Automated GC/MS Characterization of Ball Point Pen Inks by Differential Analysis and Predictive Modeling

Brain Brink, Takeshi Saner, Anthony Gray, Agilent Technologies, Santa Clara, CA

Introduction

Forensic science has an interest in determining the age of an ink sample on paper in the course of various types of investigations. Our study brings together several analytical techniques into an automated workflow to facilitate the characterization of ball point ink. Sample introduction and data acquisition is fast and easy through the use of Thermal Desorption Probe (TSP) and GC/MS. Sample classifications can be automatically and statistically predicted using a class prediction model without the need of user intervention. Statistical methods are useful in analysis as the profiles of the various inks are complex and change over time after they have been applied to paper. This makes differentiating the sometimes subtle differences a difficult task which we show can be simplified by using integrated workflow tools of chemometrics.

Experimental

Sample Prep

Square samples of 28 lbs. copper paper were prepared using a plotter. The samples were approximately 1.0 cm square. Several lines of ink were scribbled on the paper sample with one of three ball point pens. The pens were rolled up and inserted into TSP microvials. Whereupon the samples were also prepared to act as unknowns to evaluate the predictive models.

Data Analysis

Automated Mass Spectral Deconvolution and Identification System (AMDIS) is used for spectral extraction and deconvolution. Differential analysis is performed using Mass Profiler Professional 12.0 (MPP). Predictive models with various algorithms are generated and then ESP 2.0 was used in conjunction with DivaStation E.02.01 SP1 to automate workflow from data collection to final predictive report.

MPP Parameters

- Mass Accuracy: 10.006 E000
- RT Tolerance: 0.05 min
- Match Factor: 0.3

GC/MS Analysis

All samples were analyzed on the same Agilent 7890/5975C GC/MS System. The samples were introduced into the GC/MS using a TSP attached to a Multimode inlet (MMI). The temperature programming capability of the MMI allowed for consistent introduction of the samples to the GC system. The pertinent instrument parameters and conditions are here outlined:

- Columns: DB-5HT 15 m, 1.0 μm film, He
- Injection: Initial 100°C held 6.1 min
- Ramp at 60°C/min to 280°C
- Detector: Initial 70°C held 1.5 min
- Ramp at 10°C/min to 330°C
- Hold for 1.5 min

- Source: 250°C
- Quadrupole: 150
- Source: 240

Scan Range: 40-570 m/z

- Sensitivity: Medium
- Shape requirements: Medium
- Gain factor: 1
- Sampling: 2
- Match Factor: 0.3

Experimental

The workflow used in the creation of the class prediction model is described in Figure 2. After acquisition, a library was created using AMDIS. The AMDIS sample reports were imported into MPP and the Significance Testing and Full/Check Change workflow workflow was followed to generate an entity list. The entity list was filtered to increase the reliability of class prediction model creation. The filtering is outlined in Figure 3 and consists primarily of removing entities that were present in blank paper samples. Further filtering was performed using one way ANOVA with a p-value <0.15. This regimen of filtering reduced the number of entities from 341 to 87 entities of interest.

Results and Discussion

Results and Discussion

Five sample class prediction models were created based on different statistical methods. The model based on Support Vector Machine was added to the data analysis methods in DivaStation. This allowed for the automation of the sample reporting with the only user intervention at the time of sample introduction and entry into the sequence list.

Table 1: Summary of results from testing various prediction models for five unknowns.

<table>
<thead>
<tr>
<th>Model Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>P00</td>
</tr>
<tr>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>Decision Tree</td>
</tr>
<tr>
<td>Naive Bayes</td>
</tr>
<tr>
<td>Neural Network</td>
</tr>
<tr>
<td>Hierarchical Clustering</td>
</tr>
<tr>
<td>One way ANOVA (p-value &lt;0.15)</td>
</tr>
</tbody>
</table>

The evaluation of the different class prediction models was summarized in Table 1. Five different unknowns were analyzed using this automated workflow. Of the five models evaluated, two were able to correctly predict all five samples with varying confidence levels, i.e. Partial Least Squares Discrimination, and Support Vector Machine. Partial Least Squares Discrimination provided the best confidence levels ranging from 95% to 85% for the unknowns. Neural Network and K-nearest cases correctly identified 2 of the 5 unknowns while Decision Tree was successful in correctly identifying 4.

Conclusions

We demonstrate that by using statistical class prediction modeling we can differentiate ink samples from different types of pens and correctly predict the amount of time since the ink was applied to the paper. The analysis workflow was automated to improve reliability and ease of use.

To maximize the accuracy of prediction, the quality of the data is crucial. ESP provides the best predictions where the sample data is properly filtered and an appropriate predictive algorithm is used. Using different multiple models on the same entity list allows for the optimization of the class prediction model for a specific application.