

AIで予想をするだけでなく説明もしよう。

概要

機械学習による予測を利用する場合に、なぜその予測をしたのか説明を求められる場合がある。

例えば、外の気温と湿度、室内の気温の学習データがある場合に外の気温と湿度から重回帰分析で室内の温度を予想するケースを考える。外の気温と湿度の寄与率を見た場合に、外の気温の寄与のほうが湿度のそれより大きいといったことを予測と一緒に確認したい。

機械学習のモデルは複雑で予測が当たらない原因に対して問題の切り分けが難しい。完璧な説明をせずとも予測のパフォーマンスや寄与の大きい部分を表示できれば、問題の切り分けのための説明に利用できる可能性がある。

先の例であれば直観に反して湿度の寄与率が外の気温より大きいなら予測モデルや学習データについて詳細にみて振り返りを行うなどして適切な予測モデルと学習データなのか確認をするために利用する。

本パターンは予測と同時に予測への寄与の大きい属性や予測の説明のためのデータを出力し、予測と同時にその説明を行うものである。

パターンの説明

1. 名前

- a. AIで予想をするだけでなく説明もしよう。

2. 文脈

- a. 機械学習において学習データと予測のためのモデルがある。作成したモデルをつかって予測するが、予測結果の妥当性の確認や予測が当たらない場合の解析のため予測の根拠となったものに対して説明が必要になる。

3. 問題

- a. DNNなど学習のためのパラメータが膨大で、予測が当たらない場合に、原因の説明や問題の切り分けが難しい機械学習のアルゴリズムがある。
- b. 機械学習はプログラムのバグ、汚れた学習データ（ノイズが多いデータなどが考えられる）、不適切なモデル（線形な関係でないものを線形予測するなど）、時間の経過による傾向の変化（株価の予測であれば上場前と上場後で傾向がまったく違って来るなど）などさまざまな要因で適切な予測ができない場合があり予測根拠の確認をし、問題を切り分けたい。

4. フォース

- a. 複雑なモデルではなく線形回帰や決定木など簡単に説明のしやすいモデルがある。
- b. 詳細に説明せずとも予測にもっとも寄与したデータの属性や画像の部位がどれなのかをみることがある。

5. 解決

- a. 予測だけでなく、データやモデルの予測に大きく寄与した部分・属性、閾値や予測のパフォーマンスを確認・可視化する。
- b. 予測とともに説明のためデータを出すシステムにする。

6. 適用例

- a. 外の気温と湿度、室内の気温の学習データがある場合に外の気温と湿度から重回帰分析で室内の温度を予想するケースを考える。
予測する室内の温度だけでなく、MLのソフトは外の気温と湿度の寄与率同時に出力する。
仮に梅雨の時期の学習データだけがあり、外の湿度の寄与が外の気温の寄与より高いケースがあったとする。
すると夏や冬の湿度が梅雨より低い季節ではその予想が当たらなくなる場合が考えられる。
そのような場合に寄与率を確認し、湿度の寄与が気温の寄与より大きいならなら他の季節の学習データの収集と再学習の必要性が示唆される。
- b. WEBサーバーの台数を実際に高負荷になる前に増加させる要求がある。
ゲームのイベントの開催前、日内変動などさまざまな要因でサーバーへのアクセス頻度が変わる。これから起こるイベントや時間によってアクセス頻度の予測をし、サーバーの台数を事前に増加させる。
MLのソフトは予測だけでなく、アクセス頻度の増加の予測の根拠（イベントによるものか、日内変動によるものか）を提示する。
仮にゲームのイベントが発生していないが、イベントの発生の予測が失敗したことによる突然WEBサーバーの台数が増加する場合に、予測の根拠がなければイベントが発生の予測の失敗か日内変動によるものなのか判断ができないため、なにを改善すると予測がうまくいくのか切り分けができない。
- c. 顔画像認識で男か女か判別するケースを考える。MLのソフトは男か女か判別の予想をするだけでなく、予想に重要な部位を強調する感度マップ (sensitivity map) をユーザーに提示する。
仮に使用した学習データに問題があり、写真の背景色が男か女かで違う状況で、感度マップがあれば学習に背景色を使っているというのがわかれば、学数データの見直しの判断に利用できる。
参考資料 : <https://arxiv.org/abs/1706.03825>

7. 結果

- a. 予測の寄与の大きい属性や画像の部位を確認することで予測結果を確認し、予測の当たらない直接の原因の説明をし、学習モデルの改善を行う。