

Deep Learning for Odor Impression Prediction

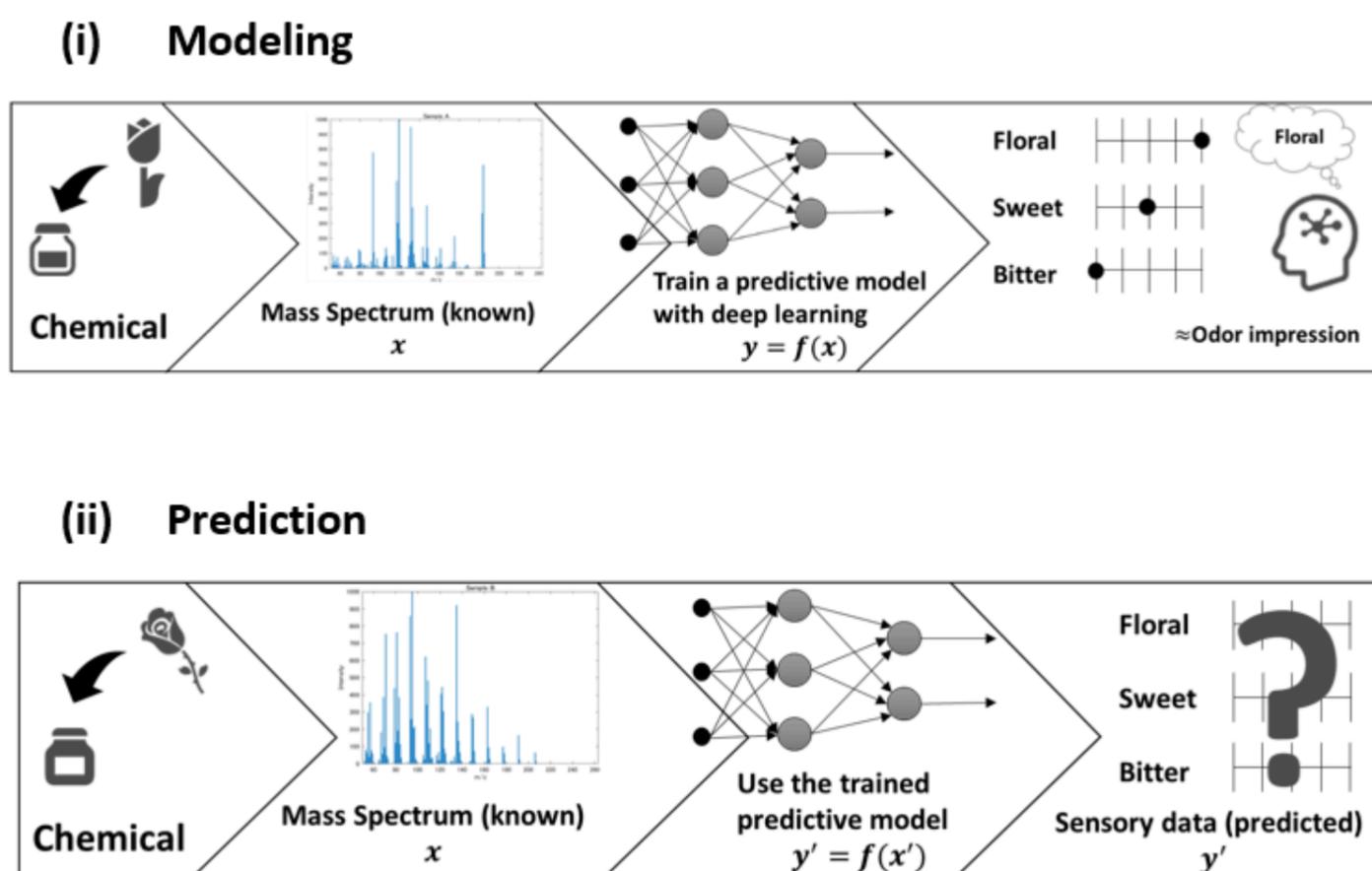
Introduction

Out of five major senses, smell and taste are responsible for experiencing flavors or fragrances. Despite the importance of smell in our daily lives, it has not received the same kind of attention from researchers as other domains such as computer vision, speech recognition & natural language processing.

Our lab studies machine olfaction to predict the smell impression of the target odor using deep learning technique.

Odor Impression Prediction using Mass Spectra

A dataset for target odor signals and a dataset for input odor signals should be prepared separately. We use the results of sensory evaluation test conducted by A. Dravnieks et al. on a scale of 0 to 5 (referred to as 'Sensory Data' or 'Y') for the target signal and its mass spectrum for input signals (referred to as 'Mass Spectra data' or 'X') as shown in fig. 1.



A 9 layer neural network as shown in fig. 2 using an auto-encoder was used for predicting the characteristic of target odorant. We use sensory evaluation data and mass spectra data by evaluating 160 odorants with 146 descriptors.

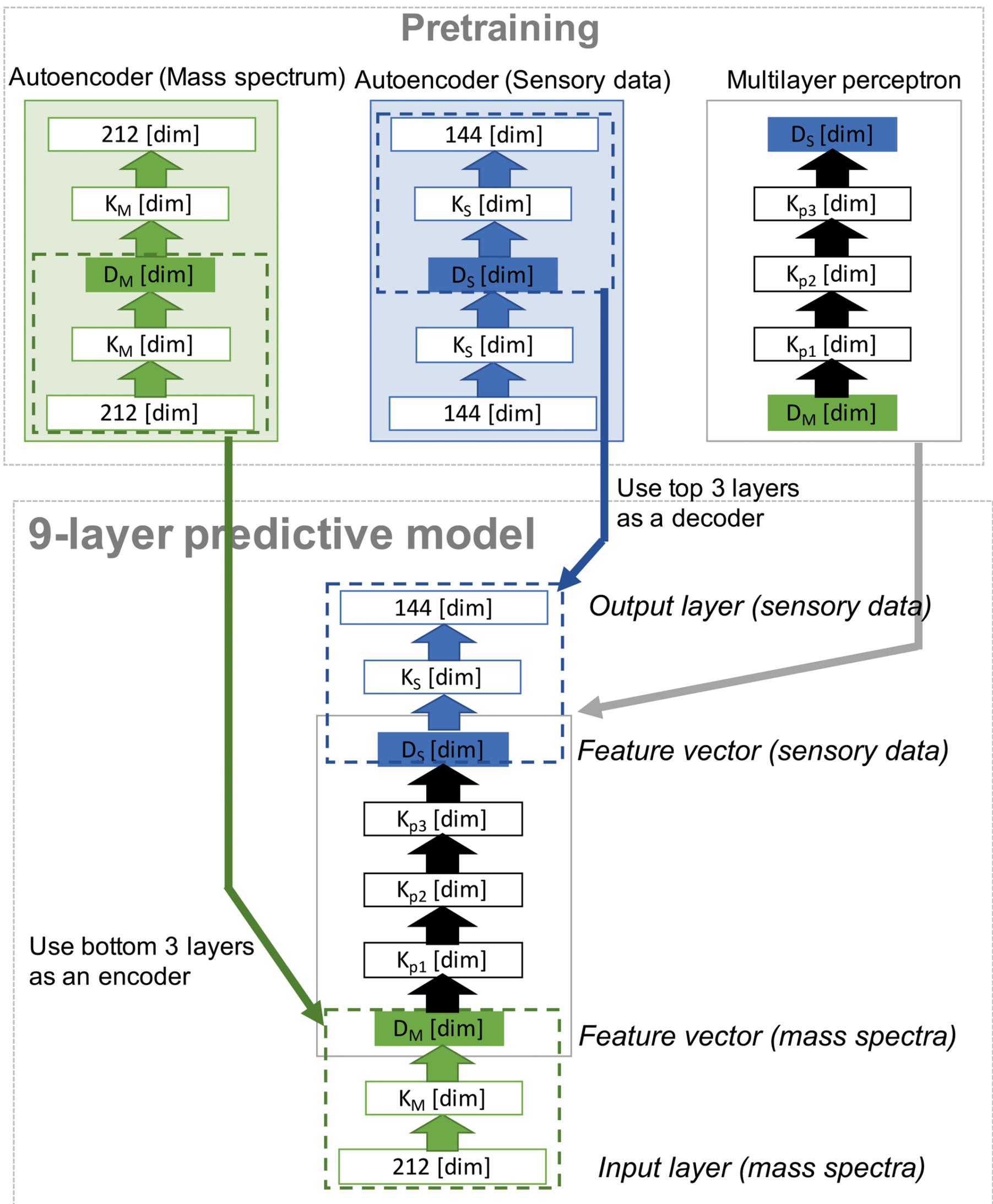


Fig.2: Schematic diagram of 9 Layer predictive Model [2]

Input data (mass spectrum data) is 121×212 , the target data (sensory evaluation data) is 121×144 . 121 is the number of odorants, 212 is the number of dimensions of each sample (51 m / z-262 m / z), and 144 is the number of descriptors.

The results of predictions obtained by the proposed new method have notable accuracy ($R \cong 0.76$) in comparison with a conventional method like partial least square (PLS) is ($R \cong 0.61$).

Odor Character Clustering

Our research team also proposed an approach for predicting the odor characters of chemicals using clustering with larger granularity containing similar odor descriptors.

For these task, we visualized the positional relationship of the descriptors in two-dimensional space by multidimensional scaling (MDS) as shown in Fig. 3 on Sigma-aldrich catalog and NIST dataset in a binary value form that represents the existence of odor impression only.

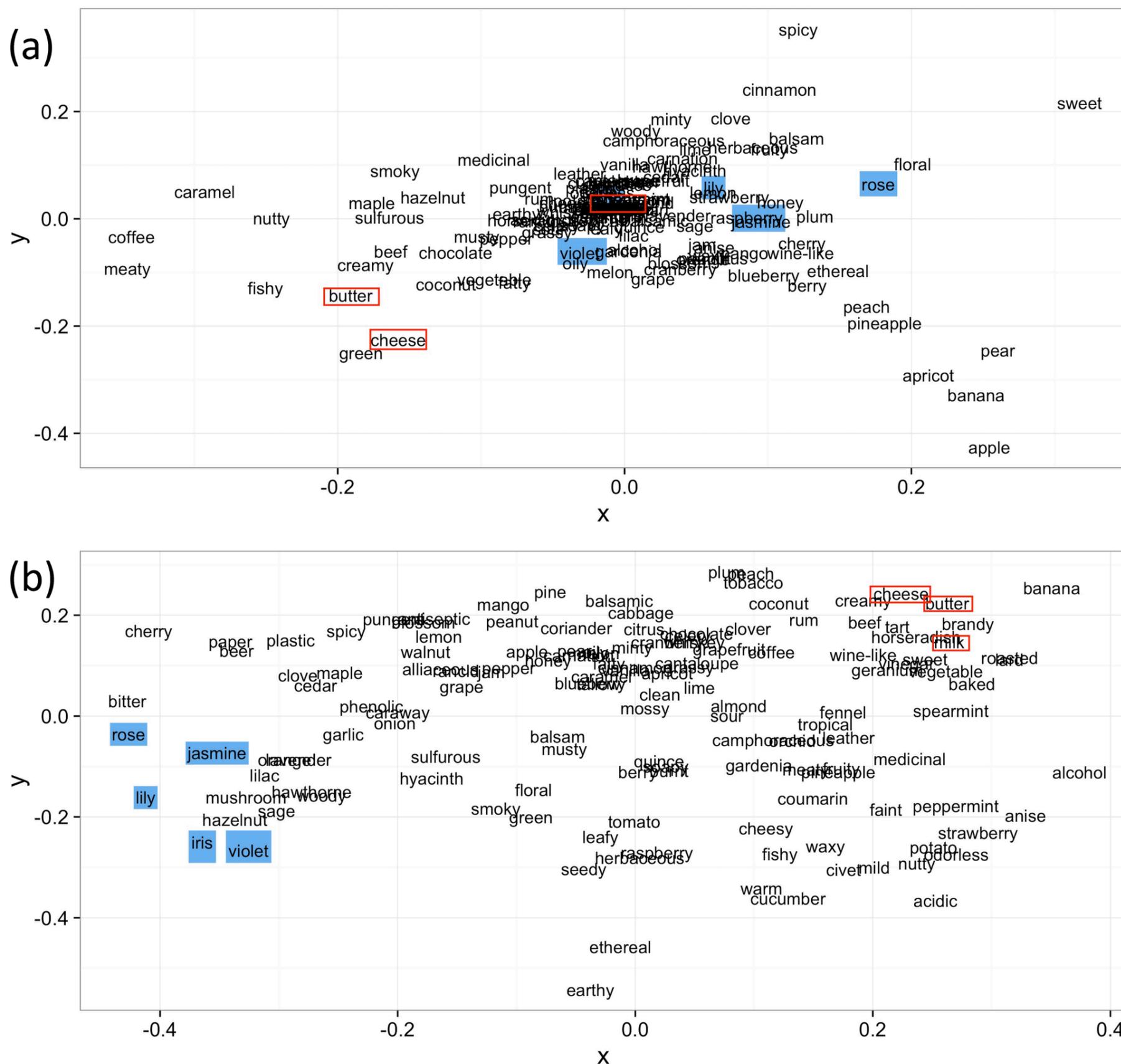


Fig.3: MDS diagrams based on (A) correlation coefficient similarity using binary value, (B) cosine distance based on similarity of words using natural language modeling method called Word2Vec [3]

It was confirmed that the calculated similarities between the descriptors were greatly different between the two methods. As shown in Fig 3(B), in the two-dimensional diagram obtained from Word2Vec, descriptors such as "rose", "jasmine", "lily", "iris", and "violet", and descriptors such as "milk", "cheese", and "butter" are placed closer than in the diagram based on the dissimilarity matrix calculated from the correlation coefficient (Fig 3(A)).

Fig. 4 shows the predictive model with clustering procedure. When any of the descriptors in a cluster has a value of 1, the value of the corresponding cluster that includes the sample is set to 1.

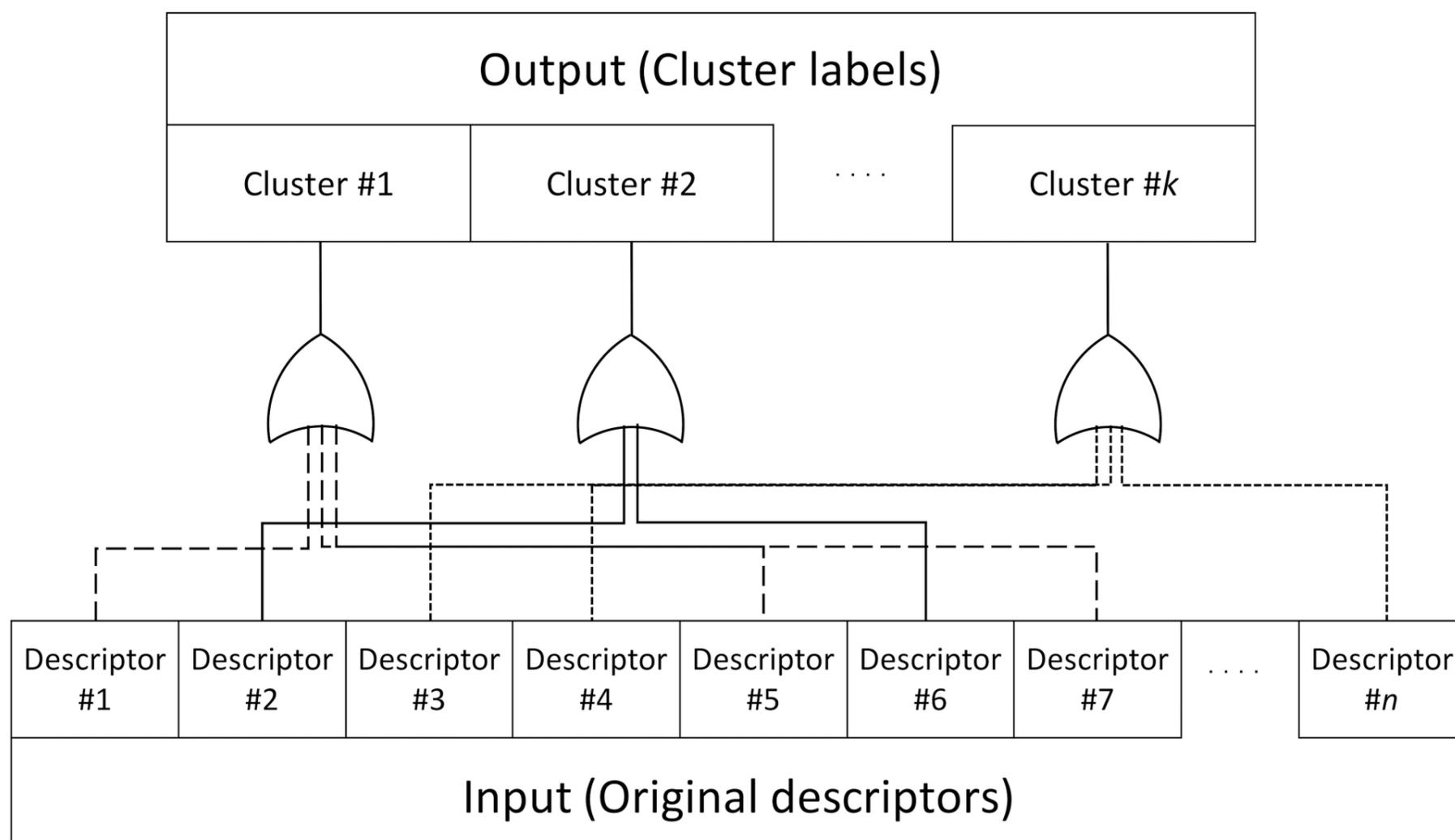


Fig.4: Odor Character clustering [3]

For example, when "rose", "lavender", and "iris" belong to a certain cluster, a sample having at least one of the three descriptors in the original catalog data is regarded as having the odor character of the cluster.

References

1. Nozaki Y., Nakamoto T. (2019) An Olfactory Sensor Array for Predicting Chemical Odor Characteristics from Mass Spectra with Deep Learning. In: Fitzgerald J., Fenniri H. (eds) Biomimetic Sensing. Methods in Molecular Biology, vol 2027. https://doi.org/10.1007/978-1-4939-9616-2_3
2. Nozaki Y, Nakamoto T (2016) Odor Impression Prediction from Mass Spectra. PLoS ONE 11(6): e0157030. <https://doi.org/10.1371/journal.pone.0157030>
3. Nozaki Y, Nakamoto T (2018) Correction: Predictive modeling for odor character of a chemical using machine learning combined with natural language processing. PLOS ONE 13(12): e0208962. <https://doi.org/10.1371/journal.pone.0208962>