

# Simplified Labeling Process for Medical Image Segmentation

Mingchen Gao<sup>1</sup>, Junzhou Huang<sup>2</sup>, Xiaolei Huang<sup>3</sup>, Shaoting Zhang<sup>1</sup>, and  
Dimitris Metaxas<sup>1</sup>

<sup>1</sup> CBIM Center, Rutgers University, Piscataway, NJ, 08554

<sup>2</sup> Department of Computer Science and Engineering, University of Texas at  
Arlington, TX, 76019

<sup>3</sup> Computer Science and Engineering Department, Lehigh University, PA, 18015

**Abstract.** Image segmentation plays a crucial role in many medical imaging applications by automatically locating the regions of interest. Typically supervised learning based segmentation methods require a large set of accurately labeled training data. However, the labeling process is tedious, time consuming and sometimes not necessary. We propose a robust logistic regression algorithm to handle label outliers such that doctors do not need to waste time on precisely labeling images for training set. To validate its effectiveness and efficiency, we conduct carefully designed experiments on cervigram image segmentation while there exist label outliers. Experimental results show that the proposed robust logistic regression algorithms achieve superior performance compared to previous methods, which validates the benefits of the proposed algorithms.

## 1 Introduction

To assist doctors locate pathologies, automatic segmentation of different regions of medical images is very useful. Supervised learning based segmentation method, which use manually segmented training data as references, has superior performance. Those methods perform well as long as the feature space sufficiently distinguishes each label. They are relatively computationally efficient and not sensitive to parameters. Many popular learning methods have been applied to solve challenging medical problems, such as, but not limited to support vector machine (SVM) [12], neural network [11, 2], conditional random field (CRF) [5], logistic regression (LR) [3].

One disadvantage of the learning-based segmentation is the requirement of manual interaction for obtaining training data. Manual segmentation by different people are subjective due to the lack of standard when performing the manual segmentation. In reality, the segmentation training data are not correctly, perfectly labeled. However, all the previous methods ignored the imperfection of training data. In such cases the training set is misleading, the guidance given by the labels may not be reliable. Consequently learning results may not totally reliable. On the other hand, the precisely labeling process is time consuming,

laborious and herpaps not necessary. In this paper, we take the advantage of learning methods handling label outliers and propose a simplified lableing process. The proposed simplified labeling process could significantly reduce doctors time and effort in labeling the training data. And the proposed learning methods would handle the labeling errors introduced by the new labeling process. While learning in the idealized setting has been thoroughly studied, learning in the presence of labeling errors has not been well explored. And nothing has been done for medical image segmentation with labeling error.

According to the practical consideration mentioned above, we are the first to systematically study logistic regression and sparse logistic regression method with outliers and apply it on cervigram image segmentation. Logistic regression (LR) is a classical method for classification and has been widely used in many applications. Sparse logistic regression, which is the logistic regression with  $l_1$ -norm regularization, is a very effective method for classification on large scale data with high dimension. Both methods could be adopted in image segmentation.

In this paper, we propose a robust sparse logistic regression method to handle the classification problem with label outliers. In our method, a latent variable is introduced for the true correct labels, and then we estimate the probability of labels being flipped. Conjgate gradient method and Lassplore algorithm [8] are used respectively to minimize the loss function under the noise assumption. After optimization, the probability of the flipped labels is updated using the new classification estimation. The algorithm can quickly converge after several iterations.

To demonstrate the effectiveness of the proposed algorithm, we apply the proposed method on the task of automatically segmenting the biomarker Ace-toWhite (AW) regions in an archive of 60,000 images of the uterine cervix [4]. The most important observation in a cervigram image is the AW region, which is caused by whitening of potentially malignant regions of the cervix epithelium, following application of acetic acid to the cervix surface. Since the texture, size and location of AW regions have been shown to correlate with the pathologic grade of disease severity, accurate identification and segmentation of AW regions in cervigrams have significant implications for diagnosis and grading of cervical lesions. Carefully designed experiments on the cervigram images with label outliers demonstrate the superior performance of the proposed method and validate the benefits of the proposed algorithm in practical applications.

## 2 Related Work

In this section, we briefly review logistic regression and sparse logistic regression.

Logistic regression is a conditional probability of the label  $y$ , given a sample  $x \in \mathbb{R}^n$ , and a weight vector  $w \in \mathbb{R}^n$ :

$$p(y = 1 | x, w) = \sigma(w^T x) = \frac{1}{1 + e^{-w^T x}}. \quad (1)$$

Given a set of training examples  $D = \{(x_1, \hat{y}_1), \dots, (x_i, \hat{y}_i), \dots, (x_m, \hat{y}_m)\}$ , we want to learn a classifier  $y = f(x)$ . Here  $x_i \in \mathbb{R}^n$  denotes a sample, and  $\hat{y}_i$  is the corresponding observed class label. The likelihood function associated with these  $m$  samples is defined as

$$\begin{aligned} f(w) &= \log \prod_{i=1}^m p(\hat{y}_i | x_i, w) \\ &= \sum_{i=1}^m \frac{\hat{y}_i + 1}{2} \log(p(\hat{y}_i = 1 | x_i, w)) + \frac{1 - \hat{y}_i}{2} \log(p(\hat{y}_i = -1 | x_i, w)), \end{aligned}$$

which is a smooth and convex function. We can determine  $w$  by minimizing the logistic loss:

$$\hat{w} = \underset{w}{\operatorname{argmin}} f(w). \quad (2)$$

This is a smooth convex optimization problem. Regularization is usually used to avoid the overfitting problem. The  $l_2$ -norm regularization leads to a smooth and differentiable unconstrained convex optimization problem. Standard optimization algorithm such as Newton method and conjugate gradient method can be applied for solving such optimization problem [1, 9].

The  $l_1$ -norm regularization is used to obtain a sparse model. However, the  $l_1$ -norm regularization term is non-differentiable. There are many algorithms proposed in the past for solving the  $l_1$  regularized logistic regression [7, 10]. Liu et al. proposed the Lassplore algorithm for solving large scale sparse logistic regression. They formulate the problem as the  $l_1$ -ball constrained logistical regression formulation, in which the objective function is continuously differentiable, and the problem domain set is closed and convex [8].

### 3 Proposed Algorithm

We propose a new simplified labeling process. Doctors would label a bounding box around the region of interest rather than precisely label the boundaries. Please see Figure 2 (a,b,c) as illustration of the proposed labeling process. Figure 2(a) is the image for segmentation. As we can see from (b), the groundtruth, the boundary of the region of interest is nontrivial and complicated. Our labeling process is illustration in (c), which would roughly include the region of interest in a bounding box. The simplified labeling process would significantly save time and effort of doctors in labeling the training data.

Considering the new labeling process, which introduces some labeling noise, none of the previous work has been able to handle label outliers. In this paper, we propose a robust logistic regression method to handle the classification problem on large scale data with label outliers.

### 3.1 Robust Logistic Regression

If there exists some label noise, Eq. 2 is no longer valid. We introduce a latent variable  $y_i$ , which is the true label. The probability of the observed label  $p(\hat{y}_i | x_i, w)$  is written as the following:

$$\begin{aligned} S_i^k &\doteq p(\hat{y}_i = k | x_i, w) \\ &= \sum_{j=0}^K p(\hat{y}_i = k | y_i = j) p(y_i = j | x_i, w). \end{aligned} \quad (3)$$

		Observed Label	
		-1	1
True Label	-1	$\gamma_{00}$	$\gamma_{01}$
	1	$\gamma_{10}$	$\gamma_{11}$

**Table 1.** Probabilistic relationship between observed label and true label

Here  $p(\hat{y}_i = k | y_i = j)$  represents the probability that the label has flipped from the true label  $j$  to the observed label  $k$ . The probability relationship between observed label and true label is represented in Table. 1. By expanding the terms in Eq. 3 we find that,

$$\begin{aligned} S_i^0 &= p(\hat{y}_i = -1 | x_i, w) \\ &= p(\hat{y}_i = -1 | y_i = -1) p(y_i = -1 | x_i, w) + p(\hat{y}_i = -1 | y_i = 1) p(y_i = 1 | x_i, w) \\ &= \gamma_{00}(1 - \sigma(w^T x)) + \gamma_{10} \sigma(w^T x) \\ S_i^1 &= p(\hat{y}_i = 1 | x_i, w) \\ &= p(\hat{y}_i = 1 | y_i = -1) p(y_i = -1 | x_i, w) + p(\hat{y}_i = 1 | y_i = 1) p(y_i = 1 | x_i, w) \\ &= \gamma_{01}(1 - \sigma(w^T x)) + \gamma_{11} \sigma(w^T x). \end{aligned} \quad (4)$$

By substituting Eq. 4 into Eq. 2, we find:

$$f(w) = \sum_{i=1}^m \frac{\hat{y}_i + 1}{2} \log(S_i^1) + \frac{1 - \hat{y}_i}{2} \log(S_i^0). \quad (5)$$

This formulation remains smooth and differentiable. It can be solved using standard optimization algorithm such as conjugate gradient method [9, 1].

### 3.2 Robust Sparse Logistic Regression

To handle outliers in sparse logistic regression, we need to optimize Eq. 5 subject to the  $l_1$ -ball constrain. We use the Lassplore algorithm to solve the sparse

logistic regression problem [8]. The Lassplore algorithm is a first-order black-box oracle method that evaluates the function value and the gradient at each iteration. The gradient of Eq. 5 is the following:

$$\nabla_w f(w) = \sum_{i=1}^m \frac{\hat{y}_i + 1}{2S_i^1} \nabla_w S_i^1 + \frac{1 - \hat{y}_i}{2S_i^0} \nabla_w S_i^0, \quad (6)$$

where  $\nabla_w S_n^0$  and  $\nabla_w S_n^1$  are:

$$\begin{aligned} \nabla_w S_i^0 &= (\gamma_{10} - \gamma_{00}) \sigma(w^T x_i) (1 - \sigma(w^T x_i)) \cdot x_i, \\ \nabla_w S_i^1 &= (\gamma_{11} - \gamma_{01}) \sigma(w^T x_i) (1 - \sigma(w^T x_i)) \cdot x_i, \end{aligned} \quad (7)$$

$\sigma(w^T x_n)$  is the conditional probability of  $\hat{y}_i$  given a sample  $x_i$ .

An initial estimation of the label flipping probability can be given to start the optimization. After solving the Lassplore algorithm, label flipping probability is updated as the following:

$$\begin{aligned} e_i &= \sigma(w^T x_i), \\ \gamma_{00} &= \frac{(e_i < 0.5) \& (\hat{y}_i = -1)}{e_i < 0.5}, \gamma_{01} = 1 - \gamma_{00}, \\ \gamma_{11} &= \frac{(e_i > 0.5) \& (\hat{y}_i = 1)}{e_i > 0.5}, \gamma_{10} = 1 - \gamma_{11}, \end{aligned} \quad (8)$$

To conclude, an iterative framework is proposed to update the label-flipping probability. First given an initialization of the label-flipping probability, the Lassplore method is used to minimize the loss function. Then, the label-flipping probability is estimated again using the new classification results. Using this scheme we iteratively update the label-flipping probability and solve the optimization problem.

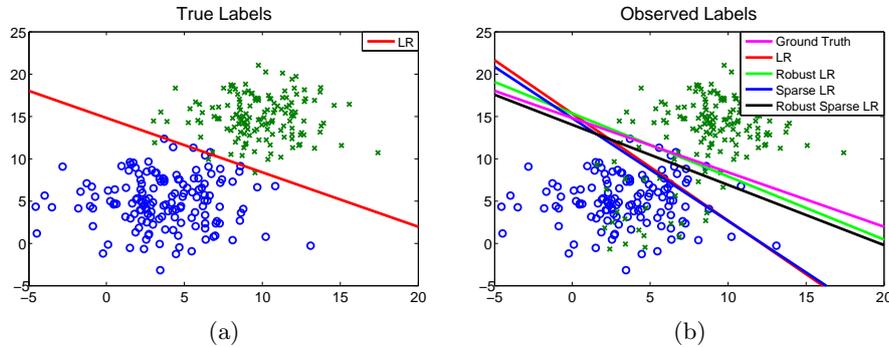
## 4 Experimental Results and Discussion

We first use a synthetic dataset to illustrate the effectiveness of our algorithm. The examples are sampled from two multivariate normal distributions differing in mean and covariance. Here we focus on the non-uniform flipping of labels, where there are negative data being labeled as positive ones. This setting is to imitating the proposed labeling process in medical images. All the data are training on the corrupted data and then tested on the true labels. We quantitatively evaluate the classification results in Table. 2. Compared to logistic regression (LR) and sparse logistic regression (SLR) methods, both robust logistic regression (rLR) and robust sparse logistic regression (rSLR) perform better on sensitivity but worse on specificity. Dice similarity coefficient (DSC), which is defined as  $(2 * \text{true positive}) / (2 * \text{true positive} + \text{false positive} + \text{false negative})$ , measures

Method	Sensitivity	Specificity	DSC
<b>LR</b>	$0.9307 \pm 0.0207$	$0.9987 \pm 0.0027$	$0.9570 \pm 0.0158$
<b>rLR</b>	$0.9920 \pm 0.0040$	$0.9687 \pm 0.0031$	<b><math>0.9809 \pm 0.0013</math></b>
<b>SLR</b>	$0.9153 \pm 0.0248$	$0.9987 \pm 0.0027$	$0.9508 \pm 0.0158$
<b>rSLR</b>	$0.9840 \pm 0.0033$	$0.9807 \pm 0.0020$	<b><math>0.9833 \pm 0.0015</math></b>

**Table 2.** Trained on the corrupted training labels, this table shows testing measurements on true labels. The proposed methods show better performance.

the consistency of results and groundtruth. DSC is a more comprehensive measurement of the coincidence of the classification result with the groundtruth. The proposed methods perform better on this measurement.



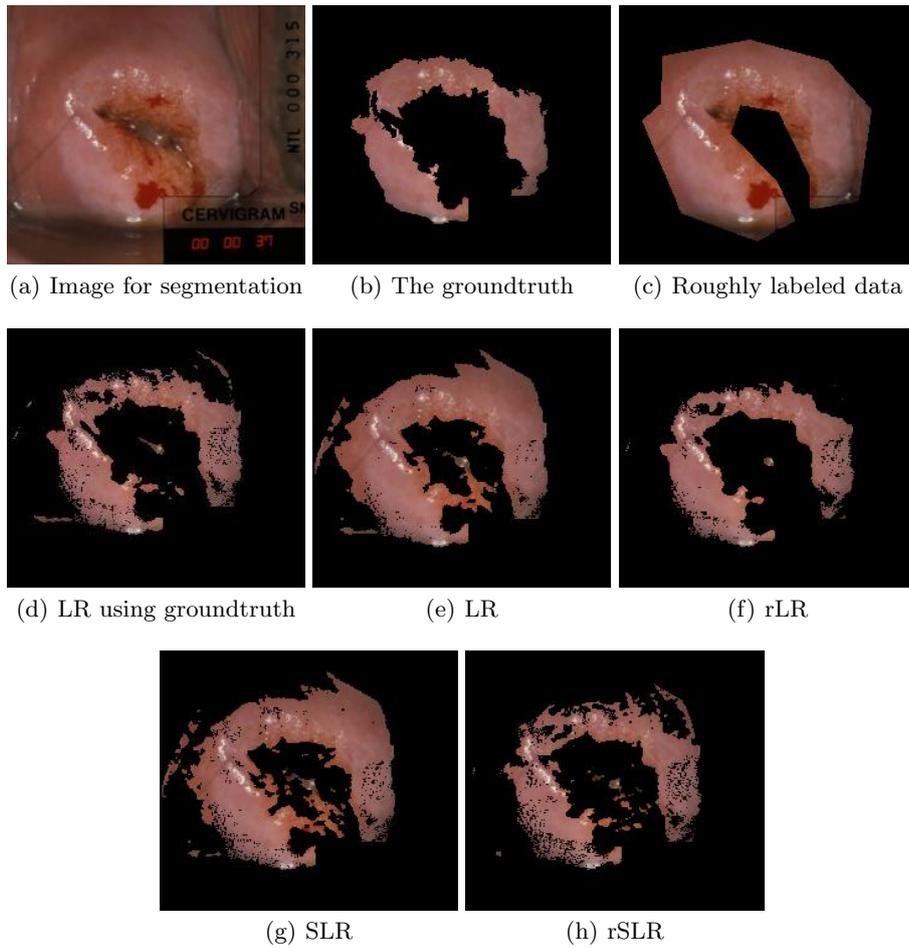
**Fig. 1.** (a) shows the samples with true labels and use Logistic Regression (LR) for separation. (b) shows the samples with observed labels, where about 20% labels from negative class are flipped to the positive class. Robust algorithms push the classification line to the true boundary line on true labels.

We apply the proposed algorithm to classify pixels in optical images into two classes, namely normal and abnormal tissues while not given accurate segmentation training labels. The experiment is designed as the following: We have experts precisely labeled the segmentation, which are considered as groundtruth. We also have experts roughly labeled the segmentation, using our proposed simplified labeling process. The roughly labeled data is used for training, and then we test our segmentation results against the precisely labeled segmentation. Our proposed labeling process and segmentation results would achieve competitive results compared using the the groundtruth for training. In our experiments, HSV feature is employed for classification.

Table 3 shows the results of different classifiers measured by sensitivity, specificity and DSC. Our method consistently achieves significantly better performance in terms of DSC. The results from our method also compares favorably with other state-of-the-art methods in this application [13].

## 5 Conclusion

This paper proposed a novel method on handling labeling noise. The proposed method has been demonstrated on cervigram segmentation problem. Using a bounding box around the region of interest as labels other than accurate labels, we can still get satisfied results. In the future, we want to extend our work using group sparsity [6].



**Fig. 2.** (d) uses the groundtruth and LR in training for comparison. The simplified labeling data, as shown in (c) are used for training for robust algorithm. (f) and (h) include lots of negative points since we use some of the negative points in training. Use the robust algorithms, the result expelled many negative points.

Method	Sensitivity	Specificity	DSC
<b>LR using groundtruth</b>	0.8351	0.8820	0.8157
<b>LR</b>	0.9486	0.5660	0.6950
<b>rLR</b>	0.8175	0.8825	<b>0.8058</b>
<b>SLR</b>	0.9354	0.5931	0.7012
<b>rSLR</b>	0.7121	0.8986	<b>0.7517</b>

**Table 3.** Testing errors on segmenting the cervigram data.

## References

1. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004.
2. K.-S. Cheng, J.-S. Lin, and C.-W. Mao. The application of competitive hopfield neural network to medical image segmentation. *Medical Imaging, IEEE Transactions on*, 15(4):560–567, aug 1996.
3. D. Cobzas, N. Birkbeck, M. Schmidt, M. Jagersand, and A. Murtha. 3D variational brain tumor segmentation using a high dimensional feature set. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, oct. 2007.
4. S. Gordon, G. Zimmerman, R. Long, S. Antani, J. Jeronimo, and H. Greenspan. Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams, 2006.
5. C. hoon Lee, M. Schmidt, A. Murtha, A. Bistriz, J. S, and R. Greiner. Segmenting brain tumor with conditional random fields and support vector machines. In *in Proceedings of Workshop on Computer Vision for Biomedical Image Applications at International Conference on Computer Vision*, 2005.
6. J. Huang, X. Huang, and D. N. Metaxas. Learning with dynamic group sparsity. In *ICCV*, pages 64–71, 2009.
7. K. Koh, S. Jean Kim, S. Boyd, and Y. Lin. An interior-point method for large-scale  $l_1$ -regularized logistic regression. *Journal of Machine Learning Research*, 2007, 2007.
8. J. Liu, J. Chen, and J. Ye. Large-scale sparse logistic regression. In *In ACM SIGKDD International Conference On Knowledge Discovery and Data Mining*, 2009.
9. T. P. Minka. A comparison of numerical optimizers for logistic regression. Technical report, 2003.
10. A. Y. Ng. Feature selection,  $l_1$  vs.  $l_2$  regularization, and rotational invariance. In *In ICML*, 2004.
11. S. Shen, W. Sandham, M. Granat, and A. Sterr. MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *Information Technology in Biomedicine, IEEE Transactions on*, 9(3):459–467, sept. 2005.
12. S. Wang, W. Zhu, and Z.-P. Liang. Shape deformation: Svm regression and application to medical image segmentation. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 209–216 vol.2, 2001.
13. S. Zhang, J. Huang, D. N. Metaxas, W. Wang, and X. Huang. Discriminative sparse representations for cervigram image segmentation. In *ISBI*, pages 133–136, 2010.