

Study on Molecular Information Processing for Volatile Organic Compounds

鄭, 曉帆

<https://hdl.handle.net/2324/7329497>

出版情報 : Kyushu University, 2024, 博士 (学術), 課程博士
バージョン :
権利関係 :



氏 名 : 鄭 曉帆

論 文 名 : Study on Molecular Information Processing for Volatile Organic
Compounds (ガス分子の情報処理に関する研究)

区 分 : 甲

論 文 内 容 の 要 旨

A wealth of information is contained in gaseous molecules. Living organisms use smell to find food resources, transmit signals, etc. Gas sensors inspired by how living organisms utilize smells can be used for further applications such as food quality assessment, gas source tracing, and so on. For the engineering application of smells, sensing and identification of odors are fundamental technologies. Comprehending the relationship between molecular structures and the corresponding odorous characteristics can result in the promotion of gas sensor development. In order to achieve this, the data-driven approach is an option where the model for predicting odor characteristics is trained using the data composed of molecules labeled with odor descriptors and is analyzed to investigate the structure-odor relationship, other than studying odorant-receptor interactions in biochemistry. On the other hand, multiple odor substances are mixed in the actual situation where odor information is used, it is necessary to extract the sensor response of individual component gases from one of the mixed gasses. This thesis includes two aspects of (1) constructing the model for predicting odor descriptors from molecular structure with the ultimate goal of elucidating the relationship between molecular structures and the corresponding odorous characteristics; and (2) individual gas measurement data extraction.

As the first aspect, the artificial neural networks for predicting odor descriptors such as 'fruity' and 'sweet' of input molecules were constructed in Chapter 2 and 3. The molecular structure representations of molecular graphs and SMILES (simplified molecular-input line-entry system) were used as model inputs respectively.

In Chapter 2, the model that used molecular graphs (i.e. atomic features and interatomic relationship of chemical bonds as well as distance) as inputs was proposed by modifying the Transformer model so that the attention mechanism can cooperate with the interatomic relationship. The decoder module in the model was used for result interpretation. In the experiment, 98 kinds of odor descriptors were predicted with the proposed model, and substructures related to the perception of 19 odor descriptors were summarized through visualizing the attention matrix in the decoder module.

In the previous studies on the model using SMILES as inputs, SMILES was often addressed in the same way as natural language sentences and the differences between them were ignored because SMILES is a sequence of symbols. In Chapter 3, 2-encoder model, a pretraining model,

was proposed by focusing on the lower dependency of symbols to the contextual environment in a SMILES than one in a natural language sentence and the corresponding of one compound to multiple SMILES. The pretrained model was used to predict 98 kinds of odor descriptors as well as 22 therapeutic-related molecular properties. For odor descriptors prediction and more than half of therapeutic-related molecular properties prediction tasks, the model pretrained with the proposed model achieved better results compared to BERT which is adept at natural language. And these results showed the proposed pretraining model was more robust in predicting a variety of molecular properties after fine-tuning.

As the second aspect, a method was developed to extract the sensor response and spatial concentration distribution for each component gas from the output of a localized surface plasmon resonance (LSPR) gas sensor device in the environment of a mixture of gases using the matrix decomposition in Chapter 4 and 5. The method includes transforming the sensor output into the processed response, extracting the spatial concentration distribution of the target gas with the processed response of the target gas and estimating the number of component gases for practical applications.

In Chapter4, the linear proportionality between gas concentration and the processed response of gas to sensor as a requirement for applying matrix decomposition, was verified to hold approximately through the measurement experiment. In the experiment using real measurement data, diffusion traces of target gases were visualized individually through matrix decomposition.

Determining the number of component gases is a critical issue for matrix decomposition in practical applications. In Chapter5, through the experiment using simulation data for evaluating the estimated spatial concentration distribution with an unknown number of component gases, the results suggested that (1) if the response of the target gas to the gas sensor is measured in advance, estimating the number of component gases correctly is not necessary, simply choosing a relatively large number of component gases can achieve a well estimated concentration distribution of the target gas; (2) the number of component gases can be estimated correctly with information criterion if the noise possessed by measurement data is relatively small.