*Journal & Proceedings of the Royal Society of New South Wales*, vol. 157, part 2, 2024, pp. 307–308. ISSN 0035-9173/24/020307-02

## Thesis abstract

# Generalizable and robust machine learning with imperfect data supervision

**Shuo Yang**

Abstract of a thesis submitted to University of Technology Sydney

Modern machine-learning systems, particularly deep neural networks, have driven advancements in many artificial-intelligence domains. Their success largely depends on ample, high-quality labelled data for training. However, consistently accessing such data isn't always possible due to privacy, ethical, or economic barriers. Real-world situations often present data that is limited in quantity or compromised in quality. For example, medical datasets are sparse due to privacy concerns, and multi-modal datasets are frequently noisy, as they are routinely sourced from the internet or labelled via crowd-sourcing platforms. These imperfect data conditions lead to suboptimal deep-learning models that are prone to overfitting, biased or noisy data distributions. This thesis strives to provide theoretical understanding, empirical analysis, and methodological solutions for training resilient and universally applicable deep-learning models under such imperfect data-resource supervision.

In addressing the problem of generalizable learning with scarce data, the distribution bias issue is first analysed. A few data points usually form a biased data distribution. A discrepancy between the biased and ground-truth data distributions is identified as the root cause of poor generalisation in deep models. A distribution calibration technique is introduced to rectify this bias, helping models trained on sparse data to maintain high performance. Subsequently, a technique, named dataset pruning, is proposed to determine the minimum necessary training data size, ensuring consistent performance between models trained on the full and pruned datasets. Observations from pruning several large-scale datasets show that a small portion can nearly match the original's performance, underscoring the surprising capability of minimal training data points.

Turning attention to the problem of robust learning with noisy data, the investigation begins with label noise in the classification task, wherein label noise manifests in the form of categorical annotation errors. A parametrical model is proposed to bridge the distribution gap between noisy labels and clean labels and significantly improve the robustness of the learned model. It is further shown that a classifier trained on the noisy dataset will asymptotically converge to the Bayes optimal classifier with an optimal convergence rate. The label-noise problem is then extended to a more realistic and challenging context, namely, multi-modal learning, where the label noise refers to alignment errors in paired data. To tackle this, a general framework termed BiCro (Bidirectional Cross-modal similarity consistency) is proposed. This framework can be conveniently integrated into exist-

ing multi-modal learning models, thereby augmenting their resilience to noisy data.

URL: https://opus.lib.uts.edu.au/handle/10453/173602

Dr Shuo Yang
Department of Computer Science
The University of Hong Kong

E-mail: syang98@hku.hk