

Určení druhové skladby lesa z družicových dat

Zadáním k vypracování této diplomové práce bylo přispět k poznatkům o metodických postupech používaných při určování výskytu zvolených druhů stromů v modelovém lesním porostu z optických družicových dat. Tato dlouhodobě zkoumaná problematika je bohatě rozpracovaná a proto bylo k dosažení zadaného cíle získat znalosti o aktuálních metodách, které se k řešení tohoto tématu používají. Posoudit současné způsoby klasifikace a specifikovat procesy ovlivňující kvalitu výsledku klasifikace vhodných dat.

To znamenalo provést literární rešerši s dostatečně širokým záběrem prací. Student tomuto úkolu věnoval nemalé úsilí, jehož výsledkem je pětadvacetistránkový souhrn relevantních metodických přístupů. Jejich rozбором byly zformulovány argumenty pro rozhodnutí použít k přípravě zpracovatelského postupu metodu Random Forest.

Popis principu i postupu aplikace metody Random Forest je rovněž věnována nemalá část práce. Do určité míry může posloužit k seznámení s touto metodou i případným novým uživatelům. Text však postrádá potřebná vysvětlení pro konstatování ocitovaná z několika prací. To na jedné straně sťažuje srozumitelnost a pochopení popisovaného postupu, na druhé straně by podrobnější výklad vyžadoval jít za rámec diplomové práce.

Užitečný je popis situace v českých lesích z hlediska druhového složení a jeho prostorového rozložení po území státu. Pro zkoumání klasifikační metody byly vybrány čtyři druhy (smrk, borovice, buk a dub) představující v souhrnu více než 80% všech lesních dřevin. Jednotlivé druhy homogenně pokrývají jen malá území a v malém počtu výskytů, převážně jde o různé prostorově smíšené lesní útvary, nicméně s definovanou dominancí dvou ze čtyř základních dřevin. Stanovení testovacích území bylo vedle požadavku dostatečně velké souhrnné plochy výskytu klasifikovaných objektů, vedeno i podmínkou jejich umístění na jedné družicové scéně. Za srovnávací referenční informaci posloužila databáze porostových typů lesa ÚHUL. Pro každý ze čtyř typů bylo vybráno 6 lokalit s čistým zastoupením příslušné dřeviny v severozápadní části Čech.

Použitá data pořídily skenery na družicích Landsat 8 a Sentinel 2. Ke zpracování byly vybrány od každého nosiče 3 letní scény s minimálním výskytem oblačnosti, v případě scén Landsatu ve dvou úrovních předzpracování. Data byla získána z obvyklých veřejných serverů. U dat družice Sentinel je zmíněn proces mozaikování dlaždic do scén, chybí však bližší vyjádření zda byla tato úprava provedena po získání dat nebo již před tím.

Vlastní postup vedoucí ke klasifikaci čtyř druhů lesa je popsán v kapitole 5. K jejímu obsahu jsou zaměřeny i následující otázky a komentáře.

Kapitola 5.1.3 popisuje formátování dat což je tvorba vektorových polygonů přičemž hranice lesa prochází pixelem družicových dat. Je to v relaci s přesností prostorového ztotožnění obou vrstev?

Při informování o přejímání závěrů z literatury jsou použity zkratkovitě popoisy, kterých chybí alespoň minimální objasnění, např. u zkratky AdaBoost na str. 61.

Bylo provedeno testování různých hodnot jednotlivých parametrů použitých v metodě RF. Průběh změny přesnosti klasifikace v závislosti na parametru $N_{estimator}$ pro jednotlivé scény (tab.10, str. 64 a následné grafy) je jiný než u srovnání téže závislosti pro dvě hodnoty parametru $criterion$ (Tab. 11, str. 66).

Čím se dá vysvětlit odlišný průběh závislosti přesnosti klasifikace na parametru $min_samples_leaf$ pro různě sestavené trénovací množiny (graf 6, str. 68), a stejně i u parametru $min_samples_split$ (graf 7, str. 69).

Byly množiny vybírány z jedné scény nebo z několika scén? U ověřování vlivu velikosti trénovací množiny na přesnost klasifikace není jasné co je míněno označením "kvalifikovaný výběr" a z čeho vychází

konstatování, že 1536 pixelů je dostatečná velikost trénovací množiny. Pro vlastní klasifikaci byla použita trénovací množina o velikosti 2560 pixelů pro Landsat a 27478 pro Sentinel. Složení trénovací množiny sledováno není.

V korelační matici jednotlivých spektrálních kanálů na str. 72 není definován parametr "Třída".

V následující kapitole jsou souhrnně uvedeny dosažené výsledky. U výsledků klasifikace je v odstavci 6.4 uvedeno, že nejvyšší byla přesnost u třídy smrk, čemuž ale neodpovídají hodnoty v tabulce 17.

Při popisu vlivu spektrálních kanálů na výsledek klasifikace není vysvětleno, jak byly získány hodnoty v tabulce 22.

V závěrečné kapitole jsou klasifikační výsledky porovnány s hodnotami dosažených v publikovaných pracích se závěrem, že výsledky dosažené v této práci jsou lepší. To je jistě dobrý výsledek podporující zvolenou modifikaci metody RF, i když není blíže uvedeno do jaké míry je tato přednost limitována dalšími podmínkami měření.

Práce obsahuje pasáže které patrně již neprošly závěrečnou formulační a redakční kontrolou, ale bez závažných obsahových chyb. Práce je ve svém celku nadprůměrně rozsáhlá. Vedle vlastního textu obsahuje čtyřicet stránek přílohy s tabulkami výsledků klasifikace a mapovými výstupy klasifikace všech testovacích ploch a také skripty připravenými pro úpravu dat i klasifikačního programu RF v programovém balíku ESRI. Přináší vlastní pohled na komplexní problematiku, a současně i nemálo otázek, které by si zasloužily zmínky v hodnocení, např. komentář či vysvětlení některých mezních hodnot ve výsledcích.

V souhrnu práce splňuje požadavky na diplomovou práci a hodnotím ji stupněm velmi dobře.

17.5.2018

Doc.Ing. Jan Kolář, CSc
vedoucí diplomové práce