

Verslag Meeting II

Ma 10/03/2014 - 18u-19u - MIRC

Aanwezig

- Dr. ir. Pieter Slagmolen (research manager)
- Ir. David Robben (doctoraatsstudent)
- Ir. Sven Van Hove
- Ir. Kim Nuyts

Agenda

Bespreking literatuurstudie

- Probleem: literatuurstudie geeft geen directie indicatie dat RF de beste methode is
- Mail Pieter: overzicht commerciële toepassingen nodule detectie (in scanners)
- Mail David: papers
 - Caruana: vergelijking classificatie methodes
 - Van Ginneken en Murphy: ANODE challenge
- ANODE challenge:
 - Standaard dataset → interessant voor validatie
 - Objectieve methodevergelijking
 - Participeren in wedstrijd?
- Zie ook: The Medical Image Computing and Computer Assisted Intervention Society (MICCAI)

Bespreking conceptueel plan

- Pieter: benader project als commercieel product
 - Bv: real time verwerking, max 15-30min
- Nog verder uitwerken? Kwantificeren?

Bespreking tijdschema (Gantt chart)

- Deadline basic classification opschuiven (einde paasvakantie)

Bespreking technische problemen

- MeVisLab module ontwerpen i.p.v. stand-alone python script?
 - Nee, gemakkelijker om in python te programmeren dan in MeVislab
 - MeVisLab debugger is niet bij de beste
 - Eventueel wel voor classificatie te gebruiken
 - Later code omzetten in MeVisLab module [extra]
- Data inladen in Python: consistent zijn met dataformaat
 - Slice per slice
 - Als 3D volume
 - Verschillende bestandsformaten (bv. NIFTY)
- Meeste datasets die onder longen geclassificeerd staan bevatten gaan annotaties
 - Gebruik Lung Image Database Consortium (LIDC) dataset

- Annotaties bevatten Malignancy probability per nodule, relevant of nuttig?
 - o In eerste instantie niet gebruiken
 - o Vervolgens eventueel opdeling in klassen maken [extra]
- Annotaties bevatten verslagen van meerdere radiologen
 - o Neem simpelweg de eerste (consistent zijn)
 - o Gebruiken indien wij fouten maken → maken zij ze ook?

Waarom kiezen voor Random Forests?

- Support Vector Machines
 - o Vanaf 20 000 te traag
 - o Noodzakelijk om features met dezelfde grootheden te gebruiken
 - Anders lineariseren, herschalen enz. → slecht
 - o Problemen met ruis
 - o Probleem indien features nog niet op voorhand gekend
 - Hoe clustering doen (invloed op rekentijd!)
 - Eventueel filters voor features
- Random Forests
 - o Kan belangrijke features zelf bepalen
 - o Beter voor grote datasets
 - o Niet zo gevoelig aan ruis
- Boosted trees
 - o Op eerste zicht beter dan RF (Carauna)
 - o Problemen
 - Slecht bestand tegen ruizige/ambigue problemen (onze beelden)
 - Keuze van parameters wordt hierdoor gevoeliger en moeilijker (boomdiepte, aantal bomen, learning rate...)
 - Training is inherent sequentieel (trager dan RF)
- Methode E, isi-cad (Murphy)
 - o KNN classifiers: suboptimale keuze (opportuniteit om state-of-the-art te kloppen)
 - o Zelfde probleem als SVM indien features verschillende grootheden hebben
- Algemene opmerkingen overzicht methodes
 - o Preprocessing en feature selectie (arbitrair???) zeer belangrijk
 - o Goede rekentijd: gebruik cascaded classifiers

Vragen

- Python versie 2.7 of 3.3?
 - o Gebruik 2.7 wegen backwards compatibility problemen in 3.x
- Gebruik chest X-rays voor eerste localisatie?
 - o Nee, zou ons te ver leiden. Mogelijk ook niet allemaal zichtbaar.
- Annotaties: nodules vs non-nodules vs other?
 - o Check bijbehorend paper (non-nodules = kanker maar geen nodulevorm?)

Volgende meeting

Donderdag 20/03/2014 @ MIRC, 17u00