

数理手法Ⅳ 「統計的機械学習入門」

数理情報学専攻
中川裕志

機械学習小史：計算機以前

- ◆ 人間の知能に関する研究は言語を媒体とし、言語で古代から進んできた。
 - ◆ ソクラテスの以降の言語名称目録説（モノが名前を付けるという考え）
 - ◆ アリストテレスの論理学
 - ◆ アンダルシアの西方イスラムを経てパリに伝わる
- ◆ 近世、計算を知能と考え始めた。
 - ◆ ライプニッツ etc etc
 - ◆ 論理至上主義 ヴィトゲンシュタイン
- ◆ ソシュールの革命
 - ◆ 言語名称目録説から脱却し、言語を閉じたシステムとして、その構造を分析する方向を示した。（現代のテキスト処理の基礎）

井筒俊彦「イス
ラーム思想史」
アベロエスが有名

機械学習小史：計算機以後

◆ 計算機の発明：1940年代

- ◆ 計算の道具（弾道計算に使った）
- ◆ 言語を分析する道具：1950年代には既に文書要約のアルゴリズムが研究されていた。

◆ 人工知能の開花

- ◆ 数値を計算するマシンから記号計算をするマシンへ
- ◆ 人工知能は人間を模倣する方向で研究された。
- ◆ 言語処理、パタン認識、記号処理
- ◆ こういった背景の下に人間が知能や知識を獲得していく過程である学習を計算機で実現しようとした → 機械学習

機械学習小史：1980年代以前

- ◆ 人工知能の時代の機械学習
- ◆ 与えられたアルゴリズムを変換して効率の良いアルゴリズムを導出
 - ◆ 効率の悪いプログラムを一定の方法でプログラム変換するソフト技術として定着
- ◆ 溜め込んだ事実の集合から機能的推論で知識を獲得
 - ◆ 死ぬ(ソクラテス) & 人間(ソクラテス) & 死ぬ(アリストテレス) & 人間(アリストテレス) → 死ぬ(x) & 人間(x)
つまり、同じ属性を持つ定数を変数化して一般的知識とする。なんだか学習した感じがしない！
- ◆ 少数のタネ知識を組み合わせて知識ベースを自動構築
 - ◆ 数学の知識体系を人工知能で再構築しようという試み
 - ◆ Automatic Mathematician などというシステムが作られたが、実用的な結果には至らず。
- ◆ ユーザとのインタラクションで言語能力を計算機が獲得しようとする試み

機械学習小史：前世紀終盤

◆ 人工知能の時代の機械学習

◆ 全て失敗した

◆ → 当時(1970年代から80年代)の弱体な計算機では無理

◆ → そもそも方法論的に間違った目標設定？

◆ 計算機能力の飛躍的向上

◆ 90年代の計算機能力の向上：**ディスク容量、CPU性能、インターネット、対象となるデータ量の増大と多様化**がある種の臨界点に達した

◆ 人工知能の時代とは違った機械学習の出現

◆ いわゆるデータマイニングとして注目を集めた。

◆ スーパーの売り上げデータから、赤ちゃん用の粉ミルクを買う客は同時にスポーツ紙を買うという傾向を発見し、商品の並べ方を工夫

◆ Amazon など、ある本を買おうとすると、「この本を買った方は、こういう本も買っています」などとrecommendしてくる

機械学習小史：今世紀

- ◆ 大規模かつ多様なデータから、新規のデータの性質や振る舞いを予測する統計モデルを機械学習
 - ◆ 大規模データの処理→統計学の利用：統計的機械学習
 - ◆ 大規模なデータの背後にある数理的構造の認識と利用
 - ◆ P)背後に潜む数理的構造を事前知識として利用しない手はない。
 - ◆ A)取得したデータと事前知識を統合してより確かな数理モデルへ：事後知識
 - ◆ P)+A)を可能にする数理モデル→ベイズ統計
- ◆ ベイズ統計に基づく機械学習の応用
 - ◆ パターン認識(画像認識、音声認識)
 - ◆ 文書分類
 - ◆ スпамメールフィルタ、フィッシング対策ソフト
 - ◆ ユーザの行動に適応して進化するモデル
 - ◆ インターネットのrecommendation
 - ◆ データマイニング
 - ◆ 構造をもつデータベースからの知識マイニング
 - ◆ テキストマイニング
 - ◆ 半構造データからの知識獲得、評判分析

機械学習小史：近未来

- ◆ 環境問題、省エネ＋問題の大規模化
 - ◆ 大容量メモリで長時間の機械学習を行う場合は、消費エネルギーが大きい。
 - ◆ それでも、実用規模に十分なスケールでの学習は困難
 - ◆ 従来のバッチ学習は全データをメモリに乗せて繰り返し計算で学習した → 省エネ、実用性の点で苦しくなってきた
 - ◆ しかも、改善幅は1%とか0.1%....
- ◆ 実際の対象は、大規模データ、ストリームデータ
 - ◆ 小さなメモリで学習 → オンライン学習 → バッチ学習と同程度の精度を目指す
 - ◆ ストリームデータ(時系列)をone-pathで処理(小さなメモリ) → ストリームマイニング
- ◆ 教師データ作成の手間が膨大。
しかも、教師データはタスク依存的
 - ◆ 教師ではない大量の生データも利用 → 半教師あり学習

授業の内容

1. 導入
2. 数学のおさらい
3. Bayes推論
4. 線形回帰と識別
5. 学習データと性能予測
6. カーネル法とサポートベクターマシン
7. 最適化と学習アルゴリズム
8. オンライン学習
9. クラスタリング
10. モデル推定
11. マルコフ連鎖モンテカルロ法