

時系列分類モデルに対する部分系列での反実仮想説明

奥村裕幸[†]長尾智晴[‡][†]横浜国立大学 大学院環境情報学府[‡]横浜国立大学 大学院環境情報研究院

1 はじめに

深層学習モデルを用いた1次元時系列データの分類精度が高く、実応用が期待されている。例えば[1]では、心電図の波形データから心筋梗塞の傾向がみられる波形を深層学習モデルによって分類している。心電図のデータに限らず、実応用時にモデルの分類結果を信頼して活用するためには、分類精度だけでなくモデルの判断根拠を提示する必要がある。

時系列データは部分系列が分類に寄与するといわれており[2]，これに基づいた説明を提示することがエンドユーザへの理解に繋がる。また，部分系列は同一の時系列データ内に複数存在するため[1]，これを同時に分類根拠として提示する必要があると考えられる。

本稿では，モデル内部に現れる特徴量から，分類への寄与度に基づいて部分系列(パッチ)を取得し，各パッチの分類結果が別のクラスに変化していく過程を生成する説明手法を提案する。本手法と同様の「反実仮想説明」と呼ばれる従来手法との比較を行い，有効性を検証する。

2 反実仮想説明に求められる性質

反実仮想説明とは，あるテストインスタンス(Query; Q)に対するモデルの予測結果を変えたいとき，Qをどのように変更すればよいかを示す説明方法である。この説明インスタンスをExplanation(E)とすると，QとEが類似していることが望まれる。これを近接性と呼ぶ。また，Eが実際のデータ分布に従っている必要がある。これを妥当性と呼ぶ。反実仮想説明では，この2つの性質を同時に満たすことが求められる。

3 提案手法

提案手法は，パッチの取得と反実仮想説明の生成に分けられる。概要を図1に示す。本手法の分類モデルは，[1]でも用いられている Fully Convolutional Networks(FCN)を想定している。

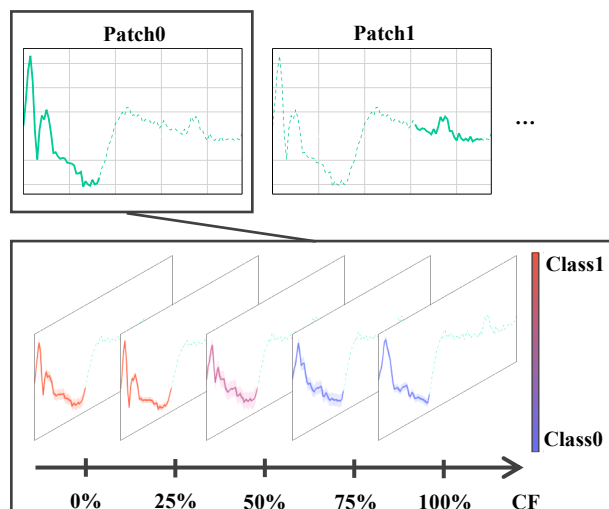


図1:パッチの取得と反実仮想への連続変化

(上)Q をパッチに分割したときの様子。(下)0 番目のパッチに対して Q を反実仮想(CF)へ連続的に変化させたときの様子。x 軸は反実仮想への変化率を表している。

3.1 パッチの取得

FCN の畳み込み後の特徴量と全結合層の重みを用いて，Class Activation Mapping(CAM)を取得する。CAM に対して Gaussian Mixture Model (GMM) のフィッティングを行うことで，特徴量を時間方向のクラスターに分ける。クラスタリング結果と FCN の畳み込みの受容野から，Q をパッチに分割する。同時に CAM から各パッチの分類に対する寄与度を取得する。

3.2 反実仮想説明の生成

Q の分類に寄与するパッチに対して，近接性・妥当性を満たす少量の学習データのパッチを取得する。これらのパッチを用いて Gaussian Process Latent Variable Model(GPLVM) [3] を学習する。GPLVM の潜在変数を変化させることで，分類結果が別のクラスに連続的に変化していく過程を生成することができる。この様子を図1(下)に示す。反実仮想への変化率 100%の波形を反実仮想説明とする。

Time Series Explanations with Plausible Patch-Level Counterfactual Instances

[†]Hiroiyuki Okumura [‡]Tomoharu Nagao

[†]Graduate School of Environment and Information Sciences, Yokohama National University

[‡] Faculty of Environment and Information Sciences, Yokohama National University

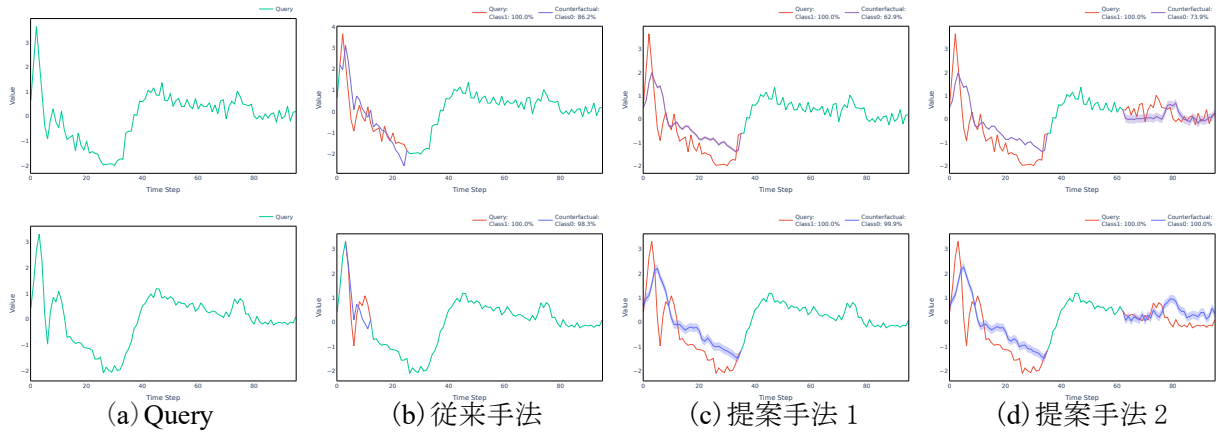


図 2: 類似した 2 つの Query(上段と下段. ともに Class1)に対する各手法の反実仮想説明
青は反実仮想説明(Class0), 赤は Query の変更箇所を示している。

表 1: 近接性・妥当性・代替性の評価

	近接性(L1)		妥当性(IM1)		代替性	
	#1	#2	#1	#2	#1	#2
(b)	0.2	0.27	1.22	1.03	0.2	0.32
(c)	0.66	1.4	0.63	0.62	0.56	0.9
(d)	1.01	1.5	0.57	0.39	0.66	0.93

4 実験設定

UCR Archive の 2 クラス分類データセットで、図 2 で示した ECG200(#1)と GunPoint(#2)を使用し、従来手法[1]との比較を行った。FCN のパラメータは[1]を参考にして設定した。評価指標として、近接性では L1 距離から算出される指標、妥当性では IM1 と呼ばれる学習済み Autoencoder による再構築誤差から算出される指標を用いた。この 2 指標は値が小さいほど良いとされる。それに加えて、説明データを用いて学習した k 近傍法の分類精度から算出される評価指標を用いた。この指標は代替性と呼ばれ、反実仮想説明が分類に寄与する波形特徴を取得できているかを評価することができる。

5 実験結果

図 2 に同一クラスの類似した Q に対する各手法の説明を示す。提案手法では波形全体の分類結果が変更する最小数のパッチでの反実仮想説明(c に該当)、分類への寄与度がしきい値を超えるパッチでの反実仮想説明(d に該当)の結果を示した。この図から 2 つの類似した Q に対して、従来手法(b に該当)の説明は異なっていることが分かる。それに対して提案手法では類似した一貫性のある説明が得られている。

表 1 に近接性・妥当性・代替性の評価結果を示す。近接性の観点では、従来手法が優れている

ことが分かる。パッチの中の一部を変更することに該当する従来手法と比べて、提案手法ではパッチ全体を変更するためである。しかしこれは、妥当性の観点ではデータ分布に従った説明を得ることに繋がっている。実際に提案手法では(c), (d)ともに従来手法よりも良い結果を確認した。また代替性の観点では、従来手法よりも k 近傍法の分類精度が高くなることが確認された。このことから、FCN が学習した分類に寄与する波形特徴を説明によって取得できていることが分かる。同時に反実仮想説明による Data Augmentation への応用可能性を示唆している。

6 まとめ

本稿では時系列分類モデルに対して部分系列ごとに反実仮想説明を生成する手法を提案し、近接性・妥当性・代替性の観点から有効性を検証した。今後、時系列データの専門家の評価を基に、提案手法における分かりやすい説明の提示方法について検討する。また提案手法を用いた効果的な Data Augmentation についても検討を行う。

参考文献

- [1] Delaney, Eoin, Derek Greene, and Mark T. Keane. "Instance-based counterfactual explanations for time series classification." International Conference on Case-Based Reasoning. Springer, Cham, 2021.
- [2] Ye, Lexiang, and Eamonn Keogh. "Time series shapelets: a novel technique that allows accurate, interpretable and fast classification." Data mining and knowledge discovery 22.1 (2011): 149-182.
- [3] Lawrence, Neil. "Gaussian process latent variable models for visualisation of high dimensional data." Advances in neural information processing systems 16 (2003).