STATISTICAL ANALYSIS AND DATA MINING: THE ASA DATA SCIENCE JOURNAL

CONTENTS

RESEARCH ARTICLES

Minimizing information loss in shared data: Hiding frequent patterns with multiple sensitive thresholds
B. Ergenç Bostanoğlu and A. C. Öztürk

Knot selection in sparse Gaussian processes with a variational objective function
N. Garton, J. Niemi and A. Carriquiry

Composite quantile-based classifiers
D. A. Pritchard and Y. Liu

GRATIS: GeneRAtion TIme Series with diverse and controllable characteristics
Y. Kang, R. J. Hyndman and F. Li

Objective identification of local spatial structure for material characterization
Y. Yu, B. P. Gorman and A. S. Hering

RAPID RESEARCH ARTICLE

Weighted k-nearest neighbor based data complexity metrics for imbalanced datasets
D. Singh, A. Gosain and A. Saha

RESEARCH ARTICLE

MR plot: A big data tool for distinguishing distributions
O. M. Ardakani, M. Asadi, N. Ebrahimi and E. S. Soo
Scope of Statistical Analysis and Data Mining: The ASA Data Science Journal

The Journal addresses the broad area of data analysis, including data mining algorithms, statistical approaches, and practical applications. Topics include problems involving massive and complex datasets, solutions utilizing innovative data mining algorithms and/or novel statistical approaches, and the objective evaluation of analyses and solutions. Of special interest are articles that describe analytical techniques, and discuss their application to real problems, in such a way that they are accessible and beneficial to domain experts across science, engineering, and commerce.

The focus of the journal is on papers which satisfy one or more of the following criteria:

- Solve data analysis problems associated with massive, complex datasets
- Are application and solution oriented with a focus on solving real problems
- Describe innovative data mining algorithms or novel statistical approaches
- Compare and contrast techniques to solve a problem, along with an objective evaluation of the analyses and the solutions

The goals of this interdisciplinary journal are to encourage collaborations across disciplines, communication of novel data mining and statistical techniques to both novices and experts involved in the analysis of data from practical problems, and a principled evaluation of analyses and solutions. All submitted manuscripts are evaluated for their scientific content and significance by the Editors. Manuscripts that are deemed appropriate for the journal are also evaluated by an Associate Editor and 2–3 independent reviewers.

Copyright and Copying: Copyright © 2020 Wiley Periodicals, LLC. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organisation (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for republication, for creating new collective works or for resale. Permissions for such reuse can be obtained using the RightsLink “Request Permissions” link on Wiley Online Library. Special requests should be addressed to: permissions@wiley.com

Disclaimer: The Publisher, Statistical Analysis and Data Mining: The ASA Data Science Journal and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, Statistical Analysis and Data Mining: The ASA Data Science Journal and Editors, neither does the publication of advertisements constitute any endorsement by the Publisher, Statistical Analysis and Data Mining: The ASA Data Science Journal and Editors of the products advertised.

Production Editor: Balasubramanian (Email: samprod@wiley.com)

For submission instructions, subscription and all other information visit: wileyonlinelibrary.com/journal/SAM
CONTENTS

RESEARCH ARTICLES

Minimizing information loss in shared data: Hiding frequent patterns with multiple sensitive support thresholds
B. Ergenç Bostanoğlu and A. C. Öztürk ................................................................. 309

Knot selection in sparse Gaussian processes with a variational objective function
N. Garton, J. Niemi and A. Carriquiry ................................................................. 324

Composite quantile-based classifiers
D. A. Pritchard and Y. Liu .................................................................................... 337

GRATIS: GeneRAting TIme Series with diverse and controllable characteristics
Y. Kang, R. J. Hyndman and F. Li ........................................................................ 354

Objective identification of local spatial structure for material characterization
Y. Yu, B. P. Gorman and A. S. Hering ................................................................ 377

RAPID RESEARCH ARTICLE

Weighted k-nearest neighbor based data complexity metrics for imbalanced datasets
D. Singh, A. Gosain and A. Saha .......................................................................... 394

RESEARCH ARTICLE

MR plot: A big data tool for distinguishing distributions
O. M. Ardakani, M. Asadi, N. Ebrahimi and E. S. Soofi ...................................... 405