From simple classification methods to machine learning for the binary discrimination of beers using electronic nose data

Mahdi Ghasemi-Varnamkhasti, Seyed Saeid Mohtasebi, Maryam Siadat, Hojat Ahmadi, Seyed Hadi Razavi

1. Introduction

Electronic nose technology is used to get a fingerprint of volatile compounds present in the headspace of a food sample using an array of semi-selective sensors, e.g. metal oxide semiconductor or MOS-based sensors. This technology is currently under consideration by food engineers for food authentication and evaluation (Villanueva et al., 2006; Gualdron et al., 2007; Barbri et al., 2008; Apetrei et al., 2010; Concina et al., 2010; Ghasemi-Varnamkhasti and Aghbashlo, 2014). Recently, the work conducted in this field of research has been reviewed by Peris and Escuder-Gilabert (2009). According to the literature, benefits of electronic nose technology are the need of only relatively small amounts of sample and the speed of analysis (Rock et al., 2008). However, this screening tool employs sensors which are not very selective to specific compounds thus preventing the real recognition or quantification of individual compounds present in a food sample (Reid et al., 2006; Ghasemi-Varnamkhasti et al., 2010).

During the last years, many attempts have been reported about using the electronic nose in beer quality evaluation but because of the many challenges arisen in this field of research, more researches are still needed (Sikorska et al., 2007; Li et al., 2007; Ghasemi-Varnamkhasti et al., 2011a,b). The necessity for electronic nose measurements in brewing industry is growing due to the adaptability and ease of use of such tools makes them appropriate for a fast and precise analysis of beers and even to quality monitoring in the production process. The capability of the e-noses in breweries could be increased by the employment of advanced computational techniques. The Support Vector Machine (SVM) technique, a machine learning technique for regression or classification, has been of interest to researchers in the last decade. Interesting papers on the combination of electronic nose with SVM have been reported in the literature (Brudzewski et al., 2004; Pardo and Sberveglieri, 2005; Acevedo et al., 2007; Gualdron et al., 2007; Wang et al., 2009; Barbri et al., 2009; Phaisangittisagul et al., 2010).

Only few applications of electronic nose to alcoholic beverages are found in the literature. Recent cases have been reported by Ragazzo-Sanchez et al. (2006, 2008, 2009). However, to date, no research is reported concerning the employment of SVM to binary beer discrimination and this can be addressed as a novel idea behind of this research. Therefore, this study was aimed to use this
technique for the binary discrimination of beer types. However before that the visualization of the data set is made by PCA, while LDA as a classifier technique is addressed on the important variables. SIMCA and PLS-DA were also used in this study to compare the binary discrimination capabilities of those methods with both SVM and LDA.

2. Materials and methods

Five tin oxide-based gas sensors purchased from Figaro Engineering (Glenview, USA) and FIS (Osaka, Japan) was used in the sensor array (Ghasemi-Varnamkhasti, 2011). The names and related specifications of the sensors are given in Table 1. A 10 V circuit voltage supplies the sensors which are located in a half bridge and a 5 V heating voltage was used to supply the operating temperature according to the Figaro Engineering and FIS data sheets. In addition to these five sensors, in order to check the created environmental conditions, two other types of sensors were included in the array to control the atmospheric conditions: a humidity sensor (Humirel, National Semiconductor, Osaka, Japan) and a temperature device (National Semiconductor Osaka, Japan). A schematic representation of the electronic nose and the measurement system components are depicted in Fig. 1. More illustrations and descriptions about the system used have earlier been reported (Delpha et al., 2004; Ghasemi-Varnamkhasti et al., 2011a,b).

Two beer types were used in this research: alcoholic (4 brands: Belzebuth (11.2% ethanol), Atlas (7.2%), Bière Abbaye (6.2%), Leffe (6.6%)) and non-alcoholic (2 brands: Buckler and Kronenbourg). Static headspace sampling followed by a dynamic injection was designed for the experimental protocol. The carrier gas was synthetic air for preserving the beers. Around 50 ml of beer sample was maintained in a 250 ml bottle at 25 °C for 45 min to provide a vapor phase in equilibrium with the liquid. In this system, the synthetic air is brought into a sample container based on the bubbling principle and then mixed with the beer headspace. The mixture portion is controlled by a mass flow controller (M.F.C.) (Brooks, Serv Instrumentation, Irigny, UK). The portions were determined as 100 (synthetic air) and 80 (synthetic air) + 20 (sample headspace) ml/min in purging and injection phases, respectively. In purging stage, the electric valves were switched during 1300 s and then dynamic injection of the beer headspace was accomplished for 320 s. This experimental protocol was carried out randomly in seven replicates for each beer brand. To remove the influence of sensor drift, the sensor array was calibrated with a blank solution of ethanol (Fluka) in deionized water (4–11% (v/v)) for different alcoholic beer brands. This approach has been documented for wine aging detection (Gutierrez-Osuna, 2002; Lozano et al., 2008a,b). Data acquisition board (LabView, National Instruments) was used to record the responses of the sensors. One of the signals recorded is shown in Fig. 2. After data collection, the features related to individual sensors were extracted. In feature extraction, the steady state of the signals was addressed (Pearce et al., 2003; Lozano et al., 2008a,b). Then, the following equation was used.

$$ F = \frac{R_{\text{sample}}}{R_{\text{calibration}}} \quad (1) $$

where $R_{\text{sample}}$ is the minimum resistance of the sensor during measurement and $R_{\text{calibration}}$ is that of the sensor exposed to an ethanol solution. Afterwards, autoscaling was exploited for data preprocessing (Pearce et al., 2003). This is the most frequently used scaling technique which standardizes a variable according to Eq. (2):

$$ Z_j = \frac{(x_j - \bar{x}_j)}{s_j} \quad (2) $$

where $Z_j$ is the value of $x_j$ after autoscaling, $x_j$ is the value for the variable before scaling, $\bar{x}_j$ is defined as the variable mean and $s_j$ as the standard deviation of the variable. As a result, the consequence is a variable $Z$ with mean zero and standard deviation one.

3. Data analysis

For analyzing the data collected, several chemometric tools were considered. Chemometrics implies the use of multivariate data analysis based on the fact that complex systems require multiple parameters to be described. As a consequence, more information about the system under analysis can be retrieved through a multivariate approach. Explanations of the chemometric tools used in this study are given below. Matlab v7.6 (The Mathworks, Natick, MA, USA) and The Unscrambler 9.2 (Camo, Norway) were the softwares used for data analysis in this study.

In this work, Principal Components Analysis (PCA) is employed as a data reduction technique for dimensionality reduction (number of variables) of the data set to retain most of the original variability (information) in the data. Score and loading plots show the PCA results and express the data structure. The PCA results were then confirmed by use of the classification techniques of Linear Discriminant Analysis (LDA), Partial Least Square Discriminant Analysis (PLS-DA), Soft Independent Modeling of Class Analogy (SIMCA) and Support Vector Machine (SVM). At the following sections, some explanations on the chemometric tools used are presented briefly. For more study, the readers are referred to the literature (Otto, 2007; Brereton, 2007).

The Kennard and Stone algorithm is used to split the data set into two sets, the so called training and test sets. The Kennard and Stone algorithm is one of the most popular techniques to representatively split a data set into different subsets (Kennard and Stone, 1969). First of all, the number of samples that has to be into the training and test sets should be decided. Then the two samples are found that are furthest away from each other, based upon the input variables, regarding some metric, such as the Euclidean distance. Then these two samples are eliminated from the original data set and inserted into the training set. The third sample selected is the one furthest from the first two objects, etc. This procedure is repeated until the desired number of samples has been reached in the training set.

3.1. LDA

One of the most popular supervised linear statistical techniques, which aims to find a linear combination of variables which best explains the data either for two classes or for multiclass classification, is Linear Discriminant Analysis (LDA). Class borders from this technique are straight linear lines (Wu et al., 1996).

Table 1
Sensor array used in the electronic nose system.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP0: SPMW0</td>
<td>Tin dioxide, monitors the changes of various gases emitted during cooking</td>
</tr>
<tr>
<td>CP1: SMAQ1</td>
<td>Tin dioxide semiconductor sensor, very high sensitivity to volatile organic compounds (VOCs) and solvents</td>
</tr>
<tr>
<td>CP2: TGS2620</td>
<td>Alcohol, toluene, xylene, other volatile organic vapors</td>
</tr>
<tr>
<td>CP3: TGS825</td>
<td>Hydrogen sulfide</td>
</tr>
<tr>
<td>CP4: TGS880</td>
<td>Volatile vapors from food (alcohol)</td>
</tr>
</tbody>
</table>
3.2. PLS-DA

This is a frequently used classification technique, which is based on the PLS approach (Barker and Rayens, 2003). PLS-DA is a variant of PLS that is used when the predictor variable $y$ is binary. The concept of PLS-DA consists first in the building of a PLS regression model on variables which are indicators of the groups. The second step of a PLS-DA method is to classify observations from the PLS regression results on indicator variables (prediction).

3.3. SIMCA

This method is known as a supervised pattern recognition or classification approach (Galtier et al., 2011). This classification technique uses linear discriminant functions which are derived from disjointed PCA of the data. This means each group of the objects is subjected independently to PCA keeping its complexity. Then, the area of acceptance, for the model is taken into account. A set of functions is derived from each group under consideration by
calculating the group mean and a certain number of PCs. Then, objects are classified into the group that its own PC model best reproduces the data. Only data points which are members of a specified group are considered in determining the functions of the model for that group. Moreover, the importance of each feature to classification is revealed by its contribution to the group covariance matrices.

3.4. SVM

Support Vector Machine is one of the recently popular machine learning methods that has been extensively used and investigated because of its ability in prediction, not only for classification but also for regression. The main goal of SVM is to discover decision boundaries separating data points of different classes using hyperplanes as shown in Fig. 3 (Vapnik, 1998).

This subsection briefly introduces SVM, which can also be used for binary classification. Given the training data \((x_1, y_1), \ldots, (x_l, y_l)\), where \(x_i\) are input vectors and \(y_i \in \{-1, +1\}\) are the associated class labels of \(x_i\). SVM is based on the maximum margin principle, and aims at constructing a hyper-plane with maximal distance between the two classes, with the following optimization problem:

\[
f(x) = \text{sign}(W^T \Phi(x) + b)
\]

\[
\min_{w:b} \frac{1}{2}w^T w + C \sum_{i=1}^{l} \xi_i
\]

Subject to \(y_i (w^T \Phi(x_i) + b) \geq 1 - \xi_i\)

\(\xi_i \geq 0, \quad i = 1, \ldots, l\)

where \(l\) denotes the number of samples, \(\Phi(x_i)\) are data mapped to a higher dimensional space by the kernel function \(\Phi, \xi_i\) represents the distance of \(\Phi(x_i)\) with respect to the decision boundary. Samples with \(\xi = 0\) are on the margins or outside the margins boundary of being correctly classified; samples with \(0 < \xi < 1\) are between the margins and are correctly classified, and samples with \(\xi \geq 1\) are misclassified. The error cost \((C)\) defines the trade-off between a large margin and misclassification error (i.e. empirical risk minimization) and the mapping function (also called kernel function).

In fact, kernel functions perform non-linear mapping between the input space and a feature space (to map the data \(x\) into a high-dimensional feature space). In this study a Gaussian (radial basis function) kernel is used, which can be shown as

\[
k(x_i, x) = \exp \left( - \gamma \| (x_i - x) \|^2 \right)
\]

where \(x_i\) is the testing data point and \(x\) the support vector and gamma \((\gamma)\) in equation is a Gaussian kernel parameter which sometimes parameterized using \(\gamma = 1/2\sigma^2\).

4. Results and discussion

The authors tried to separate the beer types in different classes. A set of new orthogonal variables or axes is generated by PCA meaning the principal components. They are linear combinations of the original variables (Ding et al., 2010). The PCA results of the data collected by the electronic nose are illustrated in Fig. 4. Alcoholic and non-alcoholic types of beers are well separated along PC1. The maximum amount of variance in the original dataset is, by definition, accounted in the first two principal components. The principal components are considered as the new axis to plot the beer samples. This plot is known as a score plot and the PC1–PC2 plot is given in Fig 4a. The PCA score plot of the two first principal components accounts for 78% of the data variance. Clearly discriminated groups are observed along PC1. The first group which appears in the left side of the score plot corresponds to non-alcoholic beer brands. The other group that appears on the right side of the score plot corresponds to alcoholic beer brands. Apart from the beer samples, the sensors (variables) could be also displayed in a plot by the values of their coefficients in the eigenvector equations, named loadings. As shown in Fig. 4b, the loading plot indicates the contribution of the sensors used in electronic nose system to each principal component: the higher the loading of a certain sensor (e.g. CP4 and CP3) on a principal component, the higher the sensor's contribution to this component. The chemical meaning of a component can be derived from the sensors mainly contributing to it. The loading plot of this data could give us information about which sensors can be removed from the array, when we want to reduce the fabrication costs of the sensor array of the electronic nose system (Ghasemi-Varnamkhasti et al., 2012). The division between the N and A groups along PC1 is mainly caused by CP4, CP3 and CP2. For CP2 and CP4, alcohol was already mentioned in the description in Table 1.

CP2 is negatively correlated to CP4 \((r^2 = 0.724)\). Obviously, CP2 and CP3 are not highly correlated \((r^2 = 0.154)\) nor CP3 with CP4 \((r^2 = 0.261)\). Thus only CP2 has a good correlation with CP4. Therefore their effect on the result separately and simultaneously should be different because they are independent from each other. Combination of CP2–CP4 (correlated) and CP3–CP4 (hardly correlated) are plotted. Fig. 5 shows that both beers groups are classified based on the CP2–CP4 (correlated) and CP3–CP4 (hardly correlated) are plotted. In fact, they can even be discriminated based on CP4. The importance of CP2, CP3, and CP4 was already discussed above. That of CP0 is also not surprising when observing Fig. 4b.

In fact, when one or two variables allow the discrimination between both classes, then in fact for this data set the more complex classification methods are not needed. However, we considered these methods. Applying linear discrimination analysis (LDA), 100% classification accuracy was obtained. Using the LDA technique to build two models based on only two variables, i.e. CP2–CP4 and CP3–CP4, respectively, 100% accuracy not only for the training set but also for the test set were obtained.
In this study, in order to evaluate the capability of the entire electronic nose system in predicting the type of beer, prediction models based on PLS-DA and SIMCA were constructed as well. Classification accuracy through SIMCA was found to be 100% and 100% for training and test sets, respectively. Quantitative data of PLS-DA of both beer types are collected in Table 2. Both the calibration and validation values for the alcoholic and non-alcoholic types involved a modeling performance.

Now, the authors wanted to see how well a sensor discriminates between the two groups of beers. For this purpose, it is necessary to fit each sample to both group models. For instance, fit alcoholic brands to the PCA models of both groups. Therefore, the residual matrices are computed, to determine the discrimination power. This criterion was computed according to Otto (2007). Fig. 6 shows the result of the SIMCA in terms of discrimination power of the sensors used in the electronic nose. The larger the value, the higher is the discrimination power. These results confirm the loading plot of PCA as discussed above.

This could be useful information when, for example, one would like to reduce the number of sensors because of some reasons in the fabrication stage of the test chamber of the electronic nose. As already discussed, the instrument builder could choose the most important and effective sensors which have the most discriminating contribution on the beer measurements (Ghasemi-Varnamkhasti et al., 2012).

In this study, accuracy of classification using SIMCA classifier was 100%, 100%, and by PLS-DA, 100%, 100%, for training and test sets, respectively. Both PLS-DA and SIMCA classification results were the same and similar to those obtained by LDA.

Even though the classification accuracies obtained by SIMCA are 100%, 100%, and by PLS-DA, 100%, 100%, for training and test sets, respectively, other computational techniques, for instance, non-linear techniques can be studied as well, but in fact the authors are evaluating the use of a non-linear technique for a situation that is easily linearly separable. Therefore this study shows that sometimes using a complex non-linear or linear method is not needed when a perfect result can be obtained by a simple method.

Support vector machine has become increasingly popular because of its application in different fields, such as data mining, pattern recognition and bioinformatics. Nowadays the application
of SVMs for analyzing chemical datasets has been recommended (Belousov et al., 2002) and a great interest has been shown to this machine learning method for both classification and regression problems. However, the prediction ability of SVM using a Gaussian (Radial Basis Function) kernel function is affected by the optimization of the meta-parameters $C$ and gamma. Several methods can be used to perform this optimization, such as grid search (Hsu et al., 2010) or gradient descent algorithm (Chapelle et al., 2002). In this study, an exhaustive grid search is used because, it is easy and especially for RBF kernel, which requires to optimize two parameters, it can be performed fast (Devos et al., 2009).

Different values of $C$ ranging from 1 to 500 and of gamma ranging from 0.1 to 2 are searched which is shown in Fig. 7. When the values of $C$ and gamma increased, the classification rate decreased. In fact when $C$ is small, margin maximization is bolded due to large margin and smooth boundary while when $C$ is large, the error minimization is predominant because of more complex boundary and smaller margins. However, for the finally selected SVM the number of support vectors (SVs), which correspond to data points on the margin borders was 14, $C$ 100, gamma 0.19. Full accuracy (100%) for both training and test sets was again achieved for the beer classification.

In summary, the result of LDA showed the classification can be done in a simple way and that such a complex SVM model is not needed. For the SVM, as can be seen higher, several parameters should be optimized which is also a challenge. This work strongly suggests performing a simple classifier for classification purposes instead of using immediately a complicated method when the simple method is appropriate. When the results are insufficient, then another technique can be used.

At the end of this paper, it is worth mentioning the results obtained in this study showed the capability of the electronic system to distinguish between alcoholic and non-alcoholic beers. As emphasized in the literature (Bamforth, 2006; Briggs and Brookes, 2004; Hardwick, 1994), alcohol has a significant contribution in
flavor attributes of beer in such a way the aroma compounds, not only in terms of volatile compounds composition but also in concentration, are different in both types of beer. This means the electronic nose used in the current study could recognize these differences or the differences in alcohol content. Signal changes caused by alcoholic beers were larger than those of the non-alcoholic beers which this is in close agreement with the findings of Bartolome et al. (2000) who underscored that non-alcoholic beers have a weaker aroma than alcoholic beers. They suggested it seems reasonable to suspect that the dealkoholization processes might also affect the phenolic composition of the beers. Getting discrimination capability with highest accuracy is more important when the electronic nose would be used to recognize among the beer samples with low differences in aroma. For instance, in the brewery fermentation stage and after that, aging is a very critical step for obtaining the final beer quality. In these stages and even after packaging, the changes in aroma of beer may be very little but important in the generation of off-flavors in beer. An unexpected off-flavor in beer is always a critical problem in brewery marketing and the governing commercial rule is that beers have to be free of defects (Ragazzo-Sanchez et al., 2009; Ghasemi-Varnamkhasti and Forina, 2014). Detection of off flavor may involve detection of one or several compounds normally absent in beer or the presence of normal flavor components in excessive concentrations. Having an electronic nose system with the highest capability in classification might be helpful to detect these changes. At the present time, many researchers are working to rectify the problems limiting the capability of the electronic nose (Bruins et al., 2013; Gutiérrez and Horrillo, 2014) and the extensive application of these instruments will be occured in the breweries.

5. Conclusion

The necessity of electronic nose applications in brewery has been increased because of the reliability and speed of analysis of such tools. Therefore, it sounds they are suitable alternatives for quick and precise analysis of beers and even to quality authentication in the production lines. From this study, the PCA results visualize that CP2, CP3 and CP4 are the most effective variables for classification. When the dataset is not complex, using a simple method, such as LDA, can be preferred against complicated techniques to discriminate between different classes. The electronic nose used in the current study, showed to have the capability to perform the classification of alcoholic and non-alcoholic beer samples in a fast and reliable way. SIMCA and PLS-DA showed the same results for the binary discrimination of the beer types. To enhance the distinguishability between alcoholic and non-alcoholic beers, SVM was used as an advanced computational technique by which a full accuracy in classification was obtained. However, when discrimination can be done with a simple method such as LDA, using SVM is not needed. The promising results of this study indicated the capability of the electronic nose used to detect and differentiate the beer aromas. Therefore, this system will be used in other parts of our project, i.e. to measure aging fingerprints in beers.

Acknowledgments

The authors would like to thank the supports of Université de Lorraine-Metz (France) and University of Tehran (Iran) for this research. The constructive comments and kind helps of Dr. Goodarzi are also acknowledged.

References


Ghasemi-Varnamkhasti M. Design, development and implementation of a metal oxide semiconductor (MOS) based machine olfaction system and bioelectronic tongue to quality change detection of beers coupled with pattern recognition analysis techniques [PhD dissertation]. Karaj, Iran: University of Tehran; 2011.

Fig. 7. The optimization grid (for C and gamma) results for classification rate using 5-fold cross validation.


