

ニューラルネットワークのトレーニングとプロファイリング のためのパラメトリック難易度を持つデータの生成

情報科学研究科
情報工学専攻

准教授 **ルカック マーティン** Lukac Martin

教員情報



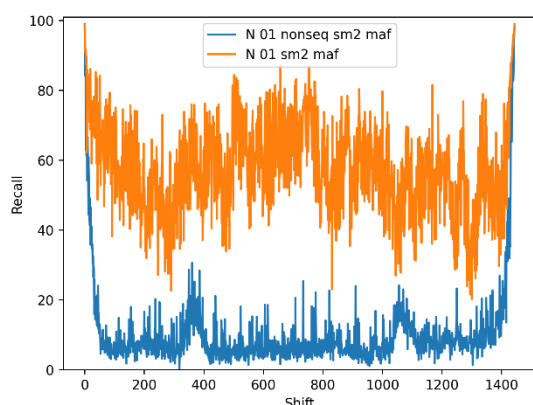
キーワード

機械学習のためのデータ生成、パラメータ化された難易度

研究シーズの概要

既存の標準的なニューラルモデルを一律に評価することは容易ではありません。なぜなら、使用されるデータセットは、特定の属性の分布など、部分的に制御可能なパラメータで構成されているからです。本研究では、特徴セットの代数的性質に基づいて、難易度を制御可能なデータセットを生成する方法を提案します。これらの特徴セットを操作することで、所望の難易度のデータセットを作成できます。

研究シーズの詳細



学習および汎化の難易度が特定されたデータセットを生成する提案手法は、関数の代数的な性質に基づいています。複数の入力に対し、線形従属・独立性に基づいてサンプルをグループ化し、グループシフトを行います。例として、2 入力の 3 値関数の集合を取り上げます。具体的には、ベント関数をそれ以外のすべての関数から区別する分類問題、および対称関数をそれ以外のすべての関数から区別する分類問題について検討します。一連の関数を取り上げ、NPN 分類に基づいてグループ化し、それらの NPN グループ間でシフトを行います。これにより、難易度の異なるデータセットが生成されます。

基本的な考え方は、この「シフト」のプロセスが定式化され理解された段階で、画像処理やテキスト処理といった対象データセットに対してこの手法を適用するというものです。その目的は、意味情報の学習の難易度を、論理関数の性質によって推定できるかどうかを明らかにすることにあります。

図に実験例を示します。これは、NPN グループに応じて関数が変化する様子を示しています。Zltas 曲線は、グループが変化するにつれて分類精度がどのように変化するかを示しています。

応用例・セールスポイント

- ◆ LLM のテスト用データを作成する
- ◆ 手でデータを準備することなく、学習済みモデルのプロファイリングを行う
- ◆ 自然データにおける特定の困難を学習するためのデータ準備

【問い合わせ先】 広島市立大学 地域共創センター 〒731-3194 広島市安佐南区大塚東 3-4-1 (情報科学部棟別館 1F)
TEL:082-830-1764 / FAX:082-830-1555 / E-mail:ken-san@m.hiroshima-cu.ac.jp