Data Mining for Aiding Diagnosis of Attention Deficit Hyperactivity Disorder by a Multilevel Approach

Sarah ITANI SARAH.ITANI@STUDENT.UMONS.AC.BE
Philippe FORTEMPS PHILIPPE.FORTEMPS@UMONS.AC.BE
Fabian LECRON FABIAN.LECRON@UMONS.AC.BE
Department of Computer Science and Management Engineering
Faculty of Engineering, University of Mons, rue de Houdain, 9, 7000, Mons, Belgium

Keywords: ADHD, decision trees, fMRI signals, multilevel approach

Abstract

In the last years, neuroimaging has shown ability to be used in the detection of mental diseases. However, a pathophysiological model of Attention Deficit Hyperactivity Disorder (ADHD) hasn't been established yet. This work aimed to aiding diagnosis of ADHD from the ADHD-200 collection launched in the context of a worldwide competition in 2011. The heterogeneous dataset, regarding on nearly one thousand patients assessed in eight research sites, includes both phenotypical and neuroimaging data. Through this work, we propose to integrate a multilevel approach to our hierachical structure of classification in order to : (1) adress the heterogeneity of the ADHD-200 collection, (2) provide praticians with a convenient and understandable diagnosis tool through decision trees, (3) raise a subset of cerebral regions of interest as biomarkers of the trouble.

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) amounts approximately to five percent of children worldwide. This neuropsychological disorder is defined by three components, the importance of which depends on patients: hyperactivity, impulsiveness and inattention. ADHD, which is likely to persist into adulthood, has impact on children's well-being on the academic, psychological and relational planes. The assessment of ADHD is based on reports from the patients' environment (parents and teachers mainly) but the neurophysiological bases of the disorder have not been fully defined yet. Therefore, as no biological information can be exploited concretely to complete the assessment, the diagnosis of ADHD is unmistakably subjective.

In 2011, a medical dataset of nearly one thousand patients was publicly made available in the context of the ADHD-200 competition (Milham et al., 2012). It includes clinical data and magnetic resonance images (MRI) that are related to the structure and the resting state functional activity of the patients' brain[1]. Scientists from all spheres were challenged to develop an algorithm able to predict diagnosis based on neuroimaging and possibly, on clinical data. With a best prediction rate amounting to 61% on the test set (Eloyan et al., 2012), the competition results are quite encouraging but there is scope for even greater progress.

The ADHD-200 collection gathers data from eight neuroimaging sites located in China, the Netherlands and the United States. This involves various sources of heterogeneities across sites, notably through the measurement conditions and instrument calibrations, the gender representativeness as well as the healthy and pathological cases proportions. Moreover, it has been shown that socio-economic factors influence the local prevalence of ADHD (Akinbami et al., 2011; Sciberras et al., 2011; Bøe et al., 2012; Russell et al., 2014). This complex aspect of the data processing was raised by previous works (Brown et al., 2012; Colby et al., 2012; Olivetti et al., 2012; Sidhu et al., 2012) and the or-

Appearing in *Proceedings of Benelearn 2016*.

[1] See `http://fcon_1000.projects.nitrc.org/indi/adhd200/`

ganizers of the competition ADHD-200 in its outcome (Milham et al., 2012). In addition to this particular feature of the dataset, the model response is structured in a hierarchical way. Indeed, the classification task should distinguish typically developing children from those who are affected by ADHD. In case of positive diagnosis, the ADHD type should also be raised between inattention (ADHD-I) and combined patterns of inattention-hyperactivity-impulsivity (ADHD-C). ADHD hyperactivity-impulsivity type (ADHD-HI) can't be predicted as the associated population is weakly represented in the dataset (1.4%).

2. Related work

2.1. Data and Features Extraction

Among clinical data, age, gender, handedness and IQ were used for training. As far as neuroimaging data is concerned, we used extracted resting-state fMRI signals processed according to *The Athena* pipeline[2] by the Neuro Bureau. The signals measure the time course of each region of interest (ROI). Brains are parcelled into ROIs according to a standard ; Automated Anatomical Labeling (AAL) atlas was considered in this work, which involves a set of 116 time courses. We focused exclusively on functional brain images rather than structural ones since recent studies showed the functional brain activity involvement as significant in neurobiological phenomenons (Purdon et al., 2011; Sidhu et al., 2012).

There are various ways of extracting features from MRI signals. The most common pipelines in neuroimaging seem to consider the computation of a functional correlation matrix (Marrelec et al., 2006; Rubinov & Sporns, 2010; Telesford et al., 2010; Smith et al., 2011; Colby et al., 2012; Dai et al., 2012). It is then possible to deduct some graph metrics (Colby et al., 2012). Signals can also be directly treated by principal component analysis, and the resulting coefficients, used for training (Sidhu et al., 2012). In this work, we considered another pipeline. From fMRI signals, we computed statistics such as **variance**, **kurtosis** and **skewness**. Besides, applying Discrete Fourier Transform (DFT) lead to consider frequency features as the **line of maximal amplitude** and the **centroid of Fourier Spectra**. The latter information is useful in music classification (Tzanetakis & Cook, 2002), measuring its timbral texture and has been also used for speech recognition (Le et al., 2011). In this context, we hoped that using this feature may reveal a distinct signature of ADHD.

A set of 116 features by modality (variance, skewness, kurtosis, frequency, spectral centroid) was computed, which means that biomarkers accounted for 580 features altogether, in addition to the four clinical attributes. A feature extraction was clearly required. This was achieved thanks to a correlation-based feature subset selection (Hall, 1999) to detect attributes that weakly correlates but are highly correlated with the prediction variable.

2.2. A Multilevel Approach

At the first level of the structure of classification, a predictive model is developed by neuroimaging site to separate the groups TD and ADHD, all types confused. The significant parameters for the classification are potentially different for each site. New-York and Peking sites were studied in this work as they are associated to the largest proposed datasets. The second level of the structure of classification allowed to separate the groups ADHD-I and ADHD-C, all sites confused within a whole training set. We so proceed for two reasons. The first one is of technical order: the size of the datasets per site (approximately from 25 to 125 instances) is low so that it is not possible to establish a reliable classifier on a restrictive mass of information. The second reason is that we wished to lead an investigation regarding the efficiency of crossing the information in a global level.

Support vector machine (SVM) is generally used for neuroimaging data learning, providing high accuracies (Fair et al., 2012; Wee et al., 2012; Strigo et al., 2013). However, this *glass model* doesn't allow human validation, and can't be concretely used by praticians. That's why we privileged decision trees that are preferred as diagnosis tools (Tanner et al., 2008).

3. Results and Conclusion

A multilevel approach allowed to raise very different conclusions between levels of classification and neuroimaging sites, in particular as for the explanatory variables implication. The low number of cerebral zones involved in the models suggests that the biological model associated to ADHD could be less complex than expected. On the test sets, the acquired accuracies for New-York and Peking sites accounted for 58% and 66.7% against 35.2% and 51% by ADHD-200 competitors in average[3] as well as 37% and 57% by a similar work of the litterature (Colby et al., 2012).

[2]See `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline`

[3]See `http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html`

References

Akinbami, L. J., Liu, X., Pastor, P. N., & Reuben, C. A. (2011). Attention deficit hyperactivity disorder among children aged 5-17 years in the united states, 1998-2009. nchs data brief. number 70. *Centers for Disease Control and Prevention*.

Bøe, T., Øverland, S., Lundervold, A. J., & Hysing, M. (2012). Socioeconomic status and childrens mental health: results from the bergen child study. *Social psychiatry and psychiatric epidemiology*, *47*, 1557–1566.

Brown, M. R., Sidhu, G. S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P. H., Greenshaw, A. J., & Dursun, S. M. (2012). ADHD-200 global competition: diagnosing ADHD using personal characteristic data can outperform resting state fmri measurements. *Frontiers in systems neuroscience*, *6*, 10–3389.

Colby, J. B., Rudie, J. D., Brown, J. A., Douglas, P. K., Cohen, M. S., & Shehzad, Z. (2012). Insights into multimodal imaging classification of ADHD. *Frontiers in systems neuroscience*.

Dai, D., Wang, J., Hua, J., & He, H. (2012). Classification of ADHD children through multimodal magnetic resonance imaging. *Frontiers in systems neuroscience*, *6*, 1–8.

Eloyan, A., Muschelli, J., Nebel, M. B., Liu, H., Han, F., Zhao, T., Barber, A., Joel, S., Pekar, J. J., Mostofsky, S., et al. (2012). Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging.

Fair, D. A., Nigg, J. T., Iyer, S., Bathula, D., Mills, K. L., Dosenbach, N., Schlaggar, B. L., Mennes, M., Gutman, D., Bangaru, S., et al. (2012). Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. Doctoral dissertation, The University of Waikato.

Le, P. N., Ambikairajah, E., Epps, J., Sethu, V., & Choi, E. H. (2011). Investigation of spectral centroid features for cognitive load classification.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., & Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional mri. *Neuroimage*, *32*, 228–237.

Milham, M. P., Fair, D., Mennes, M., Mostofsky, S. H., et al. (2012). The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, *6*, 62.

Olivetti, E., Greiner, S., & Avesani, P. (2012). ADHD diagnosis from multiple data sources with batch effects. *Frontiers in systems neuroscience*, *6*, 10–3389.

Purdon, S. E., Waldie, B., Woodward, N. D., Wilman, A. H., & Tibbo, P. G. (2011). Procedural learning in first episode schizophrenia investigated with functional magnetic resonance imaging. *Neuropsychology*, *25*, 147.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*, 1059–1069.

Russell, G., Ford, T., Rosenberg, R., & Kelly, S. (2014). The association of attention deficit hyperactivity disorder with socioeconomic disadvantage: alternative explanations and evidence. *Journal of Child Psychology and Psychiatry*, *55*, 436–445.

Sciberras, E., Ukoumunne, O. C., & Efron, D. (2011). Predictors of parent-reported attention-deficit/hyperactivity disorder in children aged 6–7 years: A national longitudinal study. *Journal of abnormal child psychology*, *39*, 1025–1034.

Sidhu, G. S., Asgarian, N., Greiner, R., & Brown, M. R. (2012). Kernel principal component analysis for dimensionality reduction in fMRI-based diagnosis of ADHD. *Frontiers in systems neuroscience*, *6*, 1–16.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., & Woolrich, M. W. (2011). Network modelling methods for fmri. *Neuroimage*, *54*, 875–891.

Strigo, I., Matthews, S., & Simmons, A. (2013). Decreased frontal regulation during pain anticipation in unmedicated subjects with major depressive disorder. *Translational psychiatry*, *3*, e239.

Tanner, L., Schreiber, M., Low, J. G., Ong, A., Tolfvenstam, T., Lai, Y. L., Ng, L. C., Leo, Y. S., Puong, L. T., Vasudevan, S. G., et al. (2008). Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Negl Trop Dis*, *2*, e196.

Telesford, Q. K., Morgan, A. R., Hayasaka, S., Simpson, S. L., Barret, W., Kraft, R. A., Mozolic, J. L., & Laurienti, P. J. (2010). Reproducibility of graph

metrics in fMRI networks. *Frontiers in neuroinformatics*, *4*, 117.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, *10*, 293–302.

Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., Wang, L., & Shen, D. (2012). Identification of mci individuals using structural and functional connectivity networks. *Neuroimage*, *59*, 2045–2056.