

RUTGERS UNIVERSITY
DEPARTMENT OF STATISTICS AND BIostatISTICS
HILL CENTER #501, BUSCH CAMPUS, PISCATAWAY

www.stat.rutgers.edu

Seminar

Speaker: Kelly Zou, Pfizer, Inc., New York, New York

Title: Validation of Image Segmentation via an EM-Algorithm and Various Metrics

Date: Wednesday April 8, 2009

Time: 3:20 PM

Place: 552 Hill Center

Abstract

The accuracy and precision of segmentations of medical images has been difficult to quantify in the absence of a ground truth or reference standard. Although physical or digital phantoms can help by providing a reference standard, they do not allow the reproduction of the full range of imaging and anatomical characteristics observed in clinical data. An alternative assessment approach is to compare with manual segmentations generated by domain experts. Typically, these segmentations differ due to intra-rater and inter-rater variability.

We present an expectation-maximization algorithm for simultaneous truth and performance level estimation (STAPLE) of the latent ground truth. The algorithm considers a collection of segmentations and computes a probabilistic estimate of the true segmentation and a measure of the performance level represented by segmentation. The source of segmentation in the collection may be an appropriately trained human rater or raters, or may be an automated segmentation algorithm. The probabilistic estimate of the true segmentation is formed by estimating an optimal combination of the segmentations, weighting the segmentation, depending upon the estimated performance level, and incorporating a prior model for the spatial distribution of structures, as well as spatial homogeneity constraints. STAPLE readily enables assessment of the performance of an automated image segmentation algorithm, and enables direct comparison of human rater and algorithm performance.

We examine the automated segmentation accuracy based on three two-sample validation metrics against the estimated composite latent gold standard, which is combined from several experts' manual segmentations via STAPLE. We estimate parametrically the corresponding receiver

operating characteristic curve, Dice similarity coefficient, and mutual information, over all possible decision thresholds. Based on each validation metric, an optimal threshold is computed via maximization. The performances of these validation metrics are investigated via Monte-Carlo simulation. Spatial correlation structures are incorporated using a Markov random field model.

Keywords: ground truth, expectation-maximization algorithm, sensitivity; specificity; receiver operating characteristic curve; dice similarity coefficient; mutual information.

This investigation was based on joint work with Professors Simon K. Warfield, William M. Wells and Ron Kikinis of Harvard Medical School and Massachusetts Institute of Technology. Funding sources included NSF ITR 0426558, a research grant from CIMIT, RG347A2/2 from the NMSS, NIH grants R01LM007861, R01GM074068, R03CA126466, R01EB008015, R01RR021885, P30HD018655, U41RR019703, and P41RR013218.