Sample complexity of learning multi-value opinions in social networks

篠田 正人（奈良女子大学）
櫻井 祐子（名古屋工業大学） 小山 聡（北海道大学）

January 22, 2023

研究の概略

Conitzer らの論文において、ソーシャルネットワーク上での情報伝達構造を有向グラフで表し、部分的に与えられた情報が有向辺を通じて各点に伝わった結果を推測する PAC Learning に関する研究がなされている。

この研究では、有向グラフ上の各点が情報を持つか持たないかを表す 2 値モデルから数値の大小で表す多値モデルに拡張し、その全体の状態を推測するための PAC Learning における必要な学習量を 2 値の場合と比較して考察する。

Conitzer,V., Panigrahi,D. and Zhang,H. Learning Opinions in Social Networks, In *Proceedings of the 37th International Conference on Machine Learning*, 2122–2132 (2020).

Shinoda,M., Sakurai,Y. and Oyama,S. Sample Complexity of Learning Multi-value Opinions in Social Networks, In *PRIMA 2022: Principles and Practice of Multi-Agent Systems*, 192–207 (2023).

研究の目的

$G = (V, E)$ を有向グラフとする。

有向グラフ G 上で設定するこの学習モデルでは、各点 $u \in V$ が人を表し、情報が有向辺に沿って人から人へ伝わる状況にある。

最初に V の一部に与えられた情報が有向辺を通して広がっていった結果全体を、サンプルとして抽出する m 個の点の情報からよい精度で推測することを目標とし、そのために必要な学習データ数 m を知りたい。

その目的のため、まず PAC Learning (Probably Approximately Correct Learning) の一般的な枠組みについて説明し、その後で有向グラフ上での設定を具体的に述べる。

学習とは

集合 X の各点 x に、ラベル $f(x) \in W$ が与えられている。

X 上のラベル全体の値 f は W^X の要素（すなわち X から W への関数 $f : X \to W$）とみなされる。

学習者は X から m 個の点を、X 上の確率分布 $\mathcal{D}$ に従って独立に選んでそれぞれの点での f の値を知り、そのサンプルデータからラベル値の全体 f の推測として h を、事前に与えられた仮説集合 $\mathcal{H}(\subset W^X)$ から１つ選ぶ。

学習結果の評価と PAC Learning の目標

学習データからの推測の良さを評価するため、真の状態を表す f とその推測として選ばれた h に対しその違いを表す値

$$L_{\mathcal{D},f}(h) = P_{\mathcal{D}}(f(x) \neq h(x))$$

を定める。$P_{\mathcal{D}}$ は点 $x \in X$ を $\mathcal{D}$ に従って選ぶときの確率を表す。つまり、$L_{\mathcal{D},f}(h)$ は f を正解としたときの h の（$\mathcal{D}$ で測った）誤答率を表している。

PAC Learning の目標は、事前に与えられた $\varepsilon, \delta > 0$ に対して、$L_{\mathcal{D},f}(h) \leq \varepsilon$ をみたす $h \in \mathcal{H}$ を $1-\delta$ 以上の確率で定めることにある。

推測の具体的な手順

1. 学習者に、W^X の部分集合である仮説集合 $\mathcal{H}$ が与えられる。推測すべき f については、W^X の要素であること以外は知らされない。
2. 学習者は、X から選ばれた m 個の点 $(x_1, x_2, \ldots, x_m) \in X^m$ での f の値（学習データセット）$(f(x_1), f(x_2), \ldots, f(x_m))$ から f 全体の推測として $h \in \mathcal{H}$ を 1 つ定めるアルゴリズム A を定める。
3. $f \in W^X$ と X 上の確率分布 $\mathcal{D}$ が定まる。
4. X から m 個の点が確率分布 $\mathcal{D}$ に従って独立に選ばれ、学習者は選ばれた点 $(x_1, x_2, \ldots, x_m)$ および各点での値 $(f(x_1), f(x_2), \ldots, f(x_m))$ を知り、アルゴリズム A に従って推測 $h \in \mathcal{H}$ を定める。
5. 得られた推測 h の評価を $L_{\mathcal{D},f}(h)$ で行い、この値が ε 以下であれば推測に成功したものとする。

推測アルゴリズムに関する注意

- 2. および 4. において選ばれる m 個の点 $x_1, x_2, \ldots, x_m$ は重複する場合も考える必要がある。
- 学習者はアルゴリズム A の決定を、3. の手順より前、すなわち $f, \mathcal{D}$ についての情報を持たずに行わなければならない。すなわち A は $(X \times W)^m$ から $\mathcal{H}$ への写像である。

このように推測を行って成功する確率が、$\mathcal{D}, f$ によらず一様に $1 - \delta$ 以上となるアルゴリズム A が定められる最小の整数 m を $m_L(\varepsilon, \delta)$ とする。

なお上記手順の 3. において与えられる f が特に $\mathcal{H}$ の要素であるという仮定に相当する条件、すなわち $\min_{h \in \mathcal{H}} L_{\mathcal{D},f}(h) = 0$ を realizability condition という。本研究ではこの条件を仮定した上での PAC Learning を行う。

VC dimension

$W = \{0,1\}$ という 2 値モデルにおいて、一般的に $\mathcal{H}$ の複雑さを測る指標になる VC dimension（Vapnik-Chervonenkis dimension）を導入する。

Definition 1 (VC dimension)

X を空でない集合とし、$\mathcal{H} \subset \{0,1\}^X$ であるとする。X の部分集合 S が $\mathcal{H}$ によって shatter されるとは、$\mathcal{H}$ を S に制限した $\mathcal{H}|_S$ が $\{0,1\}^S$ に一致することを言う。$\mathcal{H}$ の VC dimension とは、$\mathcal{H}$ によって shatter される S の要素の個数の最大値であり、$d_{VC}(\mathcal{H})$ で表す。

例. $X = \{1,2,3\}$、$\mathcal{H} = \{(0,0,0),(0,0,1),(1,0,1),(1,1,0)\}$ のとき
$S_{12} = \{1,2\}$, $S_{13} = \{1,3\}$ とすると
$\mathcal{H}|_{S_{12}} = \{(0,0),(1,0),(1,1)\}, \mathcal{H}|_{S_{13}} = \{(0,0),(0,1),(1,1),(1,0)\}$ であって S_{13} が $\mathcal{H}$ によって shatter される。よって $d_{VC}(\mathcal{H}) = 2$ である。

VC dimension の例

例. $X = \{1, 2, 3, \ldots, n\}$, $\mathcal{H} = \{0, 1\}^X$. このとき $d_{VC}(\mathcal{H}) = n$.

例. $X = \{1, 2, 3, \ldots\}$, $\mathcal{H} = \{(x_1, x_2, x_3, \ldots) \mid x_4 = x_5 = \cdots = 0\}$.
このとき $d_{VC}(\mathcal{H}) = 3$.

例. $X = \{1, 2, 3, \ldots\}$, $\mathcal{H} = \{(x_1, x_2, x_3, \ldots) \mid \sum_{i=1}^{\infty} x_i \leq 3\}$.
このとき $d_{VC}(\mathcal{H}) = 3$.

例. $X = \{1, 2, 3, \ldots\}$,

$$\mathcal{H} = \left\{ (x_1, x_2, x_3, \ldots) \;\middle|\; \begin{array}{l} \text{ある } a, b \text{ が存在して } x_1 = \cdots = x_{a-1} = 0, \\ x_a = \cdots = x_b = 1, x_{b+1} = \cdots = 0 \end{array} \right\}.$$

このとき $d_{VC}(\mathcal{H}) = 2$.

VC dimension による必要サンプル数の評価

この VC dimension を用いて、2 値モデルでの PAC Learning において必要な学習データ数 $m_L(\varepsilon, \delta)$ は、realizability condition の仮定の下で下記のように上下からオーダーを評価することができる。

Theorem 2 (たとえば Shalev-Shwartz and Ben-David の教科書参照)

$\mathcal{H} \subset \{0,1\}^X$ とし、$d_{VC}(\mathcal{H}) < \infty$ とする。このとき

$$C_1 \frac{d_{VC}(\mathcal{H}) + \log(1/\delta)}{\varepsilon} \le m_L(\varepsilon, \delta) \le C_2 \frac{d_{VC}(\mathcal{H}) \log(1/\varepsilon) + \log(1/\delta)}{\varepsilon}$$

をみたす正定数 C_1, C_2 が存在する。

この定理から、たとえ X が無限集合であったとしても $\mathcal{H}$ の VC dimension が有限であれば PAC Learning が可能であることがわかる。

Multitype model

ここからは、ラベルの集合が $W = \{0, 1, 2, \ldots, k-1\}$ の k 種類となる多値モデルについて述べる。

多値モデルでは、学習データからの推測 h の良さを評価する際に、誤答率 $L_{\mathcal{D},f}(h) = P_{\mathcal{D}}(f(x) \neq h(x))$ に加えて

$$l_{\mathcal{D},f}(h) = E_{\mathcal{D}}(|f(x) - h(x)|)$$

を定める。この 2 つの誤差評価には

$$L_{\mathcal{D},f}(h) \leq l_{\mathcal{D},f}(h) \leq (k-1)L_{\mathcal{D},f}(h)$$

の不等式が成り立ち、$k = 2$ のときには f, h のとり得る値が 0 と 1 のみであるため $L_{\mathcal{D},f}(h) = l_{\mathcal{D},f}(h)$ である。

VC dimension の拡張に向けて

多値モデルでの PAC Learning において、得られた推測 h の評価を $l_{\mathcal{D},f}(h)$ または $L_{\mathcal{D},f}(h)$ で行い、この値が ε 以下であれば推測に成功したとする。

そしてこの推測に成功する確率が $1-\delta$ 以上となるために必要な学習データ数をそれぞれ $m_L(\varepsilon,\delta), m_l(\varepsilon,\delta)$ と書く。

$\mathcal{H}$ の複雑度を測るための VC dimension の拡張として仮に、$\mathcal{H}$ を S に制限した $\mathcal{H}|_S$ が $\{0,1,\ldots,k-1\}^S$ に一致するような S の要素の個数の最大値として $d_{VC}^{(k)}(\mathcal{H})$ を定めると、$k>2$ において $d_{VC}^{(k)}(\mathcal{H})<\infty$ であっても $m_L(\varepsilon,\delta), m_l(\varepsilon,\delta)=\infty$ となる場合が生じる。

例. $|X|=\infty$, $k=3$, $\mathcal{H}=\{0,1\}^X\subset\{0,1,2\}^X$ とするとき、$d_{VC}^{(k)}(\mathcal{H})<\infty$ であるが $m_L(\varepsilon,\delta), m_l(\varepsilon,\delta)=\infty$.

Natarajan dimension

ここから、$\mathcal{H}$ に対して新たな次元を定義する。

Definition 3 (Natarajan dimension)

X を空でない集合とし、$\mathcal{H} \subset \{0, 1, 2, \ldots, k-1\}^X$ であるとする。X の部分集合 S が $\mathcal{H}$ によって N-shatter されるとは、S 上の $\{0, 1, 2, \ldots, k-1\}$ 値関数 f_1, f_2 で、S のすべての部分集合 T に対し、$g \in \mathcal{H}$ で

$$x \in T \text{ ならば } g(x) = f_1(x), \quad x \notin T \text{ ならば } g(x) = f_2(x)$$

をみたすものが存在することを言う。$\mathcal{H}$ の Natarajan dimension とは、$\mathcal{H}$ によって N-shatter される S の要素の個数の最大値であり、$d_N(\mathcal{H})$ で表す。

この Natarajan dimension は $\mathcal{H} \subset \{0, 1\}^X$ のときは VC dimension に一致し、定義から一般の k に対して $d_{VC}^{(k)}(\mathcal{H}) \leq d_N(\mathcal{H})$ が成り立つ。

Natarajan dimension の例

Natarajan dimension では、$x \in X$ ごとに各ラベルの値を

$f_1(x)$ ならば 0、$f_2(x)$ ならば 1、それ以外なら $*$ (null) に置き換える

例. $\mathcal{H} = \{(0,0,0),(0,2,2),(1,0,0),(1,2,2),(2,1,0)\}$

ここで $f_1(x_1) = 0, f_2(x_1) = 1, f_1(x_2) = 0, f_2(x_2) = 2, f_1(x_3) = 0, f_2(x_3) = 2$ として置き換えると

$\mathcal{H}' = \{(0,0,0),(0,1,1),(1,0,0),(1,1,1),(1,*,0)\}$,

$\mathcal{H}'|_{\{x_1,x_2\}} = \{(0,0),(0,1),(1,0),(1,1)\}$ なので $d_N(\mathcal{H}) = 2$

Graph dimension

Definition 4 (Graph dimension)

X を空でない集合とし、$\mathcal{H} \subset \{0,1,2,\ldots,k-1\}^X$ であるとする。X の部分集合 S が $\mathcal{H}$ によって G-shatter されるとは、S 上の $\{0,1,2,\ldots,k-1\}$ 値関数 f で、S のすべての部分集合 T に対し、$g \in \mathcal{H}$ で

$$x \in T \text{ ならば } g(x) = f(x), \quad x \notin T \text{ ならば } g(x) \neq f(x)$$

をみたすものが存在することを言う。$\mathcal{H}$ の Graph dimension とは、$\mathcal{H}$ によって G-shatter される S の要素の個数の最大値であり、$d_G(\mathcal{H})$ で表す。

この Graph dimension は $\mathcal{H} \subset \{0,1\}^X$ のときは VC dimension に一致し、定義から一般の k に対して $d_{VC}^{(k)}(\mathcal{H}) \leq d_N(\mathcal{H}) \leq d_G(\mathcal{H})$ が成り立つ。

Graph dimension の例

Graph dimension では、$x \in X$ ごとに各ラベルの値を

$f_1(x)$ ならば 0、それ以外なら 1 に置き換える

例. $\mathcal{H} = \{(0,0,0),(0,2,2),(1,0,0),(1,1,2),(2,1,0)\}$

ここで $f(x_1) = 0, f(x_2) = 0, f(x_3) = 0$ として置き換えると

$\mathcal{H}' = \{(0,0,0),(0,1,1),(1,0,0),(1,1,1),(1,1,0)\}$,

$\mathcal{H}'|_{\{x_1,x_2\}} = \{(0,0),(0,1),(1,0),(1,1)\}$ なので $d_G(\mathcal{H}) = 2$

Multitype model での必要なサンプル数の評価

これらの次元を用いることで $m_L(\varepsilon, \delta)$ の値は以下のように評価される。

Theorem 5 (Daniely,Sabato,Ben-David and Shalev-Shwartz 2015)

$d_G(\mathcal{H}) < \infty$ を仮定する。このとき正定数 C_1, C_2 が存在し、

$$C_1 \frac{d_N(\mathcal{H}) + \log(1/\delta)}{\varepsilon} \leq m_L(\varepsilon, \delta) \leq C_2 \frac{d_G(\mathcal{H}) \log(1/\varepsilon) + \log(1/\delta)}{\varepsilon}$$

が成り立つ。

Ben-David,Cesa-Bianchi,Haussler and Long (1995) では

$$d_N(\mathcal{H}) \leq d_G(\mathcal{H}) \leq 4.67 \log_2 k \ d_N(\mathcal{H})$$

が示されており、$d_N(\mathcal{H})$ と $d_G(\mathcal{H})$ が有限であることは同値である。

有向グラフ上での情報伝達モデルの設定

$G = (V, E)$ を有向グラフとする。

$(u, v) \in E$ は点 u から v への有向辺を表すこととする。2 点 $u, v \in V$ に対し、この 2 点を有向辺で結ぶ V の頂点の列 $u, u_1, u_2, \ldots, u_{l-1}, v$ （$u_0 = u, u_l = v$ とするとき $1 \leq i \leq l$ について $(u_{i-1}, u_i) \in E$）が存在するとき $u \to v$ と表記する。

2 点を最短の有向辺列で結んだときの辺数を $\rho(u, v)$ とする。$u \to v$ でなければ $\rho(u, v) = \infty$ である。

なお、2 点を結ぶ頂点の列を V の部分集合 U のものに制限したときには $\rho_U(u, v)$ と書く。$u, v \in U$ に対して $\rho(u, v) \leq \rho_U(u, v)$ である。

有向グラフ上のラベルの制約条件

各頂点に与えられるラベルの集合を $W = \{0, 1, 2, \ldots, k-1\}$ とする。各点 u に与えられたラベルの値を $f(u)$ と表す。

このラベルの値の大小はある事柄に関する知識の大小を表し、この知識はグラフの有向辺を伝わって広がる、すなわち

$$u \to v \quad \Longrightarrow \quad f(u) \le f(v) \qquad\qquad (\#)$$

が成り立つものとする。Conitzer らのモデルでは各点の状態は知識を持っているかいないかの 2 値（$k=2$）である。

各点 u は有向辺で結ばれた点以外からも情報を得るものとし、$u \to v$ かつ $f(u) < f(v)$ という状態は許すことにしている。

仮説集合 $\mathcal{H}$

V 上のラベルのすべての状態集合 W^V のうち、ラベルの大小関係の条件 $(\#)$ をみたすものを $\mathcal{H}$ で表す。また V の部分集合 V' に $\mathcal{H}$ を制限したものを $\mathcal{H}|_{V'}$ と表す。

例. $V = \{v_1, v_2, v_3\}, E = \{(v_1, v_2), (v_2, v_3), (v_3, v_2)\}, k = 3$ とする。

このとき $\mathcal{H}$ の要素を $(f(v_1), f(v_2), f(v_3))$ の形で表すことにすると、

$\mathcal{H} = \{(0,0,0), (0,1,1), (0,2,2), (1,1,1), (1,2,2), (2,2,2)\}$

となり、$V' = \{v_1, v_3\}$ とすれば

$\mathcal{H}|_{V'} = \{(0,0), (0,1), (0,2), (1,1), (1,2), (2,2)\}$ である。

2 値モデルでの評価

この情報伝達モデルでは、まず学習者に有向グラフ $G = (V, E)$ の情報が与えられる。学習者はこの時点で条件 (#) をみたす要素の集合 $\mathcal{H}$ を W^V の部分集合として定めることができる。

$W = \{0, 1\}$ の場合、条件 (#) をみたすものとして定まった $\mathcal{H}$ については、その VC dimension はグラフ内で互いに影響しあわない点集合の最大個数として決定される。すなわち

$$d_{VC}(\mathcal{H}) = \sup_{U \subset V} \{ \, |U| \mid {}^{\forall} u, v \in U, \rho_U(u, v) = \infty \}$$

である。

この次元の値は G から定まるものであるため、上式の右辺の値を $d_{VC}(G)$ と書くこととする。

有向グラフと半順序集合

有向グラフ G において、ある点 $u \in V$ を出発して u に戻る有向辺列のサイクルが存在する場合、条件 $(\#)$ によってそのサイクル内での点のラベルの値はすべて同一にならなければならない。

よって G における同一サイクル内の点をすべて 1 点に縮約した有向非巡回グラフを考えれば十分である。有向非巡回グラフは半順序集合ともみなすことができる。

半順序集合において、どの要素も大小比較が可能でない部分集合は反鎖（antichain）と呼ばれ、その最大要素数は半順序集合の幅（width）と呼ばれる。この幅の値は $d_{VC}(G)$ そのものである。

従って、定理の中の必要な学習データ数評価内の d の値は、有向グラフ G からサイクルを取り除いた有向非巡回グラフを半順序集合とみなしたときの、その幅でそのまま置き換えることができる。

多値モデルと 2 値モデルの比較

有向グラフ $G = (V, E)$ の各点のラベルの値は $W = \{0, 1, 2, \ldots, k-1\}$ の k 種類とし、V 全体のラベル値の候補である $\mathcal{H}$ が条件 $(\#)$ をみたす f 全体の集合として定まる場合を考える。

ここでは、realizability condition は必然的にみたされることになる。

f が $\mathcal{H}$ の要素のいずれであるかを推測するとき、与えられる ε, δ に関する PAC Learning に（$l_{\mathcal{D},f}(h), L_{\mathcal{D},f}(h)$ について）必要な学習データ数を、ラベルの種類数 k を明示してそれぞれ $m_l^{(k)}(\varepsilon, \delta), m_L^{(k)}(\varepsilon, \delta)$ と書く。

Multitype model における評価

k 値モデルの場合の $m_l^{(k)}(\varepsilon, \delta)$ を、2 値モデルの場合と比べて評価する。本研究では以下の評価式を得た。

Theorem 6

$k \geq 2$ とする。*PAC Learning* に必要な学習データ数 $m_l^{(k)}(\varepsilon, \delta)$ について

$$m_l^{(2)}\left(\frac{\varepsilon}{k-1}, \delta\right) \leq m_l^{(k)}(\varepsilon, \delta) \leq m_l^{(2)}\left(\frac{\varepsilon}{k-1}, \frac{\delta}{k-1}\right)$$

が成り立つ。

この定理から、k 値モデルの PAC Learning は 2 値モデルの場合に比べておよそ $k \log k$ 倍のサンプル数があれば十分であることがわかる。

誤答率に関する $m_L^{(k)}(\varepsilon, \delta)$ についても類似の評価式を得られる。(省略)

ラベルが実数値の場合

ラベルのとり得る値を実数値まで拡張した場合の PAC Learning についても考察する。$W = [a, b]$ とし、値の大小に関する条件 (#) をみたす V から W への関数の集合を $\mathcal{H}$ とする。

PAC Learning において $E_{\mathcal{D}}(|f(u) - h(u)|) \leq \varepsilon$ をみたす $h \in \mathcal{H}$ を選ぶことができる確率が $1 - \delta$ 以上となる最小のサンプル数を $m_l^{[a,b]}(\varepsilon, \delta)$ とする。

この設定では、$|\mathcal{H}| = \infty$ であっても 2 値モデルでの PAC Learning が可能であれば k 値モデルでも可能であることが以下で示される。

Theorem 7

$$m_l^{(2)}\left(\frac{\varepsilon}{b-a}, \frac{\delta}{b-a}\right) \leq m_l^{[a,b]}(\varepsilon, \delta) \leq m_l^{(2)}\left(\frac{\varepsilon^2}{2(b-a)}, \frac{\varepsilon\delta}{b-a}\right)$$

が成り立つ。

有向グラフにおける dimension

V 上のラベルがとり得る状態すべての集合 $\mathcal{H}$ の複雑さを測る指標として VC dimension および Natarajan dimension と Graph dimension が導入された。多値モデルではこれらの次元の値を用いて（再掲）

$$C_1 \frac{d_N(\mathcal{H}) + \log(1/\delta)}{\varepsilon} \leq m_L(\varepsilon, \delta) \leq C_2 \frac{d_G(\mathcal{H}) \log(1/\varepsilon) + \log(1/\delta)}{\varepsilon}$$

と必要な学習データ数の上下界が評価されている。

本研究では、有向グラフ $G = (V, E)$ 上の多値モデルで条件 $(\#)$ によって定まる $\mathcal{H}$ について、Natarajan dimension と Graph dimension の値が G から具体的に定まることを示す。

有向グラフと dimension

Theorem 8

有向グラフ $G=(V,E)$ の k 値モデルで定まる $\mathcal{H}$ について

$$d_N^{(k)}(\mathcal{H}) = d_G^{(k)}(\mathcal{H}) = \sup_{U\subset V}\{\,|U|\mid {}^{\forall}u,v\in U, \rho_U(u,v)\neq k-1\}.$$

以上の通り定まる次元は有向グラフ G からわかる値であるので、これらを $d_N^{(k)}(G), d_G^{(k)}(G)$ と書くことにする。

また有向グラフ G 上での多値モデルでは、前に述べた $d_{VC}^{(k)}(\mathcal{H})$ の値は $d_{VC}(G)$ に一致することがわかり、これらをまとめると

$$d_{VC}(G) = d_N^{(2)}(G) = d_G^{(2)}(G) \leq d_N^{(k)}(G) = d_G^{(k)}(G)$$

である。

有向グラフモデルでの評価式まとめ

Theorem 9

有向グラフ $G = (V, E)$ について、定理 *8* によって定まる値 $d_N^{(k)}(G)$ が有限であるとする。G 上の、V から $W = \{0, 1, \ldots, k-1\}$ への関数 f が条件 *(#)* をみたすとき、この f を推測する *PAC Learning* において、

$$C_1 \frac{d_N^{(k)}(G) + \log(1/\delta)}{\varepsilon} \leq m_L(\varepsilon, \delta) \leq C_2 \frac{d_N^{(k)}(G) \log(1/\varepsilon) + \log(1/\delta)}{\varepsilon}$$

をみたす正定数 C_1, C_2 が存在する。

すなわち、$d_N^{(k)}(G)$ が有限であれば PAC Learning 可能である。

$d_N^{(k)}(G)$ の上界評価

さらにこの有向グラフの設定では、G と k から定まる $d_N^{(k)}(G)$ には VC dimension の値を用いた上界評価が定まることがわかる。

Proposition 10

有向グラフ $G = (V, E)$ で条件 $(\#)$ をみたす集合として定まる $\mathcal{H}$ に対し、$d_{VC}(G) < \infty$ であれば

$$d_N^{(k)}(G) \leq (k-1) d_{VC}(G)$$

が成り立つ。

この Proposition から、有向グラフ上のモデルでは $d_{VC}(G) < \infty$ かつ $d_N^{(k)}(G) = \infty$ となることはなく、有向グラフに対応する半順序集合の幅が有限であることが PAC Learning 可能であることの必要十分条件となることがわかる。

まとめ

ソーシャルネットワーク上での情報伝達構造を有向グラフで表したモデルにおいて、グラフ上の各点が持つ情報の大小を数値の大小で表すときの各点の値を推測する PAC Learning に必要な学習データ数について考察し、各点のとり得る値が 2 値の場合と多値の場合を比較した。

このデータ数は有向グラフから定まる Natarajan 次元の値を用いて上下から評価できることがわかり、頂点数が無限個の場合に PAC Learning が有限個の学習データによって可能かどうかもグラフの構造から定まるが、その条件は 2 値モデルと多値モデルで同一であることがわかった。