Forensic Science International 236 (2014) 84–89

Contents lists available at ScienceDirect

Forensic Science International

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

Application of self-organizing feature maps to analyze the relationships between ignitable liquids and selected mass spectral ions

Jessica L. Frisch-Daiello, Mary R. Williams, Erin E. Waddell, Michael E. Sigman *

National Center for Forensic Science and Department of Chemistry, College of Science, University of Central Florida, P.O. Box 162367, Orlando, FL 32816-2367, United States

ARTICLE INFO

Article history:
Received 31 May 2013
Received in revised form 18 December 2013
Accepted 20 December 2013
Available online 7 January 2014

Keywords:
Ignitable liquids
Fire debris
Self-organizing feature maps
Gas chromatography–mass spectrometry
Total ion spectrum
Extracted ion spectrum

ABSTRACT

The unsupervised artificial neural networks method of self-organizing feature maps (SOFMs) is applied to spectral data of ignitable liquids to visualize the grouping of similar ignitable liquids with respect to their American Society for Testing and Materials (ASTM) class designations and to determine the ions associated with each group. The spectral data consists of extracted ion spectra (EIS), defined as the time-averaged mass spectrum across the chromatographic profile for select ions, where the selected ions are a subset of ions from Table 2 of the ASTM standard E1618-11. Utilization of the EIS allows for inter-laboratory comparisons without the concern of retention time shifts. The trained SOFM demonstrates clustering of the ignitable liquid samples according to designated ASTM classes. The EIS of select samples designated as miscellaneous or oxygenated as well as ignitable liquid residues from fire debris samples are projected onto the SOFM. The results indicate the similarities and differences between the variables of the newly projected data compared to those of the data used to train the SOFM.

1. Introduction

The total ion spectrum (TIS), defined as the time-averaged mass spectrum across the chromatographic profile [1], has previously been used for classification and discrimination of fire debris samples and prediction of classification error rates [1–4] based on the classes defined in the American Society for Testing and Materials (ASTM) standard E1618-11 [5]. Unlike the total ion chromatogram (TIC), the TIS is time-independent, easily allowing for inter-laboratory comparisons. In this work, a subset of ions chosen from Table 2 of ASTM E1618-11 is used to generate extracted ion spectra (EIS). These EIS are then analyzed by the artificial neural networks method of self-organizing feature maps (SOFMs). For this work, the method is used to determine if the unsupervised clustering of time-independent EIS data results in groupings of the data according to their ASTM class designations.

Ignitable liquid residues are classified by analysts in the United States according to the standard ASTM E1618-11 by pattern recognition of the TICs, extracted ion profiling, and target compound analysis [5]. Ignitable liquids are largely comprised of six major types of compounds: normal alkane, branched alkane, cycloalkane, aromatic, polynuclear aromatic, and oxygenates [5]. Characteristic ions compiled in Table 2 of ASTM E1618-11 are used for extracted ion profiling [5]. The chromatographic profile pattern along with the presence and relative amounts of the major types of compounds are used to classify ignitable liquids into seven major classes [5]: gasoline (GAS), petroleum distillates (PD), isoparaffinic products (ISO), aromatic products (AR), naphthenic–paraffinic products (NP), normal alkane products (NA), and oxygenated solvents (OXY). Those ignitable liquids that do not fall into one of the aforementioned classes or may fall into multiple classes are assigned to a miscellaneous category (MISC) [5]. Subdivisions of the classes, with the exception of the gasoline class, are based on the range of *n*-alkanes present and are defined as light (L), medium (M), and heavy (H) corresponding to C_4–C_9, C_8–C_{13}, and C_9–C_{20+}, respectively [5].

Visual pattern recognition is dependent on the interpretation of the chromatographic data by the analyst [1,6] and does not readily lend itself to automation; therefore, non-subjective chemometric techniques, which involve the extraction of information from chemical data, have been explored [7]. Many chemometric techniques allow for data visualization by reducing the dimensionality of the data [4,8] and clustering or grouping of data based on similar characteristics [9]. Kohonen's SOFMs also known as self-organizing maps (SOMs), is an artificial neural networks technique designed as a method for abstraction, clustering, and visualization of high-dimensional data [10–12]. This is usually accomplished through nonlinear mapping onto a two-dimensional grid space of a predefined

* Corresponding author. Tel.: +1 407 823 3420.
E-mail addresses: jessica.frisch-daiello@ucf.edu (J.L. Frisch-Daiello), mary.williams@ucf.edu (M.R. Williams), ewaddell@knights.ucf.edu (E.E. Waddell), michael.sigman@ucf.edu (M.E. Sigman).

0379-0738/$ – see front matter © 2014 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.forsciint.2013.12.026

number of neurons, where each neuron contains a weight vector that is comprised of the same number of components as the number of variables (or dimensionality) of the input space [13]. The weight vectors are adjusted through an unsupervised learning process resulting in the neurons being arranged according to patterns in the input signal [11]. When a sample is introduced to the map, a distance is calculated between the sample variable vector and the weight vector for each neuron, where the Euclidean distance is often used. The neuron with the smallest distance to the sample is determined to be the "winning neuron", and the weight vector for that neuron is adjusted to be more similar to the sample vector. The weight vectors of neighboring neurons that are a given distance from the winning neuron are also updated, which preserves the neighborhood relationships of the data within the input space [13]. This training process is repeated for each of the samples resulting in the completion of one full cycle or epoch, and this process is repeated iteratively for a predefined number of epochs [14]. A batch algorithm may also be used [10,15]. The amount of adjustment to the weight vectors and the neighborhood size are varied throughout the training process [16]. In the last step of training, the input data is re-introduced to the grid, and the samples are mapped onto the winning neurons [14]. This process results in a two-dimensional map of neurons, where sample data is clustered based on similarity while preserving the neighborhood relationships of the data [8,10,15].

SOFMs have previously been used for clustering and/or classifying samples analyzed by GC–MS. Strawberry varieties were studied using solid-phase micro-extraction GC–MS data and clustered according to type by SOFM [17]. Crude oil [18] and weathered crude oil samples [14] were geographically classified. Weathered and unweathered lighter fluids were classified based on manufacturer [6], and classification according to product type and brands were performed for weathered and unweathered medium petroleum distillates [19]. A set of 150 ions resulting from pyrolysis mass spectrometry was analyzed with SOFMs to classify plant seeds [16]. While SOFMs have been used for classification, Kohonen points out that self-organizing feature maps are a beneficial unsupervised method for clustering, visualization, and abstraction, but are not meant for statistical pattern recognition [10]. A method that is "particularly suitable for statistical pattern recognition" is a supervised version of SOFM known as Learning Vector Quantization (LVQ) [10]. In addition to clustering, a SOFM allows visual associations between individual variables and clustered samples in order to determine feature attributes for natural clusters within the input data space.

In this work, 313 ignitable liquids from varying ASTM classes are grouped using the unsupervised SOFM technique, and ignitable liquid residues are analyzed based on the SOFM model. The natural clusters of the input data and the relationships between these clusters and their spectral variables are examined with respect to their ASTM class designations. Many chemometric techniques for analyzing ignitable liquids have utilized TIC data [6,14,17–19]; however, in this work, the EIS are utilized to make the results applicable across multiple laboratories by overcoming inter-laboratory retention time shifts that may occur using chromatographic data [1]. The SOFMs are created using these EIS of primarily unweathered and some weathered samples obtained from data in the Ignitable Liquids Reference Collection (ILRC) [20]. The EIS of select samples designated as MISC or OXY, as well as ignitable liquid residues from fire debris samples, are projected onto the SOFM demonstrating the similarities and differences between the variables of the newly projected data compared to those of the data used to train the SOFM.

2. Materials and methods

2.1. Samples

The SOFM training dataset consisted of EIS for liquids in the ILRC database, which were designated as GAS, AR, ISO, NA, NP, LPD, MPD, HPD and OXY ASTM classes. The OXY samples used for training contained predominantly oxygenated compounds and did not have chemical characteristics of other ASTM classes. The ILRC was created and is maintained by the National Center for Forensic Science (NCFS) in collaboration with the Scientific Working Group for Fire and Explosions (SWGFEX). Classification of the ignitable liquids in the ILRC was performed by a committee of practicing fire debris analysts. The training dataset used in this work was comprised of 313 ignitable liquid samples including 289 unweathered ignitable liquid samples as well as 24 weathered GAS samples. The training samples designated as ISO were further sub-classified as light, medium, and heavy (LISO, MISO, and HISO, respectively) based on their carbon range. AR samples were also further sub-classified as light and medium (LAR and MAR) based on their carbon range. The fire debris dataset contained 116 EIS of ignitable liquid residues extracted from fire debris samples that were produced in laboratory and large-scale burns. The samples were extracted following the ASTM E1412-07 method [21], and the ignitable liquid residue patterns were observed in the chromatograms. The class designations for the fire debris samples corresponded to the ASTM class of the unweathered ignitable liquid used in the burn. A third dataset was compiled, where the samples had chemical characteristics of multiple ASTM classes. This dataset was comprised of 33 MISC and seven OXY samples from the ILRC. Sample preparation and instrumental analysis for all of the datasets are described in previous work [3].

The EIS for all datasets were comprised of 29 ions chosen as a subset from Table 2 of ASTM E1618-11 [5]. These ions which are typically produced in the EI mass spectrum are listed in Table 1 according to their mass-to-charge (m/z) ratios and their corresponding compound type. These ions represent compound types commonly observed in ignitable liquids within the seven major ASTM classes. Each EIS was normalized so the most abundant ion had an intensity of one.

2.2. Self-organizing feature map

2.2.1. Training of the self-organizing feature map

SOFM calculations were performed using the Neural Network Toolbox™ 7 with MATLAB R2011b (MathWorks, Natick, MA, USA). Several hexagonal grid sizes were investigated. The 15×15 unit hexagonal grid resulted in the minimal number of neurons being associated with more than one ASTM class. A total of 112,500 epochs, 500 times the number of neurons were calculated, where the ordering phase of learning consisted of 1000 epochs and an initial neighborhood size of 13 [11]. Therefore, in the first epoch of the ordering phase, the winning neuron and the neurons within a 13 neuron radius from the winning neuron had their weight vectors adjusted. The neighborhood size decreased from 13 to one as the number of epochs in the ordering approached 1000. The subsequent tuning phase had a neighborhood size of one, so only the weight vector of the winning neuron was adjusted. The default batch training method in the Matlab Neural Network Toolbox™ 7 was used, where for each epoch, the whole dataset was presented to the network. The winning neuron for each input vector was determined, and the weight vectors were moved to the average position of all input vectors for which the weight vector was a winner or in the neighborhood of a winner [15]. In the batch training mode, a learning-rate parameter was not used [10].