

Association for Information Systems

AIS Electronic Library (AISeL)

NEAIS 2022 Proceedings

New England Chapter of Association for
Information Systems

10-29-2022

Parameterizing Topic Models for Empirical Research

Xiaoping Liu

Northeastern University, xia.liu@northeastern.edu

Xiaobai Li

University of Massachusetts Lowell, xiaobai_li@uml.edu

Follow this and additional works at: <https://aisel.aisnet.org/neais2022>

Recommended Citation

Liu, Xiaoping and Li, Xiaobai, "Parameterizing Topic Models for Empirical Research" (2022). *NEAIS 2022 Proceedings*. 33.

<https://aisel.aisnet.org/neais2022/33>

This material is brought to you by the New England Chapter of Association for Information Systems at AIS Electronic Library (AISeL). It has been accepted for inclusion in NEAIS 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Parameterizing Topic Models for Empirical Research

Completed Research Paper

Xiaoping Liu
Northeastern University
xia.liu@northeastern.edu

Xiao-Bai Li
University of Massachusetts Lowell
Xiaobai_li@uml.edu

ABSTRACT

Machine learning techniques have been increasingly employed in business research to discover or extract new simple features from large and unstructured data. These machine learned features (MLFs) are then used as independent or explanatory variables in the main econometric models for empirical research. Despite this growing trend, there has been little research regarding the impact of using MLFs on statistical inference for empirical research. In this paper, we undertake parameter estimation issues related to the use of topics extracted by Latent Dirichlet Allocation, a popular machine learning technique for text mining. We propose a novel method to extract features that result in the minimum-variance estimation of the regression model parameters. This enables a better use of unstructured text data for econometric modeling in empirical research. The effectiveness of the proposed method is validated with an experimental evaluation study on real-world text data.

Keywords

Machine learning, text mining, latent Dirichlet allocation, regression.