

PhD position at IMT Atlantique : Information Theory for Machine Learning over Compressed Data

PhD position at IMT Atlantique, Brest, France. Expected starting date: fall 2024.

Advisors:

- Elsa Dupraz, IMT Atlantique (elsa.dupraz@imt-atlantique.fr),
- Philippe Mary, INSA Rennes (philippe.mary@insa-rennes.fr)

Context

Join a dynamic research team focused on exploring the interplay between Machine Learning and Information Theory. Our research lab is situated within IMT Atlantique, one of French's top engineering schools, nestled in the vibrant city of Brest - a hub of academia, culture, and maritime heritage, often referred to as the 'end of the earth' and a land of storms.

This opportunity arises from a collaborative effort between IMT Atlantique and INSA Rennes. As a PhD student based in Brest, you will have the advantage of being part of a stimulating and interdisciplinary academic environment, with regular visits to our partner institution in Rennes.

Motivation of the PhD

In the emerging field of "goal-oriented communications" [1], the aim of the receiver is no longer to reconstruct data, but instead to perform a specific learning task (classification, decision-making, semantic analysis, etc.) on the received data. To significantly improve transmission efficiency, it is essential to address this task and its specific performance criteria when designing the communication system.

Such an approach is even more critical today as the amount of data available online is huge: every minute, 500 hours of video are uploaded to YouTube, and 240,000 images are sent on Facebook. Therefore, it is essential to resort to advanced learning methods to process the data, sort it, or recommend it to users.

In this PhD, we will aim to consider the common case of a storage server containing a large amount of compressed multimedia data (images, video, etc.), issued, for example, from a social network. The goal of the PhD will be to design compression systems dedicated to learning tasks to be applied on the coded data. We will focus particularly on unsupervised learning tasks, *i.e.*, those for which no labels are available for learning. These labels can indeed be difficult to obtain in such a context.

Challenges to be addressed during the PhD

A strategy that would involve decompressing all the data before performing the learning task would be extremely costly in terms of computational resources. Therefore, during the PhD, we will first aim to address the following question:

Question 1 (no prior decoding): Is it possible to encode the data in such a way as to be able to perform the learning task without any prior decoding operation?

Furthermore, most of the time, the learning task that will be performed later is not known at the time of data compression. Therefore, we also target to study a second question:

Question 2 (universality): Is it possible to design a universal coding scheme, *i.e.*, one that would enable the application of different learning tasks (regression, classification, etc.) on the same compressed data?

A coding scheme that meets the constraints of the two previous questions might result in a loss of performance compared to learning after decompression. To characterize this loss, which we hope will be limited, a first step will involve an **information theory analysis** [2], [3], [4], to determine the achievable performance limits of coding systems dedicated to a particular learning task. This analysis will take into account the aforementioned constraints and should allow for a comparison with learning after decompression. Furthermore, the information theory analysis should provide useful insights and design tools for a second step, which will involve the development of **practical coding schemes** for the identified problems.

Candidates Profile

The candidate should have earned an MSc degree, or equivalent, in one of the following fields: Telecommunications, Signal Processing, Applied Mathematics. Some knowledge about Information Theory and/or Machine Learning would also be appreciated.

How to apply

Application should contain: a full CV, complete academic records (from Bachelor to Msc), as well as contact details of one or two referees (former internship advisor, etc.). No need to include a motivation letter, but please explain in a few words in the e-mail why you apply for the position, and why you think your background is relevant. **Please send your application to Elsa Dupraz (elsa.dupraz@imt-atlantique.fr) and Philippe Mary (philippe.mary@insa-rennes.fr).**

References

- [1] E. C. Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.
- [2] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6228–6237.
- [3] P. A. Stavrou and M. Kountouris, "A rate distortion approach to goal-oriented communication," in *IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 590–595
- [4] Jiahui Wei, Philippe Mary, Elsa Dupraz, Rate-Loss Regions for Polynomial Regression with Side Information, in *International Zurich Seminar (IZS) 2024*.