The Concept of Conditional Method Agreement Trees with Single Measurements per Subject

Conditional Method Agreement Trees

Siranush Karapetyan[1] and Alexander Hapfelmeier[1,2]

[1]Institute of General Practice and Health Services Research, School of Medicine, Technical University of Munich, Munich, Germany siranush.karapetyan@mri.tum.de

[2]Institute of AI and Informatics in Medicine, School of Medicine, Technical University of Munich, Munich, Germany

Abstract

The concept of conditional method agreement is introduced and respective solutions are proposed to define homogeneous subgroups in terms of mean and variance of the differences in measurements.

1 INTRODUCTION

A well-established method for assessing the agreement between two measurement methods is the Bland-Altman plot [1], which illustrates the differences against the mean values of the measurements. Bland and Altman mainly discuss the question whether the agreement between the methods depends on the magnitude of the measured values and rely on the idea, that agreement is identical for all patient groups. However, the methods may not be equally accurate or comparable in every situation and for every patient. We assume that some factors, such as patient characteristics and measurement settings, may affect the ability of the methods to produce accurate measurements and thus alter the agreement between the methods.

In the present work, we therefore aim to investigate the agreement conditional on patient caharcteristics by introducing the concept of conditional method agreement and propose a corresponding analysis method for the case of single measurements per subject. The new method, called Conditional Method Agreement Tree (COAT), is a simple combination of recursive partitioning and Bland-Altman plot. It takes advantage of the fact that agreement can be modeled by correctly specified tree-based models, which allows the definition of subgroups with different agreement in terms of the mean and variance of the measurements. The ability of the methods to define the subgroups is investigated in a simulation study. Application to real data demonstrates the relevance to research.

2 MODELLING APPROACHES

To define homogeneous subgroups concerning the mean and the variance of the differences, the following tree-based algorithms are used: conditional inference tree (Ctree) [2], distributional tree (Disttree) [3] and model-based recursive partitioning (MOB) [4]. The general algorithm of all these methods is similar, and in order to partition the data based on the potential split variables, they all proceed in two steps, namely, selecting the split variables based on the association of some goodness-of-fit (e.g. model scores) with each potential split variable, and choosing the split point so that the goodness-of-fit is maximized in the resulting subgroups. This procedure is repeated until a certain stopping criterion is reached, e.g., a minimum number of observations in terminal nodes or no more significant associations are detected. Finally, the agreement in each resulting subgroup is evaluated using a Bland-Altman plot. The aforementioned models differ only in the implementation of the individual steps.

2.1 Ctree

The decision to split is based on the test statistic which measures association between the outcome and the split variable, in our case: $T_j = vec(\sum_{i=1}^n Z_{ji} \ (Y_i, (Y_i - \bar{Y})^2)^T)$, where Z_j is a potential split variable and Y_i is the outcome. Such a transfomation of the outcome $(Y_i, (Y_i - \bar{Y})^2)^T$ allows the observations to be split based not only on the association between split variables and mean differences, but also on the basis of the association with the variance of differences.

2.2 Disttree

Another approach is to use Disttree to model the parameters of the prespecified distribution of the differences. In the present case the normal distribution can be used to model the average difference between methods and the corresponding variance. Test statistic for split decision is given by $T_j = vec(\sum_{i=1}^n Z_{ji} \ s(\hat{\theta}, Y_i))$, where $s(\hat{\theta}, Y_i)$ are the model scores and $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ parameter estimates.

2.3 MOB

Like Disttree, MOB is also based on a parametric model and corresponding score. Here we consider the normal distribution and define the residual variance as another model parameter guiding the splitting decision. The latter is based on the empirical fluctuation process given by

$W_j(t) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{n_t} s_{\pi(z_{ji})}$, which measures the deviations of the model scores from their zero mean when the potential split variable is ordered.

3 METHODS

To investigate the performance of the models in terms of identifying the true split variable and the existing subgroups a simulation study is conducted. A total of 10000 simulations are performed, considering sample sizes of n = (50, 100, 200, 400, 500). The differences are generated from a normal distribution $Y \sim N(\mu, \sigma^2)$. The scenarios differ in how the mean and variance are defined. In the first three scenarios, we vary the mean once, the variance once, and finally both based on one split variable. Scenarios 4 and 5 are more complex, as both mean and variance vary based on two split variables. We evaluate the models based on the Adjusted Rand Index (ARI), the measure of similarity between the true and predicted tree structure. We additionally examine a null case to investigate the tendency of methods to detect false subgroups when there are actually none.

To demonstrate the relevance of the new method in epidemiological research, we apply it to real data collected in a validation study [5] and provided by our project partner. In this study, 50 participants were recruited to simultaneously wear different accelerometers for an entire day and night. Daily variables were then created from the accelerometer measurements, including light physical activity (PA), moderate PA, vigorous PA, and activity EE (AEE). Additional information on age, sex, height, and weight was used to test our assumption of whether participant characteristics affect accelerometer agreement.

4 RESULTS

4.1 Simulation study

The results of the simulation study show that with the increasing sample size and in the more complex environment, the models are able to detect the true subgroups. In particular, for scenarios 2-5 and in the case of a higher sample size (400-500 observations), the ARI takes values between 0.65 and 0.95 (values close to 1 indicate the best fit), with the MOB being inferior to the other methods.

In a null case Ctree and Disttree falsely detect subgroups from 4.1-5.4% of cases. On the contrary, MOB yields error rates of only 0.3-3.1% for smaller sample sizes (50-100 observations). However, for larger sample sizes (200, 400, 500 observations), it performs slightly worse, detecting false subgroups between 5.6% and 5.7% of cases.

4.2 Application

Figure 1 is an illustration of the agreement of measurements evaluated with the new method. As can be seen, COAT is able to identify the subgroups with different agreement in terms of mean and variance. It identifies age as a relevant covariate. From the tree structure, it is seen that better agreement is obtained for older patients (age>41).
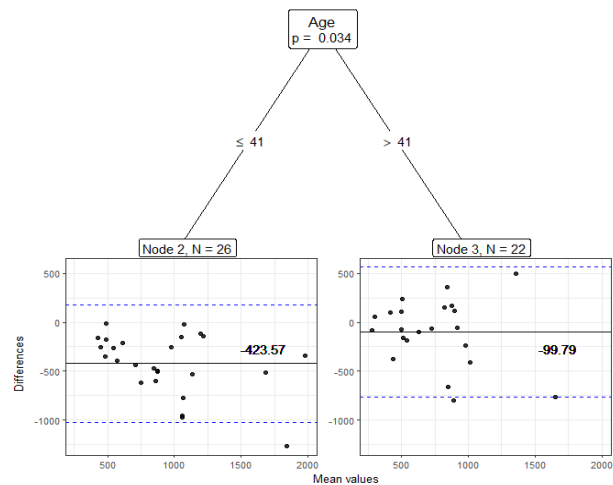

Figure 1. Conditional Method Agreement Tree using Model-Based Recursive Partitioning for AEE measured wth ActiGraph Hip (triaxal) and Actiheart upper.

5 DISCUSSION

With COAT, we are able to estimate conditional agreement and agreement dependence on covariates in a single framework. The simulation study shows that the models are able to identify the true split variable and the underlying tree structure. All three tree-based methods were able to detect heterogeneity between the mean and variance of the agreement. In addition, the real data example shows that COAT is well suited to define subgroups and evaluate conditional agreement in a unified methodological framework. However, it should be noted that the new method is rather exploratory in its nature and evaluation on the real data example is not performed.

6 REFERENCES

[1] D. G. Altman and J. M. Bland. Measurement in medicine: the analysis of method comparison studies. Journal of the Royal Statistical Society: SeriesD (The Statistician), 32(3):307–317, 1983.

[2] T. Hothorn, K. Hornik, and A. Zeileis. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics, 15(3):651–674, 2006.

[3] L. Schlosser, T. Hothorn, R. Stauffer, and A. Zeileis. Distributional regression forests for probabilistic precipitation forecasting in complex terrain.The Annals of Applied Statistics, 13(3):1564–1589, 2019.

[4] A. Zeileis, T. Hothorn, and K. Hornik. Model-based recursive partitioning. Journal of Computational and Graphical Statistics, 17(2):492–514, 2008.

[5] A. Henriksen, S. Grimsgaard, A. Horsch, G. Hartvigsen, L. Hopstock, et al. Validity of the polar m430 activity monitor in freeliving conditions: validation study. JMIR formative research, 3(3):e14438, 2019.

7 ACKNOWLEDGEMENT