Predicting human olfactory perception from chemical features of odor molecules

Xiufeng Zhao(x6zhao@ucsd.edu)

Haocheng Zhu(haz097@ucsd.edu)

Junliang Yu(juy001@ucsd.edu)

## biological problem

As we know, olfaction is very important for us in life. For example, the smell of perfume will make people pleasant and happy, and smell can also tell us the food is spoiled or not. To be more specific, it is the chemical structure of molecules affects our olfaction. For this project, we want to study the relationship between the chemical structure and the Olfaction and compare different machine learning model to predict the smell of a molecule based on its chemical structure.

## Dataset

Dataset for this project is provided in the DREAM Olfaction Prediction Challenge[3], which consists of two parts: perceptual data and molecular descriptor data for 476 different odor molecules. As for the perceptual data, 21 kinds of perceptual description were provided by each of 49 subjects for each odor molecules. As for the molecular descriptor data, Dragon[4] is used to generate 4870 molecular descriptors for each of the 476 molecules.

Among all these 476 molecules, 388 of them are used as training set, 69 of them as validation set and 69 of them as test set.

## Methods to be compared

In this project, we will use Linear Regression, Random Forest[1], SVR with polynomial kernels and AdaBoost[2] to model the relationship between molecule features and perceptual data and predict perceptual data of a new molecule given its features.

## Reference

[1] Keller et al., Science 355, 820–826 (2017)

[2] Supervised learning, <https://scikit-learn.org/stable/supervised_learning.html>

[3] DREAM Olfaction Prediction Challenge, <https://www.synapse.org/#!Synapse:syn2811262/wiki/78368>

[4] Dragon, <http://www.talete.mi.it/help/dragon_help/index.html>