Assessing Reactive Groups and Reactivity Alerts in the CAMEO Chemicals Database

Nikhil Venkatachalam, Seongjin Bae, Yixun Lu, Nicole Bain, Xinyi Guan, Paul Clingen, Daniel Harradine

PURDUE UNIVERSITY | The Data Mine

CAS A division of the American Chemical Society

Pistoia Alliance

Intro/Background/Motivation

The Chemical Safety Library contains numerous reports of hazardous incidents experienced by lab users, highlighting the need for an analytical approach to effectively process this data. Thus, our research investigates the relationship between chemicals' reactive groups, and their associated reactivity alerts. The study classifies chemicals by reactivity, including explosiveness, flammability, and oxidizing potential, and compares them across chemical groups like alcohols, nitrates, and peroxides. This approach aims to enhance understanding of chemical reactivity patterns, contributing to safer laboratory practices and more informed chemical management strategies.

Research Methodology

Two of PubChem's annotated datasets (in JSON format, converted to CSV in some cases for ease of access) from the CAMEO Chemicals database, Reactivity Alerts and Reactive Groups, at https://pubchem.ncbi.nlm.nih.gov/source/11944 were downloaded for this data analysis. Focus was given on types of alerts/groups during data analysis. This can be seen within Fig 1 and Fig 2, which show the reactive group categories and reactive alert counts, giving the user a concise picture of the data. After this, the datasets were combined and merged for cross-correlation analysis, as can be seen in Figs 3, 4, and 5. Fig 3 shows which reactivity alerts are related to which reactivity groups, which shows a small picture of correlation. Figs 4 and 5 elaborate upon this, as Fig 4 shows a better picture of how each alert is correlated for each group. Fig 5 focuses on a specific reactivity alert as an example, and shows correlation in a greater detail. Figs 6 and 7 are instances of further data analysis using Machine Learning, specifically the elbow method and PCA visualization. Both give greater insight into how reactive alerts and reactive groups are connected. All data analysis was conducted using Python.

Fig 1

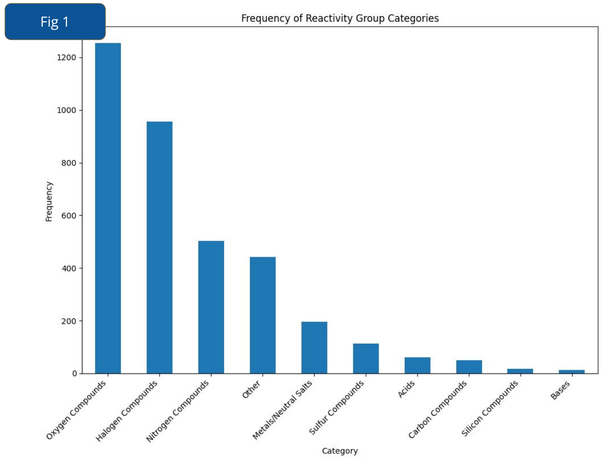
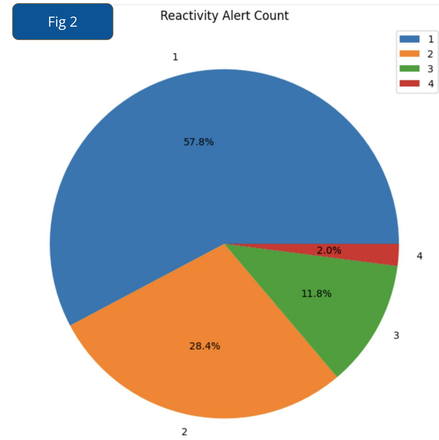


Fig 2



Conclusions

- Strong Oxidizing Agents:
  - "Nitrates" and "Peroxides" are highly reactive, acting as strong oxidizing agents.
  - They readily donate oxygen to other substances, potentially causing explosive reactions.
- Explosive Nature:
  - "Alkali Metals" and "Diazonium Compounds" are noted for their explosive characteristics.
  - Alkali metals react intensely with water, while diazonium compounds are unstable in certain conditions.
- Flammability:
  - "Alcohols and Polyols" and "Aliphatic Saturated Hydrocarbons" exhibit high flammability.
  - They can ignite under appropriate conditions.
- Reactivity Trends:
  - Organic compounds (e.g., alcohols, ethers) tend to have higher frequencies of explosive and flammable characteristics compared to inorganic compounds.
  - Certain groups like "Acrylates and Acrylic Acids" or "Organometallics" display varied reactivity profiles, suggesting multiple potential hazards.
  - Chemicals with reactive groups: Borate Esters, Sulfate Esters, Esters, Thiophosphate Esters, and Phosphate Esters are more closely clustered than other chemicals.

Fig 3

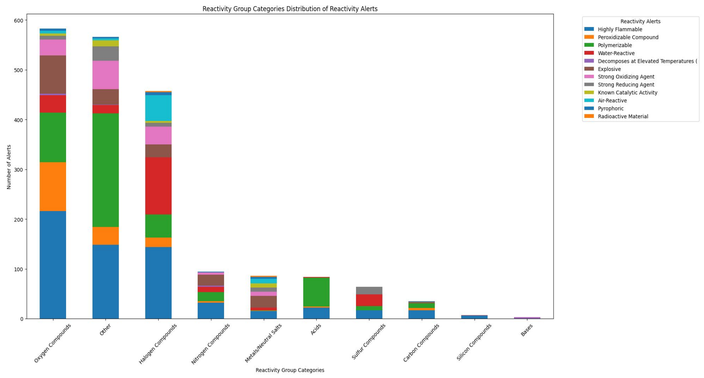


Fig 4

Cross-Correlation between Reactivity Group Categories and Reactivity Alerts

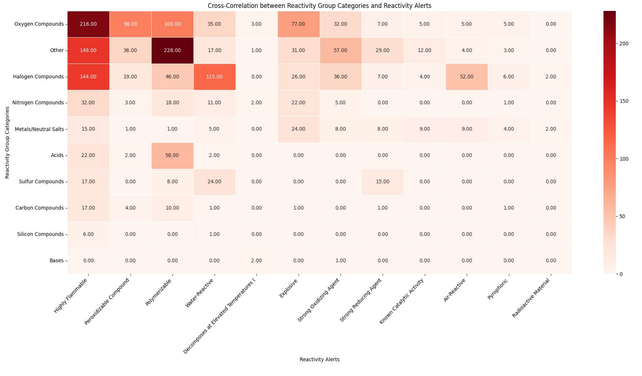


Fig 5

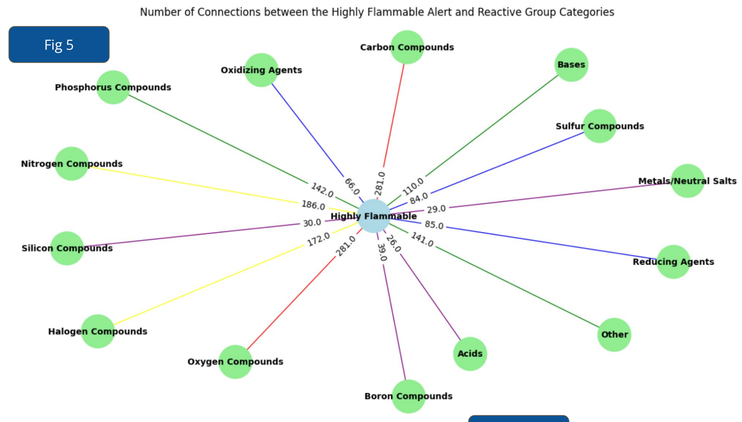


Fig 6

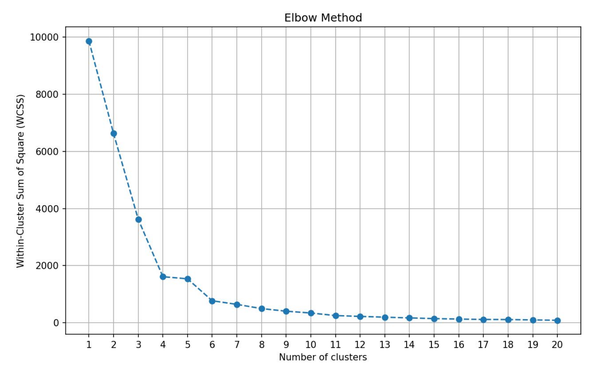
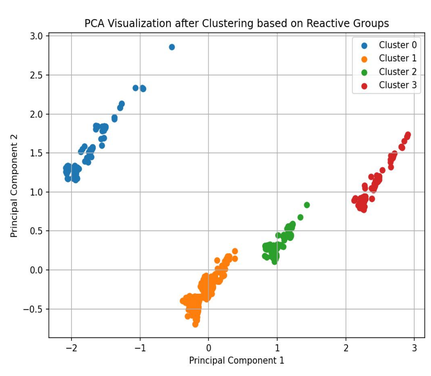


Fig 7



Future Plans

- Expand dataset:
  - Integrate more chemical groups and reactivity alerts
  - Incorporate additional experimental data, literature reviews, and incident reports
- Data Verification and Update:
  - Verify existing data for heatmap creation
  - Review current literature, safety databases, and reports
  - Update analysis with new findings or corrections
- Safety Protocol Development:
  - Develop basic safety protocols for handling common chemicals
  - Focus on highly reactive or hazardous substances
  - Ensure accessibility and comprehensibility for non-experts
- Regular Review and Preparation:
  - Review lab material regularly
  - Utilize online resources and tutorials for Python application enhancement

Acknowledgements

Jimmy Dinh - Corporate Partners TA

Grace Baysinger - CSL Advisor, Stanford, Emeritus

Thomas Vickery - CSL Advisor, Merck

Gabrielle Whittick - CSL Project Manager, Pistoia Alliance

David Zwicky - Corporate Partners Mentor, Purdue

Qiong Yuan - Corporate Partners Sponsor, CAS Innovation

PubChem, CAMEO Chemicals - datasets access and means to analyze reactivity and hazardous reactions