území nebylo ideální množství a kvalita vodních ploch a výsledky tím byly do určité míry ovlivněny. To ale není problém studentky. Přispělo k tomu asi i suché léto v roce 2016, kdy byl pořízen snímek. Land cover modelového území byl manuálně vektorizován do základních kategorií a posloužil jako referenční data pro hodnocení přesnosti.

Studentka provedla poměrně zajímavý experiment, kdy byl s využitím připraveného skriptu měněn podíl trénovacích a validačních bodů a pro každý podíl bylo provedeno 1000 klasifikací a vyhodnocení přesnosti, aby výsledek nebyl náhodný. Získané výsledky jsou zajímavé, v prvním experimentu bylo skutečně dosaženo nejvyšší přesnosti klasifikace pro podíl 1/3 trénovacích a 2/3 validačních dat. Dále se ale ukázalo, že změna velikosti validačního datasetu při zachování stabilní velikosti trénovacího datasetu má vliv na stabilitu výsledku hodnocení přesnosti klasifikace. Byl také potvrzen předpoklad, že pomocí algoritmu SVM je pro dosažení stejné/podobné přesnosti klasifikace potřeba nižší počet trénovacích bodů než v případě klasifikačního algoritmu Maximum Likelihood.

Celková úroveň práce trpí určitými nedostatky. Jedna ze stěžejních částí práce – Diskuse – je do určité míry spíše souhrnem vlastních výsledků a zamýšlením se nad nimi. Větší pozornost mohla být věnována diskusi výsledků s literaturou. Po formální stránce lze práci vytknout řadu drobných nedotažení a nepřesností, jako například popisky os některých grafů. Problematické jsou mnohé formulace a úroveň odborného vyjádření. To jsme s autorkou řešily už v průběhu práce a zcela se nepodařilo tyto problémy odstranit. Zmíněné nedostatky snižují úroveň a hodnotu práce, která je po experimentální stránce kvalitní.

Autorka vyvodila správné závěry opřené o robustní výsledky. Cíle práce byly splněny a práci Barbory Chytilové doporučuji přijmout jako diplomovou práci. Vzhledem k uvedeným nedostatkům hodnotím známku **velmi dobře** až **dobře** (dle obhajoby).



V Líbeznicích, 4. září 2019

RNDr. Lucie Kupková, Ph.D.