Thesis proposal

Nikita Tokovenko

July 2020

1 Problem statement

Information-based models for deep learning has become a sought-after subject of ongoing research. Many recent methods for self-supervised and unsupervised representation learning train feature extractors using an estimate of the mutual information. Such an approach proved its ability to reach state-of-the-art performance in many applications. We believe that investigating the Information Bottleneck principle will give useful insights to find a key to understanding how to build more efficient and accurate algorithms to solve the tasks of deep learning.

The target goal of this thesis is to define a procedure to make one able to reason about the generalization capability of different encoding procedures. In order to do this we proposed a framework to make a comparison of various supervised encoding techniques.

2 Research questions

Our hypothesis: deep learning models that incorporate information bottleneck principle can learn high-level representations from labeled data of not much lower (or even higher) quality while working at higher compression rates when compared to other ways to increase the robustness of features learnt (e.g. different regularization techniques).

In order to validate this hypothesis we aim to answer following research questions:

1. How true is that usage of information bottleneck method helps to compress the data without the loss of its expressiveness?

2. How can we quantitatively evaluate the quality of representations with respect to the amount of mutual information preserved?

3. How does the choice of the model architecture can affect this metric?

4. How does using information bottleneck objective relate to applying regularization to the model trained?

5. Can we design a more effective method to discard only irrelevant information with respect to VIB?

3 Literature Review, positioning your research in the relevant background and related research

Initially Information Bottleneck was formulated by []. In this work authors presented a general formulation of the information theoretic approach for finding representations of the signal in a manner to capture its relevant structure. Back in the days there was no way to expressively use it in application to real-world problems in machine learning - the issue of applying Information Bottleneck (IB) directly to training deep neural networks is that estimation of mutual information, which is crucial for definition of the target function, is a tough and opened task stand-alone. Later in [**?**], after the deep learning revolution, the adaptation of the information bottleneck to deep neural networks has been proposed. Authors provided a theoretical overview of IB as a special type of Rate Distortion problem to be solved by neural networks. However, proposed procedure relied on optimizing objective using iterative Blahut Arimoto algorithm, which is unfeasible in application to deep neural networks.

Recent works in [**?**] showed some promising results on bounding the estimate of the true mutual information. It is not entirely clear how tight these bounds should be in order to provide a solution for building high-quality representations. Also investing a lot of computational power might be unfeasible because building high-confidence bound on entropy requires an exponentially big sample size. An explanation on why maximizing MI does not necessarily lead to useful representations [**?**].

Recently proposed Variational Information Bottleneck (VIB) [**?**] addresses the problem of fitting stochastic encoders using the variational inference. Such approach showed to be more robust to overfitting while still achieving the state-of-the-art. In the supervised learning literature, their work is related to confidence penalty methd [**?**], while in the unsupervised learning is closely related to the work of [**?**] on variational autoencoders. VIB, same as VAE, regulates compression and fitting with only difference that target labels are available during training time.

4 Methodology which methods you are using to systematically research your topic, and by which you are answering your research questions. This helps to check the validation and verification of your research.

In order to be able to reason about the quality of the encoding procedure we need to run distinct experiments by build the entire framework.

Framework consists of 3 major procedures: training the encoder by following the chosen training method, quantitaive evaluation of its generalization capability, building the mutual information estimationTraditionally, generalization capability has been assessed by measuring the accuracy for representations to solve the down-stream classification task. We build our metric from a perspective of dependence of the prediction quality on the amount of additional data points available to build this classifier (say, logistic regression).

In order to obtain mutual information estimation we use NWJ estimator, since using unnormalized bounds is attractive because they provide tractable estimators. However, in practice they show to have high variance.

5 Planning your planning on experimenting, evaluation, writing etc.

1) February-March: Project setup and literature exploration
1.1) Literature review
1.2) Definition of the benchmark tasks
1.3) Implementation and comparison of mutual information estimators
1.4) Implementation of the generalization evaluation metrics
1.5) Finalization of the standardized benchmarking framework

2) March-May: Comparison of the different architecture
2.1) Implementation of different neural network architectures
2.2) Experimental data collection phase
2.3) Analysis and visualization of the empirical results

3) June-August: Thesis writeup

Steps 2.1 to 2.3 will be repeated multiple times on different models. The total amount of iterations depends on time availability and relevant findings. The time required for step 1.3) might also vary depending on the variance registered on the empirical measurements as accurate Mutual Information estimation is an open problem in the research field.