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

Brian Hom, Takeshi Serino, Anthony Gray. Agilent Technologies, Santa Clara, CA

ASMS 2012
MP573

Agilent Technologies

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 Separation Probe (TSP) and GC/MS. Samples classification can be automatically and statistically predicted using a class prediction model without the need of user intervention. Statistical methods are useful in this 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 20 lb. copier paper were prepared using a paper cutter. The samples were approximately 1.5 cm square. Several lines of ink were scribbled on the paper sample with one of three ball point pens. The samples were rolled up and inserted into TSP microvial. Whereupon 1 µL internal standard was added immediately prior to analysis. For the evaluation of ink profile changes over time, samples were prepared and stored in a plastic container until analysis. Three replicate samples were prepared for each of the pens at each of the time intervals. Additional samples were also prepared to act as unknowns to evaluate the predictive models.

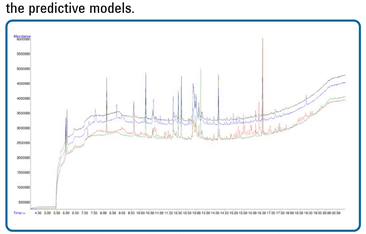

Figure 1: Overlays of TIC Chromatograms demonstrating the difficulty of direct visual differentiation. The complexity of the samples require advanced statistical tools to reliably evaluate the differences between the samples.

Experimental

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 Multi-mode inlet (MMI). The temperature programing capability of the MMI allowed for consistent introduction of the samples to the GC/MS. The pertinent instrument parameters and conditions are here outlined:

Column:	DB-5HT 15m, 1.2 mL/min He
Inlet:	Initial 100°C; hold 0.1 min
	Ramp at 600°C/min to 280°C
GC Oven:	Initial 70°C; hold 1.5 min
	Ramp at 15°C/min to 340°C
	Hold for 1.5 min
Source:	240°C
Quad:	150°C
Transfer:	300°C
Scan Range:	40-570 m/z
Sampling:	2
Gain factor:	1
Tune:	autotune

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 SCP 2.0 was used in conjunction with ChemStation E.02.02 SP1 to automate workflow from data collection to final predictive report.

AMDIS Parameters

Resolution:	Low
Sensitivity:	Medium
Shape requirements:	Medium

MPP Parameters

Minimum Abundance:	10,000 Counts
RT Tolerance:	0.05 min
Match Factor:	0.3

Results and Discussion

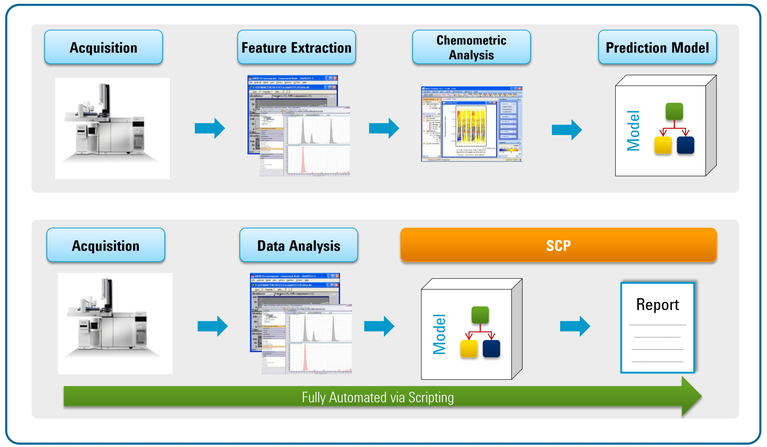

Figure 2: Schematic of workflow for generating predictive model and of automated generation of reports using the sample class prediction model.

Workflow

The workflow used in the creation of the class prediction model is sketched out in figure 2. After acquisition, a batch was created using AMDIS. The AMDIS sample reports were imported in MPP and the *Significance Testing and Fold Change* workflow wizard 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 <5%. This regiment of filtering reduced the number of entities from 341 to 87 entities of interest.

Models were generated using the *Class Prediction: Build and Test Model* wizard using the filtered entity list.

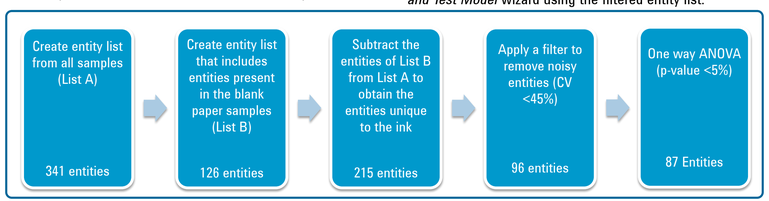

Figure 3: Schematic of the filtering to reduce the noise

Results and Discussion

	A		B		C		D		E	
Model	Predicted	CL	Predicted	CL	Predicted	CL	Predicted	CL	Predicted	CL
Support Vector Machine	[P00]	46%	[P90]	67%	[PM00]	39%	[PM90]	45%	[Z00]	97%
Neural Network	[P90]	37%	[P90]	37%	[P90]	20%	[Z00]	57%	[Z00]	69%
Partial Least Squares Discrimination	[P00]	64%	[P90]	62%	[PM00]	53%	[PM90]	65%	[Z00]	85%
Decision Tree	[P00]	100%	[P90]	100%	[PM00]	100%	[PM90]	100%	[PM00]	100%
Naive Bayes	[P00]	100%	[Z00]	100%	[P90]	100%	[PM90]	100%	[P90]	100%
Actual	P00		P90		PM00		PM90		Z00	

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

Automation of models

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 method in ChemStation. 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. Each of the unknown samples was reevaluated with the other four models simply by requantifying the data file. This standard prediction model allows the use of the same model on several instruments without having to generate a model on each system thereby increasing total sample throughput.

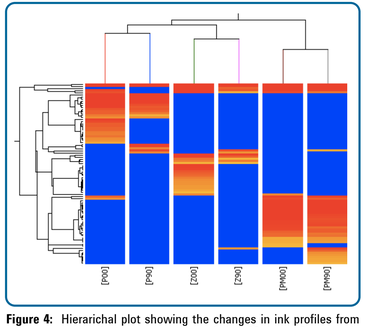

Figure 4: Hierarichal plot showing the changes in ink profiles from various pen manufacturers over the period of 90 days. Compounds that are present initially are not present later, and compounds that are not initially present are present later.

Evaluation of models

The evaluation of the different class prediction models in 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 53% to 85% for the samples. Neural Network and Naïve Bayes correctly identified 2 of the 5 unknowns while Decision Tree was successful in correctly identifying 4.

The confidence level of the models could be improved by increasing the number of samples used in training the prediction models.

Conclusions

We demonstrate that by using statistical class prediction modeling we can differentiate ink 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. SCP provides the best predictive results when the sample data is properly filtered and an appropriate prediction algorithm is used. Evaluating multiple models on the same entity list allows for the optimization of the sample class prediction for a specific application.