



Venue: Novotel Sydney Parramatta, Australia

Call for Participation

We invite you to join us in **6th International Conference on Data Science and Applications (DSA 2025)**.

This conference will act as a major forum for the presentation of innovative ideas, approaches, developments, and research projects in the areas of Data Science and Applications.

Highlights of DSA 2025 Includes:

- Big Data Algorithms
- Big Data Applications
- Big Data Fundamentals
- Big Data Management and Frameworks
- Big Data Search
- Big Data security
- Bioinformatics
- Bio-Metrics
- Consistent Data Model
- Data and Knowledge Representation
- Data Streams Mining, Graph Mining, Spatial Data Mining
- Databases
- Deep Learning
- Exploring Data Analysis
- Financial Modeling
- Forecasting, Classification, Clustering
- Image Analysis
- Inference of Causes, Prediction, Evaluating, Consolidating and Explaining Discovered Knowledge
- Infrastructures for Big Data
- Integrating Constraints and Knowledge in the KDD Process
- Integration of Data Warehousing
- Interactive Data Exploration / Visualization and Discovery
- Knowledge Discovery Framework and Process, Including Pre- and Post-Processing

- Languages and Interfaces for Data Mining
- Learning in knowledge-intensive systems
- Learning Problems
- Machine Learning Applications
- Mining from Low-Quality Information Sources
- Mining Trends, Opportunities and Risks
- OLAP and Data Mining
- Parallel and Distributed Data Mining Algorithms
- Pre-Processing Techniques, Visualization
- Security and Information Hiding in Data Mining
- Social Networks, Educational Data Mining
- Statistical Techniques for Generation a Robust
- Text video, Multimedia Data Mining, Web Mining

Registration Participants

Non-Author/Co-Author/Simple Participants (no paper)

100 USD for Online (Without Proceedings)

490 USD for Face to Face (With Proceedings)

Here is where you can reach us: dsa@acity2025.org or dsaconf@yahoo.com

A Theoretical Framework for Edge-native URLLC Traffic Classification: Bridging AI Models and Real-world Constraints

Osama AlQahtani, College of Engineering and Computer Science, University of Jazan, Jazan, Saudi Arabia

Abstract

Ultra-Reliable Low Latency Communications (URLLC) applications in 5G and beyond networks demand unprecedented performance levels with sub-millisecond latencies and 99.999% reliability. While AI-based traffic classification has emerged as a critical enabler for intelligent network management, existing approaches focus primarily on algorithmic improvements without adequately addressing the practical constraints of edge deployment environments. This paper presents a novel theoretical framework for edge-native URLLC traffic classification that systematically bridges the gap between AI model capabilities and real-world deployment limitations. The framework comprises four interconnected components: resource constraint modeling, latency decomposition analysis, reliability-performance trade-off optimization, and edge-cloud orchestration principles. A systematic four-phase methodology guides

practitioners through system characterization, model selection and optimization, deployment strategy determination, and performance validation. Theoretical case study analysis across three representative scenarios—5G base station edge computing, industrial IoT gateways, and vehicular edge computing nodes—demonstrates the framework’s effectiveness in diverse deployment environments. Framework projections indicate potential achievement of URLLC targets with 0.7ms average latency and 99.997% reliability for high-resource scenarios, and 400µs latency with 99.9993% reliability for constrained industrial applications. The theoretical analysis shows consistent resource optimization potential of 45-78% while maintaining acceptable classification accuracy. This work provides the foundation for systematic deployment of AI traffic classification systems in edge environments, offering both theoretical rigor and practical guidance for next-generation URLLC applications.

Network Security: Safeguarding the Digital Infrastructure

Nikitha Merilena Jonnada, University of the Cumberland, USA

Abstract

In this paper, the author discusses about the significance of network security as it continues to grow and how digital infrastructure is becoming an integral part of our daily life. Modern networks are facing more threats like the malware, ransomware, phishing, advanced persistent threats (APTs), and vulnerabilities in the cloud and Internet of Things (IoT) environments. These threats compromise data confidentiality, integrity, and availability, posing risks to individuals, businesses, and the government. This paper presents a comprehensive review of contemporary network security principles, technologies, and practices, while proposing an Integrated Network Defense Framework (INDF) that combines technical, administrative, and policy-driven measures for holistic protection. The paper examines the foundational CIA triad (confidentiality, integrity, availability), common attack vectors, and modern defense mechanisms, such as firewalls, intrusion detection and prevention systems, endpoint detection and response, and encryption methods.

Keywords

Network Security, Cybersecurity, Zero Trust, Artificial Intelligence, Threat Mitigation.

A Web-based Bone Marrow Disease Detection System using Convolutional Neural Networks

Bornoma Halima, Farouk Hafsa Mu’azu, Okonta Ehijesumuan, Diadie Sow, Ignace Djitog and Ekpe Okorafor, Nigerian British University, Nigeria

Abstract

Bone marrow diseases are illnesses that affect the bone marrow of an individual. The bone marrow is a mesh-like organ situated inside the bones of a human being entirely in charge of producing blood and all blood components i.e. lymphocytes, erythrocytes, platelets and plasma. Any disease affecting the bone marrow affects the production of blood which can lead to loss of blood or cancers which eventually lead to death. This paper proposes a new web-based application integrated with convolutional neural networks algorithm, a machine learning

approach, to automate an early diagnosis of bone marrow diseases without much hassle contrary to common manual processing which is exceedingly labor-intensive and costly. While the dataset was thoroughly examined, features that fit in with patient characteristics living with sickle cell disease in Nigeria were extracted to carry out the analysis. The dataset contains 11 classes of different bone marrow disease cell types, and a number of performance measures such as area under the curve (AUC), precision, and recall were generated and analyzed with the following results 98.38%, 87.12%, and 77.12% respectively. In terms of diagnosing sickle cell diseases with patients in Nigeria, the proposed model surpassed all existing learning models. The resulting model was saved in a specific file format then successfully imported into the developed web portal for instant analysis and wider access by authorized personnel.

Keywords

Machine Learning, Convolutional Neural Networks (CNN), Bone-Marrow Disease, Hematology, Smear, Prediction, Web Application.

A Multi-Agent Retrieval-Augmented Framework for Work-in-Progress Prediction

Yousef Mehrdad Bibalan¹, Behrouz Far¹, Mohammad Moshirpour², and Bahareh Ghiyasian³, ¹ University of Calgary, Canada, ² University of California, Irvine, USA

Abstract

Work-in-Progress (WiP) prediction is critical for predictive process monitoring, enabling accurate anticipation of workload fluctuations and optimized operational planning. This paper proposes a retrieval-augmented, multi-agent framework that combines retrieval-augmented generation (RAG) and collaborative multi-agent reasoning for WiP prediction. The narrative generation component transforms structured event logs into semantically rich natural language stories, which are embedded into a semantic vector-based process memory to facilitate dynamic retrieval of historical context during inference. The framework includes predictor agents that independently leverage retrieved historical contexts and a decision-making assistant agent that extracts high-level descriptive signals from recent events. A fusion agent then synthesizes predictions using ReAct-style reasoning over agent outputs and retrieved narratives. We evaluate our framework on two real-world benchmark datasets. Results show that the proposed retrieval-augmented multi-agent approach achieves competitive prediction accuracy, obtaining a Mean Absolute Percentage Error (MAPE) of 1.50% on one dataset, and surpassing Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM), and persistence baselines. The results highlight improved robustness, demonstrating the effectiveness of integrating retrieval mechanisms and multi-agent reasoning in WiP prediction.

Keywords

Predictive Process Monitoring, Work-in-Progress, Retrieval-Augmented Generation, Large Language Models, Multi-Agent Framework

Authenticity Completeness in Email Systems: A Document-oriented Digital Signature Scheme with Ephemeral Per-message Keys

Ye Li, University of Canberra, Australia

Abstract

Phishing persists because current email architectures cannot guarantee individual-level sender authenticity. Domain-based defenses (SPF, DKIM, DMARC) validate organizational infrastructure yet admit execution paths where messages from compromised accounts are accepted without verifying the named individual. We introduce authenticity completeness, a global property requiring that every execution of the email transmission process terminate either in verified authenticity or explicit rejection, thereby forbidding any form of unverified acceptance. We formalize this property via a state-machine model, prove that domain-level mechanisms are structurally incapable of satisfying it, and present a document-oriented digital signature scheme that uses ephemeral per-message keys. The design satisfies four reinforcing properties—proactive defense, individual-level authenticity, transparency of operation, and localized deploy ability—and we prove that it achieves authenticity completeness. This yields the first rigorous end-to-end guarantee against individual level impersonation within email systems.

Keywords

Email Authenticity, Phishing, Digital Signatures, Authenticity Completeness.

Fortress: A Case Study in Stabilizing Search Recommendations via Temporal Data Augmentation and Feature Pruning

Milind Jagre, Jia Huang, Dayvid V. R. Oliveira, Zhinan Cheng, Babak Seyed Aghazadeh, Puja Das, Chris Alvino, Jinda Han, Kailash Thiyagarajan, USA

Abstract

In search and recommendation systems, predictive models often suffer from temporal instability when certain features introduce volatility in output scores. This instability reduces reliability and user experience - especially in multi-stage systems where consistent predictions are critical. We introduce Fortress, a general framework that enhances model stability and accuracy by identifying and pruning features that cause inconsistent scores over time. Fortress leverages temporally partitioned historical snapshots to capture score fluctuations for the same entity and follows a four-step process: (1) collect snapshots, (2) detect unstable predictions, (3) isolate instability-inducing features, and (4) retrain models with stable features. While semantic features from LLMs improve generalization, they often lack full coverage; engagement features add predictive power but introduce volatility. Fortress suppresses this instability while preserving value, yielding more stable and accurate models. Validated on a query-to-app relevance model in a large marketplace, Fortress shows notable gains in stability and PR-AUC.

Keywords

recommendation system, relevance stability, feature pruning, information retrieval .

Natural Language Processing for Big Data: Challenges, Architectures, and Next-generation Applications

Prakhar Rai, Indian Institute of Technology Guwahati, India

Abstract

Natural Language Processing (NLP) for Big Data has become one of the most challenging frontiers in computer science and engineering. The exponential growth of heterogeneous, multi-modal, and noisy data has pushed NLP beyond classical statistical methods into the realms of distributed deep learning, knowledge-enhanced reasoning, and quantum-inspired architectures. This paper critically analyzes the integration of NLP with Big Data ecosystems, highlighting the interplay of scalability, semantic representation, and distributed optimization. We propose an advanced taxonomy of techniques, comparative analyses of architectures, and future research directions such as neurosymbolic fusion, federated multi-lingual embeddings, and quantum variational NLP models. Our aim is to stimulate the development of next-generation NLP systems capable of thriving in petabyte- scale, privacy-aware, and real-time environments..

Keywords

NLP for Big Data, Deep Learning, Semantic Systems, Quantum NLP, Dis-tributed Architectures.

Action Recognition Based on 3d Object Detection and Normalized Pose Estimation

Satsuki Maeda, Bismark Kweku Asiedu Asante, and Hiroki Imamura, Soka University of Japan, Japan

Abstract

Action recognition has many practical applications, but the task still faces significant challenges. A major challenge is the variation in human action poses across different viewpoints, which complicates determining the ideal pose for action recognition. To address these viewpoint-invariant issues, we propose a pose normalization approach combined with object-based action recognition to classify actions in videos. In this method, the normalized pose is compared with a reference pose to identify the action being performed. The objective of this research is to develop a three-dimensional (3D) object-associated action recognition framework that leverages the stereo camera's ability to capture accurate distance information. This approach offers three main advantages: (1) action recognition that incorporates object context, (2) resolving occlusion problems, and (3) improving recognition accuracy through precise distance information. Experimental results show that our proposed approach achieves 70% classification accuracy across ten selected action categories, independent of viewpoint or camera angle.

Keywords

Object Recognition, Behavior Recognition, AI, Stereo Camera, Active Detection.

A Multi User Seamless Simulation System for Interior and Exterior Design using VR

Takaki Ohoka, Bismark Kweku Asiedu Asante and Hiroki Imamura, Soka University of Japan, Japan

Abstract

In the interior and exterior industry, professionals go through a painstaking task to plan and create functional and aesthetically pleasing indoor and outdoor spaces by selecting materials, colors, finishes, furnitures, and landscaping elements to achieve a harmonious and appealing environment for occupants and visitors. Recently, technological advancements have provided simulation environments using Virtual Reality (VR) or Augment Reality (AR) to easily perform these tasks. However, the challenge of seamlessly switching between interior and exterior environments while simulating in multi-user platforms is still a challenge. This is a result of each of the spaces being simulated differently in VR space. To address this challenge, our research focuses on developing a seamless simulation system for interior and exterior design in a VR space. Our system demonstrates that users can easily switch between interior and exterior designs in the VR spaces in the multi-user platforms as well.

Keywords

Virtual Reality, interior design, exterior design, seamless simulations, 3D modelling, Data Visualization

Authenticity Completeness in Email Systems: A Document-oriented Digital Signature Scheme with Ephemeral Per-message Keys

Ye Li, University of Canberra, Australia

Abstract

Phishing persists because current email architectures cannot guarantee individual-level sender authenticity. Domain-based defenses (SPF, DKIM, DMARC) validate organizational infrastructure yet admit execution paths where messages from compromised accounts are accepted without verifying the named individual. We introduce authenticity completeness, a global property requiring that every execution of the email transmission process terminate either in verified authenticity or explicit rejection, thereby forbidding any form of unverified acceptance. We formalize this property via a state-machine model, prove that domain-level mechanisms are structurally incapable of satisfying it, and present a document-oriented digital signature scheme that uses ephemeral per-message keys. The design satisfies four reinforcing properties—proactive defense, individual-level authenticity, transparency of operation, and localized deploy ability—and we prove that it achieves authenticity completeness. This yields the first rigorous end-to-end guarantee against individual-level impersonation within email systems.

Keywords

Email Authenticity, Phishing, Digital Signatures, Authenticity Completeness.

Fortress: A Case Study in Stabilizing Search Recommendations via Temporal Data Augmentation and Feature Pruning

Milind Jagre, Jia Huang, Dayvid V. R. Oliveira, Zhinan Cheng, Babak Seyed Aghazadeh, Puja Das, Chris Alvino, Jinda Han, Kailash Thiyagarajan, USA

Abstract

In search and recommendation systems, predictive models often suffer from temporal instability when certain features introduce volatility in output scores. This instability reduces reliability and user experience - especially in multi-stage systems where consistent predictions are critical. We introduce Fortress, a general framework that enhances model stability and accuracy by identifying and pruning features that cause inconsistent scores over time. Fortress leverages temporally partitioned historical snapshots to capture score fluctuations for the same entity and follows a four-step process: (1) collect snapshots, (2) detect unstable predictions, (3) isolate instability-inducing features, and (4) retrain models with stable features. While semantic features from LLMs improve generalization, they often lack full coverage; engagement features add predictive power but introduce volatility. Fortress suppresses this instability while preserving value, yielding more stable and accurate models. Validated on a query-to-app relevance model in a large marketplace, Fortress shows notable gains in stability and PR-AUC.

Keywords

recommendation system, relevance stability, feature pruning, information retrieval .

Analyzing Clinical characteristics and Predicting Hospitalization of Older Emergency patients

Ala Karajeh¹ and Rasit Eskicioglu², ¹Independent Researcher, Winnipeg, Canada, ²Department of Data Science and Analytics, Atlas University, Turkey

Abstract

Older patients typically exhibit different traits compared to younger ones and often have multiple comorbidities, such as diabetes and cardiovascular diseases, which complicate their severity assessment at emergency care facilities. This research utilizes two clinical databases from Beth Israel Deaconess Medical Center to explore the clinical characteristics of this demographic based on triage information, triage scores, and disposition outcomes. Additionally, a machine-learning model is proposed to predict likely disposition outcomes, specifically whether patients are hospitalized or discharged at the end of their emergency visit. This model could be instrumental in proactively managing this critical patient segment and improving their health outcomes.

Keywords

Emergency Medicine, Hospitalization Prediction by Machine Learning, Emergency Older Patients Classification, and Emergency Older Patients Data Analytics.