

集合知プログラミング

第1章 集合知への招待

佐々木稔

はじめに

- 背景
 - Netflix
 - 貸出履歴を基にDVDを推薦 (Cinematch)
 - Google
 - Webサイト間のリンク関係によるランキング
- 集合知
 - 多くの人々から集めたデータ
 - 洗練されたアルゴリズムで新たな知見を得る

集合知とは何か？

- 集合知
 - 集団の行動、嗜好、アイデアを結び付け、新しい知見を生み出す
- 集合知の例
 - Wikipedia : 利用者が作るオンライン百科事典
 - Google : ページを好む人(リンク)の数で評価
- 今回の対象
 - 情報を知的に処理し、新たな知見を求める

機械学習とは何か？

- 機械学習
 - 与えられたデータの特徴を抽出
 - その特徴を利用して将来のデータを予測
- 機械学習アルゴリズム
 - 種類はさまざま
 - 決定木、ニューラルネットワーク

機械学習の限界(1)

- 弱点

- 手法により大量のパターンへの対応能力に差
- 与えられたデータの範囲を超えない
- 扱えないパターンは誤って解釈されやすい

- 扱う素性

- 単語やフレーズ
- 意味概念や文構造はあまり使われない
 - 性能改善の割には大きなコストが必要

機械学習の限界(2)

- 過学習
 - 少ない例を基に強い一般化を行う
 - すべての機械学習手法がこの問題を抱える
- 例：スパムフィルタリング規則の学習
 - スпамに多い単語、フレーズを学習
 - フレーズ “Online Subscription”
 - IEEE 論文誌の案内メールをスパムと判定

実生活における例

- Google : PageRank, AdWords, AdSense
- Amazon,Netflix: 購入履歴による商品推薦
- Pandora, Last.fm:
 - バンドや曲の評価による曲の推薦
- Hollywood Stock Exchange :
 - 人気により俳優、映画に株価を付ける
- eHarmony :
 - 参加者情報から相性の良い男女のペアを推薦

学習アルゴリズムの その他の使用(1)

- バイオテクノロジー
 - シーケンシング、スクリーニング
- クレジット詐欺発見
 - 不適切な取引を検知
- マシンビジョン
 - 侵入者の発見、自動車の特定、顔認識

学習アルゴリズムの その他の使用(2)

- 製品のマーケティング
 - 市場動向、流行予測
- サプライチェーン
 - 需給予測、店舗間輸送
- 株式市場の分析
 - 株価の予測
- 国防
 - 外交関係の分析