**Exploring the Limits of Language Modeling**

**Rafal Jozefowicz**

**Oriol Vinyals**

**Mike Schuster**

**Noam Shazeer**

**Yonghui Wu**

RAFALJ@GOOGLE.COM

[VINYALS@GOOGLE.COM](mailto:VINYALS@GOOGLE.COM)

[SCHUSTER@GOOGLE.COM](mailto:SCHUSTER@GOOGLE.COM)

[NOAM@GOOGLE.COM](mailto:NOAM@GOOGLE.COM)

YONGHUI@GOOGLE.COM

**Abstract**

In this work we explore recent advances in Recurrent Neural Networks for large scale Language Modeling, a task central to language understanding. We extend current models to deal with two key challenges present in this task: corpora and vocabulary sizes, and complex, long term structure of language. We perform an ex- haustive study on techniques such as character Convolutional Neural Networks or Long-Short Term Memory, on the One Billion Word Benchmark. Our best single model significantly improves state-of-the-art perplexity from 51.3 down to 30.0 (whilst reducing the number of parameters by a factor of 20), while an ensemble of models sets a new record by improving perplexity from 41.0 down to 23.7. We also release these models for the NLP and ML community to study and improve upon.

**摘要**

本文探索了语言理解的核心任务，即大规模语言建模的递归神经网络的最新进展。并扩展了当前模型以应对此任务中存在的两个关键挑战：语料库和词汇量的大小，以及长期语言结构的复杂性。本文对十亿字基准上的字符卷积神经网络或长短期记忆网络等技术进行了一项前瞻性研究。我们最好的单一模型达到了最先进的效果，显着地将困惑度从51.3降低到30.0（同时将参数数量减少了20倍），而模型集合通过将困惑度从41.0降低到23.7创造了新的记录。本文的模型已经在NLP和ML社区发布以进行接下来的研究和改进。