Probabilistic One Class Learning for Automatic Detection of Multiple Sclerosis Lesions

Yogesh Karpate, Olivier Commowick, Christian Barillot

PROBABILISTIC ONE CLASS LEARNING FOR AUTOMATIC DETECTION OF MULTIPLE SCLEROSIS LESIONS

Yogesh Karpate, Olivier Commowick and Christian Barillot

INSERM, VisAGeS U746, Inria, IRISA UMR CNRS 6074, University of Rennes-1, Rennes, France

ABSTRACT

This paper presents an automatic algorithm for the detection of multiple sclerosis lesions (MSL) from multi-sequence magnetic resonance imaging (MRI). We build a probabilistic classifier that can recognize MSL as a novel class, trained only on Normal Appearing Brain Tissues (NABT). Patch based intensity information of MRI images is used to train a classifier at the voxel level. The classifier is in turn used to compute a probability characterizing the likelihood of each voxel to be a lesion. This probability is then used to identify a lesion voxel based on simple Otsu thresholding. The proposed framework is evaluated on 16 patients and our analysis reveals that our approach is well suited for MSL detection and outperforms other benchmark approaches.

1. INTRODUCTION

Multiple Sclerosis (MS) is an acquired inflammatory, demyelinating disease of the central nervous system. MS is a major cause of disability in young adults prevalent in the northern hemisphere. Automatic detection and segmentation of MS lesions (MSL) from brain magnetic resonance images (MRI) can help to assess the progression of the disease and to evaluate the effectiveness of drug therapy. Automatic MSL detection and recognition is a topic of great importance and still challenging. Although manual lesion detection by experts is the Gold Standard, the objective evaluation of lesion becomes difficult for the radiologist when the number and the resolution of MRI sequences get larger. Consequently, several studies investigated the automatic/semi-automatic segmentation of MSL using multi-channel MR images [1]. One breed of MSL segmentation includes Gaussian Mixture Modeling (GMM) on multispectral MRI, where each multivariate Gaussian probability density function represents a tissue, e.g. cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM). The GMM enables characterization of the image intensities with a reduced number of parameters, which are estimated by a maximum likelihood estimator (MLE). For example, some authors assessed lesions as outliers [2, 3]. Spatial decision forests for MSL segmentation were also investigated [4]. A probabilistic framework for segmentation of gadolinium-enhancing lesions was developed using conditional random fields [5]. The performance of automatic segmentation is constrained by partial volume effects and noise artifacts. Manifold based learning methods were used to characterize pathological deviations from Normal Appearing Brain Tissues (NABT) [6, 7]. However, manifolds have huge computational complexity and do not scale to large data easily.

We propose to use simple intensity features extracted from multi-parametric MRI for lesion detection. The contributions of our work are two-fold:(1) to build an automatic, probabilistic framework to discriminate NABTs and MSLs based upon simple image representation i.e. bag of words features. (2) A probability map is generated from the classifier and used to guide detection based on Otsu thresholding. Note that our framework is closely related to binary classification between two classes (NABT and MSL) of observations. There exists subtle differences, though: binary classification is a symmetric setting for discrimination between two sets whereas we are interested in addressing the asymmetric problem of finding novel instances in one set relative to another. We focus on the form of the problem in which training samples (NABT) without anomalies (MSL) are provided, and we calculate lesion score (anomaly scores) for test data.

The paper is organized as follows. In section 2, the framework for MSL detection is presented. Section 3 gives the details of classifier, development of probability score and detection based on probability score. Results and conclusion of the proposed methods are presented in sections 5 and 6.

2. METHODOLOGY

2.1. Framework

We introduce a novel framework to identify MS lesions, as illustrated in Figure 1. It is based on two main stages. The first stage consists of two parts: (1) extraction of feature vector from multi-channel MRI and its dimensionality reduction using PCA; (2) training a classifier as described in current and next sections. The second stage consists of testing patches from the patient image by applying the learned model. The feature designing from the patient patches is the same as mentioned above. Testing is performed by doing full search over an image by placing a patch at every voxel.

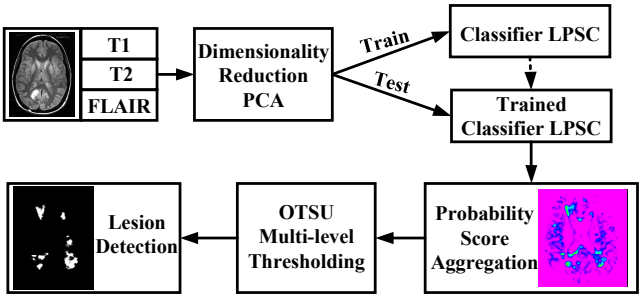

Fig. 1: Workflow for the proposed MSL detection

2.2. Probabilistic Classification

We consider training data formed at each voxel by stacking local intensities from multiple modalities inside a surrounding patch of N voxels. To reduce the dimensionality of each vector, PCA is performed on these vectors, leading to a training set $D = \left\{(\mathbf{x}_i, y_i)|\mathbf{x}_i \in \Re^d, y_i \in (-1, 1)\right\}_{i=1}^{N}$ where y_i is either -1 or $+1$, indicating the class to which the point $\mathbf{x}_i$ belongs. Each $\mathbf{x}_i$ is a d-dimensional real vector. The objective of probabilistic classification is to learn the class-posterior probability $p(y|\mathbf{x})$ of the training samples D. Based on the class-posterior probability, classification of a new sample $\mathbf{x}$ can be carried out $\widehat{y} := \arg\max_{y\in\{-1,+1\}} p(y|\mathbf{x})$ with confidence $p(\widehat{y}|\mathbf{x})$.

One approach to tackle this problem is Least Squares Probabilistic Classification (LPSC) [8]. It employs the linear combination of kernel functions and its parameters are learned by regularized least-squares fitting of the true class-posterior probability. For $y \in \{-1, +1\}$, $p(y|\mathbf{x})$ can be estimated as follows

$$\widehat{p}(y|\mathbf{x}) = \frac{q(y|\mathbf{x}, \widehat{\boldsymbol{\theta}})}{\sum_{j\in\{-1,+1\}} q(y|\mathbf{x}, \widehat{\boldsymbol{\theta}}_j)} \tag{1}$$

Where $\widehat{p}(y|\mathbf{x})$ is modeled with $q(y|\mathbf{x}; \theta_{\mathbf{y}}) = \sum_{b=1}^{B} \theta_{y,b}\phi_b(\mathbf{x}) = \boldsymbol{\theta}_y^T\boldsymbol{\phi}(x)$ and B denotes the number of parameters, $\boldsymbol{\theta_y} = (\theta_{y,1}, ..., \theta_{y,B})^T \in \Re^B$ is the basis function vector. In practice, we use a kernel model by setting $B = N$ and $\phi_b(x) = K(x, x_b)$, where $K(x, x')$ are the kernel functions. Further $\boldsymbol{\theta}_y$ can be estimated analytically as

$$\widehat{\boldsymbol{\theta}}_y = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \rho I_B\right)^{-1} \boldsymbol{\Phi}^T\boldsymbol{\pi}_y \tag{2}$$

where I_B denotes B dimensional identity matrix. In Equation 2, the effect of increasing the size of parameter ρ is both to regularize and to decrease the sensitivity to outliers. In order to estimate lesion probability of a test patch, we do not need to do any extra parameter estimation. $\boldsymbol{\Phi} = (\phi(x_1), ..., \phi(x_N))^T \in \Re^{N\times B}$ is the design matrix $\boldsymbol{\pi}_y$ is N dimensional class indicator variable defined as $\boldsymbol{\pi}_{y,n} = 1$ if $y_n = y$ and zero otherwise. Finally, a posterior is obtained by normalizing over all classes as shown in Equation 1. LPSC is a consistent estimator and is very fast to compute in practice, finding a global optimum for $\boldsymbol{\theta}$ in a single step.

3. LESION DETECTION MODEL

The framework explained in Section 2 was extended to one class learning in [9]. The basic premise holds an assumption that MSL occupy low-density regions of the data space and that a kernel model can be used to characterize the high-density regions of NABT. We consider the case where MSL are not present in training data but present in the test data. Let $y = \{-1, +1\}$ be the NABT and MSL class respectively. The task of MSL detection is to assign a value to the estimate $\widehat{p}(y = +1|\mathbf{x})$ for test data $\mathbf{x}$ given training data. The conditional probability of MSL $p(y = +1|\mathbf{x}, \boldsymbol{\theta})$ with $q(y = +1|\mathbf{x}, \theta_{+1}) = 1 - \boldsymbol{\theta}_{-1}^T\boldsymbol{\phi}(x)$ is computed as

$$q(y = +1|\mathbf{x}, \boldsymbol{\theta}_{+1}) = 1 - q(y = -1|\mathbf{x}, \boldsymbol{\theta}_{-1}) \tag{3}$$

This can be estimated as discussed in Section 2. The parameters learned to model the NABT can therefore be used for MSL detection.

3.1. Aggregate Probability Score

Since the probability map generated by the classification output scores of LPSC is noisy, we adopted the technique from [10] to rectify the probability score obtained from the classifier by smoothing it using a 3D Gaussian. For each voxel, its probability score is propagated to the neighborhood by using isotropic Gaussian kernel. It produces a weighted average of each voxel's neighborhood, with the average weighted more towards the value of the central voxels. The Gaussian kernel for a voxel has a zero mean and standard deviation defined by the probability score of that particular voxel.

3.2. Thresholding Guided Detection

Automatic thresholding is used in final lesion detection. The basic idea is to automatically select an optimal probability threshold value for separating MSLs from the NABT based on their probability distribution. A common thresholding technique, the Otsu method [11], provides satisfactory results for thresholding an image with a histogram of bimodal distribution. This method, however, fails if the histogram is unimodal or close to unimodal. We used revised Otsu method [12] for selecting optimal threshold values for both unimodal and bimodal distributions. Figure 2 shows the pictorial representation of the proposed framework.

4. EXPERIMENTS

Whole-brain MR images were acquired on 16 MS patients and 20 controls. T1-w MPRAGE, T2-w and FLAIR modalities were chosen for the experiment. Expert annotations of lesions were carried out by an expert radiologist on all MS patients. The volume size for T1-w MPRAGE and FLAIR is $256 \times 256 \times 160$ and voxel size is $1 \times 1 \times 1\ mm^3$. For

Table 1: Performance analysis for lesion detection.

	$\varphi = 0.2$		$\varphi = 0.3$		$\varphi = 0.4$		$\varphi = 0.6$	
Method	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
OSVM	0.63±0.01	0.65±0.02	0.59±0.04	0.61±0.02	0.56±0.03	0.60±0.03	0.46±0.03	0.52±0.03
Proposed	**0.79±0.02**	**0.94±0.02**	**0.74±0.03**	**0.86±0.02**	**0.7±0.02**	**0.83±0.01**	**0.63±0.03**	**0.72±0.02**
MCD	0.55±0.02	0.65±0.03	0.51±0.06	0.53±0.04	0.40±0.03	0.52±0.05	0.35±0.03	0.43±0.05

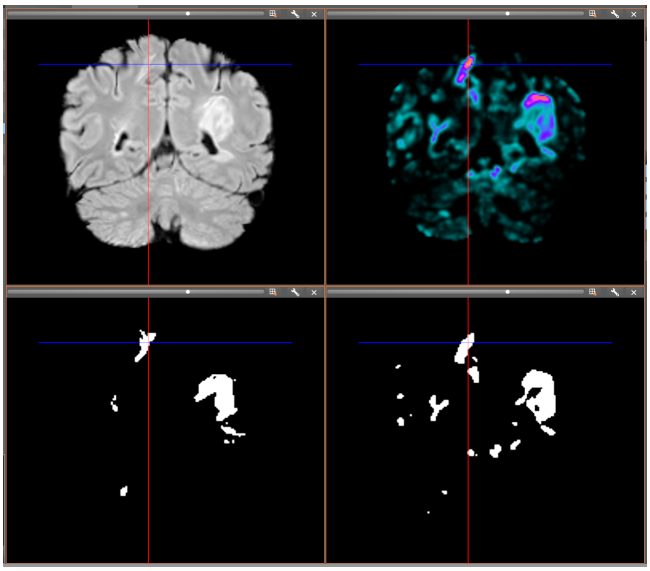

Fig. 2: From top to bottom and left to right: Slice of FLAIR, aggregated probability map, ground truth and obtained lesion mask after Otsu thresholding.

T2-w, the volume size is $256 \times 256 \times 44$ and voxel size is $1 \times 1 \times 3\ mm^3$. All images were acquired on a 3T Siemens Verio (VB17) scanner with a 32-channel head coil. MR images from each patient are de-noised [13], bias field corrected [14] and registered with respect to T1-MPRAGE volume. All images are processed to extract the intra-cranial region. We built a geometrically unbiased atlas for each sequence from the controls [15, 16]. The atlas is considered as the reference image to which all controls and patients images were aligned using intensity normalization [17].

4.1. Experimental Setup

NABT patches were collected from 20 healthy volunteers. The trained classifiers were tested on MS subjects. The features are intensity values extracted from 3^3 patches. The concatenated feature vector was formed using voxel values by extracting patches from FLAIR, T2-w and T1-w MPRAGE. The dimensionality reduction of feature vector was done using PCA. The number of components were decided by keeping 90% of the total variance.

In this experiment, we searched regularization parameter ρ and Gaussian kernel width σ over a range of $10^{[-3:1:2]}$ and $\{m/10, m/5, m/2, m, 3m/2\}$ respectively, where $m := median\left(\{\|\mathbf{x}_i - \mathbf{x}_j\|\}_{i,j=1}^{N}\right)$. We chose the best ρ according to the validation set. We compared results with other benchmark methods:One Class SVM (OSVM) [18] and the Minimum Covariance Determinant estimator (MCD) [19]. OSVM is an unsupervised algorithm that learns a decision function for novelty detection: classifying new data as similar or different to the training set. The MCD estimator is a robust, high-breakdown point estimator of covariance. Assuming that the NABT are Gaussian distributed, it will estimate the inlier location and covariance in a robust way. The Mahalanobis distances obtained from this estimate are used to derive a measure of MSL. For OSVM, the Gaussian width is set to the median distance between samples, which has been shown to be a useful heuristic [18]. We report the results for $\nu = 0.2$ for OSVM. For MCD, the amount of contamination of the data set is assumed to be 0.30.

5. RESULTS AND DISCUSSION

We report the quantitative improvement for identification of lesions: Table 1 presents the precision (Positive Predicted Value) and recall (Sensitivity) of lesion detection averaged across the 16 patients for various overlap thresholds. The lesion is said to be detected if $\frac{R_c \cap R_{GT}}{R_{GT}} \geq \varphi$ where R_c, R_{GT} and φ are respectively the candidate region in the image, the ground truth and a threshold. Table 1 reports values of precision and recall for various thresholds. As from the figures, our approach outperforms other methods. We have a very high recall of 0.94 at $\varphi = 0.2$ and even 0.72 at $\varphi = 0.6$.

Both benchmark methods have a lower performance. A potential reason for OSVM to perform slightly worse is because it is not easily scalable. Besides, the NABT class has big intra class variance. MCD also results in lower performance as it may fail to capture the multi-modal distribution of data. Our framework alleviates these problems by including a large training dataset estimating correct data distribution.

Figure 3 shows the lesion detection results for T2-w. The row represents the image and each column depicts the anatomical slice of patient MRI, corresponding ground truth, lesion detection results with OSVM, MCD and proposed framework. The white labels show the true lesions detected while red ones indicate the brain tissues which are detected as lesion but actually are not part of the lesions. This figure demonstrates visually the ability of our approach to detect lesions. As seen from the last two columns, there is considerable improvement

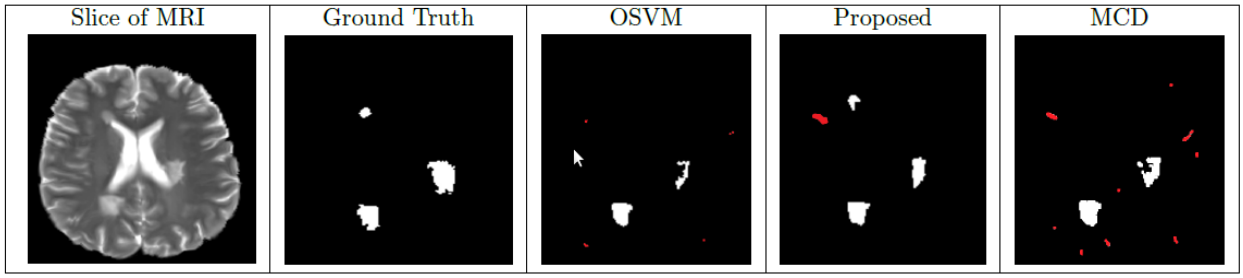

Fig. 3: Lesion detection example.

of lesion detection, thanks to the proposed framework.

6. CONCLUSION

We have introduced a novel method for MSL detection based on Least Squares Probabilistic Classifier. The efficacy of our method was evaluated though rigorous evaluation on MS patients data. We have demonstrated that our method achieves better performance compared to benchmark methods: OSVM and MCD. Our methodology is more suitable for MS lesion analysis because of its ability to capture NABT distribution correctly. This performance suggests that it can provide valuable assistance in detecting the MS lesions in clinical routine with high reliability. The proposed framework is generic in nature and can be extended beyond MSL detection e.g. strokes, tumor. The framework described here allows for exploration of additional MR sequences with or without contrast agents. For example, one can consider infusing T1-w Gadolinium and DTI.

7. REFERENCES

[1] Lladó et al., "Segmentation of multiple sclerosis lesions in brain mri: A review of automated approaches," *Inf. Sci.*, vol. 186, no. 1, pp. 164–185, Mar. 2012.

[2] Koen Van Leemput et al., "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE TMI*, vol. 20, no. 8, pp. 677–688, 2001.

[3] Daniel García-Lorenzo et al., "Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis," *IEEE TMI*, vol. 30, no. 8, pp. 1455–1467, 2011.

[4] Ezequiel Geremia et al., "Spatial decision forests for ms lesion segmentation in multi-channel magnetic resonance images.," 2011, pp. 378–390.

[5] Zahra Karimaghaloo et al., "Automatic detection of gadolinium-enhancing ms lesions in brain MRI using crf," *IEEE TMI*, vol. 31, no. 6, pp. 1181–1194, 2012.

[6] Samuel Kadoury et al., "Manifold-constrained embeddings for the detection of white matter lesions in brain MRI," in *9th IEEE,ISBI*, 2012, pp. 562–565.

[7] Kayhan Batmanghelich et al., "Multiparametric tissue abnormality characterization using manifold regularization," in *SPIE Medical Imaging,*, 2008.

[8] Masashi Sugiyama, "Superfast-trainable multi-class probabilistic classifier by least-squares posterior fitting," *IEICE Transactions*, vol. 93-D, no. 10, pp. 2690–2701, 2010.

[9] John Quinn et al., "A least-squares approach to anomaly detection in static and sequential data," *PR Letters*, vol. 40, pp. 36–40, 2014.

[10] Kishore Reddy et al., "Confidence guided enhancing brain tumor segmentation in multi-parametric MRI," in *9th IEEE, ISBI*, 2012, pp. 366–369.

[11] N Otsu, "A threshold selection method from gray-level histograms," *IEEE Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[12] Ng, "Automatic thresholding for defect detection," *PR Letters*, vol. 27, no. 14, pp. 1644–1649, Oct. 2006.

[13] Pierrick Coupé et al., "An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images," *IEEE TMI*, vol. 27, no. 4, pp. 425–441, 2008.

[14] Nicholas J et al. Tustison, "N4ITK: improved N3 bias correction," *IEEE TMI*, vol. 29, no. 6, pp. 1310–20, 2010.

[15] Alexandre Guimond et al., "Average brain models: A convergence study," *CVIU*, vol. 77, no. 77, pp. 192–210, 1999.

[16] Olivier Commowick et al., "Automated diffeomorphic registration of anatomical structures with rigid parts: Application to dynamic cervical MRI," in *MICCAI'12*, 2012.

[17] Yogesh Karpate, Olivier Commowick, Christian Barillot, and Gilles Edan, "Longitudinal intensity normalization in multiple sclerosis patients," in *Clinical Image-Based Procedures. Translational Research in Medical Imaging - Third International Workshop, CLIP 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014, Revised Selected Papers*, 2014, pp. 118–125.

[18] Bernhard Schölkopf et al., "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[19] Peter Rousseeuw et al., "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, pp. 212–223, 1998.