



チーム体制

- 畑 晃平 (チームリーダー, 九大)
Sherief Hashima (Research Scientist)
学生パートタイマー 2名

研究協力者

- 末廣 大貴 (客員, 九大)
瀧本 英二 (九大)

研究概要

動機: 大規模データの多くは離散的である一方,

離散的構造や冗長性を利用した最適化手法は乏しい

例: 材料科学における化合物, DNA配列, 圧縮データ...

目標: 離散的・構造的な性質を利用した最適化・予測問題に対するアルゴリズム

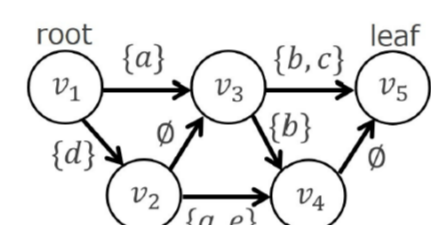
- ボトルネック: 簡潔データ構造・データ圧縮研究と最適化研究の融合
現状: 前者は検索や列挙に特化, 後者は前者を十分活用できていない

本アプローチ: 離散/オフライン/オンライン/バンディット最適化に基づく系統的な方法論

成果例: 圧縮データ上での機械学習・最適化 [TCS 20][COCOON 23]

- キーアイデア: 圧縮によるデータの構造化 + オンライン最適化 + 拡張定式化
省スペース性: データを圧縮したまま最適化
位置づけ: 従来圧縮型手法で解けなかった最適化問題の解決

展望: 離散データ上の簡潔表現に基づく最適化理論の確立



研究トピック

決定ダイアグラムに基づく最適化問題の拡張定式化(ongoing)

多様なオンライン意思決定問題に対する理論と基盤技術(ongoing)

バンディット手法の通信工学への応用 (Hashima)

マルチインスタンス学習(末廣)

ρ-insentitive損失に対するブースティングに基づく回帰手法

オンライン組合せ最適化におけるメタラウンディングの改良

バンディット手法の通信工学への応用(Hashima)

シミュレーションから実環境へのアプローチ

自動運転車のソフトウェアOTA (over-the-air) 更新

Introduction (1/3) Over the air updates?? VTC2024-Fall. Includes diagrams of OTA updates and photos of smart mobility vehicles.

マルチインスタンス学習 (末廣)

Learning from Label Proportion (LLP): 個々のデータがインスタンスの集合(バッグ)で与えられるようなマルチインスタンスデータに対し, バッグ内に含まれるインスタンスのクラスの割合のみが与えられる問題.

背景: 医療画像のWSI (Whole Slide Image) など, バッグサイズ(インスタンス数)が多い場合, バッグ内に含まれるインスタンスをすべて入力すると膨大なメモリを必要とするため, バッグ内からランダムにサンプリングして学習に用いる方法が慣習的に行われている. 一方で, サンプリングした場合とサンプリングしない場合ではクラス割合が一般に異なる.

本研究: 超幾何分布を使うことで理論的にサンプリングしたデータのクラス割合を推定, 適当に摂動させ学習するアルゴリズムを提案.

Weakly supervised learning using class label proportions within bags. Effectiveness on existing LLP methods. Includes a table of classification accuracy for various datasets and methods.

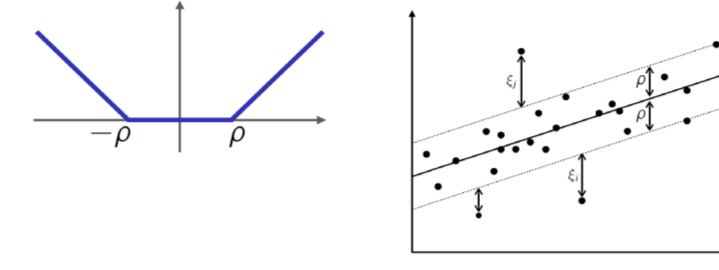
ρ-insentitive損失に対するブースティングに基づく回帰手法

ヒンジ損失に対する回帰問題

- 各点との距離をヒンジ損失として被る回帰問題.

Hinge_{ρ,w}f(x,y) = max{0, |y - f(x)| - ρ}

- 全ての点に対してヒンジ損失の総和を小さくしたい.
ヒンジ損失のパラメータ ρ ≥ 0 も同時に小さくしたい.



- ブースティングでこの回帰問題を解く.
Hinge_{ρ,w}(x,y) = max{0, |y - ∑_{h∈H} w_h h(x)| - ρ} とする.

入力: 訓練例集合 S = {(x1, y1), (x2, y2), ..., (xm, ym)} ⊂ X × R.

ヒンジ損失最適化問題

min_{ρ,w} ρ + 1/ν ∑_{i=1}^m Hinge_{ρ,w}(xi, yi) subject to ||w||_1 = 1, w ≥ 0.

出力: 最適解 w* による結合仮説 H*: x ↦ ∑_{h∈H} w*_h h(x).

関連研究と本研究の主結果

- LPBoost [Dobos] の回帰版.
LPBoost 同様, 理論保証は知られていない.
全ての点でヒンジ損失を最小化 [Dobos] (外れ値はないと仮定).
現実的でない.
過学習の危険性がある.
勾配ブースティング [FHT, CGH] 等の損失関数として設定すれば動くが, 理論保証は知られていない.
目的関数が平滑性を持たないことに起因.

本研究の成果

収束保証を持つアルゴリズムを提案.

ρ-insentitive損失に対するブースティングに基づく回帰手法(2)

弱学習者の仮定

弱学習者はどのような h ∈ H を返せば良いか?
理想: 入力分布 d に対して, 最も経験誤差を小さくする仮説 h ∈ H.

h ∈ arg min_{h∈H} ∑_{i=1}^m d_i |y_i - h(x_i)|

提案手法のアイデア

- Frank-Wolfe法に基づきアルゴリズムを設計.
ラグランジュ対称問題にエントロピー関数を加え,

max_{h∈H} min_{d∈Δ} ∑_{i=1}^m d_i |y_i - h(x_i)| - 1/η ∑_{i=1}^m d_i ln(md)

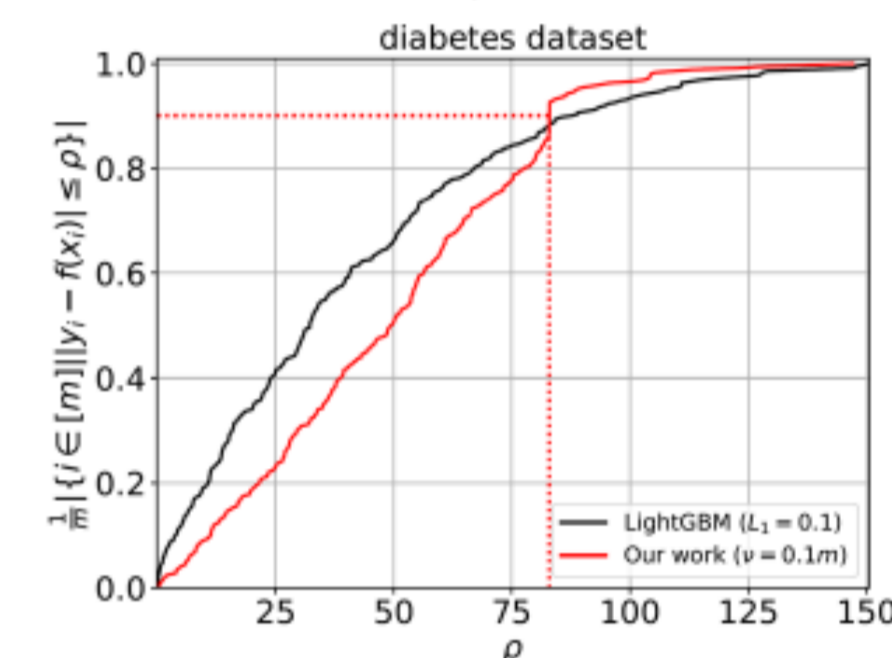
- Fenchel dual 対称問題に直す.
得られた主問題を, Frank-Wolfe法で解く.

主結果 (収束保証)

提案アルゴリズムは高々 O(ln(m/ν)/ε^2) 回の反復で収束.

実験

- LightGBM は L1 正則化付回帰.
弱学習者は深さ 1 の回帰木を返す.
高々 ν = 0.1m 個を除き, 誤差 ρ ≈ 83 以内に収まっている.

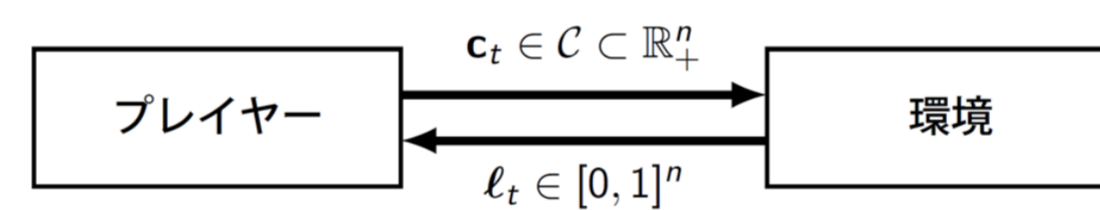


オンライン組合せ最適化におけるメタラウンディングの改良

オンライン組合せ線形最適化

決定集合 C ⊂ R^d が組合せ集合であるオンライン線形最適化.

各ラウンド t = 1, 2, ..., T において:



目標

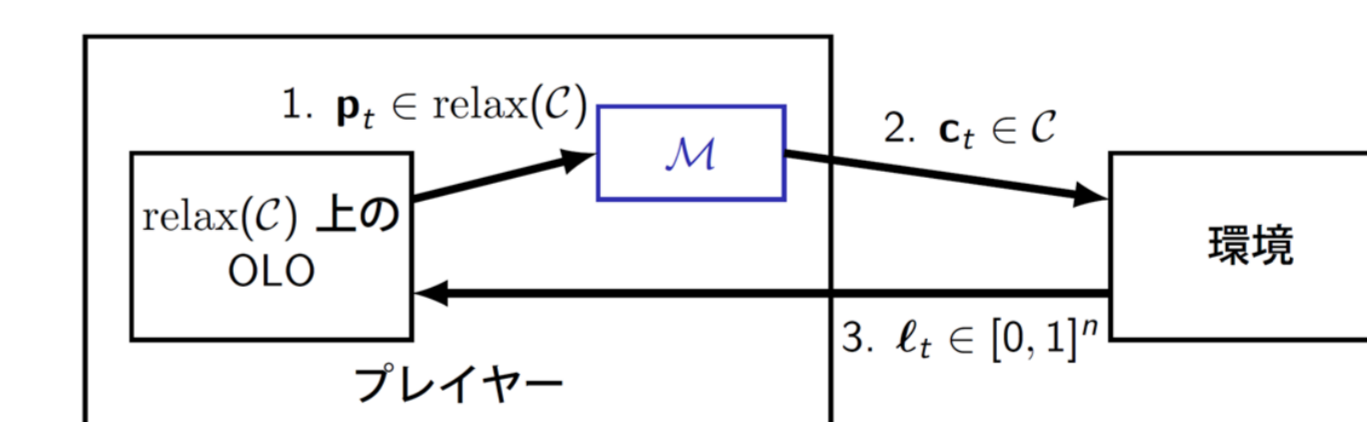
α-リグレット R_T(α) の最小化 (α ≥ 1).

R_T(α) = ∑_{t=1}^T c_t · ℓ_t - α min_{c∈C} ∑_{t=1}^T c_t · ℓ_t

- プレイヤーは何らかの近似アルゴリズムを使って良い.

[FHT13] の概要

- C を含む閉凸集合 relax(C) 上でオンラインアルゴリズムを走らせ.
得られた p_t ∈ relax(C) を C 上の点 c_t へと射影.



定理 (FHT13)

(C, relax(C)) に対するメタラウンディング M が存在 ⇒ R_T(α) = O(OLOのリグレット) となるアルゴリズムが存在.

先行研究との比較と本研究の成果

Table comparing algorithms (KKL09, G17, FHT13, This work) based on regret and oracle calls.

- diam_∞(C) = max{||c||_∞ | c ∈ C}.
[FHT13] の近似オラクルは, 上 3 つのものに比べ少し強い.
本研究は, 藤田らのものと全く同じ仮定.

本研究

ERLPBoost [WCV08] のようにエントロピー正則化.

max_{γ,ℓ} γ - 1/η ∑_{i=1}^n x_i ℓ_i ln(x_i ℓ_i / 1/n) s.t. c · ℓ ≥ γ, ∀ c ∈ C, x · ℓ = 1, ℓ ≥ 0.

C_0 = ∅ から初め, 各ラウンド k = 1, 2, ..., K において:

- 上の最適化問題を C_k 上で解き, 最適解を ℓ_k とする.
近似アルゴリズム B に ℓ_k を与え, c_{k+1} を得る.
C_{k+1} = C_k ∪ {c_{k+1}} と更新.

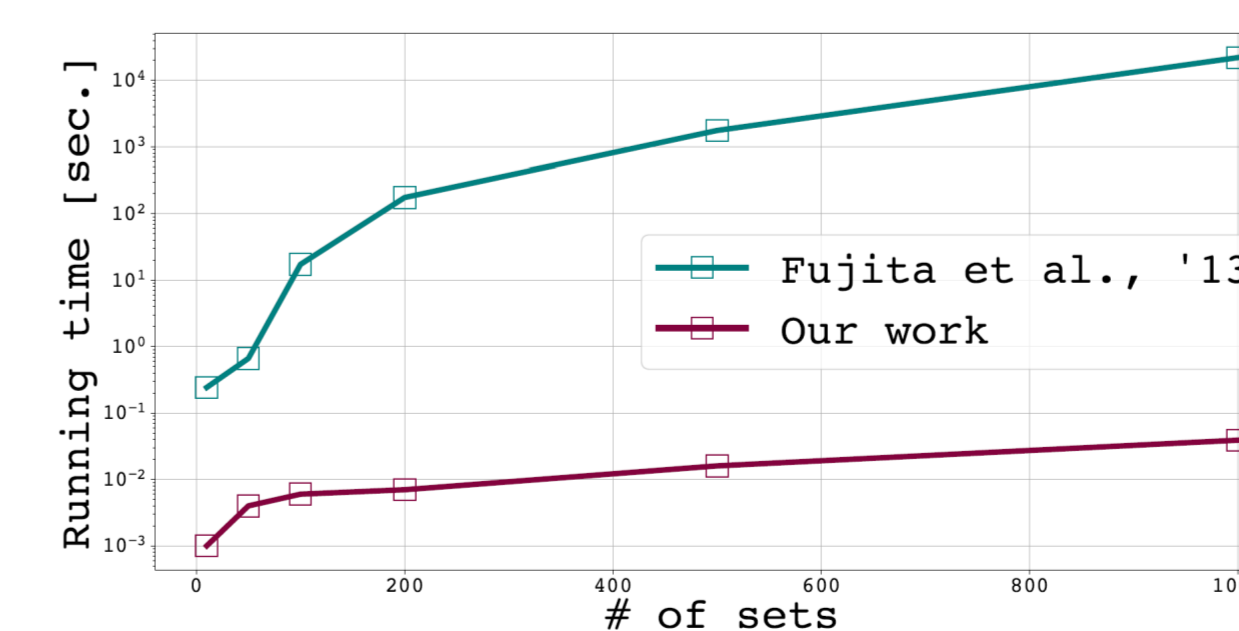
主結果 (収束保証)

提案手法は ε-近似解を O(diam_∞(C)^2 ln(n)/ε^2) ラウンドで発見.

オンライン組合せ最適化におけるメタラウンディングの改良(2)

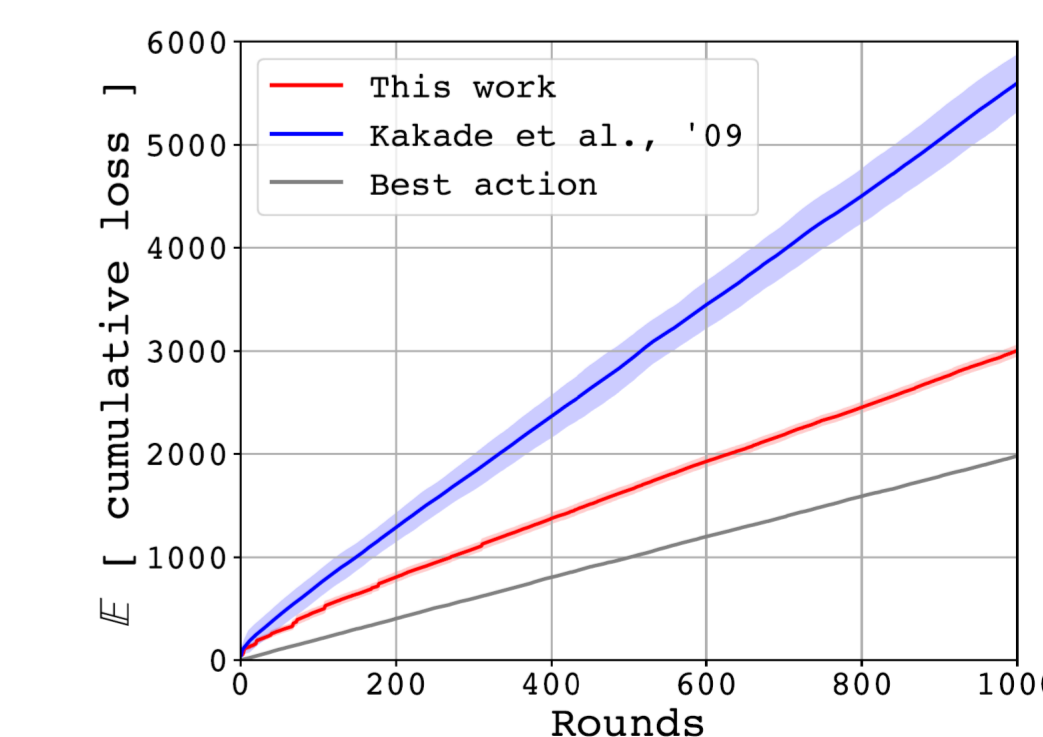
実験

- ランダムに生成した集合被覆問題.
次元 n = 20 で, 集合数 |C| を変化させて比較.
|C| ∈ {200, 400, 600, 800, 1000}



実験

- ランダムに生成した集合被覆問題.
n = 20, |C| = 100 で固定.



発表成果 (抜粋)

- 1. Sherief Hashima, Zongdian Li, Kohei Hatano, and Kei Sakaguchi, "Revolutionizing Over-the-Air Updates: Practical Dual-band V2X Measurements", VTC Fall 2024: 1-5, 2024.
2. Shunsuke Kubo, Shinnosuke Matsuo, Daiki Suehiro, Kazuhiro Terada, Hiroaki Ito, Akihiko Yoshizawa, and Ryoma Bise, "Theoretical Proportion Label Perturbation for Learning from Label Proportions in Large Bags", European Conference on Artificial Intelligence (ECAI), 2024.
3. Ryotaro Mitsuboshi, Kohei Hatano, and Eiji Takimoto, "Solving Linear Regression with Insensitive Loss by Boosting", IEICE, E107.D, Issue 3, 294-300, 2024.
4. Ryotaro Mitsuboshi, Kohei Hatano, and Eiji Takimoto, "Online combinatorial linear optimization via a Frank-Wolfe-based metarounding algorithm", IEICE, to appear.