Negative-Unlabeled Learning for Diffusion MRI

Phillip Swazinna[1], Vladimir Golkov[1], Ilona Lipp[2], Eleonora Sgarlata[2,3], Valentina Tomassini[2,4], Derek K. Jones[2], and Daniel Cremers[1]

[1]Department of Informatics, Technical University of Munich, Garching, Germany, [2]CUBRIC, Cardiff University, Cardiff, United Kingdom, [3]Department of Neurology and Psychiatry, Sapienza University of Rome, Rome, Italy, [4]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, United Kingdom

Synopsis

Machine learning strongly enhances diffusion MRI in terms of acquisition speed and quality of results. Different machine learning tasks are applicable in different situations: labels for training might be available only for healthy data or only for common but not rare diseases; training labels might be available voxel-wise, or only scan-wise. This leads to various tasks beyond supervised learning. Here we examine whether it is possible to perform accurate voxel-wise MS lesion detection if only scan-wise training labels are used. We use negative-unlabeled learning (an equivalent of positive-unlabeled learning) and achieve an AUC of 0.77.

Introduction

Diffusion MRI provides valuable information about microstructural tissue properties. The classical processing pipeline uses physical/mathematical representations that are suboptimally simplistic (discard information) and rely on unstable fitting that requires long scan times. Recent methods based on machine learning circumvent these drawbacks. They do not rely on unstable fitting[1,2,3,4,5] and extract features from diffusion-space (q-space) data in an optimized (rather than handcrafted) way[1,4,5]. Such methods exist for different situations/tasks: To predict voxel-wise microstructural properties (such as disease effects) in test data when training data with voxel-wise labels are available[1]; to detect voxel-wise "novelty" (e.g. disease effects) in test data when only "normal"-labeled (e.g. healthy) training data are available[2,3,4]; to predict scan-wise properties (e.g. disease) in test data when training data with scan-wise labels are available[5]; to get fine-grained (voxel-wise) predictions from coarse-grained (scan-wise) training labels[5].The latter corresponds to different machine-learning tasks, depending on whether the information is used that (a) voxels belonging to one scan can be grouped into a "bag" of voxels, (b) all voxels from healthy-control scans are healthy, (c) which features indicate disease may depend on context (features in other voxels). Using (a,c) but not (b) corresponds to weakly-supervised learning, e.g. class-activation maps . Using (a,b) corresponds to multiple-instance learning. Using (b) but instead of (a,c) grouping voxels into only two "bags" based on their scan-label is negative-unlabeled learning. Here we use the latter to detect multiple-sclerosis (MS) lesions without using voxel-wise training labels. Figure 1 shows an overview of methods.

Methods

Data: 94 MS patients and 26 healthy controls, each with six b=0 images and 40 diffusion-weighted images (6+40 "channels"), b_{max}=1200s/mm², SE-EPI, voxel size 1.8mm×1.8mm×2.4mm, matrix 128×128, 57 slices, TE=94.5ms, TR=16s, motion/distortion-corrected[6]. Human raters marked MS lesions. To facilitate neural network training, we perform so-called feature scaling by dividing each channel by the corresponding channel mean taken over all scans. To prevent overfitting on intensity values, we also divide each scan by its mean intensity, and multiply it by a random scalar between 0.8 and 1.2 during each training epoch.

Negative-Unlabeled Learning: We examine whether it is possible to perform accurate voxel-wise MS lesion detection if only scan-wise training labels are used. We treat every voxel as a sample, with its features being the q-space measurements. We distinguish a set of negative samples (all voxels from healthy controls) and a set of unlabeled samples (patient scans consisting of lesions and healthy voxels without labels). Using such training data is called negative-unlabeled learning[7], or equivalently (by renaming the classes) positive-unlabeled learning. When labeling the unlabeled set entirely as positive (hence introducing some "label-noise") and optimizing the area under the ROC curve (AUC) is equivalent to supervised learning with AUC optimization[8]. Here we optimize mean-squared-error instead of AUC, which usually yields similar results. We expect a prediction around 1 for lesions and around 0.78 (due to class imbalance and neural networks averaging out label-noise) for healthy voxels.

Training: We used a convolutional network: four layers with 128, 256, 512, and 1 filter, respectively, all filter sizes 1×1×1, ReLU, mean-squared-error loss (and quality evaluation) only on segmented[9,10] brain voxels. We used 60% of scans for training, 20% for early stopping, 20% for testing.

Results & Discussion

Figure 2 shows output maps from the test set together with corresponding histograms for novelty detection[2,3], negative-unlabeled learning, and voxel-wise supervised learning[1]. As expected, supervised learning[1] yields best AUC (0.91) across voxels of all test scans due to using the largest amount of information for training (correct labels for all training voxels). Negative-unlabeled learning yields overall good AUC (0.77). It uses more information (distribution of unlabeled positives) than novelty detection, but surprisingly is outperformed by it (AUC=0.89). One might suspect that "label-noise" incites negative-unlabeled learning to become sensitive to subtle MS features of normal-appearing white matter in patients, causing false-positives (in terms of lesion labels). However, high predictions also reach far into grey matter. Non-deep learning can be better for certain data, but since deep q-space novelty detection outperforms[4] its non-deep counterpart, the deep nature of negative-unlabeled learning is unlikely to be the issue in this case. We were not able to tune class-activation maps to achieve good AUC. Their cost function is similar to negative-unlabeled learning, but the difference between them appears to influence results considerably. For all these reasons we conclude that more research on these families of methods is necessary.

Acknowledgements

References

[1] Golkov et al.: "q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans" In IEEE Transactions on Medical Imaging, volume 35, 2016

[2] Golkov et al.: "Model-Free Novelty-Based Diffusion MRI" In IEEE International Symposium on Biomedical Imaging (ISBI), 2016

[3] Golkov et al.: "q-Space Novelty Detection in Short Diffusion MRI Scans of Multiple Sclerosis" ISMRM Annual Meeting, 2018

[4] Vasilev et al.: "q-Space Novelty Detection with Variational Autoencoders", In ArXiv preprint, 2018

[5] Golkov et al.: "q-Space Deep Learning for Alzheimer's Disease Diagnosis: Global Prediction and Weakly Supervised Localization" ISMRM 2018

[6] Klein et al.: "elastix: A toolbox for intensity-based medical image registration" In IEEE Transactions on Medical Imaging 29(1):196–205, 2010

[7] G. Niu: "Recent Advances on Positive Unlabeled (PU) Learning" IBISML, 2017.

[8] Zhung, Lee: "Learning Classifiers without Negative Examples: A Reduction Approach" In Third International Conference on Digital Information Management, 2008

[9] P. Kalavathi: "Brain tissue segmentation in MR brain images using multiple Otsu's thresholding technique" 8th International Conference on Computer Science & Education, 2013

[10] Garyfallidis et al.: "DIPY, a library for the analysis of diffusion MRI data" Frontiers in Neuroinformatics, vol.8, no.8., 2014

Figures

Labeled disease data available	Labels	Prediction	Method	Reference
Yes	Voxel-wise	Voxel-wise	Voxel-wise supervised	[1]
No		Voxel-wise	Novelty detection	[2,3 4]
Yes	Scan-wise	Scan-wise	Global supervised	[5]
Yes	Scan-wise	Voxel-wise explanation for scan-wise prediction of each class	Class-activation maps	[5]
Yes	Scan-wise	Voxel-wise	Negative-unlabeled learning	**Proposed**

Summary of q-space machine learning methods.
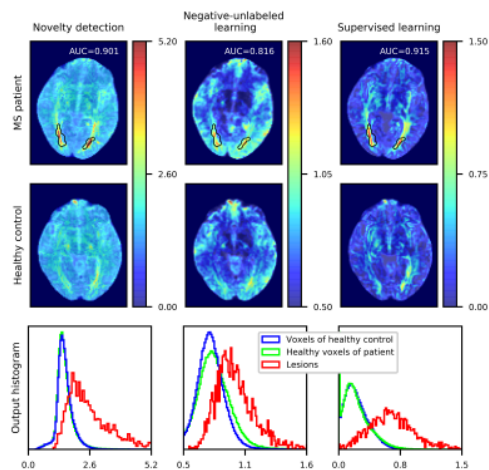
Outputs for all three methods. Lesions are well recognized in the patient scan (first row) by all methods. As expected, supervised learning yields best AUC (shown for this patient) due to using the largest amount of information for training. Surprisingly, novelty detection yields better AUC than negative-unlabeled learning, despite using less information. In the healthy-control scan (second row), all three methods mistake a small anterior distortion artifact (with likely incorrect q-space features) for a lesion, but otherwise assign low scores throughout the entire healthy-control scan. The corresponding histograms (third row) confirm quite good separability of lesions from healthy voxels.

Negative-Unlabeled Learning for Diffusion MRI

Phillip Swazinna[1]   Vladimir Golkov[1]   Ilona Lipp[2]   Eleonora Sgarlata[2,3]   Valentina Tomassini[2,4]   Derek K. Jones[2]   Daniel Cremers[1]

[1]Computer Vision & Artificial Intelligence, Department of Informatics, Technical University of Munich, Germany   [2]CUBRIC, Cardiff University, United Kingdom
[3]Department of Neurology and Psychiatry, Sapienza University of Rome, Italy   [4]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, United Kingdom
golkov@cs.tum.edu

Overview of Methods for Diffusion MRI

	Fitting of handcrafted representation (DTI, DKI, NODDI, ...)	Analytic computation of handcrafted metrics [1]	Machine learning							
			Voxel-wise supervised learning (of handcrafted metrics) [2]	Voxel-wise supervised learning (of tissue properties) [2]	Voxel-wise semi-supervised learning [hypothetical method]	Scan-wise supervised learning [3]	Weakly-supervised learning [3]	Multiple-instance learning [hypothetical method]	Novelty detection [4,5,6]	Negative-unlabeled learning [proposed]
Works with few q-space measurements	No	✓	✓	✓	✓	✓	✓	✓	✓	✓
Object of direct study	Handcrafted metrics	Handcrafted metrics	Handcrafted metrics	Tissue properties (for example abnormality)	Tissue properties (for example abnormality)	Tissue properties (for example abnormality)	Tissue properties (for example abnormality)	Tissue properties (for example abnormality)	Abnormality	Abnormality
Location of labels	None	None	Voxel-wise from fitting (which requires none)	Voxel-wise	Voxel-wise	Scan-wise	Scan-wise	Scan-wise	Any (only normal)	Any (only normal)
Location of prediction	Voxel-wise	Voxel-wise	Voxel-wise	Voxel-wise	Voxel-wise	Scan-wise	Voxel-wise clues for global prediction	Voxel-wise	Voxel-wise	Voxel-wise
Usage of unlabeled data during training	No	No	No	No	✓	No	No	No	Usually no	✓
Used knowledge — (a) Voxels from one scan belong together	N/A	N/A	N/A	N/A	N/A	✓	✓	✓		
Used knowledge — (b) All voxels from healthy-control scan are healthy	N/A	N/A	N/A	✓	N/A	✓		✓	✓	✓
Used knowledge — (c) Disease clues may depend on context (other voxels)	N/A	N/A	N/A	N/A	N/A	✓	✓			

Setting & Approach

- q-Space deep learning [2,3,4,5,6]: Prediction of tissue properties directly from q-space measurements
- Every voxel is a sample
  - Features are q-space measurements
- Only negative (healthy) and unlabeled samples are given
  - i.e. **negative-unlabeled learning** [8]
- No positive (multiple sclerosis) labels are given
  - i.e. **no knowledge about disease is required**
- Goal: distinguish negative and positive samples
- Treating unlabeled samples as positive (which introduces "label noise") is (for certain cost functions) a good method for netagive-unlabeled learning [Zhung&Lee]
 - We use a simpler cost function that yields similar results

Data

- 94 multiple sclerosis patients, 26 healthy controls
- Six b=0 images, 40 diffusion directions (b_{max}=1200s/mm^2)
- SE-EPI, TR=16s, TE=94.5ms, voxel size 1.8mm×1.8mm×2.4mm, matrix 128×128, 57 axial slices, motion/distortion-corrected [9]

Neural Network

- Feature scaling: divide each channel by its mean taken over all scans
- To prevent overfitting of intensity values: divide each scan by its mean and multiply with random scalar between 0.8 and 1.2 in every epoch
- **3D ConvNet**: ReLU, 128,256,512,1 filters 1×1×1, Adam

Results
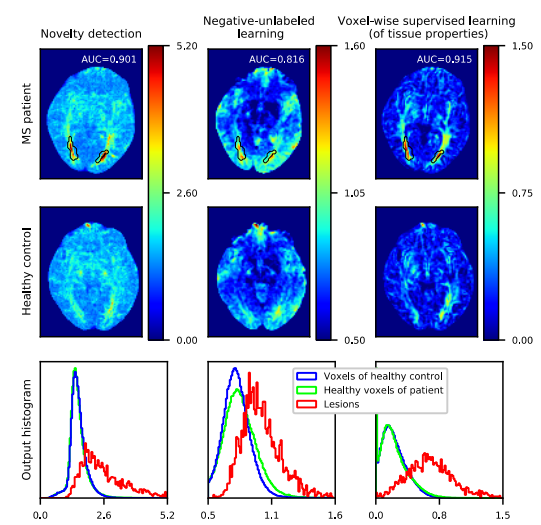
Discussion & Conclusions

- **Deep learning for diffusion MRI**:
  - **data-driven**: diagnosis directly from raw q-space data
  - **many advantages**: ultra-short scans, optimal usage of information, ...
  - applicable in **various situations**: coarse or missing labels, unknown disease effects, ...
- As expected, supervised q-Space Deep Learning yields best AUC
- Negative-unlabeled learning yields good AUC (0.77) but is surprisingly outperformed by novelty detection (0.89)
- More research is necessary

References

1. Hansen et al.: "Fast imaging of mean, axial and radial diffusion kurtosis", NeuroImage 2016
2. Golkov et al.: "q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans", IEEE TMI 2016
3. Golkov et al.: "q-Space Deep Learning for Alzheimer's Disease Diagnosis: Global Prediction and Weakly Supervised Localization", ISMRM 2018
4. Golkov et al.: "Model-Free Novelty-Based Diffusion MRI", ISBI 2016
5. Golkov et al.: "q-Space Novelty Detection in Short Diffusion MRI Scans of Multiple Sclerosis", ISMRM 2018
6. Vasilev et al.: "q-Space Novelty Detection with Variational Autoencoders", arXiv 2018
7. Zhang & Lee: "Learning Classifiers without Negative Examples: A Reduction Approach", ICDIM 2008
8. Niu: "Recent Advances on Positive Unlabeled (PU) Learning", IBISML 2017
9. Klein et al.: "elastix: A toolbox for intensity-based medical image registration", IEEE TMI 2010