APPLYING NEURO-INSPIRED MODELS TO OLFACTORY DATA

Nicole Trappe, Aminat Adebiyi, Alberto Mannari, Andrea Fasoli, Luisa Bozano

IBM

PROJECT GOAL

The project goal is to build an **electronic nose**. This consists of two parts. The first is our "C-Eva" sensor prototype (Figure 1) that detects the concentration levels of multiple odorants. The second is a machine learning model inspired by analyzing how the olfaction process in animals functions. This project is unique in that we are looking to the brain to devise a new algorithm that is both best suited to olfactory data but also challenges traditional neural networks.

SUMMARY

For the hardware, environmental and analyte conditions have contributed to large-scale drift in the C-Eva. As we need a stable flow for accurate concentration readings, we have identified temperature and humidity as two of the main culprits and built a program to preemptively adjust the flow. For the software, we first tested multiple-layer CNNs and found that olfactory data best fits shallower networks. We are now building a new model based upon the olfactory circuit of *Drosophila*—*"fruit flies"*. This model will challenege our traditional vision-inspired (dense, convergent) neural networks and better suit the data.

HARDWARE

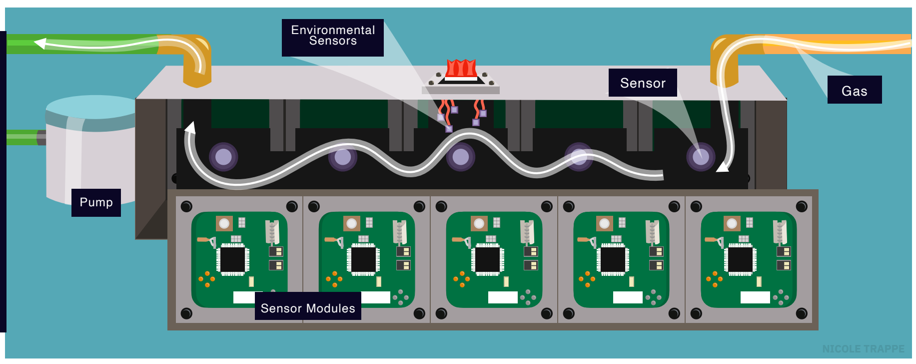

Figure 1: C-Eva*: The gas first enters through the yellow pipe and passes through the open space across the sensors of the ten sensor modules and environmental sensors before exiting via the green pipe. A pump on the exhaust controls the flow rate.*

FLOW CONTROL

A uniform flow rate through the C-Eva (**Figure 1**) is necessary for consistent, replicable concentration readings. However, temperature and humidity both affect the actual flow of a gas through the C-Eva as shown in **Figure 2.** To compensate for the difference between the *actual* and *desired* flow, we built a program to set the pump speed appropriately.

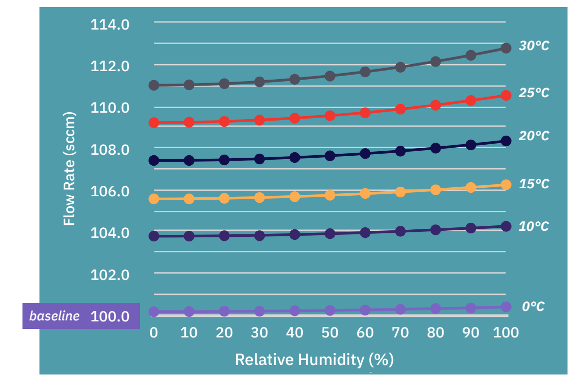

Figure 2: *Humidity and Temperature vs Flow Rate*

SOFTWARE

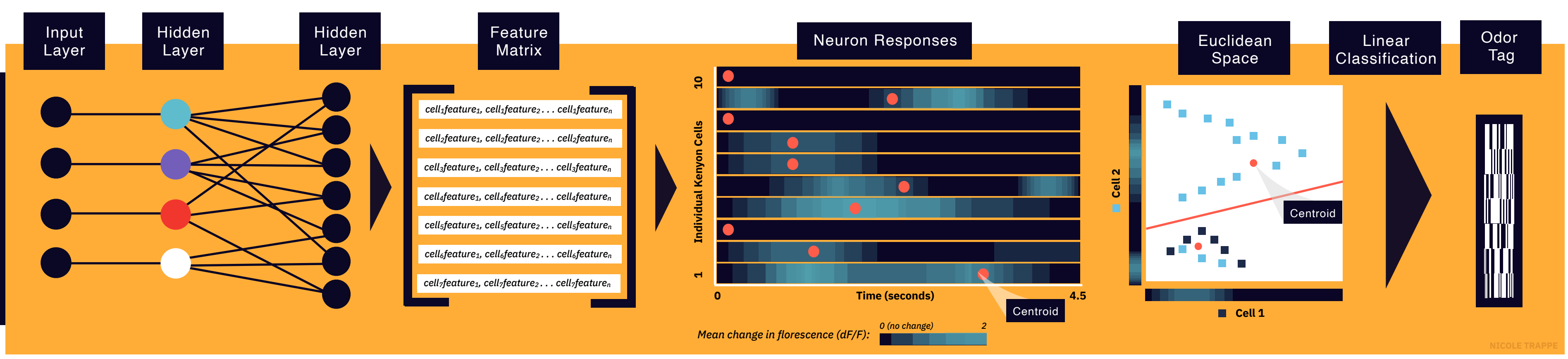

Figure 3: *The data from our setup will be entered as the input layer, randomly weighted, and sent to a hidden layer. From there, a traditional neural net would converge however, after the first hidden layer, our outputs diverge and are sent to another hidden layer. Let's think about the inputs to this next hidden layer not as concentration levels but as pixels which have different levels of florescence; the higher the concentration, the brighter the pixel is so the higher the change in florescence. Each cell has pixels that may grow brightly for a time and then quiet down again after it has been 'activated' by a substance and the peak is the concentration level. From there, we can map the neuron responses—or centroid—of each cell's florescence delta. Then, we can map the data points with each cell's centroid into Euclidean space where each axis is a cell; with 50 Kenyon cells, we would have 50-D space. Finally, linear classification will split the data points into groups based on their nearest centroid and output a matrix like the olfactory circuit hash that acts as the odor tag.*

THE MODEL

Good prediction of (X) — Predictions are nearly convergent — Unusable, predictions are unreasonable

Stats: *4 classes of beers (**y**), 40 features (**X**), 50 rows of training data, 1000 iterations, results mediand from 10 trials*

Neurons in Hidden Layer	Training Loss	Prediction Accuracy
10	0.015	Good
20	0.022	Good
5	0.037	Good
15	0.065	Good
30	0.196	Nearly convergent

Figure 4: *One-layer Neural Net (40 features)*

Stats: *4 classes of beers (**y**), 40 features (**X**), 50 rows of training data, 100 iterations, results mediand from 10 trials*

Neurons in Hidden Layer 1	Neurons in Hidden Layer 2	Training Loss	Prediction Accuracy
30	10	0.028	Nearly convergent
20	5	0.030	Good
30	5	0.031	Good
20	10	0.031	Good
15	5	0.032	Good
10	3	0.035	Nearly convergent
30	15	0.197	Unusable

Figure 5: *Two-layer Neural Net (40 features)*

Stats: *4 classes of beers (**y**), 100 features (**X**), 50 rows of training data, 100 iterations, results mediand from 10 trials*

Neurons in Hidden Layer 1	Neurons in Hidden Layer 2	Neurons in Hidden Layer 3	Training Loss	Prediction Accuracy
20	12	3	0.061	Nearly convergent
30	10	5	0.068	Nearly convergent
30	15	5	0.071	Good
30	5	2	0.080	Unusable
30	8	4	0.081	Nearly convergent
25	10	4	0.083	Unusable
30	25	10	0.146	Nearly convergent

Figure 6: *Three-layer Neural Net (40 features)*

We tested both a simple one-layer (**Fig 4**), two-layer (**Fig 5**), and three-layer (**Fig 6**) neural net on the C-Eva data. Increasing the number of hidden layers increased the training data loss—using least squares—indicating that it became worse at classifying the data. Additionally, the two- and three-layer neural nets, with many features, had worse predictions for the testing data. Many produced convergent predictions or maximas. We need to continue testing with varying parameters but we hypothesize that olfactory data is best characterized by shallow nets.

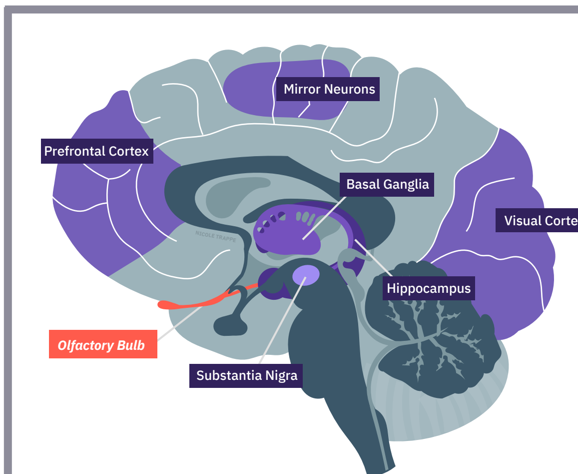

Figure 8: *Human brain illustrating the areas that have previously inspired machine learning in purple and the area we are focusing our research on in red.*

OLFACTION

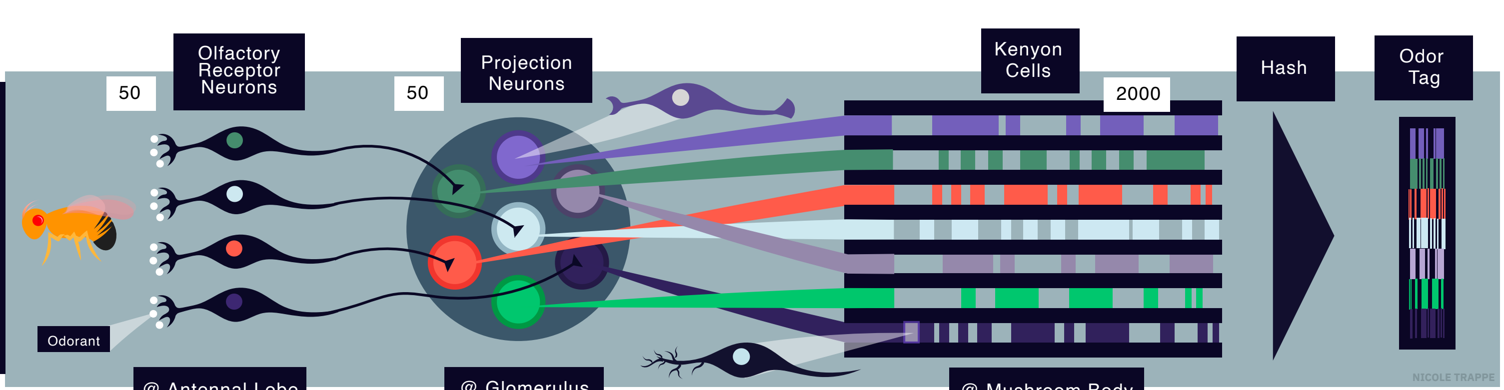

Figure 7: *Each of the 50 olfactory receptor neurons (ORNs) only activates for a specific subset of odorants. When they sense their preferred odorant, they fire and send the signal to a Projection Neuron (PN) of their same class. From there, each PN sends signals to each Kenyon cell, activating many of them. The unique pattern of activated Kenyon cells produces a hash that is the tag for the odor.*

THE BRAIN

Neurosciene has both inspired and validated machine learning (ML) models (**Figure 8**). One of most ubiquitous models, convolutional neural nets (CNNs) were based upon processing in the Visual Cortex (V1). Other popular methods include Q-networks (DQN) which are based on experience replay in the hippocampus; DNC and LSTM mimic pre-frontal reading and writing to an external memory matrix; reinforcement learning uses prediction errors in the Substantia Nigra; and even AlphaGo utilized the Basal Ganglia.