

生成AI技術関連イメージ

自然言語処理 (NLP)

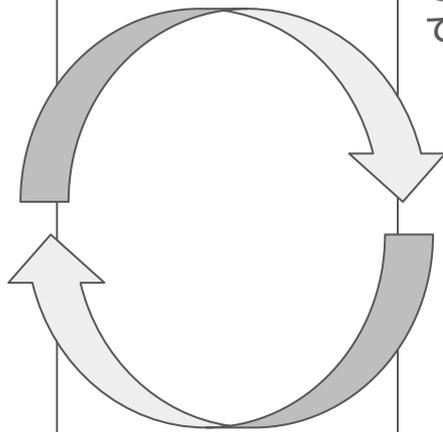
●準備:

コーパス (言語全集)
テキストデータセット
辞書
アノテーション済みデータ
等

●形態素解析:

- ・構文解析 (形態素に切分け等)
- ・依存構造解析
 - ↳ 構文解析器 (パーサー)
- ・句構造解析
- ・意味解析
 - ↳ 手法: グラウンディング, アノテーション等
- ・文脈解析

機械が学習するために必要な前処理としての自然言語処理



精度の向上

形態素解析においてディープラーニングの使用により従来のルールベース以上の成果。但し膨大な計算リソースが必要。

機械学習

データを基にしてモデルを学習させる手法全般

①ディープラーニング (深層学習) は機械学習の中でも特に強力な手法

・多層のニューラルネットワーク

CNN (畳み込みニューラルネットワーク) 画像系

RNN (再帰型ニューラルネットワーク) 言語系

GAN (敵対的生成ネットワーク) 偽vs判定

・トランスフォーマーアーキテクチャ: NLPタスク、文脈

・膨大なデータ処理 (コンピュータリソースが最も必要)

↳ 高性能GPUが必要 (NVIDIA)

②教師あり、教師なし、強化学習は学習のアプローチの種類

③LLM (大規模言語モデル) はこれらの技術を組み合わせて開発・改良される

⇒BERT、GPTモデル等

生成AI技術関連イメージ

ベン図版

自然言語処理 (NLP)

NLP (Natural Language Processing) はコンピュータが人間の言語を理解し、処理し、生成する技術や手法の総称。

- 準備：
 - コーパス (言語全集)、
 - テキストデータセット、辞書、
 - アノテーション済みデータ等
- 形態素解析：
 - ・構文解析 (形態素に切分け等)
 - ・依存構造解析
 - └ 構文解析器 (パーサー)
 - ・句構造解析
 - ・意味解析
 - └ 手法：グラウンディング, アノテーション等
 - ・文脈解析

機械学習

データを基にしてモデルを学習させる手法全般

ディープラーニング (深層学習)

- ・多層のニューラルネットワーク
 - CNN (畳み込みニューラルネットワーク) 画像系
 - RNN (再帰型ニューラルネットワーク) 言語系
 - GAN (敵対的生成ネットワーク) 偽vs判定
- ・トランスフォーマーアーキテクチャ：NLPタスク、文脈

開発

LLM

(Large Language Model)

機械学習で訓練されパターンを学習し、新しいテキストを生成したり、文脈を理解したりする能力を獲得した大規模言語モデルでGPTやBERTが代表例。

テキスト生成、質問応答、翻訳など、特定のタスクに応じたNLPタスク (自然言語処理) を行うことができる。