

教育

イリノイ大学アーバナ・シャンペーン校 (UIUC)

統計学修士; GPA: 3.84/4.00

キナン大学 / バーミンガム大学

経済学学士 GPA: 3.93/4.25

応用数学と統計学の理学士（一級学位）

アーバナ・シャンペーン, イリノイ州, アメリカ

2021年8月 - 2023年5月

広州, 中国/バーミンガム, イギリス

2017年9月 - 2021年6月

スキル

機械学習研究: 人工知能、ディープラーニング、自然言語処理、統計学習

データサイエンス: データサイエンスプログラミング手法、統計データ管理、統計コンサルティング

プログラミング: Python, R, SQL, コンピューターサイエンス

技術: Git, AWS, Shell, PyTorch, TensorFlow, NumPy, Pandas, Matplotlib, ggplot2, Tidyverse, Gradio, Docker

職歴

人工知能 / 機械学習エンジニア

Git, Shell, RunPod, LLM, GenAI

フリーランス

2023年7月 - 現在

- LLaMA モデルの包括的なテストを実施し、llama.cpp ツールキットを使用して NVIDIA の GPU と Apple Silicon を比較し、パフォーマンス評価のための革新的なクラウドベース戦略を適応させ、デバイス上でのローカル大規模言語モデル (LLM) の概念実証を確立しました。これにより、業界からの賞賛を受け、GitHub で 600 以上のスターを獲得しました。
- 様々なモデルサイズの推論速度を文書化・分析し、効率的な LLM 推論のための最適な設定に関する実用的な洞察を提供しました。

大学院コースアシスタント

Python, R, SQL, ディープラーニング, 統計学習

イリノイ大学アーバナ・シャンペーン校

2022年9月 - 2022年12月

- 教員と協力して、R、Python、SQL で魅力的な機械学習のコース内容と革新的なコーディング評価を設計しました。週に一度のオフィスアワーで学生の成績を評価し、学問的な問題を解決しました。コースに関する苦情は一切ありませんでした。

プロジェクト

ニューラル機械翻訳: RNN およびトランスフォーマーモデル | Python, PyTorch

2023年5月

- PyTorch を使用してスペイン語から英語へのテキスト翻訳に焦点を当てたニューラル機械翻訳 (NMT) システムの開発をリードしました。
- GRU とアテンションメカニズムを使用して RNN を設計・実装し、エンコーダーおよびデコーダーコンポーネントを構築しました。トランスフォーマーモデルの翻訳品質を向上させるために、マルチヘッドアテンションと位置エンコーディングを使用しました。
- RNN モデルの BLEU-4 スコア 0.058、トランスフォーマーモデルの BLEU-4 スコア 0.059 を達成しました。

スネークゲームの強化学習 | Python, Numpy

2022年12月

- ステート、アクション、リワードのマルコフ決定プロセス (MDP) フレームワークを定義し、時間差 (TD) Q 学習アルゴリズムを使用して AI エージェントがスネークゲームをマスターするようにしました。
- 新しい状態を探索することと既知の状態を利用することの間にバランスをとる探索ポリシーを実装しました。時間の経過とともに Q 値更新プロセスを最適化するために減衰学習率を使用しました。
- モデルの堅牢性と適応性を確保するために厳格なテストフェーズを通じてエージェントのパフォーマンスを分析しました。

アマゾンレビューおよび薬データセットの感情分析 | R markdown, Word2Vec

2022年12月

- Amazon および Drug Review データセットに対する感情分析を進め、四つのクラシックな埋め込みおよび NLP 方法 (BoW, Word2Vec, GloVe, fastText) を比較しました。FastText が最高のパフォーマンスを達成しました (Amazon での正確性 86.49%、Drug での正確性 78.69%)。
- テキストの正規化およびベクトル化のためのデータ前処理技術を使用しました。一般的な単語をフィルタリングするために 2 つの単純な t 検定を使用しました。分類のためにナイーブベイズおよびランダムフォレストアルゴリズムを活用しました。

CNN を用いたデータ拡張の効果の説明 | Python, PyTorch, Matplotlib, Gradio

2022年5月

- さまざまなシードで三つの異なる畳み込みニューラルネットワーク (CNN) モデルを開発・評価し、モデル性能への拡張技術の影響を特定することに焦点を当てました。二項の猫犬分類チャレンジに焦点を当てました。
- 正則化 (ドロップアウト、L1、L2) を適用して過学習を避け、二項クロスエントロピーをコスト関数として、SGD を最適化アルゴリズムとして使用しました。データ拡張を反転させることで、最高の CNN モデルの正確性は 59.6% から 74.8% に向上しました。