

施策デザインのための 機械学習入門1

廣田雄亮

はじめに

機械学習の実践における苦悩 1/2

現在，機械学習はあらゆるビジネスで活用されている

- コンテンツ推薦（例：Netflix, YouTubeなど）
- 広告のクリック確率予測

NETFLIX

You Tube



しかし,

**「機械学習を駆使しているにもかかわらず,
ビジネス上有益な変化をもたらすことができない」**

といった現実がある.

機械学習の実践に潜む落とし穴

正しい手順に則って機械学習の導入を進めたが、
望んだ結果を（継続して）得られない場合、

問題は

「機械学習に解く意味がない問題押し付けている」

原因として、

機械学習に解かせる問題の誤設定

他の落とし穴

「実環境と観測データの乖離（バイアス）の影響」

機械学習に解かせる問題の誤設定

具体例

当初の問題：ユーザに対して広告を配信すべきか否か



「広告を配信した場合のクリック確率を予測し、予測値が一定以上のユーザには広告を配信する」という流れで解く.

この場合、機械学習は「クリック確率を予測する部分」を担当する.



機械学習に解かせる問題の誤設定

具体例

当初の問題：ユーザに対して広告を配信すべきか否か

元々の問題は広告を配信すべきか否かを**決める**問題
何かを**予測する**問題ではない

「広告を配信した場合のクリック確率を予測し、予測値が
⇒機械学習に意思決定の問題を解かせるべき
関係のない問題を解かせてはいけない

この場合、機械学習は「クリック確率を予測する部分」
を担当する。

機械学習はあくまで与えられた問題をうまく解くための
1つの方法論



実環境と観測データの乖離（バイアス）の影響

具体例

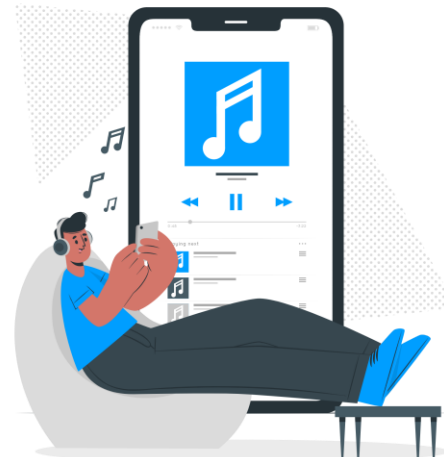
音楽のストリーミングサービスにおける推薦システム

人気アーティストの楽曲ばかりを多く推薦すると...



バイアスが発生

すでに人気のアーティストの楽曲は観測されやすい
まだ人気でないアーティストの楽曲は観測されにくい



実環境と観測データの乖離（バイアス）の影響

具体例

音楽のストリーミングサービスにおける推薦システム

人気アーティストの楽曲ばかりを多く推薦する

バイアスを無視



⇒ 人気アーティストの楽曲のみを推薦するシステム

すでに人気のアーティストの楽曲は観測されやすい

機械学習は「**観測データが実環境や予測対象をまんべんなく代表したデータである**という仮定に基づいた技術」である

アプローチ

機械学習を機能させる

機械学習を機能させるとは

**「機械学習の性能 (目的関数や評価指標の値) の改善が、
実環境で望ましい結果が得ることに繋がるための前提が
満たされている状況を導くこと」**

例

- 手元のデータがどのような経緯をたどって観測されたか

目標

目指す状態

「機械学習を機能させるために必要な手順を理解し、
それを**臨機応変に使えるようになる**」

具体的には、

「問題設定に合わせた機械学習の実践手順を
自分の意思で臨機応変に導ける状態」

構成

構成

1. 機械学習実践のためのフレームワーク
2. 機械学習実践のための基礎技術

1章では、実応用において機械学習を機能させるためにたどるべき手順をフレームワークとしてまとめている。無意識的にこなしていってしまう上流の手順を明文化している。