F-004

並列計算による機械学習の説明手法の高速化の検討
Fast Model-Agnostic Explanations with Parallel Computing

○浅野 孝平 *　　全 眞嬉 *
Kohei Asano　　Jinhee Chun

1 はじめに

近年，深層学習をはじめとする高い分類性能をもつ機械学習モデルが様々な分野に応用されている．それらのモデルの多くはブラックボックスであり，ユーザがモデルの挙動や予測の原因について知ることが困難になっており，モデルや予測結果に解釈性に関する研究が活発に行われている．予測の原因となる特徴を特定する手法として Local Interpretable Model-agnostic Explanations (LIME) がある．LIME では，複雑なモデルを局所的に解釈可能なモデルで近似することで，モデルが分類の根拠とする特徴を予測する．

本研究では，並列計算を用いた LIME の高速化手法を提案する．本手法では近似モデルの導出に用いる訓練データの生成を並列化し，高速化を実現する．そして，計算機実験によって提案手法の有効性を検証する．

2 LIME

本研究では，Ribeiro et. al. が提案した線形モデルを用いた定式化 (Linear-LIME) [1] と Asano et. al. が提案した極小パターンを用いた定式化 (Minimal Patterns LIME: MP-LIME) [2] の高速化を検討する．

2.1 Linear LIME

分類モデル $f : \mathcal{X} \to \mathbb{R}$ から得られた，インスタンス $x \in \mathcal{X}$ の予測結果 $f(x)$ に対して説明を与えることを考える．ここで，x を被説明インスタンスと呼び，$\mathcal{X}$ は x のドメインである．LIME では，はじめに被説明インスタンス x をバイナリベクトル $\mathbf{x} \in \{0,1\}^d$ に変換する．

LIME ではスパースな線形関数 $g(x) = \mathbf{w}^\top \mathbf{x}$ によって，f を x の近傍で局所的に近似する．重みベクトル $\mathbf{w}$ から，分類に大きな影響を及ぼす $\mathbf{x}$ の特徴が特定できるため，予測結果に説明を与えられる．g のように，予測結果に説明を与えるモデルを説明モデルと呼ぶ．

次に，g の生成法について述べる．はじめに $\mathbf{x}$ の近傍のデータを $\mathbf{x}$ 中の 1 となっている特徴量をランダムに 0 にすることでサンプルし，これをサンプルベクトルと呼ぶ．次に，N 個のサンプルベクトル $\mathbf{x}_k$ $(k = 1, \ldots, N)$ を生成し，それらを元のデータ表現 $x_i \in \mathcal{X}$ に戻し，$\{(\mathbf{x}_k,\, \pi(x_k) f(x_k)) : k = 1, \ldots, N\}$ を訓練データとして Lasso 回帰することで，g を導出する．ここで，π は類似度関数であり，これを用いて局所性を測る．

分類モデル f の分類コストを τ とおくと，Linear-LIME の説明にかかる計算コストは $\mathcal{O}(N\tau + Nd^2 + d^3)$ となる．

2.2 Minimal Patterns LIME

MP-LIME では $\mathbf{x}$ の非零の要素を特徴の集合とみなし，$[d] = \{1, \ldots, d\}$ と記す．また，特徴の部分集合 $e \in 2^{[d]}$ を特徴パターンと呼び，ある特徴パターン e に対応するインスタンスを x_e と記す．MP-LIME では説明モデル $\mathcal{E}_{\min}$ を極小な特徴パターンの集合として定式化する．極小な特徴パターンを定義 1 として定める．

定義 1

$$f(x_{e_{\min}}) \sim f(x) \tag{1}$$

$$\forall i \in e_{\min},\ f(x_{e_{\min} \setminus \{i\}}) \nsim f(x) \tag{2}$$

を満たす $e_{\min}$ を極小な特徴パターンとして定義する．ここで，$\sim$ は分類モデル f が x と x_e を同じクラスに分類することを表している．

定義 1 では，極小な特徴パターンを被説明インスタンス x と同じクラスに属するために必要最低限な特徴の集合として定義している．この定式化では，説明モデルに含まれるパターンは必ず被説明インスタンス x と同じクラスに属することが保証されており，分類に重要な特徴の組み合わせを表現できると考えられる．

次に，極小な特徴パターンの探索アルゴリズムについて述べる．探索はモデル f の評価値に基づいた山登り探索によって行う．はじめに $[d]$ を初期状態として，$[d]$ の近傍の特徴パターンすなわち，全ての $i \in [d]$ について $f(x_{[d] \setminus \{i\}})$ を評価する．このとき，最も評価値の高いパターンを次の状態として，同様に近傍のパターンを評価して，極小な特徴パターンを探索する．この

*東北大学大学院 情報科学研究科

探索法で複数の極小な特徴パターンを発見するためには，再度繰り返し探索が必要となる．そこで，一度評価したパターン e のうち $f(x) \sim f(x_e)$ を満たすパターンは極小パターンの候補として $\mathcal{E}_{\mathrm{cand}}$ に保存し，パターンを評価するたびに更新する．ひとつの極小な特徴パターンを発見したら，

$$\forall e_{\min} \in \mathcal{E}_{\min}, e_{\min} \subseteq e \tag{3}$$

を満たすパターン e を $\mathcal{E}_{\mathrm{cand}}$ から削除し，縮約された $\mathcal{E}_{\mathrm{cand}}$ の要素の中から最も評価値の高いパターンを次の状態として再度探索を行う．式 (3) の制約を加えて探索を行うことで，評価されるパターン候補を削減する．

説明モデルの構築において，全ての極小な特徴パターンの列挙すると，評価されうるパターンの組み合わせは 2^d 個あるため，計算時間が膨大になりうる．そこで，極小な特徴パターンを L 個の発見するまで探索を行うことによって，この問題を回避する．ひとつの極小な特徴パターンを発見するために必要な評価回数は $\mathcal{O}(d^2)$ なので，MP-LIME の説明にかかる計算コストは $\mathcal{O}(Ld^2\tau)$ となる．

3 訓練データ生成の並列化

Linear-LIME, MP-LIME では，説明の生成に膨大な近傍データを訓練データとして生成する必要がある．生成の際に，近傍インスタンスの生成は容易であるが，ラベルや評価値の計算は分類モデル f の分類コスト τ に依存する．分類コスト τ が大きい場合，訓練データの生成にかかるコストは大きくなる．各訓練データの生成は独立しているため，並列化によって説明の高速化が期待できる．

4 計算機実験

訓練データ生成の並列化の効果を，Linear, MP-LIME のインスタンスあたりの説明の計算時間を計測することで評価した．本実験ではデータセットとして，adult データ[1]を用いた．そして，adult データを学習した 3 層パーセプトロンを分類モデル f として用いた．計算時間の計測は 100 サンプルのテストデータで行い，計算時間の平均値を各計測値とした．

adult データは 8 つのカテゴリカル変数と 4 つの数値変数からなっている．カテゴリカル変数は One-hot ベクトル表現し，数値データは区間化することで，バイナリベクトル表現とした．また，LIME のパラメータは，サンプルデータ数を $N = 4000$ とし，MP-LIME では，$L = 4$ とした．

表 1: インスタンス当たりの計算時間 [sec]

手法＼並列数	1	2	4	8
Linear-LIME	42.5	26.5	15.7	13.4
MP-LIME	0.721	0.710	0.730	0.716

実験結果を表 1 に示す．これより，Linear-LIME では並列化することで，8 コアで並列化することで，およそ 3 倍の高速化が確認できる．MP-LIME では，並列化によって高速化されているものの，並列数を増やすことによる効果は確認できなかった．しかしながら，計算時間は全ての条件で MP-LIME が Linear-LIME よりも高速である．これは，MP-LIME の説明の導出のために分類モデル f の評価回数が Linear-LIME よりも少ないことが原因である．Linear-LIME では訓練データ生成のため分類モデル f を N 回呼び出す必要があるが，MP-LIME では L 個の極小パターンを発見したら探索をやめるため，評価回数が少なくなる傾向がある．また，MP-LIME は近傍のパターンを評価し探索を行うため，並列化可能な評価数はたかだか $d-1$ である．adult データは $d = 12$ であり，d が小さいため, 並列化の効果が少なくなったと考えられる．

5 まとめ・今後の課題

本研究では，並列計算による LIME の説明導出の高速化について検討した．Linear-LIME では，訓練データの生成を並列化することで，高速化が実現できた．しかしながら，MP-LIME では並列化による大幅な高速化は実現できなかった．実験より，分類モデルの呼び出し回数が計算時間を左右することがわかったため，モデルの呼び出し回数を削減することで，高速化が期待できアルゴリズムの改良や，他のデータを用いた実験評価が今後の課題となる．

参考文献

[1] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.

[2] Kohei Asano, Jinhee Chun, Atsushi Koike, and Takeshi Tokuyama. Model-agnostic explanations for decisions using minimal patterns. In *Proceedings of the 28th International Conference on Artificial Neural Netwroks*, 2019.

[1] https://archive.ics.uci.edu/ml/machine-learning-databases/adult/