Development of Coffee Classification by Feature Selection Algorithm Based on An Electronic Nose

A404
Ting Shen, Tzu-Hsiang Lo, Yu-Cheng Hsieh, Kea-Tiong Tang*
Department of Electrical Engineering National Tsing Hua University Hsinchu 30013, Taiwan
*kttang@mx.nthu.edu.tw

Abstract

This project aims to develop a feature selecting algorithm using an electronic noses (E-noses) system to classify different varieties of coffee beans based on the gas molecules in their aroma.

We present a feature selecting algorithm that combines a separability indicator (SI) and a support vector machine (SVM) to effectively select features and optimize the classification process. By testing this method on data from three distinct coffee brands, this algorithm achieve a classification accuracy of 100%.

Keywords: *Electronic Nose, gas molecules, Feature Selecting Algorithm, Classification*

Introduction

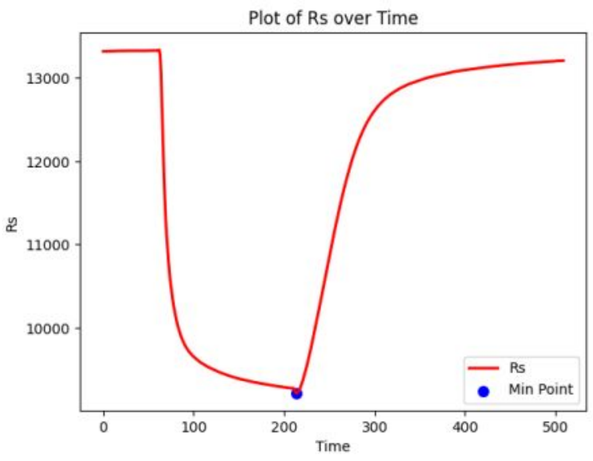
First, gas sensors capture the gas molecules of coffee beans. Next, the data undergo feature extraction, followed by the execution of a feature selection algorithm.

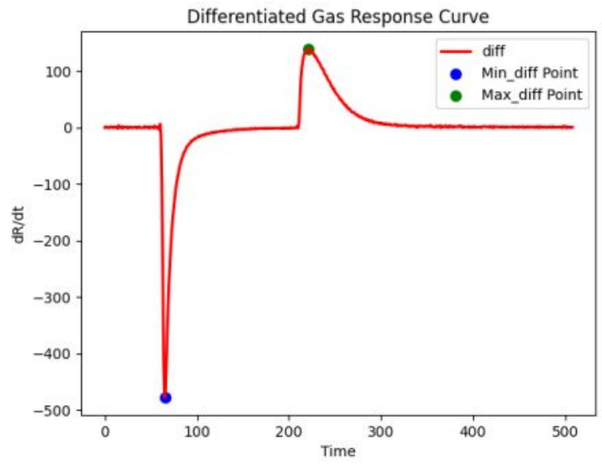
After feature extraction, this project aims to develop a feature selection algorithm. The algorithm calculates SI for different features, sequentially includes various numbers of features, using PCA to reduces data dimensions, and then applies SVM to identify the most discriminative feature combinations to optimize classification boundaries.

Last, different algorithms are executed to compare their performance.

Experimental System Architecture

1. Sensor & Dataset
   a. Sensor: 14 gas sensors consist of sensor array
   b. Dataset: total 69 with 3 varieties
2. Feature extraction
   a. Magnitude: min(R), baseline(Ro)
   b. Difference: min(R)-Ro, (min(R)-Ro)/Ro
   c. Derivative: maximum derivative R, minimum derivative R





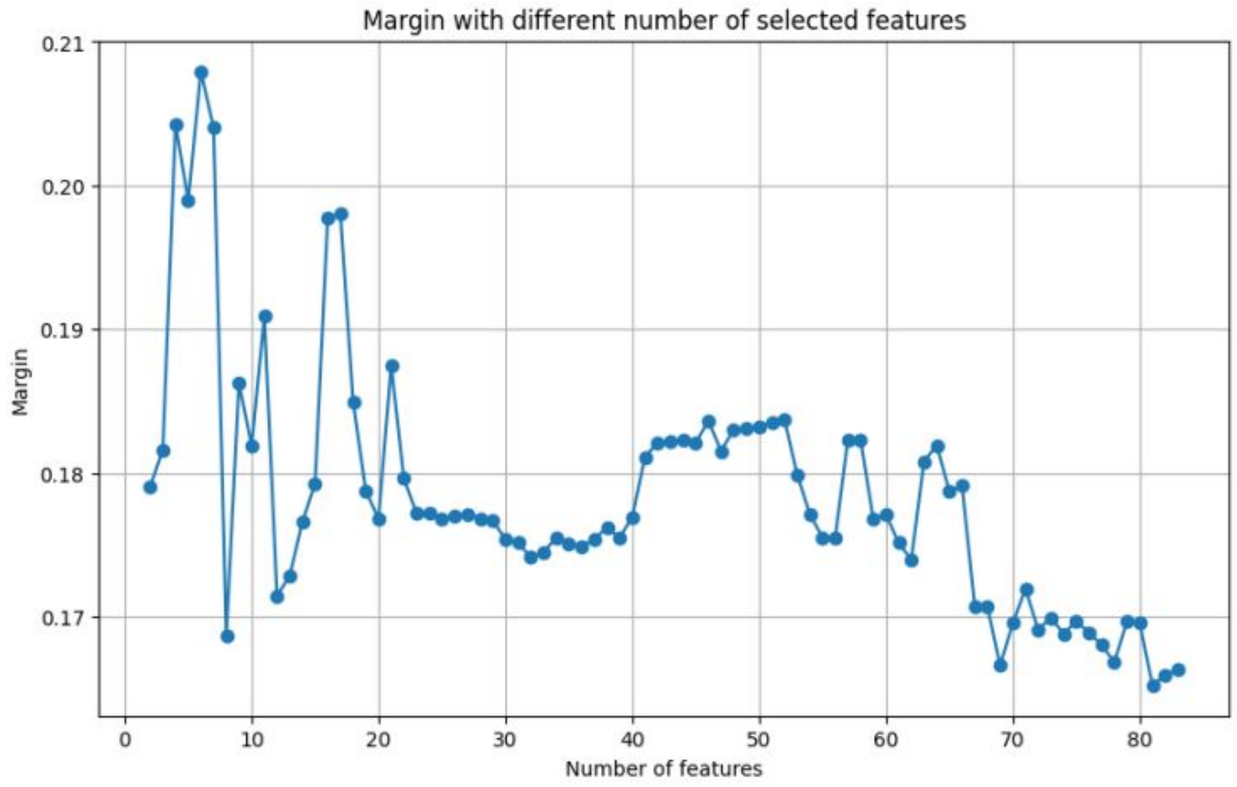
3. Feature selecting algorithm
   a. Separability indicator (SI): peformance of classification

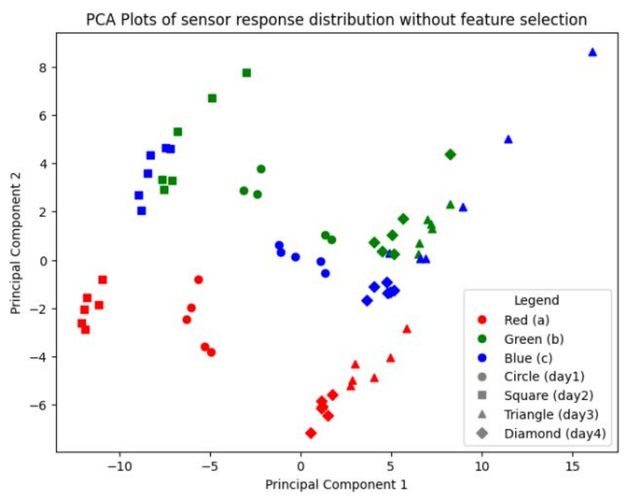
$$SI = \frac{\sigma_{bc}^2}{\sigma_{wc}^2}$$, where σ_{bc} is variance of between-class
σ_{wc} is variance of within-class

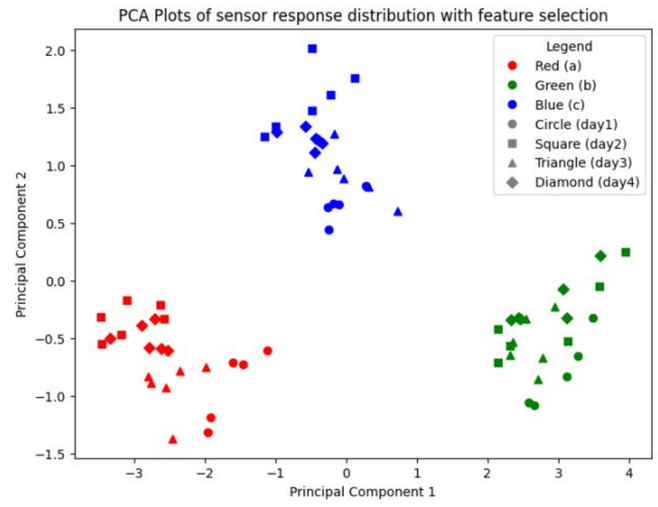
   b. Optimize feature set

Input	Data $X \in \mathbb{R}^{N*M}$, N samples with M features
Output	Optimize feature set S
Procedure	
1.	Compute the SI for each feature and sorted in descending order
2.	**For** k = 2, 3, …M **do**
3.	Include the k features with highest SI in S_k
4.	2D PCA(S_k)
5.	Send 4. result to SVM to compute the margin m_k
6.	**if** SVM has no misclassification:
7.	record (S_k, m_k)
8.	**End For**
9.	Choose largest $m_{k,max}$ among all m_k
10.	S = $S_{k,max}$
11.	**Return** S

Result







Classification accuracy with/without proposed feature selection method.

	Train	Test	Accuracy
mRMR	03/10	03/13, 03/15, 03/17	0.94
	03/13	03/10, 03/15, 03/17	0.94
	03/15	03/10, 03/13, 03/17	1.0
	03/17	03/10, 03/13, 03/15	0.98
Average			0.97
SVM-RFE	03/10	03/13, 03/15, 03/17	0.85
	03/13	03/10, 03/15, 03/17	0.80
	03/15	03/10, 03/13, 03/17	1.0
	03/17	03/10, 03/13, 03/15	0.96
Average			0.90
Feature Selection Algorithm	03/10	03/13, 03/15, 03/17	1
	03/13	03/10, 03/15, 03/17	1
	03/15	03/10, 03/13, 03/17	1
	03/17	03/10, 03/13, 03/15	1
Average			1

Conclusion

The top figure shows the results of the feature selection algorithm, depicting the margins obtained after SVM for different numbers of selected features. It can be observed that selecting the six features with the highest SI yields the maximum margin. It represents the coffee classification most effectively.

The plots of PCA show the improvement in classification accuracy achieved through feature selection algorithm. The analysis on left side using all features, indicating poor classification performance. In contrast, the plot on right side demonstrates PCA using the top six features with the highest SI values, showcasing clear differentiation between different classes.

As shown in Table, the results demonstrate the accuracy of different algorithms. The results demonstrate that the accuracy of this feature selection algorithm surpasses that of two commonly used gas recognition and classification algorithms, mRMR and SVM-RFE. Therefore, the feature selection algorithm provides a more efficient and reliable way to identify the gas of coffee beans.