Data Harmonization*

*Note: Literature Search Task

Divya Sasidharan
M.Sc. Digital Engineering
Otto von Guericke University
Magdeburg, Germany
divya.sasidharan@st.ovgu.de

***Abstract*—This document describe data harmonization in the context of medical MR images and why it is necessary for analysing data from multiple sources.**

***Index Terms*—harmonization, MRI, neuroimaging**

I. Introduction

Medical field have huge requirement for data for their analytics and research works . The data is collected from various neuroimaging research organizations which helps in development of diagnostics studies . The main challenges faced is in combining imaging data from multiple studies and sites. There is lack of standardization in image acquisition protocols, scanner hardware, and software. Inter-scanner variations has been demonstrated to affect reliable combination of different dataset and measurements obtained for downstream analysis. Inclusion of low-quality dataset is another factor effecting the medical research group currently. When combining multi-site data their demographic differences are also important factors to be considered. To overcome such issues we need to develop various data harmonization techniques which will remove multiple site data collection effects discussed above [2].

II. Data Harmonization in medical MRI

The Intra and Inter scanner variations are influenced by various factors like field and gradient strength differences, receiver coils count and sensitivity,even positioning of the participant, software version etc [4] [1]. There are various harmonization techniques used to remove scanner and protocol specific differences on the MRI data. one of method mentioned in paper [1] was to analyse each site data separately and to them perform a meta analysis using z-score for each subjects.Another approach was to use ENIGMA-DTI using site-specific analysis. MRI data harmonization can be acheived with help of various deep learning algorithms like CycleGAN [5] which can learn scanner-invariant features and help in image to image translation or encoder/decoder model [6]. For example in case of CycleGAN , which is a deep generative network with 2 GANs, harmonization is performed as follows. The first one transform image1(image we want to harmonize) to image2 (reference image) and second GAN transform image2 back to image1.we use learned forward mapping from image1 to image 2 to apply harmonization and learned reverse mapping can be used to maintain the semantics of image1. The harmonized data image1 can then be used as inputs for other models [5].

III. Why analysing data from multiple source?

Combining data-sets from different scanners and/or acquired at different time points could dramatically increase the statistical power of clinical studies, and facilitate multi-centre research [3].The main aim of analysing data from multiple sources is to provide better opportunity for healthcare institutions to integrate external data for advanced knowledge gain. For example , research associated with the rare disease require larger dataset which might not be available within a single health organization. Data harmonization comes in as the first step towards collaborative research in healthcare domain .

Without data harmonization researches will need to analyse each data source separately which is not possible when we are thinking of collaborative research . It will be very long and huge effort required process. This results in waste of resources in data wrangling rather than using them for healthcare researches which will be huge benefit for public health.

References

[1] H. Mirzaalian, L. Ning, P. Savadjiev, O. Pasternak, S. Bouix, O. Michailovich, G. Grant, C.E Marx, R.A. Morey, L.A. Flashman, M.S. George, T.W. McAllister, N. Andaluz, L. Shutter, R. Coimbra, R.D. Zafonte, M.J. Coleman, M. Kubicki, C.F. Westin, M.B. Stein, M.E. Shenton, Y. Rathi, Inter-site and inter-scanner diffusion MRI data harmonization, NeuroImage, Volume 135, 2016, Pages 311-323, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2016.04.041.

[2] Raymond Pomponio, Guray Erus, Mohamad Habes, Jimit Doshi, Dhivya Srinivasan, Elizabeth Mamourian, Vishnu Bashyam, Ilya M. Nasrallah, Theodore D. Satterthwaite, Yong Fan, Lenore J. Launer, Colin L. Masters, Paul Maruff, Chuanjun Zhuo, Henry Völzke, Sterling C. Johnson, Jurgen Fripp, Nikolaos Koutsouleris, Daniel H. Wolf, Raquel Gur, Ruben Gur, John Morris, Marilyn S. Albert, Hans J. Grabe, Susan M. Resnick, R. Nick Bryan, David A. Wolk, Russell T. Shinohara, Haochang Shou, Christos Davatzikos, Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. NeuroImage, Volume 208, 2020, 116450, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2019.116450.

[3] Chantal MW. Tax, Francesco Grussu, Enrico Kaden, Lipeng Ning, Umesh Rudrapatna, C. John Evans, Samuel St-Jean, Alexander Leemans, Simon Koppers, Dorit Merhof, Aurobrata Ghosh, Ryutaro Tanno, Daniel C. Alexander, Stefano Zappalà, Cyril Charron, Slawomir Kusmia, David EJ. Linden, Derek K. Jones, Jelle Veraart, Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms, NeuroImage, Volume 195, 2019, Pages 285-299, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2019.01.077.

[4] Mirzaalian H. et al. (2015) Harmonizing Diffusion MRI Data Across Multiple Sites and Scanners. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9349. Springer, Cham. https://doi.org/10.1007/978-3-319-24553-9_2

[5] Bashyam, Vishnu et al. "Medical Image Harmonization Using Deep Learning Based Canonical Mapping: Toward Robust and Generalizable Learning in Imaging." ArXiv abs/2010.05355 (2020): n. pag.

[6] Dewey B.E. et al. (2020) A Disentangled Latent Space for Cross-Site MRI Harmonization. In: Martel A.L. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science, vol 12267. Springer, Cham. https://doi.org/10.1007/978-3-030-59728-3_70

[7] J. Wrobel, M.L. Martin, R. Bakshi, P.A. Calabresi, M. Elliot, D. Roalf, R.C. Gur, R.E. Gur, R.G. Henry, G. Nair, J. Oh, N. Papinutto, D. Pelletier, D.S. Reich, W.D. Rooney, T.D. Satterthwaite, W. Stern, K. Prabhakaran, N.L. Sicotte, R.T. Shinohara, J. Goldsmith, Intensity warping for multisite MRI harmonization, NeuroImage, Volume 223, 2020, 117242, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2020.117242.