

Smart Innovation, Systems and Technologies 289

V. Sivakumar Reddy

V. Kamakshi Prasad

D. N. Mallikarjuna Rao

Suresh Chandra Satapathy *Editors*



# Intelligent Systems and Sustainable Computing

Proceedings of ICISSC 2021



# **Smart Innovation, Systems and Technologies**

Volume 289

## **Series Editors**

Robert J. Howlett, Bournemouth University and KES International,  
Shoreham-by-Sea, UK

Lakhmi C. Jain, KES International, Shoreham-by-Sea, UK

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago, DBLP.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <https://link.springer.com/bookseries/8767>

V. Sivakumar Reddy · V. Kamakshi Prasad ·  
D. N. Mallikarjuna Rao · Suresh Chandra Satapathy  
Editors

# Intelligent Systems and Sustainable Computing

Proceedings of ICISSC 2021

 Springer



*Editors*

V. Sivakumar Reddy  
Malla Reddy University  
Hyderabad, India

D. N. Mallikarjuna Rao  
Federation University at IIBIT  
Sydney, NSW, Australia

V. Kamakshi Prasad  
Department of Computer Science  
Engineering  
Jawaharlal Nehru Technological University  
Hyderabad, India

Suresh Chandra Satapathy  
KIIT University  
Bhubaneswar, India

ISSN 2190-3018

ISSN 2190-3026 (electronic)

Smart Innovation, Systems and Technologies

ISBN 978-981-19-0010-5

ISBN 978-981-19-0011-2 (eBook)

<https://doi.org/10.1007/978-981-19-0011-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# **Conference Committee**

## **Chief Patron**

Sri. CH. Malla Reddy, Founder Chairman, MRGI

## **Patrons**

Sri. CH. Mahendar Reddy, Secretary, MRGI

Sri. CH. Bhadra Reddy, President, MRGI

## **Conference Chair**

Dr. V. Sivakumar Reddy, Vice-Chancellor, Malla Reddy University, Hyderabad

## **Convener**

Prof. K. Kailasa Rao, Dean, School of Engineering

## **Publication Chair**

Dr. Suresh Chandra Satapathy, Professor, KIIT, Bhubaneswar

## **Co-convener**

Prof. P. Sanjeeva Reddy, Dean, International Studies

## **Organizing Chair**

Prof. G. S. Naveen Kumar, HOD, CSE (Data Science), MRUH

## **Organizing Secretaries**

Dr. E. V. Reddy, HOD, CSE, MRUH

Dr. Thayyaba Khatoon Mohammed, HOD, CSE (AI & ML), MRUH

Dr. Meeravali Shaik, HOD, CSE (Cyber Security), MRUH

## **Coordinators**

Dr. Rajasekar Rangasami, Professor, CSE (AI & ML)

Dr. B. V. V. Siva Prasad, Professor, CSE

Dr. Sudheer Nidamanuri, Professor, CSE (DS)

Dr. Arun Singh Chauhan, Professor, CSE (CS)

Dr. K. Vijaya Sekhar Reddy, Professor, MBA

## **Organizing Committee**

Dr. V. Dhanunjana Chari, Dean, School of Sciences

Dr. P. S. V. Srinivasa Rao, HOD, I Year Engineering

Dr. M. Ravikanth, Professor, CSE

Dr. Magesh Kumar, Professor, CSE (AI & ML)

Dr. Pranayath Reddy, Professor, CSE (AI & ML)

Dr. K. Srikanth, Professor, CSE (Data Science)

Prof. V. Ramachandran, CSE (AI & ML)

Mr. T. Rama Rao, CSE

Dr. G. Nanda Kishor Kumar, Professor, CSE

Ms. K. Lakshmi Madhuri, CSE (Data Science)

Mr. T. Vinay Simha Reddy, CSE (AI & ML)

Mr. T. Sanjeeva Rao, School of Engineering

## **Session Chairs**

Prof. Yu-Dong Zhang, School of Informatics, University of Leicester, UK  
Dr. Jinshan Tang, George Mason University, Virginia, United States  
Dr. Sanju Tiwari, Universidad Autonoma de Tamaulipas, Mexico  
Dr. Naga Mallikarjuna Rao Dasari, Federation University @IIBIT, Australia  
Dr. Rama Murthy Garimella, Mahindra University, Hyderabad, India.  
Dr. Bhanu Murthy Bhaskara, Ex-Professor, University of Majmaah, Saudi Arabia  
Dr. Midhun Chakkaravarthy, Lincoln University College, Malaysia  
Prof. K.Venkat Rao, Dean Academics, Andhra University, Visakhapatnam, India  
Dr. Ilaiah Kavati, National Institute of Technology, Warangal, India  
Dr. E. Suresh Babu, National Institute of Technology, Warangal, India  
Dr. M. Ramesh, Professor, RVR & JC College of Engineering, Guntur, India

# Preface

The International Conference on Intelligent Systems and Sustainable Computing (ICISSC-2021) was successfully organized by Malla Reddy University during September 24–25, 2021, at Hyderabad. The objective of this conference was to provide opportunities for the researchers, academicians and industry persons to interact and exchange the ideas, experience and gain expertise in the cutting-edge technologies pertaining to soft computing and signal processing. Research papers in the above-mentioned technology areas were received and subjected to a rigorous peer-reviewed process with the help of program committee members and external reviewers. The ICISSC-2021 received a total of 605 papers out of which 38 papers are from twenty-one countries of abroad, each paper was reviewed by more than two reviewers, and finally, 60 papers were accepted for publication in scopus-indexed Springer book series “Smart Innovation, Systems and Technologies (SIST).”

Our sincere thanks to our Chief Guest **Prof. D. N. Reddy**, Hon’ble Chancellor, MRUH, and our Guests of Honor and Keynote Speakers **Dr. Suresh Chandra Satapathy**, Professr and Dean R&D, KIIT; **Dr. Aninda Bose**, Senior Editor, Springer Publications, India; **Prof. Yu-Dong Zhang**, University of Leicester, UK; **Dr. Jinshan Tang**, George Mason University, Virginia, USA; **Dr. Mufti Mohhamad**, Department of Computing and Technology, Nottingham Trent University, UK; **Dr. Naga Mallikarjuna Rao Dasari**, Federation University @IIBIT, Australia; **Dr. Bhanu Murthy Bhaskara**, Ex-Professor, University of Majmaah, Saudi Arabia; **Dr. Midhun Chakkaravarthy**, Lincoln University College, Malaysia; **Dr. Sanju Tiwari**, Universidad Autonoma de Tamaulipas, Mexico; **Prof. Ganapati Panda**, Professorial Fellow, IIT Bhubaneswar, India; and **Dr. Rama Murthy Garimella**, Professor, Mahindra University Ecole Centrale School of Engineering, India, for extending their support and cooperation.

We would like to express our gratitude to all session chairs, viz. **Dr. M. Ramakrishna Murthy**, ANITS, Visakhapatnam; **Prof. K. Venkat Rao**, Dean Academics, Andhra University, Visakhapatnam; **Dr. Ilaiah Kavati**, National Institute of Technology, Warangal; **Dr. E. Suresh Babu**, National Institute of Technology, Warangal; **Dr. M. Ramesh**, Professor, RVR and JC College of Engineering, Guntur; **Dr. K. Mallikharjuna Lingam**, Professor and Head, ECE, MRCET Campus, Hyderabad;

**Dr. G. Sharada**, Professor and Head, CSE, MRCET Campus, Hyderabad, for extending their support and cooperation.

We are indebted to the program committee members and external reviewers who have produced critical reviews in a short time. We would like to express our special gratitude to publication chair **Dr. Suresh Chandra Satapathy**, KIIT, Bhubaneswar, for his valuable support and encouragement till the successful conclusion of the conference.

We express our heartfelt thanks to our Chief Patron **Sri. CH. Malla Reddy**, Founder Chairman, MRGI; Patrons **Sri. CH. Mahendar Reddy**, Secretary, MRGI; **Sri. CH. Bhadra Reddy**, President, MRGI; Conference Chair Dr. V. Sivakumar Reddy, Vice-Chancellor, MRUH; Convener **Prof. K. Kailasa Rao**, Dean, School of Engineering; and Co-Convener Prof. P. Sanjeeva Reddy.

We would also like to thank the organizing chair **Prof. G. S. Naveen Kumar**, HOD, CSE (Data Science), Organizing Secretaries; **Dr. E. V. Reddy**, HOD, CSE; **Dr. Thayyaba Khatoon Mohammed**, HOD, CSE (AI & ML); and **Dr. Meeravali Shaik**, HOD, CSE (Cyber Security), for their valuable contribution. Our thanks also go to all coordinators **Dr. Rajasekar Rangasami**, Professor, CSE (AI & ML); **Dr. B. V. V. Siva Prasad**, Professor, CSE; **Dr. Sudheer Nidamanuri**, Professor, CSE (DS); **Dr. Arun Singh Chauhan**, Professor, CSE (CS); **Dr. K. Vijaya Sekhar Reddy**, Professor, MBA, and the organizing committee as well as all the other committee members **Dr. V. Dhanunjana Chari**, Dean, School of Sciences; **Dr. P. S. V. Srinivasa Rao**, HOD, I Year Engineering; **Dr. Pranayath Reddy**, Professor, CSE (AI & ML); **Dr. Magesh Kumar**, Professor, CSE (AI & ML); Dr. K. Srikanth, Professor, CSE (DS); Dr. G. Nanda Kishor Kumar, Professor, CSE; Dr. M. Ravikanth, Professor, CSE; Ms. K. Lakshmi Madhuri; Prof. V. Ramachandran; Mr. T. Rama Rao; Mr. T. Vinay Simha Reddy; and Mr. T. Sanjeeva Rao for their contribution in successful conduct of the conference.

We are indebted to the program committee members and external reviewers who have produced critical reviews in a short time. Last, but certainly not least, our special thanks to all the authors without whom the conference would not have taken place. Their technical contributions have made our proceedings rich and praiseworthy.

Hyderabad, India

V. Sivakumar Reddy  
Vice Chancellor, MRUH

# Contents

<b>Unsupervised Learning for the Automatic Management of Scientific Aspect in Academic Conferences</b> .....	1
Abdeldjaouad Nusayr Medakene, Abdelhadi Mohammed Ali Eddoud, and Khadra Bouanane	
<b>Context Sensing from Deployed IoT Sensors for Improved Public Transport Systems</b> .....	15
Evariste Twahirwa, James Rwigema, and Raja Datta	
<b>College Admissions Counseling Using Intelligent Question Answering System</b> .....	27
Bui Thanh Hung	
<b>Juvenile Crime Categorization with EM Clustering</b> .....	39
Lalitha Saroja Thota, Ravinder Reddy Baireddy, Suresh Babu Chandalasetty, and Rambabu Pemula	
<b>Review of Interoperability Issues Influencing Acceptance and Adoption of Cloud Computing Technology by Consumers</b> .....	49
Gabriel Terna Ayem, Salu George Thandekkattu, and Narasimha Rao Vajjhala	
<b>Time Series Analysis Using Random Forest for Predicting Stock Variances Efficiency</b> .....	59
Parnandi Srinu Vasarao and Midhun Chakkaravarthy	
<b>Prediction of COVID-19 Severity Level Using XGBoost Algorithm: A Machine Learning Approach Based on SIR Epidemiological Model</b> .....	69
Labeba Tahsin and Shaily Roy	

**A Blockchain-Based Architecture for Implementing Job Recruitment Operations** ..... 79  
P. C. Sherimon, Vinu Sherimon, Arjun B. Dev, Mammen Jacob Ambrayil, and Moonesah Abeer

**Analysis of the Interest in the Profession of Tutor on the Internet (A Case Study of Google Trends)** ..... 89  
Oleh Karyy, Ihor Kulyniak, Oksana Ivanytska, Liubov Halkiv, and Ivan Zhygalo

**A PID Temperature Controller for Passive Solar Water Heaters with a Low-Cost Three-Way Valve** ..... 101  
Chanh-Nghiem Nguyen, Ngo Thanh The, and Luong Vinh Quoc Danh

**MATLAB/Simulink-Based Simulation Model for Minimizing the Torque Ripple Reduction in Brushless DC Motor** ..... 115  
Issa Etier, C. Arul Murugan, and Nithiyananthan Kannan

**Improving Accuracy in Facial Detection Using Viola-Jones Algorithm AdaBoost Training Method** ..... 127  
Kodeti Haritha Rani and Midhun Chakkaravarthy

**Comparative Analysis Model for Lambda Iteration, Genetic Algorithm, and Particle Swarm Optimization-Based Economic Load Dispatch for Saudi Arabia Network System** ..... 139  
Abdulrhman Alafif, Youssef Mobarak, and Hussain Bassi

**A Novel Approach for Providing Security for IoT Applications Using Machine Learning and Deep Learning Techniques** ..... 155  
M. V. Kamal, P. Dileep, and M. Gayatri

**Analysis of Bangla Keyboard Layouts Based on Keystroke Dynamics** ..... 165  
Shadman Rohan, Koushik Roy, Pritom Kumar Saha, Sazzad Hossain, Fuad Rahman, and Nabeel Mohammed

**Improved Efficiency of Semantic Segmentation using Pyramid Scene Parsing Deep Learning Network Method** ..... 175  
Pichika Ravikiran and Midhun Chakkaravarthy

**Fractional Approach for Controlling the Motion of a Spherical Robot** ..... 183  
Quoc-Dong Hoang, Le Anh Tuan, Luan N. T Huynh, and Tran Xuan Viet

**Progression of Metamaterial for Microstrip Antenna Applications: A Review** ..... 193  
Renju Panicker and Sellakkutti Suganthi



**Natural Language Processing of Text-Based Metrics for Image Captioning** ..... 203  
 Sudhakar Sengan, P. Vidya Sagar, N. P. Saravanan, K. Amarendra, Arjun Subburaj, S. Maheswari, and Rajasekar Rangasamy

**Li-Fi Enables Reliable Communication of VLT for Secured Data Exchange** ..... 213  
 N. Venkata Ramana Gupta, G. Rajeshkumar, S. Arun Mozhi Selvi, Sudhakar Sengan, Arjun Subburaj, S. Priyadarsini, and Rajasekar Rangasamy

**Design of Dual-Stack, Tunneling, and Translation Approaches for Blockchain-IPv6** ..... 221  
 Srihari Varma Mantena, S. Jayasundar, Dilip Kumar Sharma, J. Veerappan, M. Anto Bennet, Sudhakar Sengan, and Rajasekar Rangasamy

**Wavelet-Based Aerial Image Enhancement Tools and Filtering Techniques** ..... 233  
 P. Ramesh, V. Usha Shree, and Kesari Padma Priya

**Efficiency Optimization of Security and Safety in Cooperative ITS Communication for QoS Service** ..... 243  
 M. Mohan Rao, Sreenivas Mekala, Rajaram Jatothu, G. Ravi, and Sarangam Kodati

**Classification of Medicinal Plants Using Machine Learning** ..... 255  
 Rohit Sunil Meshram and Nagamma Patil

**Performance Measures of Blockchain-Based Mobile Wallet Using Queueing Model** ..... 269  
 R. Kavitha, R. Rajeswari, Pratyusa Mukherjee, Suchismita Rout, and S. S. Patra

**Conventional Data Augmentation Techniques for Plant Disease Detection and Classification Systems** ..... 279  
 Srinivas Talasila, Kirti Rawal, and Gaurav Sethi

**Directionality Information Aware Encoding for Facial Expression Recognition** ..... 289  
 A. Vijaya Lakshmi and P. Mohanaiah

**MRI Breast Image Segmentation Using Artificial Bee Colony Optimization with Fuzzy Clustering and CNN Classifier** ..... 303  
 R. Sumathi and V. Vasudevan

**Human Action Recognition in Still Images Using SIFT Key Points** ..... 313  
 M. Pavan and K. Jyothi

<b>Detecting Credit Card Fraud Using Majority Voting-Based Machine Learning Approach</b> .....	327
V. Akshaya, M. Sathyapriya, R. Ranjini Devi, and S. Sivanantham	
<b>Evaluation and Performance Analysis of Magnitude Comparator for High-Speed VLSI Applications</b> .....	335
Madhusudhan Reddy Machupalli, Krishnaveni Challa, Murali Krishna Bathula, G. Rajesh Kumar, and Raja Posupo	
<b>Artificially Ripened Mango Fruit Prediction System Using Convolutional Neural Network</b> .....	345
V. Laxmi and R. Roopalakshmi	
<b>Quality and Dimensional Analysis of 3D Printed Models Using Machine Vision</b> .....	357
S. H. Sarvesh, Kempanna Chougala, J. Sangeetha, and Umesh Barker	
<b>Traffic Density Determination Using Advanced Computer Vision</b> .....	375
Narayana Darapaneni, M. S. Narayanan, Shakeeb Ansar, Ganesh Ravindran, Kumar Sanjeev, Abhijeet Kharade, and Anwesh Reddy Paduri	
<b>Analyzing and Detecting Fake News Using Convolutional Neural Networks Considering News Categories Along with Temporal Interpreter</b> .....	387
Viraj Desai, Neel Shah, Jinit Jain, Manan Mehta, and Simran Gill	
<b>Automatic Student Attendance and Activeness Monitoring System</b> .....	405
Naveena Narayana Poojari, J. Sangeetha, G. Shreenivasa, and Prajwal	
<b>Greenput Algorithm for Minimizing Power and Energy Consumption in Hybrid Wireless Sensor Networks</b> .....	417
S. Beski Prabakaran and Saira Banu	
<b>Cloud-Based Face and Face Mask Detection System</b> .....	427
V. Muthumanikandan, Prashant Singh, and Rithwik Chithreddy	
<b>Design of Delay-Efficient Carry-Save Multiplier by Structural Decomposition of Conventional Carry-Save Multiplier</b> .....	435
M. Venkata Subbaiah and G. Umamaheswara Reddy	
<b>Speech Intelligibility Quality in Telugu Speech Patterns Using a Wavelet-Based Hybrid Threshold Transform Method</b> .....	449
S. China Venkateswarlu, N. Uday Kumar, D. Veeraswamy, and Vallabhuni Vijay	
<b>Landslide Susceptibility for Communities Based on Satellite Images Using Deep Learning Algorithms</b> .....	463
Aadityan Sridharan, A. S. Remya Ajai, and Sundararaman Gopalan	

**Comparative Study on Estimation of Remaining Useful Life of Turbofan Engines Using Machine Learning Algorithms** ..... 473  
 Eshika Shah, Nishil Madhani, Aditya Ghatak, and Abhiram Ajith Kumar

**Machine Learning-Based Framework for Malware Analysis in Android Apps** ..... 485  
 Anil Kumar Naik and K. V. Pradeepthi

**A Survey on Byzantine Agreement Algorithms in Distributed Systems** ..... 495  
 P. C. Sachin and Dua Amit

**An Autonomous Intelligent System to Leverage the Post-harvest Agricultural Process Using Localization and Mapping** ..... 507  
 Amitash Nanda and Deepak Ahire

**Zero-Day Attack Detection Analysis in Streaming Data Using Supervised Learning Techniques** ..... 517  
 B. Ida Seraphim and E. Poovammal

**Comparative Evaluation of Machine Learning Methods for Network Intrusion Detection System** ..... 531  
 Sunil Kumar Rajwar, Pankaj Kumar Manjhi, and Indrajit Mukherjee

**Rare Pattern Mining from Data Stream Using Hash-Based Search and Vertical Mining** ..... 543  
 Sunitha Vanamala, L. Padma Sree, and S. Durga Bhavani

**A Novel ARDUINO Based Self-Defense Shoe for Women Safety and Security** ..... 553  
 Swarnalatha Pasupuleti, Sattibabu Gummarekula, V. Preethi, and R. V. V. Krishna

**A New Supervised Term Weight Measure Based Machine Learning Approach for Text Classification** ..... 563  
 T. Raghunadha Reddy, P. Vijaya Pal Reddy, and P. Chandra Sekhar Reddy

**Machine Learning-Based Human Activity Recognition Using Smartphones** ..... 573  
 A. Vinay Kumar, M. Neeraj, P. Akash Reddy, and Ameet Chavan

**A Feature Selection Technique-Based Approach for Author Profiling** ..... 583  
 D. Radha and P. Chandra Sekhar

**Detection of Fake News Using Natural Language Processing Techniques and Passive Aggressive Classifier** ..... 593  
 K. Varada Rajkumar, Pranav Vallabhaneni, Krishna Marlapalli, T. N. S. Koti Mani Kumar, and S. Revathi

**Efficiency Analysis of Pre-trained CNN Models as Feature Extractors for Video Emotion Recognition** ..... 603  
Diksha Mehta, Janhvi Joshi, Abhishek Bisht, and Pankaj Badoni

**Mapping of Computational Social Science Research Themes: A Two-Decade Review** ..... 617  
Agung Purnomo, Nur Asitah, Elsa Rosyidah, Andre Septianto, and Mega Firdaus

**Diagnosis of Pneumonia with Chest X-Ray Using Deep Neural Networks** ..... 627  
E. Venkateswara Reddy, G. S. Naveen Kumar, G. Siva Naga Dhipti, and Baggam Swathi

**High Performance Algorithm for Content-Based Video Retrieval Using Multiple Features** ..... 637  
G. S. Naveen Kumar and V. S. K. Reddy

**Smart Recruitment System Using Deep Learning with Natural Language Processing** ..... 647  
Ranganath Ponnaboyina, Ramesh Makala, and E. Venkateswara Reddy

**Early Diagnosis of Age-Related Macular Degeneration (ARMD) Using Deep Learning** ..... 657  
Pamula Udayaraju and P. Jeyanthi

**Recent Trends in Calculating Polarity Score Using Sentimental Analysis** ..... 665  
K. Srikanth, N. Sudheer, G. S. Naveen Kumar, and Vijaysekhar

**Author Index** ..... 675

## About the Editors

**V. Sivakumar Reddy** is Professor in Computer Science and Engineering, Vice Chancellor, Malla Reddy University, and has an experience of more than 25 years in Teaching and Industry put together. He is an alumnus of Indian Institute of Technology, (IIT) Kharagpur, obtained his Ph.D., and he is versatile in multidisciplinary specializations in Signal Processing & Communications as well as Computer Science Engineering. His laurels include more than 150 Publications in reputed National and International Conferences and Journals. He is Fellow of IETE, Life Member of ISTE, Member of IEEE, and Member of CSI. He was awarded as “Best Teacher” in three consecutive academic years with citation and cash award. He is the recipient of “India Jewel Award” for outstanding contribution to research in the field of Engineering and Technology. He was Member of Board of Studies for JNT University, Hyderabad, India.

**V. Kamakshi Prasad** completed his Ph.D. in speech recognition at the Indian Institute of Technology Madras and his M.Tech. in Computer Science and Technology at Andhra University in 1992. He has more than 22 years of teaching and research experience. His areas of research and teaching interest include speech recognition and processing, image processing, pattern recognition, ad-hoc networks, and computer graphics. He has published several books, chapters, research papers in peer-reviewed journals, and conference proceedings. He is also Editorial Board Member of the *International Journal of Wireless Networks and Communications* and Member of several academic committees.

**D. N. Mallikarjuna Rao** is currently working as Professor in Faculty of Science and Technology, Federation University @IIBIT, Australia. He obtained his Ph.D. in Computer Science at University of South Australia, Australia, and Master of Technology at Jawaharlal Nehru Technological University, India. He has more than 20 years of teaching and research experience. He has experience in developing mathematical models, designing and developing algorithms and data analysis, and expertise in computational techniques and computer networks. He published more than

15 reputed journal papers in computation, networking principles, and graph theoretical analysis. He has experience in designing computational methodologies to solve engineering problems.

**Suresh Chandra Satapathy** is currently working as Professor, KIIT Deemed to be University, Odisha, India. He obtained his Ph.D. in Computer Science Engineering from JNTUH, Hyderabad, and Master's degree in Computer Science and Engineering from National Institute of Technology (NIT), Rourkela, Odisha. He has more than 27 years of teaching and research experience. His research interest includes machine learning, data mining, swarm intelligence studies, and their applications to engineering. He has more than 98 publications to his credit in various reputed international journals and conference proceedings. He has edited many volumes from Springer AISC, LNEE, SIST and LNCS in the past, and he is also Editorial Board Member in few international journals. He is Senior Member of IEEE and Life Member of Computer Society of India. Currently, he is National Chairman of Division-V (Education and Research) of Computer Society of India.

# Unsupervised Learning for the Automatic Management of Scientific Aspect in Academic Conferences



Abdeldjaouad Nusayr Medakene , Abdelhadi Mohammed Ali Eddoud, and Khadra Bouanane 

**Abstract** An academic conference is of great importance in the scientific community as it provides a platform for scientists and researchers to share their research findings, exchange ideas and establish new cooperative relationships with research groups around the world. Tasks related to the management of a scientific conference can be categorized according to two main aspects: the technical or scientific aspect and the administrative and logistical aspect. While tasks in both aspects are challenging and time/effort consuming, the scientific aspect of a conference requires in addition, a high level of expertise. Despite the great support that conference management systems provide to conference organizers, most of them focus mainly on the administrative and logistical aspect, while tasks of the scientific aspect are less considered. In this work, we make use of several unsupervised algorithms to automate some tasks from the scientific aspect, where expertise, time and efforts are required. Namely, the detection of out of scope papers and the creation of an accurate conference program.

## 1 Introduction

The organization of an academic conference is a prestigious event in the scientific community as it provides a platform for scientists and researchers to share their research findings, exchange ideas and establish new cooperative relationships with research groups around the world.

However, making such an event exceptional requires time and tremendous efforts since there is a plethora of must achieved tasks in a limited amount of time, especially if the main concern of organizers is to preserve or build a high reputation for their conference. This makes the web-based information management systems, commonly named *Conference Management Systems (CMSs)*, an indispensable tools that aim to assist the organizers in performing many time-consuming and complex tasks.

---

A. N. Medakene · A. M. A. Eddoud · K. Bouanane (✉)  
Department of Computer Science and IT, Kasdi Merbah University, Ouargla, Algeria  
e-mail: [bouanane.khadra@univ-ouargla.dz](mailto:bouanane.khadra@univ-ouargla.dz)

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022  
V. S. Reddy et al. (eds.), *Intelligent Systems and Sustainable Computing*,  
Smart Innovation, Systems and Technologies 289,  
[https://doi.org/10.1007/978-981-19-0011-2\\_1](https://doi.org/10.1007/978-981-19-0011-2_1)

On the other hand, the organization of a conference involves generally several people in different roles performing several tasks. A detailed description of these roles and related tasks can be found in [21, 23].

Tasks related to the management of a scientific conference can be categorized according to two main aspects: the technical or scientific aspect, and the administrative and logistical aspect. While tasks in both aspects are challenging and time and effort consuming, those related to the scientific aspect require in addition, a high level of expertise. These tasks are usually achieved by the chairmen or members of the Technical Program Committee (TPC). For instance, we can mention:

**Technical program committee composition:** the selection of the technical program committee members is highly related to the conference scope and topics. In this task, the PC chairmen must look for experts with relevant profiles to take part of the TPC.

**Plagiarism and out of scope papers detection:** while multiple plagiarism checkers can be used to detect plagiarism in submitted manuscripts, detecting out of scope papers is usually done manually. To do so, TPC members are requested to take a look to each submitted manuscript and decide whether it fits the scope of the conference or not. Basically, this requires a high expertise and time, especially with a large number of submissions, which makes this task more tedious.

**Assignment of papers to reviewers:** a good assignment allows an impartial assessment of submitted manuscripts which positively impacts the conference reputation. However, finding an appropriate, equitable and fair assignment for reviewers and papers is considered as a challenging and time-consuming task.

**Preparation of the conference program:** this task requires a high level of expertise and needs to be accomplished in two steps. Based on available resources (time and locations), the TPC chairmen must first organize accepted papers into sessions for oral and poster presentations. Each session should contain papers that share some common topics and must be labeled accordingly. Once this step is finished, the schedule is established where several sessions can be planned to be held in parallel. However, one should consider potential conflicts. For instance, sessions that involve presentations from the same author must not be scheduled in the same time slot. Furthermore, it is better to schedule sessions with related topics in different time slots to allow participants benefit from the event.

While the automation of paper-reviewer assignment task is integrated in many existing CMSs and has been the focus of extensive studies over decades [3, 7, 10, 12, 15, 19, 24, 26], the automation and integration of the other tasks remain less considered.

In this work, we are concerned with two tasks from those described above. Namely, the detection of out of scope manuscripts and the preparation of the conference program schedule. For this purpose, we propose to use efficient unsupervised methods to automate the aforementioned tasks.

The remainder of the paper is structured as follows. Section 2 provides an overview of related work. The automation of the aforementioned tasks is described in Sect. 3. In Sect. 4, the results of computational experiments are presented then the paper is concluded in Sect. 5.



## 2 Related Work

As mentioned above, CMSs become an necessary tool that aim to assist the organizers in performing many tasks and thus make the organization process easier. For instance, we can cite EasyChair [8], Conftool [6], Openconf [16], ConfMaster [5], Confious [18] and many others.

The main tasks that are usually handled by CMSs include email notification, online submission, collection and download of papers and abstracts, reviewers management, conflict of interest management, assignment of papers, collection of reviews, the preparation of the proceedings or even registration and payment of conference fees.

On the other hand, many existing works propose to enhance such systems by automating and integrating other tasks. In addition to a new assignment algorithm, authors in [20] present methods for conflict of interest detection, the poster setup plan and a conference participant support tasks and in [21], the same authors employed hierarchical clustering techniques to automatically propose a compilation of the papers into scientific sessions. Reinhardt et al. [23] propose to integrate common features of social media in CMSs to make scientific event management more social and awareness supporting. In [27], authors performed modifications on Openconf to include registration and payment acceptance.

## 3 Methods' Description

In the current work, we make a focus on two tasks related to the scientific aspect of the conference. Usually, these tasks are accomplished manually by the TPC chairmen or members and thus require time, efforts and a high expertise:

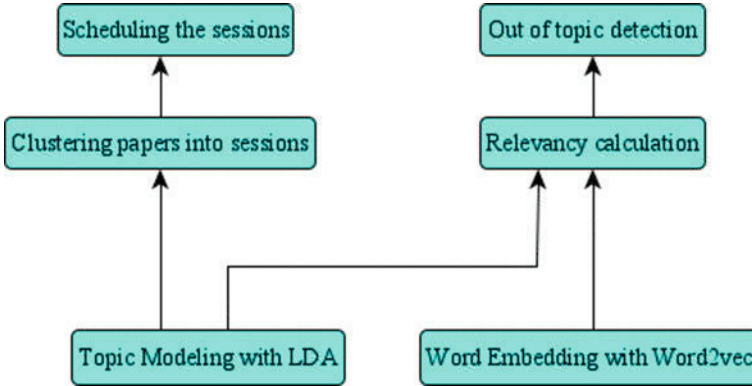
1. Detection of out of scope papers: in many conferences, TPC members or chairmen are requested to check out the submitted papers and make a decision about their fitness with regard to the topics of the conference.
2. Preparation of the conference program: which consists first on creating sessions for both oral talks and posters according to common themes then the schedule is established based on available resources.

To do so, several steps are performed as shown in Fig. 1.

Given a conference (which will be referred to as `CONF`), to perform any task for `CONF`, our framework first performs the following operations:

### 3.1 Topic Modeling

This involves modeling `CONF`'s topics by training a Latent Dirichlet Allocation (LDA) topic model, as described in [13, 14, 22], on a dataset of documents that are



**Fig. 1** A global description of the framework

relevant to `Conf`'s topics. In our case, we trained an LDA model on the document-term matrix of the pre-processed training set of the PeerRead dataset [11], which consists of over 14K papers submitted in top-tier venues including ACL, NIPS and ICLR.<sup>1</sup> For an arbitrary conference, the system could be modified to dynamically build a corpus specifically for the conference, given the conference's list of topics, which can be chosen from a predefined and updatable list of topics.

The LDA model infers a set of  $q$  ( $q$  is given) topics  $T = \{t_k \mid k \in [q]\}$  from its training corpus, and for each submitted paper in  $P = \{P_i \mid i \in [n]\}$ , it predicts  $P_i$ 's topics probability distribution vector, denoted  $z_i$ , given `Conf`'s topics  $T$ .

### 3.2 Word Embedding

In addition to the topic model, the semantic information of words is obtained by carrying out training of a Word2vec word embedding model on the combined corpus of `Conf`'s LDA model corpus (in our case PeerRead) and `Conf`'s pre-processed submitted papers corpus.

The Word2vec model outputs, for each word in the model's vocabulary, its learned word vector, that will be used in calculating the semantic similarity between words using a suitable similarity measure (such as cosine similarity).

Once the topic and word models have been built, the framework then proceeds to perform the main tasks as follows:

<sup>1</sup> <https://github.com/allenai/PeerRead>.

### 3.3 Detection of Out of Scope Papers Task

As mentioned previously, this is a tedious task that is usually accomplished by the TPC chairmen or members, manually. In this paper, we propose an algorithm which adapts the method presented in [22] for this task.

Originally, authors in [22] used an LDA model to extract topics and feature words for each topic, and a Word2vec model to express the semantic information of words in documents. They trained both models solely on the subject documents, and used both of LDA's probabilities and Word2vec's vectors to calculate what they referred to as the "relevancy" of a document.

Instead of training the LDA and Word2vec models on the subject documents only (in our case the submitted papers), we propose to train the LDA model using the corpus described in Sect. 3.1, since its documents are guaranteed to be in-topic.

On the other hand, we train the Word2vec model on the corpus described in Sect. 3.2. By doing so, the Word2vec model will grasp the submitted papers' words' semantics from their context in the submitted papers and allows it to integrate well with the LDA model by modeling its vocabulary as well.

The topic and word models together will prove useful in determining the relevancy of a submitted paper.

**Submitted Papers' Relevancy Calculation** First, for each word  $w_j$  of every submitted paper, vectors extracted by Word2vec are used to calculate the cosine similarity between  $w_j$  and the feature word  $w_f$  under topic  $t_k$ .

The relevancy between the word  $w_j$  and the topic  $t_k$ , denoted  $\text{Rel}_{w,t}(w_j, t_k)$ , is then defined as the probability weighted sum of the cosine similarities between  $w_j$  and a number  $N_f$  of feature words under topic  $t_k$ :

$$\text{Rel}_{w,t}(w_j, t_k) = \sum_{f=1}^{N_f} P(w_f | t_k) \cos(w_j, w_f) \quad (1)$$

where  $\cos$  denotes the cosine similarity between the Word2vec feature vectors of words  $w_j$  and  $w_f$ .

Then, the relevancy between the word  $w_j$  and the submitted paper  $P_i$ , denoted as  $\text{Rel}_{w,p}(w_j, p_i)$ , is defined as the probability weighted sum of the relevancy between  $w_j$  and all topics of  $P_i$ :

$$\text{Rel}_{w,p}(w_j, p_i) = \sum_{k=1}^q P(t_k | p_i) \text{Rel}_{w,t}(w_j, t_k) \quad (2)$$

After that,  $\text{Rel}_{w,p}(w_j, p_i)$  of all words  $w_j$  in submitted paper  $P_i$  are accumulated and divided by the number of words  $J_i$  in paper  $P_i$  to obtain the total relevancy of the submitted paper  $P_i$ , denoted  $\text{Rel}_p(p_i)$ , where

$$\text{Rel}_p(p_i) = \frac{1}{J_i} \sum_{j=1}^{J_i} \text{Rel}_{w,p}(w_j, p_i) \quad (3)$$

Finally, off-topic submitted papers are filtrated according to a predetermined threshold for  $\text{Rel}_p$ .

We also note that in their original paper, the authors in [22] defined  $\text{Rel}_p(p_i)$  without the factor  $\frac{1}{J_i}$ . The reason we added that factor was to normalize a document's relevancy with respect to its size.

### 3.4 Sessions Creation and Conference Program Schedule

As mentioned in the beginning of this section, the preparation of the conference program goes through two steps, first, the generation of sessions for oral and poster presentations then the scheduling of these sessions in a meaningful way. Considering available time slots and locations, we must group, in each session, a predefined number of papers that share a common theme. Thus, sessions can be generated by performing a clustering under size constraints on clusters. In this paper, the generation of sessions is accomplished by a new algorithm based on Bregman Hard Clustering [2]. A cluster labeling method is then proposed.

This new algorithm uses the BregMeans++ with Local Search method [25] as an initializer, and is modified to take into account clusters' size constraints, since each session's (i.e., cluster's) size is specified by the PC Chairmen beforehand.

Once the sessions are generated, establishing the conference program can be formulated as a coloring problem, where any of the many existing algorithms may be used to solve it. The description of our approach is detailed as follows:

**Clustering the Accepted Papers into Sessions Task** To cluster the accepted papers into sessions, we propose an algorithm based on Bregman Hard Clustering (BHC), introduced by authors in [2].

Since the feature vector used to represent a paper  $P_i$  is its topics probability distribution vector  $z_i$ , we make use of a suitable divergence measure for these features. In our case, we choose the Jensen–Shannon divergence, denoted JSD.

The proposed clustering algorithm consists of three steps:

- S1. Initialization Step: For this step, we use the method proposed in [25], BregMeans++ with Local Search, which is an improvement of BregMeans++ [1]. BregMeans++ with Local Search combines BregMeans++ with the local search strategy, whose main purpose is to further improve the initial centroids assignment by an iterative process based on a single swap operation.
- S2. Assignment Step: Standard BHC doesn't take into account prior information about cluster sizes, and given the PC Chairmen would want to specify exactly how many papers are fit into each session, standard clustering doesn't suffice. For this reason, we propose to incorporate the constraint defined in [9] into BHC.

We denote with  $\zeta_s$  the number of papers in session  $s$ , and with  $N_s$  the number of sessions. The algorithm then assigns the point (paper)  $z$  to the nearest available cluster (session)  $C_s$  such that:

$$s = \operatorname{argmin}_{s' \in \{s'' \in [N_s] : |C_{s''}| < \zeta_{s''}\}} \operatorname{JSD}(z, \mu_{s'}) \quad (4)$$

S3. Update Step: For each cluster  $C_s$ , its new centroid  $\mu_s$  is computed as:

$$\mu_s = \frac{\sum_{z \in C_s} z}{|C_s|} \quad (5)$$

The termination or convergence criterion for this algorithm, depends on the centroids convergence. Other approaches use a maximum number of iterations  $\operatorname{Iter}_{\max}$  as a termination criterion. For our work, we combine the two criteria into the following termination criterion:

- If for all  $s \in [N_s]$ ,  $\operatorname{JSD}\left(\mu_s, \frac{\sum_{z \in C_s} z}{|C_s|}\right) \leq \xi$ , for some convergence threshold  $\xi \geq 0$ , stop the main (assignment and update) loop. After that, run another assignment and update loop for an additional number  $\operatorname{Iter}_+$  of iterations. We use this because the threshold criterion might terminate the algorithm prematurely, when it could improve further, if the choice of the value of  $\xi$  is poorly made (similarly for  $\operatorname{Iter}_{\max}$ ). So to make the best of both criteria, we use  $\xi$  to indicate that an initial minimum is attained, and use  $\operatorname{Iter}_+$  to allow the algorithm to try and improve that minimum further.

**Labeling the Clusters** The labeling method we propose labels a cluster  $C_s$  as follows:

1. First, for a point  $z_i$  in  $C_s$ , let  $T_i^{\max}$  denote the topic with maximum probability in paper  $P_i$ , i.e.,

$$T_i^{\max} = \operatorname{argmax}_T P(T | p_i) \quad (6)$$

2. then, let  $W_i^{\max}$  be the set of feature words that are most prevalent in  $T_i^{\max}$  and having  $|W_i^{\max}| = N_W$  for a given number  $N_W$ , i.e.,

$$W_i^{\max} = \operatorname{argmax}_{W \subseteq T, |W|=N_W} \sum_{w \in W} P(w | T_i^{\max}) \quad (7)$$

3. finally, collect the  $W_i^{\max}$  of each  $z_i \in C_s$  in a bag of words  $\mathcal{B}_s$ , and label  $C_s$  with  $\mathcal{B}_s$ , i.e.,

$$\mathcal{B}_s = \{(w, f_s(w)) \mid w \in \bigcup_{z_i \in C_s} W_i^{\max}\} \quad (8)$$

where  $f_s(w)$  denotes the frequency of the word  $w$  and is defined as:

$$f_s(w) = \sum_{z_i \in C_s} \mathbf{1}_{W_i^{\max}}(w) \quad (9)$$

with  $\mathbf{1}_{W_i^{\max}}$  denoting the indicator function of the subset  $W_i^{\max}$  of the set of all the words in the LDA training corpus, defined as:

$$\mathbf{1}_{W_i^{\max}}(w) = \begin{cases} 1, & \text{if } w \in W_i^{\max}, \\ 0, & \text{if } w \notin W_i^{\max}. \end{cases} \quad (10)$$

**Scheduling the Sessions** Establishing the conference program is the final task of TPC chairmen before the conference.

Once the sessions are created, we can schedule them according to the available resources. Usually, several sessions can be scheduled in parallel.

However, we must consider eventual conflicts in this process. For instance, sessions that involve presentations from the same author must not be scheduled in the same time slot. Also, to allow participants gaining the most benefit from the event, it is preferable to schedule sessions on similar topics in different time slots.

Under the aforementioned conditions, the problem can be formulated as a typical coloring problem where sessions are vertices and two vertices are joined by an edge if either there is a conflict between the sessions or their corresponding topics are similar. The problem can then be solved using any coloring technique.

Formally, the problem of scheduling the sessions can be formulated, for a graph  $G_C = (V, E)$  where  $V = \{C_s \mid s \in [N_s]\}$  and  $E$  is the set of edges:

- $\{C_s, C_k\}$  if  $\text{JSD}(\mu_s, \mu_k) < \varepsilon$ , for  $\varepsilon \geq 0$ .
- $\{C_s, C_k\}$  if  $(\bigcup_{z_i \in C_s} \text{Authors}(P_i)) \cap (\bigcup_{z_j \in C_k} \text{Authors}(P_j)) \neq \emptyset$ .

where  $\text{Authors}(P_i)$  is the set of authors of paper  $P_i$ , and  $\text{JSD}(\mu_s, \mu_k)$  is the distance between clusters  $C_s$  and  $C_k$ , defined as the Jensen–Shannon divergence between their centroids.  $\varepsilon$  is the threshold such that if  $\text{JSD}(\mu_s, \mu_k) < \varepsilon$  then the sessions  $C_s$  and  $C_k$  are considered to be in the same topic. The problem then is to find a proper coloring of  $G_C$  where each color class represents the sessions that can be executed in parallel.

## 4 Experimental Results

To evaluate our approach, we generated a dataset  $D$  of 120 papers that were uniformly sampled from the entire test set of the PeerRead dataset, whose dominant topics include Natural Language Processing, Artificial Intelligence, Machine Learning, and Computation. Another dataset  $D'$  was also created which contains the papers in  $D$  including an additional 10 papers that are off-topic, and that were randomly selected from the test set of the arXiv dataset of scientific (mainly Physics and Chemistry related) papers introduced in [4].

## 4.1 Out of Scope Detection

We used as an evaluation method the Receiver Operating Characteristic (ROC) Curve, and the Area Under the Curve (AUC) to analyze the performance of the in-/off-topic classification as the threshold for  $Rel_p$  is varied (Fig. 2). The curve is created by plotting the true positive rate against the false-positive rate at various threshold settings.

However, ROC curves can present an overly optimistic view of an algorithm's performance if there is a large skew in the class distribution, which is the case with  $D'$  (the number of in-topic documents is much larger than that of off-topic documents), and most likely is the case with actual conferences. For this reason, we also used the precision–recall (PR) Curve (Fig. 3), which is usually used when there is a moderate to large class imbalance.

The high AUC values in Figs. 2 and 3 indicate a very good performance of the proposed out of scope detection algorithm.

## 4.2 Clustering Papers into Sessions

Since the number of sessions and their size are determined by the PC Chairmen, rendering the clustering algorithm's parameters to be unimprovable, internal validation methods become unreliable to evaluate the quality of the clustering. We will however, calculate some internal validation indices. But the main evaluation we will

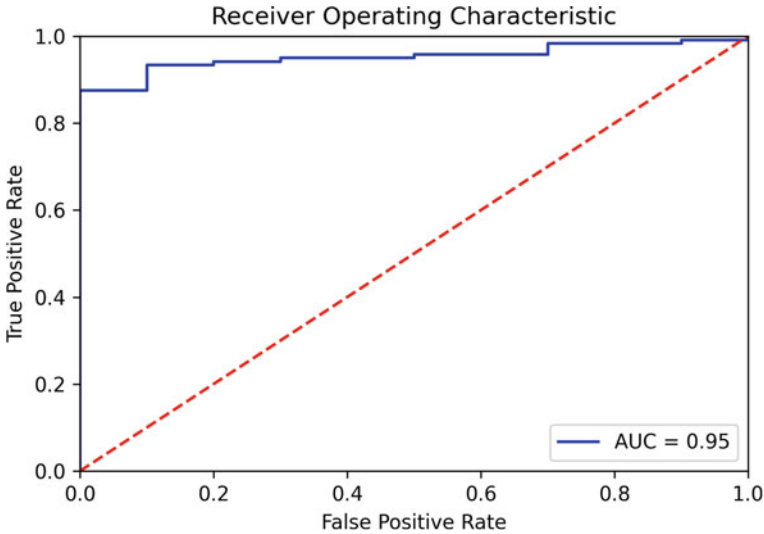
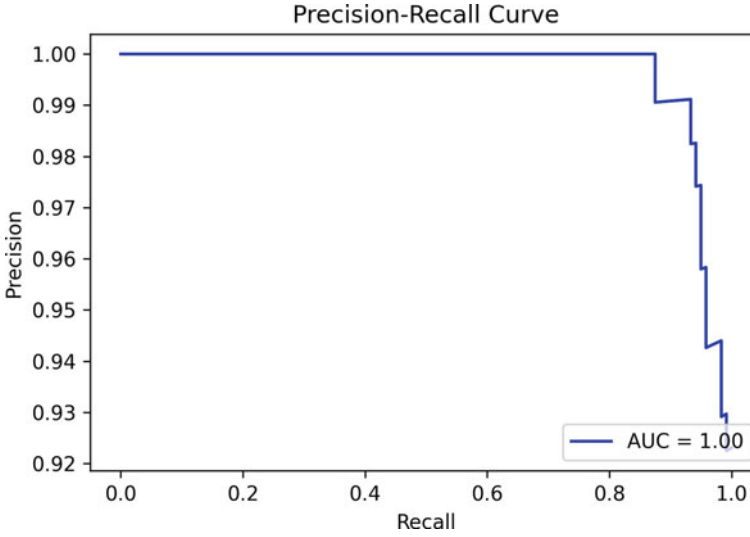


Fig. 2 ROC curve and AUC of  $Rel_p$  for  $D'$



**Fig. 3** PR curve and AUC of  $\text{Rel}_p$  for  $D'$

**Table 1** Internal evaluation indices for the clustering of  $D$

Index	Silhouette index	Calinski-Harabasz index	Davies–Bouldin index
Value	0.1003	8.3011	1.9429

focus on here, is to label the clusters using the method described in Sect. 3.4, and then based on the labeling, an expert may assess the clustering.

For the clustering of  $D$ , we used a number of sessions  $N_s = 15$ , and, the cluster size parameter  $\zeta_s$  took the value of 8 for all sessions  $s \in [N_s]$ .

The internal validation indices we used to evaluate the clustering of  $D$  are the Silhouette index, the Calinski-Harabasz index, and the Davies–Bouldin index, whose values are shown in Table 1:

- The Silhouette index (SI) is defined in the interval  $[-1, 1]$  for each point. An SI closer to 1 indicates a high separation between clusters. An SI closer to  $-1$  is an indication of overlapping clusters [17]. An SI closer to 0 is an indication that the data are uniformly distributed.
- The Calinski-Harabasz index (CHI) is the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters. The higher the CHI, the better the performance.
- The Davies–Bouldin index (DBI) is defined as the average similarity between each cluster and its closest cluster. A lower DBI relates to a model with better separation between the clusters.



The Silhouette index value of 0.1003 suggests that the data might be uniformly distributed, which further suggests a uniformity of the topic distribution vectors of the papers. That is probably the case with dataset  $D$ , given it was uniformly selected from a dataset with a small number of dominant general topics. The high value of the Calinski-Harabasz and the low value of the Davies–Bouldin indices however indicate a good performance of the clustering.

Table 2 shows the results of labeling the clustering of  $D$ .  $N_W = 5$  for the labeling, and for display reasons,  $\mathcal{B}_s$  was cut to include the 5 words with maximum count in  $\mathcal{B}_s$ . Column 1 of Table 2 represents the cluster to be labeled. Column pairs (2, 3), (4, 5), ...etc., represent the pairs  $(w_j \in \mathcal{B}_s, f_s(w_j))$ .

The labels of the labeling method described in Sect. 3.4 could be summarized (interpreted) using one term labels by an expert. For instance, for the labels of the clusters in Table 2 (ignoring the words that could not fit here), one can say that cluster  $C_1$  is about “reinforcement learning”, cluster  $C_6$  is about “object detection”, cluster  $C_{14}$  is about “question answering” and so on. We can notice from Table 2 the cluster labels’ meaningfulness and significance, which also shows the algorithm’s good performance. However, the true performance of the algorithm can only be measured if its parameters could change freely, which isn’t the case with conferences having many time and resource constraints. This necessitates a method to plan a program schedule under the time and resource constraints, while maximizing the clustering performance.

**Table 2** Excerpts of the labels of clusters of  $D$

Cluster	$w_1$	$f_s(w_1)$	$w_2$	$f_s(w_2)$	$w_3$	$f_s(w_3)$	$w_4$	$f_s(w_4)$	$w_5$	$f_s(w_5)$
$C_1$	action	7	polic	7	learn	7	reward	7	agent	7
$C_2$	model	8	word	8	embed	8	vector	8	represent	8
$C_3$	algorithm	5	bound	3	lemma	3	theorem	3	proof	3
$C_4$	set	4	definit	4	defin	4	proposit	4	proof	4
$C_5$	model	8	languag	7	translat	7	word	7	sentenc	7
$C_6$	imag	4	featur	4	segment	4	object	4	recognit	4
$C_7$	vector	4	data	4	matrix	3	kernel	3	method	3
$C_8$	network	6	node	4	graph	4	tree	4	edg	4
$C_9$	process	4	data	3	time	3	pattern	3	method	3
$C_{10}$	document	4	topic	4	text	4	sentiment	4	term	4
$C_{11}$	model	4	algorithm	3	distribut	3	sampl	3	probabl	3
$C_{12}$	network	8	train	8	layer	8	model	8	neural	8
$C_{13}$	learn	4	action	3	polic	3	reward	3	agent	3
$C_{14}$	model	7	question	3	answer	3	task	3	attent	3
$C_{15}$	entiti	5	annot	5	tag	5	relat	5	extract	5

## 5 Conclusion

In this paper, we discussed the problem of the automatic management of the scientific aspect of an academic conference. We focused in this work on the study of two tasks that are seldom studied in literature and neglected in most conference management systems, namely, the detection of out of scope submitted papers, and the preparation of the conference program.

We used LDA and Word2vec to model the submitted papers topics and words' semantics. The probabilities obtained from LDA along with the vectors obtained from Word2vec were used to calculate a submitted paper's relevancy score, then out of scope papers were identified by specifying a threshold for this score.

We proposed a new algorithm to cluster the papers into sessions with cluster size constraints that is based on the Bregman Hard Clustering algorithm. We also proposed a scheme to label the resulting clusters using the LDA model. As for the problem of scheduling the sessions into a technical program, we formulated it as a graph coloring problem where many algorithms exist to solve it.

An experimental study on data randomly sampled from the PeerRead and arXiv datasets was conducted to evaluate the proposed algorithms. For the evaluation of the results of the out of scope detection algorithm, the AUC of the ROC and PR curves indicated a good performance of the algorithm. As for the evaluation of the clustering algorithm, we used some internal validation metrics, whose values showed good performance of the proposed algorithm. But given the parameters of the algorithm are unimprovable, we relied mainly on experts' assessment of the cluster labels, which showed the coherence of the clusters.

Despite the good performance of the proposed method, it mainly relies on the use of LDA and Word2vec. It would be interesting to make use of another topic modeling and word embedding tools seeking for more accurate results.

## References

1. Ackermann, M.R., Blömer, J.: Coresets and approximate clustering for bregman divergences. In: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1088–1097. SODA '09, Society for Industrial and Applied Mathematics, USA (2009)
2. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J.: Clustering with bregman divergences. *J. Mach. Learn. Res.* 6(58), 1705–1749 (2005)
3. Charlin, L., Zemel, R.: The toronto paper matching system: an automated paper-reviewer assignment system. In: 30 th International Conference on Machine Learning, vol. 28. JMLR (2013)
4. Cohan, A., Deroncourt, F., Kim, D.S., Bui, T., Kim, S., Chang, W., Goharian, N.: A discourse-aware attention model for abstractive summarization of long documents (2018)
5. ConfMaster. <https://www.confmaster.net>
6. Conftool. <https://www.conftool.net>
7. Conry, D., Koren, Y., Ramakrishnan, N.: Recommender systems for the conference paper assignment problem. In: Proceedings of the third ACM Conference on Recommender Systems, pp. 357–360 (2009)