Decision list compression by mild random restrictions

Kewen Wu
Peking University

Joint work with

Shachar Lovett
UCSD

Jiapeng Zhang
UCSD $\rightarrow$ Harvard

Section 1

Introduction

Decision list (DL)

Decision list $L = ((C_1, v_1), (C_2, v_2), \ldots, (C_m, v_m))$ is

If $C_1(x) = \texttt{True}$ **then** output v_1,
else if $C_2(x) = \texttt{True}$ **then** output v_2,
...,
else if $C_m(x) = \texttt{True}$ **then** output v_m.

- C_i is a conjunction of literals: $C_i = y_1 \wedge y_2 \wedge \cdots$
- The last rule is default value: $C_m \equiv \texttt{True}$
- Its size is the number of rules
- its width is the maximal number of literals in C_i

Decision list (DL)

Let $L = ((C_1, v_1), \ldots, (C_m, v_m))$ be some width-w DL.

- L generalizes width-w DNF.

If $v_1 = \cdots = v_{m-1} = 1, v_m = 0$, then $L = C_1 \vee \cdots \vee C_{m-1}$.

- L generalizes width-w CNF.

If $v_1 = \cdots = v_{m-1} = 0, v_m = 1$, then $L = \neg C_1 \wedge \cdots \wedge \neg C_{m-1}$.

- Actually L is *strictly* more expressive than width-w DNF/CNF.

Main result

Definition (ε-approximation)

f is ε-approximated by g if $\Pr_{x\sim\{0,1\}^n}[f(x) \neq g(x)] \leq \varepsilon$.

Theorem (Decision list compression)

Width-w DL can be ε-approximated by a width-w size-s DL.

- Lovett and Zhang 2019: $s = (1/\varepsilon)^{O(w)}$.
- Now: $s = \left(2 + \frac{1}{w}\log\frac{1}{\varepsilon}\right)^{O(w)}$ and this is tight.

Section 2

Application

DNF sparsification and junta theorem

Corollary (DNF sparsification)

Width-w DNF can be ε-approximated by a width-w size-s DNF.

- Gopalan, Meka and Reingold 2013: $s = (w \log(1/\varepsilon))^{O(w)}$.
- Lovett and Zhang 2019: $s = (1/\varepsilon)^{O(w)}$.
- Now: $s = \left(2 + \frac{1}{w} \log \frac{1}{\varepsilon}\right)^{O(w)}$ and this is tight.

Corollary (Junta theorem)

Width-w DL can be ε-approximated by $\left(2 + \frac{1}{w} \log \frac{1}{\varepsilon}\right)^{O(w)}$-junta.

A k-junta is a function depending on at most k variables.

Learning small-width DNFs

Theorem (Jackson's harmonic sieve 1997)

Size-s n-variate DNFs is (ε, δ)-PAC learnable under the uniform distribution with $q = \mathrm{poly}(s, n, 1/\varepsilon, \log(1/\delta))$ membership queries.

Corollary (Learning small-width DNFs)

Width-w n-variate DNFs is (ε, δ)-PAC learnable under the uniform distribution with $q = \mathrm{poly}(s, n, 1/\varepsilon, \log(1/\delta))$ membership queries, where $s = \left(2 + \frac{1}{w} \log \frac{1}{\varepsilon}\right)^{O(w)}$.

Section 3

Proof overview

More definitions

Let $L = ((C_1, v_1), \ldots, (C_m, v_m))$ be a DL.

Definition (Index function)

Ind$L(x)$ is the index of first satisfied rule. (worst case: $v_i = i$)

Definition (Useful index)

Index i is useful if there exists x such that Ind$L(x) = i$.
#useful (L) is the number of useful indices.

Example

Assume $C_1 = x_1, C_2 = x_1 \wedge x_2, C_3 = 1$. Then Ind$L(1,1) = 1$ and #useful $(L) = 2$.

Step 1.1: randomness kills structure

We should be able to compress L (in some form) under restrictions.

Lemma (Håstad's switching lemma 1987)

Let f be a width-w DNF, $\alpha \in (0,1)$, and d be an integer. If ρ randomly restrict each input bit to $0, 1, *$ w.p. $(1-\alpha)/2, (1-\alpha)/2, \alpha$, then

$$\Pr_{\rho}[\mathrm{DT}(f\restriction_{\rho}) \geq d] \leq (5\alpha w)^d.$$

- Meaningful only when $\alpha \leq O(1/w) \implies$ most bits are fixed.
- Prove by encoding bad ρ.

Step 1.2: mild randomness also kills structure

Let's analyze L's size under restrictions.

Lemma

Let L be a width-w DL, $\alpha \in (0,1)$, and s be an integer. If ρ randomly restrict each input bit to $0, 1, *$ w.p. $(1-\alpha)/2, (1-\alpha)/2, \alpha$, then

$$\Pr_{\rho}\left[\#\textit{useful}\,(L \restriction_{\rho}) \geq s\right] \leq \frac{1}{s}\left(\frac{4}{1-\alpha}\right)^{w}.$$

- Meaningful for all kinds of α.
- Prove by encoding bad ρ and useful index i, $(\rho, i) \to (\rho', \text{aux})$. ρ' activates $*$'s in $C_i \restriction_{\rho} \implies \text{Ind}L(\rho') = i$.

Step 2: intuition for compression

Let $L = ((C_1, v_1), \ldots, (C_m, v_m))$ be a width-w DL.

- If index i is not useful, we can safely remove the i-th rule.
- Assume $p(i) = \Pr_x[\mathsf{Ind}L(x) = i]$ is decreasing in i.
  If $p(i)$ decreases fast, we only need to keep the top few.
- Let $L' = ((C_1, v_1), \ldots, (C_t, v_t), (C_m, v_m))$. Then

$$\Pr\left[L(x) \neq L'(x)\right] \leq \Pr\left[\mathsf{Ind}L(x) > t\right] = \sum_{i>t} p(i).$$

Now what?

Let $L = ((C_1, v_1), \ldots, (C_m, v_m))$ be a width-w DL.
Let ρ_α denote the random restriction with $*$-probability α.

- What we can do so far?
  We can analyze $q(\alpha, i) = \Pr\left[\text{index } i \text{ is useful in } L \restriction_{\rho_\alpha}\right]$, since

$$\sum_i q(\alpha, i) = \mathbb{E}\left[\#\text{useful}\left(L \restriction_{\rho_\alpha}\right)\right].$$

- What we want to do next?
  We want to bound $p(i) = \Pr\left[\text{Ind}L(x) = i\right]$, since

$$\Pr\left[L(x) \neq L'(x)\right] \leq \sum_{i>t} p(i).$$

Step 3: noise stability

Let's introduce noise stability to relate $p(i)$ and $q(\alpha, i)$.

Definition (Noise distribution $\mathcal{N}_\beta$)

$y \sim \mathcal{N}_\beta(x)$ is sampled by taking $\Pr[y_i = x_i] = (1+\beta)/2$.

Then for $x \sim \{0,1\}^n, y \sim \mathcal{N}_\beta(x)$, we can also do it by sampling

- 1. common restriction ρ with $\Pr\left[\rho_i = *\right] = 1 - \beta$.
- 2. x' by uniformly filling out $*$'s in ρ, and set $x = \rho \circ x'$.
- 2. y' by uniformly filling out $*$'s in ρ, and set $y = \rho \circ y'$.

Step 4: bridging lemma

Let $L = ((C_1, v_1), \ldots, (C_m, v_m))$ be a width-w DL and fix i.
Sample $x \sim \{0,1\}^n, y \sim \mathcal{N}_\beta(x)$, which can be seen as $\rho_{1-\beta}, x', y'$.

Define $\mathsf{Stab}(\beta, i) = \Pr\left[\mathsf{Ind}L(x) = \mathsf{Ind}L(y) = i\right]$ and recall
$p(i) = \Pr[\mathsf{Ind}L(x) = i]$, $q(\alpha, i) = \Pr[\text{index } i \text{ is useful in } L \restriction_{\rho_\alpha}]$.

Fact (Hypercontractivity)

$\mathit{Stab}(\beta, i) \le \left(\Pr\left[\mathit{Ind}L(x) = i\right]\right)^{\frac{2}{1+\beta}} = (p(i))^{\frac{2}{1+\beta}}$.

We also have

$$\begin{aligned}
\mathsf{Stab}(\beta, i) &= \Pr\left[\mathsf{Ind}L(x) = \mathsf{Ind}L(y) = i, \text{index } i \text{ is useful in } L \restriction_\rho\right] \\
&= q(1-\beta, i) \Pr\left[\mathsf{Ind}L(x) = \mathsf{Ind}L(y) = i \mid \text{index } i \text{ is useful in } L \restriction_\rho\right] \\
&\ge q(1-\beta, i) \left(\Pr\left[\mathsf{Ind}L(x) = i \mid \text{index } i \text{ is useful in } L \restriction_\rho\right]\right)^2 \\
&= (p(i))^2 / q(1-\beta, i).
\end{aligned}$$

Section 4

Open problems

Upper bound sparsification

Assume L is a width-w DNF.
L' is constructed by removing rules of L, thus $L'(x) \leq L(x)$.

Now we want some DNF L'' such that $L''(x) \geq L(x)$.

Problem (Upper bound compression)

L can be ε-approximated by a width-w size-s DNF from above.

- Gopalan, Meka and Reingold 2013: $s = (w \log(1/\varepsilon))^{O(w)}$.
- Lovett, Solomon and Zhang 2019: restricted to monotone case, $s = (\log w/\varepsilon)^{O(w)}$ implies improved sunflower lemma.

Upper bound sparsification

- What if we allow approximator to be width-$O(w)$?
- Can we show improved sunflower lemma implies it?
- How can we $1/100$-approximate this example from above?

Example

Let $m = \log w$ and define

$$f(x) = \bigvee_{i_1=1}^{m} \bigvee_{i_2=1}^{m} \cdots \bigvee_{i_w=1}^{m} (x_{1,i_1} \wedge x_{2,i_2} \wedge \cdots \wedge x_{w,i_w}).$$

Hitting set

Let $f : \{0,1\}^n \to \{0,1\}$ and write it as $f(x) = \sum_S f_S \prod_{i \in S} x_i$.

Problem (Hitting set)

Let $\mathcal{F} = \{S \mid f_S \neq 0\}$. Then there exists $T \subset [n]$ of size s, such that $T \cap S \neq \emptyset$ holds for any $S \in \mathcal{F}$.

- Best known bound is $s \approx \sqrt{|\mathcal{F}|}$.
- Can we show $s = (\log |\mathcal{F}|)^{O(1)}$?

Fooling small-width DNFs

Problem (Fooling small-width DNFs)

There exists an explicit pseudorandom generator $G : \{0,1\}^r \to \{0,1\}^n$ that δ-fools all width-w DNFs.

- Gopalan, Meka and Reingold 2013: $r \approx w^2 \log^2 w$.
- Potential approaches:
  - Upper bound compression
  - Mansour's conjecture
  - fractional PRG
  - ...

Thanks!