



Language Models are Unsupervised Multitask Learners

DSCI560

Instructor: Young Cho, Ph.D.

Yifan Yang

8386626867

INTRODUCTION



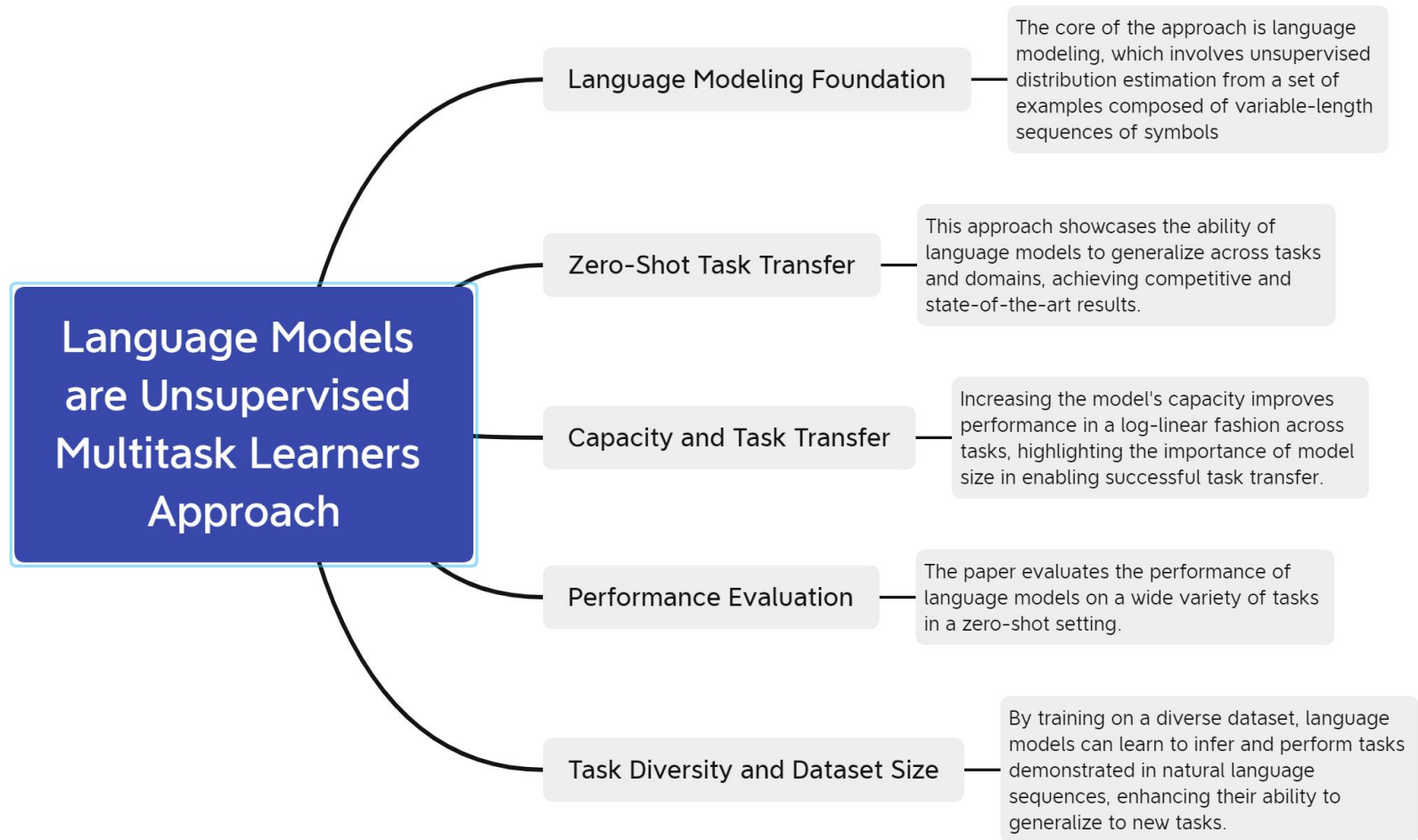
Highlights from what I read:

Language models have the **transformative potential** as **unsupervised multi-task** learners, capable of **inferring and performing** a wide range of tasks without explicit supervision.

The paper introduces a novel perspective by demonstrating that language models can begin to learn these tasks **without explicit supervision** when trained on a new dataset called WebText, which contains millions of web pages.

Language model capabilities play an important role in **enabling zero-shot task** transfer, and performance improvements scale logarithmically across tasks.

Language models are used for **specific tasks such as common sense reasoning and sentiment analysis**, as well as a trend towards more general transfer learning methods.



RESULTS



This experimental part is designed to be relatively basic and simple in my opinion. The purpose is to evaluate the performance of language models performing various tasks in **zero-shot settings**. The experiment adopts a zero-shot setting, that is, directly testing the performance of the language model when performing tasks without receiving training on a specific task. This setting is designed to evaluate the **generalization ability and transfer learning effect** of language models on unseen tasks.

Generalization vs Memorization

Focus on the language model's ability to **generalize** to unseen data after training. Pay attention to whether the language model **over-memorizes** training data. Over-memory may result in a model that performs well on the test set but lacks generalization ability in real-world applications.

The degree of **overlap** between training data and test data was analyzed. If there is a lot of overlap in the test data with the training data, **the model may favor memorization over generalization**.

Discussions and Conclusions



Highlights from what I read:

Overmemorizing training data can result in a model that performs well on the test set but **lacks generalization ability**.

Highly overlapping data can cause the model to overmemory, **affecting** its performance on new data.

Increasing model capacity may improve memory ability, but may also **decrease generalization ability**.

Q1: How to balance memory ability and generalization ability of language models during the training process is a key challenge. how to respond?

A1: If a model overmemorizes the training data, we can artificially increase the diversity of the training data through regularization techniques. During training, once performance on the validation set starts to degrade, training is stopped to avoid overfitting. For data overlapping issues, we need to consider data deduplication, multi-task learning, and increase model capacity.