Introduction to Transformers for NLP

With the Hugging Face Library and Models to Solve Problems

Shashank Mohan Jain

Apress®

Introduction to Transformers for NLP: With the Hugging Face Library and Models to Solve Problems

Shashank Mohan Jain
Bangalore, India

ISBN-13 (pbk): 978-1-4842-8843-6 ISBN-13 (electronic): 978-1-4842-8844-3
https://doi.org/10.1007/978-1-4842-8844-3

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin Suresh John
Development Editor: James Markham
Coordinating Editor: Shrikant Vishwakarma

Cover designed by eStudioCalamar

Cover image by and machines on Unsplash (www.unsplash.com)

Distributed to the book trade worldwide by Apress Media, LLC, 1 New York Plaza, New York, NY 10004, U.S.A. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub (https://github.com/Apress). For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

Table of Contents

About the Author

Shashank Mohan Jain has been working in the IT industry for around 22 years mainly in the areas of cloud computing, machine learning, and distributed systems. He has keen interests in virtualization techniques, security, and complex systems. Shashank has many software patents to his name in the area of cloud computing, IoT, and machine learning. He is a speaker at multiple reputed cloud conferences. Shashank holds Sun, Microsoft, and Linux kernel certifications.

About the Technical Reviewer

Akshay Kulkarni is a renowned AI and machine learning evangelist and thought leader. He has consulted several Fortune 500 and global enterprises on driving AI and data science–led strategic transformation. Akshay has rich experience in building and scaling AI and machine learning businesses and creating significant impact. He is currently a data science and AI manager at Publicis Sapient, where he is a part of strategy and transformation interventions through AI. He manages high-priority growth initiatives around data science and works on various artificial intelligence engagements by applying state-of-the-art techniques to this space. Akshay is also a Google Developers Expert in machine learning, a published author of books on NLP and deep learning, and a regular speaker at major AI and data science conferences. In 2019, Akshay was named one of the top "40 under 40 data scientists" in India. In his spare time, he enjoys reading, writing, coding, and mentoring aspiring data scientists. He lives in Bangalore, India, with his family.

Introduction

This book takes the user through the journey of natural language processing starting from n-gram models to neural network architectures like RNN before it moves to the state-of-the-art technology today, which is known as the transformers. The book details out the transformer architecture and mainly explains the self-attention mechanism, which is the foundation of the transformer concept.

The book deals with the topic of transformers in depth with examples from different NLP areas like text generation, sentiment analysis, zero-shot learning, text summarization, etc. The book takes a deep dive into huggingface APIs and their usage to create simple Gradio-based applications. We will delve into details of not only using pretrained models but also how to fine-tune the existing models with our own datasets.

We cover models like BERT, GPT2, T5, etc., and showcase how these models can be used directly to create a different range of applications in the area of natural language processing and understanding.

The book doesn't just limit the knowledge and exploration of transformers to NLP but also covers at a high level how transformers are being used in areas like vision.

Source Code

All source code used in this book can be found at `github.com/apress/intro-transformers-nlp`.