Bachelor thesis
Computer Science

Radboud University

Finding hyperintensities in the brain using unsupervised feature learning

Author:
Koen Vijverberg
s4132858

First supervisor/assessor:
Mohsen Ghafoorian, MsC
mohsen.ghafoorian@radboudumc.nl

Second supervisor:
Prof. dr Tom Heskes
t.heskes@science.ru.nl

January 20, 2015

Abstract

In this thesis we try to find out if unsupervised feature learning can be used to find white matter lesions in brain-MRI scans. Results show that unsupervised feature learning algorithm performs similar to classification using regular features. In some cases it performs even better. Downside is that unsupervised feature learning is computationally more expensive.

Contents

Chapter 1

Introduction

In this thesis we are trying to solve the problem of detecting hyperintensities in Magnetic Resonance Imaging (MRI) images of the brain. MRI is one of the techniques by which an image of the internals of an organism can be made. This is used by clinicians to determine the healthiness of the organs from a patient. In this thesis, we are using FLAIR weighted MRI[1] scans of the brain. In an MRI FLAIR scan, white matter lesions (WML) emerge with a higher signal and thus brighter areas; that is why they are also called white matter hyper-intensitiies (WMH)". WMHs is one of the symptoms which is used to diagnose people with cerebral small vessel disease (SVD). Improvement of isolating WMHs helps the diagnosis of SVD.

Up to now the identification and annotation of white matter hyper-intensities is done by an analysis of a neurology expert. This handwork by the doctor can be automated by using image analysis and machine learning techniques. The manual annotation has several drawbacks: they are usually very time consuming, subjective and error prone. Thus this automation can be very useful. The usual method for this automation is to use conventional machine learning techniques that make use of hand-crafted features to train the classifier. In this learning scheme, the overall performance strictly depends on careful domain-dependent choice of features. Instead we attempt to use an algorithm called 'unsupervised feature learning', which, as the name says, independently learns specific features of the object it is trying to detect. Another reason we chose this algorithm is because recently some very good results were achieved on well-known datasets [9]. The goal is to create a WMH detection system without any hand crafted-feature that has similar or better results than a regular classifier.

[1]MRI scans can be transformed/weighted in several ways, FLAIR weight nulls the cerebrospinal fluids. E.a. this makes the cerebrospinal fluids appear black in the MRI-scan.

Chapter 2

Preliminaries

2.1 Small vessel disease

In this research we develop and evaluate a computer-aided detection algorithm to detect WMHs related to patients diagnosed with small vessel disease (SVD). These lesions appear in the cerebral cortex and are common for elderly people. SVD is linked to cognitive decline and loss of other functionality such as problems with gait and speech [13, 1]. Hyperintesities, together with lacunar infarcts and brain microbleeds are symptoms of SVD and are used to diagnose people with this disease. Figure 1 shows a healthy brain on the left side, and a brain with WMHs on the right side.

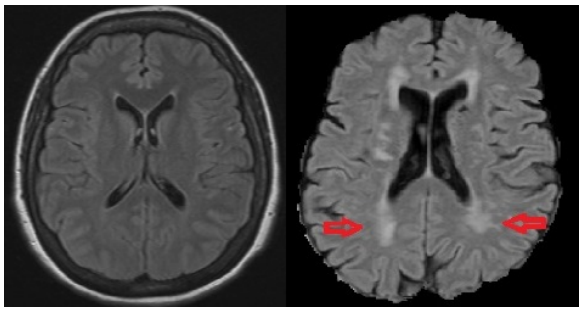

Figure 1: Healthy vs. non-healthy brain.

2.2 Unsupervised feature learning

Unsupervised Feature Learning (USFL) is a technique to learn some discriminative features from the input data and to then use those features to classify a given image. This means that the created features differ from dataset to dataset. In the following section we explain in more detail which steps are involved.

2.2.1 Input

A possible input for the unsupervised learning algorithm is images. In this thesis these are patches of FLAIR MRI-scans from the RUN DMC study [12]. Images are often pre-processed before they are used. A common way to do this is by subtracting the mean and dividing by the standard deviation for each pixel. After the normalization it is possible to whiten the image but it is not required. Our dataset is normalized but not whitened.

2.2.2 Algorithm

There are many variations in USFL and there are some parameters that need to be chosen manually. The one described here is a commonly used one [4, 5]. At some points we deviate from the common method to accommodate for specifics of our dataset.

In general there are two steps in USFL: 1. generating the feature mapping and 2. creating the training dataset. The first step, generating the feature mapping consists of three steps as also described in [4].

- Getting patches from all the images
- Pre-processing the patches (optional)
- Learning a feature mapping

Learning the feature mapping can be done in multiple ways. For example using sparse auto-encoders [3], sparse restricted Boltzmann machines [8], K-means clustering [7] and Gaussian mixtures [10]. We use K-Means clustering which is easy to implement and gives good results [4].

The second step, creating the training dataset, also consists of 3 sub-steps:

- Extracting patches from the training images and map them to features
- Pooling features from certain areas together so the dimensionality is reduced.
- Assigning labels to the patch feature vector based on the ground truth.

Once the dataset is created one can choose an arbitrary supervised machine learning algorithm to train and test the data. An algorithm with low complexity is advised because: 1. the created features are usually discriminating enough not to demand complex classifiers and 2. the data has so many variables and data points (in our case 1200 variables and 64958 data points) that it is computationally infeasible to train non-linear algorithms on it.

Chapter 3

Research

3.1 Generating dataset

We want to determine for each candidate-lesion whether or not this is a real lesion (or a false positive). This can be achieved by approaching this as a binary classification problem, we chose using the label 1 for true lesions and label 0 for false-positives. Research conducted by Ghafoorian et al [6]. resulted in a probability maps for each patient in the earlier mentioned RUN DMC study. The probability map gives for each position in the MRI-scan the probability that that part of the brain is WML tissue. Thresholding his probability maps for each patient on an eighty or higher percent chance resulted in 64958 candidate lesions/regions. Pre-selecting this way gives us a high probability of having a WMHs in each region. Another option would have been to slide a window over each slice of the MRI scan and each patient, but this would have created an excessive amount of negative sampless. Pre-selecting regions prevents this. Since most methods for USFL in the literature classify images in their entirety, e.g. classifying if there is a cat or no cat in the picture the algorithm required a slight modificatio since we want to classify parts of the image to see if there is a WML in it or not. We chose a size of this sub-image -from now on called intermediate patch- of 16×24 pixels. The chosen window size of 16×24 is based on a lesion x,y-size graph, a heatmap of lesion size and a lesion distance histogram. General idea is to have a window size, such that the average lesion fits exactly. We chose 16 pixels width because this covers most lesions as can be seen in the lesion frequency versus size image. This is also small enough to make sure that there will not be two or more lesions within one window as can be seen in the lesion distance histogram. The reason we chose a larger height than width is because WMLs tend to appear more elongated in the vertical direction. This is clearly visible in the heatmap.

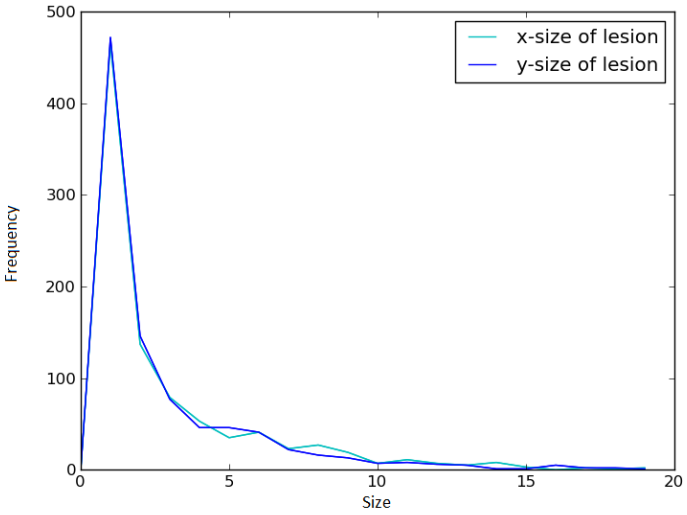

Figure 2: Width and height histogram for 64958 candidate lesions generated based on the probability map by Ghafoorian et al [6].

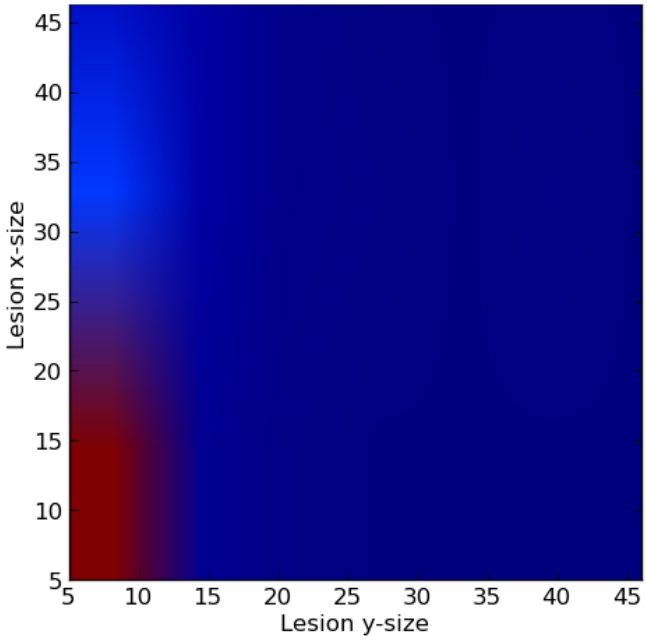

Figure 3: Heatmap of lesion size for 64958 candidate lesions generated based on the probability map by Ghafoorian et al [6].

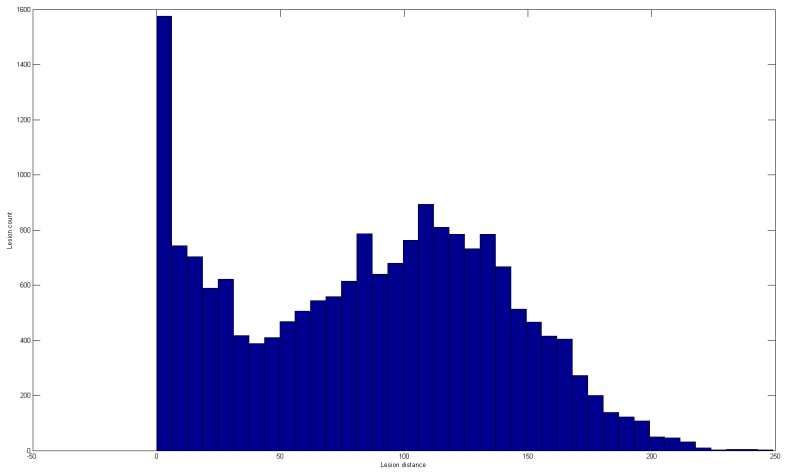

Figure 4: Histogram of lesion distance for 64958 candidate lesions generated based on the probability map by Ghafoorian et al [6].

3.2 Creating feature mapping

Normal machine learning approaches would collect information about the intensities of each patch e.g. min intensity, max intensity average intensity. We will be doing something different: use the USFL algorithm. This algorithm takes the intensities into account, but it can also learn shapes. If WMHs always have a certain shape this should be picked up by the algorithm and improve the classification performance. USFL algorithm starts by learning the feature mapping. To create features from the data we must choose a patch size of n-by-n which will be used to create the feature maps, note that this is *not* the intermediate patch size. Usual values for this parameter are 5×5 or 6×6. In this case we chose patches of 5×5. To create the feature map, the MRI-images need to be cut into patches of the chosen size. It is also possible to consider a parameter for the space between each patch, which is called the 'stride'. A small value for the stride leads to the best classification performance because this way you miss the least data, but it is computationally more expensive since more patches are created. This creates M patches. Each patch is represented by a vector of $n \times n$ (each voxel intensity is a value in the vector). After this, K-means runs over the M vectors to create K centroids. The resulting centroids which were trained from our input data can be seen in the dictionary image. Using a higher K yields better results [4], but also makes the algorithm computationally more expensive, and thus slower. We chose a K of 200. Now for each new patch the distance to each centroid K^i can be computed and be concatenated into

a vector. This concludes the description of our procedure to map an input patch into a set of features.

Figure 5: Dictionary image of the trained features.

3.2.1 Pooling

Pooling is a technique to reduce the number of features, proposed by Mitra et al [11]. To do this, we first needs to choose the amount of regions to pool over. After this, the patches inside each pool are averaged and taken as a single vector instead of each patch separately. For instance, an intermediate patch of 100×100 pixels with a patch size of 5×5 and stride of 1 pixel and a K of 200 would give $95 \times 95 \times 200 = 18050000$ features. Now if we pool over four regions this is reduced to $4 \times 200 = 800$ features. Because of the larger vertical size of the intermediate patches (16×24) we chose to pool over 6 regions. This corresponds to $6 \times 200 = 1200$ features.

3.2.2 Labeling

In the RUN DMC study two human experts marked WML regions in the MRI images, they each marked different parts of the dataset which we used as 'ground truth'. A candidate region is classified as a WML if it overlaps for eighty or more percent with the ground truth. Otherwise it is labelled as healthy/non-WML tissue. It should be noted that this is not ideal as marking lesions can't be done consistently by the same by expert and especially not by two different experts. Although the human experts are not always marking every lesion (there are in fact a lot of small ones missed), this was still the data we had to use. It took the two experts nine months of full time work to mark all scans.

3.3 Classifying the dataset

For training we have the earlier mentioned 64958 candidate lesions out of 312 MRI-scans. A linear support vector machine (SVM) was trained on this using Azure Machine Learning [1]. For testing we use 32 MRI scans containing 6419 candidate regions.

[1] Azure Machine Learning is a toolset which contains several machine learning algorithms for use on large-scale datasets.

3.4 Evaluating the results

To compare results we also created a dataset with regular features. These features are: minimum intensity, maximum intensity, average intensity and standard deviation of the intensity in each candidate region. These were taken for each candidate lesion, the same ones used for the USFL algorithm, for both training and test. On both training sets (USFL and regular features) a linear SVM is trained and subsequently tested on 32 MRI scans which were not used in the trainingset. The USFL SVM scores better with a low rate of false positives. The regular SVM scores better with high rates of false positives. Overall the two SVM's are in the same league, there is no large difference in performance. The exact results can be found in table 1. Accuracy is calculated with $accuracy = \frac{TP+TN}{P+N}$, this gives the percentage of the dataset that was correctly classified. Precision does the same, but only for the positives: $precision = \frac{TP}{TP+FP}$. Recall: $recall = \frac{TP}{(TP+FN)}$, this gives an indication how good the algorithm can identify a WML correctly. F1 score combines precision and recall by taking their harmonic mean: $F1score = \frac{2\times(precision\times recall)}{precision+recall}$.

	Unsupervised feature learning	Hand-crafted features
True Positive	371	343
False Positive	78	107
True Negative	5207	5187
False Negative	754	782
Accuracy	0.869	0.862
Precision	0.810	0.762
Recall	0.330	0.305
F1 score	0.469	0.463

Table 1: Unsupervised feature learning and hand-crafted features results.

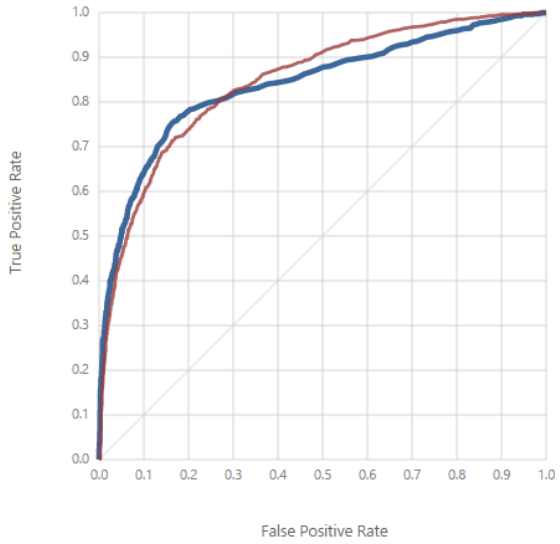

Figure 6: Comparison of unsupervised feature learning approach (blue) to conventional learning with hand-crafted features (red).

Chapter 4

Conclusions

Using the unsupervised feature learning algorithm to classify white matter lesions performs more or less the same as using regular features. The proposed method performs better on lower false positive rate, which is more practical in clinical use. Using unsupervised learning is computationally more expensive since this method has more features per instance, but it makes the process independent of selection of a number of domain-dependent features.

There are some possible ways to improve the results of the USFL algorithm, the number of clusters can be increased - the K parameter -. This way each region can better match one of the extracted features, and the algorithm will have better results [4]. Another option is to pre-process the input images by whitening them first [2]. Stride can't be reduced, that was already at the minimum of one pixel, the patch size might be varied, but further research is required to see if this increases classification performance.

A completely different option is to combine both methods (USFL and handcrafted features), there might be some features that we already know that might be very important, such as location information. This hybrid approach can help to better decide whether or not a sub-image contains a white matter lesion.

Bibliography

[1] John G Baker, Amy J Williams, Catalina C Ionita, Peterkin Lee-Kwen, Marilou Ching, and Robert S Miletich. Cerebral small vessel disease: cognition, mood, daily functioning, and imaging findings from a small pilot sample. *Dementia and geriatric cognitive disorders extra*, 2(1):169–179, 2012.

[2] Isaac N Bankman, Kenneth O Johnson, and Wolfger Schneider. Optimal detection, classification, and superposition resolution in neural waveform recordings. *Biomedical Engineering, IEEE Transactions on*, 40(8):836–841, 1993.

[3] Adam Coates and Andrew Y Ng. The importance of encoding versus training with sparse coding and vector quantization. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 921–928, 2011.

[4] Adam Coates, Andrew Y Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *International Conference on Artificial Intelligence and Statistics*, pages 215–223, 2011.

[5] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2, 2004.

[6] Mohsen Ghafoorian, Nico Karssemijer, Frank-Erik de Leeuw, Tom Heskes, Elena Marchiori, and Bram Platel. Automated Detection of White Matter Lesions in Small Vessel Disease. *in preparation.*

[7] Cüneyt Güler, Geoffrey D Thyne, John E McCray, and Keith A Turner. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology journal*, 10(4):455–474, 2002.

[8] Geoffrey Hinton. A practical guide to training restricted Boltzmann machines. *Momentum*, 9(1):926, 2010.

[9] Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–97. IEEE, 2004.

[10] D-S Lee. Effective Gaussian mixture learning for video background subtraction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):827–832, 2005.

[11] Pabitra Mitra, CA Murthy, and Sankar K. Pal. Unsupervised feature selection using feature similarity. *IEEE transactions on pattern analysis and machine intelligence*, 24(3):301–312, 2002.

[12] Anouk GW van Norden, Karlijn F de Laat, Rob AR Gons, Inge WM van Uden, Ewoud J van Dijk, Lucas JB van Oudheusden, Rianne AJ Esselink, Bastiaan R Bloem, Baziel GM van Engelen, Machiel J Zwarts, et al. Causes and consequences of cerebral small vessel disease. The RUN DMC study: a prospective cohort study. Study rationale and protocol. *BMC neurology*, 11(1):29, 2011.

[13] Anouk GW van Norden, Heleen AC van den Berg, Karlijn F de Laat, Rob AR Gons, Ewoud J van Dijk, and Frank-Erik de Leeuw. Frontal and Temporal Microbleeds are related to cognitive function The Radboud University Nijmegen Diffusion Tensor and Magnetic Resonance Cohort (RUN DMC) Study. *Stroke*, 42(12):3382–3386, 2011.