Computers and Electronics in Agriculture 150 (2018) 402–410

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Optimizing prediction of human assessments of dairy odors using input variable selection

Fangle Chang[a], Paul H. Heinemann[b,*]

[a] Department of Agricultural and Biological Engineering, The Pennsylvania State University, 105 Agricultural Engineering Building, University Park, PA 16802, USA
[b] Department of Agricultural and Biological Engineering, Pennsylvania State University, 105 Agricultural Engineering Building, University Park, PA 16802, USA

ARTICLE INFO

Keywords:
Odor evaluation
Electronic nose
Artificial neural networks
Feature selection
Prediction

ABSTRACT

Use of instruments instead of human panels to assess odors can make the collection and measurement process more efficient and reliable. Odor-emitting samples from dairy farms, including manure, feed, and bedding materials, were collected and assessed by an electronic nose and a human panel. Artificial neural networks based on the Levenberg-Marquardt Back-propagation algorithm were used to build prediction models to predict human response to odor pleasantness. Feature selection methods, including Forward Selection (FS), Gamma Test (GT), and Principal Component Analysis (PCA), were applied to reduce the dimensionality of the measurements, potentially eliminating noise. Out of the 28 variable candidates (eNose sensors), 10 variables were selected when PCA was applied, and 16 variables were selected when either FS or GT approaches were applied. The model developed using GT provided the lowest mean square error of 0.56 (2.5%) hedonic scale units for separate validation. The GT-based model was able to predict the human assessments within 10% of the target for 81% of the independent validation samples and within 5% of the target for 63% of the independent validation samples.

1. Introduction

Odor emitted from dairy operations may raise living quality concerns of the farm neighbors. To balance the rights of producers and non-farmer neighbors, odor evaluation and mitigation may be required. In some states, agricultural odors are regulated by state and local government (Wal, 2001). The human nose is a valuable tool for odor detection. However, it is difficult to describe the odors (over 10,000) that humans can detect; there are over 160 associated with agriculture (O'Neil and Phillips (1992)). Wheeler et al. (2012) describe a method to qualitatively categorize odors related to an offensiveness level using a general hedonic (pleasantness) scale. A trained panel assesses odors using the scale, which ranges from −11 (extremely unpleasant) to +11 (extremely pleasant). Experimental results showed assessments of the same odor by different people can be vary considerably, including both pleasant and unpleasant responses to the same odor. However, trained panelists are more likely to give individually consistent responses for the same odor.

The use of instruments to predict human assessments has many advantages, such as avoiding the highly subjective nature of human perception, reducing the cost of human panels, and saving time. Two general approaches used to measure odors are utilizing instruments to detect odorous gas concentrations and human assessments. Because the relationship between the presence of gaseous compounds and odor is not obvious (Ostojic and O'Brien, 1996), human sensory methods, such as olfactometry and scentometer (Brandt et al., 2011; Henry et al., 2011), are the most commonly used. Trained panelists are expected to provide consistent responses to an odor sample's intensity and hedonic tone. Using humans has some weaknesses, however; getting several people together on a short notice can be difficult, and the high subjectivity of human sense of smell can reduce the reliability of the results. Research and development of new, unbiased, and lower-cost measurement methods is ongoing.

An electronic nose (eNose) is considered to be a proven device to classify volatile sample patterns. The device is made of multiple electronic sensors. Various sensors can be employed based on specific needs. In most studies, the measurements from electronic noses did not show correlation with human responses (Wheeler et al., 2012). Additional tools may be needed to connect the output of the instruments with odors. Artificial Neural Networks (ANNs), compared with traditional networks, can handle non-linear data and outliers without losing the overall relationship (Smith, 2003). Combining ANNs with an eNose can visualize odorant characteristics using mathematical concepts and language. ANNs may be used to model difficult phenomena, and it may be possible to leverage ANNs in modeling human assessments of odors (Sohn et al., 2006).

* Corresponding author.
E-mail address: hzh@psu.edu (P.H. Heinemann).

https://doi.org/10.1016/j.compag.2018.05.017
Received 13 December 2017; Received in revised form 17 March 2018; Accepted 13 May 2018
0168-1699/

Electronic sensors used for data collection are sensitive to particular components. In the case of datasets with large numbers of input variables, irrelevant, redundant, and noisy information may be included (Noori et al., 2009). Such information can be discarded without the loss of significant information in the dataset (King and Jackson, 1999). However, the most common disadvantages for modeling using AI techniques (including ANNs) are that they do not determine the best subsets of inputs (Noori et al., 2011; Kecman, 2005). Because of this, input selection methods are recommended for the development of ANN prediction models. Unlike dimensionality reduction, which tends to create new combinations of attributes, the feature selection methods usually choose and discard the attributes without changing them in the dataset (Kohavi and John, 1997).

Jollife (1972) indicated that in multivariate analyses, more efficient subsets can be chosen when the variable number is larger than 10, and any variable can be selected from several variables if these variables are correlated. There are different methods for feature selection, such as principal component analysis (PCA), forward selection (FS) (Mao, 2002), Support Vector Machine (Degroeve et al., 2002), and Gamma test (GT) (Agalbjorn et al., 1997; Tsui et al., 2002). In this work, three variable selection methods, PCA, FS, and GT, were used. PCA has been widely applied to classification and regression analysis for feature selection, and the application in prediction models is becoming prevalent (King and Jackson, 1999; Lu et al., 2007; Noori et al., 2011; Noori et al., 2010a). Choi and Park (2001) compared the accuracy of two methods, multivariable linear regression (MLR) and ANNs, to predict the influent concentration of total kjedahl nitrogen (TKN, sum of organic nitrogen) wastewater irrigation. In both methods, inputs composed of original data and PCA data were examined through the models. The results showed the superiority of ANN models with PCA. Normally, the FS method is used in linear regression. The application in prediction models has been successfully developed (Wang et al., 2006; Noori et al., 2010b; Noori et al., 2011; Prakash et al., 2012). However, few studies showed the use of the combination of FS with ANNs to predict human assessments. The application of GT to animal odor evaluation has not been reported since its introduction by Agalbjorn et al. (1997). The GT is also called the near neighbor test. Through finding the nearest neighbor of every point, the subset with the best possible performance can be estimated without the need to train the model. A few studies involving the application of this method in variable selection (Corcoran et al., 2003; Noori et al., 2010b; Noori et al., 2011) showed high estimation accuracy. Corcoran et al. applied GT to enhance the accuracy of ANN models for predicting geo-temporal variations of crime and disorder. The results showed noise was successfully excluded using GT.

Chang (2016) and Chang and Heinemann (2018) combined electronic nose measurements of dairy odors with human assessments utilizing three types of neural network architectures to predict hedonic tone of the odors. The optimal number of neurons and architecture type for best prediction accuracy was determined.

The overall goal of this work was to develop an odor assessment system that would provide a prediction of hedonic tone in the same way a human panel would, but without requiring the use of the actual human panel. Such a successful system would allow for rapid evaluation of odor quality. This paper presents the further refinement of the prediction system based on odor sources emitted from dairy operations utilizing data reduction techniques. Human responses were the model targets, and the measurements by instrument were the model predictors. Feature selection techniques were applied to improve the predictive accuracy. The objectives of this study were to:

(1) Develop ANN models utilizing subsets determined by variable selection methods (PCA, FS, and GT) for prediction of human assessments based on odor from dairy farms.
(2) Compare the accuracy between the original data set model with the models developed from variable selection model subsets, through mean square error and correlation coefficient analysis.

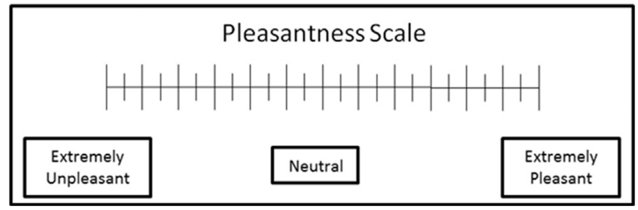

Fig. 1. Pleasantness scale (Wheeler et al., 2012).

2. Development of variable selection techniques

2.1. System design and data collection

Human perception of odor was measured using a general hedonic (pleasantness) scale (Fig. 1). The middle point presents neutral (0), from which to the right side the pleasant feeling increases until it reaches the extremely pleasant point (+11), and the left side represents the opposite feelings (−11).

Nine solid samples that included manure, bedding, and feed materials were collected in dark glass jars on five sampling occasions; three from different commercial farms and two from the Penn State dairy barns (Chang and Heinemann, 2018). These 47 samples were representative of the array of odors emitted from these operations. One extra sample was collected from sampling occasions three and five to widen the pleasantness range. The samples were placed in 1 L amber glass bottles, and transported to the assessment lab. Five trained panelists assessed the odor and provided a response to each sample for three cycles, and the average of these five assessments for each sample provides one target. Glass syringes were used to extract odor from the sample jar headspace and were presented to each panelist. The targets were developed into a vector with 141 observations from all five sampling occasions.

The Cyranose 320 (Sensigent, Baldwin Park, CA) electronic nose was used for this work. The Cyranose 320 consists of an array of 32 internal polymer sensors. The sensors swell and retract based on the absorption of particular volatile compounds. A current runs through each sensor, and the swelling and shrinking of the sensors creates a change in resistance. This change in resistance ($R_{Max} - R_{Baseline}$) is then divided by the initial baseline resistance ($R_{Baseline}$). Water vapor was present in the samples, but water vapor does not contribute to the odor array, which can add noise to the instrument measurements. Li et al. (2007) and Williams et al. (2010) found that four sensors (5, 6, 23, and 31) within the eNose were sensitive to water vapor and were therefore shut down during the experiment. Measurements from four sampling occasions were used for training of the neural networks, and measurements from the fifth sampling occasion were used for separate validation.

2.2. Data reduction methods

The dataset created by the eNose measurements contained 141 observations with 28 variables (sensor readings), in which 114 observations (80%) from four farm sampling occasions were used for training, and 27 observations (20%) from the fifth sampling occasion were used for separate validation. This provided a sufficient number of samplings for training, and enough independent validation values for statistical purposes. The large number of the variables potentially reduces the accuracy of the prediction models, so determination of optimal subsets composed of the principal sensor readings was considered an efficient solution for the improvement of the prediction accuracy. An optimal subset does not necessarily need to be unique, because similar accuracy may be achieved using different sets of features (e.g., the feature can be replaced when another feature is perfectly correlated to it) (Kohavi and John, 1997). Most of these methods select variables by