AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Computer and Systems Engineering

Multivariate Calibration and Classification Modeling in Spectroscopy Applications

A Thesis submitted in partial fulfillment of the requirements of
Master of Science in Electrical Engineering
(Computer and Systems Engineering)

by
Mohamed Hossam El-Din Mohamed Mokhtar
Bachelor of Science in Electrical Engineering
(Computer and Systems Engineering)
Faculty of Engineering, Ain Shams University, 2014

Supervised By
Prof. Dr. Mohamed Watheq Ali Kamel El-Kharashi
Prof. Dr. Amr Galal el-din Ahmed Wassal

Cairo, 2019

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Computer and Systems Engineering

Multivariate Calibration and Classification Modeling in Spectroscopy Applications

by

Mohamed Hossam El-Din Mohamed
Bachelor of Science in Electrical Engineering
(Computer and Systems Engineering)
Faculty of Engineering, Ain Shams University, 2014

Examiners' Committee

Name and affiliation	**Signature**
Prof. Dr. Mohsen Abd-ElRazek Ali Rashwan Department of Electronics and Electrical Communications Engineering Faculty of Engineering, Cairo University.	....
Prof. Dr. Hoda Korashy Mohamed Department of Computer and Systems Engineering Faculty of Engineering, Ain Shams University.	....
Prof. Dr. Amr Galaleldin Ahmed Wassal Department of Computer Engineering Faculty of Engineering, Cairo University.	....
Prof. Dr. Mohamed Watheq Ali Kamel El-Kharashi Department of Computer and Systems Engineering Faculty of Engineering, Ain Shams University.	....

Date:26 10 2019

Statement

This thesis is submitted as a partial fulfillment of Choose Degree in Choose Branch, Faculty of Engineering, Ain shams University. The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

Mohamed Hossam El-Din Mohamed

Signature

....

Date: 26 10 2019

Researcher Data

Name: Mohamed Hossam El-Din Mohamed
Date of Birth: 05/08/1991
Place of Birth: Cairo, Egypt
Last academic degree: Bachelor of Science
Field of specialization: Computer and Systems Engineering
University issued the degree: Ain Shams University
Date of issued degree: June, 2014
Current job: Embedded Software Engineer in Si-ware Systems

Abstract

Recent technology trends to miniaturize spectrometers have opened the doors for mass production of spectrometers and for new applications that were not possible before and where the spectrometer can possibly be used as a ubiquitous spectral sensor. However, with the miniaturization from large reliable bench-top instrument to chip-size miniaturized spectrometers leads to new issues and challenges which are introduced into building multivariate spectroscopic models based on these spectrometers.

The purpose of this thesis is to study the feasibility of building multivariate models based on miniaturized Fourier Transform Near-Infrared (FT-NIR) spectrometers, and determine the issues emerged and propose appropriate solutions for them. The thesis presents some classification models with different natures, and each model introduces a new challenge associated with our proposed handling, the models are textiles type classification, coffee classification according to caffeine level, species type classification and milk classification according to fat level models. In addition, the thesis presents two regression applications with commercial standards, namely, milk regression application and health care application.

In this work, we proposed a calibration transfer technique to mitigate the effect of unit-to-unit variations, variations due to changing the measurement setup and variations due to changing the measurement medium. The technique shows a significant improvement in the performance of the models subjected to this kind of variations, like in the milk classification model and the regression models of the health care application.

Summary

This thesis aims to study the feasibility of building multivariate models based on miniaturized Fourier Transform Near-Infrared (FT-NIR) spectrometers, and determine the issues emerged and propose appropriate solutions for them. New technologies in miniaturizing spectrometers have enabled spectrometers to enter the mass production phase and opened the doors in front of new applications that were not possible before where the spectrometer can be used as a ubiquitous spectral sensor. However, with the miniaturization from large reliable bench-top instrument to chip-size miniaturized spectrometers, new issues and challenges are introduced into building multivariate spectroscopic models based on these spectrometers.

In the thesis, we contributed to the field of spectroscopic applications by presenting several classification models based on miniaturized FT-NIR spectral sensors. Also, we presented complete regression models to predict the quantity of fat, protein and lactose in raw cow milk, and the models in the health care application. In addition, we proposed a solution for the unit-to-unit variations that might appear while using miniaturized spectrometers, like NeoSpectra micro DVK and NeoSpectra micro sensors in our case.

The thesis is divided into six chapters as described below:

Chapter 1 is an introduction to the thesis and the motivation of utilizing near infrared spectroscopy in materials inspection. The chapter also provides an overview on the modeling challenges and sheds light upon the contribution of the thesis. At the end of the chapter, it elucidates the structure of the thesis.

Chapter 2 covers the most important concepts and topics related to spectroscopy and modeling from a theoretical point of view. The chapter was divided into two main sections, namely, "Spectroscopy and Instrumentation" and "Modeling". In the spectroscopy and instrumentation section, we covered the theoretical background related to the vibrational spectroscopy which is the core concept on which all the thesis work based, abstracted working theory of the instrumentations, the common measuring configurations, and the specifications and features of the instrument we used in this work.

In the modeling section, we covered the bits and pieces required to build a spectroscopic model in both of its forms, classification or regression.

Chapter 3 reviews both common and new techniques used in building spectroscopic models. The chapter was divided into two parts to cover the classification techniques in the first one and the regression techniques in the second part. The chapter covered some new techniques that weren't mentioned in chapter 2, such as PLS-DA, SIMCA, SVM, ANN and CNN. The chapter also presented some of the applications that were built using the mentioned techniques.

Chapter 4 we have introduced in this chapter the classification applications built using spectral sensors and the details of the models of each application and how we collected its data. The applications introduced in this chapter belong to diverse nature and contain different materials, namely, the textiles application, coffee application, spices application, and milk application. The chapter also introduced a common issue in the models based on miniaturized spectrometers which is model generalization, and we proposed the calibration transfer algorithm to solve this issue. The calibration transfer algorithm showed a significant improvement in the milk model according to the three experiments conducted to test the algorithm.

Chapter 5 we have presented our work in building two commercial level applications, the milk application and the health case application. In the milk application, we have showed the procedure we followed in collecting our dataset and all the details concerning the model building and validating. Also, we introduced an enhancement in the milk model which showed an improvement in the models' repeatability and accuracy which reached 19% improvement in prediction error and 20% improvement in the coefficient of determination.

On the other hand, we presented the model built for the health care application which was built using two Micro sensors and the tested versus a testing dataset which comprises data from several different Micro sensors. The model used the same concept of the calibration transfer to be able to accommodate all the different sensors.

Chapter 6 states the conclusion of the thesis and review its main points. The chapter also will list the potential topics that can be explored to extend this work in the future.

Keywords: Chemometrics, Cross Validation, Detrending, Fourier Transform Near Infrared Spectrometer, Gaussian Process Classification, Model Transfer, Multiplicative Scatter Correction, NeoSpectra, Orthogonal Signal Correction, Partial Least Squares, Principle Component Analysis, Spectroscopy, Standard Normal Variate , Unit-to-unit Variations.

Acknowledgment

All praise and thanks to Allah, lord of the worlds who taught by the pen, taught the human what he never knew, for granting me the power and ability to complete this work. I would like to express my gratitude to my supervisors, Prof. Mohamed Watheq Ali Kamel El-Kharashi and Prof. Amr Galal el-din Ahmed Wassal for their effort, guidance and patience throughout the course of this work. Also, I'd like to thank Si-Ware Systems for their support they gave me, there are no words can really describe my appreciation, and I'd like to thank all my colleagues and my managers there, especially Eng. Mostafa Medhat and Eng. Ahmed Fadeel. Last but not least, I want to dedicate my last words to thank my parents and my family for their unconditional love and continuous support and encouragement.

Mohamed Hossam El-Din Mohamed
Department of Computer and Systems Engineering
Faculty of Engineering
Ain Shams University
Cairo, Egypt
October 2019

Contents