# Integrated Gradients を用いたガウス過程の解釈

氏　　名　張　凡

研 究 室 名　計算システムズ生物学研究室

主指導教員名（論文博士の場合は推薦教員名）　金谷重彦

内 容 梗 概（1ページ目に収めること）

ガウス過程回帰はノンパラメトリックな確率モデルであり、予測値だけでなく、予測の信頼度を表す予測標準偏差をも計算することができる。また、カーネル関数を設計することで自由自在に非線形化でき、尤度関数を変えることで外れ値に頑健なモデルにしたり、分類モデルに拡張したりでき、非常に柔軟性の高いアルゴリズムということができる。このため、ガウス過程は材料科学や創薬研究でよく用いられている。さらに近年では、深層学習とガウス過程回帰を組み合わせた深層カーネルガウス過程やガウス過程回帰を複数組み合わせた深層ガウス過程なども提案されており、これらはガウス過程回帰より高い精度でモデリングできることが報告されている。

しかしながら、ガウス過程回帰はノンパラメトリックであるため、解釈するのが難しい。近年では、様々な観点から解釈可能な AI が求められている。解釈可能性にはグローバルな解釈性とローカルな解釈性がある。ガウス過程回帰をグローバルに解釈する方法としてカーネル関数のスケールパラメータや ARD カーネルの length パラメータを比較する方法などがある。しかし、ARD カーネルは目的変数と線形な関係にある説明変数を重要でないと判断してしまうことを指摘し、感度分析を用いて特徴量重要度を計算する手法を提案している。一方で近年、モデルに依存しないローカルな解釈を行う手法が数多く提案されている。LIME や kernel SHAP のような解釈可能な代理モデルを用いて解釈する手法があるが、これらは乱数によって結果が変化することや、高次元データに対しては計算コストが高くなる欠点がある。他にも勾配を用いて解釈する感度分析がある。しかし、感度分析は Sensitivity と Implementation invariance という２つの公理を満たさないことが示されており、これらの公理を満たす勾配を用いた手法として、Integrated Gradients が提案され、深層学習モデルを解釈する際によく用いられる。しかし、これらの手法をそのまま使用しただけではガウス過程の予測値は解釈できるが、予測標準偏差を解釈することは難しい。GPR と同等で解釈可能な確率モデル GPX も開発されているが、GPX では深層学習などと組み合わせてしまうと解釈するのが難しくなってしまう。

そこで、本研究では GPR はもちろん、DKLGPR や GPC なども解釈できる Integraed Gradinets ベース手法を開発したので報告する。