Linear model for odor predictions in DREAM olfaction prediction challenge

Authors: Guillermo Cecchi, Pablo Meyer   
Affiliations; IBM research

Summary Sentence

We developed a linear model with a second layer cubic correction based on a reduced version of the Dragon molecular descriptors to predict the average perceptual descriptors for all individuals.

Background/Introduction

* The motivation for the approach was based on previous (but limited) attempts to model perceptual responses of humans.
* The underlying methodology used is a multi-linear regression for each of the descriptors based on the responses of all individuals and the molecular descriptors of each odor.
* The only pre-processing of the data we did was dimensionality reduction of the number of features of the Dragon molecular descriptors, and a log transformation of the values.

Methods

Training

(1) Eliminate 'non responses'

(2) Keep only 'high' concentrations

(3) Identify 1/1,000 concentrations to test Intensity

(4) Identify odors with replicates and without 1/1,000 trials

(5) Transform molecular values using logarithm

(6) Reduce with PCA the dimension of molecular descriptors to 40

(7) Build perceptual matrix, averaging over subjects, the underlying model is that:

K=M\*W

where K is the perceptual matrix for the 338 odors, M is the molecular descriptor matrix for the odors and W is the linear model

(8) Learn a linear model and a 2nd layer correction using a cubic for each perceptual value

Testing

(1) Build perceptual matrix

(2) Mean of features for each odor

(3) Predict K using W learned with training set and M for the 69 new odors.

(4) Add 2nd layer of processing

Train/Test Partition

(1) Based on the above, we chose a random partition that yields good average predictive accuracy

Conclusion/Discussion

We chose the partitions for the leaderboard and test set based on the distribution of median correlation over test odors obtained with the model, for different random partitions. The median correlation across odors selected for the selected partition is above 0.33.

References

Kahn et al The Journal of Neuroscience, September 12, 2007 • 27(37):10015–10023

Mandairon N, Poncelet J, Bensafi M, Didier A (2009) Humans and Mice Express Similar Olfactory Preferences. PLoS ONE 4(1):

Snitz K, Yablonka A, Weiss T, Frumin I, Khan RM, et al. (2013) Predicting Odor Perceptual Similarity from Odor Structure. PLoS Comput Biol 9(9):

Authors Statement

Guillermo Cecchi designed and coded the model.