IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligence

Special Track: DBKDA 2025

Chair Prof. Basabi Chakraborty,
Iwate Prefectural University, Iwate, Japan,
Madanapalle Institute of Technology and Science, AP, India
basabi@iwate-pu.ac.jp drbasabic@mits.ac.in

About Me

- Received B. Tech, M. Tech and Ph. D in Radio Physics and Electronics from Calcutta University
- Worked in Indian Statistical Institute, Kolkata as Computer Engineer
- Visiting Researcher in AIC (Advanced Intelligent Communication) Systems in Sendai, Japan (1991-1993)
- Doctoral and Post doctoral research In RIEC (Research Institute of Electrical Communication) , Tohoku University, Japan (1993-1998)
- Ph. D in Information Science (1996)
- Faculty in Dept. of Software and Information Science ,Iwate Prefectural university (1998-2022)
- Visiting faculty in Dept. of Electrical and Computer Engineering, University of Western Ontario, Canada (Oct. 2006 -March 2007)
- Professor Emeritus and Distinguished Professor in Iwate Prefectural University
- Dean, School of Computing, Madanapalle Institute of Science and Technology

Explainable Artificial Intelligence

- Recently eXplainable Artificial Intelligence (XAI) has gained significant attention in the development of trustworthy AI systems.
- Performance of AI/ML models in terms of average accuracy is important but it is one dimensional metric that creates problem with highly complex real world systems.
- Risk sensitive systems – Medical diagnosis, Financial prediction etc
- Safely critical systems : Vehicle control, autonomous decision systems etc.
- To design and develop AI models, transparency should be promoted instead of black box models.

Problems with real world system

- Complex real-world systems
  - Risk-sensitive systems
    - E.g. Medical diagnosis, Financial modeling/prediction
  - Safety-critical systems
    - E.g. Cockpit decision support

Cost of a bad decision can be very high

- Accuracy is not the only objective
- Need for a multi-dimensional perspective

In general, it seems like there are few fundamental problems –

- We don’t trust the models
- We don’t know what happens in extreme cases
- Mistakes can be expensive / harmful
- Does the model makes similar mistakes as humans ?
- How to change model when things go wrong ?

Interpretability is one way we try to deal with these problems

Explainable Machine Learning

- Interpretability : Comprehending what the model is doing (How ?)
- Explainability: Summarizing the reasons for model behavior, causes of decision (Why?)
- Explainable models are interpretable by default, but the reverse is not always true
- Building Interpretable model is the first step.

Interpretable Feature Selection

- Optimal feature subset selection is an integral preprocessing step of building machine learning models.

- Interpretable features lead to interpretable and finally explainable decision models for real world problems.

Incorporating Interpretability

- Global vs Local
- Inherent or Ad-hoc vs Post-hoc
- Model based vs Model Agnostic
- Theory or Applications

IFSEAI Contents ： IFSEAI I

- 58001 Exploring Latent Concepts in SHAP Values -A New Approach Using Singular Value Decomposition – Yukari Shirota, Gakushuin University, Japan yukariyoshiura@gmail.com
- 58002 Route Planning in Wildfire Areas by Integrating a Modified A* Algorithm with Deep Learning - Manavjit Singh Dhindsa, University of WaterLoo, Canada ms2dhind@uwaterloo.ca
- 58003 Time-Series Topic Analysis of Large-Scale Social Media Data using Two-stage Clustering - Takako Hashimoto, Chiba University of Commerce, Japan

takako@cuc.ac.jp

IFSEAI II

- 58004 Evaluating the Potential of SHAP-Based Feature Selection for Improving Classification Performance – Basabi Chakraborty, MITS, AP, India and IPU, Japan

  basabi@iwate-pu.ac.jp
- 58005 Visualizing Proximity of Audio Signals from Different Musical Instruments - A Two Step Approach - Cedric Bornand, University of Applied Sciences, HES-SO, Switzerland cedric.bornand@heig-vd.ch

(From general session)

- 58012 Privacy-preserving data sharing collaborations: architectural solutions and trade-off analysis.
  Michiel Willocx, DistriNet, KU Leuven, Belgium