Joseph Viviano <joseph@viviano.ca>

Re: Within-Group Heterogeneity

1 message

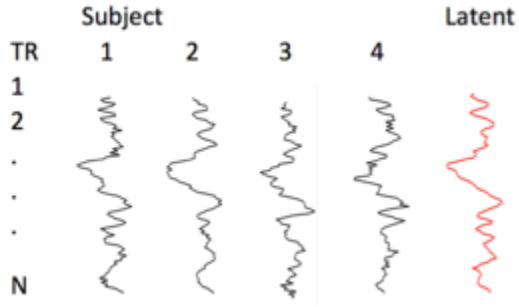
Michael Chow <machow@princeton.edu> Mon, Nov 30, 2015 at 12:32 PM
Reply-To: machow@princeton.edu
To: Joseph Viviano <joseph@viviano.ca>
Cc: Christopher Honey <christopher.honey@gmail.com>, Joseph Viviano <joseph.d.viviano@gmail.com>

Hey again Joseph!
Sorry for not getting back earlier--I spent some time thinking about it, but didn't get a chance to work out a decent response before Thanksgiving.

Out of the gate, I should mention that my focus with ISC is not classification (e.g. are individuals in these groups different?), but producing models that explain the differences between groups (e.g. how are individuals in these groups different?). To the extent that you just want to produce an accurate and generalizable prediction, without needing to interpret the parameters of whatever models are generating predictions (a reasonable goal), then the rest of this email might not be as useful.

However, since you mentioned a template to compare the time-series of each member of a group to, and within vs between group analyses, it sounds like the models I've been developing may be useful.

I would say that current ISC methods fit well with a latent variable model where each participant within a group is represented as a noisey measure of the latent variable for their group. For one group, you could depict it as..



So, if you have 2 groups, then the timecourse for each participant is a noisey measure of their groups latent timecourse. From this perspective, there are different questions you could ask about the groups.

1. **Internal consistency:** Are participants within a group more noisey measurements of their latent time course? I'll use the term internal consistency to refer to participants within a group as having stronger signals (and less noise).
2. **Group similarity:** Are the latent time courses of the two groups the same? (or better, how correlated are they?)

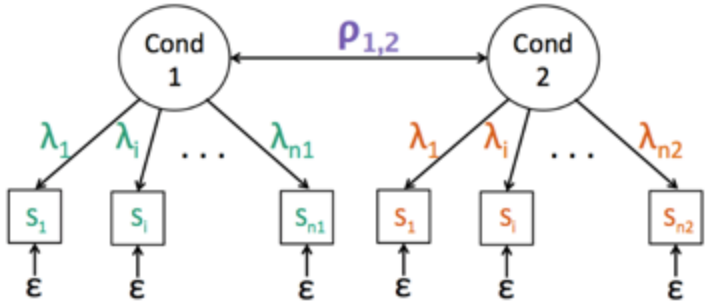
So when you mention a gold standard to compare new participants to, I would say both of these questions could be used to classify them.

Same latent timecourse across groups

For example, suppose that the two latent timecourses are identical. This means that, realistically speaking, there is one underlying timecourse. You still may be able to classify participants as healthy controls or part of your clinical population based on how noisey they are off of that timecourse. A good example of this is the argument made in Uri's autism paper, that autistic individuals have strong person-specific component, but also a weak group-specific component, and the group-specific component is similar for autistic and control subjects.

Different latent timecourse across groups
Suppose that the two timecourses are different. In this case, the more dissimilar the groups are, the lower the correlations across groups should be. However, the correlations across group will also be a function of how internally consistent the groups are. If you just want to classify subjects into groups, this may not matter. However, if you want to say whether differences between groups are due to differing levels of internal consistency, group similarity, or both, then you need to model them correctly.

Modeling internal consistency and group similarity
This problem is highly related to work that has been done in individual differences research in psychology for the past century. As it turns out, you can model it using structural equation models with latent variables (or their equivalent bayes nets):



(path diagram, if you're familiar with SEM; lambdas are the strength of the latent signal in participants, rho is the correlation between signals. Subjects are the boxes.)

I have some R code that measures internal consistency / group similarity using these models, but haven't put much documentation up. If it seems relevant to your question, though, it's something that I need to do, so I'm happy to work on implementing an example that might fit your use-case. I've used it to do voxelwise analyses, but it takes about 10 seconds per voxel, so some parallelization is needed.

If you have an ROI with the timecourses for the two groups, I'm happy to write an example script for implementing the model. Alternatively, I put up this simple shiny app that allows people to upload a .mat (and maybe a csv file?), and then spits out the fitted model. If you want to try that out, I can send an example .mat file / tweak the app to report whatever is useful.

I'm trying to wrap up some bayesian forms of the models, which might be particularly useful for your problem, since it seems like you should be able to take the posterior, pass a new participant and classify which latent they came from and their noise levels. I think it would take a fair amount of thought / work to implement, though. They take longer to fit, but I'm going to take a pass at implementing an approximation that should approximate the posteriors much faster. Probably not a good solution for now.

I've attached a poster from SfN on the issue, and am happy to explain any parts of it.

Cheers,
Michael

On Mon, Nov 16, 2015 at 3:31 PM, Joseph Viviano <joseph@viviano.ca> wrote:

> Thanks Chris,
>
> Hi again Michael, it's been some time.
>
> We're doing some ISC analysis in a healthy control / disease setting. We're using classification (random forest) to evaluate the relative performance of the HRF/GLM approach vs the ISC approach, and we're getting a massive boost in performance using ISC.
>
> Right now, our ISC values are derived within-group. Part of me feels like that is cheating a little. Do you have any experience with group-based analysis, comparing all subjects in both groups to a 'template'? In our case, it would be some group of healthy controls we expect to be able to perform our task very well. So our ISC metric in this case would be the similarity of the time series of the members of each group to some gold-standard template group. This is attractive from a biomarker-development standpoint because this template should generalize and would allow for one-off testing.
>
> Any thoughts or caveats would be really appreciated. I'm also happy to share more about what we've done and the code we've written to do it if any of this interests you.
>
> Best,
>
> On Thu, Nov 12, 2015 at 4:20 PM, Christopher Honey <christopher.honey@gmail.com> wrote:
>
>> Dear Joe,
>>
>> I am putting you in touch with Michael (cc'd) who is an expert on how to handle issues of group heterogeneity when comparing response timecourse reliability within and across groups.
>>
>> Converse away!
>>
>> Cheers from Toronto,
>> CH
>
> --
> Joseph

2015_SFN.pdf
6187K