修士論文 要旨

有向ハイパーグラフ上の全変動正則化

数理情報学専攻 48-196232 山中遥太
指導教員 岩田覚教授

1 はじめに

機械学習において構造正則化は，モデルのパラメータに特定の構造を持たせる正則化であり，過学習を避けモデルの解釈性を向上させる．無向ハイパーグラフ上の全変動正則化 [3] は，パラメータの成分にグループ単位で同じ値を持たせる構造正則化である．

本研究では，無向ハイパーグラフ上の全変動正則化の拡張として，有向ハイパーグラフ上の全変動正則化，およびその正則化項の近接作用素計算手法を提案する．有向ハイパーグラフ上の全変動正則化は，グループ単位でのパラメータの増加性を導く．論文の分野推定の数値実験では，有向ハイパーグラフ上の全変動正則化により著者の特徴を捉えたパラメータ推定が可能になり，無向ハイパーグラフ上の全変動正則化に比べ推定精度が改善されることを確かめた．

2 既存手法：無向ハイパーグラフ上の全変動正則化

機械学習において正則化では，損失関数 $f : \mathbb{R}^n \to \mathbb{R}$ と正則化項 $\Omega : \mathbb{R}^n \to \mathbb{R}$ の和を最適化する．つまり

$$\min_{\boldsymbol{w} \in \mathbb{R}^n} f(\boldsymbol{w}) + \lambda \Omega(\boldsymbol{w})$$

を解くことで，モデルのパラメータの推定を行う．特に，パラメータに特定の構造を持たせる正則化は構造正則化と呼ばれる．モデルの特徴に合わせた構造正則化を行うことで，過学習の防止や，モデルの解釈性向上が期待できる．

構造正則化の一種として，無向ハイパーグラフ上の全変動正則化 [3] がある．各頂点がモデルのパラメータの各成分に対応している無向ハイパーグラフ $H = (V, E)$ を考える．H 上の全変動正則化では，

$$\mathrm{TV}_H(\boldsymbol{w}) = \sum_{e \in E} c_e (\max_{i \in e} w_i - \min_{i \in e} w_i)$$

を正則化項として用いる．c_e は枝 e の重みである．TV_H は，ハイパーエッジ内の最大成分を小さく，最小成分を大きくすることで，ハイパーエッジ内の成分全てに同じ値を持たせる．本研究ではこれを拡張し，有向ハイパーグラフ上の全変動正則化を提案する．

3 提案手法：有向ハイパーグラフ上の全変動正則化

3.1 有向ハイパーグラフ上の全変動正則化

有向ハイパーグラフ $D = (V, E)$ では，各有向ハイパーエッジ $e \in E$ が始点集合 $T(e) \subseteq V$ と終点集合 $H(e) \subseteq V$ を持つ．各頂点がモデルのパラメータの各成分に対応している有向ハイパーグラフ D を考える．D 上の全変動正則化項を

$$\mathrm{TV}_D(\boldsymbol{w}) = \sum_{e \in E} \max\{c_e (\max_{i \in T(e)} w_i - \min_{i \in H(e)} w_i), 0\}$$

で定義する．c_e は枝 e の重みである．TV_D は始点集合の最大成分と終点集合の最小成分の差を小さくするため，推定されるパラメータは始点集合から終点集合にかけて，グループ単位で増加性を持ったものとなりやすい．

3.2 近接作用素の計算

TV_D の近接作用素

$$\mathrm{prox}_{\lambda \mathrm{TV}_D}(\boldsymbol{u}) = \operatorname*{argmin}_{\boldsymbol{w} \in \mathbb{R}^n} \frac{1}{2} \|\boldsymbol{w} - \boldsymbol{u}\|_2^2 + \lambda \mathrm{TV}_D(\boldsymbol{w})$$

の計算手法を提案する．有向ハイパーグラフのカット関数 κ_D は，$\delta_D(A) = \{e \in E \mid T(e) \cap A \neq \emptyset, H(e) \cap (V \setminus A) \neq \emptyset\}$ として $\kappa_D(A) = \sum_{e \in \delta_D(A)} c_e$ で定義される．本研究では次の定理を証明した．

定理 3.1. $B(\kappa_D) = \{\boldsymbol{x} \in \mathbb{R}^V \mid \sum_{i \in A} x_i \leq \kappa_D(A) (\forall A \subset V), \sum_{i \in V} x_i = \kappa_D(V)\}$ として，

$$\mathrm{prox}_{\lambda \mathrm{TV}_D}(\boldsymbol{u}) = -\min_{\boldsymbol{s} \in B(\kappa_D)} \sum_{i \in V} \frac{1}{2} (\lambda s_i - u_i)^2 + \frac{1}{2} \|\boldsymbol{u}\|_2^2.$$

この最適化問題は基多面体上の分離凸関数最小化問題であり，以下のパラメトリック劣モジュラ関数の最小解が切り替わるような $\alpha \in \mathbb{R}$（ブレークポイント）を全て求める問題と等価であることが知られている [1]：

$$\kappa_D^\alpha(A) = \kappa_D(A) - \sum_{i \in A} \frac{u_i}{\lambda} - |A| \alpha \ .$$

定理 3.2. D の各ハイパーエッジに対して，図 1 のように補助頂点と補助枝を加え有向グラフ G を構築する．この時，

$$\kappa_D^\alpha(A) = \min_{B \subseteq W} \kappa_G(\{s\} \cup A \cup B) - \sum_{i \in V} \frac{u_i}{\lambda} - |V| \alpha \ .$$

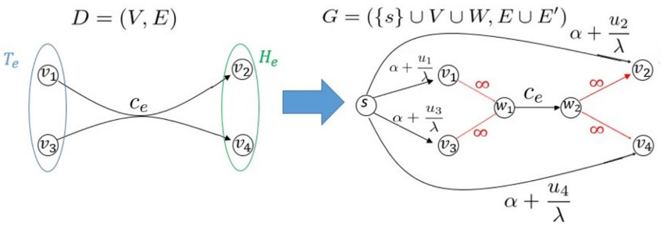

図 1. κ_D^α のグラフ表現.

したがって，κ_D^α は，G に最大流アルゴリズムを適用することで，最小化が可能である．この時，パラメトリック最大流アルゴリズム [2] を用いることで，一つの最大流問題を解くのと同じ最悪計算量で κ_D^α の全てのブレークポイントを計算する事が可能である．

4　数値実験：論文の分野推定

4.1　実験設定

機械学習分野の論文 (9274 件) のラベルを 1，データベース分野の論文 (6915 件) のラベルを-1 とし，いくつかの論文に正解ラベルが与えられたもとで，残りの論文のラベル推定を行った．損失関数はラベルベクトル $\boldsymbol{l} \in \{-1, 0, 1\}^V$ との二乗誤差 $f(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{l} - \boldsymbol{w}\|_2^2$ とし，無向ハイパーグラフ $H = (V, E)$ 上と，3 種類の有向ハイパーグラフ $\overrightarrow{D} = (V, E \cup \overrightarrow{E}), \overleftarrow{D} = (V, E \cup \overleftarrow{E}), \overleftrightarrow{D} = (V, E \cup \overleftarrow{E} \cup \overrightarrow{E})$ 上で全変動正則化を行う．ここで，無向ハイパーエッジ $e \in E$ は各著者に対応し，その著者の全ての論文から構成される．$\overrightarrow{E}, \overleftarrow{E}$ はいずれも有向ハイパーエッジのみを含む．各著者の論文を 2011 年以前に採択されたグループと，2012 年以降に採択されたグループに分割し，$\overrightarrow{E}$ では前者を始点集合，後者を終点集合とし，$\overleftarrow{E}$ では後者を始点集合，前者を終点集合とした．事前に与えたラベル数は $500, 1000, ..., 5000$ とした．

4.2　結果

結果は図 2 のようになった．$\mathrm{TV}_{\overrightarrow{D}}$ がラベルの増加性を導き，ラベルが増加している著者の論文に対して精度の良い推定をしている．同様に減少著者の論文に対して $\mathrm{TV}_{\overleftarrow{D}}$ が低い誤答率を記録している．両方向に枝を貼った $\mathrm{TV}_{\overleftrightarrow{D}}$ は，増加・減少著者の両方で精度の良い推定ができており，データの特徴とは逆方向の方向性を導く枝を張っても，推定精度はそれほど悪化しない事が分かる．また，ラベル一定の著者の論文に対しての $\mathrm{TV}_{\overleftrightarrow{D}}$ の低誤答率は，増加・減少の効果が打ち消し合い，パ

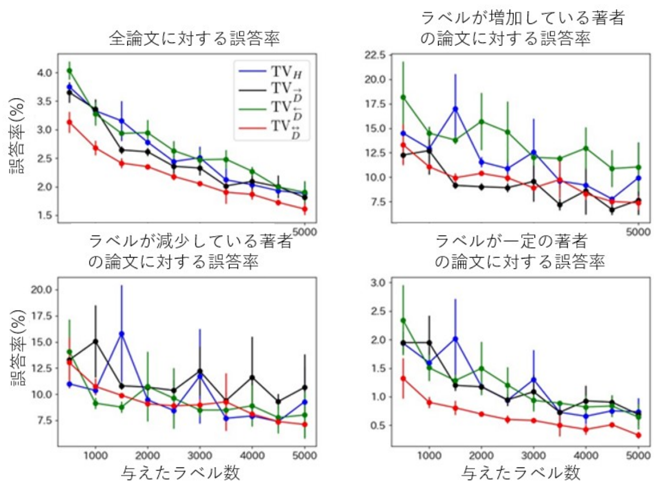

図 2. 各正則化の誤答率．誤答率計算の対象とした論文は，左上：全論文，右上：過去（2011 年以前）から直近（2012 年以降）にかけ平均ラベルが増加している著者の論文，左下：過去から直近にかけ平均ラベルが減少している著者の論文，右下：ラベルが全て同一の著者の論文.

ラメータに方向性が導かれなかったためだと考えられる．全てのタイプの著者で精度の良い推定ができたため，$\mathrm{TV}_{\overleftrightarrow{D}}$ が既存手法の TV_H より低い誤答率を記録できたと考察できる．

5　結論

有向ハイパーグラフ上の全変動正則化，およびその正則化項の近接作用素計算手法を提案した．論文の分野推定の数値実験では，各著者の過去の論文と直近の論文との間に両向きに有向ハイパーエッジを張ることで，各著者の特徴を捉えたラベル推定が可能になり，無向ハイパーグラフ上の全変動正則化に比べ推定精度が改善されることを確かめた．本実験では損失関数を二乗誤差としたが，損失関数が二乗誤差以外の場合でも近接勾配法を用いた最適化が可能であり，様々な問題設定への応用が可能である．

参考文献

[1] Satoru Fujishige. *Submodular Functions and Optimization.* North-Holland, 1991.

[2] Giorgio Gallo, Michael D. Grigoriadis, and Robert E. Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM Journal on Computing*, 18(1):30–55, 1989.

[3] Matthias Hein, Simon Setzer, Leonardo Jost, and Syama S. Rangapuram. The total variation on hypergraphs-learning on hypergraphs revisited. In *Advances in Neural Information Processing Systems*, pages 2427–2435, 2013.