Automatic multiple sclerosis lesion segmentation with P-LOCUS

Senan Doyle, Florence Forbes, Michel Dojat

Automatic multiple sclerosis lesion segmentation with P-LOCUS

Senan Doyle[1], Florence Forbes[2], and Michel Dojat[3]

[1] Pixyl Medical, Grenoble, France
[2] Inria Grenoble Rhône-Alpes, Grenoble, France
[3] INSERM U1216, GIN, Grenoble, France

Abstract. P-LOCUS provides automatic quantitative neuroimaging bio-marker extraction tools to aid diagnosis, prognosis and follow-up in multiple sclerosis studies. The software performs accurate and precise segmentation of multiple sclerosis lesions in a multi-stage process. In the first step, a weighted Gaussian tissue model is used to perform a robust segmentation. The algorithm avails of complementary information from multiple MR sequences, and includes additional estimated weight variables to account for the relative importance of each voxel. These estimated weights are used to define candidate lesion voxels that are not well described by a normal tissue model. In the second step, the candidate lesion regions are used to populate the weighted Gaussian model and guide convergence to an optimal solution. The segmentation is unsupervised, removing the need for a training dataset, and providing independence from specific scanner type and MRI scanner protocol.

1 Introduction

MS brain lesion segmentation is important for diagnosis, prognosis, and patient follow-up. Typically, this task is performed manually by a medical expert, however automatic methods are sought to alleviate the tedious, time consuming and subjective nature of manual delineation. Automatic methods are motivated by the demand for large-scale multi-center clinical research studies that require precise, repeatable and cost-efficient analysis. Automatic brain image segmentation remains a challenging task for a number of reasons, including the presence of artefacts, and the heterogeneity of MRI scanner protocol.

Automated or semi-automated MS brain lesion detection methods can be classified according to their use of multiple sequences, *a priori* knowledge about the structure of normal brain, and the specific tissue segmentation model. In most approaches, normal brain tissue prior probability maps are used to help identify lesion as an outlier.

Existing methods frequently avail of complementary information from multiple sequences. For example, lesion voxels may appear hyperintense in one modality and normal in another. This is implicitly used by neurologists when examining data. In an statistical framework, complementary information from different sequences can help to better discriminate data generated by different probabilistic

distributions in a multi-dimensional space. Intensity distributions are commonly modeled as multi-dimensional Gaussian distributions. This provides a way to combine the multiple sequences in a single segmentation task but with all the sequences having equal importance. Given that the information content and discriminative power to detect lesions varies between different MR sequences, we adopt a weighted data model, originally proposed by Forbes *et al* [4], that allows for the identification of atypical lesion voxels and the subsequent inclusion of the lesion class as an additional model component.

2   Weighted Model

We consider a finite set V of N voxels on a regular 3D grid. The intensities observed at each voxel are denoted by $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. Each $\mathbf{y}_i = \{y_{i1}, \ldots, y_{iM}\}$ is itself a vector of M intensity values corresponding to M different MR sequences. The goal is to assign each voxel i to one of K classes considering the observed features data $\mathbf{y}$. For brain tissue segmentation, we consider in general 3 tissues plus some possible additional classes to account for lesions in pathological data. We denote the hidden classes by $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$, and the set in which $\mathbf{z}$ takes its values by $\mathcal{Z}$. Typically, the $\mathbf{z}_i$'s take their values in $\{1 \ldots K\}$. We consider non-negative weights $\omega = \{\omega_i, i \in V\}$ in a state space denoted by $\mathcal{W}$ and with $\omega_i = \{\omega_{i1}, \ldots, \omega_{iM}\}$. In our general setting the weights are sequence and voxel-specific.

The segmentation task is recast into a missing data framework in which $\mathbf{y}$ are observations and $\mathbf{z}$ are missing variables. Their joint distribution $p(\mathbf{y}, \mathbf{z}|\omega; \psi)$ is governed by the weights $\omega \in \mathcal{W}$ and parameters $\psi \in \underline{\Psi}$, which are both unknown and need to be estimated. A prior distribution $p(\omega)$ is defined on the weights, considered additional missing variables. Denoting the parameters by $\psi = \{\beta, \phi\}$, we assume that the joint distribution $p(\mathbf{y}, \mathbf{z}, \omega; \psi)$ is a MRF with the following energy function:

$$H(\mathbf{y}, \mathbf{z}, \omega; \psi) = H_{\mathbf{Z}}(\mathbf{z}; \beta) + H_W(\omega) + \sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$$

where the energy term $H_W(\omega)$ involving only ω does not depend on ψ and the $g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$s are probability density functions of $\mathbf{y}_i$.

For the data term $\sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$ in (1), we consider M-dimensional Gaussian distributions with diagonal covariance matrices. For each class k, ${}^t(\mu_{k1}, \ldots, \mu_{kM})$ is the mean vector and $\{s_{k1}, \ldots, s_{kM}\}$ the covariance matrix components. When $\mathbf{z}_i = k$, then $\mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, s_{\mathbf{z}_i m})$ represents the Gaussian distribution with mean μ_{km} and variance s_{km}. The whole set of Gaussian parameters is denoted by $\phi = \{\mu_{km}, s_{km}, k = 1 \ldots K, m = 1 \ldots M\}$. Our data term is then defined by setting

$$g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi) = \prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, \frac{s_{\mathbf{z}_i m}}{\omega_{im}}),$$

which is proportional to $\prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, s_{\mathbf{z}_i m})^{\omega_{im}}$. Intuitively, the impact of a larger ω_{im} is to give more importance to the intensity value y_{im} in the model. A weight of one recovers the standard multivariate Gaussian case.

The missing data term $H_{\mathbf{Z}}(\mathbf{z}; \beta)$ in (1) is set to a standard Potts model, with external field ξ and spatial interaction parameter η, and whose energy is

$$H_{\mathbf{Z}}(\mathbf{z}; \beta) = \sum_{i \in V}(\xi_{i\mathbf{z}_i} + \sum_{j \in \mathcal{N}(i)} \eta \, \langle \mathbf{z}_i, \mathbf{z}_j \rangle),$$

where $\mathcal{N}(i)$ denotes the voxels neighboring i and $\langle \mathbf{z}_i, \mathbf{z}_j \rangle$ is 1 when $\mathbf{z}_i = \mathbf{z}_j$ and 0 otherwise. Parameter $\beta = \{\xi, \eta\}$ with $\xi = \{^t(\xi_{i1} \ldots \xi_{iK}), i \in V\}$ being a set of real-valued K-dimensional vectors and η a real positive value.

The weights are assumed independent from parameters ψ and independent across modalities. The simplest choice is to define a prior $p(\omega) = \prod_{m=1}^{M} \prod_{i \in V} p(\omega_{im})$ where each $p(\omega_{im})$ is a Gamma distribution with hyperparameters α_{im} (shape) and γ_{im} (inverse scale). Thus

$$H_W(\omega) = \sum_{m=1}^{M} \sum_{i \in V}((\alpha_{im} - 1) \log \omega_{im} - \gamma_{im} \, \omega_{im}).$$

In practice, the set of hyperparameters is fixed so that the modes of each prior $p(\omega_{im})$ are located at some *expert weights* $\{\omega_{im}^{exp}, m = 1 \ldots M, i \in V\}$ accounting for some external knowledge, if available. Formally, we set $\alpha_{im} = \gamma_{im} \, \omega_{im}^{exp} + 1$ to achieve this. The expert weights can be chosen according to the specific task. For example, when voxels with typical lesion intensities are not numerous enough to attract a model component, increasing the expert weight for some of them will help in biasing the model toward the identification of a lesion class.

A solution to the model is found using the Expectation-Maximization (EM) framework [2] combined with a variational approximation for tractability in the presence of Markov dependencies. In particular, the mean field principle provides a deterministic way to deal with intractable MRF models[1].

3   Method

Of the four possible input sequences available, the method uses only the unprocessed T1-weighted and Flair. The images are masked, co-registered and corrected for inhomogeneities using the N4 algorithm.

The segmentation process consists of two stages, as detailed in [3]. In the first step, we set $K = 3$, considering only the three normal tissue classes (with all ω_{im}^{exp} and γ_{im} set to 1). The ξ parameters in the MRF prior are set to $\xi_{ik} = \log f_{ik}$ where f_{ik} is the normalized value given by a normal tissue atlas. The interaction parameter η is estimated using a stochastic gradient descent method as specified in [1]. The estimated weights for the Flair sequence are thresholded

at a value of one to identify outlier regions corresponding to candidate lesion regions. This candidate region is refined using additional intensity, location and size constraints, as in [7, 5, 6]. Retained lesions are hyperintense in Flair, confined to white matter and greater than $5mm^3$.

In the second step, the candidate region is used to specify the parameters of the weight distribution in a $K = 4$ segmentation setting. We set γ_{im} according to: $\gamma_{im} = \gamma_{\mathcal{L}}$ for all $i \in \mathcal{L}$ and $\gamma_{im} = \gamma_{\bar{\mathcal{L}}}$ for all $i \notin \mathcal{L}$, where $\gamma_{\mathcal{L}}$ and $\gamma_{\bar{\mathcal{L}}}$ are values to be specified. We set $\gamma_{\bar{\mathcal{L}}} = 1000$ to express our a priori trust in the estimation of the normal brain tissue classes from the preliminary first step, and set $\gamma_{\mathcal{L}} = 10$ to allow some flexibility in the weight estimation.

The expert weight is fixed to $\omega_{\mathcal{L}} = 2$ and $\omega_{\bar{\mathcal{L}}} = 1$. Large values of $\omega_{\mathcal{L}}$ make the lesion class more representative and handle the possibility of very small lesions, while a small $\omega_{\bar{\mathcal{L}}}$ ensures that the weighting of a large candidate lesion region does not affect the estimation of other classes. A post-processing step removes artefacts based on spatial location.

4   Conclusion

The adaptive weighting model facilitates accurate and robust MS lesion segmentation from T1-weighted and Flair sequences. The advantage of this approach is that the weights, and therefore the 'outliers' are obtained in a multi-sequence framework that provides a more robust estimation of normal tissue parameters. The method is independent of MRI scanner, and does not require training data.

References

1. G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pat. Rec.*, 36(1):131–144, 2003.
2. A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39:1–38, 1977.
3. F. Forbes, S. Doyle, D. Garcia-Lorenzo, C. Barillot, and M. Dojat. A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation. In *AISTATS, Sardignia, May 13-15*, 2010.
4. Florence Forbes, Senan Doyle, Daniel García-Lorenzo, Christian Barillot, and Michel Dojat. Adaptive weighted fusion of multiple mr sequences for brain lesion segmentation. In *ISBI*, pages 69–72. IEEE, 2010.
5. O. Freifeld, H. Greenspan, and J. Goldberger. Lesion detection in noisy MR brain images using constrained GMM and active contours. In *IEEE ISBI*, pages 596–599, 2007.
6. D. Garcia-Lorenzo, L. Lecoeur, D.L. Arnold, D. L. Collins, and C. Barillot. Multiple Sclerosis lesion segmentation using an automatic multimodal graph cuts. In *MICCAI*, pages 584–591, 2009.
7. K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE trans. Med. Ima.*, 20(8):677–688, 2001.