Machine Learning Techniques for Text

Introduction

Dr. Nikos Tsourakis

Overview

<packt>

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

1ST EDITION

- Crafting machines that can learn from data to perform intelligent decisions is becoming the dominant paradigm in many areas of technology

- Acquiring the necessary skill set to perform this task will definitely boost your skillset

- **Machine Learning Techniques for Text** aims to help you in this endeavor, focusing specifically on text data and human language

- Between resources presenting complex theoretical concepts or focus disproportionately on Python code, this seminar steers a middle path to keep the right balance between theory and practice

- No *Python* or *machine learning knowledge* is required

Course outline

<packt>

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

- **Module 0: Python Crash Course**
- **Module 1: Intro to Machine Learning**
- **Module 2: Detecting Spam Emails**
- **Module 3: Classifying Topics of Newsgroup Posts**
- **Module 4: Extracting Sentiments from Product Reviews**
- **Module 5: Recommending Music Titles**
- **Module 6: Teaching Machines to Translate**
- **Module 7: Summarizing Wikipedia Articles**
- **Module 8: Detecting Hateful and Offensive Language**
- **Module 9: Generating Text in Chatbots**
- **Module 10: Clustering Speech-to-Text Transcriptions**

Course objectives

‹packt›
Machine Learning Techniques for Text
Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

After completing this course, you should be able to:

- Analyze text data, get started with machine learning, and work effectively with the Python libraries often used for these tasks, such as pandas, NumPy, matplotlib, seaborn, and scikit-learn
- Work with state-of-the-art deep learning frameworks such as TensorFlow, Keras, and PyTorch
- Identify at least one practical usage for the method or technique presented

Machine Learning Techniques for Text

Section 1: Introducing machine learning for text

The intersection of text and machine learning

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

1ST EDITION

- ***Text analysis*** involves extracting valuable insights and information from textual data, a task critical for understanding and decision-making.
- ***Machine learning*** empowers computers to learn patterns and make predictions without explicit programming, unlocking the potential to derive meaningful insights from large volumes of data
- The marriage of text analysis and machine learning enhances our capacity to extract meaningful information from the vast landscape of textual data, driving advancements in various fields
- Text can be ***structured*** (tables, forms, and databases with labeled fields and columns) or ***unstructured*** (plain text documents, social media posts, and most natural language content)

Peculiarities of human language

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

- **Complexity and Ambiguity**: Human language is inherently complex and often ambiguous. Words can have multiple meanings, and context plays a crucial role in interpretation
- **Variability and Evolution**: Languages evolve over time, and variations exist across regions, communities, and generations
- **Creativity and Expressiveness:** Humans use language not just for conveying information but also for creativity and expression (metaphors, idioms, etc.)
- **Ambivalence and Emotional Tone**: Human language often conveys emotional undertones and ambivalence
- **Cultural Influence**: Language is deeply intertwined with culture

Key techniques

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

- **Natural Language Processing** (NLP)**:** NLP equips machines to comprehend, interpret, and respond to human language, enabling applications like sentiment analysis and language translation
- **Text Classification:** This technique involves categorizing text into predefined groups or classes, allowing for automated sorting and organization of textual data
- **Named Entity Recognition** (NER)**:** NER identifies and classifies entities, such as names, locations, and organizations, within text, facilitating information extraction
- **Text Generation:** Advanced models like GPT can generate coherent and contextually relevant human-like text, revolutionizing content creation and language generation tasks

Applications

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

- Answering natural language questions—For example, IBM Watson
- Summarizing documents—analyzing documents and producing short summaries
- Speech synthesis (speech-to-text) and speech recognition (text-to-speech)
- Text-to-text translation and voice-to-voice translation
- Collaborative filtering—used to implement recommender systems
- Text classification—for example, classifying news articles by categories, etc.
- Topic modeling—finding the topics discussed in documents
- Sarcasm detection—often used with sentiment analysis
- Text simplification—making text more concise and easier to read
- Speech to sign language and vice versa—to enable a conversation with a hearing-impaired person
- Lip reader technology—convert lip movement to text or speech to enable conversation
- Closed captioning—adding text captions to video

Applications

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

Answering natural language questions—For example, IBM Watson

Summarizing documents—analyzing documents and producing short summaries

Speech synthesis (speech-to-text) and speech recognition (text-to-speech)

Text-to-text translation and voice-to-voice translation

Collaborative filtering—used to implement recommender systems

Text classification—for example, classifying news articles by categories, etc.

Topic modeling—finding the topics discussed in documents

Sarcasm detection—often used with sentiment analysis

Text simplification—making text more concise and easier to read

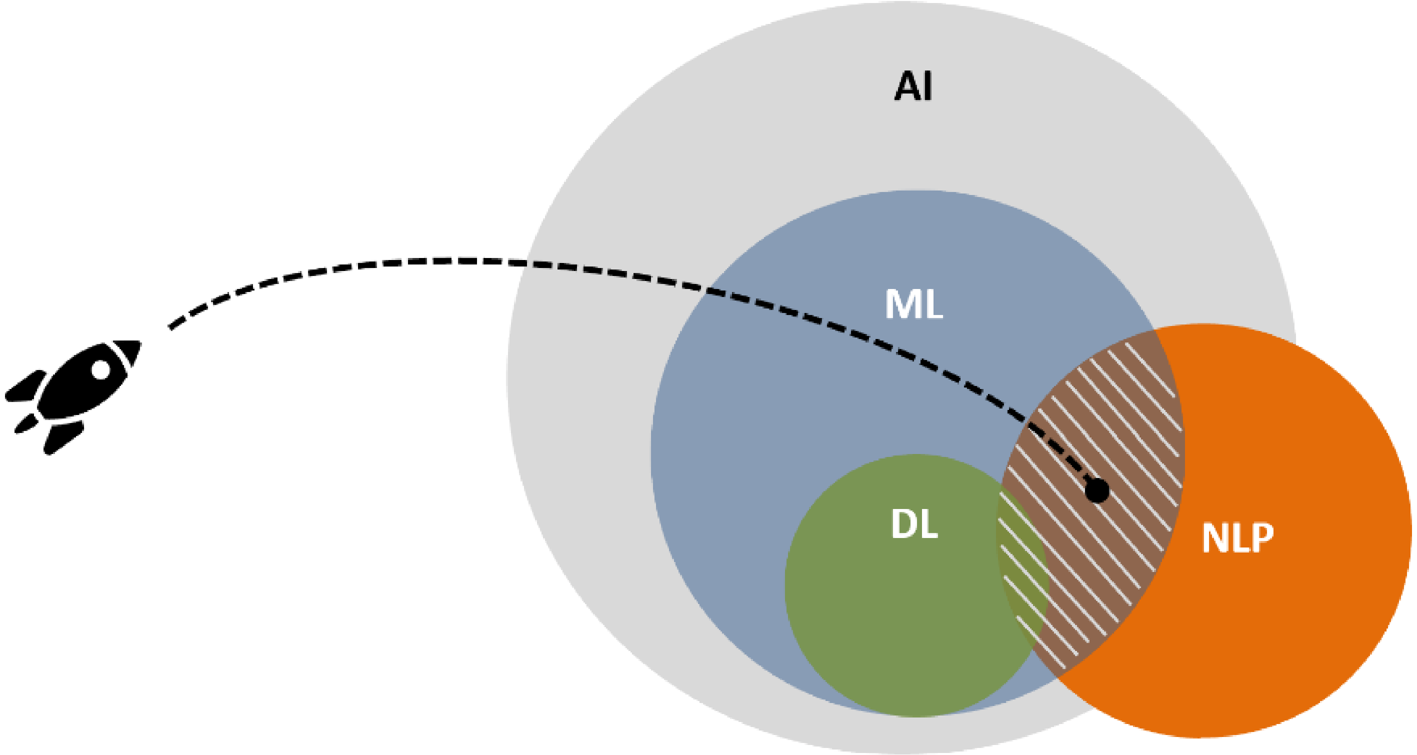
Speech to sign language and vice versa—to enable a conversation with a hearing-impaired person

Lip reader technology—convert lip movement to text or speech to enable conversation

Closed captioning—adding text captions to video

… and many more!

Our landing point



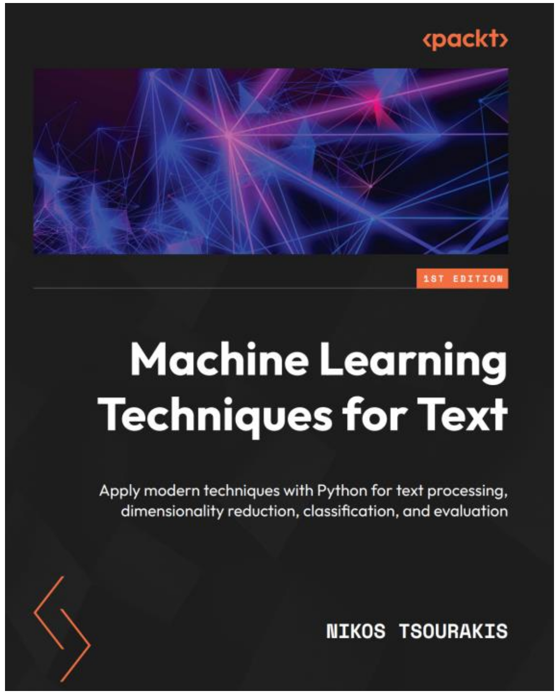
AI – Artificial intelligence, **ML** – Machine Learning, **DL** – Deep Learning, **NLP** – Natural Language Processing

Machine Learning Techniques for Text

Section 2: The book

Overview

- A book for professionals in the area of computer science, programming, data science, informatics, business analytics, statistics, language technology, and more who aim for a gentle career shift in ***machine learning for text***
- Students in relevant disciplines that seek a textbook in the field will benefit from the practical aspects of the content and how the theory is presented
- Professors teaching a similar course will be able to pick pertinent topics in terms of content and difficulty
- Beginner-level knowledge of ***Python*** programming is needed to get started with this book



‹packt›

1ST EDITION

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

NIKOS TSOURAKIS

amazon https://a.co/d/856lejK

GitHub https://github.com/PacktPublishing/Machine-Learning-Techniques-for-Text

Table of contents

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

1. Introducing Machine Learning for Text
2. Detecting Spam Emails
3. Classifying Topics of Newsgroup Posts
4. Extracting Sentiments from Product Reviews
5. Recommending Music Titles
6. Teaching Machines to Translate
7. Summarizing Wikipedia Articles
8. Detecting Hateful and Offensive Language
9. Generating Text in Chatbots
10. Clustering Speech-to-Text Transcriptions

Table of contents

‹packt›

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

1.	Introducing Machine Learning for Text	*“What is ML for text?”*
2.	Detecting Spam Emails	*“How can I know that an email is spam?”*
3.	Classifying Topics of Newsgroup Posts	*“How can I identify the topic of a news post?”*
4.	Extracting Sentiments from Product Reviews	*“How can I extract the customer’s sentiment?”*
5.	Recommending Music Titles	*“How can I recommend music titles to users?”*
6.	Teaching Machines to Translate	*“How can I translate a piece of text?”*
7.	Summarizing Wikipedia Articles	*“How can I summarize Wikipedia articles?”*
8.	Detecting Hateful and Offensive Language	*“How can I detect inappropriate language?”*
9.	Generating Text in Chatbots	*“How can I create a chatbot?”*
10.	Clustering Speech-to-Text Transcriptions	*“How can I cluster similar transcriptions?”*

Machine Learning Techniques for Text

Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

Let’s practice!

- *Google* ***Colab***, short for Colaboratory, is a free, cloud-based platform provided by Google for writing and executing Python code
- Widely used in education for teaching and learning purposes
- However, resources provided in the platform are not unlimited

https://colab.research.google.com/

Machine Learning Techniques for Text

Questions?