Advancements in Multi-label Learning

Kaan Demir

CDSAI & SECS
Victoria University of Wellington
PO Box 600, Wellington, New Zealand, 6140

IEEE
NZ Central Section Postgraduate Symposium

Introduction

Multi-Label Learning:

- Predicting a **subset** of K **categories** (0 or 1 for each category)
- $f(\mathbf{x}) = \hat{\mathbf{y}}, \hat{\mathbf{y}} \in \{0,1\}^K, 1 \leq \sum_{y \in \hat{\mathbf{y}}} \leq K$

Introduction (Contd.)

Multi-Label Learning Caveats:

1. Scalability to high-dimensional and large-scale datasets
2. Handling label dependencies and interactions
3. Inconsistency between surrogate loss functions and evaluation metrics

Advancements: Scalability

Challenge: Scalability to High-Dimensional and Large-Scale Datasets

Advancements:

- **Representation Learning:** Reduce data complexity. Focus on relevant information.
  - Demir, et. al., Co-operative Co-evolutionary Many-objective Embedded Multi-label Feature Selection with Decomposition-based PSO, ACM GECCO, 2023.
  - Hu, et. al., Dynamic subspace dual-graph regularized multi-label feature selection, Neurocomputing, 2022.
  - Jian, et. al., Multi-label informed feature selection, IJCAI, 2016.

Advancements: Label Dependencies

Challenge: Handling Label Dependencies and Interactions (*the* hot topic!)

Advancements:

- **Graph-Based Models:** Represent labels as nodes and relationships as edges in a graph.
  - Demir, et. al., Dual Sparse Structured Subspaces and Graph Regularisation for Particle Swarm Optimisation-based Multi-label Feature Selection, IEEE CIM, (Accepted).
  - Yuan, et. al., Graph Attention Transformer Network for Multi-label Image Classification, ACM TOMM, 2023.
  - Hang, et. al., Collaborative Learning of Label Semantics and Deep Label-Specific Features for Multi-Label Classification, TPAMI, 2022.
  - Wang, et. al., Cross-Modality Attention with Semantic Graph Embedding for Multi-Label Classification, AAAI, 2020.
  - Wang, et. al., Multi-Label Classification with Label Graph Superimposing, AAAI, 2020.

State-of-the-art Methods

Benefits:

- Enhanced ability (and efficiency) to handle complex real-world scenarios.
- Improved prediction accuracy by adapting to varying label dependencies.

Advancements: Consistency

Challenge: Surrogate Loss vs. Evaluation Metrics
Many multi-label learning models optimize surrogate loss functions during training, which may not always align perfectly with the desired evaluation metrics (e.g., Hamming loss, Ranking loss).
Advancements:

- Actually, we have proven that consistency cannot be achieved in many cases (in-fact, only partial consistency):
  - Gao and Zhou, On the consistency of multi-label learning, PMLR, 2011.
  - Liu, et. al., The Emerging Trends of Multi-Label Learning, TPAMI, 2022.

What *is* "consistency" (and why should we care about it)?

Usually, the learning behaviour of model f is prescribed by the gradients (or direction) of a carefully designed loss function ψ.

$$\begin{array}{ccc} \text{Loss} & \searrow & \\ & & \text{Alignment} \\ \text{Metric} & \nearrow & \end{array}$$

$$\psi(p(\mathbf{x}), f) = \sum_{\mathbf{y}\in\mathcal{Y}} p(\mathbf{y}|x)\psi(f(\mathbf{x}), \mathbf{y}) \tag{1}$$

$$R_\psi(f) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{P}}[\psi(p(\mathbf{x}), f)] \quad R_\psi^B(f) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{P}}[\inf_{f'}[\psi(p(\mathbf{x}), f')]] \tag{2}$$

Multi-label Consistency

Suppose ψ is the surrogate loss approximating $\mathcal{L}$, then given some sequence $f^{(n)}$ if:

$$\lim_{n\to\infty} R_{\psi}(f^{(n)}) \to R_{\psi}^{B}(f) \quad \text{then} \quad R_{\mathcal{L}}(f^{(n)}) \to R_{\mathcal{L}}^{B}(f) \tag{3}$$

$\exists f \in \Omega$.

Multi-label Consistency (Contd.)

This is not the case for almost all multi-label metrics!
Please refer to Gao and Zhou, On the consistency of multi-label learning, PMLR, 2011.

Conclusion

Multi-label learning has seen significant advancements:

- Scalability enabling efficient processing of large datasets.
- Handling label dependencies and interactions.
- Consistency remains an open topic.

Further exploration and collaboration are crucial for progress in this field.

IEEE
Advancing Technology for Humanity

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA
ECRG
Evolutionary Computation Research Group

Thank you for your attention! Any questions?