Journal of Analytical and Applied Pyrolysis xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Analytical and Applied Pyrolysis

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

# Forensic comparison of pyrograms using score-based likelihood ratios

Agnieszka Martyna[a,*], Grzegorz Zadora[a,b], Daniel Ramos[c]

[a] University of Silesia in Katowice, Institute of Chemistry, Department of Analytical Chemistry, 9 Szkolna, Katowice 40-006, Poland
[b] Institute of Forensic Research in Krakow, 9 Westerplatte, Krakow 31-033, Poland
[c] Audias: Audio, Data Intelligence and Speech, Escuela Politecnica Superior, Universidad Autonoma de Madrid, C/Francisco Tomas y Valiente 11, 28049 Madrid, Spain

ARTICLE INFO

*Keywords:*
Comparison/classification
Likelihood ratio
Data dimensionality reduction
Pyrolysis gas chromatography
Forensic science

ABSTRACT

The comparative analysis of chromatographic profiles of materials is the subject of interest in many scientific fields, including forensic science. Plastic microtraces collected during hit-and-run accidents and examined with pyrolysis gas chromatography mass spectrometry (Py-GC–MS), may serve as an example. The aim of comparing the recovered and control samples is to help reconstruct the event by commenting on their common, or not, sources. The objective is to report the evidential value of data in the context of two competing hypotheses: $H_1$ – both samples share common origins (e.g. car) and $H_2$ – they do not share common origins. The likelihood ratio approach (LR) addresses this idea as an acknowledged method within the forensic community. However, conventional feature-based LR models (using e.g. signal intensities of the chromatographically separated compounds) suffer from the curse of multidimensionality. Their considerable complexity can be reduced in the score-based LR models. In this concept the evidence expressed by the score, computed as a distance between the recovered and control samples characteristics, is evaluated using LR. A score solely based on a distance or a measure of similarity, without taking into account typicality, may not reflect the differences between similar samples clearly in a highly multidimensional space. Here we show that boosting the between-samples variance (**B**) whilst minimising the within-samples variance (**W**) helps distinguish between samples and improves the score-based LR models performance. Instead of computing the distances in the feature space, the authors use the space defined by ANOVA simultaneous component analysis, regularised MANOVA and ANOVA target projection that find directions with the magnified differences between **B** and **W**. The concept was successfully illustrated for 22 plastic containers and automotive samples, examined using Py-GC–MS. The research shows that this so-called hybrid approach combining chemometric tools and score-based LR framework yields a performing solution for the comparison problem for Py-GC–MS chromatograms.

## 1. Introduction

Chromatography plays an important role in the forensic evidence analysis to detect the organic compounds. It is either used to identify the unknown substances, e.g. drugs, or to record chromatographic profiles (usually from the pyrolysis gas chromatography) of the microtraces of plastics, automotive paints, tires, fire debris, explosives and fibres. In the latter case the chromatograms are usually recorded for two samples, namely recovered and control. Then the task is to compare them and assess whether they may be two pieces of the same object. This issue is known as the comparison task.

In the era of developing society, the road transport holds an important place. This is also a subject of interest in the forensic field, where the experts frequently face the problem of inferencing in the car accidents cases. Among many questions arising in the hit-and-run car accidents, experts may be asked to find the connections between the scene of the car accident and the car driven by the suspected perpetrator of the accident. The task is resolved by comparing the physicochemical characteristics, e.g. chromatograms, of the material recovered from the scene of the car accident, collected in the form of microtraces of glass, automotive paints or plastics used for car body elements production (e.g. bumpers) and control material from the suspected car. Then the question is raised whether the recovered material (e.g. found on the victim clothes) and control material may have come from the same source (i.e. suspected car) or not [1,2]. Such considerations refer only to the common or separate source of data collected for case assessment. This so-called *source level* is the first step within the *hierarchy of propositions*, which backbone embodies the source, activity and offence levels [3–5]. In the comparison problem of plastics collected in hit-and-run car accidents investigation such source-generic hypotheses can be expressed as:

---

* Corresponding author.
*E-mail addresses:* agnieszka.martyna@us.edu.pl (A. Martyna), gzadora@ies.krakow.pl (G. Zadora), daniel.ramos@uam.es (D. Ramos).

https://doi.org/10.1016/j.jaap.2018.03.024
Received 12 January 2018; Received in revised form 29 March 2018; Accepted 29 March 2018

Please cite this article as: Martyna, A., Journal of Analytical and Applied Pyrolysis (2018), https://doi.org/10.1016/j.jaap.2018.03.024

- $H_1$: compared recovered and control plastic fragments come from the same vehicle body.
- $H_2$: compared recovered and control plastic fragments come from two different vehicles.

In the forensic experts practice the chromatographic profiles are typically compared visually for detecting the similarities and discrepancies between the location and shape of the leading peaks. Such a naked-eye comparison can only be credible for visually distinguishable profiles. For highly similar chromatograms this approach lacks the objectivity and precludes expressing the degree of similarity in a quantitative manner. For objectifying the methodology the analytical results must be interpreted and reported according to the recommendations of the interpretation schemes acknowledged in the forensic sciences [6], i.e. using the likelihood ratio approach (LR) [1,2,7]. This methodology provides a way for expressing the evidential value of the compared profiles in a reliable manner in view of two contrasting hypotheses ($H_1$ and $H_2$). Generally, the LR is computed as the probability of data characterising the evidence $E$, given the propositions, $H_1$ and $H_2$:

$$\mathrm{LR} = \frac{\Pr(E|H_1)}{\Pr(E|H_2)}. \tag{1}$$

$H_1$ is supported by the LR values larger than 1 and the support is strengthening with increasing LR. Conversely, the $H_2$ is more likely when LR is below 1 and the support for this hypothesis reinforces with the LR values approaching 0. Both hypotheses are equally likely when $LR = 1$. The LR models reliably express the evidential value by accounting for:

- similarities and discrepancies between the physicochemical data of the compared samples;
- typicality (rarity) of the data. Observing the similarity between rare features assigns greater evidential value than when rife features demonstrate similar resemblance;
- the sources of uncertainty including the within- and between-samples variability computed from the relevant background population. If we are comparing two plastics of a particular kind, the background population is the set of chromatograms of plastics of that kind, recorded using the same methods as for the evidence material. If all replicate measurements for an object (e.g. car bumper) form a sample, then the between-samples variance is the variation of the averages of the samples. The variation of the replicates within each sample, that is averaged over all samples, represents the within-samples variance;
- statistical dependencies between the measured variables/features.

Considering the above aspects and viewing the data in the context of two contrasting, but equivalent, hypotheses, makes LR approach more suitable for forensic data interpretation than the typically used statistical tests (e.g. *t*-test) or chemometric methods. Also, it follows the rules of probability, and integrates in a Bayesian decision framework in a natural way, allowing straightforward decision-making (Eq. (2)) [8,9]. The Bayesian theory can be seen as an illustration of the trial course. The prior assumptions ($\Pr(H_1)$ and $\Pr(H_2)$) about the hypotheses ($H_1$ and $H_2$) stated before the evidence analysis are modified by the LR values computed after collecting more information in the course of the evidence examination. Prior assumptions updated by LR values are expressed in the form of ratio of posterior probabilities $\Pr(H_1|E)$ and $\Pr(H_2|E)$. These probabilities are the basis for a further decision by the fact finder.

$$\frac{\Pr(H_1)}{\Pr(H_2)}\cdot\frac{\Pr(E|H_1)}{\Pr(E|H_2)} = \frac{\Pr(H_1)}{\Pr(H_2)}\cdot\mathrm{LR} = \frac{\Pr(H_1|E)}{\Pr(H_2|E)}. \tag{2}$$

A large body of literature exists [1,7,10–15] in which the solutions to the comparison problem of physicochemical data using the LR approach were developed. A great majority of them use the measured features of the samples to construct the feature-based LR models, which require that:

(i) the number of samples substantially exceeds the number of variables they are described by to make the matrix algebra feasible. Hence highly multivariate data, such as chromatograms, delivering thousands of variables understood as signal intensities measured in time as the elution process continues, need some data dimensionality reduction to make the LR models computationally available;
(ii) the (co)variance within each sample is constant and (multi)variate normal;
(iii) the average variance of the replicate measurements within each sample is much lower than the variance of samples averages. This condition ensures that the samples are easily distinguished i.e. replicate measurements for each sample are recognised as same-source, while these from different samples are indicated as coming from different sources.

The obvious solution to (i), which is examining more samples than the number of variables, involves considerable time and money consumption and it is often infeasible. Practical way for muddling through the problem is by compressing the data dimensionality. Apart from the principal component analysis, the most efficient data dimensionality reduction is by moving from feature representation to the (dis)similarity (or score) representation [16–19]. In feature representation both compared samples are characterised by a set of parameters (features/variables), which resemblance, variability, rarity and dependencies is then studied using the feature-based LR. Contrary to that, in score representation individual multivariate observations are replaced by the pairwise measure of their mutual (dis)similarity and possibly typicality, using the scores [19,20]. The so-called score-based LR approach is then used for studying whether pairwise scores between observations support the hypothesis that they originate from the same source ($H_1$) or different sources ($H_2$). The concept of score-based LR models simplifies the typical approach for solving the comparison problem in the feature space found in [1,7,10–15]. If the typicality is skipped, the (dis)similarity is simply defined by the distances between observations, which are computed in the same way for common or rare features. As has been noted in recent literature [20,21], this may lead to a severe loss of information and degradation of the discrimination abilities of the score-based LR models. However, for forensic likelihood ratios, calibration is a critical measure of performance to be considered beyond discriminating power [1,22–24], and good calibration can be achieved by using distance-only models. Consequently, sometimes distance-only models outperform feature-based models, despite the loss of discriminating power. Nevertheless, it is recommended to include typicality information in any score.

An additional advantage of the score-based LR models is that the requirement (ii) is of no importance for computing the distances. However, the concept of score-based LR models is reasonable only when the features are much closer to each other among observations from the same source than between different sources. This is equivalent to requirement (iii), thus the condition to keep greater between-sample variance than the within-sample variance for the features still holds for score-based LR models.

The comparison task that needs LR models to be engaged focuses on the data that are visually hardly distinguishable. As a consequence of huge chemical similarity of the studied polymers, which contain mostly the same constituents after the pyrolysis degradation, the analysed chromatograms differ only in small time ranges and a substantial part of each chromatogram is identical throughout the database. For this reason the variance of the majority of the variables is comparable within and between samples. The concept of finding the directions along which the data within each sample are similar and differ between samples is easily accomplished using chemometric techniques. In this