International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Validating urea-nitrogen CARS spectra with collaborative filtering

Varun Patankar *

Juanita High School and Atonarp, 8801 142nd Ave NE Redmond WA 98052. United States of America.

International Journal of Science and Research Archive, 2024, 13(02), 1638–1642

Publication history: Received on 17 October 2024; revised on 23 November 2024; accepted on 26 November 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.2.2304

Abstract

As someone passionate about improving end stage renal disease (ESRD) treatment, I covered a program that I developed which enhances the success of dialysis. ESRD is a condition where the kidneys lose function, requiring dialysis to remove urea nitrogen from the blood via dialysate fluid. Accurate analysis of dialysate samples is important, as it reveals excretory success, and Coherent Anti-Stokes Raman Spectroscopy (CARS) is ideal for this purpose. CARS measures the molecular composition of a sample by using light interactions to produce a molecule vibrational signal, revealing urea nitrogen concentrations. To ensure the validity of CARS results, as measurements can be inaccurate, I've integrated collaborative filtering techniques into a program. These techniques find patterns from user input to identify valid samples. My program has two algorithms: Byzantine Generals and Iyengar, the latter providing more in-depth analysis. Both algorithms compare user input parameters—such as wavelengths, intensities, and concentrations—to a database of lab values from Atonarp's pilot study. The tool interface allows users to input their spectral data, select an algorithm for validation, and get results with comparison graphs. If no data is available, sample data can be copied and pasted. Tested with real-life data from Atonarp, this tool can be used to make the dialysis process safer and more accurate.

Keywords: CARS Spectroscopy; Collaborative Filtering; Engineering; ESRD

1. Introduction

Imagine battling a disease so severe that it's not only difficult to treat but also increases your risk of heart failure, infections, and even death. This is the reality for those suffering from end-stage renal disease (ESRD), a growing health crisis with severe consequences. ESRD [1] is the final stage of kidney disease when the kidney cannot function independently, requiring patients to undergo costly and frequent dialysis to remove waste products via a special type of fluid. The process necessitates regular monitoring and analysis of this fluid, specifically urea nitrogen and uric acid levels, which are essential for measuring excretory success. Urea nitrogen is a waste product formed from protein breakdown, while uric acid is a waste product from the breakdown of purines found in certain foods and cells in the body. Both substances need to be removed during dialysis. Traditional molecular monitoring involves dialysate samples being sent to labs for analysis, which can take time. Dialysate is the name of the fluid that draws the waste products from the blood. Due to the need for a more efficient process, an innovative form of technology called Optical Spectroscopy is in development [3]. This technology uses Coherent Anti-Stokes Raman Spectroscopy (CARS) [2], an advanced analytical technique used to measure the molecular composition of a sample.

CARS involves the interaction of four photons, a particle representing light, with any molecule. Two of these photons are from a component called the pump beam, one is from the Stokes beam, and the fourth is from the Stokes photon. The pump and Stokes beams are tuned such that their frequency difference matches the vibration of the molecule. This results in a recognizable result, allowing for the detection of specific molecular vibrations and directions in the presence of any environment. These beams interact with the molecules in the sample, exciting them and causing them to emit light at different frequencies. However, to ensure the accuracy of the CARS measurements, which can sometimes be

* Corresponding author: Varun Patankar

Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

anomalous, it is essential to validate the incoming spectra against known lab values. Accurate validation of the CARS spectra ensures that the data collected is trustworthy and can be used to make smart decisions about patient treatment.

2. Materials and Methods

This is where collaborative filtering techniques come in. Collaborative filtering is a technique commonly used in recommendation systems to predict user preferences by noting the history of many other users. In the context of validating urea nitrogen spectra, collaborative filtering can help by aggregating/analyzing data from multiple samples to identify patterns and validate measurements. For example, if multiple samples provide high value, while some provide low values, collaborative filtering can be used to cross-validate data, ensuring that the measurements are consistent. This technique helps in identifying outliers or anomalies in the data, which can then be flagged for further investigation. Several collaborative filtering techniques can be used for Urea Nitrogen validation.

The Brooks-Iyengar algorithm is a technique for filtering that provides fault-tolerant estimation by combining data from multiple sources. [4] This method starts with collecting spectral measurements, each of which may have varying levels of accuracy and susceptibility to errors. The algorithm aggregates these measurements into a dataset, applying statistical techniques to mitigate the faulty data. This includes identifying and filtering out outliers—data points that deviate significantly—as well as using weighted averaging, where sources with higher reliability or accuracy are given more influence in the final estimate. By focusing on the consensus among most sources and reducing the impact of any outliers, the Brooks- Iyengar algorithm ensures that the final validated spectra are accurate and representative of the true urea nitrogen levels, making it highly effective for applications in dialysate analysis and other medical uses.

RBMs are another technique that recognizes patterns and learns features in nitrogen urea spectra samples. An RBM has two parts: the visible layer, which holds the input data, and the hidden layer, which finds patterns within this data. The connections between these two layers are adjusted during training using something called Contrastive Divergence and determined by a process called Matrix Factorization. Here's how the process works: First, the RBM analyzes the visible layer data, which in this case is the nitrogen urea spectra. Each unit in the hidden layer calculates the probability of being activated, which means determining how strongly the existing pattern correlates to the input data. This process transforms the spectral data into a probability distribution over the hidden units. Then, it tries to recreate the visible layer data from these hidden activations by reversing the process: summing valid inputs from the hidden layer and applying the sigmoid function to generate visible unit probabilities. The differences between the original data and recreated data are used to adjust the weights between the layers using a learning rate to minimize error, making the model better at recognizing patterns. Once the RBM is trained, it can analyze new spectral data to see how well it fits the patterns. If the new data significantly deviates, it is flagged as a potential anomaly. This helps to filter out incorrect data, improving the accuracy of the spectra validation process. RBMs have been successfully applied in various domains, including video platforms. For instance, Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton demonstrated that RBMs could efficiently handle large datasets, such as the Netflix dataset with over 100 million user/movie ratings. They also showed that RBMs could outperform Singular Value (SV) models. By combining the predictions of multiple RBM and SV models, they achieved an error rate of over 6% better than Netflix's system.

Using RBMs, one can better validate nitrogen urea spectra samples, ensuring that the data analyzed is accurate and reliable. The same principles that allow RBMs to excel in collaborative filtering can be applied to improve the recognition and validation of spectral data, highlighting the adaptability and quality of RBMs across different types of data and applications. In matrix factorization for RBM, the rating matrix R is approximated by the product of two matrices: U (the user feature matrix) and V (the movie feature matrix). In this context, the visible layer of the RBM corresponds to the spectral data (R), and the hidden layer corresponds to the learned patterns (U and V). If U is a 3x2 matrix (where 3 is the number of users and 2 is the number of features), and V is a 4x2 matrix, the product of U and V's matrix dimensions results in a 3x4 matrix, providing the predicted ratings for 3 users and 4 movies. By minimizing the difference between the actual ratings in R and the predicted ratings from the product of U and V, the model learns to capture the patterns in user preferences and movie characteristics, like how RBMs learn patterns in spectral data. [5]

The Byzantine Generals' Problem provides another method for achieving consensus in distributed systems where some participants may be unreliable. The solution to the Byzantine Generals' Problem [6] consists of determining how many samples need to be valid to achieve a consensus despite faulty data. Specifically, let us say a system evaluates a total of 3m + 1 samples. In that case, it can tolerate up to m faulty samples while still achieving consensus if the other samples are valid. For example, a system with 7 samples can handle up to 2 faulty samples, meaning that 5 out of 7 samples must be valid to ensure a reliable consensus. This principle helps to filter out unreliable data and achieve a validation result by ensuring that most samples are dependable. Using all these collaborative filtering methods, a program was developed to validate Urea Nitrogen CARS spectra. Before getting into it, it’s important to consider previous studies that provided

valuable lab information comparisons, most notably, the pilot study by Atonarp, a mass spectroscopy company based in Japan. In the study, dialysate samples were collected from seven patients during a four-hour dialysis session. The aim was to measure urea nitrogen and uric acid levels using the ATON System, an in vitro diagnostic analyzer based on CARS technology. Samples were collected at ten different time points, allowing analysis of the dialysate over time. Each sample was analyzed by the ATON System and compared to laboratory values. The ATON System is a diagnostic platform designed for precise and efficient analysis of biological samples. This cutting-edge technology uses the principles of Raman scattering to deliver quantitative measurements of molecular concentrations, specifically targeting molecules such as urea nitrogen and uric acid.

The system consists of three subsystems: the Excitation Module, which uses two ultra-fast lasers— the high-energy pump beam and a lower-energy Stokes beam—to spark an optical process in the sample; Free-Space Optics, which aligns and focuses these beams onto the dialysate sample in a glass cuvette, generating the Anti-Stokes Raman signal through optical management; and the Detection Module, which captures the signal with a high-resolution spectrometer and converts it into a spectrum analyzed by software algorithms to quantify molecule concentrations. These profiles are compared to laboratory reference values to ensure accuracy. By offering real-time, direct detection of molecules with high sensitivity, the ATON System provides a more efficient, cost-effective alternative to traditional methods for kidney disease management, ultimately aiming to streamline the diagnostic process.

3. Results and Discussion

The results of the study were quite interesting. The CARS spectra for urea nitrogen demonstrated high consistency across the seven patients, with a strong correlation factor to lab values (R^2 = 0.9888). This indicates that the ATON System provided reliable measurements of urea nitrogen. However, the spectra for uric acid showed more variability, which was credited to spectral overlap with bicarbonate in the dialysate, which interferes with molecular vibration. The correlation factor for uric acid measurements was less strong (R^2 = 0.3521), and the uric acid concentrations measured by the lab method were below 2 mg/dL, which was below the limit of measurement. Graphs comparing the incoming CARS spectra to the reference spectra were generated by Atonarp. These graphs demonstrated the correlation between the measured spectra and the reference spectra from the study. The pilot study also confirmed that the ATON System is capable of accurately measuring urea nitrogen in dialysate, proving it to be a viable alternative to laboratory methods for this molecule. However, improvements are needed to enhance the sensitivity for measuring uric acid and other molecules. Therefore, the program that I will discuss at the end currently only validates urea nitrogen spectra, as data is available for this molecule. However, future development will focus on increasing the system's sensitivity through modifications to the sample molecule, like the uric acid.

In support of Atonarp's study, a Python-based algorithm study was initiated, inspired by the Byzantine Generals' Problem and the Brooks-Iyengar Algorithm, to validate Nitrogen Urea CARS spectra. The goal of the algorithms was to ensure the accuracy of the CARS measurements by comparing incoming spectral data to a set of reference values in Atonarp's study and discarding any data points that were invalid. The study began by researching the Byzantine Generals' Problem. Understanding its application to data validation, an initial algorithm based on this concept was implemented to ensure that incoming spectral data could be reliably validated even in the presence of anomalous or corrupted data points. The Byzantine-inspired algorithm compared incoming spectra against multiple reference spectra, using parameters such as wavelengths, frequencies, and intensities, allowing it to detect and reject outlier data that did not align with most reference points. After successfully implementing the Byzantine-based approach, the study explored the Brooks-Iyengar Algorithm, which is effective at handling uncertainty with noise in data. To further enhance the validation process, the Brooks-Iyengar Algorithm was integrated into the existing framework. This algorithm also uses the same parameters—wavelength, frequency, and intensity—but introduces specific intervals around the expected values for these parameters. For each incoming spectral sample, the Brooks-Iyengar algorithm calculates whether these values fall within their respective intervals compared to reference data. If all parameters (wavelength, frequency, and intensity) fall within these intervals, the sample is considered valid; otherwise, it is invalid and discarded. The code for both algorithms was developed, ensuring it was well-documented and structured for clarity. The combined approach was validated by comparing its performance against real-life customized values, which were obtained directly from Atonarp.

To calculate the correlation between the threshold and input samples for the mentioned parameters, the program used comparison equations. Both algorithms calculated the Euclidean distance [7] between the incoming and reference wavelengths to determine validity. If the distance exceeds a set threshold in the study, the data is considered invalid. Euclidean distance measures the straight-line distance between two points in a multidimensional plane, making it useful for comparing the similarity between incoming and reference wavelengths. The programs perform similar comparisons with other thresholds, most notably using correlation coefficients and signal-to-noise ratio (SNR), to compare

concentration and intensity parameters. The correlation coefficient [8] quantifies the strength and direction of the relationship between two variables, allowing the algorithms to assess the alignment between molecule quantities; a high correlation indicates that changes in one variable are consistently associated with changes in the other, which is crucial for valid measurements.

Signal-to-Noise Ratio (SNR) [9] is used to evaluate the intensities of the photons reflected, as it measures the level of the desired signal relative to background noise. A high SNR indicates a clear, strong spectral signal, necessary for measurements like intensity, where signal strength is critical. The algorithms also generate graphs comparing incoming and reference spectra based on the input parameters, helping to visualize discrepancies and validate measurement accuracy. To input values into the program, a user must enter the parameter values in the following order: 7 wavelengths separated by commas, 7 intensities separated by commas, and 7 concentrations separated by commas. The parameter values in a specific position, combined with the other parameter values in the same position, make up one sample. For example, if a wavelength of 700 is entered first in its group, an intensity of 0.054 is entered first in its group, and a concentration of 20 is entered first in its group, these values collectively represent the properties of sample 1. Upon entering the spectra, the user selects which algorithm they want to use and clicks "validate." The outcome, along with the graphs, is displayed.

4. Conclusion

The study demonstrated that Optical Spectroscopy can accurately measure urea nitrogen in dialysate, offering a promising alternative to traditional laboratory methods. The successful implementation of a Byzantine, Brooks-Iyengar-inspired algorithm significantly improved the reliability of the CARS data validation process. Using reference spectra from the ATON study [10], the accuracy of incoming data was ensured through a structured comparison and graphical analysis. Moving forward, the focus will be on refining the validation algorithm and expanding the range of molecules input to further advance the capabilities of molecular detection technologies and improve patient care in dialysis clinics and beyond.

Compliance with ethical standards

Disclosure of conflict of interest

Varun Patankar could have potential conflict of interests with the Atonarp study, as he has interned with them before and has a vested interest towards them.

References

[1] Eckardt, K. U., Coresh, J., Devuyst, O., et al. (n.d.). Evolving importance of kidney disease: From subspecialty to global health burden.

[2] Camp, C. H. Jr., Lee, Y. J., Heddleston, J. M., et al. (n.d.). High-speed coherent Raman fingerprint imaging of biological tissues. Nature Photonics.

[3] Cheng, J. X., & Xie, X. S. (n.d.). Coherent anti-Stokes Raman scattering microscopy: Instrumentation, theory, and applications. Journal of Physical Chemistry B.

[4] Brooks, R. R., & Iyengar, S. S. (n.d.). Robust distributed computing and sensing algorithm. IEEE Computer.

[5] Salakhutdinov, R., Ruslan, et al. (n.d.). Restricted Boltzmann machines for collaborative filtering.

[6] Lamport, L., Shostak, R., & Pease, M. (n.d.). The Byzantine generals problem. ACM Transactions on Programming Languages and Systems.

[7] Euclidean distance: Advantages & limitations. (2019). Available online: https://botpenguin.com/glossary/euclidean-distance (accessed on 20 Nov. 2019).

[8] Correlation coefficients. (2005). Available online: https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm (accessed on 18 July 2005).

[9] Signal-to-noise ratio formula. (2022). Available online: https://www.geeksforgeeks.org/signal-to-noise-ratio-formula/ (accessed on 20 Mar. 2022).

[10] Plot study: Real-time measurements of uremic toxins in dialysate. (n.d.). Atonarp.

[11] De Smet, R., Van Kaer, J., Van Vlecka, B., et al. (n.d.). Novel methods for estimating urea distribution volume for non-linear kinetic analysis. American Journal of Kidney Diseases.

[12] Gregorek, J. K., Wright, A. I., Slabacki, W., et al. (n.d.). Ultraviolet Raman spectroscopy for detecting renal disease biomarkers in human serum. Analytical Chemistry.

[13] Hinton, G. E., & Salakhutdinov, R. R. (n.d.). Reducing the dimensionality of data with neural networks. Science.

[14] Koren, Y., Bell, R., & Volinsky, C. K. (n.d.). Matrix factorization techniques for recommender systems. Computer.

[15] Li, S., Li, Y., Boguyszk, E., Liu, L., & Qu, J. (2020). Coherent anti-Stokes Raman scattering microscopy and its applications. Frontiers in Physics, 8. https://doi.org/10.3389/fphy.2020.598420.

[16] Dangerous dialysis. (2024). Available online: https://psnet.ahrq.gov/web-mm/dangerous-dialysis (accessed on 13 Aug. 2024).

[17] Iyengar, S., Sitharana, et al. (2019). Fusion of the Brooks–Iyengar algorithm and blockchain in decentralization of the data-source. Journal of Sensor and Actuator Networks, 8(1), 6. https://doi.org/10.3390/jsan8010017.

[18] [18] James, L. (2024). James 2. Available online: https://jameslike.com/writes/ns-sys-part-4 (accessed on 13 Aug. 2024).

[19] Optical spectrometer. (2020). Available online: https://en.wikipedia.org/wiki/Optical_spectrometer (accessed on 13 June 2020).

[20] Practical Byzantine fault tolerance (PBFT). (2019). Available online: https://www.geeksforgeeks.org/practical-byzantine-fault-tolerance-pbft/ (accessed on 14 Jan. 2019).

[21] Science X. (2024). Novel time-of-flight-resolved stimulated Raman scattering microscopy enables high-resolution bioimaging. Available online: https://phys.org/high-raman-microscopy-enables-high.html (accessed on 13 Aug. 2024).

[22] Science Direct topics in computer science. (n.d.). Available online: https://www.sciencedirect.com/topics/computer-science.

[23] Sheehan, T., & Yost, R. A. (2015). What's the most meaningful standard for mass spectrometry: Instrument detection limit or signal-to-noise ratio? Spectroscopy Online, 16(16), 2216–2222.

[24] Singaravelu, S. (n.d.). Practical guide of deep learning networks.

[25] Wang, M. C., Fleury, B., Shapiro, N. I., et al. (n.d.). Coherent anti-Stokes Raman scattering (CARS) microscopy for the diagnosis of early kidney injury. Journal of the American Society of Nephrology.