neuroCombat applied to ADNI

DRC-CMIC Workshop

April 30th, 2022

Table of Contents

- What is Harmonization?
- Introduction to neuroCombat
- Going through neuroCombat applied to ADNI

  https://github.com/Zeena-Shawa/neuroCombat-on-ADNI-Workshop

Harmonization

“the action or process of making something consistent or compatible.”
–Oxford Languages

Why Harmonize?

- **Multisite international neuroimaging datasets** are increasing
- Such datasets have **increased non-biological variability** in the data due to:
  - Scanner hardware: field strength, manufacturer, gradient nonlinearity
  - Image acquisition protocols
  - Other: subject positioning, longitudinal drift
- These properties **increase bias and variance** in neuroimaging analyses
- May result in spurious findings

Harmonization

Statistical Methods

“the action or process of making something consistent or compatible.”
–Oxford Languages

Removing systematic differences from medical images

Harmonization vs. Covariate Adjustment

- Systematic variations that you cannot correct
- Non-biological

- Covariate - “an independent variable that can influence the outcome of a given statistical trial, but which is not of direct interest”
- Biological
- Can increase the statistical power

neuroCombat

Adjusting batch effects in microarray expression data using empirical Bayes methods

W. EVAN JOHNSON, CHENG LI*

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA and Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA
cli@hsph.harvard.edu

ARIEL RABINOVIC

Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, MA, USA

doi:10.1093/biostatistics/kxj037

neuroCombat

Adjusting batch effects in microarray expression data using empirical Bayes methods

W. EVAN JOHNSON, CHENG LI*

Department of Biostatistics and Computational Biology,

d Department of Biostatistics,

ston, MA, USA

ool of Public Health, Boston, MA, USA

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

ELSEVIER

Harmonization of cortical thickness measurements across scanners and sites

Check for updates

Jean-Philippe Fortin[a,1], Nicholas Cullen[b,c,1], Yvette I. Sheline[c,d,e], Warren D. Taylor[f], Irem Aselcioglu[c], Philip A. Cook[c,d], Phil Adams[g], Crystal Cooper[h], Maurizio Fava[i], Patrick J. McGrath[g], Melvin McInnis[j], Mary L. Phillips[k], Madhukar H. Trivedi[h], Myrna M. Weissman[g,l,m], Russell T. Shinohara[a,c,*]

https://doi.org/10.1016/j.neuroimage.2017.11.024

Method (in Brief)

Residuals

$$y_{ijv} = \alpha_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

$$y_{ijv}^{\text{Res}} = y_{ijv} - \mathbf{Z}_{ij}^T \widehat{\theta}_v$$

Method (in Brief)

Residuals

$$y_{ijv} = \alpha_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

residual terms

k x 1 vector of the coefficients associated with **Z** for feature v

average CT for the reference site for feature v

n x 1 vector of CTs for imaging site I, participant j and feature v

$$y_{ijv}^{\text{Res}} = y_{ijv} - \mathbf{Z}_{ij}^T \widehat{\theta}_v$$

k x n matrix of site indicators (deviations from a baseline site)

Method (in Brief)

Residuals

$$y_{ijv} = \alpha_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

$$y_{ijv}^{\text{Res}} = y_{ijv} - \mathbf{Z}_{ij}^T \widehat{\theta}_v$$

Method (in Brief)

Residuals

$$y_{ijv} = \alpha_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

$$y_{ijv}^{\text{Res}} = y_{ijv} - \mathbf{Z}_{ij}^T \widehat{\theta}_v$$

Adjusted Residuals

$$y_{ijv} = \alpha_v + \mathbf{X}_{ij}^T \beta_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

$$y_{ijv}^{\text{Adj}} = y_{ijv} - \mathbf{Z}_{ij}^T \tilde{\theta}_v$$

p x 1 vector of coefficients
associated with X for feature v

p x n matrix of biological
covariates of interest

Method (in Brief)

Residuals

$$y_{ijv} = \alpha_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

$$y_{ijv}^{\text{Res}} = y_{ijv} - \mathbf{Z}_{ij}^T \widehat{\theta}_v$$

Adjusted Residuals

$$y_{ijv} = \alpha_v + \mathbf{X}_{ij}^T \beta_v + \mathbf{Z}_{ij}^T \theta_v + \varepsilon_{ijv}$$

$$y_{ijv}^{\text{Adj}} = y_{ijv} - \mathbf{Z}_{ij}^T \widetilde{\theta}_v$$

ComBat

describes the multiplicative site effect of the j-th site on voxel v

$$y_{ijv} = \alpha_v + \mathbf{X}_{ij}^T \boldsymbol{\beta}_v + \mathbf{Z}_{ij}^T \theta_v + \delta_{iv} \varepsilon_{ijv},$$

$$y_{ijv} = \alpha_v + \mathbf{X}_{ij}^T \boldsymbol{\beta}_v + \gamma_{iv} + \delta_{iv} \varepsilon_{ijv},$$

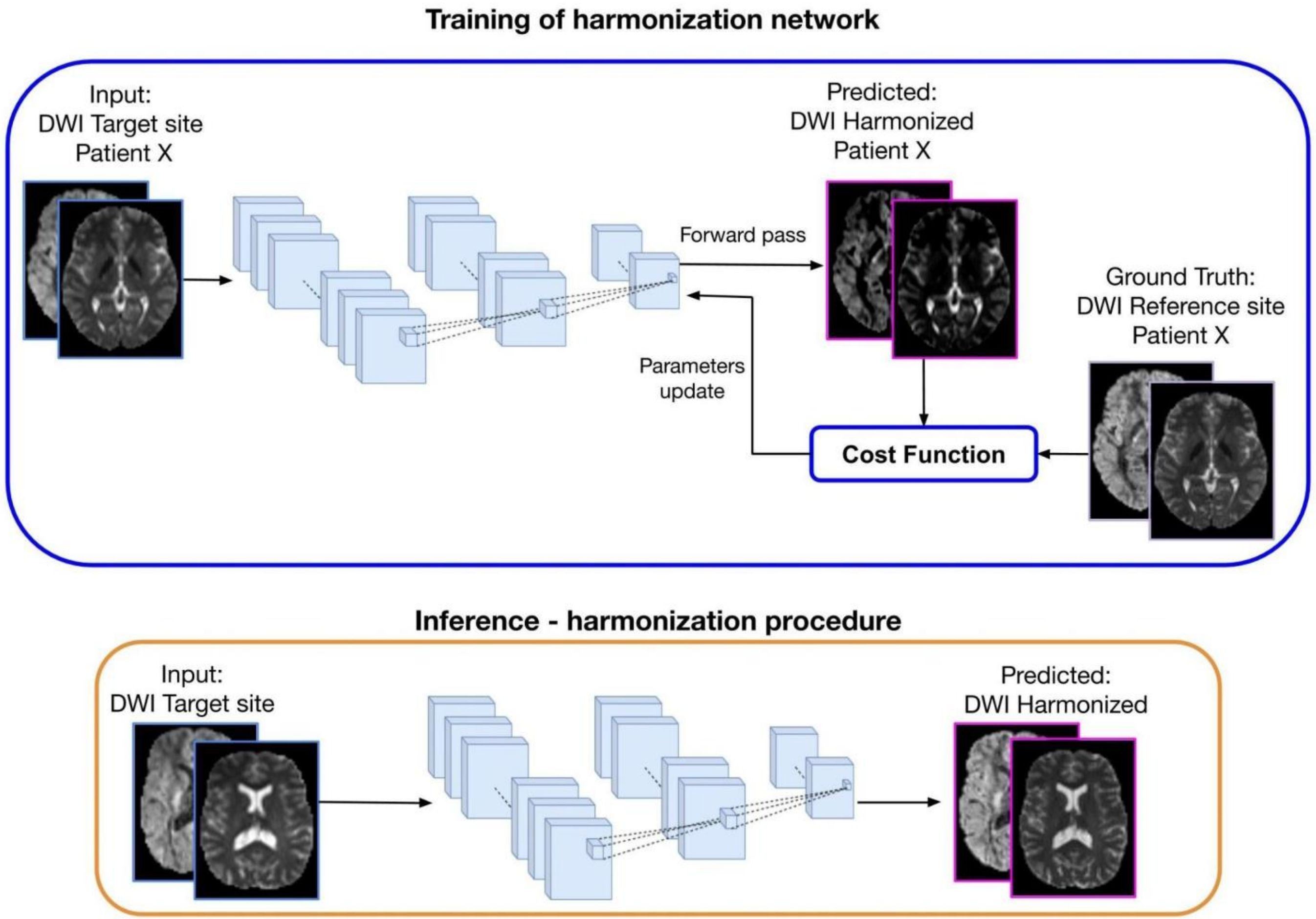
$$y_{ijv}^{\text{ComBat}} = \frac{y_{ijv} - \widehat{\alpha}_v - \mathbf{X}_{ij} \widehat{\boldsymbol{\beta}}_v - \gamma_{iv}^*}{\delta_{iv}^*} + \widehat{\alpha}_v + \mathbf{X}_{ij} \widehat{\boldsymbol{\beta}}_v$$

Paper Concluding Remarks

- **Removes unwanted** sources of **scan variability** and **preserves covariates**
- **Increases** the **power and reproducibility** of further analyses
- Useful for combining imaging data to study brain life-span trajectories
- **Improved stability** in **small samples**
- **Minimal computational overhead** (scales linearly with number of features)
- **versatile** as does not make specific assumptions on type of imaging measures

Competitors

- application-specific: diffusion MRI, DTI, functional MRI, etc.
- RAVEL, Neuroharmony, Deepharmony
- New or more specific versions: longitudinal, GAM, ComBat-RAVEL



[Pinto et al., 2020, https://doi.org/10.3389/fnins.2020.00396](https://doi.org/10.3389/fnins.2020.00396)

Onto the Demonstration!

https://github.com/Zeena-Shawa/neuroCombat-on-ADNI-Workshop

Links to Relevant Resources

- neuroCombat github
- neuroCombat paper
- original ComBat paper

- Dave Cash's HealthBioscienceIDEAS/demon-imaging-harmonisation: A Binder-ready GitHub repo to show the effects of using different scanners and accompanying guide