Chapter 5
Enhancing Web Data Mining:
The Study of Factor Analysis

Abhishek Taneja
S. A. Jain College, India

ABSTRACT

An enormous production of databases in almost every area of human endeavor particularly through web has created a great demand for new, powerful tools for turning data into useful, task-oriented knowledge. The aim of this study is to study the predictive ability of Factor Analysis a web mining technique to prevent voting, averaging, stack generalization, meta- learning and thus saving much of our time in choosing the right technique for right kind of underlying dataset. This chapter compares the three factor based techniques viz. principal component regression (PCR), Generalized Least Square (GLS) Regression, and Maximum Likelihood Regression (MLR) method and explores their predictive ability on theoretical as well as on experimental basis. All the three factor based techniques have been compared using the necessary conditions for forecasting like R-Square, Adjusted R-Square, F-Test, JB (Jarque-Bera) test of normality. This study can be further explored and enhanced using sufficient conditions for forecasting like Theil's Inequality coefficient (TIC), and Janur Quotient (JQ).

INTRODUCTION

Factor analysis is a collection of techniques employed to explore underlying latent variables/factors which influence the outcomes on a number of measured variables. All of the techniques use common factors in their underlying model which is shown in Figure 1.

This model describes in Figure 1 that in a factor based model every observed measure/prediction from measure 1 to measure 5 is influenced by the underlying latent variables/common factors. These common factors i.e., A1 to A5 are also described latent variables and demonstrates the correlation among the different factors because of the more factors in (Kim, Jae-on., Mueller, Charles W., 1978).

Factor based techniques are actually a one-sample technique (Rencher C. Alvin, 2002). For example, the author thinks a sample X_1, X_2, X_n from an identical population with a mean vector μ and covariance matrix $\sum$. Factor based model represents each variable as a linear collection of essential *common factors*

Figure 1. Factor model
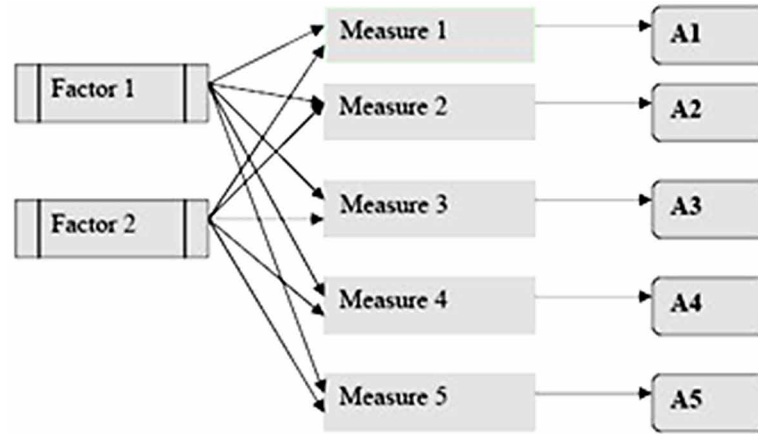
f_1, f_2, f_m, with a linked residual term to account for that part of the variable that is unique. For X_1, X_2, X_p in any observation vector X, the model is as follows:

$$X_1 - \mu_1 = \lambda_{11}f_1 + \lambda_{12}f_2 + \cdots + \lambda_{1m}f_m + \varepsilon_1$$

$$X_2 - \mu_2 = \lambda_{21}f_1 + \lambda_{22}f_2 + \cdots + \lambda_{2m}f_m + \varepsilon_2$$

...

$$X_p - \mu_p = \lambda_{p1}f_1 + \lambda_{p2}f_2 + \cdots + \lambda_{pm}f_m + \varepsilon_p.$$

If possible, m should be considerably smaller than p; or else the author have not achieved a sensible explanation of the variables as functions of a few underlying factors (Kim, Jae-on & Charles W. Mueller, 1978). In the above equation, f's in random variables that make the X's. The loadings which serve as weights are the coefficients i.e., λ_{ij}. They display how every X_i independently depends on the f's. λ_{ij} *describes* the significance of the jth factor f_j to the ith variable X_i and can also be used in explanation of f_j. The author explains f_2, for example, by examining its coefficients, λ_{12}, λ_{22}, λ_{p2}. The larger loadings associate f_2 to its corresponding X's. From these X's, the author deduce a meaning or description of f_2. After estimating the coefficients λ_{ij}'s, it is assumed that they will segregate the variables into parts equivalent to factors. Initially it appears that the MLR and factor analysis are similar techniques but there are fundamentally different because f's in above equations are unobserved and equations above represents one observational vector, whereas MLR represents all n observations.

MAIN FOCUS OF THE CHAPTER

An enormous production of databases in almost every area of human endeavor particularly with the introduction of web has created a great demand for new, powerful tools for turning data into useful,

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/enhancing-web-data-mining/162891

Related Content

Exploration of College Students' Learning Adaptability Under the Background of Wisdom Education
Henan Zhangand Xiangzhe Liu (2024). *International Journal of Information Technology and Web Engineering (pp. 1-12).*
www.irma-international.org/article/exploration-of-college-students-learning-adaptability-under-the-background-of-wisdom-education/336486

Block-Chain-Based Security and Privacy in Smart City IoT: Distributed Transactions
Thangaraj Muthuraman, Punitha Ponmalar Pichiahand Anuradha S. (2019). *Handbook of Research on Implementation and Deployment of IoT Projects in Smart Cities (pp. 134-148).*
www.irma-international.org/chapter/block-chain-based-security-and-privacy-in-smart-city-iot/233270

Metamodeling Architectures and Interoperability of Web-Enabled Information Systems
Marie-Noella Terrasse, George Beckerand Marinette Savonnet (2003). *Web-Enabled Systems Integration: Practices and Challenges (pp. 1-18).*
www.irma-international.org/chapter/metamodeling-architectures-interoperability-web-enabled/31407

Design and Prototyping of a Smart University Campus
Vincenzo Ciminoand Simona E. Rombo (2019). *Handbook of Research on Implementation and Deployment of IoT Projects in Smart Cities (pp. 228-252).*
www.irma-international.org/chapter/design-and-prototyping-of-a-smart-university-campus/233275

Toward a More Robust Home Automated System: Introducing a Voice Activation Technique via Pubnub
Segun Aina, Samuel Dayo Okegbile, Adeniran Ishola Oluwaranti, Oghenerukome Brenda Okoroand Tayo Obasanya (2019). *International Journal of Information Technology and Web Engineering (pp. 78-92).*
www.irma-international.org/article/toward-a-more-robust-home-automated-system/227689