Generative AI, Data compression

Developing learning theory for uncertainty using Information theory

Department of Systems Innovation, Graduate School of Engineering science

Associate Professor Futoshi Futami

Researchmap https://researchmap.jp/ffutami?lang=en

Abstract

Probabilistic inference in machine learning is used not only to achieve good predictive accuracy, but also plays a central role in quantifying predictive uncertainty and in extracting and generating knowledge through latent variable models. However, the theory for these roles remains limited. This project develops a learning theory for probabilistic algorithms grounded in information theory and PAC-Bayesian analysis. We study uncertainty manifested as predictive dispersion, confidence or credible intervals, class posterior probabilities, and latent representations produced by encoder-based models. By analyzing these quantities directly, rather than only prediction error, we obtain principled guarantees on both the quality of uncertainty quantification in prediction and the behavior of latent variables in deep generative models.

Background & Results

Probabilistic inference were originally developed within a Bayesian framework and have long played a central role in statistics. Combined with deep learning, they now support knowledge extraction through latent variables, uncertainty aware prediction, and decision making under complex environments. Modern probabilistic inference, however, is not limited to exact Bayesian updates. Approximate schemes such as variational Bayes and generalized Bayesian procedures with task specific update rules are widely used in practice.

Existing theory mainly assumes independent and identically distributed data and focuses on prediction error, offering limited insight into latent variable models or refined uncertainty quantification. Consequently, many methods still rely on heuristics, which can undermine robustness and trustworthiness in high-stakes settings.

This project provides a theoretical formalization and guarantees for uncertainty measures and encoder based latent variables that had previously been treated mostly empirically. By clarifying links between information theoretic quantities and classical statistical learning theory, we build a framework in which generalization performance and uncertainty are discussed jointly via information measures and loss smoothness within a uniform convergence perspective. Our analysis further shows that deep latent variable models such as Variational Autoencoders can be handled naturally in this framework by treating latent variables themselves as theoretical objects, extending the scope of learning theory for generative models.

Significance of the research and Future perspective

This research addresses a core requirement for reliable machine learning: principled quantification and theoretical understanding of uncertainty in model predictions and latent representations. By integrating information theory with PAC Bayesian methods, the project forms the basis of a learning theory centered on uncertainty, beyond accuracy focused analyses. We plan to extend this framework to non i.i.d. and interactive settings and to derive practical design guidelines for high reliability AI systems that must operate safely under uncertainty.

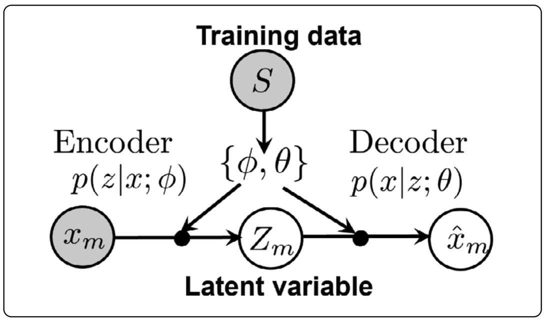

Fig.1: Latent variable models with encoders (e.g., variational autoencoders)

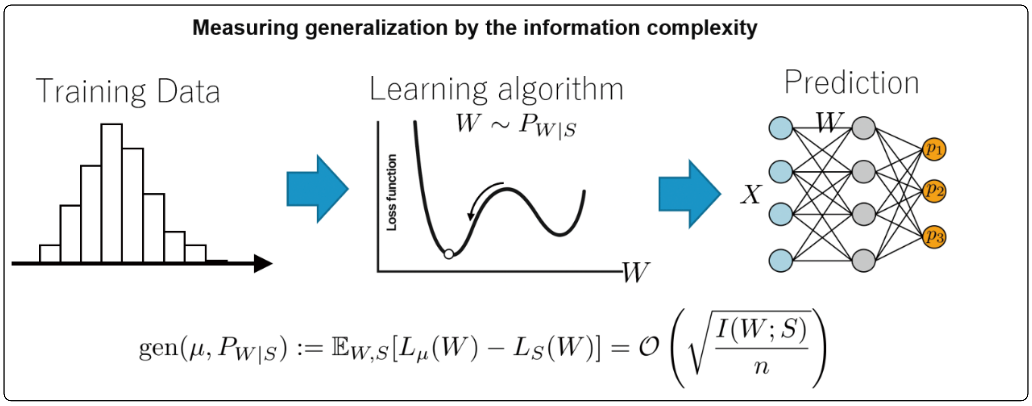

Fig.2: Relationship between statistical learning theory and information theory

Patent

Treatise

Futami, Futoshi & Fujisawa, Masahiro. Information-theoretic Generalization Analysis for VQ-VAEs: A Role of Latent Variables. Accepted at Advances in Neural Information Processing, 2025.

Fujisawa, Masahiro & Futami, Futoshi. PAC-Bayes Analysis for Recalibration in Classification. Proceedings of 42nd International Conference on Machine Learning, 267, 17986-18023, 2025. doi: 10.48550/arXiv.2406.06227

Futami, Futoshi. Epistemic Uncertainty and Excess Risk in Variational Inference. Proceedings of 28th International Conference on Artificial Intelligence and Statistics, 258, 568-576, 2025.

Futami, Futoshi & Fujisawa, Masahiro. Information-theoretic Generalization Analysis for Expected Calibration Error. Advances in Neural Information Processing Systems 37, 84246-84297, 2024. doi: 10.48550/arXiv.2405.15709

Futami, Futoshi & Iwata, Tomoharu. Information-theoretic Analysis of Bayesian Test Data Sensitivity. Proceedings of 27th International Conference on Artificial Intelligence and Statistics, 238, 1099-1107, 2024.

URL https://sites.google.com/view/futoshifutami/home

Keyword statistical learning theory, machine learning, generative AI, information theory, Bayesian inference