E4 Sparse Modeling for the Classification of Evidence From Spectral Data

Dayla C. Rich, BS, University of South Carolina, Columbia, SC 29208*

Learning Overview: After attending this presentation, attendees will have a deeper appreciation and understanding of the role multivariate statistical techniques, in particular sparse methods, can play in the interpretation of spectral data of trace evidence.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by illustrating how statistical analyses, specifically sparse chemometrics, can aid in classifying evidence from their respective spectral data.

Ultraviolet/Visible (UV/Vis) data of dyed fibers and Fourier Transform Infrared (FTIR) spectroscopy data of lipstick samples will be explored in this presentation. Trace evidence can play a critical role in successfully solving a criminal case by allowing for associations to be made between suspects, victims, and the crime science. A common technique utilized to analyze the dyes on fibers is UV/Vis Microspectrophotometry (MSP).[1] Further, lipsticks are analyzed through a plethora of techniques. FTIR is commonly used in forensic science and offers a quick and non-destructive method that yields vital information about the chemical nature of the sample.[2] However, classification of these two sample types via spectral analysis (spectral overlay) can be difficult due to the complexity of the data (i.e., large number of wavelengths or wavenumbers in spectral data).[3] Multivariate statistical methods can help in drawing conclusions about qualitative forensic analyses. Sparse chemometrics simplifies the predictive model by shrinking unimportant variable coefficients to zero. Thus, only the most significant variables are retained in the final model.[4] This research achieves sparsity using logistic regression with Least Absolute Shrinkage and Selection Operator (LASSO).

When compared to the more common statistical approach of principal component Analysis (PCA) combined with Linear Discriminant Analysis (LDA), the sparse method outperformed in prediction accuracy. Logistic regression with LASSO achieved 96.6% prediction accuracy for the fiber data set, while PCA with LDA yielded only 89.7% prediction accuracy. For the lipstick samples, prediction accuracy also increased. PCA with LDA achieved 60.0 % accuracy, while logistic regression with LASSO achieved 66.7% prediction accuracy. Limiting predictive models to only consider the most critical variables (sparsity) also greatly enhances model interpretability. Rather than considering hundreds of wavelengths or wavenumbers, only the most important and informative variables are retained in the model. As statistical analyses increase in popularity and become more prominent tools in the forensic science community, sparsity can help overcome the limitations of traditional modeling techniques and aid in the classification of various types of trace evidence.

Reference(s):

1. Goodpaster, John V., and Elisa A. Liszewski. Forensic Analysis of Dyed Textile Fibers. *Analytical and Bioanalytical Chemistry* 394, no. 8 (2009): 2009-2018.
2. Chopi, Rito, Spriha Sharma, Sweety Sharma, and Rajinder Singh. Trends in the Forensic Analysis of Cosmetic Evidence. *Forensic Chemistry* 14 (2019).
3. Gareth, James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2017.
4. Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton: CRC Press, 2015.

Sparse Chemometrics, Fibers, Lipsticks