A Domain-Change Approach to the Semantic Labelling of Remote Sensing Images

Chandrabali Karmakar, Gottfried Schwartz, Corneliu Octavian Dumitru, and Mihai Datcu

EGU 2022

Machine Learning for Earth system Modelling

24.05.2022

Semantic Labelling in Remote Sensing Images

"**Semantic labeling is the process of mapping attributes in data sources to classes in an ontology** " [1]

Recent approaches in semantic labelling of remote sensing images:

Supervised : Supervised Deep Learning models, high accuracy but high cost of annotation

Semi-supervised: Transfer learning, Active learning (Human in the loop), less expensive

Unsupervised: Based on latent data structure, may require expert's validation, least expensive

Our research explores an active learning method based on a tool developed by [2] and an unsupervised method based on Latent Dirichlet Allocation (LDA) as proposed by [3]. The two approaches are combined to create semantic labels for large datasets.

[1] Pham, Minh, et al. "Semantic labeling: a domain-independent approach." International Semantic Web Conference. Springer, Cham, 2016.

[2] W. Yao, C.O. Dumitru, and M. Datcu, "An active Learning Tool for the Generation of Earth Observation Image Benchmarks", IGARSS, Brussels, Belgium, (2021): 1-4.

[3] K. Chandrabali, C.O. Dumitru, and M. Datcu, "Feature-free explainable data mining in SAR images using latent Dirichlet allocation." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2020): 676-689.

Workflow combining Active Learning (semi-supervised) and LDA (unsupervised)

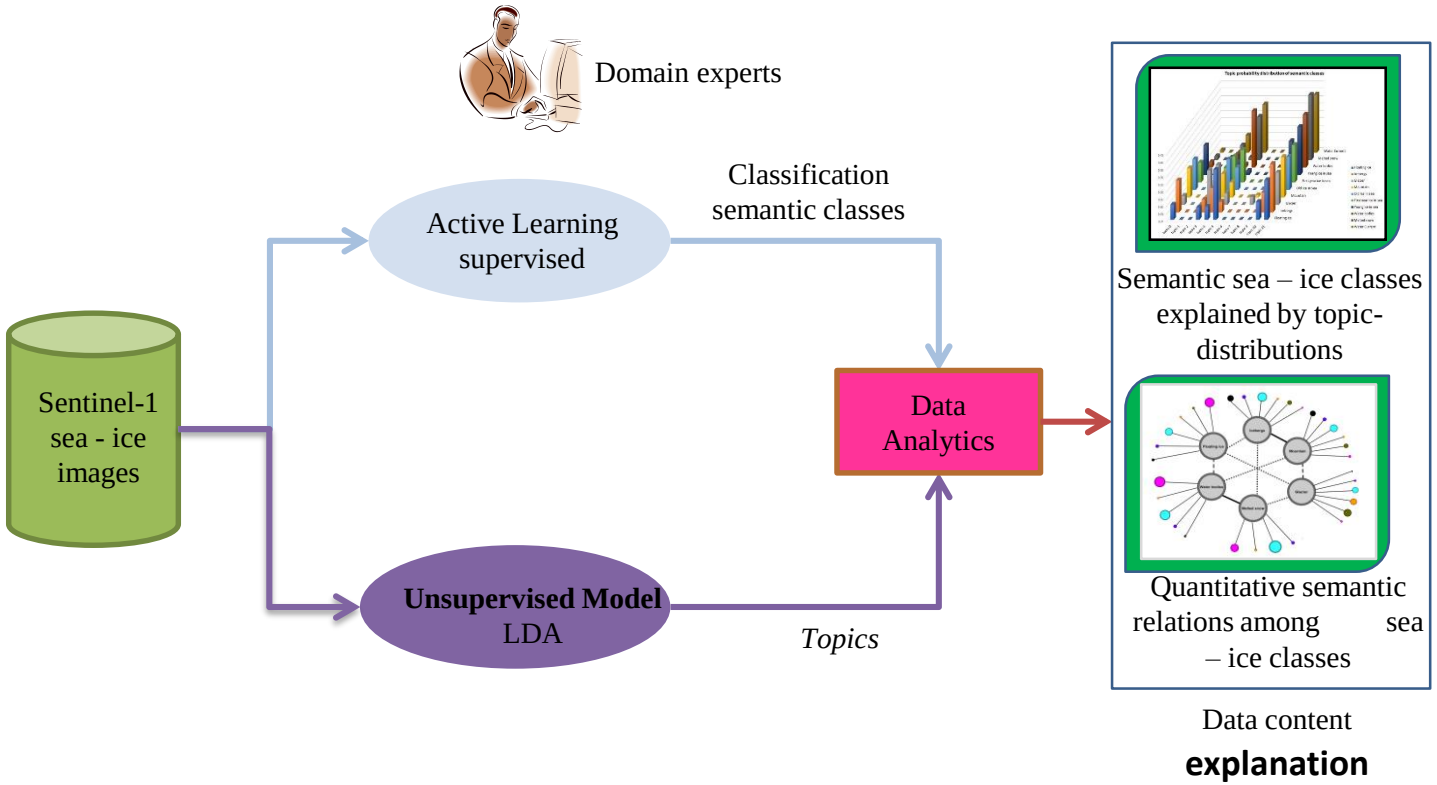
The Active Learning tool creates labelling for limited set of data

LDA is applied to discover latent structures called *'topics'* in an unsupervised manner

A relation between the labels and topics are established

Requires less amount of annotation

Integrates domain knowledge



Active Learning work mechanism

Techniques

SVM

Relevance feedback

Active learning

Machine interactively inquires a user for new knowledge

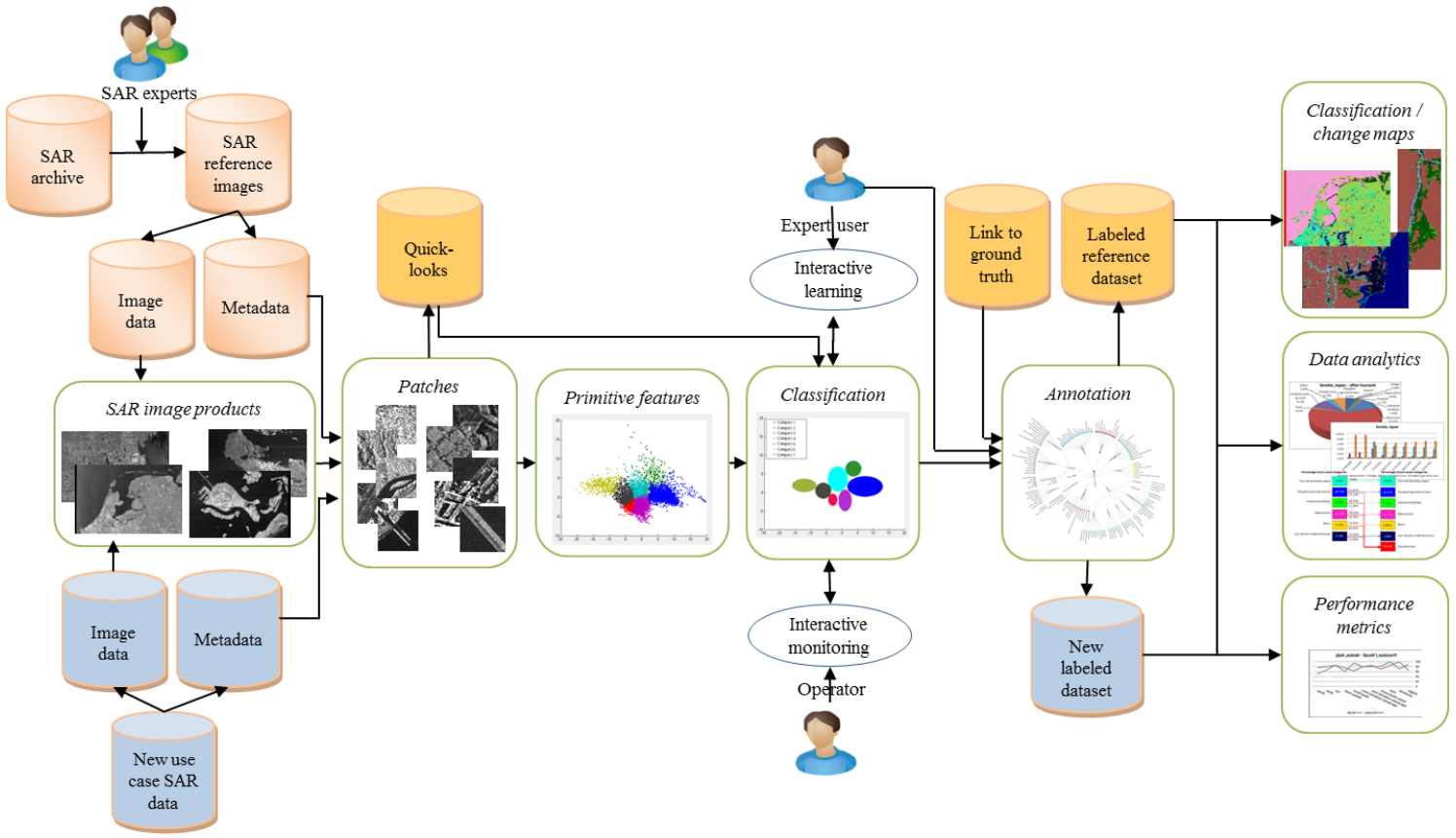
User select the semantic labels from a list or it defines itself

The user obtains a labelled dataset learned with few samples

Cascaded learning

Multi-grid layer classification

Multi-product image mining



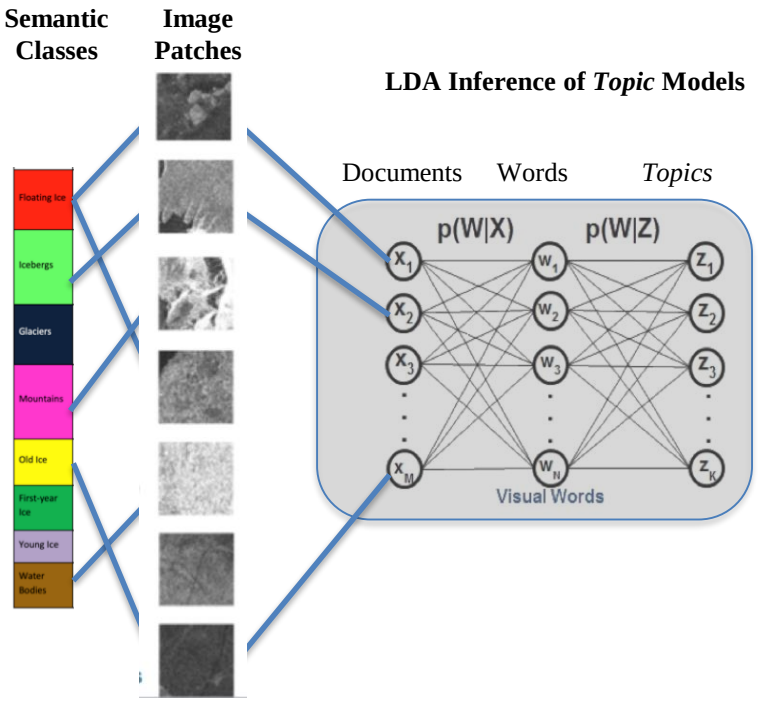
LDA work mechanism

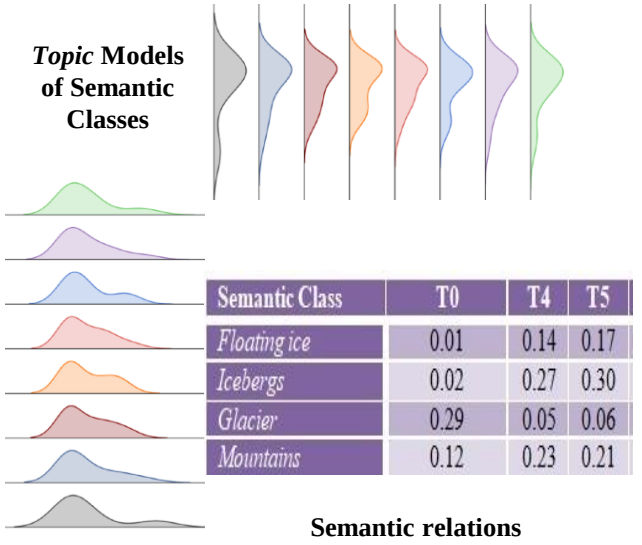
Works on top a **bag-of-words** representation of the images

Finds **latent structures** (*topics*) in the data by observing distribution of words

Topics are **mapped to semantic classes** retrieved with active learning

An **explainable approach** to semantic labelling of remote sensing imagery



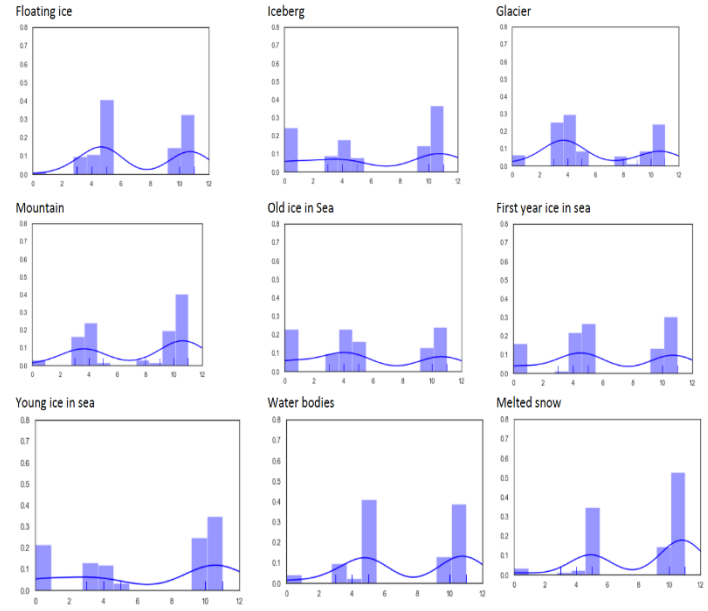


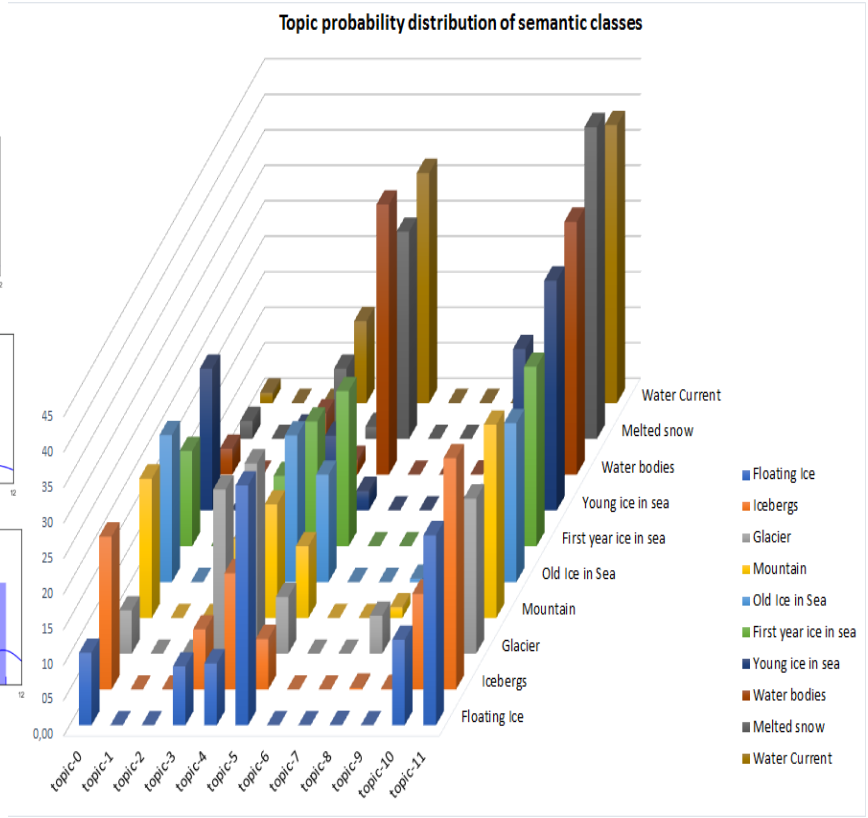
Data analytics

Semantic relations

With the **sea-ice classes** from **Active Learning**

derived by using **topic distributions**





Ref: K. Chandrabali, C.O. Dumitru, and M. Datcu, "Feature-free explainable data mining in SAR images using latent Dirichlet allocation." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2020): 676-689.

DLR

Thank you!

Contact: chandrabali.karmakar@dlr.de