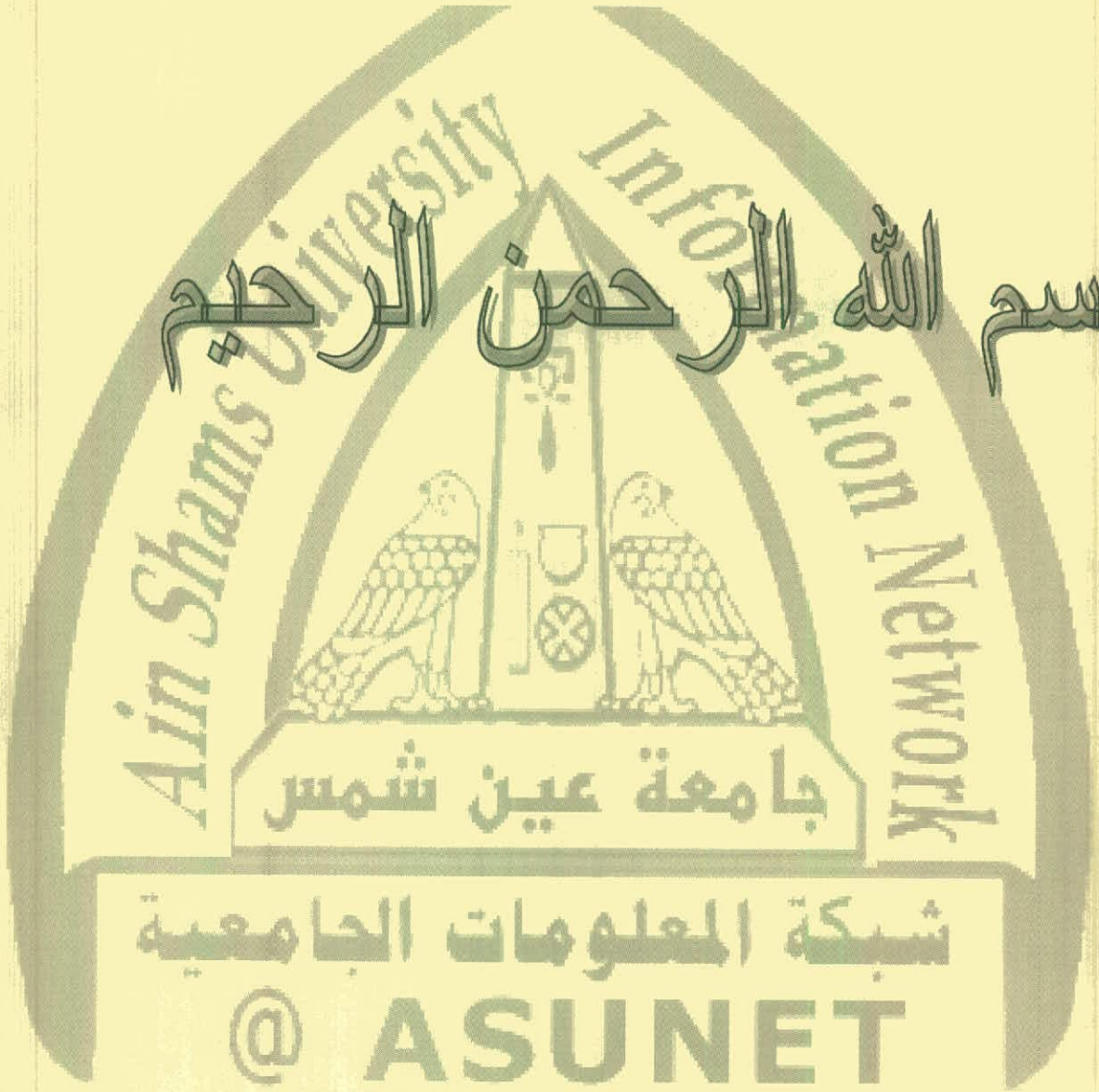
شبكة المعلومات الجامعية



بسم الله الرحمن الرحيم

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأفلام قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأفلام بعيدا عن الغبار

في درجة حرارة من ١٥-٢٥ مئوية ورطوبة نسبية من ٢٠-٤٠%

To be Kept away from Dust in Dry Cool place of 15-25- c and relative humidity 20-40%

بعض الوثائـــــــق الاصليـــــة تالفـــة

بالرسالة صفحات لم ترد بالاصل

Cairo University
Faculty of Economics and Political Science
Department of Statistics

A Mathematical Programming Approach for Nonmetric Multidimensional Scaling

Prepared by

Nada Mohammed Hafez Ibrahim

Supervised by

Dr. Afaf Ali Hassan El-Dash
Prof. of Statistics
Head of Dept. of Mathematics, Insurance and Applied Statistics
Faculty of Commerce
Helwan University

Dr. Ali El-Hefnawy
Associate Prof. of Statistics
Dept. of Statistics
Faculty of Economics and Political Science
Cairo University

A Thesis Submitted to the Department of Statistics, Faculty of Economics and Political Science in Partial Fulfillment of the Requirements for the M.Sc. Degree in Statistics

2011

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Afaf El-Dash, who gave me valuable advice, comments, criticism, and suggestions which enriches the thesis.

My deepest appreciation and thanks go to my supervisor, Dr. Ali El-Hefnawy, I am indebted to him for his critical reading, his thoughtful suggestions, and morale support throughout the work in the thesis.

Also I would like to thank Dr. Nadia Makary and Dr. Hegazy Zaher on their critical reading and comments that enriches the thesis.

My special thanks go to my family, specially my parents, for their care and support during my work in this thesis.

I am very grateful to my colleagues and friends for their support and help.

I would like to extend my thanks and appreciation to all my professors in faculty of Economics and political science.

A Mathematical Programming Approach for Nonmetric Multidimensional Scaling

Abstract

Multidimensional Scaling (MDS) is a data visualization method for identifying structure or clusters of objects depending on proximity measures between pairs of objects. In this thesis, nonmetric multidimensional scaling is considered, which is prone to local minimum solution when minimizing the stress function, and we do not guarantee obtaining the global optimum. For this reason, a new approach for solving nonmetric MDS problems is proposed based on mathematical programming. Two mathematical programming models are proposed, namely; nonlinear and mixed integer programming models. The proposed nonlinear programming model is compared with the classical nonmetric approach (Shepard-Kruskal algorithm) using the data of the corruption perceptions index (CPI) for 2010 for 19 Middle East countries. The comparison is based on stress value, Shepard diagram, residual diagram and the structure of derived configuration. Although Shepard-Kruskal algorithm came out with very good solution for this data set, the nonlinear programming model showed its superiority, especially for stress value and the structure of configuration. As for the proposed mixed integer programming model, it minimizes the stress function based on city block metric instead of Euclidean distance. It is linear in both objective function and constraints, so that it guarantees obtaining the global minimum solution.

Key Words

Nonmetric Multidimensional Scaling — *Mixed Integer Programming*

Nonlinear Programming — *Corruption Perceptions Index*

Supervised by

Dr. Afaf Ali Hassan El-Dash

Prof. of Statistics

Head of Dept. of Mathematics, Insurance and Applied Statistics

Faculty of Commerce

Helwan University

Dr. Ali El-Hefnawy

Associate Prof. of Statistics

Dept. of Statistics

Faculty of Economics and Political Science

Cairo University

A Thesis Submitted to the Department of Statistics, Faculty of Economics and Political Science in Partial Fulfillment of the Requirements for the M.Sc. Degree in Statistics

2011

Name: Nada Mohammed Hafez Ibrahim Nationality: Egyptian
Date and Place of Birth: 16/1/1985, Alexandria, Egypt.

Degree: Master

Specialization: Statistics

Supervisors:

Dr. Afaf Ali Hassan El-Dash
Prof. of Statistics
Head of Dept. of Mathematics, Insurance and Applied Statistics
Faculty of Commerce
Helwan University

Dr. Ali El-Hefnawy
Associate Prof. of Statistics
Dept. of Statistics
Faculty of Economics and Political Science
Cairo University

Title of the Thesis:

A Mathematical Programming Approach for Nonmetric Multidimensional Scaling

Summary of the Thesis:

Multidimensional Scaling (MDS) is a graphical representation method. The key objective of multidimensional scaling procedures is the creation of a low-dimensional picture or map for objects, subjects, or stimuli based on proximities between these items. In other words, objects are represented by points in the space and the distance between these points represents the dissimilarity between objects. The aim is to find representative configuration of points in low dimensions, especially 2 or 3 dimensions to enable visualization of data structure.

The purpose of using MDS differs according to the problem of the study. MDS could be used as a method of data exploration, and for discovering the hidden structure of data, or it could be used as a method for testing structural hypothesis.

There are two types of multidimensional scaling procedures; the metric and nonmetric multidimensional scaling. The difference between these two types depends on the nature of dissimilarities. If the dissimilarities are quantitative, we use metric MDS

scaling, whereas if the dissimilarities are qualitative (ordinal or nominal), we use nonmetric MDS scaling. This thesis is concerned with nonmetric MDS.

MDS techniques suffer from the possibility of obtaining a local minimum solution when minimizing a loss function (called stress). Therefore, many attempts in literature were presented for constructing MDS models using mathematical programming approach, in order to overcome this problem. However, all these attempts dealt only with metric MDS, and there is nothing in literature about a mathematical program for solving a nonmetric MDS problem. For this reason, this thesis presents a new approach for solving nonmetric MDS problems using mathematical programming instead of the classical approach (Shepard-Kruskal method). Two mathematical programming models were proposed to solve nonmetric MDS problems, dealing with Euclidean distance and city block metric as measures of distance between points on the MDS configuration.

Also, the problem of corruption in Middle East countries is considered, trying to create a map of middle east countries in terms of their level of corruption via the proposed nonmetric MDS analysis. The performance of the proposed mathematical programming models is compared with the classical approach for nonmetric MDS by applying them on corruption perceptions index 2010 for 19 Middle East countries.

This thesis is divided into five chapters and two appendices. The first chapter presents an introduction to multidimensional scaling procedure, its importance, the different types of MDS, and a brief discussion of the stress function and MDS interpretability.

The second chapter presents mathematical programming models that were presented in literature for solving metric multidimensional scaling problems.

The third chapter presents the two suggested mathematical programming models for solving nonmetric MDS problems. The first model is a nonlinear programming model, which was proposed to minimize the stress function of the classical approach based on Euclidean distance, and it was solved using Generalized Reduced Gradient (GRG) method. The second model is a mixed integer programming model for nonmetric

MDS based on city block metric, the model is in a linear form and it was solved using Branch and Bound method.

In the fourth chapter, the proposed models for nonmetric MDS are applied on the data set of corruption perceptions index 2010, and their performance is compared with the classical algorithm of nonmetric MDS (ALSCAL). The comparison is based on stress value, Shepard diagram, residual diagram and the structure of derived configuration.

The last chapter summarizes the most important conclusions of this thesis and some points for future work. The most important study conclusions are as follows:

By comparing ALSCAL algorithm and NLP model, we found that the NLP model produced configurations with better stress values than the ALSCAL algorithm for both unidimensional and two-dimensional scaling. As for the unidimensional scaling, both methods produced, in general, solutions that are representative to data, and could be described as excellent fits, with approximately zero residuals, even though the stress value for NLP model is better than that of ALSCAL algorithm. When two-dimensional scaling is concerned, the stress value for ALSCAL becomes greater than the one for unidimensional scaling, which means that the second dimension does not add too much to the configuration. However, adding the second dimension for the NLP model illustrates certain classification of countries; indicating highly clean and medium clean countries in one direction, and medium corrupted and highly corrupted countries in the other direction.

As for the proposed mixed integer programming model, it minimizes an objective function based on the city block metric instead of the Euclidean distance, therefore we cannot compare the stress values with the other nonlinear models directly. However, its linear formulation guarantees a global minimum solution for the case of city block metric, and it came out with a representative configuration for CPI 2010.

Moreover, two appendices are included, appendix (A) gives the commands of the software packages- GAMS and SPSS- used in this thesis, and Appendix (B) gives values of disparities and other output values obtained from the suggested models and ALSCAL algorithm.