LETTER *Special Section on Medical Imaging*

Ensemble Learning Based Segmentation of Metastatic Liver Tumours in Contrast-Enhanced Computed Tomography

Akinobu SHIMIZU[†a)], ***Member***, **Takuya NARIHIRA**[†*], ***Nonmember***, **Hidefumi KOBATAKE**[†**], ***Fellow***, **Daisuke FURUKAWA**[†***], **Shigeru NAWANO**[††], ***and*** **Kenji SHINOZAKI**[†††], ***Nonmembers***

SUMMARY This paper presents an ensemble learning algorithm for liver tumour segmentation from a CT volume in the form of U-Boost and extends the loss functions to improve performance. Five segmentation algorithms trained by the ensemble learning algorithm with different loss functions are compared in terms of error rate and Jaccard Index between the extracted regions and true ones.

key words: *CT image, liver tumour, segmentation, ensemble learning, U-boost*

1. Introduction

Segmentation of metastatic liver tumours from a contrast-enhanced computed tomography (CT) volume is a crucial process for computer-aided surgery and diagnosis. A number of algorithms for liver tumour extraction have been proposed [1]–[5]; some of them were pitted against each other in the competition at the Medical Image Computing and Computer-Assisted Intervention (MICCAI) workshop held in 2008 [5], in which ensemble learning based segmentation (ELBS) trained by AdaBoost [6] presented the best automated segmentation [7]. A segmentation algorithm consists of several hundred weak classifiers (or hypotheses) [4], [7]–[9], called weak segmentation processes in this paper. In a series of training rounds, an AdaBoost algorithm repeatedly calls a weak segmentation process and combines the called processes into one to construct a strong segmentation algorithm with a minimised loss function. The best algorithm in the MICCAI competition combined two ELBS algorithms, one of which was trained using tumours whose mean radius is larger than 24 mm and the other by smaller tumours, by a logical sum operation. Because characteristics of CT value distribution differ according to tumour size. This paper also adopts the same procedure.

Although ELBS has excellent features that can deal with large variations in CT values, some false positives and false negatives occur in the resultant images when compared to manually labelled boundaries set by human observers. One possible reason for such errors is outliers of input feature vectors; another is mislabelling of tumours used for training. AdaBoost is known to be incapable of managing outlier and mislabelling problems appropriately because of its exponential loss function [10]–[12].

Several improved ensemble learning algorithms have been proposed [10]–[13], such as modified AdaBoost (MadaBoost) [10] and Most B-robust η-Boost (MBRη-Boost) [11], known to be robust against outliers of feature vectors and/or mislabelling of tumours. Each algorithm has a different loss function, leading to different characteristics in segmentation. The best loss function depends on the specific application. Although a comparative study of loss functions is important, there is no report regarding liver tumour segmentation. Additionally, the outlier and mislabelling problems might occur more often around tumour boundaries than in other locations, but this is not taken into account in the existing ensemble learning algorithms.

The contributions of this paper are three-fold. First, we present an ensemble learning algorithm in the form of U-Boost [12], which is a class of ensemble learning algorithms including not only AdaBoost but also MadaBoost and MBRη-Boost. The definition in the form of U-Boost makes the differences in algorithms clear, because the differences can be summarised in terms of the differences in the loss functions. Second, extensions of the loss functions of MadaBoost and MBRη-Boost are proposed to improve segmentation performance. Third, we compare the five ELBS algorithms for metastatic liver tumour segmentation trained by the five ensemble learning algorithms, or AdaBoost, Modified AdaBoost (MadaBoost), Most B-robust η-Boost (MBRη-Boost) and extensions of the MadaBoost and MBRη-Boost, to show the difference between AdaBoost and the other algorithms, the effectiveness of the proposed extension and determine the optimal algorithms.

2. Methods

2.1 An Ensemble Learning Algorithm

Consider a segmentation that distinguishes two different classes in a label set $L = \{-1, 1\}$. Here, a voxel in a tu-

Manuscript received June 15, 2012.
Manuscript revised September 6, 2012.

[†]The authors are with Tokyo University of Agriculture and Technology, Koganei-shi, 184–8588 Japan.
[††]The author is with International University of Health and Welfare, Tokyo, 108–8329 Japan.
[†††]The author is with Kyushu Cancer Center, Fukuoka-shi, 811–1395 Japan.
[*]Presently, with Sony Corp., Japan.
[**]Presently, with Institute of National Colleges of Technology, Japan.
[***]Presently, with Canon Research Centre France S.A.S., France.
a) E-mail: simiz@cc.tuat.ac.jp
DOI: 10.1587/transinf.E96.D.864

Copyright © 2013 The Institute of Electronics, Information and Communication Engineers

mour has a value of 1. The segmentation process of a three-dimensional image includes mapping from a three-dimensional position vector $\vec{x}_i$ to the labels $h_t(\vec{x}_i)$: $Z^3 \rightarrow L$. Pseudo-code for U-Boost is provided in Fig. 1.

U-Boost finds an optimum weak segmentation process h_t $(t = 1, \ldots, T)$ repeatedly in a series of training rounds, which minimises the weighted error ε_t computed from the training dataset. Finally, the sequence of weak segmentation processes is integrated into final output H with weight α_t, which intuitively measures the importance assigned to h_t, and α_t increases as the error ε_t of h_t decreases. Here, a weak segmentation process $h_t(\vec{x}_i)$ is defined as follows.

$$\begin{cases} f(\vec{x}_i) \geq T_f & \Rightarrow \quad h_t(\vec{x}_i) = 1 \ \ (tumor) \\ else & \Rightarrow \quad h_t(\vec{x}_i) = -1 \ \ (other\ tissue) \end{cases} \tag{1}$$

The symbol $f(\vec{x}_i)$ is the value of a feature at location $\vec{x}_i$ and T_f is a threshold value. We prepared 80 features, which are divided into three groups (see Table 1). The first group contains CT-value-based features, such as the variance and skewness, measured in a region near the voxel of interest. Features in the second group are based on the output of a convergence index (C.I.) filter [3] used for enhancing tumours. The spatial filter computes a convergence index of gradient vectors in a spherical mask region centred at the voxel of interest, resulting in a high convergence index around the centre of the tumour. The last group consists of other features, such as the output of a Sobel filter and a matched filter. The matched filter is designed in terms of the CT value profile of the tumour whose size is estimated by a C.I. filter [3].

AdaBoost, MadaBoost and MBRη-Boost comprise a class of U-Boost; each has a different loss function $U(z)$, as shown in Table 2, resulting in different properties against outliers and mislabelling of liver tumour labels. Since the loss function of AdaBoost is exponential, it is sensitive to data contaminated by outliers and/or mislabelling. In contrast, MadaBoost minimises the gross error sensitivity [12], [14], and thus minimises the influence of outliers on input feature vectors. MBRη-Boost is robust against not only outliers, but also mislabelling [11]. In general, a class label is determined by an observer with varying degrees of uncertainty, resulting in mislabelling around a tumour boundary. In this boosting, a contamination model was introduced to describe the occurrence of mislabelling, where a conditional probability $P(l\,|\,\vec{x})$ is contaminated as

$$(1-\eta)P(l\,|\,\vec{x}) + \eta P(-l\,|\,\vec{x}) \tag{2}$$

where a parameter η represents the probability of mislabelling.

Given $(\vec{x}_1, l_1)$, $(\vec{x}_2, l_2)$, ..., $(\vec{x}_N, l_N)$, where $\vec{x}_i \in Z^3$, $l_i \in L = \{-1, 1\}$,
N : # of voxels to be processed in a 3D CT volume.
- Initialization $H_0(\vec{x}_i) = 0 \quad (\forall i \in \{1, \cdots, N\})$
- Repeat for $t = 1, ..., T$
  (1) Update the distribution $D_t(i) = U'(-H_{t-1}(\vec{x}_i) l_i) / Z_t$
  where Z_t is a normalizing factor
  (2) Get a weak segmentation by $h_t = \mathrm{argmin}_{h \in H}\, \varepsilon_t(h)$
  where $\varepsilon_t(h) = \Sigma_{i=1,\ldots,N}\, D_t(i)\, I(h(\vec{x}_i) \neq l_i)$
  (3) Set the confidence of the weak segmentation
  $\alpha_t = \mathrm{argmin}_{\alpha}\, \Sigma_{i=1,\ldots,N}\, U(-l_i\, H_{t-1}(\vec{x}_i) - \alpha\, l_i\, h_t(\vec{x}_i))$
  (4) Update $H_t = H_{t-1} + \alpha_t\, h_t$
- Output the final segmentation $H(\vec{x}) = \mathrm{sign}\,[\,\Sigma_{t=1,\ldots,T}\, \alpha_t h_t(\vec{x})\,]$

Fig. 1 Illustrations of ensemble learning algorithm for liver tumour segmentation based on U-Boost.

Table 1 Details of features for weak segmentation processes.

type	features		parameters
CT value based features	original CT value		-
	max CT value	in a spherical region	radius of a spherical region = 3,5,7
	min CT value		
	opening		size of a structural element = 3,5,7
	closing		
	closing => opening		
	opening => closing		
	stdev	of CT values in a cubic region	size of a cubic : 3^3,5^3,7^3
	variance		
	skewness		
	kurtosis		
3D Convergence Index (C.I) filter based features	C.I.		thickness of a filter mask = 3 radius of a filter mask = 5~33 number of directions = 62
	C.I. + diffusion		
	absolute C.I		
	absolute C.I. + diffusion		
other features	sobel		size of a filter mask = 3^3,5^3,7^3
	matched filter		size of a filter mask = estimated leison size by a C.I. filter
	matched filter + diffusion		

2.2 Extension of the Loss Functions

In this study, we modified the loss functions of MBRη-Boost and MadaBoost.

First, η of MBRη-Boost is extended to a function $\eta(\vec{x})$ at location $\vec{x}$. With this we can set the possibility of mislabelling to be high at a tumour boundary, where mislabelling

Table 2 Loss functions of the ensemble learning algorithms compared in this study.

Boosting algorithm	Loss function U(z)
AdaBoost	$\exp(z)$
MadaBoost	$z \quad (z \geq 0)$ $0.5\exp(2z) - 0.5 \quad (z < 0)$
Most B-robust η-Boost	$z \quad (z \geq 0)$ $\{\,(1-\eta)\,\Omega + (2\eta - 1)\log(1+\Omega)\,\}/\eta^2 \quad (z < 0)$ $(\,\Omega = (\exp(z) - 1)\,\eta\,)$

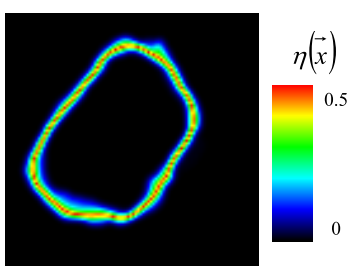

Fig. 2 Example of extended function η for liver tumour segmentation. Voxels with high η, coloured in yellow or red, correspond to the tumour boundary.

is expected to occur. Figure 2 shows an example of the extended function $\eta(\vec{x})$ for liver tumour segmentation, which is computed by the following equation.

$$\eta(\vec{x}) = 0.5 - |0.5 - G_\sigma * I(\vec{x})| \quad (3)$$

G_σ is a Gaussian function with standard deviation σ; $*$ is a convolution operation and $I(\vec{x})$ is a binary image in which tumour voxels have a value of 1 and others have 0. The maximum of $\eta(\vec{x})$ is 0.5 at the tumour boundary, and the minimum is 0 inside or outside of the tumour. In principle, mislabelling might be more likely to occur at the tumour boundary than in other regions. Setting the $\eta(\vec{x})$ of voxels near the boundary to be large will improve segmentation performance. This boosting algorithm is named "*modified MBRη-Boost*" in this paper.

Second, a voxel-wise cost is introduced in the loss function of MadaBoost. A cost-sensitive boosting was presented in [15] and the cost is extended to voxel-wise cost in this paper. We set the cost of the tumour boundary according to Eq. (4) to a small value, such that the undesirable fluctuation of samples in the feature space does not cause the learned ELBS algorithm to suffer from over-learning. The loss function U of MadaBoost is multiplied by the cost $C(\vec{x})$, which derives "*cost-sensitive MadaBoost*".

$$C(\vec{x}) = 1 - 2\eta(\vec{x}) \quad (4)$$

3. Experiments

3.1 Materials

The materials for validation were 40 contrast-enhanced CT volumes of 40 patients, including a total of 133 metastatic tumours. Mean number of tumours per patient scan is 3.3. Volumes were obtained using multi-detector row CT, which provided several hundred 512×512 pixel slice images with 12-bit accuracy. The voxel size of the CT volume was 0.542–0.865 mm in the axial plane and 0.5–5 mm in the z direction. Mode of voxel size in the z direction is 1 mm and the mean is 1.37 mm. Tumours were divided into two groups according to mean radius: tumours larger than 24 mm radius on average and smaller tumours. The reason of the grouping is that gray value characteristics of a small tumour differ from those of a larger one. For example, a small tumour has a uniform gray value while a larger one has higher diversity in gray value. Consequently different ELBS algorithms are required for different size of tumours [7]. The large tumour group included 22 tumours, and the small tumour group consists of 111 tumours. ELBS algorithms were trained and validated for each group. True tumour regions for the training and validation were manually labelled by an author and approved by a radiologist.

3.2 Results

The performance of the five ELBS algorithms trained by AdaBoost, MadaBoost, MBRη-Boost, cost-sensitive MadaBoost (CS-MadaBoost) and modified MBRη-Boost (M-MBRη-Boost) were evaluated by a 10-fold CV test for each group, the large tumour and small tumour groups. The maximum number of weak segmentation processes was set to 200 for the large tumour group, and 1,000 or 1,500 for the small tumour group, σ of the Gaussian for $\eta(x)$ was 1.8 mm, η for MBRη-Boost was an average of $\eta(x)$ used in its extended version, or 0.0331 for large tumours and 0.0512 for small tumours, and minimisation in the computation process of confidence α_t of step 3 in Fig. 1 was performed using the Brent algorithm [16].

Figure 3 (a) shows the transition of error rates for large tumours evaluated by 10-fold CV. The minimum error rates are marked by red circles. AdaBoost is inferior to MadaBoost and MBRη-Boost, whose performances are comparable to each other. The two extensions proposed in this paper yielded different results. M-MBRη-Boost failed to reduce its error due to a local minimum of the loss function, resulting in a high error rate, while CS-MadaBoost was confirmed as the best, slightly better than its original ensemble learning algorithm; it achieved the minimum error among all the algorithms.

Figure 4 presents the regions extracted by AdaBoost, MadaBoost and CS-MadaBoost, where the number of weak segmentation processes was set to an optimal level, according to the red circles labelled a, b and c in Fig. 3 (a). Semi-transparent coloured regions correspond to the extracted re-

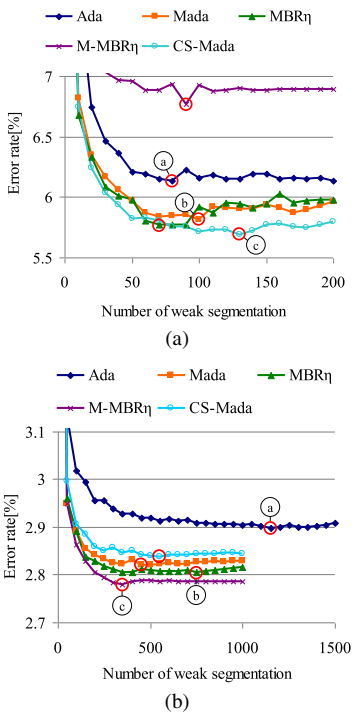

Fig. 3 Error rate transition for (a) the large tumour group and (b) the small tumour group as the number of weak segmentation processes increases. Red circles indicate the number with minimum error rates; the results of three algorithms denoted by a, b and c will be presented in Fig. 4 and 5.

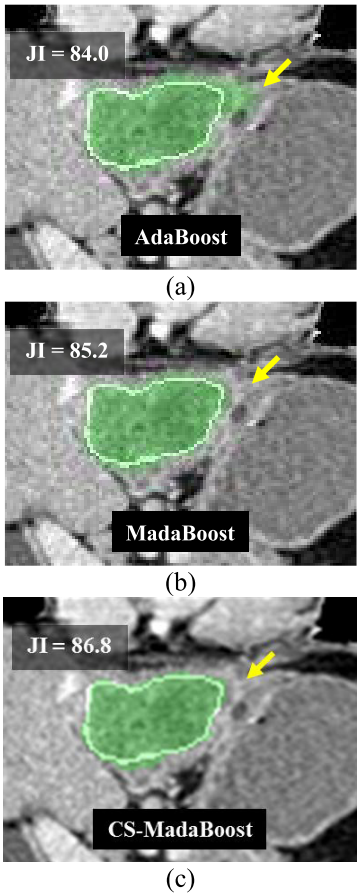

Fig. 4 Results of the segmentation processes trained by (a) AdaBoost, (b) MadaBoost and (c) CS-MadaBoost. White lines are manually labelled true boundaries, and semi-transparent coloured regions are extracted regions. Yellow arrows indicate a false positive in AdaBoost that was eliminated by both MadaBoost and CS-MadaBoost.

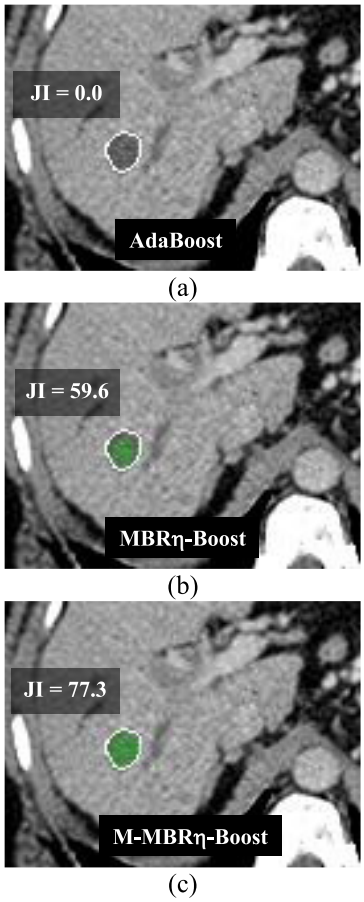

Fig. 5 Segmentation results of the three ELBS algorithms trained by (a) AdaBoost, (b) MBRη-Boost and (c) M-MBRη-Boost. White lines show true boundaries of tumours, and semi-transparent coloured regions are extracted regions. MBRη-Boost and M-MBRη-Boost succeeded in extracting the tumour missed by AdaBoost.

gions, and manually defined true boundaries are shown in white. A difference was observed in the region indicated by yellow arrows, where false positives were eliminated by MadaBoost and CS-MadaBoost.

To evaluate the segmentation performance quantitatively, we computed the Jaccard Index (JI) between an extracted region and a true region, which was manually labelled by an author and approved by a radiologist. The JI of AdaBoost was improved by 1.2 pt with MadaBoost and 2.8 pt with CS-MadaBoost. The average improvement in JI for all materials was 0.092 pt by MadaBoost and 0.51 pt by CS-MadaBoost.

Figure 3 (b) presents the transition of error rates for small tumours evaluated by 10-fold CV. Minimum error rates were achieved at the positions marked by red circles. This figure shows that AdaBoost is the worst, while the extension of MBRη-Boost is better than its original version and achieved the best performance.

Figure 5 demonstrates that MBRη-Boost and M-MBRη-Boost could extract a tumour missed by AdaBoost when the number of weak segmentation processes was set as optimal. The JI of AdaBoost was greatly improved by MBRη-Boost and M-MBRη-Boost. The average improvement in JI for all materials was 2.0 pt by MBRη-Boost and 2.9 pt by M-MBRη-Boost, respectively.

We focus on AdaBoost as well as the best ensemble learning algorithm for each type of tumours and its original version, whose minimum error rates are annotated by a, b and c in the Fig. 3. A Wilcoxon test was carried out to evaluate the difference in performance statistically. The null hypothesis H_0 is that no significant difference exists between the two distributions of error in a CT volume. In the large tumour group, the results told us that the difference between AdaBoost and MadaBoost, and that between AdaBoost and CS-MadaBoost were statistically significant with risk $p < 0.01$. A possible reason for their superiority to AdaBoost is their robustness against outliers and/or mislabelling of tumours. It was also found that extending the loss functions could boost segmentation performance, due mainly to the fact that outliers and/or mislabelling tend to happen more often around tumour boundaries than in other positions.

Similar findings were also observed in the small tumour group. The difference between AdaBoost and

MBRη-Boost, and that between AdaBoost and M-MBRη-Boost whose error was minimum were statistically significant with risk $p < 0.01$ tested by a Wilcoxon test. Moreover, the difference between the JI values of MBRη-Boost and M-MBRη-Boost was statistically significant with risk $p < 0.05$. The result suggests that extending the loss function has the potential to improve the performance, particularly in small tumours.

4. Conclusions

The five different loss functions for ensemble learning based liver tumour segmentation were presented in the form of U-Boost. The performance of segmentation algorithms trained by the different loss functions were compared and optimal one was determined for each tumour group, or small tumour group and large one.

References

[1] L. Soler, H. Delingette, G. Malandain et al., "Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery," Computer Aided Surgery, vol.6, no.3, pp.131–142, 2001.

[2] K. Seo and T. Chung, "Automatic boundary tumor segmentation of a liver," Computational Science and Its Applications — ICCSA, vol.3483, pp.836–842, 2005.

[3] A. Shimizu, T. Kawamura, and H. Kobatake, "Proposal of computer-aided detection system for three dimensional CT images of liver cancer," Proc. Computer Assisted Radiology and Surgery, pp.1157–1162, 2005.

[4] D. Pescial, N. Paragios, and S. Chemouny, "Automatic detection of liver tumors," Proc. ISBI, 2008.

[5] W. Niessen, T. Walsum, M. Schaap et al., "3D segmentation in the clinic: A grand challenge," Proc. MICCAI Workshop, 2008.

[6] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," J. Comput. Syst. Sci., vol.55, pp.119–139, 1997.

[7] A. Shimizu, T. Narihira, D. Furukawa et al., "Ensemble segmentation using AdaBoost with application to liver lesion extraction from a CT volume," Proc. MICCAI Workshop "3D segmentation in the clinic: A grand challenge", 2008.

[8] V. Singh, L. Mukherjee, and M. Chung, "Cortical surface thickness as a classifier: boosting for autism classification," MICCAI; Part I, pp.999–1007, 2008.

[9] J. Morra, Z. Tu, L. Apostolova et al., "Automatic subcortical segmentation using a contextual model," MICCAI; Part I, pp.194–201, 2008.

[10] C. Domingo and O. Watanabe, "MadaBoost: A modification of AdaBoost," Proc. Thirteenth Annual Conference on Computational Learning Theory, pp.180–189, 2000.

[11] T. Kanamori, T. Takenouchi, S. Eguchi et al., "Robust loss functions for boosting," Neural Comput., vol.19, pp.2183–2244, 2007.

[12] N. Murata, T. Takenouchi, T. Kanamori et al., "Information geometry of U-Boost and Bregman divergence," Neural Comput., vol.16, pp.1437–1481, 2004.

[13] J. Aslam, "Improving algorithms for boosting," Proc. 13th Annual Conference on Computational Learning Theory, pp.200–207, 2000.

[14] F. Hampel, P. Rousseeuw, E. Ronchetti et al., Robust Statistics, John Wiley & Sons, 1986.

[15] H. Masnadi-Shirazi and N. Vasconcelos, "Asymmetric boosting," Proc. 24th International Conference on Machine Learning, pp.609–619, 2007.

[16] W.H. Press, B.P. Flannery, and S.A. Teukolsky, Numerical Recipes in C, Cambridge University Press, Cambridge, 1989.