

Aroma-based Localization in GNSS-denied Environments

Saiful Islam⁽¹⁾ Elena-Simona Lohan⁽²⁾ Philipp Müller⁽²⁾ Mohammad Zahidul Hasan Bhuiyan⁽¹⁾

⁽¹⁾ Finnish Geospatial Research Institute FGI,
National Land Survey NLS
Geodeetinrinne 2, FI-02430 Masala, Finland
firstname.lastname@nls.fi
⁽²⁾ Tampere University
33720 Tampere, Finland
firstname.lastname@tuni.fi

Abstract

This paper studies infrastructure less localization solutions using aroma fingerprints. These fingerprints are collected under varying conditions from different indoor locations using Ion Mobility Spectrometry based Electronic Noses. A supervised machine learning algorithm for data processing location estimation is proposed. The non-parametric system is trained with data from all locations, and its performance evaluated using data from the same locations collected under different environmental conditions. Five different classifiers are studied and tested for location estimation. The Stochastic Gradient Descent classifier achieved the highest accuracy, with the k NN with Euclidian distance also performing reliably under different conditions.

1 Introduction

Electronic Noses (eNoses) have been used in the literature to identify gas leakage, for food quality analysis, and other identification tasks. However, eNose based localization was first investigated in [1]. This paper investigates the results and suggestions mentioned in [1] [2] further. Datasets studied in this paper were taken from [1]. The eNoses used for this research contain an ionization chamber and 14 electrodes for measuring currents of ionized molecules. Two datasets were taken from the same locations, under different environmental conditions. Two aroma fingerprint databases were prepared using the datasets [1]. A supervised Machine Learning (ML) algorithm is used to predict the location of an object. The concept was to train the ML model using a dataset with measurements from different locations. A second dataset from the same locations but different environmental conditions was then used for testing the performance of the ML model. In addition, we tested the performance of the model with the first dataset when it was trained using the second dataset. Five classifiers were used to predict the location of the object, namely the K-Nearest Neighbor (kNN), the Linear Discriminant Analysis (LDA), multiclass Support Vector Machines (SVM), the Random Forest (RF), and the Stochastic Gradient Descent (SGD). Based on a given dataset, each classifier offers a probabilistic output. In addition, we study how the accuracy depends on the changing surroundings and the external noise.

2 Data collection and Processing

Ion Mobility Spectrometry (IMS) is designed to identify and evaluate various airborne chemicals very rapidly. The data sets used in this study were obtained from Chempro 100i, a handheld chemical detector including an IMS. Multidimensional sensors (16 electrodes) can simultaneously detect a wide range of chemicals. For data collection, only 14 of the 16 electrodes were active. Electrodes 8 and 16 were used only for airflow control and therefore did not provide useful information for classification [1]. Most of the simulation part of this study has been performed in the Python platform. A well-defined structure is needed to provide the

best coding solution. The Python libraries (Sklearn) and frameworks are well-enriched to ensure less complexity and faster programming.

3 Data Description

Two data sets of approximately 600 seconds were collected from seven different locations with 1 Hz frequency. This means for each location two times ~600 samples were taken (see [1] for details). During the weekend, the first data set ‘Data Empty’ was collected to ensure that no (almost) individuals were present at those locations. Another dataset ‘Data Crowded’ was collected during the weekday from the same locations, as illustrated in Figure 1.

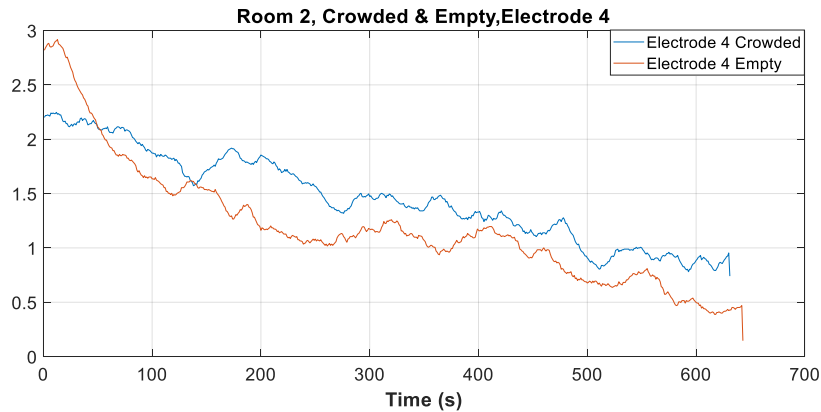


Figure 1: Ion mobility plot in both environments

From Figure 1 it can be seen that there is a rapid decrease in the measurement electrode 4 at the beginning (about 120s). After that, the values of the electrode continue to decrease over time until the end of the measurements. It is observed that, under both environmental conditions, the values of the electrode decay almost equally, which is a good sign of data reliability. The ML algorithm can easily predict them during the matching stage, depending on their similarities. The mean plot in Figure 2 shows the upper, lower, and average values of all the electrodes in Room 2.

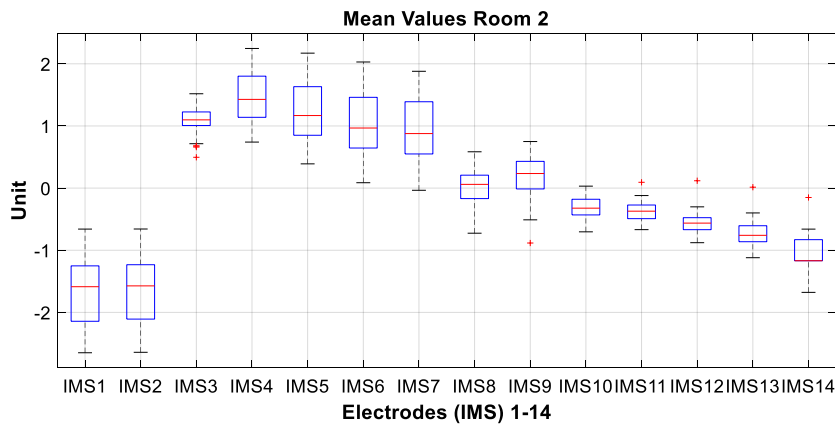


Figure 2: Mean plot of Room 2

4 Result and Analysis

Multiple experiments were carried out using different parameters in the analytical section. In the same experiment, each classifier is trained and tested with the same training and test data. The experimental results are the results of a master thesis work of the main author [3]. Table 1

summarizes the results: without Principal Component Analysis (PCA) in the upper table, with PCA (in the middle table) and with the Stochastic Gradient Descent (SGD) (in the lower table).

Table 1. Experimental results based on three experiments

Experiment 1				
Classifier	Distance	Value of k	Classification Rate (Process 1)	Classification Rate (Process 2)
kNN	Euclidean	5	37.56%	29.96%
kNN	Euclidean	3	37.33%	30.21%
kNN	Minkowski	5	37.53%	29.57%
kNN	Minkowski	3	37.21%	29.89%
kNN	Manhattan	5	36.89%	25.95%
kNN	Cityblock	5	36.94%	25.94%
kNN	Canberra	5	29.83%	29.66%
kNN	Cosine	5	29.97%	34.12%
LDA	--	--	35.42%	31.47%
SVM	--	--	34.05%	31.31%
RFC	--	--	29.55%	22.97%

Experiment 2, PCA Applied			
Classifier	Training Size	Classification Rate (Process 1)	Classification Rate (Process 2)
kNN	75%	29.16%	7.90%
kNN	25%	32.05%	12.06%
LDA	75%	28.68%	16.04%
LDA	25%	28.18%	10.83%
SVM	75%	40.83%	8.16%
SVM	25%	37.28%	12.70%
RFC	75%	33.86%	9.53%
RFC	25%	27.05%	13.23%

Experiment 3				
Classifier	Training Size	Classification Rate	Status	Remarks
SGD	75%	53%	Trained as crowded and test as empty	loss=squared_hinge
SGD	25%	47%		

The first experiment was carried out under two processes or scenarios: (i) in Process-1, 'Data Crowded' was used to train the model and Data Empty was used to test the model; (ii) in Process-2, 'Data Empty' was used to train the model and Data Crowded used to test the model. In both cases, the training size was 75% of the test data. Experiment 1 shows that the kNN classifier with Euclidean distance, and k=5 predicted the accurate locations around 38% in Process-1. On the other hand, Process-2 in most cases achieved less accuracy than Process-1. The potential reason for the overall performance degradation in Process-2 is that 'Data Crowded' contains more environmental information than 'Data Empty'. The model trained with more environmental data would, therefore, be more accurate. Moreover, the experiment shows that different values of 'k' have no significant impact on the accuracy of the classifier. However, different distance metrics have some impact on accuracy.

The PCA method was implemented in the second experiment. PCA is a dimensionality-reduction technique most commonly used to decrease the dimensionality of a bulky dataset. The process methods (1&2) remained the same as in the first experiment. The second experiment shows that the SVM classifier achieved approximately 41% accuracy when the training size was 75 percent. However, the rest of the classifiers did not perform well in Process-2.

The third experiment was performed using the Stochastic Gradient Descent (SGD) classifier, which can be obtained directly from the Scikit-Learn Library. Based on previous experiences, Data Crowded was used to train the model. Several loss functions are available in the Scikit-Learn library. The square hinge loss function was used to train the SGD classifier. It achieved a

maximum probabilistic accuracy of 53%. The experimental process is similar to Experiment 1: instead of multiple classifiers, only a unique classifier has been used to focus more on the findings of several experiments.

The final experiment was conducted to study the impact of the amount of training data on the classification accuracy. Figure 4 shows the accuracies of all classifiers for variable test sizes from 1% to 99%. It can be seen from this final experiment that the SGD classifier in general outperforms the other classifiers clearly. However, when the test size is very high (> 90 %) and the training size is very low (< 10 %), the accuracy of the SGD classifier drops dramatically. On the other hand, even with small training size, the output of the kNN classifier with Euclidean distance and $k=5$ stays almost stable all the time. Significantly, very few training data are generally available in real cases. In that case, the kNN classifier could still perform reliably.

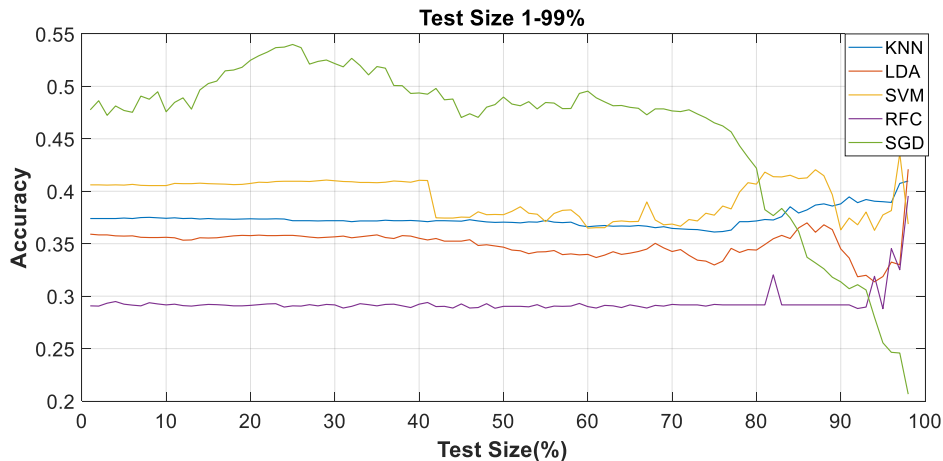


Figure 4: Accuracy plot with variable testing and training size

5 Conclusion and Future Work

It is possible to improve accuracy of aroma fingerprint-based localization by using advanced machine learning algorithms, considering primary ideas and suggestions from [1]. Before using aroma fingerprinting as a trusted localization method, however, some issues still need to be addressed. Moreover, in classification problems, no single algorithm wins all the time. Different classifiers react with varying datasets individually. In conclusion, the IMS base eNoses system may have enormous possibilities in localization applications when the limitations regarding composition and processing errors have been removed.

Use of the AdaBoost algorithm along with other classifiers, can be considered as future work. The AdaBoost classifier merges a set of classifiers that are weak or poorly managed into a powerful classifier. The new classifier's accuracy is expected to be greater than that of any individual classifier.

References

- [1] Müller, Philipp & Lekkala, Jukka & Ali-Löytty, Simo & Piché, Robert. (2017). Indoor Localisation using Aroma Fingerprints: A First Sniff. 10.1109/WPNC.2017.8250046.
- [2] Minaev, Georgy & Müller, Philipp & Visa, Ari & Piché, Robert. (2018). Indoor Localisation using Aroma Fingerprints: Comparing Nearest Neighbour Classification Accuracy using Different Distance Measures. 397-402. 10.1109/ICoSC.2018.8587811.
- [3] Islam, Saiful (2019). Infrastructure-less based positioning: Localization in GNSS-denied environments, Master's thesis. Tampere University. <http://urn.fi/URN:NBN:fi:tuni-201908263020>