Responses to the comments

The new version of the paper partly answered my previous concerns. However, the presentation still needs more precision to highlight the technical value of the paper.

1) The first version of the paper states that “We can see the PCA and LDA 14 analysis proposed in this paper have better performance to detect stenches”, while the second version the authors state that “We can see random forest in this paper has better performance to detect stenches”. How can this be explained?

   Thank you for the comment.In fact, the first version using PCA and LDA analysis has strict conditions.We first use PCA and SVM to marked the methylcyclopentanone, γ -undecalactone, 2-methylindole.Then we use LDA analysis to classify the other three odors.But in the second version of the paper ,we only use PCA (SVM) or LDA(SVM) for all data.So the accuracy is lower than the first version using PCA and LDA analysis.And the second version of the paper wrong use of cross-validation.It causes that the training set accuracy and the test set accuracy is similar. The third version also use PCA (SVM) and LDA(SVM) not the combination of PCA and LDA analysis.

2) Page 3, line 82. The authors state “we change the sensors types in sensor array to solve the new problem”. Which exactly is the new problem? Which sensors were changed in the sensor array and why?

   Thank you for the comment.We used to use similar system to detect the odors of paraffin wax. Compared to the old electronic nose, we change the sensors types in sensor array to detect the odors. All sensors are new in the sensors array.Because different odors need to find the available sensors.

   As for these five stenches, we find the voltages change little in negative pressure. So we use two needles. One needle is to sample the odors, the other is to input the air. And we adjust the experimental process and acquisition mode. We use the new algorithm and new hardware to solve the detection problem. (line 82)

3) The graphics in figure 3 are too small. It is impossible to distinguish between colors and shapes of curves. The authors could separate each graphic into several ones with a small number of curves each one. Plotting 16 curves on a small graphic is not useful for interpretation.

   Thank you for the comment.All line has been thickened. The size of graphics has been changed.And I choose nine sensors to do the experiment.The rule is that the same sensors whose voltages are higher.And we discard sensors that only change a little in

methylcyclopentanone and change little in other odors.So I can clearly show the response curve. (line 100)

4) The graphics in figure 4 are too small to analyze. The same remark as before.
   Thank you for the comment. All line has been thickened. The size of graphics has been changed.(line 150)
5) Page 5, line 151. I do not agree with the remark “Figure 4a displays that all sensors will have a big change in voltage value in 1-60”. It seems that three of four sensors have a small change in voltage value. However, doe to the size of the graphic, which is too small, it is not possible to say which ones.
   Thank you for the comment. All line has been thickened. The size of graphics has been changed.(line 150). I must note that green line and blue line coincide with the red line.
6) Page 5, line 152. The authors state that “Figure 4b displays that four sensors will have a big change in voltage value at all times. So we can separate the data into two parts and recognize the sample one by one.” Why one has to separate data in two parts if four sensors have big change … Please explain this statement.
   Thank you for the comment.I have delete this statement.This statement is based on the first version of the paper.Becasue the four sensors have big change,we can marked γ -undecalactone, 2-methylindole independently.
7) The comments of the authors related to figure 4 b, state that only 4 sensors are selected. Could it be more useful to build a sensor network using only the 4 sensors identified as more performant? Does the other sensors have a significate impact on the performance of the classification?
   Thank you for the comment.
   The consideration of fewer sensors to detect the odors makes the third version paper.
   During the experiment, we find sensors(2602,826) are sensitive to 2-methylindole and methylcyclopentanone.Sensor826 is also sensitive to γ -undecalactone and 2-phenylethyl alcohol. Figure 5 uses no.3 sensor and no.8 sensor to draw the response curve in 1—120s.We must note that the blue full line,green full line and blue full line coincide with the full red line.Beacuse the shape of the curves is similar,we find that only using these two sensors will be difficult to classify isovaleric acid and steam.But these two sensors will contribute to the classification of 2-methylindole and methylcyclopentanone.(line 153)

8) Page 5, sub-section 2.4. Why does the authors need to randomize the data before performing the 5-fold cross validation?

Thank you for the comment.I forget that performing the 5-fold cross validation will randomize the data.So I change the description.(line 136)

9) Figure 5 is too small. It is absolutely impossible to read the legend.
Thank you for the comment.I reduce the number of pictures. All line of shapes has been thickened. The size of graphics has been changed. (line 166)

10) Page 6, line 160. “Figure 5 uses no.6 sensor and no.13 sensor to draw the response curve in 60—120s.” However, the graphic in figure 5 draws much more than two curves.
Thank you for the comment.In fact,we have 6 kinds of odors.Each odor has two lines.

11) Figure 6 is too small too.
Thank you for the comment.Figure 6 has been deleted. The reason is: When we use LDA to reduce data dimension,the eigenvalues of samples which is sampled from 1 to 60 seconds are [0.9979,0.9674,0.9158,0.8644,0.3126]. When we use LDA to reduce data dimension,the eigenvalues of samples which is sampled from 60 to 120 seconds are [0.9926,0.9760, 0.9104,0.8596,0.3754].So we reduce the dimension of data to four dimensions. (line 190)
So 2D picture is not helpful.

12) It is not clear how does the authors use the % variation explained by the first two components.
Thank you for the comment. The data variances of PC1 and PC2 are marked in the picture.(Figure 6)
The first two components, PC1 and PC2,account for 93.32% of the data variance. Figure 6b shows the distribution of the samples which is sampled in 60-120s. The first two components, PC1 and PC2,account for 91.08% of the data variance. The first two components of PCA are all bigger than 85%.So reducing the data dimension to two dimensions still remains the feature information.We also find that 2-methylindole, methylcyclopentanone can be easily classified using simple machine learning algorithms.And γ -undecalactone is mixed with 2-phenylethyl alcohol and isovaleric acid.(line 172)

13) Page 6, line 181. The authors state that “Figure 4a shows the reason for the separation of methylcyclopentanone.” The authors must provide clear justifications for their statements.
Thank you for the comment.In fact,we can compare the picture of methylcyclopentanone with other odors.It is special.All sensors have big voltages.(line 149)

14) Concerning the use of the SVM to define the boundary in figure 6d, the authors provide the values of the parameters. However, the authors must provide an accurate description of the data used for training the

SVM. If I understood well, they use a 5-fold cross validation approach. For instance, which is the size of the data set?

Thank you for the comment.

Then we use grid search method to find the best penalty parameter and kernel parameter gamma for the SVM using RBF kernel. The specific multi-classification tool is SVMs using libsvm function in Matlab. We use 5-fold cross-validation ten times to get the training set accuracy and testing set accuracy.The standard deviation of classification rate is the average of 10 times of 5-fold cross validation.(line 179)

The data set is consisted of 15 samples of 2-phenylethyl alcohol, isovaleric acid, methylcyclopentanone, 2-methylindole and 25 samples of γ -undecalactone.All samples is a 1x28 matrix and the twenty-eighth column shows the stench label.So the data set is a 100x28 matrix.(line 233)

15) The graphics in figure 7 are too small. Please consider previous remarks.

Thank you for the comment.The size of graphics has been changed.

16) Page 8, lines 209, 210. The authors must recall the name of three algorithms they analyze in subsection 3.2.

Thank you for the comment.I check the name of three algorithms.

17) Page 8, line 226. “The training set accuracy is 0.9904, the testing set accuracy is 0.9913.” It is strange that the test accuracy is better that the training accuracy. Is it correct?

It isn’t correct.And I redo the experiment.I add the standard deviation of classification rate.It will show the balance of the accuracy10 times of 5-fold cross-validation.