Combining Analytical Chemistry and Machine Learning to Detangle Mixtures

NIEHS Superfund Research Program (SRP)-funded researchers demonstrated a significant step toward identifying individual chemical components in complex mixtures. Their approach uses advanced analytical techniques and sophisticated machine learning approaches while overcoming the time-consuming separation steps in traditional chemical analysis.

The researchers from the Baylor College of Medicine-Rice University SRP Center developed a new strategy to identify individual polycyclic aromatic hydrocarbon (PAH) compounds in complex mixtures. PAHs are a family of pollutants composed of fused benzene rings generated during incomplete combustion. PAHs are known carcinogens, but often occur in the environment as complex mixtures, which makes detecting and identifying them difficult.

The team's strategy combines the ultrasensitive molecular fingerprinting capabilities of surface-enhanced Raman spectroscopy (SERS) with the complex signal separation and detection capabilities of machine learning. SERS is a type of chemical analysis based on chemical bonds' interaction with light. The resulting spectra of wavelengths is used to infer chemical structures.

Their machine learning method, Characteristic Peak Extraction (CaPE), can be used with SERS in an approach they called computational chromatography. In short, they used algorithms that de-mix SERS spectra from complex mixtures into individual PAH spectra. Then they applied CaPE to extract characteristic peaks. Importantly, this approach helps address components of mixtures that occur at low concentrations by focusing on those that are most unique within the dataset, creating a compressed, or simplified molecular fingerprint. Their strategy can also handle small frequency shifts, common to SERS, which confound other approaches.

To evaluate their method, the team collected SERS spectra for six different two-component PAH mixtures and for more complex mixtures of four PAHs at different concentrations.

They compared several common de-mixing algorithms with and without CaPE to see which performed best at pulling out individual chemicals, called de-mixed components. Then, the scientists used another algorithm to match unknown de-mixed components to specific PAHs in a PAH spectra library based on characteristic features.

The research team aimed to observe how their approach performed with PAH mixtures of different complexity. In mixtures of two PAHs, their approach performed best for the anthracene and pyrene mixture and the anthracene and benzo(a)anthracene mixture, although in some cases minor features in the SERS spectra were not present in the resulting de-mixed components. Analysis of other mixtures contained some errors, however, the team explained that none prevented the de-mixed components from being matched visually or computationally to the correct PAH.

In more complex mixtures, with four PAHs, the de-mixed components did not match as closely with the SERS spectra. The best results were obtained for pyrene where the five most intense peaks matched the most intense peaks from SERS well. Despite the slightly less robust

performance, their matching algorithm was still able to pair the de-mixed components to the correct PAHs.

When comparing different de-mixing methods with and without CaPE, they found that adding CaPE consistently improved performance, particularly for the most difficult mixtures. In particular, existing methods only matched about half of the correct PAHs, whereas CaPE extracted and assigned characteristic peaks effectively.

According to the team, their results show that combining SERS and machine learning can more accurately, rapidly, and effectively de-mix compounds so that the individual chemicals can then be identified. This strategy can be used for rapidly detecting and identifying diverse chemicals in complex mixtures based on key molecular structures with no prior knowledge of their identity, the authors noted.

If you'd like to learn more about this research, visit the Superfund Research Program website at niehs.nih.gov/srp. From there, click on the Research Brief title under the banner, and refer to the additional information listed under the research brief. If you have any questions or comments about this month's podcast, send an email to srpinfo@niehs.nih.gov.

Join us next month as we discuss more exciting research and technology developments from the Superfund Research Program.