Prediction of organoleptic properties of chemical compounds

Literature analysis

VirtualTaste: a web server for the prediction of organoleptic properties of chemical compounds

- Basic information:
  - Authors: Franziska Fritz, Robert Preissner, Priyanka Banerjee
  - Published: 27 April 2021
  - Nucleic Acids Research, Volume 49, Issue W1, 2 July 2021, Pages W679–W684
  - Published by Oxford University Press

https://academic.oup.com/nar/article/49/W1/W679/6255693

What problem does the research solve?

Predict three different taste endpoints of chemical compounds - sweet, bitter and sour.

Motivation:

- Taste is an important organoleptic property and is one of the crucial senses involved in the perception of food by humans.
- Taste is simulated when fundamental nutrients or harmful compounds, such as toxic molecules, activate specialized receptors located in taste buds.
- Taste of a chemical compound present in food stimulates us to take in nutrients and avoid poisons.
- The taste prediction of a compound is of considerable interest in the food industry.

Main difficulties

- The chemical structure of sweet tasting compounds is known to be incredibly diverse. There is a long list of sweet tasting compounds and there are at least a few different structural classes.
- Bitter agonists include plant-derived and synthetic compounds.
- The models are only as good as the data they are based on – lack of large amount of clean data.
- Choose the best suited model, tune hyperparameters, choose good evaluation metrics, etc.

Other methods, researches and tools for solving the problem

- Traditionally, discovering the taste of compounds is done using the human taste panel or cell-based high throughput screening.
- Tools for prediction/classification of bitter and sweet chemical compounds:
  - Bitter or not? BitterPredict, a tool for predicting taste from chemical structure - a machine learning classifier, BitterPredict, which predicts whether a compound is bitter or not, based on its chemical structure.
  - e-Sweet: A Machine-Learning Based Platform for the Prediction of Sweetener and Its Relative Sweetness - five machine-learning methods and conformational-independent molecular fingerprints to derive the classification and regression models for the prediction of sweetener. A user-friendly platform “e-Sweet” for the automatic prediction of sweetener.
  - BitterSweetForest: A Random Forest Based Binary Classifier to Predict Bitterness and Sweetness of Chemical Compounds - the development and validation of a machine learning model based on molecular fingerprints to discriminate between sweet and bitter taste of molecules.

Datasets used for developing VirtualTaste

- Sweet data: SuperSweet database
- Bitter data: BitterDB
- Sour data: extracted from public databases ChEMBL and PubMed
- Bitter Receptor data: extracted from ChEMBL and BitterDB

Features that make the solution better from the others

- Predict three different tastes: sweet, bitter and sour.
- Very good results for different measurement methods:

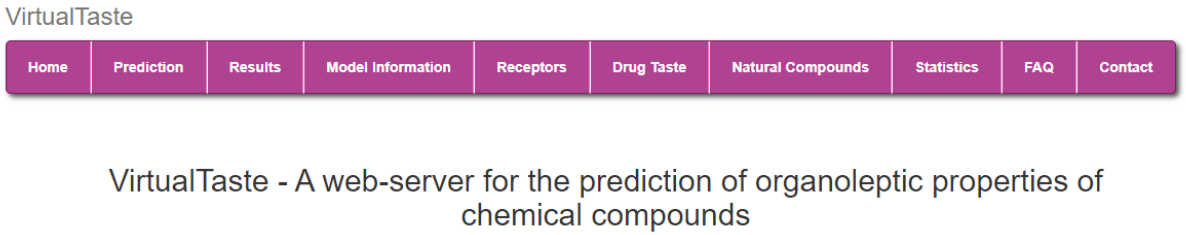
Table 1. Performance statistics for the VirtualTaste models applied to cross-validation and external validation sets

VirtualTaste models Data sampling method		VirtualSweet SMOTETC	VirtualBitter SMOTE VDM	VirtualSour AugRandOS
Cross-validation	*Prediction accuracy*	0.88	0.94	0.98
	Sensitivity	0.97	0.94	0.94
	Specificity	0.96	0.92	0.97
	ROC–AUC	0.99	0.97	0.97
	F-measure	0.87	0.94	0.98
External validation	*Prediction accuracy*	0.89	0.90	0.97
	Sensitivity	0.86	0.88	0.80
	Specificity	0.92	0.97	0.99
	ROC-AUC	0.95	0.96	0.99
	F-measure	0.88	0.88	0.84

- Free to use web platform available with built in tool for taste prediction.
- Many additional information presented on the platform that are related to the work.

Results and benefits of the research

- Free to use web platform that can be used by researchers for the prediction of organoleptic properties of chemical compounds.



- Help for the experimental food chemist to predict compounds of three different tastes in a fast and easy way.
- The publication that describes the research and gives the scientists an overview of the related work, available datasets and sources with many useful information related to this subject.