Oponentní posudek diplomové práce

Klasifikace vybraných zemědělských plodin v modelovém území Kutnohorska s využitím časové řady dat Sentinel-2 a PlanetScope

Autor: Bc. Tomáš Kuthan
Vedoucí práce: RNDr. Lucie Kupková, Ph.D.

Předložená diplomová práce zpracovává téma využití časových řad snímků pořízených z recentních systémů DPZ poskytujících data vysokého časového i prostorového rozlišení pro klasifikaci sedmi vybraných zemědělských plodin. Cílů si autor klade několik, a to analýzu spektrálních charakteristik vybraných plodin a následné porovnání přesnosti jejich klasifikace z mono- a multitemporálního kompozitu multispektrálních dat Sentinel-2 a PlanetScope. Tato data se liší svým prostorovým rozlišením (10 m a 20 m pro Sentinel-2 vs. 3,125 m pro PlanetScope) a spektrálním rozlišením (13 spektrálních pásem Sentinelu-2 vs. 4 pásma PlanetScope). Dalším cílem je pak porovnat přesnost výsledků řízené klasifikace založené na algoritmu maximum likelihood (ML) a support vector machine (SVM) na pixelové úrovni a klasifikace založené na objektovém přístupu. Zvolené téma považuji za aktuální a cíle v obecné rovině odpovídající úrovni magisterské práce.

Textová část zabírá 84 stran včetně obrázků a tabulek a je tradičně členěna do osmi kapitol. Autor pracuje s velkým množstvím převážně zahraniční literatury, kterou v práci řádně cituje. Úvodní kapitola seznamuje s důvody výběru tématu, jasně vytyčuje cíle práce a stanovuje hypotézu, že využití multitemporálních dat pořízených během jedné sezóny by mělo vést ke zvýšení přesnosti klasifikace vybraných plodin. Kapitola Rešerše a úvod do problematiky v prvních čtyřech sekcích seznamuje s pojmy a přístupy spojenými se zpracováním multispektrálních a multitemporálních dat (korekce, analýza hlavních komponent, NDVI, klasifikace). V několika případech se zde vyskytují neobratná či zavádějící vyjádření jako např. v případě atmosférických korekcí, kdy z textu vyplývá, že poloha slunce a intenzita slunečního záření patří mezi vlivy atmosféry. Geometrické korekce sice „transformují soustavu obrazového záznamu“, ale bez jediného příkladu či bližšího vysvětlení toto vyjádření čtenáři podstatu neobjasňuje. V případě algoritmů pro atmosférické korekce jsou v rešerši zmiňovány dva zdroje, které se neshodují v hodnocení modulu Sen2Cor v případě travnatých porostů. Vzhledem k tomu, že je později tento modul dále využíván pro předzpracování dat Sentinel-2, předpokládala bych v rešerši další práce, které tuto neurčitost objasní. V případě klasifikátoru ML není správné izolinie stejných pravděpodobností v n-rozměrném příznakovém prostoru nazvat elipsami (str. 20). Vysvětlení přístupů klasifikace SVM „one against the rest“ a „one against one“ na str. 21 není srozumitelné. Sekce 2.5 týkající se definice fenologie plodin, jejího mapování a vztahu ke spektrálním příznakům je kvalitně zpracovaná. Stejně tak k následující a pro práci stěžejní sekci 2.6, která shrnuje studie o klasifikaci zemědělských, plodin nemám zásadní připomínky. Pouze bych nevyčleňovala termíny ozim a jařina do zvláštních podsekcí; některé převzaté obrázky mají sníženou čitelnost.

Třetí kapitola stručně popisuje modelové území z hlediska fyzicko-geografického a využití půdy. Čtvrtá kapitola přehledně popisuje použitá data DPZ a ostatní data, zejména databázi LPIS. Kapitola Metodika je rozdělena do čtyř sekcí, z nichž první je věnována předzpracování dat. Autor objasňuje, že pro data Sentinel-2 použil modul Sen2Cor. U dat PlanetScope zmiňuje atmosférickou korekci v ENVI pro jeden termín. Byla data nějak upravována v případě multitemporálního kompozitu 28 pásem? Popis jednotlivých plodin bych doporučovala nadřadit do sekce 5.1.3 a dále bych ji nedělila na podsekce. V sekci 5.1.4 se nachází tabulka 7 (str. 48) obsahující sedm kombinací původních pásem použitých snímků a jejich derivátů. Bylo by vhodné vzhledem ke všem použitým variantám, které se ještě navyšují s počtem použitých klasifikací, v úvodu metodické části čtenáře krátce seznámit s tím, jak se budou pásma a metody kombinovat a proč. Počet pásem uvedený v tabulce 7 je nekonsistentní (pokud se jednalo u multitemporální komposit, mělo by být 24 „pásem“ u PCA datasetů). Kvalitně je popsán výběr trénovacích a kontrolních ploch. V práci by mělo být vyjádřeno, jaká konkrétní metrika byla použita pro výpočet separability trénovacích množin. Druhá část metodiky se týká analýzy spektrálních příznaků a klasifikace. V části 5.2.1 jsem čekala více než konstatování,

že pro analýzu spektrálních příznaků bude využito průměru a NDVI. V části 5.2.2 jsou uvedeny použité kombinace datasetů a klasifikačních metod. Pro přehlednost by bylo vhodné doplnit slovní popis o schéma. V rozporu se zadáním je aplikace metody ML na všechny termíny snímků Sentinel-2. Neměl být na základě spektrální analýzy vybrán termín, kdy jsou jednotlivé plodiny nejlépe odlišitelné a pak teprve provést klasifikaci? Oceňuji i) snahu o optimalizaci výběru parametrů klasifikace, ač z ní vyplývá, že nejlepších výsledků bylo dosaženo při defaultním nastavení, ii) testování normality rozdělení trénovacích množin, ač by se zde nabízelo statistické testování hypotézy o normalitě daného souboru než použití pouze grafického vyjádření. Zbytek kapitoly popisuje metodu hodnocení klasifikace a použitý software.

Šestá kapitola podrobně seznamuje se získanými výsledky. První sekce v grafické formě ukazuje průběh NDVI odvozený z obou typů dat a hodnot odrazivosti (v grafech je ponecháno DN hodnoty) v pásmech R, G, B a NIR dat Sentinel-2 pro každou plodinu. Grafy jsou podrobně komentovány. Zásadní závěry vidím v konstatování, že během vegetační sezóny vykazují data Sentinel-2 a PlanetScope srovnatelné hodnoty NDVI. Co se týká oddělitelnosti tříd, pak z grafického porovnání vycházejí stejné termíny jako při použití míry separability trénovacích tříd. Otázkou je, zda by tedy tato metrika nebyla pro získání odpovědi dostatečná než hůře uchopitelné množství grafů. Za přínosný považuji graf 12 znázorňující odvozené fenofáze jednotlivých plodin. K doplnění popisu z předchozích kapitol by bylo názorné na příkladu jedné plodiny ukázat, jak byl graf z hodnot NDVI konstruován. Další sekce seznamuje s výsledky jednotlivých klasifikací. Vzhledem k již zmiňovanému počtu kombinací je poněkud obtížné se v textu orientovat. Kapitola je doplněna mapovými výstupy zobrazujícími výsledek jednotlivých klasifikací, které bych zařadila spíše do příloh. Diskuse k získaným výsledkům je nadstandardně zpracovaná, co se týká obsahu i rozsahu. Detailně se vrací k použitým metodám, hodnotí výsledky a porovnává s literaturou. Odpovídá na otázky, proč data Sentinel-2 vykazují lepší výsledky klasifikace vybraných plodin oproti datům PlanetScope, hodnotí výhody multitemporálního přístupu. Závěrem jsou shrnuty výsledky práce a nastíněno možné rozšíření řešení o kombinaci s daty Sentinel-1, což je ale již na jiných pracovištích průběžně řešeno.

Po formální stránce nemám k práci kromě již zmíněných drobností zásadní připomínky, pouze upozorňuji na některé nevhodné překlady a slovní kombinace jako *preparation* – příprava, nikoli „preparace“ půdy (str. 26), senzory satelitních družic (str. 31), „vegetační indicie“ – mělo se jednat o překlad *vegetation indic*es = vegetační indexy (str. 34). Dále mám několik otázek do diskuse:

- V případě porovnávání výsledků klasifikace ze dvou typů dat je vhodné použít stejný počet komponent nebo stejné % informace, které nesou? Viz str. 45 a tabulka 5.
- Mohl byste vysvětlit metodu PCA použitou při objektové klasifikaci (str. 54)?
- Bylo možné u objektového přístupu použít metodu ML (pro lepší porovnatelnost objektového a pixelového přístupu)? V diskusi mi chybí stručné porovnání výsledku klasifikace SVM získaného pixelovým a objektovým přístupem.

Diplomová práce Bc. Tomáše Kuthana splnila předložené cíle, a proto ji **doporučuji** k obhajobě. Vzhledem k výše uvedeným připomínkám navrhuji hodnocení **velmi dobře**.

V Kutné Hoře, 17.1.2019 Ing. Markéta Potůčková, Ph.D.