



TAPE

NIPS

19

Property prediction

Benchmark

- 5 semi-supervised tasks
 1. Secondary structure: helix, strand, other
 2. Structure: each pair is in contact or not
 3. Remote homology detection
 4. Fluorescence
 5. Stability
- Transformer > LSTM > ResNet > UniRep

Transformer

mLSTM

Evaluating Protein Transfer Learning with Tasks Assessing Protein Embeddings (TAPE)

Protein modeling is an increasingly popular area of machine learning research. Semi-supervised learning has emerged as an important paradigm in protein modeling due to the high cost of acquiring supervised protein labels, but the current literature is fragmented when it comes to datasets and standardized evaluation techniques. To facilitate progress in this field, we introduce the Tasks Assessing Protein Embeddings (TAPE), a set of five biologically relevant semi-supervised learning tasks spread across different domains of protein biology. We curate tasks into specific training, validation, and test splits to ensure that each task tests biologically relevant generalization that transfers to real-life scenarios. We benchmark a range of approaches to semi-supervised protein representation learning, which span recent work as well as canonical sequence learning techniques. We find that self-supervised pretraining is helpful for almost all models on all tasks, more than doubling performance in some cases. Despite this increase, in several cases features learned by self-supervised pretraining still lag behind features extracted by state-of-the-art non-neural techniques. This gap in performance suggests a huge opportunity for innovative architecture design and improved modeling paradigms that better capture the signal in biological sequences. TAPE will help the machine learning community focus effort on scientifically relevant problems.

Add paper
(Google Form)

Design and Engineering

Deep learning and architecture

Annotation

Color box
Reviewed!
Click to see the summary

Grey box
Not yet reviewed

Representation

Property prediction

Augmentation

Folding / Inverse folding

