Chemical Identification with Chemical Sensor Arrays

Quintin Stedman

Introduction

An electronic nose composed of small, inexpensive chemical sensors would have many potential applications in areas such as breath analysis, hazardous gas detection and industrial process monitoring. An electronic nose, in anology to a biological nose, consists of an array of non-specific chemical sensors. With pattern recognition, the system can recognize individual chemicals or scents. Capacitive micromachined ultrasonic transducer (CMUT) chemical sensors are a particularly promising chemical sensor technology. They are micromachined devices, so they can be inexpensively batch fabricated. In addition, they have extraordinary sensitivity; A limit of detection of 50.5 parts-per-trillion (ppt) and sensitivity of 34.5 ppt/Hz to DMMP, a sarin gas simulant have been demonstrated [2]. In this work, I demonstrate the application of machine learning to an array of four CMUT chemical sensors to discriminate between six different chemicals present in low concentrations in nitrogen.

Typically, only the equilibrium response of chemical sensors is used in chemical identification. I show that using the response of the sensor over time provides extra information that improves classification performance.

I also study the robustness of the system with respect to variations in sensor sensitivity, and test its ability to determine the concentration of the detected chemical once it is identified.

The data used in this work has been previously been shown at a conference, with some machine learning tested on it [8], but the feature selection and machine learning work in this paper is new, and shows significant improvement over the previous results.

Data Set

A CMUT chemical sensor consists of a conductive plate over a vacuum cavity. The device is actuated electrostatically by applying a voltage between the plate and the substrate. The top of the membrane is coated with a sorbent coating. Chemicals in air absorb into the coating, increasing the mass of the plate and thus decreasing its resonant frequency according to

$$\Delta f/f = -\Delta m/2m$$

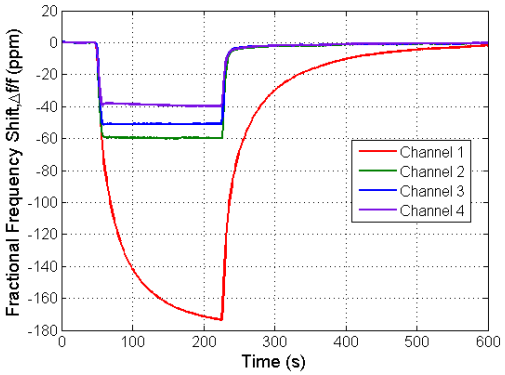

Figure 1 – Sensor response to 2318 ppm acetone

where f is the plate frequency, Δf is the change in frequency, m is the effective mass of the plate and Δm is the change in effective mass.

In this work, an array of 4 CMUT chemical sensors was used as an electronic nose. Three of the chips were coated with different polymers and one was left uncoated. Sensors on different chips were used to make the polymer application easier.

121 chemical detection experiments were performed with 6 different chemicals diluted to <1% in nitrogen. These are summarized in Table 1. The devices were placed in a small test chamber. Nitrogen was flowed through the chamber for 45 seconds. Then, the chemical of interest was flowed through the chamber diluted in nitrogen, maintaining the same flow rate as for the pure nitrogen. Finally, pure nitrogen was flowed through again until the sensors returned to equilibrium. The frequency of each sensor was recorded over time throughout the experiment. The data from one of the experiments is shown in Figure 1.

Table 1 – Summary of chemical identification experiments

Chemical	Concentrations Tested	Number of Measurements
Methane	199-794	14
Carbon Dioxide	2953-4921	33
Water	20-203	11
Ethanol	38-786	18
Acetone	185-3245	15
Ethyl Acetate	161-1608	30

Feature Selection

Some pre-processing was done on the data before extracting the features. First, linear regression was used to remove drift. The frequency of the sensors can drift slowly over time due to variations in room temperature, pressure or due to electrostatic effects in the devices. To remove this, linear regression is performed to the first 45 seconds and last 20 seconds of the data, and the resulting fit is subtracted from the data.

Next, the portion of the data where the chemical is present is selected. I omit the first 4 seconds of data after the chemical turns on. The test chamber takes about 2 seconds to purge, so the first few seconds of data may contain features that are a result of the way the chamber fills rather than the true responses of the sensors to the chemical.

The sensor system cycles through the four sensors, recording the frequency and the time of measurement for each one sequentially. A data point is obtained roughly every 1 second for each sensor, but the time between measurements is not precisely uniform. I used locally-weighted linear regression to estimate the frequency at the precise times I wish to use as features. I used a Gaussian weighting function and a bandwidth parameter of $\tau = 2$ s.

I used two different set of feature sets in this work, corresponding to two different types of sensor operation. In the first data set, I used only the final frequency shift of the sensor just before the chemical turns off. This corresponds to a situation where the sensors are continuously exposed to chemicals and we use the equilibrium response of the sensors to identify them. In some linear models, I also use the squares of the final response as features in order to capture the small non-linearity of final response as a function of concentration.

In the second data set, I estimated the frequency every 2 seconds. This corresponds to a "sniffing" chemical sensor system where the sensors are presented with the sample at a known time. In this situation, we can use the time-dependent response of the sensors to the chemicals, giving us extra information that allows better identification performance than we can get from the equilibrium frequency shifts alone.

Finally, I divide the sensor responses by the average equilibrium response of all the sensors to that chemical. This is to ensure that the algorithms cannot "cheat" by using the fact that different chemicals were presented at different concentrations. This also has the advantage of making the features roughly 1 or less, which tends to make the models easier to train.

Models

Classification

Four models were used for chemical classification: logistic regression, support vector machine (SVM) with linear kernel and support vector machine with Gaussian kernel.

The LIBLINEAR software package [3] was used to implement of L2-regularized logistic regression was used. Multi-class classification is done using a one-vs-rest approach. The regularization parameter was chosen by testing a range of values and choosing the one which maximized performance on the test data.

The LIBLINEAR software package was also used to implement an L2-regularized linear kernel SVM. Again, the regularization parameter was chosen to give the best performance on the test data.

The LIBSVM software package [4] was used to implement the Gaussian kernel SVM. The Gaussian variance σ^2 was chosen jointly with the regularization parameter to give the best performance on the test data.

The final classification algorithm tested was a feedforward artificial neural network trained with backpropogation. In this algorithm, nodes are objects with a sigmoid transfer function:

$$g(z) = \frac{1}{1 + e^{-z}}$$

The nodes are organized into layers. The input layer has one node for every feature, and each node outputs the value of the feature. Each node in the first hidden layer receives a linear combination of these outputs. Each subsequent layer receives a linear combination of the outputs of the previous layer. The last layer (the output layer) has one node corresponding to each category. The network is trained so that the output nodes output 1 if the inputs correspond to their category and 0 otherwise. The network is trained by minimizing the error of the network using the backpropogation algorithm. I used a network with one hidden layer, and used regularization to prevent overfitting.. The number of nodes in the hidden layer and the value of the regularization parameter were chosen to give the best performance on the test data. It is important to note that optimizing a neural network is a non-convex problem, so it is possible for it to get stuck in local minima.

Concentration Determination

Once the chemical is identified, it is useful to determine its concentration. By plotting the final

frequency shifts of the sensors versus concentration, it can be seen that the final frequency shifts are a slightly nonlinear function of concentration. Since the dependence of the frequency shift on concentration is so clear, we do not need a complex model. I chose to use linear regression, with the final frequency shifts and their squares used as features. The inclusion of the squares captures the non-linearity.

The amount of data for each chemical is small, so to avoid overfitting I used weight decay [7]. With weight decay, the linear regression cost function is

$$\frac{1}{2}\sum_{i=1}^{m}(y_i - \theta^T x_i)^2 + \frac{\lambda}{2}\left(\theta'^T \theta'\right)$$

where *i* indicates the training example, *y* is the concentration, *x* is the features and θ is the parameters. θ' is defined to be the parameter vector θ with the parameter corresponding to the offset term removed. I use θ' rather than θ so that the offset term is not penalized for being large. The regularization parameter λ was chosen to minimize the test data error.

Sensitivity Variation

The sensitivity of the sensors may vary for a number of reasons. The sensitivity may drift as the sensors age. I the manufacturing process, the thickness of the chemical sensitive layers will have some variation, and it would be advantageous not to have to calibrate every sensor individually. To study this problem, I varied the sensitivity of the sensors and calculated the effect on classification performance. The sensitivities of sensors 1-3 were multiplied by a factor drawn from a normal distribution with mean 1 and variance σ^2. σ was varied from 0 to 0.5 to simulate different severities of sensor variation. The system was trained on the uncorrupted data using the parameters obtained by optimized test data performance. For each value of σ, 2000 experiments were simulated. The results are shown in Figure 2. The sensitivity of sensor 4 was not varied since it has no chemical-sensitive layer deposited on the surface and thus should show very little aging or manufacturing-related sensitivity changes.

Results

Classification

Table 2 shows the training and test performance of the algorithms tested. The SVM worked well with both a linear kernel and Gaussian kernel. Table 3 shows the performance using the equilibrium frequency shifts. The algorithms were tested using leave-three-out cross-validation.

Table 2 – Performance of classification algorithms

Model	Features	Training Accuracy	Test Accuracy
Logistic Regression	Full Response	86.3%	86.2%
SVM, Linear Kernel	Full Response	96.8%	97.5%
SVM, Gaussian Kernel	Full Response	96.8%	96.8%
Neural Network	Full Response	99.7%	100%
Logistic Regression	Final Response	84.7%	82.9%
SVM, Linear Kernel	Final Response	92.0%	87.0%
SVM, Gaussian Kernel	Final Response	97.6%	95.9%
Neural Network	Final Response	99.2%	99.2%
Logistic Regression	Final Response and Squares	83.7%	84.5%
SVM, Linear Kernel	Final Response and Squares	93.7%	93.5%

Concentration Determination

Table 3 shows the results from for concentration determination. Testing was done using leave-one-outcross-validation.

Table 3 – Performance of linear regression for concentration determination

Chemical	RMS Training Error	RMS Test Error
Methane	10.7%	15.7%
Carbon Dioxide	5.9%	9.4%
Water	1.3%	6.2%
Ethanol	1.5%	4.5%
Acetone	2.1%	5.7%
Ethyl Acetate	2.5%	5.2%

Robustness to Sensitivity Variation

The results of the sensitivity variation experiments are shown in Figure 2. Of the algorithms that used the full response, the Gaussian kernel SVM showed the most robustness to sensitivity variations. Of the algorithms using the final responses, the linear kernel SVM with the squared responses included as features showed the least reduction in sensitivity. However, the Gaussian kernel SVM showed higher performance with uncorrupted data. So, the Gaussian kernel SVM gives the best performance for small amounts of sensitivity variation while the linear SVM

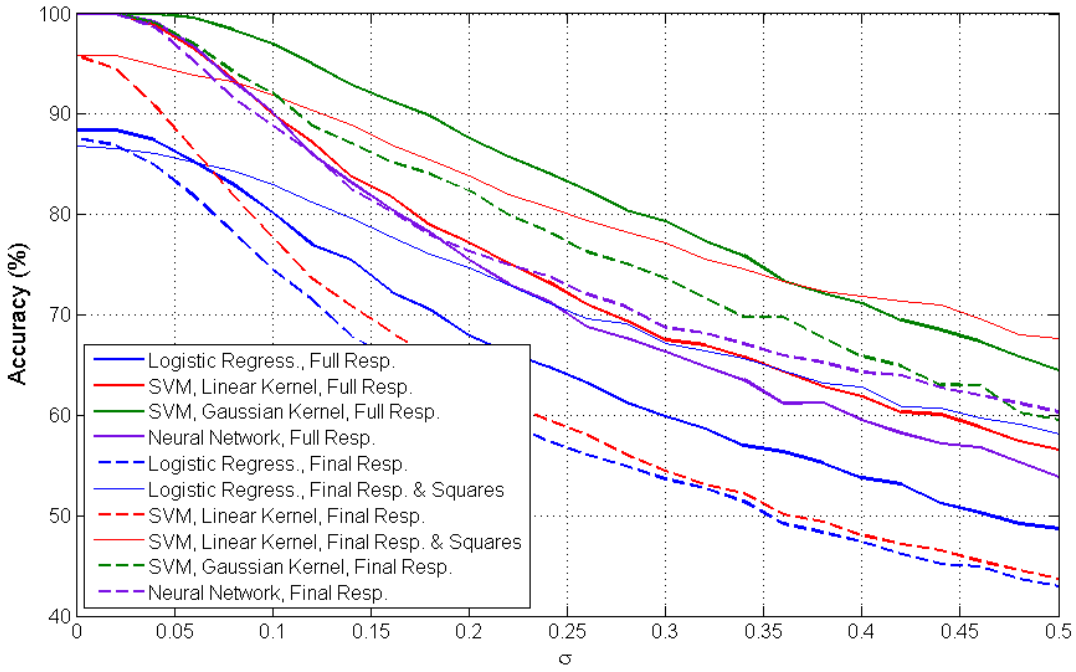

Figure 2 - Performance of chemical classification algorithms as sensitivity variation increases

with squared features gives the best performance for larger sensitivity variation.

Discussion

When the full sensor responses were used, excellent classification performance was obtained from the linear kernel SVM, Gaussian kernel SVM and neural network. When only the final responses were used, good performance was obtained from the Gaussian kernel SVM, the linear kernel SVM with the squared responses included, and the neural network. Performance was better when the full response was used. This demonstrates that incorporating the extra information from the time-dependence of the response can improve performance of the sensor network.

Different models showed different robustness to sensor variation. Although neural networks performed excellently on the uncorrupted data, they generally showed worse performance with aging than the SVM models. It makes sense that the SVMs are less susceptible to aging since they are optimal margin classifiers and thus place their decision boundaries as far as possible from the data points. For this data, it is also important that the model by non-linear. The linear-kernel SVMs did not perform well in the presence of sensor variation even though they performed well on the uncorrupted data. This is likely due to the fact that the sensor responses are mildly non-linear with respect to concentration. When the sensors are uncorrupted, the linear models can find decision boundaries that separate the data since the nonlinearity is not too large, but the margin is not as large as it can be with a non-linear model.

In general, the concentration was determined reasonably accurately, with the worst performance being for methane and carbon dioxide. This is not surprising since these chemicals produced relatively small responses from the sensors at the concentrations tested.

While these results demonstrate that it is important to consider sensor degradation, they do not necessarily mean that the algorithms that were most resistant to sensitivity variation here will be the most resistance to sensitivity variation in all circumstances. It may be that if more training data is available, more complex models like the neural network perform better while with less training data simpler models such as logistic regression may be superior.

Determining the concentration of the chemicals once they were identified worked well. The physics of the sensors means that the responses will increase smoothly and monotonically as the concentration increases. Examining the data, it could be seen that the data did not look exactly like this and that there was some noise due to sensor error. So, the current limiting factor on the concentration measurement

accuracy is most likely the sensors and not the regression algorithm.

Conclusions

I successfully demonstrated the ability of several machine learning algorithms to classify chemicals. I showed how using the time-dependent response can provide improved performance compared to using just the final, equilibrium responses. Additionally, I illustrated that it is important to consider drift in the data when choosing machine learning models, if drift may be present. Some models may perform well with perfect data, but perform much worse when the data is flawed. For this problem, neural networks appear very promising based on their performance on uncorrupted data, but they are much less robust to sensor variation than some of the SVM models. This fact, combined with the fact that neural networks are trickier and more time-consuming to optimize, makes SVMs the better choice for this problem.

Future

In this work, I was able to find models that were very effective at chemical classification and concentration determination. To improve the performance of the system, the next step will be to improve the sensors and collect more data.

The polymer coatings used as chemical-sensitive layers for the devices in this work were not optimized for this chemical identification task. In future, I plan to try a larger set of coatings, including ones tailored to strongly absorb specific chemicals of interest. It is not possible to design perfectly selective coatings, but selectivity can be improved by engineering the coatings for the desired task [5]. Feature selection could be used to find the optimal set of sensors for a given task.

In this work, the sensors only had to deal with single chemicals. In future, I plan to test the system on mixtures of chemicals. The difficulty of the problem will depend on whether the response to multiple chemicals is equal to the sum of the responses to the chemicals individually, or if there are interaction effects.

Another type of application I hope to try is using a sensor array to distinguish complex scents. For example, chemical sensors have been used to distinguish between people with lung cancer and healthy controls [6].

References

1. K. K. Park, H. J> Lee, G. G. Yaralioglu, A. S. Ergun, O. Oralkan, M. Kupnik, C. F. Quate, B. T. Khuri-Yakub, T. Braun, J.-P. Ramseyer, H. P. Lang, M. Hegner, Ch. Gerber, and J. K. Ginzewski, "Capacitive micromachined ultrasonic transducers for chemical detection in nitrogen," *Appl. Phys. Lett.*, vol. 91, pp. 094102, 2007.

2. H. J. Lee, K. K. Park, M. Kupnik, O. Oralkan, and B. T. Khuri-Yakub, "Chemical Vapor Detection Using a Capacitive Micromachined Ultrasonic Transducer," *Anal. Chem.*, vol. 83, no. 24, pp. 9314-9320, 2011.

3. R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research* vol. 9, 2008.

4. C.-C. Chang and C.-J. Lin. "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2 , no. 3, 2011.

5. H. J. Lee, K. K. Park, M. Kupnik, N. A. Melosh, and B. T Khuri-Yakub, "Mesoporous Thin-Film on Highly-Sensitive Resonant Chemical Sensor for Relative Humidity and CO_2 Detection," *Anal. Chem.*, vol. 84, no. 7, pp. 3063-3066, 2012.

6. G. Peng, U. Tisch, O. Adams, M. Hakim, N. Shehada, Y. Y. Brova, S. Billan, R. Abdah-Bortnyak, A. Kuten, and H. Haick, "Diagnosing lung cancer in exhaled breath using gold nanoparticles," *Nature Nanotechnology*, vol. 4, no. 10, pp. 669-673, 2009.

7. C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006, pp. 144-145.

8. Q. Stedman, K. K. Park, and B. T. Khuri-Yakub, "Distinguishing Chemicals Using CMUT Chemical Sensor Array and Artifical Neural Networks," *Presented at the IEEE Ultrasonics Symposium*, Chicago, Illinois, Sept. 3-6, 2014.