Hierarchical Conditional Random Fields for Detection of Gad-Enhancing Lesions in Multiple Sclerosis

Zahra Karimaghloo¹, Douglas L. Arnold³,
D. Louis Collins², and Tal Arbel¹

¹ Centre for Intelligent Machines, McGill University, Canada

² Montreal Neurological Institute, McGill University, Canada

³ NeuroRx Research, Montreal, Canada

Abstract. The detection of gad-enhancing lesions in brain MRI of Multiple Sclerosis (MS) patients is of great interest since they are important markers of disease activity. However, many of the enhancing voxels are associated with normal structures (i.e. blood vessels) or noise in the MRI, making the detection of gad-enhancing lesions a challenging task. Furthermore, these lesions are typically small and in close proximity to vessels. In this paper, we present an automatic, probabilistic Hierarchical Conditional Random Field (HCRF) framework for detection of gad-enhancing lesions in brain images of patients with MS. In the first level, a CRF with unary and pairwise potentials is used to identify candidate lesion voxel. In the second level, these lesion candidates are grouped based on anatomical and spatial features, and feature-specific lesion based CRF models are designed for each group. This lesion level CRF incorporates higher order potentials which account for shape, group intensities and symmetries. The proposed algorithm is trained on 92 multimodal clinical datasets acquired from Relapsing-Remitting MS patients during multicenter clinical trials and is evaluated on 30 independent cases. The experimental results show a sensitivity of 98%, a positive predictive value of 66% and an average false positive count of 1.55, outperforming the CRF and MRF frameworks proposed in [1].

1 Introduction

Multiple Sclerosis (MS) is a disorder of the central nervous system which is characterized by focal, inflammatory lesions of the white matter (WM) which appear hyperintense on T2-weighted MRI (i.e. *T2w lesions*.) A subset of MS lesions showing active inflammation can be identified in T1-weighted (T1w) scans acquired after injection of gadolinium-based contrast agents (i.e. *gad-enhancing lesions*). The number and volume of gad-enhancing lesions are important biomarkers of disease activity and can be used in the development of drugs for MS. Gad-enhancing lesions are generally segmented manually, a laborious task subject to intra- and inter-rater variability. As clinical trials for MS treatments usually involve enormous amounts of data from multiple centers, it is desirable

Fig. 1. (a), (b) and (c) respectively show the post-contrast T1w (T1c), the binary mask of the voxels having sufficient enhancement to be considered as lesions and the manual labels (in green). In (b), red and green color mark some non-lesional and lesional enhancements respectively. (d) and (e) show zoomed in images of one of the active lesions without and with the label, respectively.

to have an automatic segmentation method that is robust to data variability due to different scanners and protocols. Unfortunately, automatic identification of enhancing lesions is very challenging. This is due, in part, to variability in size (as small as 3 voxels), texture, intensity and location across patients. Furthermore, the majority of the enhancing voxels are associated with non-lesional, normal structures (e.g. mostly blood vessels) and MRI noise (Fig.1). These factors make the clinical objective of detecting *all* of the gad-enhancing lesions while maintaining low False Positive (FP) counts very difficult. Existing methods for gad-enhancing lesion segmentation are either not fully automatic [2,3], depend on non-standard MRI acquisition sequences [3,4], or require the prior segmentation of T2w lesions in order to remove FPs [4,5]. A Conditional Random Fields (CRF) [1] classifier was recently developed for this task, without relying on the pre-segmentation of T2w lesions and using only commonly acquired MRI sequences (i.e. pre- and post-contrast T1w, T2w, PDw and FLAIR). The CRF was shown to outperform standard MRF, SVM and linear regression models in terms of FPs, virtually eliminating the number of False Negative (FN) lesions. However, although some shape information was used, this model used mainly local, voxel-level unary and pairwise potentials for classification, and was not powerful enough to remove all the FP lesions. Higher order potentials can express more complex image features [6]. However, incorporating them is computationally expensive when considering all enhancing voxels.

In this work, we introduce a probabilistic Hierarchical CRF (HCRF) model for the automatic detection of gad-enhancing lesions which allows context to be incorporated at multiple levels sequentially. In particular, our framework includes: (1) *a voxel level CRF* with unary and pairwise potentials to obtain lesion candidates through a voxel-based classification, and (2) *a regional, lesion level CRF* incorporating higher order potentials, allowing for the integration of shape based features (e.g. elongation indicating vessels), groupwise intensity characteristics and global features (e.g. symmetry).

The experimental results of applying the HCRF classifier on real, multi-centre clinical trial images acquired from 122 patients with Relapsing Remitting MS (RRMS) yields a 98% sensitivity rate, 0.66 positive predictive value (PPV) and an average of 1.55 FP counts per patient when compared to a set of “silver standard” manual labels attained by expert consensus¹. The results also show that the new HCRF method outperforms standard MRF and previous CRF methods.

2 Method

The goal of image classification is to assign to each voxel in the image a label. Let us denote $\mathbf{x}_i \in \mathcal{R}^d$ as the observation vector at the voxel i and $y_i \in \{1, 0\}$ as the label of the i^{th} instance. We now describe the 2 main levels of inference in our framework.

2.1 Voxel Level CRF

At the first level of inference, we develop a Conditional Random Field (CRF) [7] classifier at the voxel level. A CRF is a discriminant graphical model that directly estimates the parameters of the conditional posterior, $p(Y|X)$, by learning a mapping from observations to class labels, generally formulated as:

$$p(Y|X) = \frac{1}{Z} \exp\left(\sum_{i=1}^n \phi_v(y_i|X) + \sum_{i,j,i \neq j} \varphi_v(y_i, y_j|X)\right) \quad (1)$$

where $X = \{\mathbf{x}_i\}_1^n$ and $Y = \{y_i\}_1^n$. n and Z indicate the total number of pixels and the normalization term, respectively. ϕ_v and φ_v are the *unary* and the *pairwise* potentials at the voxel level, evaluating the likelihood of a voxel taking a particular label and a pair of neighbouring voxels taking on different labels, respectively. The unary potential is modeled as:

$$\phi_v(y_i|X) = \log p(y_i|\mathbf{x}_i) = \log(\sigma(g(\mathbf{x}_i))) \quad (2)$$

where $\mathbf{x}_i$ is the observation vector at voxel i (e.g. intensity values from different MRI sequences) and σ denotes the sigmoid function. Similar to [1], we use a Relevance Vector Machine (RVM) classifier to model $p(y_i|\mathbf{x}_i)$.

The pairwise potential in a CRF model permits learning the relationships between the labels of neighbouring nodes given the observed data. This potential is usually modeled based on the absolute difference in neighbouring observations. In this work, we incorporate the sign of the difference as well in order to account for the directional relationship between the enhancing lesion voxels and the surrounding non-enhanced voxels (e.g. lesion voxels should typically be hyper intense in T2w and hypo intense in T1w pre-contrast comparing to the surrounding non-enhanced voxels). Therefore, to evaluate the likelihood of an

¹ It is important to note that errors in manual labeling are very likely present.

observed edge, not only should the gradient in the associated observations be high but the sign of the gradient should also support the labelling configuration. We model the pairwise potential as:

$$\varphi_v(y_i, y_j | \mathbf{x}_i, \mathbf{x}_j) = -[\alpha_{y_i y_j} f_{y_i y_j}(\mathbf{x}_i, \mathbf{x}_j) + \beta_{y_i y_j}](1 - \delta(y_i - y_j)) \quad (3)$$

where $\alpha_{y_i y_j}$ and $\beta_{y_i y_j}$ are model parameters and δ is the Kronecker delta function. The function $f_{y_i y_j}$ is the data dependant term.

Training at the voxel level of inference is adapted to eliminate the possibility of having any FNs, at the expense of additional FPs. A lesion level CRF will now be developed to remove the remaining FPs based on higher level spatial and structural information.

2.2 The Regional Lesion-Based CRF

After the first level of CRF, voxels with the same label are grouped together to form lesion candidates (each lesion candidate is surrounded by a bounding box, B). As lesion characteristics are highly non-homogeneous throughout the image volume, lesion candidates are grouped in accordance with their spatial and anatomical characteristics. Features are specifically chosen in order to perform this grouping. Once grouped, a feature specific CRF model with the incorporation of higher order potentials is learned for each grouping to optimally distinguish false positive lesion candidates from true lesions.

The spatial and anatomical features used for the grouping includes: average location of each lesion candidate along the axis perpendicular to the axial plane (z -axis), average WM and partial volume (PV) values of each region obtained from spatial probabilistic atlases registered to each patient. The range of these three features are divided to three parts yielding 27 groups overall. Empirically, it was observed that these features are able to efficiently represent the diversity of lesion candidates. After grouping lesion candidates based on these features, we then learn a feature specific CRF model for each group as follows:

$$p(Y|X) = \frac{1}{Z} \exp\left(\sum_{i=1}^n \phi_r(y_i|X) + \sum_{i,j,i \neq j} \varphi_r(y_i, y_j|X) + \sum_{b \in B} \psi_r(y_b|X)\right) \quad (4)$$

where ϕ_r and φ_r are modeled similar to Eq. 2 and 3, respectively. The higher order potential takes the form of the robust P^N model (for details on model, see [6]):

$$\psi_r(y_b|X) = \begin{cases} N(y_b) \frac{1}{Q} \gamma_{max} & \text{if } N(y_b) \leq Q . \\ \gamma_{max} & \text{otherwise.} \end{cases} \quad (5)$$

where b denotes the voxels within the bounding box, B , and $N(y_b)$ denotes the number of voxels in B not taking on the dominant label. Q is the truncation parameter which controls the rigidity of the higher order potential. γ_{max} is determined by evaluating the quality of the lesion candidate within each bounding box and is based on the average intensity values of the region, the shape of the

underlying enhanced structure (to distinguish enhancing lesions from vessels) and assessing if the enhanced region has any symmetrical pair (to distinguish FPs causing from the enhancements at the choroid plexus regions). For lesion candidates where the assessed quality of the candidate is high, γ_{max} has a lower value allowing for the voxels within the bounding box to take on different labels. Otherwise, γ_{max} is high and the model favours all voxels within the bounding box to take on the same label (i.e. we favoured label zero to eliminate FPs).

We use the piecewise learning approach, to learn the parameters of the model [8]. This technique makes the computation tractable by breaking the model into disjoint pieces, each of which is trained independently and integrated afterwards. The parameters of the unary potential at both levels are learned within a Bayesian inference, while the parameters of the pairwise model are learned within a cross validation method. The parameters of the higher order potential are manually selected to minimize the error on the training set. In the inference stage, considering the CRF model at each level and its learned parameters, we seek the most probable labelling that maximizes the conditional probability of Eq. 1 and 4. Graph Cuts are chosen to solve this optimization problem primarily because of their ability to find globally optimal solutions for binary classifications [9]².

3 Experiments and Results

3.1 Data Pre-processing

The data was acquired from 122 patients with RRMS as part of a multi center clinical trial (31 centers). The patients had varying levels of gad-enhancing lesion loads, located in different areas of the brain WM, and showed varying amounts of brain atrophy. All MRI volumes consist of 3mm thick axial slices with 1mm $\times$ 1mm intra-plane resolution. Each acquisition was composed of five sequences: pre- and post-contrast T1w, T2w, PDw and FLAIR. For the particular data set that we had access to, the “silver standard” manual labels were determined using a protocol where two trained experts label the data separately, followed by consensus agreement. Prior to classification, pre-processing steps including bias-field inhomogeneity correction as well as removal of non-brain regions from the MRI are performed. Furthermore, all intra-subject sequences are registered to a common coordinate space and the intensity histogram of all sequences is normalized [10].

The HCRF classifier is trained on 92 randomly selected MRI volumes and tested on the remaining 30 cases. There is at least one patient from each center in the training set³. Fig. 2 shows an example of the performance of the various components of the HCRF framework. While the classification results obtained from the unary potential and unary + pairwise potentials at the voxel level show many FPs, the proposed HCRF model in (2(e)) successfully captured all four enhancing lesions without any FPs.

² We used the Matlab wrapper for robust higher order potentials by Shai Bagon: [http://www.wisdom.weizmann.ac.il/~sim\\$Bagon/matlab.html](http://www.wisdom.weizmann.ac.il/~sim$Bagon/matlab.html).

³ We had insufficient data from each center to provide a statistical analysis per center.

Fig. 2. The performance of the different components of the proposed HCRF classifier. The shown images are: the post contrast T1w (a), the binary mask (b), the classification results of using only the unary potential, $\phi_v(y_i|X)$ (c), using both the unary and pairwise potentials, $\phi_v(y_i|X) + \varphi_v(y_i, y_j|X)$ (d), and the HCRF model (e). Green and cyan show TP and FP regions respectively.

We compare the performance of the HCRF model with the MRF and CRF models described in [1]. Here, the MRF is modeled similar to the CRF model at the voxel level where the pairwise potential is modeled as: $\varphi_v^{MRF}(y_i, y_j) = \beta y_i y_j$ and β is the Ising parameter. In [1], a post-processing step is used to remove any regions with fewer than 3 connected voxels. We found that this can result in missing small lesions, specifically when only 1 or 2 voxels of the lesion are captured by the model. Therefore, there is no post-processing step in our HCRF model. Instead, we try to effectively learn the neighbourhood characteristics of small lesions through the higher order potentials to better distinguish between lesional enhancements and FPs. Fig. 3 shows a qualitative example of the results for a case with one small enhancing lesion in the shown slice, which only the HCRF framework successfully captures. It is missed by the MRF due to the over-smoothing effect of the Ising model, and because it does not consider the interaction of neighbouring observations. For CRF [1], even though 2 voxels of the lesion were captured, the aforementioned post-processing step has removed the lesion.

The performance of our HCRF model is also quantitatively evaluated against the MRF and CRF [1] in Table 1. For a fair comparison, the CRF [1] without the aforementioned post-processing step ($CRF[1]_{no\ post}$) is also included. Comparisons are based on: the sensitivity rate ($\frac{TP}{TP+FN}$ for each scan), the average FP number and the average PPV ($\frac{TP}{TP+FP}$ for each scan). Note that the goal is high lesion detection rates and not the exact delineation of lesion boundaries. Hence, if only one voxel of an enhancing lesion is captured, that is counted as a TP⁴. If all of the voxels of an enhancing lesion are missed, that is counted as an FN and any candidate that does not correspond to an enhancing lesion is counted as an FP. The results show that the HCRF model has the lowest FP count and the highest sensitivity rate over all methods. The pairwise potential in the first

⁴ For the detected lesions, 70% of the area of enhancing lesions were captured on average.

Fig. 3. Comparison of the performance of the HCRF classifier against the MRF and CRF [1] models. The shown images are: the post-contrast T1w (a), the binary mask (b), the classification results of the MRF (c), CRF (d) and HCRF (e). Red and green show FN and TP regions respectively.

Table 1. Quantitative comparisons of the performance of different models. The first and second columns show the voxel-level classification results for the CRF using only the unary ($\text{CRFv}(\phi)$) and the combination of the unary and pairwise potentials ($\text{CRFv}(\phi, \varphi)$) respectively. The third column (HCRF) is the proposed hierarchical model. The forth and fifth columns show results using the CRF model found in [1], with ($\text{CRF}[1]$) and without post processing ($\text{CRF}[1]_{no\ post}$) respectively. The last column shows the classification result using an MRF model.

	$\text{CRFv}(\phi)$	$\text{CRFv}(\phi, \varphi)$	HCRF	$\text{CRF}[1]$	$\text{CRF}[1]_{no\ post}$	MRF
Sensitivity	0.98	0.98	0.98	0.88	0.92	0.81
Avg FPs	22.29	9.13	1.55	1.58	12.03	2.45
PPV	0.13	0.27	0.66	0.58	0.27	0.48

CRF level has decreased the average FP count by 59% (while maintaining high sensitivity) over using the unary potential alone. For the complete HCRF, incorporation of the higher order potential in the second level has decreased the average FPs by 83%. The low sensitivity rate in CRF [1] is due to missed lesions or the 1 or 2 lesion voxels filtered by post-processing. It should be noted that even though the average FP count is almost the same for the CRF [1] and HCRF, no post-processing is performed in HCRF to remove the small regions (and the sensitivity is much higher). As we see for $\text{CRF}[1]_{no\ post}$, the FP count increases drastically without any post-processing step. The Ising parameter for the MRF model was set to give an average FP count close to that of the HCRF model. Even with a higher FP count, the MRF has lower sensitivity as a result of oversmoothing.

4 Discussion

In this paper, we propose a new a Hierarchical CRF model to detect gad-enhancing lesions in brain MRI that embeds contextual information at multiple

levels. At the first level, a CRF model with up to pairwise potentials are used to detect the lesion candidates. In the second level, we group the similar lesion candidates according to features. For each group, a new CRF model embedding higher order potentials is learned, allowing for the integration of high level features (e.g. shape and symmetry) and boosting the discrimination power of our model. Our classifier is tested on 30 multi-centre clinical data set from RR MS patients with varying loads of gad-enhancing lesions. The experimental results show the advantage of the method, showing higher sensitivity rate while maintaining a very low false positive rate over other approaches. In the future, a full cross validation will be performed to precisely study the sensitivity of the parameters to different training sets. More sophisticated models for the higher order potential and investigation of the performance of the CRF classifier for other contexts are among the future works as well.

References

1. Karimaghloo, Z., Shah, M., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T.: Automatic detection of Gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields. *IEEE Trans. Med. Imag.* 31(6), 1181–1194 (2012)
2. Miki, Y., Grossman, R., Udupa, J., Samarasekera, S., Buchem, M., Cooney, B., Kolson, D., Constantinescu, C., Pollack, S., Polansky, M., Mannon, L.: Computer-assisted quantitation of enhancing lesions in multiple sclerosis: correlation with clinical classification. *Am. J. Neuroradiol.* 18, 705–710 (1997)
3. Bedell, B., Narayana, P.: Automatic segmentation of Gadolinium-enhanced multiple sclerosis lesions. *Magn. Reson. Med.* 39, 935–940 (1998)
4. He, R., Narayana, P.: Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. *Med. Phys.* 29, 1536–1546 (2002)
5. Datta, S., Sajja, R., He, R., Gupta, K., Wolinsky, S., Narayana, A.: Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis. *J. Magn. Reson. Imaging* 25, 932–937 (2007)
6. Kohli, P., Ladicky, L., Torr, P.: Robust higher order potentials for enforcing label consistency. *Int. J. Comput. Vision*, 302–324 (2009)
7. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *CML*, pp. 282–289 (2001)
8. Sutton, C., McCallum, A.: Piecewise pseudo-likelihood for efficient training of conditional random fields. In: *ICML*, pp. 863–870 (2007)
9. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE T. Pattern Anal.*, 1124–1137 (2004)
10. Nyul, L., Udupa, J.: On standardizing the MR image intensity scale. *Magn. Reson. Med.* 42, 1072–1081 (1999)