離散最適化ユニット
Discrete Optimization Unit

前原貴憲
Takanori Maehara

ミッション

離散構造論・離散最適化手法に基づく
機械学習・人工知能研究に活用可能な
理論・手法を開発する

コアメンバー

- 前原貴憲 (UL)
- 谷合竜典
- Tiphaine Viard
- 波多野大督
- 福永拓郎

他，パートタイム学生 4 名

2019年度主要成果

- 滑らかでない分布を推定する統計理論
  [前原他, AISTATS'18 Best Paper]
- 光の反射モデルを組み込んだ深層学習モデルによる照度差ステレオ
  [谷合・前原, ICML'18]
- 一般因果グラフに対するバンディットアルゴリズム
  [福永・波多野他, ICML'18]
- 優モジュラゲームの最小コアを求めるアルゴリズム
  [波多野他, IJCAI'18]
- ストリームグラフ構造を利用した推薦システム
  [Viard他, Computer Networks]
- モジュラ束上の劣モジュラ関数に基づく貪欲法で解ける部分空間選択問題の特徴づけ
  [中島 (パートタイム)・前原, AAAI'19]

上記を含む全 26 本の論文誌・国際会議発表

滑らかでない分布を推定する統計理論

分布推定問題

入力：未知の分布からサンプルされた点の集合
出力：元の分布を近似する分布

基本的には パラメタ θ をもつ分布族 $\rho(x \mid \theta)$ を考え
サンプルに最もよくあうように θ を調整する

- どのような分布族を考えればよい？
- サンプル数 n vs 精度の関係は？

滑らかな関数の推定

既知の定理 (see: Tsybakov の教科書)
適当な仮定のもと D 次元上の分布が β 回微分可能のとき，カーネル密度推定はサンプル数に対して誤差 $O(n^{-2\beta/(2\beta+D)})$ を達成

「滑らかなほど推定が簡単」というのは直感的
関数が滑らかでない場合はどうすればよい？
（金融系等，打切りが発生するデータは自然に不連続になるため，実用上も重要な課題）

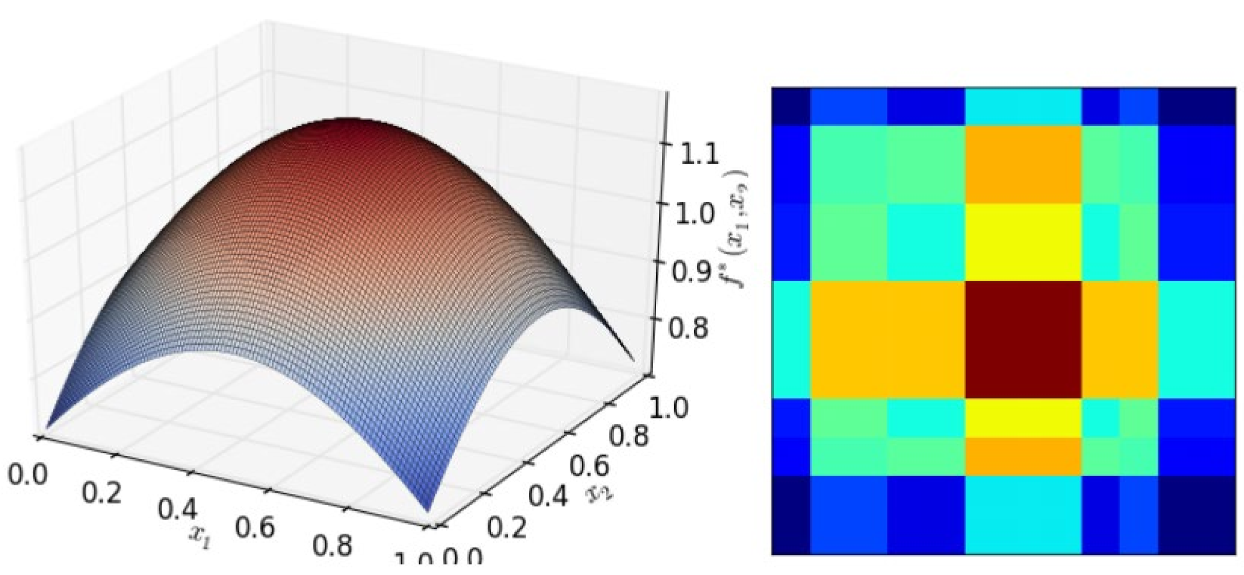
本研究の結果

定理 [Imaizumi, Maehara, and Yoshida, AISTATS'18]
密度関数が有界である任意の分布に対して誤差 $O(1/\sqrt{\log n})$ の推定手法が存在する
このレートはミニマックス最適

理論の鍵：Szemeredi の正則性補題

分布間の全変動距離: $TV(\rho,\rho') = \sup_S |\rho(S) - \rho'(S)|$

定理 (Szemeredi の正則性補題 [Szemeredi'75])
任意の単位立方体上の有界分布関数 ρ に対して
単位区間の分割 $[0,1] = I1 \cup \dots \cup Ik$ が存在し，
この区間から誘導されるセル上の区分定数分布関数で
元の関数を TV の意味で $O(1/\sqrt{D} \log k)$ 近似する
ものが存在



実際に分布を推定するには，サンプル点だけを見て
背後の分布に対する上定理の単位区間分割を見つける
ことで行う（アルゴリズムも提案）

コメント

Szemeredi の正則性補題は元々「十分大きなグラフはランダムグラフで近似できる」という形で述べられる
近代グラフ理論で最も重要な定理の1つ
最近も巨大グラフ解析の文脈では注目されつつあったが，分布推定の分野に初めて導入したのが本ユニットらしい成果といえる