

Here is a summary of the paper "SPECTER: Document-level Representation Learning using Citation-informed Transformers":

The paper introduces SPECTER, a new method for generating document-level embeddings of scientific papers by pretraining a Transformer language model (SciBERT) using citation graph information.

Key points:

- SPECTER leverages citation relationships between papers as a signal of document-level relatedness during pretraining, using a triplet loss objective to learn embeddings where cited papers are closer in the embedding space.
- Unlike most pretrained language models that require task-specific fine-tuning, SPECTER produces general-purpose embeddings that can be directly used for downstream tasks without additional training.
- The authors introduce SCIDOCS, a new benchmark with 7 document-level tasks (classification, citation prediction, recommendation etc.) for evaluating scientific paper embeddings.
- Experiments show SPECTER outperforms strong baselines across all tasks, with an average 3.1 point improvement over the next best method.
- The model only requires title and abstract as input at inference time, making it applicable to new uncited papers.
- The paper demonstrates the effectiveness of incorporating inter-document relationships (via citations) into Transformer-based document representations.

The approach combines the strengths of pretrained language models with citation graph information to produce high-quality document embeddings for scientific papers. The released code, models and SCIDOCS benchmark facilitate further research in document-level representation learning.