16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024)

Volume 1: KDIR

Porto, Portugal
17-19 November 2024

Editors:

Frans Coenen
Ana Fred
Jorge Bernardino

ISBN: 979-8-3313-1098-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2024) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda. at the address below.

SCITEPRESS – Science and Technology Publications, Lda.
Avenida de S. Francisco Xavier, Lote 7 Cv. C,
2900-616 Setúbal, Portugal

Phone: +351 265 520 185
Fax: +351 265520 186

info@scitepress.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

Multimodal Deep Learning in Medical Imaging 5
Carlo Sansone

Intelligent Reuse of Explanation Experiences: The Role of Case-Based Reasoning in Promoting Best Practice in Explainable AI 7
Nirmalie Wiratunga

Recent Advances in Learning from Data Streams 9
João Gama

PAPERS

FULL PAPERS

Learning to Rank for Query Auto-Complete with Language Modelling in Enterprise Search 15
Colin Daly and Lucy Hederman

Machine Learning Unravels Sex-Specific Biomarkers for Atopic Dermatitis 27
Ana Duarte and Orlando Belo

MLN-Subdue: Substructure Discovery In Homogeneous Multilayer Networks 36
Anish Rai, Anamitra Roy, Abhishek Santra and Sharma Chakravarthy

A Knowledge Map Mining-Based Personalized Learning Path Recommendation Solution for English Learning 48
Duong Thien Nguyen and Thu Minh Tran Nguyen

Positive-Unlabeled Learning Using Pairwise Similarity and Parametric Minimum Cuts 60
Torpong Nitayanont and Dorit S. Hochbaum

Efficient Visualization of Association Rule Mining Using the Trie of Rules 72
Mikhail Kudriavtsev, Andrew McCarren, Hyowon Lee and Marija Bezbradica

Predicting Post Myocardial Infarction Complication: A Study Using Dual-Modality and Imbalanced Flow Cytometry Data 81
Nada ALdausari, Frans Coenen, Anh Nguyen and Eduard Shantsila

GenCrawl: A Generative Multimedia Focused Crawler for Web Pages Classification 91
Domenico Benfenati, Antonio Maria Rinaldi, Cristiano Russo and Cristian Tommasino

Antibiotic Resistance Gene Identification from Metagenomic Data Using Ensemble of Finetuned Large Language Models 102
Syama K. and J. Angel Arul Jothi

Approaches for Extending Recommendation Models for Food Choices in Meals 113
Nguyen Thi Hong Nhung, Dao Khoa Nguyen, Tiet Gia Hong and Thi My Hang Vu

Personalization of Dataset Retrieval Results Using a Data Valuation Method 122
Malick Ebiele, Malika Bendechache, Eamonn Clinton and Rob Brennan

A Systematic Literature Review on LLM-Based Information Retrieval: The Issue of Contents Classification 135
Diogo Cosme, António Galvão and Fernando Brito E Abreu

Enhancing Answer Attribution for Faithful Text Generation with Large Language Models 147
Juraj Vladika, Luca Mülln and Florian Matthes

MERGE App: A Prototype Software for Multi-User Emotion-Aware Music Management 159
Pedro Lima Louro, Guilherme Branco, Hugo Redinho, Ricardo Correia, Ricardo Malheiro, Renato Panda and Rui Pedro Paiva

Contrato360 2.0: A Document and Database-Driven Question-Answer System Using Large Language Models and Agents 167
Antony Seabra, Claudio Cavalcante, João Nepomuceno, Lucas Lago, Nicolaas Ruberg and Sergio Lifschitz

Reviewing Machine Learning Techniques in Credit Card Fraud Detection 179
Ibtissam Medarhri, Mohamed Hosni, Mohamed Ettalhaoui, Zakaria Belhaj and Rabie Zine

SHORT PAPERS

An Index Bucketing Framework to Support Data Manipulation and Extraction of Nested Data Structures 191
Jeffrey Myers II and Yaser Mowafi

Utilizing Data Analysis for Optimized Determination of the Current Operational State of Heating Systems 200
Ahmed Qarqour, Sahil-Jai Arora, Gernot Heisenberg, Markus Rabe and Tobias Kleinert

A Core Technology Discovery Method Based on Hypernetwork 210
Chen Wenjie

Modelling of an Untrustworthiness of Fraudulent Websites Using Machine Learning Algorithms 218
Kristína Machová and Martin Kaňuch

Comparative Analysis of Topic Modelling Approaches on Student Feedback 226
Faiz Hayat, Safwan Shatnawi and Ella Haig

Reducing the Transformer Architecture to a Minimum 234
Bernhard Bermeitinger, Tomas Hrycej, Massimo Pavone, Julianus Kath and Siegfried Handschuh

Efficient Neural Network Training via Subset Pretraining 242
Jan Spörer, Bernhard Bermeitinger, Tomas Hrycej, Niklas Limacher and Siegfried Handschuh

Multi-Label Classification for Fashion Data: Zero-Shot Classifiers via Few-Shot Learning on Large Language Models 250
Dongming Jiang, Abhishek Shah, Stanley Yeung, Jessica Zhu, Karan Singh and George Goldenberg

Optimizing High-Dimensional Text Embeddings in Emotion Identification: A Sliding Window Approach 258
Hande Aka Uymaz and Senem Kumova Metin

Text-Based Feature-Free Automatic Algorithm Selection
Amanda Salinas-Pinto, Bryan Alvarado-Ulloa, Dorit Hochbaum, Matías Francia-Carramiñana, Ricardo Ñanculef and Roberto Asín-Achá 267

Flow Is Best, Fast and Scalable: The Incremental Parametric Cut for Maximum Density and Other Ratio Subgraph Problems
Dorit S. Hochbaum 275

Federated Learning for XSS Detection: A Privacy-Preserving Approach
Mahran Jazi and Irad Ben-Gal 283

Integrated Evaluation of Semantic Representation Learning, BERT, and Generative AI for Disease Name Estimation Based on Chief Complaints
Ikuo Keshi, Ryota Daimon, Yutaka Takaoka and Atsushi Hayashi 294

RUDEUS: A Machine Learning Classification System to Study DNA-Binding Proteins
David Medina-Ortiz, Gabriel Cabas-Mora, Iván Moya, Nicole Soto-García and Roberto Uribe-Paredes 302

Knowledge Discovery in Optical Music Recognition: Enhancing Information Retrieval with Instance Segmentation
Elona Shatri and György Fazekas 311

Evaluating the Suitability of Long Document Embeddings for Classification Tasks: A Comparative Analysis
Bardia Rafieian and Pere-Pau Vázquez 320

Prompt Distillation for Emotion Analysis
Andrew L. Mackey, Susan Gauch and Israel Cuevas 328

Comparing Human and Machine Generated Text for Sentiment
WingYin Ha and Diarmuid P. O'Donoghue 335

Enhancing LLMs with Knowledge Graphs for Academic Literature Retrieval
Catarina Pires, Pedro Gonçalo Correia, Pedro Silva and Liliana Ferreira 343

An Explainable Classifier Using Diffusion Dynamics for Misinformation Detection on Twitter
Arghya Kundu and Uyen Trang Nguyen 348

A Network Learning Method for Functional Disability Prediction from Health Data
Riccardo Dondi and Mehdi Hosseinzadeh 358

Enhancing Dyeing Processes with Machine Learning: Strategies for Reducing Textile Non-Conformities
Mariana Carvalho, Ana Borges, Alexandra Gavina, Lídia Duarte, Joana Leite, Maria João Polidoro, Sandra Aleixo and Sónia Dias 363

Comparative Analysis of Real-Time Time Series Representation Across RNNs, Deep Learning Frameworks, and Early Stopping
Ming-Chang Lee, Jia-Chun Lin and Sokratis Katsikas 371

Multilayer Networks: For Modeling and Analysis of Big Data
Abhishek Santra, Hafsa Billah and Sharma Chakravarthy 383

Route Recommendation Based on POIs and Public Transportation
Ágata Palma, Pedro Morais and Ana Alves 391

Hyperparameter Optimization for Search Relevance in E-Commerce 399
Manuel Dalcastagné and Giuseppe Di Fabbrizio

A Framework for Self-Service Business Intelligence 408
Rosa Matias and Maria Beatriz Piedade

Knowledge Graphs Can Play Together: Addressing Knowledge Graph Alignment from Ontologies in the Biomedical Domain 416
Hanna Abi Akl, Dominique Mariko, Yann-Alan Pilatte, Stéphane Durfort, Nesrine Yahiaoui and Anubhav Gupta

Decoding AI's Evolution Using Big Data: A Methodological Approach 424
Sophie Gvasalia, Mauro Pelucchi, Simone Perego and Rita Porcelli

MAEVE: An Agnostic Dataset Generator Framework for Predicting Customer Behavior in Digital Marketing 433
William Ferreira da Silva Filho, Seyed Jamal Haddadi and Julio Cesar dos Reis

Beyond Twitter: Exploring Alternative API Sources for Social Media Analytics 441
Alina Campan and Noah Holtke

Optimizing Federated Learning for Intrusion Detection in IoT Networks 448
Abderahmane Hamdouchi and Ali Idri

An Improved Meta-Knowledge Prompt Engineering Approach for Generating Research Questions in Scientific Literature 457
Meng Wang, Zhixiong Zhang, Hanyu Li and Guangyin Zhang

A Smart Hybrid Enhanced Recommendation and Personalization Algorithm Using Machine Learning 465
Aswin Kumar Nalluri and Yan Zhang

An End-to-End Generative System for Smart Travel Assistant 473
Miraç Tuğcu, Begüm Çıtamak Erdinç, Tolga Çekiç, Seher Can Akay, Derya Uysal, Onur Deniz and Erkut Erdem

Comparative Performance Analysis of Active Learning Strategies for the Entity Recognition Task 480
Philipp Kohl, Yoka Krämer, Claudia Fohry and Bodo Kraft

Intrinsic Evaluation of RAG Systems for Deep-Logic Questions 489
Junyi Hu, You Zhou and Jie Wang

Prediction of Response to Intra-Articular Injections of Hyaluronic Acid for Knee Osteoarthritis 497
Eva K. Lee, Fan Yuan, Barton J. Mann and Marlene DeMaio

Comparative Analysis of Single and Ensemble Support Vector Regression Methods for Software Development Effort Estimation 509
Mohamed Hosni

Software Testing Effort Estimation Based on Machine Learning Techniques: Single and Ensemble Methods 517
Mohamed Hosni, Ibtissam Medarhri and Juan Manuel Carrillo de Gea

Insights into the Potential of Fuzzy Systems for Medical AI Interpretability 525
Hafsaa Ouifak and Ali Idri

AUTHOR INDEX 533