

GPU ACCELERATION FOR T-SNE MAPS

JACKSON WALTERS

Using Google Colab, I obtained access to a NVIDIA Tesla K80 with 12GB of VRAM for use on a trial basis. With this hardware, I was able to train models using cuML¹, a GPU accelerated library including t-SNE. The data includes at least seven columns which are ‘diagnostic’, which are:

- TRAUSTREFLG=Trauma- and stressor-related disorders
- ANXIETYFLG=Anxiety disorders
- ADHDFLG=Attention deficit/hyperactivity disorder (ADHD)
- BIPOLARFLG=Bipolar disorders
- DEPRESSFLG=Depressive disorders
- SCHIZOFLG=Schizophrenia or other psychotic disorders
- PERSONFLG=Personality disorders

These are binary, 0/1 indicators of a diagnosed disorder. Since there are seven categories, this yields 128 discrete categories. This is too many for visualization purposes, so we cut them down by choosing one for ‘no-disorder’ (all zeros), one for ‘single disorder’ (one 1 and the rest zeros), and ‘multi-disorder’ (more than one 1). This yields nine categories. The t-SNE with perplexity around 10 reveals clear clusters:

¹<https://github.com/rapidsai/cuml>

FIGURE 1. num-data-points=10,000. Mental health diagnostic data mapped to 2d using t-SNE. Note there is no symptom or life-factor data in this dataset yet. There is one 'no-disorder' category, seven unique disorder categories, and one multi-disorder category. We observe eight clusters, with the multi-disorder categories appearing scattered.

