
dr hab. inż. Wojciech Kotłowski
Instytut Informatyki (Institute of Computing Science)
Politechnika Poznańska (Poznań University of Technology)
ul. Piotrowo 2, 60-965 Poznań
tel: (+48) 61 665 2936
wkotlowski@cs.put.poznan.pl



Poznań, 20 stycznia 2019 roku

Recenzja rozprawy doktorskiej
mgr. inż. Michała Romaszewskiego
Nowe metody klasyfikacji obrazów hiperspektralnych
(*New methods for hyperspectral image classification*)

1 Wprowadzenie

Rozprawa mgr. inż. Michała Romaszewskiego dotyczy problemu obrazowania hiperspektralnego (ang. *hyperspectral imaging, HSI*), techniki rejestracji obrazów w bardzo szerokim zakresie widma elektromagnetycznego, w której obraz składa się wielu (nawet setek) kanałów, odpowiadających różnym przedziałom długości fal. Dzięki coraz większej dostępności odpowiednich urządzeń rejestrujących, obrazowanie hiperspektralne dynamicznie rozwija się i znajduje coraz liczniejsze zastosowania. W tym sensie recenzowana rozprawa wydaje się być więc zdecydowanie na czasie i bardzo dobrze wpisuje się w rozwój całej dziedziny.

Autor skupił się podczas swoich studiów doktoranckich na konstrukcji algorytmów służących do klasyfikacji obrazów hiperspektralnych. Przez klasyfikację rozumiemy tu nadanie właściwej etykiety każdemu pikselowi na obrazie. Przykładowo, w obrazach obszarów rolniczych etykiety odpowiadałyby różnym rodzajom rosnących tam roślin, a w przypadku obszarów miejskich etykietami mogłyby być typy materiałów znajdujących się w danym miejscu, np. metal, cegły, ziemia, drzewa, itp. Naturalnym pomysłem na budowę algorytmów klasyfikacji, i taki właśnie pomysł realizowany jest w całej pracy, wydaje się zastosowanie tutaj paradygmatu uczenia maszynowego: zamiast próbować ręcznie utworzyć zestaw reguł pozwalających na przypisanie właściwej etykiety danemu pikselowi na podstawie jego położenia czy wykresu spektralnego, co byłoby zadaniem bardzo trudnym lub wręcz niemożliwym, algorytm powinien *nauczyć się* poprawnie klasyfikować obraz na podstawie przekazanej mu niewielkiej liczby przykładów poprawnej klasyfikacji, czyli pikseli, którym zewnętrzny ekspert był w stanie nadać właściwe etykiety. Zbiór takich przykładów nazywa się zwykle zbiorem treningowym. Celem algorytmu jest więc uogólnienie informacji zawartej w zbiorze treningowym tak, aby na jej podstawie poprawnie sklasyfikować wszystkie pozostałe piksele. Jest to zadanie o tyle trudne, że liczba wszystkich pikseli może z łatwością sięgać milionów, podczas gdy realnie możemy wymagać od eksperta uprzedniej klasyfikacji kilku do kilkudziesięciu pikseli na etykietę. Rzeczywiście, z punktu widzenia uczenia maszynowego uogólnienie z tak małego zbioru treningowego jest sporym wyzwaniem i z pewnością trzeba posiłkować się dużą dozą wiedzy dziedzinowej. Takie właśnie podejście, silnie oparte na założonych regularnościach występowania etykiet na obrazie, przyjmuje doktorant w swojej pracy.

Z matematycznego punktu widzenia, obraz hiperspektralny można wyobrazić sobie jako dwuwymiarową macierz pikseli, ale w odróżnieniu od klasycznego obrazu, w którym piksele posiadają tylko trzy kanały (czerwony, zielony, niebieski), liczba kanałów d jest bardzo duża, stąd piksele można traktować jako wektory w wysoko-wymiarowej przestrzeni. Daje to ciekawy z punktu widzenia

Ułan

reprezentacji problem: z jednej strony obraz może być traktowany jako zbiór d -wymiarowych wektorów spektralnych, indeksowanych kolejnymi numerami pikseli; z drugiej strony, możemy myśleć o obrazie hiperspektralnym jako o zbiorze d jedno-kanałowych obrazów. Pierwsza z perspektyw nazywana jest zwykle *spektralną*, zaś druga – *przestrzenną*.

Naturalną rzeczą wydaje się użycie obu perspektyw do konstrukcji algorytmów klasyfikujących obrazy. W przypadku perspektywy spektralnej można wykorzystać obserwację, że piksele o tych samych etykietach mają zwykle podobne przebiegi spektralne. Podobnie, w przypadku perspektywy przestrzennej wiedza dziedzinowa podpowiada nam, że etykiety nie zmieniają w sposób zupełnie nieuporządkowany, a raczej tworzą spójne i zwarte obszary na obrazie, często sporych rozmiarów. Wynika z tego, że dany piksel ma spore szanse (ile akurat nie leży na granicy danego obszaru) mieć taką samą etykietę, jak jego sąsiedzi. Podejście do uczenia maszynowego, w którym wykorzystuje się więcej niż jedną perspektywę nazywa się często podejściem *wielo-widokowym* (ang. *multi-view*). Co więcej, ponieważ już w trakcie procesu uczenia mamy do dyspozycji piksele znajdujące się poza zbiorem treningowym, czyli te, które chcemy sklasyfikować, mamy de facto do czynienia również z tzw. problem uczenia pół-nadzorowanego (ang. *semi-supervised learning*), w którym obok zbioru etykietowanych przykładów uczących algorytm ma do dyspozycji również (często znacznie obszerniejszy) zbiór przykładów bez etykiet.

Autor pracy słusznie zauważa, że skoro mamy do dyspozycji wiele widoków oraz ogrom danych nienadzorowanych, warto z tych cennych informacji w procesie uczenia skorzystać. Stąd główna teza zawarta w pracy brzmi: *Stosowanie wielo-widokowego i pół-nadzorowanego uczenia maszynowego polepsza trafność algorytmów klasyfikujących obrazy hiperspektralne*. W mojej opinii cele pracy są dobrze uzasadnione, interesujące, oraz wystarczająco ambitne, aby stały się przedmiotem rozprawy doktorskiej.

2 Ocena struktury i zawartości pracy

Recenzowana rozprawa jest napisana sprawnie w języku angielskim. Mimo sporej liczby wprowadzanych pojęć i wielu wyników formalnych (definicji, dowodów, twierdzeń), oraz złożonej analizy eksperymentalnej, czyta się ją, z pewnymi wyjątkami, bardzo dobrze. Liczy 130 stron i składa się z sześciu rozdziałów. Całościowo jej strukturę oceniam bardzo pozytywnie, a sam podział na rozdziały wydaje się zupełnie naturalny i rozsądny. Narracja już od wprowadzenia do pracy sprawia u czytelnika zainteresowanie tematem. Poniżej omówię zawartość poszczególnych rozdziałów, równocześnie starając się dokonać ich merytorycznej oceny.

Rozdział pierwszy ma charakter wprowadzenia. Opisany w nim został sam problem obrazowania hiperspektralnego oraz wyzwania, które ze sobą niesie. Następnie autor przybliżył definicje istotnych dla pracy problemów uczenia maszynowego: uczenia nadzorowanego, uczenia półnadzorowanego, transdukcji oraz uczenia wielo-widokowego. Definicji zostały przedstawione w sposób formalny i nie budzący żadnych wątpliwości. Na koniec opisana została struktura dalszych rozdziałów pracy i pojawiły się również dwa główne cele rozprawy:

- konstrukcja nowych algorytmów klasyfikacji hiperspektralnej,
- eksperymentalna i teoretyczna analiza pół-nadzorowanego, wielo-widokowego podejścia do klasyfikacji hiperspektralnej.

Oba cele oceniam na zasadne i ciekawe.

Rozdział drugi poświęcony jest pierwszemu z wymienionych powyżej celów, gdyż dotyczy pół-nadzorowanych algorytmów klasyfikacji hiperspektralnej przy użyciu zbioru treningowego niewielkich rozmiarów. Rozpoczyna się od ogólnego opisu zaproponowanej metodologii, inspirowanej się rozwiązaniem problemu tzw. długoterminowego śledzenia (ang. *long-term tracking*) w strumieniach wideo, o nazwie TLD. Autor zauważył bardzo interesującą analogię do klasyfikacji hiperspektralnej i postanowił przedstawić problem jako uczenie się z dwoma typami ekspertów: P oraz N. Skrótory biorą się od słów „pozytywny” i „negatywny”, i w kontekście ustalonej klasy (etykiety) dotyczą decyzji o klasyfikacji danego piksela do tej klasy (klasyfikacja pozytywna) lub do klasy przeciwnej (klasyfikacja negatywna). Ekspert P jest optymistyczny i skupia się na pominiętych przykładach potencjalnie pozytywnych, natomiast ekspert N jest pesymistą i odrzuca przykłady sklasyfikowane jako pozytywne, co do których ma wątpliwości.

Pomysł znakomicie pasuje do klasyfikacji hiperspektralnej, ponieważ ekspert P, wykorzystując zjawisko grupowania się etykiet w spójne obszary na obrazie, będzie starał się nadać danym pikselom etykietę podobną do tej, którą posiadają ich sąsiedzi w sensie przestrzennym, natomiast ekspert N będzie weryfikował podobieństwo piksela do innych z proponowanej klasy w sensie spektralnym, co może skutkować odrzuceniem potencjalnych kandydatów wspieranych przez eksperta P. Sam algorytm przebiega w kilku powtarzających się etapach, w których obaj eksperci oceniają piksele bez etykiet w kontekście przynależności do różnych klas, a następnie podzbiór tych pikseli wspierany przez eksperta P i nie budzący wątpliwości eksperta N zostaje etykietowany klasą o największej wartości funkcji oceniającej przynależność. Zaproponowany algorytm (nazwany przez autora PNGrow) został obszernie przetestowany na wielu rzeczywistych zbiorach danych, oraz porównany z najlepszymi istniejącymi metodami klasyfikacji hiperspektralnej. Algorytm PNGrow wyszedł z tego porównania zwycięsko, zwykle wygrywając pod względem trafności klasyfikacji z konkurentami. Jestem pod wrażeniem umiejętności doktoranta w zakresie analizy eksperymentalnej, która zawiera między innymi liczne wizualizacje działania algorytmu i przydziału etykiet do pikseli, scenariusze testowe modyfikujące zbiory danych celem zmniejszenia ich regularności, aby sprawdzić działanie algorytmu w trudniejszych warunkach, analizę zależności trafności klasyfikacji względem wielkości początkowego zbioru treningowego, oraz skrupulatną analizę statystycznej istotności uzyskanych wyników. Jedyne, czego zabrakło mi w eksperymencie, to porównanie złożoności obliczeniowej PNGrow z konkurencją, choć brak takiej analizy jest zapewne spowodowana wzięciem wyników innych algorytmów z poprzednich prac (domyślam się, że nie łatwo byłoby znaleźć ich implementację, której można by samodzielnie użyć). Nie zmienia to jednak mojej wysokiej oceny wyników tego rozdziału.

W rozdziale trzecim, najbardziej teoretycznej części pracy, doktorant wprowadza matematyczny model problemu klasyfikacji pół-nadzorowanej z wieloma widokami w postaci reprezentacji grafowej: każdy z widoków związany jest z pewnym grafem, które wierzchołki to przykłady uczące, a połączenia między nimi związane są z ich podobieństwem w obrębie danego widoku. Następnie opisany zostaje schemat algorytmów klasyfikacji, które oparte są na pomysłach losowego spaceru po takim grafie: wierzchołki odwiedzane są z prawdopodobieństwami proporcjonalnymi do podobieństw między sąsiadami, można więc policzyć prawdopodobieństwa trafienia do danego wierzchołka po n krokach wychodząc od pewnego rozkładu początkowego na podzbiorze wierzchołków związanym z daną etykietą. Te końcowe prawdopodobieństwa są w stanie wychwycić swego rodzaju sumaryczne podobieństwo wszystkich wierzchołków w grafie do poetykietowanego podzbioru. Następnie algorytm łączy ze sobą podobieństwa wynikające z poszczególnych grafów (widoków), np. za pomocą średniej ważonej (kombinacji wypukłej). Wierzchołki o najwyższych wartościach takiego łącznego podobieństwa otrzymują etykiety związane z klasą, do której są najbardziej podobne. Pomysł taki uważam za interesujący, a jednocześnie bardzo ogólny, pozwalający reprezentować w ten sposób wiele możliwych wariantów algorytmów klasyfikacji.

Dalsza część rozdziału skupia się na teoretycznej analizie modelu: warunkach koniecznych na przypisanie etykiety wszystkim przykładom uczącym, analizie sytuacji, w której graf składa się ze skupisk wierzchołków przynależnych do tej samej klasy, zdefiniowania warunków koniecznych i wystarczających na poprawną klasyfikację, oraz przedstawienia eksperymentalnego podejścia pozwalającego na zastosowanie i weryfikację powyższych rezultatów. Wyniki te są ciekawe i znaczące, mam jednak kilka uwag krytycznych dotyczących ich przejrzystości i sposobu ich przedstawienia, które zostały opisane w części „Uwagi drobne”. Chodzi tu przede wszystkim o pewne trudności przy śledzeniu samych twierdzeń i ich dowodów, oraz pojawiające się gdzieś niejasności. Ogólność uzyskanych wyników zmniejszają również pojawiające się czasem dość mocne założenia dotyczące struktury grafu lub samego algorytmu (wspomniane w części „Uwagi dyskusyjne i polemiczne”). Nie zmienia to faktu, że uważam wszystkie wymienione rezultaty za wartościowe.

Rozdział czwarty to eksperymentalna weryfikacja wyników uzyskanych w rozdziale poprzednim. W szczególności, celem jest zastosowanie warunków koniecznych i wystarczających na przydzielenie poprawnych etykiet do problemów klasyfikacji hiperspektralnej, oraz przetestowanie algorytmu opartego na spacerach losowych po grafie. Autor wprowadza dwie miary, SCM (ang. *sufficient condition measure*) i ESCM (ang. *extended sufficient condition measure*), które pozwalają potencjalnie zagwarantować poprawną klasyfikację wszystkich pikseli. Jest to jednak obwarunkowane bardzo silnymi założeniami, np. że wartości miary są dodatnie dla każdego piksela na obrazie, lub że pikselowi przydzielono etykietę w pierwszej iteracji, co w mojej opinii może powodować, że uzyskane gwarancje mogą zachodzić stosunkowo rzadko w praktycznych problemach obrazowania. Natomiast druga część rozdziału, czyli testy algorytmu opartego na spacerach losowych po grafach, nie

budzi zastrzeżeń. Z testów wynika, że algorytm ten radzi sobie lepiej, niż konkurencyjne algorytmy klasyfikacji hiperspektralnej, natomiast, co ciekawe, przegrywa z „wewnętrzną konkurencją” czyli algorytmem PNGrow z rozdziału drugiego. Nie zmienia to oczywiście wynikającego z testów wniosku, że autor rozprawy w jej obrębie wprowadził dwa najlepsze algorytmy klasyfikacji hiperspektralnej. Zgadzam się również z autorem, że algorytm oparty na spacerach losowych ma spory potencjał i dzięki odpowiedniemu doboru jego komponentów (grafu podobieństwa, funkcji określającej liczbę wierzchołków do etykietowania, itp.) jego wyniki z pewnością można by polepszyć. Chciałbym również w tym miejscu ponownie pochwalić doktoranta za bardzo dobrą i obszerną analizę eksperymentalną, badającą trafność algorytmu względem doboru różnych jego parametrów i względem wielkości początkowego zbioru treningowego.

Rozdział szósty dotyczy praktycznych zastosowań klasyfikacji hiperspektralnej. Podzielony jest na trzy zasadniczo niezależne części, związane z trzema publikacjami doktoranta. W pierwszej części autor opisuje ciekawy pomysł algorytmu, który w odróżnieniu od poprzednio rozważanego paradygmatu wyboru pikseli najbardziej podobnych w obu widokach, bazuje na niezgodnościach między widokami i rozszerza zbiór treningowy poprzez dobór kandydatów najbardziej niezgodnych. Pomysł okazał się bardzo trafny, ponieważ powoduje on wzrost trafności klasyfikacji algorytmów użytych w dalszym etapie na w ten sposób rozszerzonym zbiorze treningowym. W drugiej części opisano użycie algorytmów klasyfikacji do detekcji pozostałości po strzale z broni palnej, co ma istotne zastosowanie w analizach sądowych. Pomimo trudnego problemu z nieregularną strukturą poszukiwanego sygnału na obrazie autor uzyskał całkiem dobre wyniki trafności klasyfikacji. W trzeciej części rozdziału opisano problem selekcji pasm w widmach obrazów hiperspektralnych, w których wzorec, którego należy odszukać, jest znacznie mniejszy niż inne wzorce pojawiające się na obrazie, a własności materiałów związanych z poszukiwanym wzorcem przejawiają się jedynie w niewielkiej liczbie pasm widma. Użyte zostały tutaj metody bazująca na kumulantach wyższego rzędu. Z wyników eksperymentu na rzeczywistym obrazie dotyczącym wydobycia materiałów wyraźnie widać, że zaproponowane algorytmy osiągnęły sukces, zwiększając trafność klasyfikacji w stosunku do metody referencyjnej.

W mojej opinii ten rozdział w pełni potwierdza, że autor posiada umiejętność konstrukcji i stosowania metod klasyfikacji hiperspektralnej bardzo dobrze sprawdzających się w praktyce.

Ostatni rozdział pracy to podsumowanie, zawierające wnioski z badań naukowych związanych z doktoratem oraz potencjalne kierunki dalszych badań nad klasyfikacją hiperspektralną.

3 Ocena wkładu oryginalnego

Rozprawa jest oparta na czterech artykułach, których współautorem jest doktorant:

- Romaszewski, M., Głomb, P., i Cholewa, M. (2016). Semi-supervised hyperspectral classification from a small number of training samples using a co-training approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 121:60 – 76
- Głomb, P., Romaszewski, M., Cholewa, M., i Domino, K. (2018b). Application of hyperspectral imaging and machine learning methods for the detection of gunshot residue patterns. *Forensic Science International*
- Głomb, P., Domino, K., Romaszewski, M., i Cholewa, M. (2018a). Band selection with higher order multivariate cumulants for small target detection in hyperspectral images. *arXiv:1808.03513v1*
- Cholewa, M., Głomb, P., i Romaszewski, M. (2018). A spatial-spectral disagreement-based sample selection with an application to hyperspectral data classification. *Goescience and Remote Sensing Letters*, Accepted for publication

Dwa pierwsze artykuły zostały opublikowane, a czwarty przyjęty do druku, w renomowanych czasopismach (o współczynnikach *Impact Factor* wynoszących, odpowiednio, 5,994, 1,974 i 2,892). Przytaczam tutaj powyższe dane bibliograficzne, aby podkreślić, że mgr inż. Michał Romaszewski wydaje się spełniać z nawiązką wymogi formalne stawiane przez nawet ambitne jednostki badawcze doktorantom chcącym finalizować swój przewód doktorski. Dodatkowo, rzut oka na listę publikacji na stronie Instytutu Informatyki Teoretycznej i Stosowanej PAN pokazuje, że wyżej wymienione prace nie stanowią całego dorobku naukowego doktoranta, który jest znacznie obszerniejszy, i zapewne nie został włączony do pracy aby zachować pełną spójność tematyczną.

Rozprawa zawiera wiele oryginalnych wyników, które omówiłem już w poprzedniej części recenzji, w związku z tym tutaj wymienię kilka najważniejszych, moim zdaniem, osiągnięć:

- Konstrukcja algorytmu w oparciu o współdziałanie dwóch ekspertów związanych z dwoma widokami – przestrzennym i spektralnym, pozwalająca efektywnie się uczyć z zaskakująco małych zbiorów treningowych. Algorytm cechuje się wysoką trafnością klasyfikacji, przewyższającą rozwiązania z poprzednich prac (a nawet inne rozwiązanie samego doktoranta).
- Formalna definicja problem uczenia pół-nadzorowanego z wieloma widokami w postaci zbioru nieskierowanych grafów (związanych z poszczególnymi widokami), oraz wprowadzenie algorytmów klasyfikacji bazujących na spacerach losowych po grafach i je agregujących. W szczególności, autor analizował zaproponowany schemat algorytmiczny z teoretycznego punktu widzenia, podając warunki konieczne na przydzielenie etykiety każdemu przykładowe, definiując skupiska w grafie pozwalające na poprawną klasyfikację wszystkich wierzchołków, oraz wprowadzając warunki konieczne i wystarczające na poprawną klasyfikację w szerszym kontekście.
- Propozycja eksperymentalnego podejścia pozwalającego na zastosowanie i weryfikację teoretycznych rezultatów opisanych w poprzednim punkcie, włącznie z wyprowadzeniem miar gwarantujących poprawną klasyfikację, lub poprawną klasyfikację pikseli etykietowanych w pierwszym kroku działania algorytmu, a także eksperymentalną weryfikację działania algorytmów opartych na spacerach losowych po grafach widoków.
- Zastosowanie algorytmów klasyfikacji do rozwiązania praktycznych problemów obrazowania hiperspektralnego.

Krótko podsumowując wymienione osiągnięcia uznaję, że doktorantowi udało się osiągnąć cele rozprawy, jakimi była konstrukcja nowych algorytmów klasyfikacji wielospektralnej oraz analiza eksperymentalna i teoretyczna wielo-widokowego i pół-nadzorowanego podejścia do problemu.

4 Uwagi dyskusyjne i polemiczne

Nie kwestionując wartości całościowych wyników zawartych w rozprawie, chciałbym zgłosić poniżej kilka uwag krytycznych i dyskusyjnych.

- W rozdziale 2 moje wątpliwości budzi odejmowanie od siebie ocen dawanych przez ekspertów zgodnie ze wzorem (2.4). Sugerowałoby to, że obie oceny znajdują się na tej samej skali. Mimo, że obie oceny przyjmują wartości z przedziału $[0, 1]$, to jednak ich interpretacja wydaje się inna. W szczególności suma ocen eksperta N (po odjęciu od jedynki) dla różnych klas zawsze daje wartość jeden, może więc być interpretowana jako jeden minus prawdopodobieństwo przydzielenia piksela do klasy, $P(\text{klasa}|\text{piksel})$. Z kolei, oceny eksperta P nie mają takiej własności. Zbudowane na jądrowym estymatorze gęstości mają raczej wymiar rozkładu gęstości prawdopodobieństwa po obrazie pod warunkiem wybranej klasy, czyli $P(\text{piksel}|klasa)$. Sugerowałoby to konieczność odwrócenia drugiego z prawdopodobieństw za pomocą twierdzenia Bayesa.
- Również w rozdziale 2, może trochę niepokoić złożoność obliczeniowa eksperta P , który dla każdej z klas i i dla każdego z pikseli musi wyznaczyć estymator gęstości na podstawie wszystkich innych pikseli należących do danej klasy (tak wynika przynajmniej ze wzoru (2.2)). Czy da się tutaj zastosować przybliżenia lub metody obliczeniowe, które zmniejszą złożoność tej części algorytmu? Przykładowo, można by użyć jądra gęstości, które ma skończony nośnik (np. trójkątnego lub Epanecznikowa).
- Jak już wspomniano, w rozdziale 3 część twierdzeń i dowodów jest trudna w lekturze i pojawiły się pewne niejasności. Sugerowałbym skrupulatne przejście tego rozdziału.
- Również w rozdziale 3, istotność wyników we wniosku 3.7.3, twierdzeniu 3.8.2, lub wniosku 3.9.2 obniża założenie, który mówi, że klasyfikator etykietuje każdy węzeł grafu. Tymczasem autor podaje tylko warunki na to konieczne, a nie warunki wystarczające, na podstawie których można by postawić założenie sprawdzić. Oczywiście, w przypadku funkcji 1 tak się

zapewne dzieje (uwaga 3.10.1), ale już w przypadku funkcji 2 trudno cokolwiek powiedzieć. Podobnie, inne założenia poprawnej klasyfikacji związane z wartościami funkcji $S'_v(p, k)$ wydają się dość restrykcyjne.

- W rozdziale 4 moje wątpliwości budzi używanie wprowadzonych miar SCM i ESCM. Miary te są warunkami wystarczającymi na poprawną klasyfikację, ale nie koniecznymi, ich spełnienie może być w praktycznych problemach, w których grafy nie mają czystych separowalnych komponentów, dość trudne, a ich działanie jest obrazowane jedynie na bardzo regularnych przykładach (rysunki 4.4-4.7). Dodatkowo, założenie o etykietowaniu danego wierzchołka/piksela w pierwszej iteracji, które towarzyszy niektórym z wyników, powoduje, że uzyskane przy tym założeniu gwarancje poprawnej klasyfikacji dotyczą zapewne tylko niewielkiej liczby wszystkich pikseli.

5 Uwagi drobne

Poniżej przedstawiam listę drobnych uwag dotyczących mniej ważnych uchybień lub literówek:


- Sekcja 2.3.1, 3 linia: zamiast (x_1, \dots, x_n) powinno być (x_1, \dots, x_b)
- Sekcja 2.3.1: trochę (przynajmniej dla mnie) niecisły opis zbioru treningowego: w zasadzie do zbioru treningowego powinny wchodzić również opisy przestrzenne pikseli, tj. pary (r_i, c_i) dla każdego piksela $i = 1, \dots, n$.
- Sekcja 2.3.2, linia 9: pojawia się nieużywany wcześniej symbol S_L na określenie danych uczących (a może na określenie *score*?).
- Wzór (2.3): być może warto zwrócić uwagę na pewne niebezpieczeństwo numeryczne, gdyby $w(i, j)$ dla pewnej pary pikseli byłoby równe zero.
- Strona 20, linia 8: *spices-level* \rightarrow *species-level* (?)
- Strona 20, sekcja 2.4.2: opis, w jaki sposób estymowany jest parametr θ jest niezrozumiały: co oznacza zapis $i = N_l(k)$, jeśli $N_l(k)$ jest zbiorem, a k są również elementami zbioru? Czy chodzi o to, że minimum liczone jest po takich i , które należą do $N_l(k)$ dla pewnego $k \in \mathcal{T}^c$?
- Strona 24, równanie (2.5): z tego, co rozumiem, wyniki konkurujących algorytmów zostały wzięte z poprzednich prac, a więc od zmiennej $SM(D, s)$ odejmowana jest zawsze stała wartość (najlepszej) metody referencyjnej, niezależna od losowo wybranego zbioru treningowego? Równanie (2.5) pozostawia w tym sensie pewną niejasność: pierwszy człon po prawej stronie równania (jakość PNGrow) zależy od losowego zbioru treningowego, natomiast drugi jest stały i niezależny od zbioru treningowego. Być może warto byłoby to jakoś zaznaczyć.
- Strona 43, uwaga 3.2.3, pkt 2: co kryje się za stwierdzeniem „*every node in the set $\mathcal{T}_{\text{train}}$ is classified correctly*”? Czy chodzi o to, że w zbiorze treningowym informacja o etykietach jest zgodna z \mathbf{Y}_{true} ?
- Strona 44, definicja 3.3.7: czy ścieżki w definicji mierzone są względem grafu \mathcal{G} , czy indukowanego podgrafu η ? Definicja sugeruje, że względem oryginalnego grafu, ale w tej sytuacji jaka jest rola \mathcal{E}' ? (wydaje się, że można wziąć $\mathcal{E}' = \emptyset$ i definicja nadal będzie spełniona).
- Strona 49, pierwszy punkt: nie jest jasne dlaczego kombinacja wypukła jest oznaczona jako $S'_v = \alpha_v (S_v^T)^n$. Przez kombinację wypukłą n elementów x_1, \dots, x_n rozumie się sumę $\sum_i \alpha_i x_i$ (z nieujemnymi współczynnikami sumującymi się do jedynki), a nie pojedynczy wyraz tej sumy.
- Strona 50, opis algorytmu, sekcja „Arguments”: czy powinno być InitialProbabilities zamiast StartingProbabilities? Również w algorytmie pojawia się zbiór $\mathcal{T}_{\text{la}}^t$, który nie jest uaktualniany ani zdefiniowany wewnątrz algorytmu (czy powinno się pojawić $\mathcal{T}_{\text{la}}^{t+1} = \mathcal{T}_{\text{la}}^t \cup \mathcal{T}_{\text{selected}}^t$?).
- Strona 53, sekcja 3.5.2: funkcja 3 jest opisana jako dająca rozkład prawdopodobieństwa „*that starts from every labelled node with equal probability*”. Nie jest to dla mnie do końca jasne, ponieważ sugerowałoby to wszystkie wierzchołki z $\mathcal{T}_{\text{la}}^t$, podczas gdy zgodnie z funkcją 3, dla danej klasy $c \in \mathcal{C}$ przypisywany jest rozkład jednostajny na wszystkich wierzchołkach *mających etykietę tej klasy*, a nie mających jakkolwiek etykietę.

- Strona 54, Przykład 3.1: czym jest funkcja $f_v(\cdot)$? (nie jest to chyba wyjaśnione)
- Strona 55, opis przykładu 3.2: zdanie „for each point from view 1 and arbitrary one-to-one mapping was found at random in view 2” jest niezrozumiałe; nie są też podane wartości parametrów α_v .
- Strona 57, Lemat 3.6.3: czy v' z treści lematu to dokładnie v ? Jeśli tak, to może warto to zaznaczyć w lemacie?
- Mimo usilnych starań nie dałem rady zrozumieć logiki stojącej za dowodem twierdzenia 3.6.5. W szczególności, czy stwierdzenie „let there be a sequence of nodes” zakłada istnienie takiej sekwencji, czy udowadnia, że taka sekwencja istnieje? Nie jest dla mnie jasne, czym są poszczególne elementy k^t, \dots, k^0 i dlaczego $k^{i-1} \in \mathcal{Y}_c^i$ (i czym właściwie jest k^{i-1} ?). Żeby było jasne, nie podważam prawdziwości tego twierdzenia i intuicyjnie rozumiem, że jego dowód wynika z naprzemiennego stosowania lematów 3.6.3 i 3.6.4.
- Lemat 3.7.2: czy warunek drugi (tzn. $\mathcal{T}_{\text{train}} \cap V(\eta) \neq \emptyset$) nie wynika z warunku trzeciego (Algorytm 1 etykietuje każdy węzeł)? Jeśli warunek drugi nie jest spełniony, to w jaki sposób algorytm mógłby poetykietować wszystkie węzły?
- Lemat 3.8.1: czego dotyczy „if and only if”? Nie powinno ono pojawić się po wyrażeniu (3.7)? Podobnie jest w twierdzeniu 3.8.2 (rozpoczyna się od „if and only if”). Również w tym lemacie: w równaniu (3.7) mamy sumę po $k \in \mathcal{Y}_c^t$ zamiast po $k \in \mathcal{P}$ – dlaczego?
- Wniosek 3.9.1: dlaczego wszystkie prawdopodobieństwa muszą być równe? Mam wrażenie, że w dowodzie nie wykorzystuje się tego faktu (w równaniach (3.9) i (3.10) znaczenie ma tylko to, że minimum jest mniejsze (a maksimum większe) od dowolnej kombinacji wypukłej, niekoniecznie z równymi wagami). Kolejna uwaga: wydaje się, że najlepiej przyjęc $\mathcal{I} = \mathcal{T}_{1a}^t$, wtedy nierówność (3.8) jest najłatwiejsza do spełnienia.
- Strona 70: *assumption 3* \rightarrow *Assumption 3*.
- Lemat 3.10.2: również niepoprawna składnia z „if and only if” przed wyrażeniem (3.12).
- Strona 75, linia 5: zdanie „some nodes may no longer be labelled” sugeruje dla mnie, że węzły mogą w pewnym momencie przestać by poetykietowane.
- Definicja 4.2.1: skąd skrót „cli” na funkcję SCM?
- Strona 85, punkt 3 u góry strony: ten punkt chyba również wymaga założenia $\mathcal{T}_{\text{train}} \subseteq \mathcal{I}$.
- Strona 85, punkt 4: to nie wydaje mi się takie oczywiste, aby pozostawić bez dowodu.
- Strona 86, punkt 5: trzeba chyba założyć, że $p \in \mathcal{N}_c$ (tzn. powiązać klasę c z wierzchołkiem p)?
- Sekcja 4.4.1 (eksperyment): co oznacza stwierdzenie, że parametry algorytmu zostały wybrane za pomocą 5-krotnej walidacji krzyżowej? Czy mam przez to rozumieć, że zbiór treningowy (5, 10, lub 15 przykładów na klasę), został podzielony na właściwą część treningową (4, 8 lub 12 przykładów) i część testową (1, 2, lub 3 przykłady)? Część testowa wydaje mi się w tym wypadku dość mała, aby sensownie wybrać parametry.
- Tabela 4.1: nie jest powiedziane, w jaki sposób uzyskano klasyfikację dla pikseli, które nie zostały poetykietowane przez Algorytm 1. Czy użyto jakiegoś innego klasyfikatora, czy po prostu Algorytm 1 poetykietował wszystkie przykłady?
- Wykresy 4.8, 4.9, 4.12 – co oznacza „percentage of unlabelled examples labelled in every iteration”? Czy jest to związane wielkością zbioru $\mathcal{T}_{\text{selected}}^t$ w każdej iteracji?
- Sekcja 5.2.1: nie jest dla mnie do końca jasne, czy metoda MRG to dokładnie użycie SVM i MRF, a następnie wybór punktów, na których obie metody się nie zgadzają, czy jest to wstępna, odrębna metoda użyta uprzednio przed wymienionymi algorytmami.
- Tabela 5.1: nie jest opisany, w jakim sposób uzyskano wyniki i ich istotność statystyczną: czy eksperyment był powtarzany wielokrotnie i uśredniany? Przydałoby się choć jedno zdanie.

6 Konkluzja końcowa

Chciałbym zauważyć, że krytyczne uwagi oraz komentarze przedstawione w poprzednich częściach recenzji mają charakter dyskusji, mającej na celu lepsze zrozumienie rozprawy oraz pomoc w dalszych badaniach autora. Natomiast całościowo oceniam rozprawę bardzo dobrze. Problemy badawcze, z którymi zmierzył się doktorant, są ambitne i istotne dla postępu w dziedzinie obrazowania hiperspektralnego. Sama rozprawa charakteryzuje się wysokim poziomem merytorycznym i zawiera interesujące rezultaty, zarówno o charakterze eksperymentalnym, jak i teoretycznym. W mojej opinii doktorant wykazał się dobrymi umiejętnościami prowadzenia badań naukowych. Wszystkie postawione w rozprawie cele zostały osiągnięte.

W związku z tym **rozprawę oceniam jako spełniającą wymogi stawiane pracom doktorskim i wnoszę o dopuszczenie mgr. inż. Michała Romaszewskiego do dalszych etapów przewodu doktorskiego.**


dr hab. inż. Wojciech Kotłowski