JOURNAL LA MULTIAPP

VOL. 04, ISSUE 03 (092-097), 2023
DOI: 10.37899/journallamultiapp.v4i3.852

Automated Chemical Equation Balancing Using the Apriori Algorithm

Yasmin Makki Mohialden[1], Nadia Mahmood Hussien[1], Walaa A Abd Al-Rada[1]

[1]Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

**Corresponding Author: Yasmin Makki Mohialden*
Email: ymmiraq2009@uomustansiriyah.edu.iq

Article Info

Article history:
Received 10 October 2023
Received in revised form 3 November 2023
Accepted 21 November 2023

Keywords:
Chemical Equation Balancing
Apriori Algorithm
Automation in Chemistry
Machine Learning in Chemistry
Conservation of Mass

Abstract

Chemical equations must be balanced to maintain mass conservation. Traditional chemists employed manual processes with meticulous investigation and trial-and-error iterations. Automating and enhancing this difficult process is becoming more popular as machine learning (ML) progresses. We provide a novel Apriori algorithm-based chemical equation balancing method in this paper. Our solution uses the Apriori algorithm to find common itemsets of balanced reactions and translates unbalanced equations into machine-readable language. After that, it reconstructs balanced equations, automating a tedious task.

Introduction

Chemical equations must be balanced to understand and predict chemical systems. When reactants equal products, the rule of conservation of mass is met. Reaction prediction, process optimization, and stoichiometric calculations need this balancing mechanism (Lunz et al., 2021).

Traditionally, chemists balanced chemical equations through examination and trial and error. For complicated reactions with several reactants and products, these methods may be time-consuming. Manual balancing is also skill-dependent and may be inaccurate (Johar, 2020).

The integration of chemistry and machine learning (Mohialden et al., 2022) has enabled chemical equation balance automation and optimization. Scientists and computer scientists may solve a long-term problem using this approach. Machine learning can speed up chemical reaction balance for researchers, instructors, and practitioners (Suleimanov & Green, 2015).

At the intersection of chemistry and AI, might automation and data-driven methods transform chemical equation balancing? Traditional methods for balancing chemical equations include careful observation and trial and error. Automation of chemical equation balancing is a fresh problem statement at the junction of scientific principles and computer innovation (Taylor et al., 2023); (Cova Pais, 2019).

It is possible to change chemistry using this innovative problem statement. Automating chemical equation balancing might improve efficiency, accuracy, and accessibility. Streamlining a laborious procedure accelerates chemical process optimization, reaction result

ISSN: 2716-3865 (Print), 2721-1290 (Online)

prediction, and scientific discovery. Automation of chemical equation balance is the focus of this chemistry-machine learning study. We hope to demonstrate how automation may transform chemical research and education, opening new doors for practitioners and researchers. Apriori is an unsupervised learning method used in association rule mining for predictive modeling.

To find item-set association rules, the Apriori method was devised in the early 1990s. It's utilized in pattern identification and prediction, such as predicting a consumer's buying behavior after buying other goods. Using relational database transactional data, the Apriori algorithm operates. It finds frequent item sets, which are transaction-prone combinations of items. Create association rules from these things. Customers who commonly purchase items A and B together may develop an association rule suggesting that buying A improves the chance of buying B (Zhang, 2021); (Wang et al., 2016).

The Apriori algorithm helps analysts generate predictions and suggestions based on item association patterns in transactional data (Li et al., 2022). This article uses the 'apyori' library's Apriori algorithm to balance imbalanced chemical equations using 'balance_equation'. The technique finds common itemsets representing balanced chemical equations from a collection of unbalanced equations and picks the most balanced.

Related Work

"Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods" (Suleimanov & Green, 2015). This paper presents a protocol for the automated discovery of elementary chemical reaction steps using double-ended transition state optimization methods.

"Automated learning of chemical reaction networks (Wilson & Wilson, 2019): This paper proposes an approach that utilizes a mixed-integer nonlinear programming formulation to estimate and identify kinetic rate parameters for chemical reaction networks.

"A graph-based network for predicting chemical reaction pathways in solid-state material synthesis" (McDermott et al., 2021). This paper presents a chemical reaction network model for solid-state synthesis constructed from available thermochemistry data and devises a computationally tractable approach for suggesting likely reaction pathways via the application of pathfinding algorithms and a linear combination of the lowest-cost paths in the network.

"Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks (Unsleber et al., 2022): This paper presents a tool that facilitates reaction network explorations for diverse chemical problems with a wide range of goals, such as mechanism.

Table 1. Is A Comparison Table of The Methods Based on Various Factors

Factor	Freezing String and Berny Optimization (6)	Mixed-Integer Nonlinear Programming (7)	Graph-Based Network for Predicting Pathways (8)	Chemoton 2.0 (9)
Methodology	Transition state optimization	Mathematical formulation for kinetics	Graph-based network with pathfinding algorithms	Chemical network exploration tool
Automation Level	Specific task	Specific task	Broad exploration capabilities	Versatile exploration tool
Applicability	Reaction steps discovery	Kinetic parameter estimation	Solid-state material synthesis reactions	Various chemical problems
Data Requirements	Reaction data	Reaction data	Thermochemistry data	Not specified
Complexity	Complex optimization	Complex mathematical formulation	Data-driven and algorithmic	User-friendly tool

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Goal	Mechanistic analysis	Parameter estimation	Pathway prediction in solid-state synthesis	Chemical network exploration
Computational Cost	Potentially high	Potentially high	Depends on network size	Variable depending on the problem
Ease of Use	Requires optimization expertise	Requires kinetics expertise	User-friendly for non-experts	User-friendly for non-experts

Source: DPMPTSP South Tangerang City (2023)

Methods

With the ML revolution, we present a novel method for automated chemical equation balancing that makes use of the Apriori algorithm. Our method employs the Apriori algorithm to find frequently occurring itemsets that indicate balanced reactions in addition to converting unbalanced equations into a machine-readable format. The main steps for the proposed method are:-

Input

The method operates on unbalanced chemical equations, which are provided as input in the form of strings. For instance, an example input could be:

"Fe + H2SO4 -> Fe2(SO4)3 + H2".

Parsing

Upon receiving the input equation, the method proceeds to parse it. The equation is first split into reactants and products by identifying the "->" arrow as the delimiter. Then, each side of the equation is further broken down into individual items representing chemical species. This is accomplished by splitting each side on the "+" symbol.

Data Conversion

Chemical species are hashed into tuples for processing. The Pandas DataFrame stores these tuples with each row representing a chemical equation.

Transaction Creation

The data frame is transformed into a list of transactions, where each transaction corresponds to a chemical equation. This step prepares the data for subsequent analysis.

Apriori Algorithm

The strategy relies on applying the Apriori algorithm on the transaction list. Association rule mining often uses the Apriori algorithm for item mining. In this example, it finds chemical species that occur often in input equations. The minimum support level is 0.2, however, you may increase it to suit your requirements.

Balanced Equation Selection

The approach selects the most balanced frequent item set as the balanced chemical reaction after using the Apriori algorithm. The idea of "most balanced" is decided by choosing the itemset with the most items, indicating a better-balanced equation.

Result

The balanced item is converted back into the balanced chemical equation's string representation. This equation is the best effort at balancing the input chemical reaction.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Output and Savings

For convenience, the balanced chemical equation is displayed on the console. Furthermore, the unbalanced and balanced equations are recorded in separate text files for convenient access and analysis.

This technique provides a computer method for balancing imbalanced chemical equations. It uses the Apriori algorithm to discover commonly occurring chemical species combinations that are supposed to reflect balanced reactions. This method can help, but it may not always provide perfectly balanced equations, necessitating human verification.

Results and Discussion

In this study, we provide a unique approach for balancing chemical equations based on the Apriori algorithm. The suggested approach aims to use data mining methods to automate the sometimes-time-consuming operation of balancing chemical equations. In this part, we will look at the outcomes of applying our approach to a set of imbalanced chemical equations.

Our approach was applied to a set of imbalanced chemical equations, and the results confirmed its efficacy in constructing balanced equations. The following is a summary of the outcomes for each sample equation:

Example 1:

Unbalanced Equation: Fe + H2SO4 -> Fe2(SO4)3 + H2

Balanced Equation: ('Fe', 'H2SO4', 'Fe2(SO4)3', 'H2')

Example 2:

Unbalanced Equation: C4H10 + O2 -> CO2 + H2O

Balanced Equation: ('C4H10', 'O2', 'CO2', 'H2O')

Example 3:

Unbalanced Equation: KClO3 -> KCl + O2

Balanced Equation: ('KClO3', 'KCl', 'O2')

Example 4:

Unbalanced Equation: P4 + O2 -> P2O5

Balanced Equation: ('P4', 'O2', 'P2O5')

Example 5:

Unbalanced Equation: C6H12O6 -> C2H5OH + CO2

Balanced Equation: ('C6H12O6', 'C2H5OH', 'CO2')

Example 6:

Unbalanced Equation: N2H4 + N2O4 -> N2 + H2O

Balanced Equation: ('N2H4', 'N2O4', 'N2', 'H2O')

Example 7:

Unbalanced Equation: NH4NO3 -> N2O + H2O

Balanced Equation: ('NH4NO3', 'N2O', 'H2O')

Example 8:

Unbalanced Equation: Al2(SO4)3 + KOH -> Al (OH)3 + K2SO4

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Balanced Equation: ('Al2(SO4)3', 'KOH', 'Al (OH)3', 'K2SO4')

Example 9:

Unbalanced Equation: C2H4 + O2 -> CO2 + H2O

Balanced Equation: ('C2H4', 'O2', 'CO2', 'H2O')

We balanced each chemical equation by picking the most balanced frequent item set. Such as the Apriori algorithm, data mining may automate complicated chemical operations.

Our strategy for balancing chemical equations seems promising for scientists and educators. To save time and reduce errors, we automate balancing. On samples, our method worked, but more sophisticated equations may disagree. We need to investigate and refine the system to handle additional chemical processes. Customize the approach with flexible minimum support threshold modification. Controlling thresholds may enhance chemical equation results.

Conclusions

This research demonstrated a practical automatic chemical equation balancing approach with satisfactory test results. This simplifies an important chemical idea for students and researchers.

Our method shows potential, but further research is required to make it useful in chemistry. Future work ideas: (1) Chemical Element Data: Incorporate a comprehensive database of chemical elements and their properties. This step is crucial to ensuring the correctness of the balanced equations by verifying the conservation of the number of atoms in each element; (2) Web Application: Consider transforming our algorithm into a web-based application. This would enable users to balance chemical equations conveniently and efficiently online, making it accessible to a broader audience; (3) Handling Special Cases: Addressing special cases, such as redox reactions, is essential. These reactions involve electron transfer, which must be balanced alongside mass conservation. Extending our algorithm to handle such scenarios would significantly increase its versatility.

While our current approach is a promising foundation for automating chemical equation balancing, further research, refinement, and implementation of the suggested enhancements will be pivotal in realizing its full potential and making a substantial impact in the field of chemistry.

Acknowledgment

The Authors would like to thank Mustansiriyah University (https://uomustansiriyah.edu.iq/) in Baghdad, Iraq, for their support in the present work.

References

Cova, T., & Pais, A. (2019). *Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns*. *Frontiers in Chemistry, 7. * https://doi.org/10.3389/fchem.2019.00809.

Johar, D. (2020). Application of the Concept of Linear Equation Systems in Balancing Chemical Reaction Equations. **International Journal of Green and Organic Chemistry, 1**(4), 130-135. https://doi.org/10.47194/IJGOR.V1I4.48.

Li, G., Li, Z., Zheng, C., & Li, Z. (2022). Design of Middle School Chemistry Experiment Simulation System Based on Apriori Algorithm. *Journal of New Media, 4*(1).

Lunz, D., Batt, G., Ruess, J., & Bonnans, J. (2021*). Beyond the chemical master equation: Stochastic chemical kinetics coupled with auxiliary processes*. *PLoS Computational Biology, 17. * https://doi.org/10.1371/journal.pcbi.1009214.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

McDermott, M. J., Dwaraknath, S. S., & Persson, K. A. (2021). A Graph-Based Network for Predicting Chemical Reaction Pathways in Solid-State Material Synthesis. *Nature Communications, 12*(1), 1–12. https://doi.org/10.1038/s41467-021-23339-x.

Mohialden, Y. M., Hussien, N. M., Mohammed, S. J., & Mohammed, I. S. (2022). Recent hybrid machine learning algorithm applications: a review. *Journal of Information Technology and Informatics*, *2*(1), 1-4.

Suleimanov, Y. V., & Green, W. H. (2015, August 5). Automated Discovery of Elementary Chemical Reaction Steps Using Freezing Strings and Berny Optimization Methods. *Journal of Chemical Theory and Computation, 11*(9), 4248–4259. https://doi.org/10.1021/acs.jctc.5b00407.

Suleimanov, Y., & Green, W. (2015). Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *Journal of Chemical Theory and Computation, 11*(9), 4248-4259. https://doi.org/10.1021/acs.jctc.5b00407.

Taylor, C. J., Felton, K. C., Wigh, D., Jeraal, M. I., Grainger, R., Chessari, G., Johnson, C. N., & Lapkin, A. A. (2023, April 13). Accelerated Chemical Reaction Optimization Using Multi-Task Learning. *ACS Central Science, 9*(5), 957–968.

Unsleber, J. P., Grimmel, S. A., & Reiher, M. (2022, August 4). Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks. *Journal of Chemical Theory and Computation, 18*(9), 5393–5409. https://doi.org/10.1021/acs.jctc.2c00193. Apriori (n.d.). Retrieved from https://pypi.org/project/apyori/.

Wang, J., Li, H., Huang, J., & Su, C. (2016). Association rules mining based analysis of consequential alarm sequences in chemical processes. *Journal of Loss Prevention in the Process Industries*, *41*, 178-185.

Wilson, Z. T., & C, N. V. (2019). Automated Learning of Chemical Reaction Networks. *Computers & Chemical Engineering, 127*, 88–98. https://doi.org/10.1016/j.compchemeng.2019.05.020.

Zhang, W. (2021). Retracted article: the integration of fertilizer soil pollution and agricultural industry economy based on Apriori algorithm. *Arabian Journal of Geosciences*, *14*(16), 1582.

ISSN: 2716-3865 (Print), 2721-1290 (Online)