TrAC Trends in Analytical Chemistry
Volume 124, March 2020, 115796

Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering

Félix Lussier [a b 1], Vincent Thibault [a 1], Benjamin Charron [a 1], Gregory Q. Wallace [a 1], Jean-Francois Masson [a]

Show more

https://doi.org/10.1016/j.trac.2019.115796

Abstract

Machine learning is shaping up our lives in many ways. In analytical sciences, machine learning provides an unprecedented opportunity to extract information from complex or big datasets in chromatography, mass spectrometry, NMR, and spectroscopy, among others. This is especially the case in Raman and surface-enhanced Raman scattering (SERS) techniques where vibrational spectra of complex chemical mixtures are acquired as large datasets for the analysis or the imaging of chemical systems. The classical linear methods of processing the information no longer suffice and thus machine learning methods for extracting the chemical information from Raman and SERS experiments have been implemented recently. In this review, we will provide a brief overview of the most common machine learning techniques employed in Raman, a guideline for new users to implement machine learning in their data analysis process, and an overview of modern applications of machine learning in Raman and SERS.

Graphical abstract

Download : Download high-res image (219KB)
Download : Download full-size image

Introduction

Raman spectroscopy is a vibrational spectroscopy which has been extensively applied in the past decades in analytical sciences. Due to the nature of the Raman scattering, a vibrational fingerprint intrinsic to the molecule is acquired, thus enabling the identification of the analyte. As with most spectroscopic techniques, Raman requires advanced data processing to extract meaningful information from spectra. To analyze the optically rich and complex signal, linear regression and multiple multivariate data analysis methods have been developed to recognize vibrational fingerprints with high accuracy, sensitivity, and selectivity. Such data analysis typically employed a form of artificial intelligence (AI) to automate data analysis. AI, as far as data science entails, corresponds to the usage of a computing system which applies a mathematical model meticulously developed by the user (i.e. hard-coded). For instance, a linear regression is a form of simple AI where strict mathematical rules are applied and followed in order to extract information such as the slope and intercept of a curve. In some cases, an AI system requires the capability to extract its own relevant features, patterns, or knowledge from the data (i.e. the machine learns from the data set). This is usually the case when applying AI to vibrational spectroscopy to analyze vibrational fingerprints. This capability of an AI system is referred as machine learning. Thus, it was identified in the 1990s that machine learning methods, could efficiently discriminate between spectra of different molecules [1] and be more effective than linear regression methods for data analysis in Raman spectroscopy [2]. Hence, machine learning algorithms were then applied in fields such as food analysis [3], bacteria identification [4], diagnostic application [5], and material analysis [6].

The interest in machine learning methods has resurged lately, due to the rise of surface-enhanced Raman scattering (SERS) [7], increased computing power, and the availability of open source machine learning libraries (i.e. TensorFlow®, Keras and PyTorch) or many application program interface (API) available in commercial multivariate data analysis software. Several characteristics of SERS makes it ideally suited for the use of machine learning data processing. The use of plasmonic materials solved the low sensitivity issue of Raman, and SERS is thus capable of single molecule detection [8,9]. However, the Raman spectra of single molecules show strong variations in intensity and spectral profiles due to the effects of orientation of the molecules with respect to the SERS surface. This leads to the unsuitability of linear based regressions (i.e. partial least square) to capture the various possible Raman spectra of single molecules. On the contrary, machine learning based methods are well adapted to capture complex relationships within large sets of spectra from single molecule experiments. Machine learning methods can be trained to recognize features in Raman (or SERS) spectra, and assign them to the proper label, which corresponds to the identity of the analyte. Consequently, a plethora of analytical sensors employed a Raman based detection method with machine learning for molecular analysis and chemical sensing, and many of these studies will be discussed later in this review.

To identify features and employ them to perform classification or regression, an algorithm will typically employ various forms of machine learning. By definition, machine learning corresponds to a system capable of acquiring knowledge by extracting features from raw data, and then using this newly gained knowledge to tackle problems of the real world by making decisions [10]. For example, simple machine learning models, such as logistic regression, can recommend a caesarean delivery over natural delivery [11], or separate a legitimate e-mail from a malicious one [10]. However, the limited performance of various machine learning models depends mostly on the quality of the features that were presented to the model to perform its classification operation, consequently requiring extensive features extraction and selection. Such limitation was recently overcome using deep learning, a representation-learning based method which autonomously extracts relevant features, and then uses this information to perform classification or regression tasks. This type of model is considered a subset of machine learning, also respecting its definition. To achieve high performance, deep learning models employ multiple levels (or layers) of representation, of increasing abstraction, which ultimately allow models to learn very complex functions and extrapolate from its inputs. Deep learning models have thus been applied in image [12,13], [14]] and speech [15,16] recognition, analysis of data from a particle accelerator [17], predicting the activity of new therapeutics [18], and predicting the effects of potential mutations on non-coding DNA of the genes' expression and diseases state [19,20] to only name a few examples. Thus, it is unsurprising that deep learning has recently been used in many diverse fields, enabling the extraction and analysis of highly valuable information and has attracted other fields of sciences.

The most common form of machine/deep learning is known as supervised learning, which implies the training of a model by presenting multiple known examples of the different objects, classes of interest or output (i.e. target molecules, disease states, among others) measured in similar conditions. The model then learns by extracting knowledge from the data and performs a generalization of the classification problem. More specifically to the topic of this review, new Raman spectra are sequentially presented to the model, which predicts the corresponding output.

Many supervised learning algorithms have been employed to analyze Raman spectra, which can be separated into distinct sub-methods: (1) methods based on discriminant analysis such as linear discriminant analysis (LDA), partial least-squares followed by a discriminant analysis (PLS-DA), or discriminant function analysis (DFA); (2) artificial neural network (ANN) based models such as the multilayer perceptron (MLP), convolutional neural network (CNN), or support vector machine (SVM); (3) models based on regression analysis like multiple linear regression (MLR), principal components regression (PCR), or partial least squares (PLS); (4) models based on regression trees like random forests (RF), and classification and regression trees (CART); and finally (5) evolutionary based algorithms such as genetic algorithm (GA), genetic programming and computing (respectively GP and GC), evolutionary algorithm (EA), and evolutionary programming (EP) [21]. A review of the most important models is provided in the next section, while the following section will provide the reader with useful information on how to apply these models in their research.

Section snippets

Discriminant analysis-based methods

Linear discriminant analysis (LDA) and discriminant function analysis (DFA), also referred to as canonical variates analysis (CVA), are supervised learning methods largely employed in the field of spectroscopy. These methods are based on the extraction of different hyperplanes, or linear functions, that effectively enable discrimination between two or more classes in multivariate data. These discriminant algorithms search for an optimal hyperplane which maximizes the variance between the…

Getting started with machine learning in Raman spectroscopy

Now that the different models have been presented, this section of the review aims at defining some common language in machine learning as well as providing some advices to ease the entry barrier to apply machine learning for Raman data analysis. Going in-depth on every machine learning algorithm is beyond the scope of this review and therefore, most of the further discussion will focus on supervised learning methods as these are more common in contemporary Raman spectroscopy. More in depth…

Applications

In this section, we will highlight existing and emerging applications involving the combination of deep learning and artificial intelligence methods with Raman and/or SERS. We have chosen to focus on four of the most common categories: food and beverage, forensics, bacteria and viruses, and medical diagnostics. This is by no means a list of all possible applications, it is simply designed to demonstrate the broad scope of applications….

Outlook and concluding remarks

Deep learning and artificial intelligence methods have had a significant impact on the development of contemporary Raman and SERS sensors, and are poised to have a greater impact in the near future of the field. The development of open source algorithms in addition to commercial software are lowering the entry barrier of these data processing methods in the analysis of Raman and SERS spectra. The improvement of computing power from modern personal computer also facilitates the use of these…

Acknowledgements

This work was funded by the Natural Science and Engineering Research Council (NSERC) of Canada (Grant no. RGPIN/03864-2016) and by the Fonds de Recherche du Québec, Canada (Grant no. 2019-AUDC-262722). The authors would like to dedicate this paper to Prof. Frank Vogt from the University of Tennessee, who suddenly passed away in late 2018, with whom the authors had many discussions on the use of deep learning and artificial intelligence in vibrational spectroscopy. F.L. acknowledges the…

References (102)

I.R. Lewis et al.
Raman spectrometry and neural networks for the classification of wood types - 1
Spectrochim. Acta A (1994)

P.S. Gromski et al.
A tutorial review: metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding
Anal. Chim. Acta (2015)

N. Kumar et al.
Chemometrics tools used in analytical chemistry: an overview
Talanta (2016)

Z. Xiaobo et al.
Variables selection methods in near-infrared spectroscopy
Anal. Chim. Acta (2010)

J. Yong et al.
Deep learning for vibrational spectral analysis: recent progress and a practical guide
Anal. Chim. Acta (2019)

J.-L. Li et al.
Determination of trace thiophanate-methyl and its metabolite carbendazim with teratogenic risk in red bell pepper (Capsicum annuum L.) by surface-enhanced Raman imaging technique
Food Chem. (2015)

P. Reokrungruang et al.
A simple paper-based surface enhanced Raman scattering (SERS) platform and magnetic separation for cancer screening
Sens. Actuators B Chem. (2019)

J. Engel et al.
Breaking with trends in pre-processing?
Trends Anal. Chem. (2013)

R.S. Uysal et al.
Determination of butter adulteration with margarine using Raman spectroscopy
Food Chem. (2013)

H. Mohamadi Monavar et al.
Determining quality of caviar from Caspian Sea based on Raman spectroscopy and using artificial neural networks
Talanta (2013)

View more references

Cited by (345)

Faceted gold nanocrystal film over nanosphere for sensitive SERS distinction of fungicides by deep learning analysis
2024, Sensors and Actuators B: Chemical
Show abstract

Raman spectroscopy by a detector-noise-limited birefringent wedge-based Fourier transform spectrometer
2024, Optics and Lasers in Engineering
Show abstract

Revealing the adulteration of sesame oil products by portable Raman spectrometer and 1D CNN vector regression: A comparative study with chemometrics and colorimetry
2024, Food Chemistry
Show abstract

Dynamic surface-enhanced Raman spectroscopy and positively charged probes for rapid detection and accurate identification of fungal spores in infected apples via deep learning methods
2024, Food Control
Show abstract

Does the fish rot from the head? Hyperspectral imaging and machine learning for the evaluation of fish freshness
2024, Chemometrics and Intelligent Laboratory Systems
Show abstract

Unified multi-parameter predictive modeling of crude oil and its distilled fractions through Artificial Neural Networks
2024, Fuel
Show abstract

View all citing articles on Scopus

[1] These Authors contributed equally to the manuscript.

View full text

Part of special issue

New Directions in Sensor Technology
Edited by Junwen Liu

View special issue

Recommended articles

Single convolutional neural network model for multiple preprocessing of Raman…
Vibrational Spectroscopy, Volume 121, 2022, Article 10…
Jiahao Shen, …, Wei Zhang

Feature visualization of Raman spectrum analysis with deep convolutional neural…
Analytica Chimica Acta, Volume 1085, 2019, pp. 11-19
Hiroshi Fukuhara, …, Hiroyasu Itoh

Raman spectroscopy and machine-learning for edible oils evaluation
Talanta, Volume 256, 2023, Article 125276
Cornelia Berghian-Grosan, Dana Alina Magdas

Show 3 more articles

Article Metrics