Application of Conformal Predictors to Tea Classification Based on Electronic Nose

Ilia Nouretdinov[1], Guang Li[2], Alexander Gammerman[1], and Zhiyuan Luo[1]

[1] Computer Learning Research Centre, Royal Holloway, University of London
Egham Hill, Egham, Surrey TW20 0EX, UK
zhiyuan@cs.rhul.ac.uk
[2] Department of Control Science and Engineering, Zhejiang University,
38 ZheDa Road, HangZhou 310027, Zhejiang, P.R. China

Abstract. In this paper, we present an investigation into the performance of conformal predictors for discriminating the aroma of different types of tea using an electronic nose system based on gas sensors. We propose a new non-conformity measure for the implementation of conformal predictors based on Support Vector Machine for multi-class classification problems. The experimental results have shown the good performance of the implemented conformal predictors.

Keywords: Tea Classification; Electronic Nose; Conformal Predictors; Support Vector Machines; Pattern Recognition.

1 Introduction

Tea is by far the most popular drink consumed in Britain today, with over 165,000,000 cups being drunk in the UK every single day of the year [1]. The major quality attributes of tea are flavour, aroma, colour and strength. Aroma means smell of the tea and is considered as the most important of the attributes. However, characterisation of aroma of tea has been a challenge for tea scientists for long. Tea flavour is traditionally measured through the use of a combination of conventional analytical instrumentation and human or ganoleptic profiling panels. These methods are expensive in terms of time and labour and also inaccurate because of a lack of either sensitivity or quantitative information [5].

Electronic nose has been successfully used for tea classification and grading [2,5,6,12]. In general, an electronic nose is a smart instrument that is designed to detect and discriminate among complex odours using an array of sensors. The array of sensors is exposed to volatile odour vapour through suitable odour handling and delivery system that ensures constant exposure rate to each of the sensors. The response signals of these sensors are recorded and processed for analysis and classification. For tea classification, we want to discriminate the aroma of different types of tea.

Different data analysis and machine learning methods have been tried to analyse and classify the signals from the sensor array of the electronic nose system. For example, these signals were processed using Principal Components Analysis (PCA), Fuzzy C Means algorithm (FCM), Self-Organizing Map (SOM) method along with a Radial Basis Function network (RBF), Probabilistic Neural Network classifier, Back-Propagation network and Olfactory model [3,5,6,8,11,12]. Good performance has been reported in

the literature. However, the learning techniques used can only provide bare predictions, i.e. algorithms predicting labels for new examples without saying how reliable these predictions are. The reliability of a method is often given by measuring general accuracy across an independent test set. Confidence estimation is a well-studied area of both parametric and non-parametric statistics; however, usually only low-dimensional problems are considered.

This paper describes an investigation into the performance of conformal predictors for discriminating the aroma of different types of tea using an electronic nose system. In particular, we study a recently developed theory of algorithmic learning, namely conformal predictors described in [10]. One of the major advantages of these algorithms is that they can be used for solving high-dimensional problems without requiring any parametric statistical assumptions about the source of data (unlike traditional statistical techniques). The only assumption made is the i.i.d. assumption (the examples are generated from the same probability distribution independently of each other). Another advantage of the method is that it also allows to make estimation of confidence in the classification of individual examples. The remainder of this paper is arranged as follows. We briefly describe conformal predictors and present a new non-conformity measure for the implementation of conformal predictors based on Support Vector Machine for multi-class classification problems in the next section. Section 3 discusses the experiments and results of applying the implemented conformal predictors to tea classification. Finally, Section 4 concludes the paper.

2 Conformal Prediction

In machine learning problems, we usually work with examples which are pairs $z_i = (x_i, y_i)$. $x_i \in \mathbf{X}$ is an object, normally represented as a vector and its components (dimensions) are called attributes. $y_i \in \mathbf{Y}$ is its label, which may be real-valued, or taking one of several possible values [4]. For the tea classification problem using electronic nose system, x_i is a vector of response signals from the sensor array and y_i is the finite types of tea.

Let us assume that we are given a training set of l examples $\{(x_1, y_1), \ldots, (x_l, y_l)\}$. Our goal is to predict the classifications for testing examples. We make only one assumption (i.i.d.) about the examples: all the examples have been generated independently by some fixed but unknown stochastic mechanism. Suppose that for a new example x_{l+1} and its corresponding real label y_{l+1} is unknown, we are interesting in predicting its label. The basic idea behind conformal predictors is to try every possible label $y \in \mathbf{Y}$ as candidate for the new example and see how well the resulting sequence $\{(x_1, y_1), \ldots, (x_l, y_l), (x_{l+1}, y)\}$ conforms the i.i.d. assumption. If it does, we will say it is "random". The ideal case is where all $y \in \mathbf{Y}$ but one lead to sequences that are not random and y can be considered as the prediction for x_{l+1}. The general description of the conformal predictors is given in Algorithm 1 [10].

In conformal predictors, checking whether a possible new label violates i.i.d. assumption is done by a test for randomness based on a *non-conformity (or strangeness) measure* $\alpha = A(z_{l+1}, \{z_1, \ldots, z_{l+1}\})$. This is a measure of disagreement or non-conformity between a finite set (subset of $\mathbf{X} \times \mathbf{Y}$) and its element. Specific form of

this function depends on a particular machine learning algorithm to be used and can be determined for many well-known machine learning algorithm. In this paper, we investigate and propose a non-conformity function for multi-class Support Vector Machine.

Algorithm 1. Conformal Predictor

Input: example sequence $\{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$
Input: new example x_{l+1}
Input(optional): confidence level γ
Input: non-conformity measure A
for $y \in Y$ **do**
$\quad z_{l+1} = (x_{l+1}, y)$
$\quad$**for** j in $1, 2, \ldots, l+1$ **do**
$\quad\quad \alpha_j = A(z_j, \{z_1, \ldots, z_l, z_{l+1}\})$
$\quad$**end for**
$\quad p(y) = \frac{\#\{j=1,\ldots,l+1 : \alpha_j \geq \alpha_{l+1}\}}{l+1}$
end for
Output(optional): predictive set $R_{l+1}^{\gamma} = \{y : p(y) \geq 1 - \gamma\}$
Output: forced prediction $\hat{y}_{l+1} = \arg\max_y \{p(y)\}$
Output: confidence $conf(\hat{y}_{l+1}) = 1 - \max_{y \neq \hat{y}_{l+1}} \{p(y)\}$
Output: credibility $cred(\hat{y}_{l+1} = \arg\max_y \{p(y)\}$

The conformal predictor calculates $p(y)$ which is a *p-value* associated with a hypothetical completion $y_{l+1} = y$ for the test example. Using these *p-values*, predictions made by conformal predictors can be presented in following two ways:

- Either the user inputs a required degree (level) of certainty and the algorithm outputs a predictive set: a list of classifications that meet this confidence requirement.
- Or the algorithm outputs so called *forced prediction*: an individual predictions together with its confidence, which is the minimal level of certainty at which the output is certain (predictive set consists of only one label) and credibility, which indicates whether the test object is representative of the training set, in order to compare with the other conventional learning methods.

The common way of presenting the prediction results is to choose a significance level ("degree of certainty") $\gamma < 1$ and output the (γ)-*predictive set (region)* containing all labels with p-value equal or greater than $1 - \gamma$:

$$R_{l+1}^{\gamma} = \{y : p(y) \geq 1 - \gamma\} \tag{1}$$

Prediction errors can occur when the prediction set fails to contain the true label. If the size of the predictive set R_{l+1}^{γ} is 1 or 0, then this prediction is *certain*, otherwise it is uncertain and we have multiple predictions. In the context of conformal predictors, uncertain predictions are not errors, but these are indications that the amount of information is not sufficient to make a certain decision at the selected level. Naturally, the higher the confidence level is, the more multiple predictions will appear. If non-conformity measure is specified adequately for the problem at hand, the percentage of

uncertain predictions should not be too high. It has been proved that the conformal predictors have validity property [10,7] in online mode where the examples are presented one by one. Each time, the predictors take an object to predict its label and then receive feedback of the true label. The validity property implies that the successive predictions will be right 1- ϵ of the time for a given error probability ϵ, even though they are based on an accumulating dataset rather than on independent dataset.

It is important to note that the prediction region of a conformal predictor in online mode is valid for any used non-conformity measure [10]. However, the quality of prediction depends on the choice of this measure. If it is inappropriate, the confidence in a prediction will be still valid but never high. Thus a non-conformity measure for conformal predictor is usually based on an underlying method of prediction (such as nearest neighbours and Support Vector Machine) which is appropriate enough for the data set.

In the case of finite label space, there is an alternative way to represent the output: single ("forced") prediction and some measure of *confidence* in this prediction. It does not require the confidence level γ as input. The single prediction is selected by largest p-value. The maximal confidence level (degree of certainty) at which the prediction is certain, is called "confidence". It can be calculated as 1 minus the second largest p-value.

2.1 Support Vector Machine

Conformal predictors are a generic framework for making prediction and virtually any classification or regression algorithm can be transformed into a conformal predictor [10]. In this section, we consider the implementation of conformal predictors on the top of Support Vector Machine. Support Vector Machines (SVM) are effective learning systems based on statistical learning theory and have been applied successfully to many real problems [9]. For pattern recognition, SVM finds the hyperplane that separates the data set $(x_1, y_1), ..., (x_l, y_l), x_i \in \mathbf{R}^d, y_i \in \{-1, +1\}$ with maximal margin. If the data are not linearly separable one can introduce slack variables (ζ_i) in the optimisation,

$$\min \frac{1}{2} \parallel \mathbf{w} \parallel + C \sum_{i=1}^{l} \zeta_i \tag{2}$$

under the constraints $y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] \geq 1 - \zeta_i$ and $\zeta_i \geq 0$, where $\mathbf{w}$ is a normal vector that is perpendicular to the hyperplane, $\frac{\parallel \mathbf{w} \parallel}{b}$ determines the offset of the hyperplane from the origin along the normal vector $\mathbf{w}$ and C is a constant chosen a priori. This optimisation problem can then be translated into its corresponding dual problem [9]

$$\max \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \tag{3}$$

under the constraints, $0 \leq \alpha_i \leq C$, $\sum_{i=1}^{l} \alpha_i y_i = 0$, where K is a kernel function defined by $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$. The Lagrangian α_i is assigned to each example in the dual setting of SVM and it is independent on their order. It is positive if the example is a Support Vector (SV), and 0 if it is non-SV. Therefore, these Lagrangian values can be used as non-conformity measure for SVM.

SVMs are inherently two-class classifiers. The common approach to dealing with multi-class SVM is to reduce the multi-class problem into multiple binary classification problems. Then each of the problems can be solved by a binary classifier. In this paper, we propose and use the following non-conformity measure as a tool to link two-class problems together for multi-class classification problems, see Algorithm 2. For example, if we have 4 classes: A, B, C and D, then a binary class non-conformity measure can be applied to each of possible class pairs: A-B, A-C, A-D, B-C, B-D and C-D. When the class pair A-B is considered, non-conformity values are only assigned to those examples from the class A and class B and the examples from other classes are ignored. The same procedure is applied to the other class pairs. Having considered all the possible class pairs, each example now has 3 different non-conformity values. The maximal of them is considered as the final non-conformity measure of this example.

Algorithm 2. Non-conformity measure for multi-class SVM (for $q > 2$ classes)

> **Input:** Data set $\{z_1 = (x_1, y_1), \ldots, z_{l+1} = (x_{l+1}, y_{l+1})\}$, where $y_i \in Y = \{1, 2, \ldots, q\}$
> **Input:** z_j (one of examples above)
> **for** $r = 1, \ldots, q - 1$ **do**
>   **for** $s = r + 1, \ldots, q$ **do**
>   Apply dual form SVM for examples from classes r and s only
>   Extract Lagrange multipliers $\alpha_j^{SVM(r:s)}$ for j such that $y_j \in \{r, s\}$
> $\alpha_j^{SVM(r:s)} = 0$ for j such that $y_j \notin \{r, s\}$
>   **end for**
> **end for**
> $\alpha_j := \max_{r,s} \alpha_j^{SVM(r:s)}$

---

3   Experiments and Results

Metal Oxide Semiconductor (MOS) sensors are commonly used in electronic nose applications for their convenience in operating and steadiness in features. The MOS gas sensors obtained from Figaro Engineering Inc., Japan (http://www.figaro.co.jp) were used in the experiments. In particular, an array of seven gas sensors: TGS2610, TGS2611, TGS800, TGS813, TGS822, TGS826 and TGS880 were chosen for the tea classification experiments. For illustration purposes, four types of tea which are ready available from the market: Japanese Green Tea (JG), Chinese Green Tea (CG), Indian Black Tea (IB) and Chinese Black Tea (CB) were used and we label them as class 1 to 4, respectively.

Each tea sample is heated before data acquirement. There are 68 data samples in total obtained in the experiments (15 for Indian Black Tea, 19 for Chinese Black Tea, 17 for Chinese Green Tea and 17 for Japanese Green Tea). Each data sample is described by 7 features obtained from those 7 gas sensors of the electronic nose system. A typical sensor response is shown in Figure 1. Originally each of these 7 features was represented as time series and we simply take its maximum value in time for the classification problem.
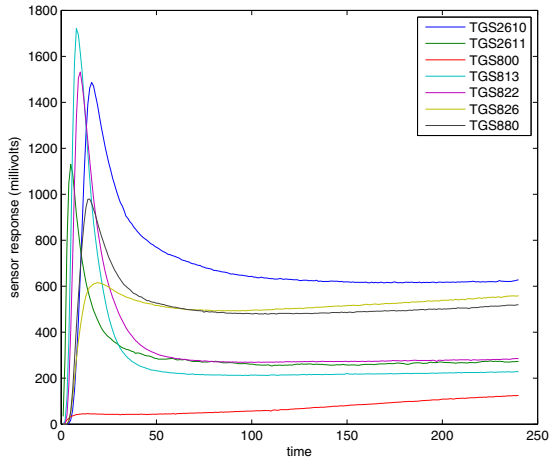
Fig. 1. An example of recorded sensor response

Table 1. Examples of individual predictions for leave-one-out cross validation

No.	T	p(1)	p(2)	p(3)	p(4)	P	Conf	Cred	No.	T	p(1)	p(2)	p(3)	p(4)	P	Conf	Cred
..	..	...	...	...	...	..	...	...	..	..	...	...	...	...	..	...	...
6	2	0.056	0.118	0.063	0.05	2	0.937	0.118	40	1	1	0.056	0.063	0.05	1	0.937	1
7	4	0.222	0.056	0.063	0.053	1	0.937	0.222	41	4	0.056	0.056	0.063	1	4	0.937	1
8	1	1	0.056	0.063	0.05	1	0.937	1	42	4	0.056	0.056	0.063	1	4	0.937	1
9	4	0.167	0.111	0.188	0.316	4	0.812	0.316	43	3	0.056	0.056	1	0.05	3	0.944	1
..	..	...	...	...	...	..	...	...	..	..	...	...	...	...	..	...	...

A 4-class SVM conformal predictor as described above (see Algorithm 2) has been implemented and applied to the tea classification dataset in leave-one-out mode. Table 1 shows example individual predictions with sample data presented in random order, where "T" is the true label, p(1) ... p(4) are calculated *p-values* for different labels, "P" is the forced prediction, "Conf" and "Cred" are the confidence and credibility associated with the forced prediction, respectively. For example, the true label is 4 and the forced prediction is 1 for the 7th sample data. The calculated confidence for the prediction is 0.937 and credibility is 0.222. High (i.e. close to 100%) confidence means that all labels except the predicted one are unlikely. Low credibility means that the whole situation is suspect; perhaps the test example is very different from all examples in the training set. The overall accuracy of forced prediction is 89.7%, as 61 of 68 predictions are correct according to the confusion matrix shown in Table 2.

The performance of forced prediction of the conformal predictor is compared with KIII model (a novel bionic neural network), a conventional artificial neural network (back-propagation network) and volunteers [11]. 30 volunteers were invited to try tea classification. All the volunteers were trained to remember the odour of each kind of tea. Then they made tea classification by smelling without seeing. Table 3 shows the comparison of forced prediction of conformal predictors with other methods on the same

Table 2. Confusion matrix for forced prediction

True Label\Forced Prediction	JG Tea (1)	CG Tea (2)	IB Tea (3)	CB Tea (4)
JG Tea (1)	14	1	1	1
CG Tea (2)	1	16	0	0
IB Tea (3)	1	0	14	0
CB Tea (4)	2	0	0	17

Table 3. Performance comparison (accuracy %)

Method	JG Tea (1)	CG Tea (2)	IB Tea (3)	CB Tea (4)	Average
KIII	93.3%	86.7%	93.3%	80%	88.3%
Back-Propagation	80%	100%	66.7%	93.3%	85%
Volunteers	80%	46.7%	83.3%	50%	65%
Conformal Predictor	82.4%	94.1%	93.3%	89.5%	89.7%

data set. Forced predictions made by conformal predictors have similar performance as those of the KIII model and BP networks. The volunteers performed not so well as the electronic nose, because of some physiological and psychological factors [5].

The predictive sets for a specific confidence level γ can be derived from the list of p-values by Equation (1). For the confidence level of 80%, 85%, 90% and 95%, the performance (in terms of error rate and multiple prediction rate) of the implemented conformal predictor is shown in Table 4. It is clear that the predictions made by the conformal predictor are valid and uncertainty rate becomes larger as the confidence level grows. Validity is true in the sense that error rate is smaller than 1 minus confidence level. The property of validity means that the conformal predictors never overrate the accuracy and reliability of their predictions. From the practical point of view, they allow us to control the error rate by selecting a suitable confidence level. For example, at 80% confidence level, the error rate is just over 16% and less 6% of predictions are uncertain. At 90% confidence level, over 23% of predictions become uncertain.

Table 4. Performance of conformal predictor at different confidence level

Confidence Level	Error Rate	Multiple Prediction Rate
80%	16.2%	5.9%
85%	11.8%	13.2%
90%	5.8%	23.5%
95%	0	100%

4 Conclusion

An investigation into application of conformal predictors to tea classification based on an array of gas sensors has been described. A new non-conformity measure for multi-class tea classification problem using SVM has been discussed. Experiments have been

carried out to evaluate the performance of the conformal predictors. Empirical results have showed the good performance of the implemented conformal predictor.

References

1. United Kingdom Tea Council, http://www.tea.co.uk/the-home-of-tea-all-you-need-to-know-about-tea (last accessed: June 2010)
2. Bhattacharyya, N., Bandyopadhya, R., Bhuyan, M., Tudu, B., Ghosh, D., Jana, A.: Electronic Nose for Black Tea Classification and Correlation of Measurement with "Tea Taster" Marks. IEEE Transactions on Instrumentation and Measurement 57(7), 1313–1321 (2008)
3. Borah, S., Hines, E.L., Leeson, M.S., Iliescu, D.D., Bhuyan, M., Gardner, J.W.: Neural network based on electronic nose for classification of tea aroma. Sensing and Instrumentation for Food Quality and Safety 2(1), 7–14 (2008)
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, Chichester (2001)
5. Dutta, R., Kashwan, K.R., Bhuyan, M., Hines, E.L., Gardner, J.W.: Electronic nose based tea quality standardization. Neural Networks 16(5-6), 847–853 (2003)
6. Gonzalez, E., Li, G., Ruiz, Y., Zhang, J.: A Tea Classification Method Based on an Olfactory System Model. In: Advances in Cognitive Neurodynamics ICCN 2007, pp. 747–751 (2008)
7. Nouretdinov, I., Vovk, V.: Criterion of calibration for Transductive Confidence Machine with limited feedback. Theoretical Computer Science, Algorithmic learning theory 364(1), 3–9 (2006)
8. Tudu, B., Jana, A., Metla, A., Ghosh, D., Bhattacharyya, N., Bandyopadhyay, R.: Electronic nose for black tea quality evaluation by an incremental RBF network. Sensors and Actuators B: Chemical 138, 90–95 (2009)
9. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
10. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer, Heidelberg (2005)
11. Yang, X., Fu, J., Lou, Z., Wang, L., Li, G., Freeman, W.J.: Tea classification based on artificial olfaction using bionic olfactory neural network. In: Proceedings of Third International Symposium on Neural Networks, pp. 343–348 (2006)
12. Yu, H., Wang, J.: Discrimination of LongJing green-tea grade by electronic nose. Sensors and Actuators B 122, 134–140 (2007)