

Morning Meeting

Showed the differences between raw, normalized (subtract min, divide by range) and standardized (subtract mean, divide by standard deviation).

- Levannia thought the standardized looked much cleaner but again brought up her concern that by setting the mean to 0 we might lose important information.
- we told her that for some of the waveforms the standard deviation was 0 so we had to divide by a very small number instead. She said she would be interested in seeing those waveforms and that something else may have gone wrong with them.

We looked at how the calibration period looks the same for pin contacts and unsuccessful and mentioned how this is making it hard to cluster them apart.

- Levannia mentioned that in pin contact they usually look for fluctuations around 0, along with CMeanQCHigh.
 - o We looked at an example of an unsuccessful reading with that return code which looked similar to the pin contact examples.
 - Main difference: values for pin contact were positive, which they never should be
 - o Another thing to look for is the difference between windows. Pin contacts usually fluctuate around 0 during calibration and have little to no response in the post/sample window.
 - o Diagnosing pin contacts are usually just with built-up knowledge over years. There's not really a consistent methodology or process.
- Our analyte is also famously noisy and the readings are usually not very accurate industry wide.
 - o We need to find a way to reduce noise. Even if we end up not clustering successfully, developing a denoising pipeline will be helpful for them.
 - o They will give us some example waveforms from a less noisy analyte that we can use as a reference while trying to filter out noise.
- After de-noising it might be useful to look at changes between windows. The additional window on the calibration side may also be useful.

Saisree asked how confident they are in the labels

- They are very confident that pin contact errors are really pin contact (they were double checked)
- Unsuccessful readings may be pin contacts as well but unlabeled.
 - o The unsuccessful readings are only investigated if a certain threshold of cards in a batch yield unsuccessful readings.

We looked at the PCA 3D representation of the waveforms

- Showed how there's no clear separation between pin contact and unsuccessful readings.

Levannia mentioned that in readings with 'Cannot Calculate' and sample detect time 0, something was probably wrong with the calibration fluid, but other cannotcalculates may be because the reading

depends on the sensors for other analytes which failed – the waveform might be fine, just can't be confident in the calculations using the other sensors information.

Is it worth just trying to understand the differences between different types of waveforms and return codes?

- Already know what flags a return code – but that's just the first error to occur. Other things might happen in the waveform as well.
- Ultimately want to understand the ways in which sensors fail and if we can identify them. Even if pin contact is not a category we can pull out, what are the categories?
 - o Eg. high noise in x window + high noise in y window might be a certain failure type.
- If we can't pull out pin contacts, maybe we take a step back and think about what are the characteristics of the waveforms themselves?
 - o Might lead us to transformations/ways of looking at waveforms that could help us pull out errors.

Another transformation we can potentially look at is computing the 1st/2nd derivatives of the waveform.

- They will send us the formulas that they have used for this in the past.
 - o Haven't been used as a return code for our analyte in the past. Will have to find our own implementation that works best for us.
- They will also give us their normalization formula.

Finding a way to consider the whole waveform will probably be useful as the operators use the entire thing when looking for pin contact errors.

They recommended we talk to Dr. Braun about noise reduction.

Bubbles and pin contacts are the biggest sources of error but sometime readings are just weird and they don't know why – they can't really give us a number of clusters to expect. Bubbles in themselves are quite variable.

Run numbers for synthetic – 1 and 2 = glue, 3-7 scraped, may 17-may 18 – leftover glue on leads.