University of Victoria | Graduate Studies

Notice of the Final Oral Examination for the Degree of Master of Applied Science

of

LUYUN GAN

BSc (Nakai University, 2012)

"Multi-label Classification with Optimal Thresholding for Multi-composition Spectroscopic Analysis"

Department of Electrical and Computer Engineering

Monday, August 26, 2019
1:00 P.M.
Engineering Office Wing
Room 430

Supervisory Committee:
Dr. Tao Lu, Department of Electrical and Computer Engineering, University of Victoria (Supervisor)
Dr. Wu-Sheng Lu, Department of Electrical and Computer Engineering, UVic (Member)

External Examiner:
Dr. George Tzanetakis, Department of Computer Science, UVic

Chair of Oral Examination:
Dr. Ajtony Csaba, School of Music, UVic

Dr. David Capson, Dean, Faculty of Graduate Studies

Abstract

Spectroscopic analysis sees plural applications in physics, chemistry, bioinformatics, geophysics, astronomy, etc. It has been widely used for detecting mineral samples, gas emission, and food volatiles. Machine learning algorithms for spectroscopic analysis focus on either regression or single-label classification problems. Using multi-label classification to identify multiple chemical components from the spectrum, is under explored. In this thesis, we implement multi-label neural networks with optimal thresholding (FNN-OT) to identify gas species among a multi gas mixture in a cluttered environment. Spectrum signals are firstly processed by a FNN model, which produces one output score for each gas. Output scores will be the input of a following optimal thresholding (OT) system. Predictions of each gas in one testing sample will be made by comparing its output score from FNN with threshold from OT. If its output score is larger than the threshold, the prediction is 1 and 0 otherwise, representing the existence/non-existence of that gas component in the spectrum.

Using infrared absorption spectroscopy and tested on synthesized spectral datasets, our approach outperforms FNN itself and conventional binary relevance - partial least squares discriminant analysis (PLS-BR). All three models are trained and tested on 18 synthesized datasets with 6 levels of SNR and 3 types of gas correlation. They are evaluated and compared with micro, macro and sample averaged precision, recall and F1 score. For mutually independent and randomly correlated gas data, FNN-OT yields better performance than FNN itself or the conventional PLS-BR, by significantly increasing its recall without much sacrifice of its precision. For positively correlated gas data, FNN-OT is more capable of capturing information of positive label correlation from noisy datasets than the other two models.