

## 01

## 膨大な情報の組合せから楽々学習

～大域的最適性を保証する低ランク回帰学習技術：CFM～

## どんな研究

複数の特徴量から目的値を予測する回帰問題を汎用的に、かつ高精度に解く研究です。提案法「Convex Factorization Machines」(CFM)は、データを表す特徴から予測に有用な組合せを見つけて高精度なモデルを作ります。提案法は、価格推定、推薦システム、遺伝子解析などの様々な応用例に適用できます。

## どこが凄い

提案法は、膨大な数の特徴の組合せを低ランク制約によって効率的に扱うことができます。さらに、初期値に依存せず大域的最適解を保証することができます。また、学習データに現れない未観測特徴の組合せの重要度(重み)も推定できるため、ユーザに新規の商品を推薦するなどの応用が可能です。

## 目指す未来

本技術を展開させ、より高次の特徴の組合せも考慮できれば、さらに予測精度の高いモデルを構築することができます。応用先として、推薦システムのほか、ゲノムデータから病気の有無や治療の効果を予測することで、その病気に影響する遺伝子の組合せの発見や新薬の開発技術などに利用可能です。

提案：重要な特徴の組合せを見つけて高精度な回帰モデルを学習する

## Convex Factorization Machines (CFM)

- 汎用的(様々な予測問題に適用可能)
- モデルのパラメタ学習が容易(解が初期値に非依存)
- 膨大な数の組合せを扱える(高次元データに適用可能)
- 未観測な特徴組合せの重要度を推定(新たな知識を発見可能)

$$\hat{y} = \underset{\text{特徴量}}{W^T} X + X^T \underset{\text{膨大な行列!}}{W} X$$

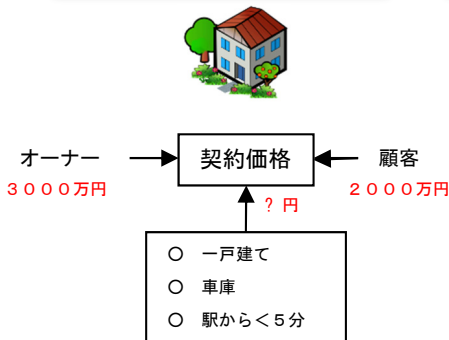
CFMは低ランクの制約によってWの大きさを抑制します

$$\begin{matrix} W_{1,1} & \dots & W_{1,d} \\ \vdots & & \vdots \\ W_{d,1} & \dots & W_{d,d} \end{matrix} = \lambda_1 \begin{matrix} p_1^T \\ \vdots \\ p_1 \end{matrix} + \dots + \lambda_k \begin{matrix} p_k^T \\ \vdots \\ p_k \end{matrix}$$

Wの固有ベクトルを固有値が大きい順で逐次的に求めていくため、効率的に計算できます

## 応用例 1

住宅の契約価格を予測



契約価格を決定する要因(条件の組合せ)を発見

## 応用例 2

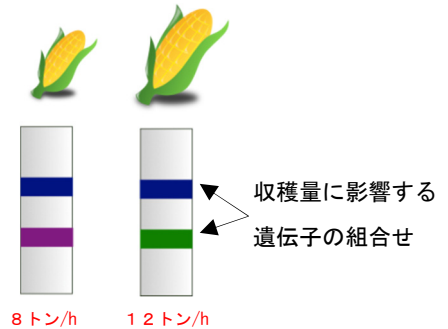
ユーザが好む書籍を推薦

Alice	☆	☆☆☆
Bob	☆☆☆	?
Charlie	?	☆☆

ユーザと書籍の組合せから評価点を予測することで推薦システムを実現

## 応用例 3

遺伝子から穀物収穫量を予測



トウモロコシの収穫量を遺伝子の組合せから予測!

## 【関連文献】

[1] M. Blondel, A. Fujino, N. Ueda, "Convex factorization machines," in *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2015.

## 【連絡先】

ブロンデル マチュー (Mathieu Blondel) 上田特別研究室  
E-mail : mathieu.blondel(at)lab.ntt.co.jp