

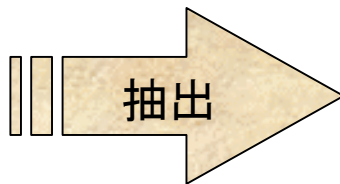
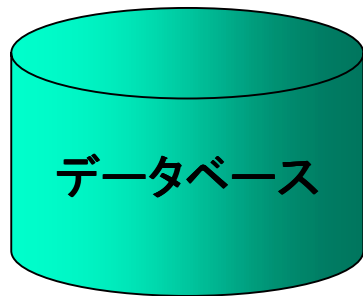


確率論を用いた類似性尺度による 事例分類に関する研究

山田泰大

世木研究室 M2

機械学習の流れ



未知の対象物



適用

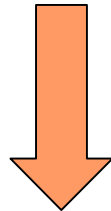


うれしいな



研究の流れ

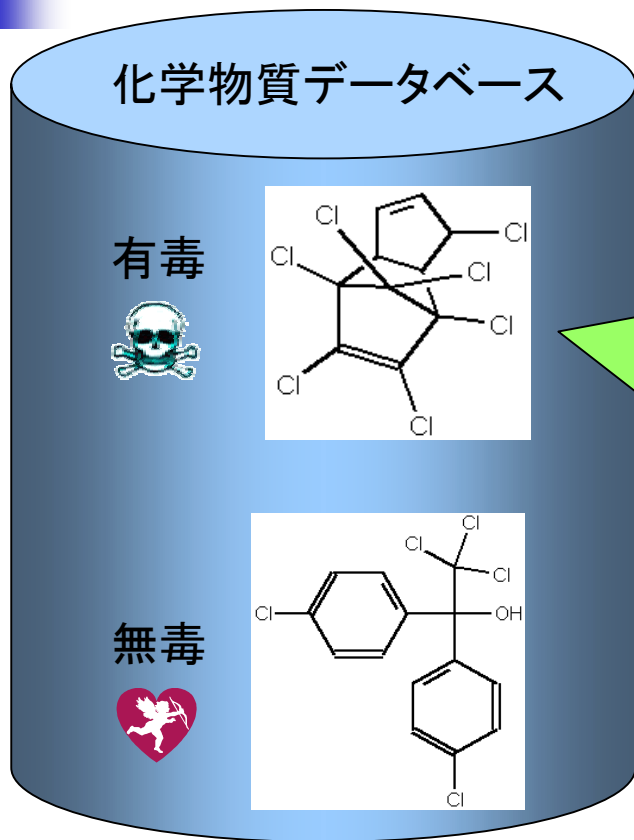
- 予測のためのアルゴリズムを考える
 - 従来法の改良
 - 新アルゴリズムの提案
- ベンチマーク用DBで実験を行う



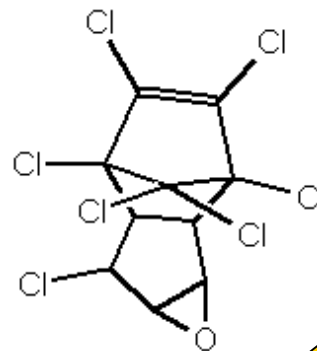
目標

- 理論的にきれいに説明でき
- 従来法を上回る予測精度

類似度に基づく分類予測



類似!



有毒? 無毒?

- 確率的な枠組みで類似度を計算



確率論を用いた類似性尺度

$$\text{sim}(x, y) = P(C_X = C_Y \mid X = x, Y = y)$$

: 2つの事例 X , Y が同じ分類に属する事後確率

- 確率を用いて類似度を定義
 - 理論的な取り扱いが可能に
 - データ形式(数値や記号etc)に非依存



おわりに



- 確率論を用いた類似性尺度を提案
- 機械学習用ベンチマーク・データベース
UCI repository上で実験
→提案法の有効性を確認
- 学会発表
ACI2002(国際)
情報科学技術フォーラム2002(国内)

