Classification of Alcohols in Cosmetic Production

Nang Nandar Htun

University of Computer Studies (Pinlon)

yinnoblewadi@gmail.com

Abstract— ***Different chemical structures of alcohols are chemical founds in cosmetic products such as lotions, shampoos, and conditioners. The correct usage of alcohol in cosmetic production is the important factor for product quality. If experts know the correct name and compound of the alcohol in short time, they can easily applied this alcohol in good quality cosmetic production. Machine learning algorithms are less costly way to classify alcohols according their attributes obtained by QCM sensors so that an informed decision about cosmetic products can be made automatically. In this paper, Support Vector Machine (SVM) learning algorithm is performed to classify type of alcohols based on their features obtained from QCM sensor for cosmetic production development. The QCM sensor dataset for alcohol classification is used in this paper and the five labels of alcohols are 1-octanol alcohols, 1-propanol alcohols, 2-butanol alcohols, 2-propanol alcohols and 1-isobutanol alcohols.***

Keywords—alcohols classification, machine learning algorithms, QCM sensors, Support Vector Machine

I. INTRODUCTION

From the very old times, alcohol is accessible to people. Biologically, alcohol is ethanol or ethyl alcohol. In addition to industrial purposes, it has been and is used for multiple reasons in daily life. Many individuals have distinct views about their usage. But in current culture, other than usage, its purposes are essential. Alcohol is suitable not only drink but also in other material production such as cooking and cosmetic production.

Quartz Crystal Microbalance (QCM) sensors, are widely used in real world applications like; liquid bio sensing, thin film monitoring and electrochemistry. The core of the QCM is the nestled piezoelectric AT-cut quartz crystal between a couple of electrodes. The quartz crystal begins to oscillate at its resonant frequency owing to the electrostatic impact when the electrodes are attached to an oscillator and an AC voltage is introduced over the electrodes. Due to the high quality of the oscillation this oscillation is usually very stable. The QCM sensors data and machine learning are successfully applied in Artificial Intelligent (AI) and other automatic system.

Machine learning is the powerful tool for automatic system. Most of machine learning algorithms are widely applied in automatic system and decision making system. In this paper, different types of machine learning algorithms are applied to successfully classify the different type of alcohol for cosmetic production. The main focus of this paper is to classify the different type of alcohol by using features from QCM sensors. The features from QCM sensors are the structural reaction of structural chemical compounds and the cost of QCM sensors are low [1]. According to these reasons, the combination of QCM sensor features and machine learning techniques are decided to apply in alcohol structure classification.

In Detecting Rip Channels from Images, various kinds of machine learning algorithms such as Support Vector Machine, Convolutional Neural Network, Viola-Jones Method, and Meta-Classifier are used. The comparison of these algorithms will assist guide scientists in selecting a suitable model for rip channel identification by monitoring error rates and false positive counts of these algorithms in experiments [2]. Logistics Regression, Liner SVM and Multinomial Naïve Bayes algorithms are applied in sentiment analysis for Public Opinion on Dockless Bike Sharing. With distinct machine learning algorithms, distinct types of subjects and characteristics are implemented to discover the best appropriate model for sentiment analysis by evaluating accuracy of n-gram TF-IDF and amount of characteristics in the experiment[5]. The issue of XSS (cross-site scripting) attacks in web applications is becoming increasingly severe with the fast rise in web applications. A comparison of machine learning algorithms for detecting XSS attacks was conducted to solve this issue by assessing detailed observation of their designs. Different models of K-means, Naïve Bayes, Decision Tree, Association Rules, Hidden Markov model, Support Vector Machine and deep learning are studies for XSS detection [8].

In recent years, Support Vector Machine (SVM) is the novel classifier for drug discovery, high resolution remote sensing images classification, network intrusion detection and brain tumor classification [3, 4, 6, 7]. Linear Discriminant Analysis (LDA) is widely used in speech recognition for continuous vocabulary, face recognition, emotion recognition and Adaptive Radar-Based Human Activity Recognition [9, 10, 11]. K-nearest neighbors (KNN) is the good algorithm for moving object detection, spatial network databases analysis and intrusion detection [12, 13, 14]. Decision Tree is used in predicting employment after moderate to severe traumatic brain injury, breast cancer classification and Predicting service industry performance [15, 16, 17].

In this paper, Section II describes the overview of Classification of Alcohols. Section III identifies how SVM classifier is built and identifies the dataset. Section IV describes experimental results of each classifier and shows the results of the experiments. Section V contains discussion the conclusions.

II. OVERVIEW OF ACOHOL CLASSIFICATION

This section describes the overview of alcohol classification. The input is the small amount of alcohol. The input alcohol passes through the different type of QCM sensors. The features are extracted from the QCM sensors and applied in classification and testing. The trained classifier with features is used for testing and unknown data. The overview of alcohol classification is shown in Figure 1.

In Figure 1, the features of alcohol are obtained by the QCM sensors and then the classification is perform to predict the name of the alcohol used for cosmetic production.

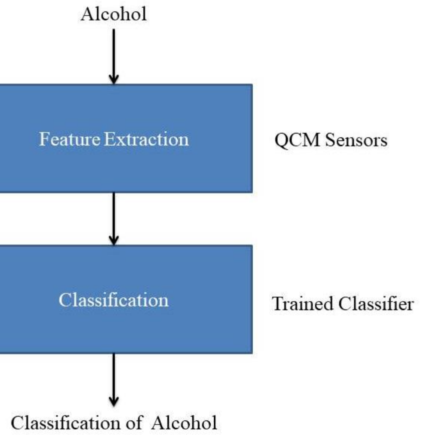
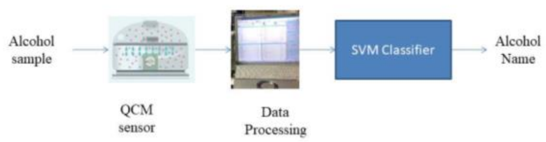


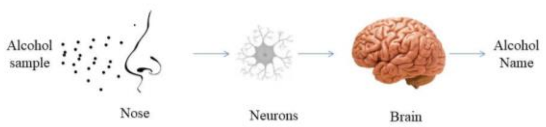
Figure 1: Flow of alcohol classification

A. QCM Sensor

Quartz crystal microbalance (QCM) sensor measures a mass variety per unit region by evaluating a quartz crystal resonator's shift in frequency. The resonance is troubled by adding or removing a tiny mass at the bottom of the acoustic resonator owing to oxide growth / decay or film deposition. The QCM can be used in vacuum, in gas phase and in liquid settings more lately. It is helpful for tracking deposition rates under vacuum in thin film deposition processes. In fluid, the attachment of molecules to surfaces fictionalized with identification locations is extremely efficient. There are also investigations into larger organizations such as viruses or polymers. QCM was also used to explore bio-molecular relationships. The usage of QCM in alcohol classification and figure of QCM sensor is shown in Figure 2a. In Figure 2, classification of alcohol is similar to the human olfactory system as shown in Figure 2 (b).



(a)Usage of QCM sensor in Alcohol Classification



(b) Human Olfactory system

Figure 2: Usage of QCM sensor in Alcohol Classification and Human Olfactory system

B. .Feature Extraction

In five distinct levels, the gas sample is carried through the sensor[1]. These levels are the proportion of air (ml) and gas (ml) as shown in Table 1. Each QCM detector produces ten alcohol gas measurement characteristics.

Table I: Air ratio and Gas ratio of QCM sensors

No.	Air ratio	Gas ratio
1	0.799	0.201
2	0.700	0.300
3.	0.600	0.400
4.	0.501	0.499
5.	0.400	0.600

There are two channels in the sensor. One of these circles forms channel 1, and the other forms channel 2[1]. MIP and MP ratios used in the QCM sensors are shown in Table 2.

Table 2: Different MIP ration and NP ratio of Sensors

Sensor name	MIP ratio	NP ratio
QCM3	1	1
QCM6	1	0
QCM7	1	0.5
QCM10	1	2
QCM12	0	1

III. SUPPORT VECTOR MACHINE (SVM)

Classification is the method of information organization by classifications agreed upon. A thoroughly scheduled classification makes it possible to use and protect critical information more efficiently across the organization and adds to risk management, legal discovery and enforcement procedures. In machine learning, classification may be an individual method before some other significant method to define the mark of an unidentified component or sub-process. If it is a sub-process, it will be easier to access the elements if they have a certain structure. The input of this stage is the extracted feature of sample and the output is the name of alcohol. In this study, Support Vector Machine (SVM) is used for Classification.

SVM's performance is very nice when the information has no clue. Even unstructured and semi-structured data such as text, images and trees work well. SVM's true power is the kernel trick. We can fix any complicated issue with a suitable kernel function. Unlike in neural networks, for local optima, SVM is not fixed. It scales to high-dimensional information comparatively well. In practice, SVM models are generalized, with less risk of over-fitting in SVM. The objective of the SVM classification is to produce hyper plane for finding the best possible separation between two labels. A classifier function f(x) is linear if it can be expressed as:

$$f(x;w,b) = (w,x) + b \qquad (1)$$

where w and b are parameters of the function and (w,x) denotes the inner product of two vectors w and x.

A. Data background and Dataset

This study uses five separate QCM gas sensors and performs five distinct gas measurements (1-octanol,1-propanol,2-butanol,2-propanol, and1-isobutanol) in each of these sensors. One channel includes molecular polymers (MIP), while the other channel includes nanoparticles (NP). Different QCM detectors are accomplished using separate MIP and NP ratios. One single experiment lasted 120 min [1] in any QCM sensor. Table 3 shows the amount of records in this experiment.

Table 3: Number of Data from Different QCM Sensors

Sensor name	MIP ratio	NP ratio	No. of Data
QCM3	1	1	25
QCM6	1	0	25
QCM7	1	0.5	25
QCM10	1	2	25
QCM12	0	1	25
Total Number of Data			125

IV. Experimental Results

In this section, the experiment is performed on the alcohols dataset. The classifier performance is measured in term of accuracy and confusion matrix. The validation structure of experiment is 4 fold-cross validations.

A. Four fold-cross Validation

The dataset is divided into 4 groups and the validation is performed for 4 times. For each validation time, three groups are used as training and the rest one is used as testing, and then measure accuracy. The average classification accuracy and confusing matrix are exported over this 4 fold-cross validation as shown in Figure 3.

Validation	Dataset				
1	Test	Train	Train	Train	Accuracy1
2	Train	Test	Train	Train	Accuracy2
3	Train	Train	Test	Train	Accuracy3
4	Train	Train	Train	Test	Accuracy4

Figure 3: Four Fold-cross validations

B. Classification Accuracy

Classification accuracy is the percentage of correct rate according to the total number of testing units. The performance of classifier is evaluated by calculating of classification accuracy:

$$\text{Classification Accuracy} = x / N \qquad (2)$$

where x is the number of correct classified data and N is the total number of testing data.

C. Confusion Matrix

The matrix of error is also called the table of contingency [21]. A confusion matrix is a table that is often used to define a classification model's efficiency (or "classifier") on a collection of sample information recognized for the real values. It enables the efficiency of an algorithm to be visualized. The matrix of confusion is displayed by a matrix representing the cases in an expected class in each row, whereas each column is displayed in an real class. One of the benefits of using this performance evaluation instrument is that if the model confuses two categories, the data mining analyzer can readily see.

The chart also demonstrates the classifier's precision as the proportion of properly categorized motifs in a class separated by that class's complete amount of motifs. The classifier's median classification precision is also assessed using the confusion matrix.

		predicted	
		Positive	Negative
actual examples	Positive	a TN - True Negative correct rejections	b FP - False Positive false alarms type I error
	Negative	c FN - False Negative misses, type II error overlooked danger	d TP - True Positive hits

Figure4: Confusion Matrix

D. Classification Setting

In our experiment, we use Classification Learner tool in MATLAB. The original data type is csv file type and needed to convert into mat file. This section describes the detail setting for Support Vector Machine (SVM) as shown in table 4.

Table 4: Classification setting for Support Vector Machine (SVM)

No.	Machine Learning Method	Classifier Setting
1.	Support Vector Machine (SVM)	**Preset**: Fine Gaussian SVM **Kernel function**: Gaussian **Kernel scale**: 0.79 **Box Constraint level**: 1 **Multiclass method**: One-vs-One

E. Results and Discussion

The classification accuracy is measure in the structure of four fold-cross validation as shown in Figure 5. Support Vector Machine (SVM) takes 12.539 sec for training time and its prediction speed has 910 obs/sec. The confusion matrices of SVM classifier is shown in Figure 5. In figure 5, 1,2,3,4 and 5 are 1-octanol alcohols, 1-propanol alcohols, 2-butanol alcohols, 2-propanol alcohols and 1-isobutanol alcohols respectively. In SVM, label 1 and label 3 are misclassified as label 5. Although label 1 and label 3 are misclassified; the classification accuracy of SVM classifier is acceptable. The classification accuracy of SVM reaches up to 98.4 %.

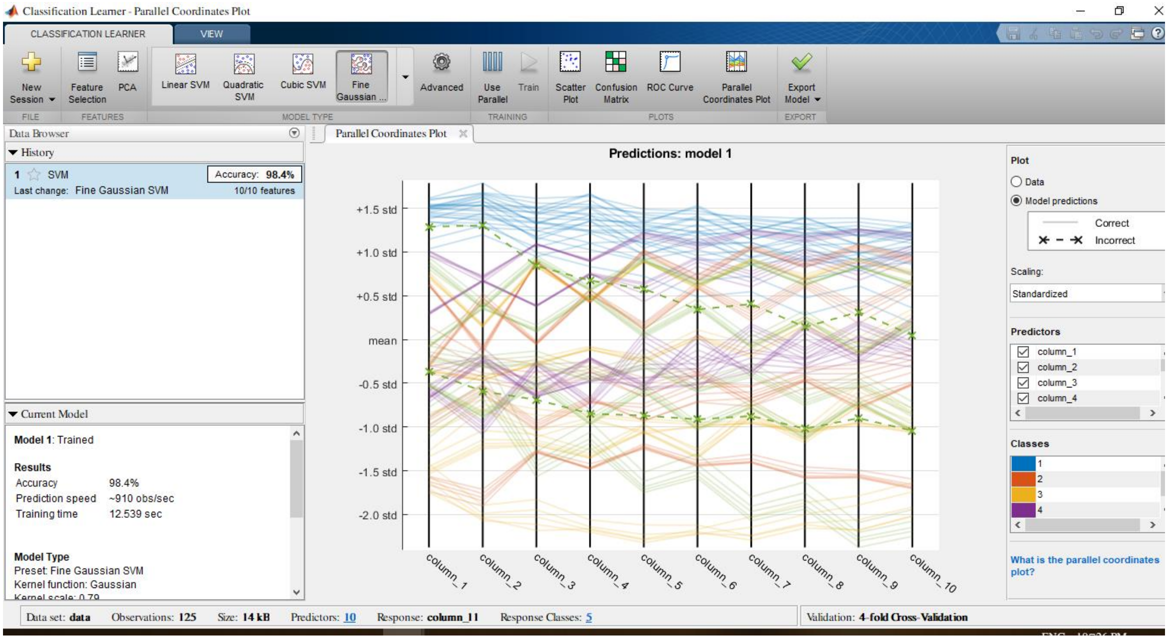
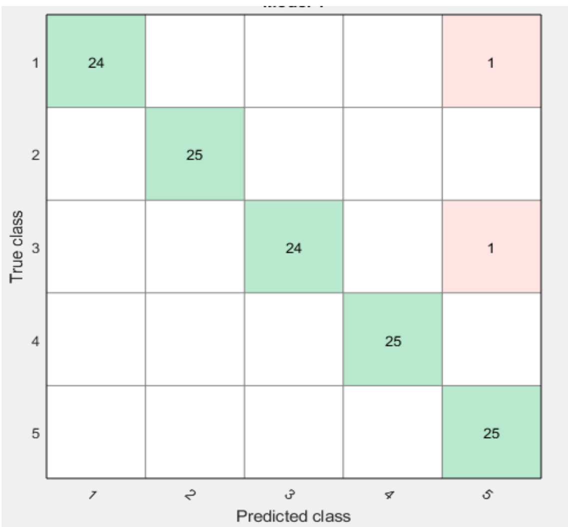


Figure 5: Classification Accuracy over 4 fold-cross validation



Labels

1= 1-octanol alcohols

2= 1-propanol alcohols

3= 2-butanol alcohols

4= 2-propanol alcohols

5=1-isobutanol alcohols

Figure 5: Confusion matrix of SVM for alcohol classification

V. Conclusion

This paper primarily presents Support Vector Machine (SVM) for machine learning in alcohol classification. While classifying alcohol, SVM machine learning algorithm has distinct implementation situations, and has the acceptable classification accuracy. The aim of this paper is to classify the alcohol structure by combination of QCM sensors and SVM classifier to assist in cosmetic production. In addition, the machine learning algorithm implemented in this paper is primarily used in alcohol classification and does not investigate how to obtain alcohol features more sensibly and efficiently, so it is also a significant research path to study how to correctly and efficiently extract alcohol feature for the machine learning algorithm.

References

[1] Adak, M.F., Lieberzeit, P., Jarujamrus, P. and Yumusak, N., 2019. Classification of alcohols obtained by QCM sensors with different characteristics using ABC based neural network. Engineering Science and Technology, an International Journal.

[2] Maryan, C., Hoque, M.T., Michael, C., Ioup, E. and Abdelguerfi, M., 2019. Machine learning applications in detecting rip channels from images. Applied Soft Computing, 78, pp.84-93.

[3] Maltarollo, V.G., Kronenberger, T., Espinoza, G.Z., Oliveira, P.R. and Honorio, K.M., 2019. Advances with support vector machines for novel drug discovery. Expert opinion on drug discovery, 14(1), pp.23-33.

[4] Sun, G., Rong, X., Zhang, A., Huang, H., Rong, J. and Zhang, X., 2019. Multi-Scale Mahalanobis Kernel-Based Support Vector Machine for Classification of High-Resolution Remote Sensing Images. Cognitive Computation, pp.1-8.

[5] Rahim Taleqani, A., Hough, J. and Nygard, K.E., 2019. Public opinion on dockless bike sharing: A machine learning approach. Transportation Research Record, p.0361198119838982.

[6] Chowdhury, M.N., 2019. A support vector machine cost function in simulated annealing for network intrusion detection.

[7] Kang, C., Huo, Y., Xin, L., Tian, B. and Yu, B., 2019. Feature selection and tumor classification for microarray data using relaxed Lasso and generalized multi-class support vector machine. Journal of theoretical biology, 463, pp.77-91.

[8] Chen, X., Li, M., Jiang, Y. and Sun, Y., 2019, July. A Comparison of Machine Learning Algorithms for Detecting XSS Attacks. In International Conference on Artificial Intelligence and Security (pp. 214-224). Springer, Cham.

[9] Haeb-Umbach, R. and Ney, H., 1992, March. Linear discriminant analysis for improved large vocabulary continuous speech recognition. In [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 1, pp. 13-16). IEEE.

[10] Lu, J., Plataniotis, K.N. and Venetsanopoulos, A.N., 2003. Face recognition using kernel direct discriminant analysis algorithms. IEEE Transactions on Neural Networks, 14(1), pp.117-126.

[11] Chen, D.W., Miao, R., Yang, W.Q., Liang, Y., Chen, H.H., Huang, L., Deng, C.J. and Han, N., 2019. A Feature Extraction Method Based on Differential Entropy and Linear Discriminant Analysis for Emotion Recognition. Sensors, 19(7), p.1631.

[12] Zhang, M., Rishe, N., Weitong, L.I.U., Lazarre, J. and Li, T., Florida International University, 2019. Efficient progressive continuous k-nearest neighbor query algorithm for moving objects with a tree-like index. U.S. Patent Application 10/331,753.

[13] Kolahdouzan, M. and Shahabi, C., 2004, August. Voronoi-based k nearest neighbor search for spatial network databases. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30 (pp. 840-851). VLDB Endowment.

[14] Liao, Y. and Vemuri, V.R., 2002. Use of k-nearest neighbor classifier for intrusion detection. Computers & security, 21(5), pp.439-448.

[15] Using Decision Tree Methodology to Predict Employment After Moderate to Severe Traumatic Brain Injury

[16] Quist, J., Mirza, H., Cheang, M.C., Telli, M.L., O'Shaughnessy, J.A., Lord, C.J., Tutt, A.N. and Grigoriadis, A., 2019. A four-gene decision tree signature classification of triple-negative breast cancer: Implications for targeted therapeutics. Molecular cancer therapeutics, 18(1), pp.204-212.

[17] Yeo, B. and Grant, D., 2018. Predicting service industry performance using decision tree analysis. International Journal of Information Management, 38(1), pp.288-300.

[18] R. Kohavi, F. Provost: Glossary of terms, Machine Learning, Vol. 30, No. 2/3, 1998, pp. 271-274.