

Factorization Machinesにおけるスパース正則化を用いた組合せ特徴選択

Factorization Machines (FM) : 特徴の2次組合せ (交互作用) を用いるモデル

- 特徴の組合せ (交互作用) = 特徴の積 (例: x_1x_2, x_1x_3)
- 特徴の数に対して線形時間で評価可能, スパースなデータセットでも上手く学習可能

組合せ特徴選択 : 全ての組合せの中から**有用な組合せのみを自動で選択**する問題

- 通常のFMは**すべての2次の組合せ**を用いる
- **何が嬉しい?** : モデルの**解釈性**や**精度の向上**, **高度な知識発見**等が可能になる
- 既存の方法 : **組合せ特徴選択が実はできていない** (特徴選択をしている)

本研究の貢献 : FMにおいて**組合せ特徴選択を行う**方法を提案

- **アイデア** : モデルのパラメータではなく**パラメータから計算される組合せ重み行列がスパースになるように正則化**を行う. 提案した正則化は様々な良い性質を持つ
- 既存のFMのアルゴリズムと同様の計算時間で動作する**高速な学習アルゴリズム**も提案
- 人工データで正しく特徴選択ができることを確認・実データで予測に役立つ**組合せ特徴の選択**ができていることを確認