AI Fairness Amidst Non-IID Data and Algorithmic Bias

Dr. Wenbin Zhang
Presenter: Wangyang Kong

November 11, 2024

1. Research Background and Motivation

- Algorithmic Bias as a Major Issue in the AI Field
- Limitations of Traditional Supervised Learning Paradigms
- Complexities of the Real World: Missing Class Labels, Non-Independent and Identically Distributed Data, Dynamic Data Changes
  - Missing Class Labels
  - Non-Independent and Identically Distributed Data
  - Dynamic Data Changes

2. Objectives of the Paper

- Review the Latest Advances in AI Fairness
- Bridge the Gap Between Theory and Practice
- Envision the Vast Potential in Practical Applications

3. AI Fairness in Non-IID Graph Structures

- Definition and Challenges of Non-Independent and Identically Distributed (Non-IID) Data
- Role of Counterfactual Graph Fairness in Identifying Sensitive Attributes and Label Biases
- Case Study: How Real Fair Counterfactual Graph Neural Networks (RFCGNN) Mitigate Bias

4. Discipline-Specific Fairness

- Unique Challenges of Discipline-Specific Fairness (e.g., Spatial Fairness)
- Ride-Sharing Pricing Example: Higher Charges in High-Population Density, Low-Income Areas
  - Higher Charges in High-Population Density
  - Low-Income Areas
  - Importance of Discipline-Specific Fairness and Policy Recommendations

5. Fairness of Large Language Models (LLMs)

- Importance and Challenges of LLM Fairness
- Sources of Bias: Data, Model Architecture, and Training Process
- Difficulties in Achieving LLM Fairness and Potential Solutions

6. Fairness-Aware Hoeffding Tree (FAHT)

- Design of FAHT: Balancing Prediction and Fairness in Stream Data Classification
- How Fair Information Gain (FIG) Reduces Bias in Splitting
- Extended Version: Fair Adaptive Random Forest (FARF) Provides Diversity and Fairness

7. Conclusion

- Summarize Key Advances in AI Fairness Research
- Emphasize the Vast Potential in Practical Applications
- Outlook on Future Work: Technology, Policy, and Ethics

7. Questions

- How can we ensure that biases in training data are neither amplified nor perpetuated within AI models?
- Should AI fairness employ a unified standard across different disciplines or fields? Why or why not?

7. Answers

- How can we ensure that biases in training data are neither amplified nor perpetuated within AI models?
  - Biases can be filtered and screened through data preprocessing, or by incorporating more diverse perspectives into the dataset. Additionally, fairness constraints or bias detection mechanisms can be introduced during the training process to reduce the propagation of biases within the model. Regularly evaluating the fairness of the model and adjusting training strategies are also important measures.
- Should AI fairness employ a unified standard across different disciplines or fields? Why or why not?
  - Different fields may require distinct fairness standards, as each discipline has its unique sources of bias and fairness requirements. For example, credit scoring in the financial sector and diagnostic models in the medical field are subject to different legal and ethical regulations. Therefore, customized fairness standards are necessary to better align with their respective practical application scenarios.