Online Statistical Inference for Large-Scale Binary Images

Moo K. Chung, Ying Ji Chuang, Houri K. Vorperian
University of Wisconsin-Madison, USA
email: mkchung@wisc.edu

Introduction

An online algorithm is one that processes its inputted data in a sequential manner (Karp, 1992). Instead of processing the entire set of imaging data from the start, an online algorithm processes one image at a time. That way, we can bypass the memory requirement, reduce numerical instability and increase computational efficiency. Since medical image processing is often done semi-automatically, the resulting images may be available at the same time. Further, modern medical imaging datasets are too large to fit into a computer’s memory. Thus, there is a need to develop an iterative analysis framework where the final statistical maps are updated sequentially each time a new image is added to the analysis.

Online algorithm for f-test

Given images $x_{m}$, $x_{n}$ an online algorithm for computing the sample mean $\mu_{m}$ is given by

$$\mu_{m} = \frac{1}{m} \sum_{i=1}^{m} x_{i}.\mu_{m-1} + \frac{m-1}{m} \mu_{m-1}$$

An online algorithm for computing the sample variance map $\sigma_{m}$ is algebraically involved (Chan et al., 1983; Knuth, 1981). It can be shown that $\sigma_{m}$ of interest and $\sigma_{m}$ are updated sequentially each time a new image is added to the analysis.

$$\sigma_{m} = \frac{1}{m} \sum_{i=1}^{m} (x_{i} - \mu_{m})^2 \cdot \frac{m}{m-1}$$

$\sigma_{m}$ is estimated sequentially using online algorithm $\sigma_{m-1} = \frac{m-1}{m} \sigma_{m-1} - \frac{1}{m} (x_{m} - \mu_{m})^2$.

The goodness-of-fit of the full model is measured by the sum of the squared errors (SSE):

$$\text{SSE}_{\text{full}} = \sum (y_{m} - \hat{y}_{m})^2$$

where $\hat{y}_{m}$ is estimated sequentially using online algorithm for regression.

Online statistical inference for large-scale binary images

Online algorithm for regression

Given data vector $y_{m}$ and $x_{m}$ an online algorithm for computing the sample mean $\mu_{m}$ is given by

$$\mu_{m} = \frac{1}{m} \sum_{i=1}^{m} x_{i} \cdot \mu_{m-1} + \frac{m-1}{m} \mu_{m-1}$$

The algorithm starts with the initial value $\sigma_{1}^2 = \frac{1}{n} \sum_{i=1}^{n} x_{i}^2 - \mu_{1}^2$. The algorithm needs to start from $\lambda_{1} = \frac{1}{n} \sum_{i=1}^{n} y_{i}^2$, an online algorithm $\lambda_{m} = \frac{m}{m-1} \lambda_{m-1} - \frac{1}{m} (y_{m} - \mu_{m})^2$.

Under null hypothesis $\beta = 0$, the test statistic at the $m$-th iteration $t_{m}$ is given by

$$t_{m} = \frac{\lambda_{m} - \lambda_{m-1}}{\sqrt{\sigma_{m}^2 / (m - 1)^2}}$$

which is the F statistic with $m - 1$ and $p - r$ degrees of freedom.

Fig. 1. (a) A representative mandibular binary segment that are affine registered to the template. (b) Gaussian kernel smoothing of segmentation with bandwidth $\sigma = 20$. Smoothing can easily patch topological artifacts such as cavities and handles. The sample mean (c) and variance (d) of the smoothed maps computed using the online algorithms.

Application

Subjects. The dataset consisted of 290 typically developing individuals. The age distribution of the subjects is $\text{age} = 19.92 \pm 3.4$ years. A total of 160 male and 130 female subjects $\pm 4.41$, 4.37 and 4.34 were considered significant in the $t$-statistics between age groups I and II, I and III, and III and I respectively at the 0.05 level. The dark red regions show positive growth (bone deposition) and dark blue regions show negative growth (bone resorption). The findings are consistent with previous studies based on 2D surface deformation (Chung et al., 2015) and landmarks (Kelly et al., 2017).

Sex effects. Within each group, we tested the significance of sexual dimorphism by performing the two-sample t-test between males and females (Fig. 2 bottom). Any region above or below $\pm 4.37, 4.89$ and $4.50$ (for group I, II and III respectively) were considered significant at 0.05 level. In group I and II, there is no gender differences. In group III, the statistical significance is localized in the regions between Condyle and Gonion in the both sides. Such findings are consistent with general findings on sexual dimorphism that become evident during puberty.

Fig. 2. Top: t-stat. maps showing mandibel growth. The elation of mandible is shown between Groups II and III, and I and III. The condyle regions show prominent growth in Group III-I comparison. At the same time, the elation is shown as negative growth (dark blue). Bottom: t-stat. maps (male - female) showing sex differences in each age group. There were no significant sex differences in groups I and II. However, pubertal and post-pubertal sex differences are evident in group III that starts at age 13.

References