

Article

Conformal Prediction Based on K-Nearest Neighbors for Discrimination of Ginsengs by a Home-Made Electronic Nose

Zhan Wang ¹ , Xiyang Sun ¹, Jiacheng Miao ¹, You Wang ^{1,*}, Zhiyuan Luo ² and Guang Li ¹

¹ State Key Laboratory of Industrial Control Technology, Institute of Cyber Systems and Control, Zhejiang University, Hangzhou 310027, China; 11732003@zju.edu.cn (Z.W.); 21532085@zju.edu.cn (X.S.); jiacheng@zju.edu.cn (J.M.); guangli@zju.edu.cn (G.L.)

² Computer Learning Research Centre, Royal Holloway, University of London, Egham Hill, Egham, Surrey TW20 0EX, UK; zhiyuan@cs.rhul.ac.uk

* Correspondence: king_wy@zju.edu.cn

Received: 8 July 2017; Accepted: 8 August 2017; Published: 14 August 2017

Abstract: An estimate on the reliability of prediction in the applications of electronic nose is essential, which has not been paid enough attention. An algorithm framework called conformal prediction is introduced in this work for discriminating different kinds of ginsengs with a home-made electronic nose instrument. Nonconformity measure based on k-nearest neighbors (KNN) is implemented separately as underlying algorithm of conformal prediction. In offline mode, the conformal predictor achieves a classification rate of 84.44% based on 1NN and 80.63% based on 3NN, which is better than that of simple KNN. In addition, it provides an estimate of reliability for each prediction. In online mode, the validity of predictions is guaranteed, which means that the error rate of region predictions never exceeds the significance level set by a user. The potential of this framework for detecting borderline examples and outliers in the application of E-nose is also investigated. The result shows that conformal prediction is a promising framework for the application of electronic nose to make predictions with reliability and validity.

Keywords: conformal prediction; electronic nose; k-nearest neighbors; ginseng

1. Introduction

Two main problems that the electronic nose (E-nose) faces are classification and regression. Lots of techniques, such as support vector machine (SVM) [1], k-nearest neighbors (KNN) [2–4], artificial neural network (ANN) [5,6], linear discriminant analysis (LDA) [2–4,7,8] and other methodologies, have been successfully applied for predictions with E-nose. However, there are still two drawbacks: (1) the lack of reliable measure of the confidence in their individual prediction; and (2) the accuracy of overall prediction is not guaranteed. Many approaches have been developed to complement these drawbacks by predicting with additional information, such as probably approximately correct learning (PAC), Bayesian learning, generalized least square regression in combination with a stepwise backward selection [9] and hold-out estimate or cross-validation.

In PAC learning, upper bounds on the probability of error often exceed one even for a relatively clean data set, which makes it useless in practical applications. In addition, PAC learning does not provide any information on the reliability of individual prediction [10]. In contrast, Bayesian learning and other probability algorithms, such as logistic regression [11] and Platt's method [12], can complement every individual prediction with information of probability to indicate how every potential label is correct. However, the main disadvantage of these algorithms is that they depend on a strong statistical assumption for the model. Once the data does not conform well to statistical

model, the prediction may be invalid and misleading. Data generated from E-nose is usually difficult to be described by a precise statistical model because most gas sensors suffer from sensor drift caused by variation of temperature and humidity, sensor aging, lack or partial selectivity and sensor poisoning [13,14]. Though lots of work have been devoted to drift calibration, there is still no effective approach to compensate the drift [15–17]. Hold-out estimate and cross-validation provide information about the accuracy of overall prediction. Usually, the dataset is divided into training set and testing set; then, the average prediction accuracy of the testing set is used to estimate the performance of the model. However, this performance estimation usually provides overoptimistic results in real-world applications, since testing samples measured afterwards may suffer from sensor drift, sensor aging or sensor poisoning. ‘External validation’ is strongly recommended for artificial olfaction by some researchers, where other testing sets generated in different experiment conditions from the training set should be used [18]. However, the performance of ‘external validation’ is still unsure and also depends on the experiment conditions. Therefore, the accuracy of overall prediction cannot be guaranteed by hold-out estimate or cross-validation.

Conformal prediction was proposed and developed by Vladimir Vovk and his co-workers since 2005 [10,11,19,20]. It is based on a consistent and well-defined mathematical framework and measures how well the new example is conformed to the group of observations. The algorithm produces each individual prediction with additional information of confidence and credibility. The most important property of conformal prediction is automatic validity under the randomness assumption: the objects and their labels are assumed to be generated from the identical probability distribution. Informally, validity means that conformal predictor never overrates the accuracy and reliability of their prediction. The randomness assumption is a much weaker condition than statistical assumption and can easily be satisfied by real-world data, such as E-nose data.

In this paper, a home-made electronic nose is introduced to discriminate nine different kinds of ginsengs and conformal prediction is used to make reliable and valid predictions. The definition of conformal prediction and nonconformity measure is introduced in Section 2. Then, the sampling experiment with a homemade E-nose and data pre-processing process is presented in Section 3. The results and discussions are shown in Section 4. Finally, we draw the conclusions of our research in Section 5.

2. Conformal Prediction

2.1. Definition

Let \mathbf{X} be a measurable space (the object space) and \mathbf{Y} be a finite set (the label space). Every sample, $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)$, is composed of a object $\mathbf{x}_i \in \mathbf{X}$ and a label $\mathbf{y}_i \in \mathbf{Y}$. The observation space is defined as $\mathbf{Z} := \mathbf{X} * \mathbf{Y}$, and $\mathbf{z}_i \in \mathbf{Z}$. We find a measurable function A that changes every sequence of observations, $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n) \in \mathbf{Z}$, to a same-length sequence $(\alpha_1, \dots, \alpha_n) \in \mathbf{R}$, which is formed by positive real numbers and is equivariant with respect to permutations: for any n and any permutation π of $\{1, \dots, n\}$

$$(\alpha_1, \dots, \alpha_n) = A(\mathbf{z}_1, \dots, \mathbf{z}_n) \rightarrow (\alpha_{\pi(1)}, \dots, \alpha_{\pi(n)}) = A(\mathbf{z}_{\pi(1)}, \dots, \mathbf{z}_{\pi(n)}).$$

The conformal predictor determined by A meets the exchangeability assumption and is defined by

$$\Gamma^\epsilon(\mathbf{z}_1, \dots, \mathbf{z}_l, \mathbf{x}) := \{\mathbf{y} | p^{\mathbf{y}} > \epsilon\}, \quad (1)$$

where $(\mathbf{z}_1, \dots, \mathbf{z}_l) \in \mathbf{Z}$ is a training sequence that is a part of observation space $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)$, \mathbf{x} is a test object, and $\mathbf{y} \in \mathbf{Y}$ is potential labels for \mathbf{x} . In this work, E-nose was used to deal with

classification problem, so \mathbf{Y} is the set of labels of all sample categories, Γ^ϵ is a corresponding predict region with a given *significance level* $\epsilon \in (0, 1)$. For each $\mathbf{y} \in \mathbf{Y}$, the corresponding p -value is defined by

$$p^y = \frac{|\{i = 1, \dots, l+1 \mid \alpha_i^y \geq \alpha_{l+1}^y\}|}{l+1}. \quad (2)$$

The corresponding sequence of *nonconformity* scores is defined by

$$(\alpha_1^y, \dots, \alpha_l^y, \alpha_{l+1}^y) = A(z_1, \dots, z_l, (x, y)). \quad (3)$$

Generally speaking, the lower the α_{l+1}^y is, the more confidence we have. The lower the p^y is, the less we can trust this prediction.

It is clearly that the predict region in a conformal predictor is nested, i.e., for any $\epsilon_1 < \epsilon_2$,

$$\Gamma^{\epsilon_2}(z_1, \dots, z_l, x) \subseteq \Gamma^{\epsilon_1}(z_1, \dots, z_l, x). \quad (4)$$

The property of validity of conformal predictor is that for any l , the probability of the event

$$y_{l+1} \in \Gamma^\epsilon(z_1, \dots, z_l, x_{l+1})$$

is at least $1 - \epsilon$, i.e.,

$$P(y_{l+1} \notin \Gamma^\epsilon(z_1, \dots, z_l, x_{l+1})) \leq \epsilon. \quad (5)$$

2.2. Nonconformity Measure

In theory, any other prediction algorithms can be modified and developed as underlying algorithms (nonconformity measure). Vladimir Vovk once used k -nearest neighbors as the underlying algorithm to compute nonconformity measure for the conformal prediction [10]; so, in this work, we implement a typical nonconformity measure based on KNN. For simplicity, we define CP-1NN and CP-3NN as conformal predictor based on 1NN and 3NN, while simple predictors are 1NN and 3NN.

Given a sequence of examples (z_1, \dots, z_n) , the nonconformity score measure by CP-KNN is

$$\alpha_i := \frac{\sum_{j=1}^k d_{ij}^+}{\sum_{j=1}^k d_{ij}^-} \quad i = 1, \dots, n, \quad (6)$$

with the example $\mathbf{z}_i = (x_i, y_i)$, where d_{ij}^+ is the j th shortest distance from x_i to other objects labelled the same as x_i , and d_{ij}^- is the j th shortest distance from x_i to other objects labelled different from x_i . The parameter k is the number of elements taken into account. The larger α_i is, the stranger \mathbf{z}_i is, and it shows \mathbf{z}_i is more non-conformal than other elements.

2.3. Prediction in Online Mode

Online learning is the most popular and practical learning protocol in machine learning. In online learning mode, examples are presented one by one. Every time, we observe a new object x_i and predict its label y_i with old examples (z_1, \dots, z_{i-1}) . Then, we observe the label y_i , and put (x_i, y_i) into the sequence of old examples. Then, we observe another new object x_{i+1} and predict its label y_{i+1} with old examples (z_1, \dots, z_i) . The quality of prediction would be improved as more and more old examples are accumulated. This is the way of how the online mode are learning.

The process of conformal prediction can be summarized by the following protocol [10]:

*c*ONLINE PREDICTION PROTOCOL :

$Err_0^\epsilon := 0, \epsilon \in (0, 1);$
 $Mult_0^\epsilon := 0, \epsilon \in (0, 1);$
 $Emp_0^\epsilon := 0, \epsilon \in (0, 1);$
 Training set = $\{(x_1, y_1), \dots, (x_i, y_i)\}$
 FOR $n = l + 1, l + 2, \dots$:
 Reality outputs $x_n \in X;$
 Predictor outputs $\Gamma_n^\epsilon \subseteq Y$ for all $\epsilon \in (0, 1)$
 Reality outputs $y_n \in Y$
 $err_n^\epsilon = \begin{cases} 1 & \text{if } y_n \notin \Gamma_n^\epsilon \\ 0 & \text{otherwise} \end{cases}$
 $Err_n^\epsilon = Err_n^\epsilon + err_n^\epsilon$
 $mult_n^\epsilon = \begin{cases} 1 & \text{if } |\Gamma_n^\epsilon| > 1 \\ 0 & \text{otherwise} \end{cases}$
 $Mult_n^\epsilon = Mult_n^\epsilon + mult_n^\epsilon$
 $emp_n^\epsilon = \begin{cases} 1 & \text{if } |\Gamma_n^\epsilon| = 0 \\ 0 & \text{otherwise} \end{cases}$
 $Emp_n^\epsilon = Emp_n^\epsilon + emp_n^\epsilon$
 Training set = $\{\text{Training set}, (x_n, y_n)\}$
 End FOR.

Combining Formula (5) and the strong law of large numbers, we can deduce that

$$\lim_{n \rightarrow \infty} \sup \frac{Err_n^\epsilon}{n} \leq \epsilon, \quad (7)$$

holding with probability one for the conformal predictor [20].

2.4. Prediction in Offline Mode

Validity of conformal prediction is only proved for the online mode, in which every example is predicted one by one and every prediction is made based on the examples that have been considered before, rather than generating a certain rule from a fixed set of examples. However, the conformal prediction can still work in the offline mode and provide additional information about reliability for every prediction.

In offline mode, we can set a significance level to force the conformal predictor to output a prediction, which is the label with the highest p -value. This approach is called *forced prediction*, similar to *simple prediction*, which is designed to output singleton label. However, in this case, the significance level varies across different examples and the validity does not hold [19].

Two indicators, *confidence* and *credibility*, are used to provide additional information about the prediction. They are defined as:

$$\text{confidence} : \sup\{1 - \epsilon : |\Gamma^\epsilon| \leq 1\},$$

$$\text{credibility} : \inf\{\epsilon : |\Gamma^\epsilon| = 0\}.$$

In the classification case, *confidence* equals 1 minus the second maximum p -value, which shows how confident we are rejecting other labels. *Credibility* equals the highest p -value, which shows how well the chosen label conforms to the rest of the set. The forced prediction is considered to be reliable if its confidence is close to 1 and credibility is not close to 0 [19].

3. Experiment and Methods

3.1. Sample Preparation

Nine categories of ginsengs (35 pieces for every category), which are showed in Table 1, were purchased from Changchun Medicinal Material Market (Changchun, China) randomly. Every ginseng sample was pulverized into powder, and 10 g powder of every sample was put into 100 mL empty glass bottles, which has been washed with clean air for 30 min separately before. Then, all of the bottles were sealed and placed at 50 °C for 30 min. Finally, 10 mL head-space gas was extracted from the top of the bottle for measurement with an injector.

Table 1. Details of the ginseng samples.

No.	Ginseng Samples	Places of Production
1	Chinese red ginseng	Ji'an
2	Chinese red ginseng	Fusong
3	Korean red ginseng	Ji'an
4	Chinese white ginseng	Ji'an
5	Chinese white ginseng	Fusong
6	American ginseng	Fusong
7	American ginseng	USA
8	American ginseng	Canada
9	American ginseng	Tonghua

3.2. E-Nose Equipment and Measurement

All of the samples were measured with a homemade E-nose consisting of 16 metal-oxide semi-conductive sensors of TGS type purchased from Figaro Engineering Inc. (Osaka, Japan). TGS sensors have been used in the field of food classification [21,22]. The response characteristics of sensors are listed in Table 2. The schematic of E-nose system is shown in Figure 1. All sensors were fixed on a printed circuit board and placed in a 200 mL stainless chamber. A three-way valve was used to switch between target gas and clean dry air. Two mini vacuum pumps were fixed for air washing with a constant air flow of 1 L/min. A data acquisition (DAQ) unit USB6211, purchased from National Instruments Inc. (Austin, TX, USA), is equipped to record the signals of different sensors and control the pumps. Heating temperature for metal-oxide semi-conductive sensor is very important, and heater voltage of 5 V DC was applied for each sensor, as recommended by Figaro Engineering Inc. to guarantee the best performance of the sensors.

Table 2. The response characteristics of sensors.

No.	Sensor Type	Response Characteristic
1	TGS800	Carbon monoxide, ethanol, methane, hydrogen, ammonia
2	TGS813	Carbon monoxide, ethanol, methane, hydrogen, isobutane
3	TGS813	Carbon monoxide, ethanol, methane, hydrogen, isobutane
4	TGS816	Carbon monoxide, ethanol, methane, hydrogen, isobutane
5	TGS821	Carbon monoxide, ethanol, methane, hydrogen
6	TGS822	Carbon monoxide, ethanol, methane, acetone, n-Hexane, benzene, isobutane
7	TGS822	Carbon monoxide, ethanol, methane, acetone, n-Hexane, benzene, isobutane
8	TGS826	Ammonia, trimethyl amine
9	TGS830	Ethanol, R-12, R-11, R-22, R-113
10	TGS832	R-134a, R-12 and R-22, ethanol
11	TGS800	Carbon monoxide, ethanol, methane, hydrogen, isobutane
12	TGS2620	Methane, Carbon monoxide, isobutane, hydrogen
13	TGS2600	Carbon monoxide, hydrogen
14	TGS2602	Hydrogen, ammonia ethanol, hydrogen sulfide, toluene
15	TGS2610	Ethanol, hydrogen, methane, isobutane/propane
16	TGS2611	Ethanol, hydrogen, isobutane, methane

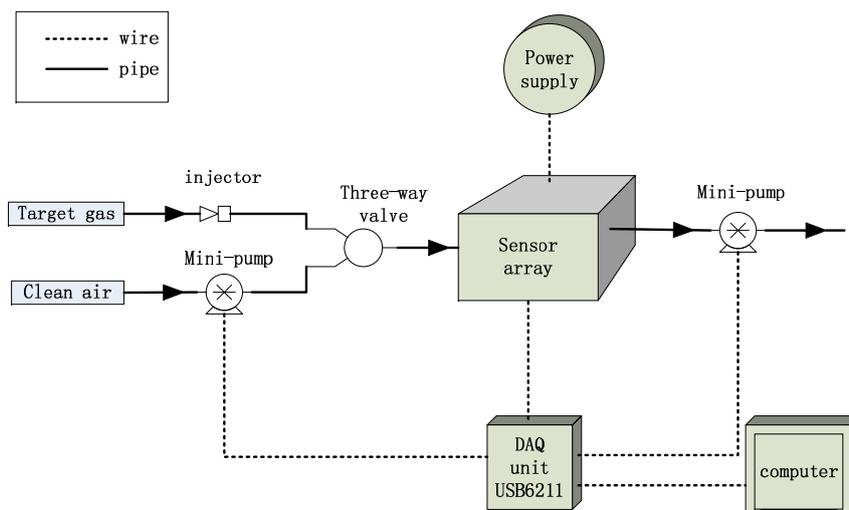


Figure 1. The schematics of the E-nose system.

The procedure of measurement is as follows: firstly, the test chamber loaded with sensor array was washed by a clean-dry-air flow of 1 L/min for 360 s to allow the sensors to return to the baseline. Such flow can wash the sensors clean in 100–200 s, just near the time of reaction and won't damage the E-nose. Then, washing-air flow was stopped and 10 mL target gas was taken and injected into the chamber through an injector. The target gas was left in the chamber obtained for 200 s, so that the gas sample diffused in the whole chamber freely to a steady condition. After that, the valve of washing-air flow was open to wash away the target gas. Responses of 16 sensors were recorded for 340 s, which included 20 s before the injection of the target gas, 200 s of the whole reaction time and 120 s after the washing flow was opened. The sampling frequency is of 10 Hz. The procedure of measurement is shown in Figure 2.

A total of 315 samples taken from nine categories of ginsengs were measured at room temperature (22–25 °C) and humidity (50–70%). During each measurement, the temperature and humidity of ambient environment were relatively stable.

Finally, all of the statistical analysis were performed by using Matlab 9.0.0.341360 (R2016a) (MathWorks, Natick, MA, USA).

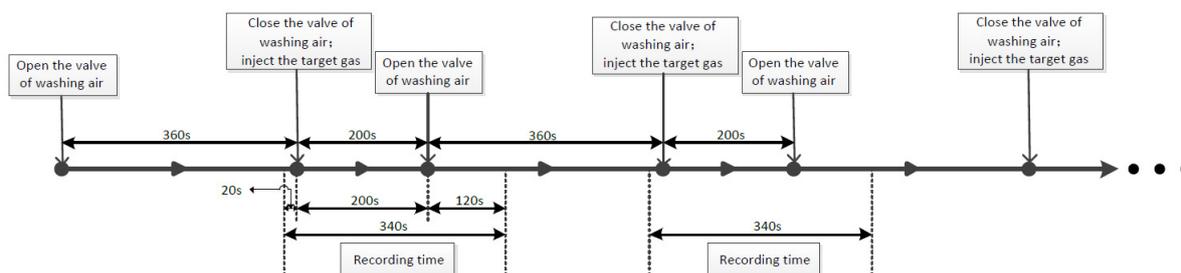


Figure 2. The procedure of measurement.

3.3. Data Preprocessing

The typical response curve of 16 sensors to a ginseng sample is shown in Figure 3. The voltage signal of each sensor was converted to resistance signal and then the resistance signal of each sensor was calibrated separately by:

$$R = (R_s - R_0) / R_0, \quad (8)$$

where R_s is the original resistance signal and R_0 is the value of baseline. Five commonly used features were extracted (a total of 80 features with 16 sensors) and discussed as follows:

1. The maximal absolute response value, $R_{max} = \max(|R|)$.
2. The area under the full response curve, $R_{int} = \int_0^T R(t)dt$, where T is the total measurement time, T = 340 s.
- 3–5. Exponential moving average of derivative of R, $E_a(R) = \max(|y(k)|)$, $k \in [1, 400]$. The discretely sampled exponential moving average is defined as $y(k) = (1 - a)y(k - 1) + a(R(k) - R(k - 1))$ with smoothing factors $a = 0.005, 0.05, 0.5$ (sampling frequency: 10 Hz). Thus, three different smooth factors “a” give us the last three features.

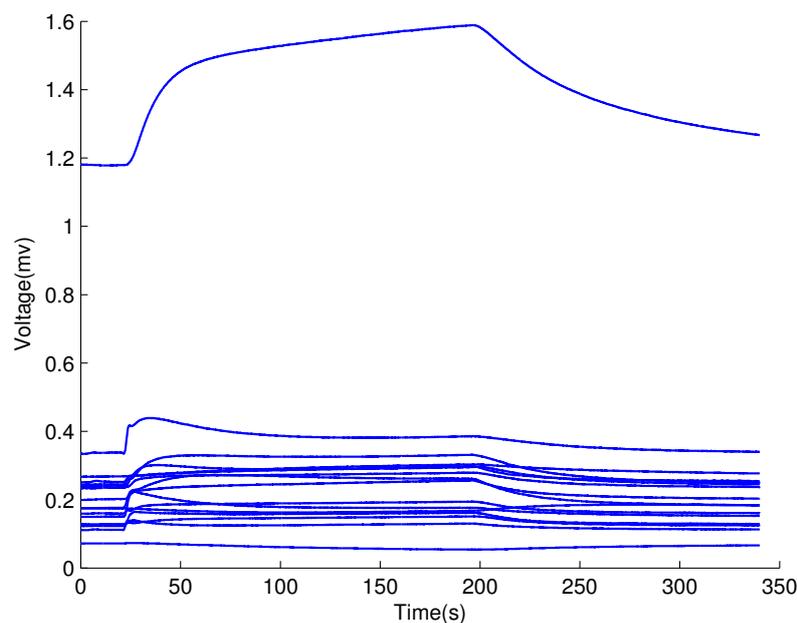


Figure 3. Typical response curves of 16 metal-oxide semi-conductive sensors to a sample.

4. Results and Discussion

4.1. Comparison of Forced Conformal Prediction with Simple Prediction

In offline mode, a forced conformal predictor can be used as a simple predictor to output only one label with the maximum p -value. It also provided indicators of confidence and credibility for the predicted label, compared to simple predictor. To compare the characteristics of forced conformal predictor and simple predictor, 1NN and 3NN were used as nonconformity measure methods and simple predictors separately.

Results of four typical individual predictions by forced predictor with CP-1NN are shown in Table 3. The p -value for each potential label, confidence, and credibility of forced conformal predictor are given. For Sample 1, both confidence and credibility are close to one, which means that we are confident of rejecting other potential labels, and the new object with predicted label conforms well to the old examples. This is the ideal case where we can strongly believe in our prediction. For Samples 2 and 3, the confidence is high (>0.90) while the credibility is rather low (<0.16). It tells us that the new object with a predicted label does not conform well to old examples. However, the new object with other potential labels conforms much worse to the old examples. Therefore, we have less confidence in rejecting other potential labels because no other label is appropriate or even close to the predicted label. For Sample 4, the confidence is low, which means that we are not so confident of rejecting other potential labels, so error prediction is very likely to occur in this case.

Table 3. Typical individual prediction with CP-1NN (conformal prediction based on 1NN).

Sample Serial	True Label	Forced Prediction	Confidence	Credibility	Simple Prediction
1	1	1	0.9968	0.9397	1
2	1	1	0.9047	0.1143	1
3	2	2	0.9587	0.1587	2
4	2	1	0.7714	0.2349	1

Compared to forced conformal prediction, simple prediction only outputs the predicted label with no comment on the reliability of the prediction.

In addition, we compare the average classification rate of forced conformal predictors and simple predictors. The average classification rate was achieved with leave-one-out validation: in every cycle, only one sample was taken as a testing set, and the remaining as a training set; the cycle was repeated until all samples had been treated as testing samples once. The result is shown in Table 4. We can see that when using forced prediction, the conformal predictors have the approximate classification accuracy as simple predictors, but provide additional information about the reliability of the prediction. Thus, we can demonstrate that the framework will not sacrifice the classification rate while it gives us prediction reliability information.

Table 4. Comparison of average classification rate of forced conformal predictors and simple predictors.

Predictors	1NN	3NN
Forced conformal predictor	84.44%	80.63%
Simple predictor	84.13%	77.46%

4.2. Validity of Online Conformal Prediction

CP-1NN and CP-3NN are used for online conformal prediction of discriminating ginseng samples according to the protocol in Section 2.3. Five samples were randomly taken from each class and treated as initial training sets. The rest of the samples were randomly sorted, predicted and added to the training set one by one. The cumulative error (Err_n^ϵ), multiple predictions ($Mult_n^\epsilon$) and empty predictions (Emp_n^ϵ) for CP-1NN with confidence level of 80% are shown in Figure 4. As the number of samples in the training set increased, the predictor is more and more stable. Increasing of multiple predictions becomes slow gradually and even stops when the number of training samples is over 133. Then, the conformal predictor only outputs singleton prediction, just like a simple predictor dose, while keeping validity, which we will discuss below. The number of empty predictions starts to show up when the size of the training set is over 159, which can be explained by the fact that with enough training samples, the conformal predictor is confident of judging that these certain samples do not belong to any other class under the confidence level of 80%, and it is very likely that they are outliers.

Two main properties of conformal predictor are validity and efficiency. The ratio of errors Err_n^ϵ/n is a measure of validity of conformal predictor. The cumulative error Err_n^ϵ for CP-1NN with different confidence levels of 80%, 85% and 90% are shown in Figure 5. For all n and different confidence levels, we have $Err_n^\epsilon < \epsilon n$, which testify to the validity of online conformal predictors.

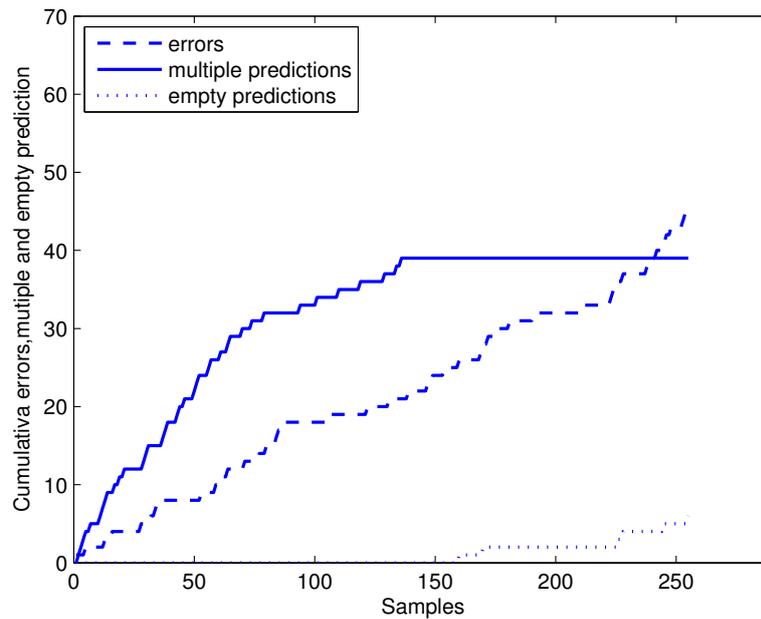


Figure 4. Online conformal prediction with confidence level of 80%.

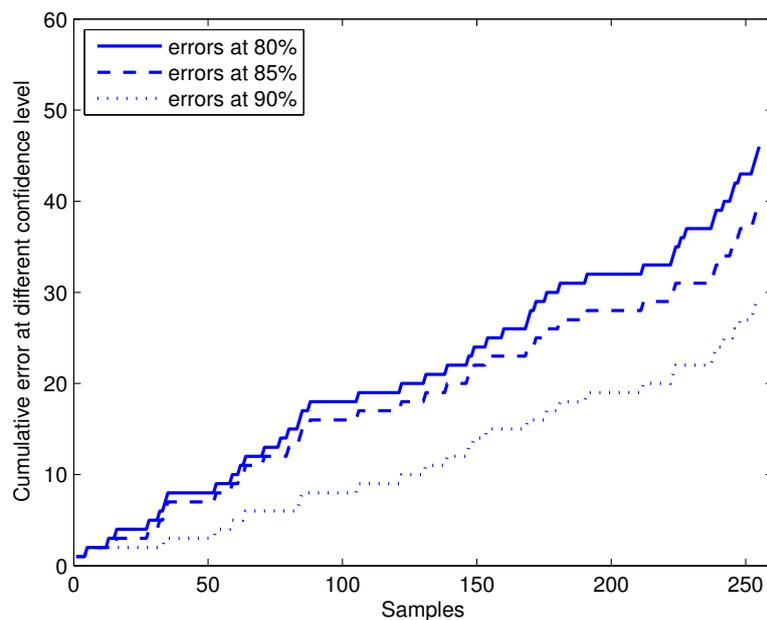


Figure 5. The cumulative errors of online prediction with CP-1NN (conformal prediction based on 1NN) at confidence levels of 80%, 85% and 90%.

Besides guaranteeing the validity of conformal prediction, multiple prediction and empty prediction can provide the information about the distribution of the samples in feature space. For binary classification, multiple prediction with both labels may seem useless, but it demonstrates that both of the labels are reliable with a certain level of confidence, and this sample is very likely to be the borderline example between the two classes. For multi-label classification, multiple prediction can exclude those labels that are not reliable with a certain confidence level, and the sample is very likely to be the borderline example among the classes in this predicted region. Empty prediction tells that, with a certain confidence level, no label is reliable for the sample, and it is very likely to be an outlier.

4.3. Efficiency of Online Conformal Prediction

By choosing different confidence levels $1 - \epsilon$, the error ratio is guaranteed to be under significance level ϵ . However, it does not mean that if we choose the higher confidence level, the predictor will perform better. As the predicted region by a conformal predictor is nested (stated in Section 2.1), higher confidence levels lead to low efficiency, which means wider predicted regions and more multiple predictions, and that is not what we expected. How the multiple predictions by online conformal predictors with CP-1NN vary with different confidence levels is showed in Figure 6. With lower confidence levels, conformal predictors can become stable with a smaller size of training set and gradually decreasing cases of multiple predictions. With higher confidence levels, the conformal predictor has to output more labels to avoid overrate errors occurring and more multiple prediction being output.

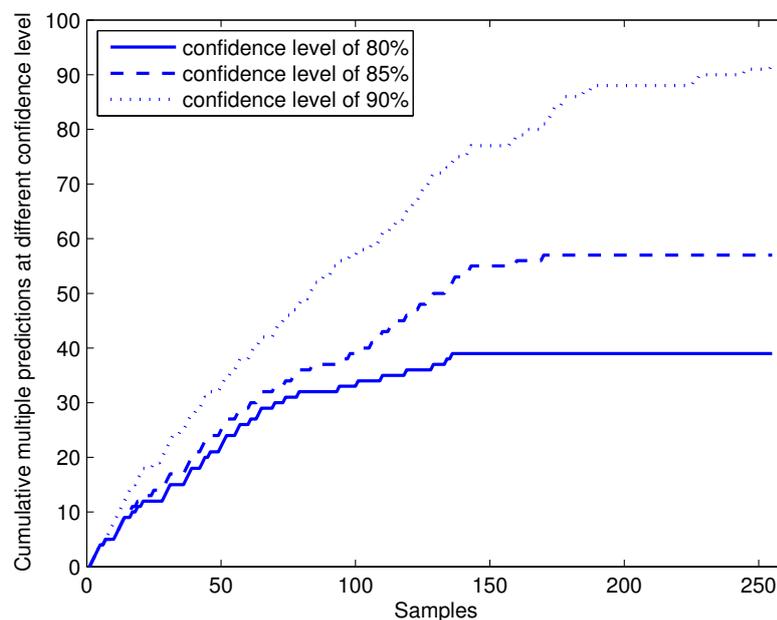


Figure 6. Cumulative multiple predictions of online conformal prediction with CP-1NN at different confidence levels of 80%, 85% and 90%.

Thus, when the nonconformity measure method is fixed, we have to balance between confidence level and efficiency. However, we can still improve the efficiency of conformal predictors under the same confidence levels by improving the nonconformity measure methods. There are two major criteria depending on confidence level for conformal predictor [23]. One is the ratio of multiple predictions in the test sequence, which is known as M (which stands for ‘multiple’) criterion. The other is the average number of labels in the predict region of multiple prediction, which is known as the E (stand for ‘excess’) criterion. The lower these two criteria are, the better the efficiency of the predictor. These two criteria of efficiency for CP-1NN and CP-3NN with confidence levels of 80%, 85%, and 90% were compared in Table 5. The CP-1NN has the lowest ratio of multiple prediction for all confidence levels, and the average number of labels in multiple predictions is just a little higher than that of CP-3NN for confidence of 80% and 85%. Therefore, we can conclude that CP-1NN achieved better efficiency than that of CP-3NN.

Table 5. Criterion of efficiency for online conformal predictors.

Confidence Level	CP-1NN		CP-3NN	
	M Criterion	E Criterion	M Criterion	E Criterion
80%	15.29%	1.23	23.92%	1.32
85%	22.35%	1.40	32.94%	1.47
90%	36.08%	1.62	45.88%	1.75

5. Conclusions

In this work, conformal prediction is introduced for discriminating nine different ginsengs with a home-made electronic nose. Nonconformity measure methods with 1NN and 3NN are completed as underlying algorithms of conformal prediction. In offline mode, conformal predictors provides singleton predictions as well as additional information on reliability on the prediction. In addition, conformal prediction achieves a classification rate 84.44% with CP-1NN and 80.63% with CP-3NN, which are better than that of simple 1NN and 3NN with a modest increase, respectively. In online mode, the validity of conformal prediction is discussed and testified. The efficiency of conformal prediction are compared for CP-1NN and CP-3NN. Both the “M criterion” and “E criterion” of CP-1NN are lower than those of CP-3NN under every confidence level in the work, such as (15.29%, 1.23) of CP-1NN being lower than (23.92%, 1.32) at the confidence level of 80%, which indicates that the efficiency of CP-1NN is better than that of CP-3NN. The potential of conformal prediction for detecting borderline examples and outliers in the application of E-nose is also discussed. In conclusion, conformal prediction can give every prediction the estimation of reliability without sacrificing classification rates, which can make the prediction more comprehensive in both offline and online modes. Future work will be focused on designing new methods of computing nonconformity measures to improve the validity and efficiency of conformal prediction for E-nose data.

Acknowledgments: The work is supported by the Natural Science Foundation of China (NSFC) (Grant No. 61403339) and the Science Fund for Creative Research Groups of the NSFC (Grant No. 61621002).

Author Contributions: Zhan Wang, You Wang and Guang Li conceived and designed the experiments; Zhan Wang and Jiacheng Miao performed the experiments; Zhan Wang, Xiyang Sun and Zhiyuan Luo analyzed the data; and Zhan Wang, Xiyang Sun and You Wang wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaur, R.; Kumar, R.; Gulati, A.; Ghanshyam, C.; Kapur, P.; Bhondekar, A.P. Enhancing electronic nose performance: A novel feature selection approach using dynamic social impact theory and moving window time slicing for classification of Kangra orthodox black tea (*Camellia sinensis* (L.) O. Kuntze). *Sens. Actuators B Chem.* **2012**, *166*, 309–319.
2. Miao, J.; Zhang, T.; Wang, Y.; Li, G. Optimal sensor selection for classifying a set of ginsengs using metal-oxide sensors. *Sensors* **2015**, *15*, 16027–16039.
3. Gebicki, J.; Szulczynski, B.; Kaminski, M. Determination of authenticity of brand perfume using electronic nose prototypes. *Meas. Sci. Technol.* **2015**, *26*, 125103.
4. Gebicki, J.; Dymerski, T.; Namiesnik, J. Investigation of air quality beside a municipal landfill: The fate of malodour compounds as a model VOC. *Environments* **2017**, *4*, 7.
5. Singh, H.; Raj, V.B.; Kumar, J.; Mittal, U.; Mishra, M.; Nimal, A.T.; Sharma, M.U.; Gupta, V. Metal oxide SAW E-nose employing PCA and ANN for the identification of binary mixture of DMMP and methanol. *Sens. Actuators B Chem.* **2014**, *200*, 147–156.
6. Longobardi, F.; Casiello, G.; Ventrella, A.; Mazzilli, V.; Nardelli, A.; Sacco, D.; Catucci, L.; Agostiano, A. Electronic nose and isotope ratio mass spectrometry in combination with chemometrics for the characterization of the geographical origin of Italian sweet cherries. *Food Chem.* **2015**, *170*, 90–96.

7. Green, G.C.; Chan, A.D.C.; Dan, H.H.; Lin, M. Using a metal oxide sensor (MOS)-based electronic nose for discrimination of bacteria based on individual colonies in suspension. *Sens. Actuators B Chem.* **2011**, *152*, 21–28.
8. Dymerski, T.; Gebicki, J.; Wardencki, W.; Namiesnik, J. Application of an electronic nose instrument to fast classification of polish honey types. *Sensors* **2014**, *14*, 10709–10724.
9. Giungato, P.; Laiola, E.; Nicolardi, V. Evaluation of industrial roasting degree of coffee beans by using an electronic nose and a stepwise backward selection of predictors. *Food Anal. Methods* **2017**, doi:10.1007/s12161-017-0909-z.
10. Gammerman, A.; Vovk, V. Hedging predictions in machine learning-The second Computer Journal Lecture. *Comp. J.* **2007**, *50*, 151–163.
11. Nouretdinov, I.; Devetyarov, D.; Vovk, V.; Burford, B.; Camuzeaux, S.; Gentry-Maharaj, A.; Tiss, A.; Smith, C.; Luo, Z.Y.; Chervonenkis, A.; et al. Multiprobabilistic prediction in early medical diagnoses. *Ann. Math. Artif. Intell.* **2015**, *74*, 203–222.
12. Zhou, C.Z.; Nouretdinov, I.; Luo, Z.Y.; Adamskiy, D.; Randell, L.; Coldham, N.; Gammerman, A. A Comparison of venn machine with platt's method in probabilistic outputs. *Artif. Intell. Appl. Innov. Pt Ii* **2011**, *364*, 483–490.
13. De Vito, S.; Piga, M.; Martinotto, L.; Di Francia, G. CO, NO(2) and NO(x) urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. *Sens. Actuators B Chem.* **2009**, *143*, 182–191.
14. Romain, A.C.; Nicolas, J. Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. *Sens. Actuators B Chem.* **2010**, *146*, 502–506.
15. Nimsuk, N.; Nakamoto, T. Study on the odor classification in dynamical concentration robust against humidity and temperature changes. *Sens. Actuators B Chem.* **2008**, *134*, 252–257.
16. Padilla, M.; Perera, A.; Montoliu, I.; Chaudry, A.; Persaud, K.; Marco, S. Drift compensation of gas sensor array data by Orthogonal Signal Correction. *Chemom. Intell. Lab. Syst.* **2010**, *100*, 28–35.
17. Ziyatdinov, A.; Marco, S.; Chaudry, A.; Persaud, K.; Caminal, P.; Perera, A. Drift compensation of gas sensor array data by common principal component analysis. *Sens. Actuators B Chem.* **2010**, *146*, 460–465.
18. Marco, S. The need for external validation in machine olfaction: Emphasis on health-related applications. *Anal. Bioanal. Chem.* **2014**, *406*, 3941–3956.
19. Vovk, V. Conditional validity of inductive conformal predictors. *Mach. Learn.* **2013**, *92*, 349–376.
20. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithm Learning in a Random World*; Springer: New York, NY, USA, 2005.
21. Haddi, Z.; Boughrini, M.; Ihlou, S.; Amari, A.; Mabrouk, S.; Barhoumi, H.; Maaref, A.; Bari, N.E.; Llobet, E.; Jaffrezic-Renault, N.; et al. Geographical classification of Virgin Olive Oils by combining the electronic nose and tongue. *Sensors* **2012**, doi:10.1109/icsens.2012.6411502.
22. Timsorn, K.; Wongchoosuk, C.; Wattuya, P.; Promdaen, S.; Sittichat, S. Discrimination of chicken freshness using electronic nose combined with PCA and ANN. In Proceedings of the International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Nakhon Ratchasima, Thailand, 14–17 May 2014; pp. 1–4.
23. Vovk, V.; Fedorova, V.; Nouretdinov, I.; Gammerman, A. Criteria of efficiency for conformal prediction. In *Lecture Notes in Computer Science*; Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Springer: Berlin, Germany, 2016; pp. 23–39.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).