Однородная по степени нижняя оценка на веса многочленов с заданной знаковой функцией

Владимир Подольский*

Аннотация

Персептроном степени d называется многочлен p степени d от n входных переменных с целыми коэффициентами (весами). Мы говорим, что персептрон вычисляет функцию $f\colon \{0,1\}^n \to \{0,1\}$, если $f(x) = 1$ тогда и только тогда, когда $p(x) > 0$. Сумма абсолютных значений коэффициентов p называется общим весом персептрона.

Для всяких n и $d \le D \le \frac{\varepsilon n^{1/5}}{\log n}$ мы построим функцию, вычислимую персептроном степени d, но требующую общего веса $2^{(\delta n)^d/D^{4d}}$ при вычислении персептронами степени D, где $\varepsilon > 0$ и $\delta > 0$ – некоторые константы. В частности, если D постоянно, наша функция требует веса $2^{\Omega(n^d)}$ при вычислении персептронами степени D. Ранее функции с такими свойствами были известны только для $d = 1$ (и произвольного D) [3] и для $D = d$ [24]. При постоянных d построенные нами функции представимы в виде полиномиальных ДНФ. Лучшая нижняя оценка на веса персептронов известная ранее была $2^{\Omega(n)}$.

1 Введение

Пороговым элементом с входными булевыми переменными $x_1, \ldots, x_n$ мы называем линейный многочлен $\sum_{i=1}^n w_i x_i - t$, где $w_1, w_2, \ldots, w_n, t$ – целые

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 09-01-00709-а), программы поддержки ведущих научных школ НШ-845.2008.1 и гранта NAFIT ANR-08-EMER-008-01.

Предварительная версия данной работы была опубликована в [18].

числа. Он вычисляет булеву функцию $f(x_1, \ldots, x_n) = sgn(\sum_{i=1}^{n} w_i x_i - t)$, где sgn обозначает функцию знака: $sgn(x) = 1$, если x положительно, и $sgn(x) = 0$ в противном случае. Числа $w_1, w_2, \ldots, w_n$ и t называются весами и порогом, соответственно. Сумма абсолютных значений всех коэффициентов $\sum_i |w_i| + |t|$ называется общим весом порогового элемента.

В данной работе мы рассматриваем булевы схемы, составленные из пороговых элементов (пороговые схемы). Более точно, мы доказываем экспоненциальную оценку на размер пороговых схем специального вида (называемых персептронами), вычисляющих конкретные функции. Существует два способа определения размера пороговых схем: с учетом весов и без учета весов. В нашей нижней оценке мы пользуемся первым вариантом. То есть, мы определяем размер пороговой схемы как сумму всех общих весов ее элементов. Это эквивалентно тому, чтобы разрешать только веса ± 1 в пороговых элементах и определять размер как суммарное число проводов в схеме.

Пороговые схемы, рассматриваемые в этой работе, имеют постоянную глубину. Первой экспоненциальной оценкой для схем постоянной глубины была оценка для схем, состоящих из отрицаний, конъюнкций и дизъюнкций (неограниченной входной степени) из [23, 9]. (Заметим, что конъюнкции и дизъюнкции являются частными случаями пороговых элементов.) Следующая экспоненциальная оценка была получена для схем постоянной глубины, состоящих из отрицаний, конъюнкций, дизъюнкций и MOD p элементов[1] (где p – простое число) [19, 22]. (Заметим, что всякий MOD m элемент может быть вычислен пороговой схемой постоянной глубины и полиномиального размера.) Для MOD m элементов с составным m не известно сверхполиномиальных оценок даже для схем глубины 2.

Для пороговых схем глубины 2 общего вида была доказана экспоненциальная оценка на размер всякой схемы, вычисляющей скалярное произведение двух n-битовых строк [8] (напомним, что при подсчете размера мы учитываем веса пороговых элементов). Позже этот результат был усилен в двух направлениях: во-первых, в [16] было доказано, что экспоненциальная оценка верна, даже если мы не учитываем веса на проводах из входных переменных (то есть, мы учитываем веса только порогового элемента, находящегося в вершине схемы); во-вторых, в [6] было дока-

[1] Элемент MOD m на входных переменных $x_1, \ldots, x_n$ выдает 1, если число единиц среди $x_1, \ldots, x_n$ делится на m.

зано, что экспоненциальная оценка верна также, и если мы учитываем только веса пороговых элементов на нижнем уровне схемы. Не известно сверхполиномиальных оценок на число элементов в пороговых схемах глубины 2, вычисляющих конкретную функцию (то есть, для случая, когда мы не учитываем все веса).

В этой работе мы рассматриваем персептроны, пороговые схемы глубины 2 специального вида. Персептроном степени d называется булева схема глубины 2 с пороговым элементом в вершине и конъюнкциями входной степени не выше d на втором уровне. Персептрон степени 1 – это просто пороговый элемент. Общим весом персептрона называется общий вес его порогового элемента. Поскольку конъюнкция булевых переменных соответствует их умножению, функция, вычисляемая персептроном степени d – это знак многочлена степени d с целыми коэффициентами. То есть, можно также определить персептрон степени d как целочисленный многочлен $p(x)$ степени d от n переменных. Тогда персептрон $p(x)$ вычисляет функцию $f\colon \{0,1\}^n \to \{0,1\}$, если $f(x) = sgn\, p(x)$ для всякого $x \in \{0,1\}^n$.

Персептроны активно изучались в шестидесятых годах двадцатого века в теории интеллектуальных систем (смотри [13]), как простейшая модель нейрона. Позже они естественным образом возникли при изучении сложностных классов [2, 3, 4], доказательстве нижних оценок на коммуникационную сложность [20, 21], а также в теории обучения [11, 17, 12].

Мы будем доказывать нижние оценки на общие веса персептронов, вычисляющих конкретные заданные функции, в терминах их степени d и числа входных переменных n. Поскольку размер персептрона, как пороговой схемы глубины 2 (с учетом весов) больше, чем общий вес, наша оценка справедлива и для размера. Заметим, что существует булева функция от n переменных, не вычислимая никаким персептроном степени меньше n (функция MOD 2, смотри [13]). В данной работе мы заинтересованы только в функциях, вычислимых персептронами маленькой степени d, и таких что всякий персептрон степени d или больше, вычисляющий эту функцию, имеет большой общий вес. Например, мы не заинтересованы в скалярном произведении двух n-битовых строк (изучавшемся в [8, 16]) – эта функция не вычислима персептроном степени меньше n, так как MOD 2 сводится к ней.

Известные результаты такого типа можно разбить на две категории. К первой относятся результаты, в которых нижняя оценка на общий вес персептрона доказывается только для какой-то фиксированной степени.

Довольно давно были известны оценки вида $2^{\Omega(n)}$ на веса персептронов степени 1, вычисляющих конкретные функции (первый такой результат был получен в [15]). Однако, они не достигали известной верхней оценки $n^{O(n)}$, доказанной в работе Мурога [14] (смотри также [10]). Позднее Хостад [10] доказал точную (вплоть до константы в экспоненте) нижнюю оценку, то есть построил функцию от n переменных, вычислимую пороговым элементом, но такую что всякий пороговый элемент, вычисляющий ее, имеет общий вес не менее $n^{\Omega(n)}$.

Для степеней персептрона больших 1 известен только один результат такого рода. В работе [24] результат Хостада [10] был обобщен на произвольное d: для всякого $d \geqslant 2$ построен персептрон степени d, такой что всякий персептрон степени d, вычисляющий ту же функцию f имеет общий вес не меньше $n^{\Omega(n^d)}$ (константа в Ω зависит от d). Эта оценка, также как и результат Хостада, точна вплоть до константы в экспоненте. Действительно, верхняя оценка Мурога для $d = 1$ несложно обобщается на все d. То есть, всякий персептрон степени d с n входными переменными эквивалентен персептрону степени d с общим весом $n^{O(n^d)}$, где константа в O зависит от d. (Будем предполагать, что все конъюнкции на нижнем уровне различны и всякая переменная встречается во всяком конъюнкте не более одного раза. Тогда существует не более $O(n^d)$ функций на нижнем уровне персептрона степени d. Будем считать их независимыми переменными. Тогда верхняя оценка на общий вес пороговых элементов дает верхнюю оценку $n^{O(dn^d)}$ для персептронов степени d.)

Ко второй категории относятся результаты, в которых для одной и той же функции доказывается нижняя оценка на веса персептронов различных степеней. До настоящего времени было известно два результата такого типа. В работе [7] была построена функция, представимая пороговым элементом, но требующая экспоненциального веса при вычислении персептронами произвольной степени (рассмотренная в этой работе функция в некотором смысле самая сложная функция, среди всех функций, представимых пороговыми элементами). Второй результат такого типа принадлежит Бейгелю [3]. Им была построена функция от n переменных, вычислимая пороговым элементом с общим весом 2^n, и такая что для всякого D всякий персептрон степени D, вычисляющий эту функцию имеет общий вес $2^{\Omega(n/D^2)}$ (здесь константа в $\Omega(n/D^2)$ не зависит от D). Этот результат был получен в связи с разделением некоторых сложностных классов. Нижняя оценка Бейгеля не точна: она отличается от верхней оценки Мурога (для случая $D = 1$) в том, что основание степени

равно 2, а не n. С другой стороны, Бейгель доказал нижнюю оценку для всякой степени D, а не только для $D = 1$.

Можно ли обобщить результат Бейгеля на произвольные степени персептронов? А именно, верно ли, что для всяких d и n существует функция f, вычислимая персептроном степени d, и такая что для всякого D всякий персептрон степени D, вычисляющий f имеет общий вес $n^{\Omega(n^d)}$ (или хотя бы $2^{\Omega(n^d)}$, как в результате Бейгеля)? Здесь, константа в $\Omega(n^d)$ не зависит от n, но может зависеть от D и d? Этот вопрос открыт для всех $d > 1$.

В данной работе, мы делаем шаг в направлении положительного ответа на этот вопрос. Для всех n и $d \leqslant D \leqslant \frac{\varepsilon n^{1/5}}{\log n}$, где ε – некоторая положительная константа, мы строим функцию, вычислимую персептроном степени d, и требующую общего веса $2^{(\delta n)^d / D^{4d}}$ при вычислении персептронами степени не выше D, где δ – некоторая положительная константа. Этот результат все еще не достигает цели по следующим причинам: (1) основание экспоненты в оценке – константа, а не n, и (2) построенная функция зависит от D. Однако, достигнут существенный прогресс по сравнению с результатами [3] и [24]. Действительно, мы доказываем нижнюю оценку $2^{\Omega(n^d)}$ для всех d, а не только для $d = 1$, как в [3], и наша оценка верна для большого числа степеней персептронов, а не только для персептронов степени d, как в [24].

Коэффициент в экспоненте в нашем результате имеет вид δ^d / D^{4d}. Это делает наш результат содержательным и для непостоянных D. Например, мы можем зафиксировать произвольное d и положить $D = \log n$. Тогда мы получаем последовательность функций $f_n \colon \{0,1\}^n \to \{0,1\}$, вычислимых персептроном степени d, и таких что всякий персептрон степени не выше $\log n$, вычисляющий f_n, имеет общий вес $2^{\Omega(n^d / \log^{4d} n)}$. В частности, эта оценка верна для всех персептронов любой постоянной степени.

Получение оценок на веса персептронов интересно также и для функций из специальных классов. Мы рассмотрим один из самых простых таких классов (в том смысле, что функции, лежащие в нем, вычислимы булевыми схемами очень ограниченного вида), а именно дизъюнктивные нормальные формы (ДНФ) полиномиального (от числа переменных) размера. Этот класс активно изучался в теории обучения, в частности и с точки зрения минимальных степеней и весов персептронов (смотри работу [11] и ссылки в ней). Известно, что всякую полиномиальную ДНФ можно представить в виде персептрона степени $O(n^{1/2} \log n)$ с весом

$2^{O(n^{1/2}\log^2 n)}$, а также в виде персептрона степени $O(n^{1/3}\log n)$ [11]. Упоминавшийся выше результат Бейгеля [3] также был доказан для функции, представимой в виде полиномиальной ДНФ.

Мы докажем, что если d постоянно, то функции, для которых мы доказываем нашу оценку, также являются полиномиальными ДНФ. Это дает первую более чем экспоненциальную оценку (большую, чем $2^{\Omega(n)}$) на веса персептронов для функций, представимых в виде полиномиальных ДНФ.

Оставшаяся часть работы посвящена формулировке и доказательству нашего результата. В Разделе 2 мы строим наши функции. В Разделе 3 мы рассматриваем вопросы, связанные с представлением наших функций в виде ДНФ. В разделе 4 мы формулируем и доказываем основной результат.

2 Построение функции

Обозначим через $[n]$ множество первых n натуральных чисел $\{1, 2, \ldots, n\}$.

Наша функция будет обобщать следующую функцию, использованную Бейгелем [3] в похожих целях:

Определение 1 ([3]). *Для $x \in \{0,1\}^n$ булева функция* $\mathrm{OMB}(x)$[2] *принимает значение* 1 *тогда и только тогда, когда самая правая единица в x расположена на нечетной позиции. (Если x не содержит единиц, полагаем* $\mathrm{OMB}(x) = 0$.*)*

Нетрудно заметить, что функция $\mathrm{OMB}(x)$ представима в виде ДНФ полиномиального размера. А именно, если $x = (x_1, \ldots, x_n)$, то

$$\mathrm{OMB}(x) = \bigvee_{i \in [n],\ i-\text{нечетно}} x_i \wedge \overline{x}_{i+1} \wedge \overline{x}_{i+2} \wedge \ldots \wedge \overline{x}_n.$$

Теперь мы перейдем к определению нашей функции. Она будет зависеть от трех натуральных параметров n, d и D, где $d \leqslant D$. Первые два параметра задают число входных переменных (более точно, функция будет имеет nd переменных). Смысл параметров d и D следующий: функция будет вычислима персептроном степени d (с большим весом), но не вычислима никаким персептроном степени не выше D с маленьким

[2] OMB – это сокращение от ODD-MAX-BIT, использованного Бейгелем.

общим весом (мы уточним, что значит "маленький" и "большой" в этом контексте, позже).

Разобьем dn входных переменных на d групп одинакового размера n: $(x^1, x^2, \ldots, x^d)$, где $x^i = (x^i_1, x^i_2, \ldots, x^i_n)$. Всякий набор значений переменных x^i задает множество натуральных чисел $X_i = \{j | x^i_j = 1\}$. И наоборот, всякое подмножество X_i множества $[n]$ задает единственный набор значений переменных x^i. Так что, далее мы будем предполагать, что наша функция отображает кортежи $(X_1, \ldots, X_d)$ длины d подмножеств $[n]$ в $\{0, 1\}$.

Аналогично функции $\mathrm{OMB}(x)$, значение нашей функции на кортеже $(X_1, \ldots, X_d)$ зависит только от максимального (в некотором линейном порядке на множестве $[n]^d$) кортежа $(\alpha_1, \ldots, \alpha_d)$, лежащего в множестве $X_1 \times \cdots \times X_d$. Порядок на множестве $[n]^d$ будет довольно сложным и будет играть ключевую роль в доказательстве. Перейдем к определению этого порядка.

Сначала мы определим $D+1$ различный линейный порядок на множестве $[n]$. Первый порядок $<_1$ – это обычный порядок: $1, 2, \ldots, n$. Остальные порядки получаются из первого циклическими сдвигами. Более точно, разобьем множество $[n]$ на $D+1$ блок $u_1, \ldots, u_{D+1}$ одинакового размера так, чтобы каждый блок состоял из последовательных элементов: $u_i = (i-1)\frac{n}{D+1} + 1, \ldots, i\frac{n}{D+1}$ (формально, это определение годится только для n, делящихся на $D+1$; если n не делится на $D+1$, будем считать, что блоки устроены также, и их размеры отличаются не более чем на 1). Для всякого $i = 2, \ldots, D+1$ порядок $<_i$ получается из порядка $<_1$ перестановкой первых $i-1$ блоков с последними $D+2-i$ блоками: $u_i, \ldots, u_{D+1}, u_1, \ldots, u_{i-1}$[3]. Обозначим через $\mathrm{num}_i(t)$ порядковый номер числа t в порядке $<_i$ (здесь $i = 1, \ldots, D+1$ и $t \in [n]$, порядковый номер наименьшего элемента равен 1).

Теперь у нас есть $D+1$ различный порядок на множестве $[n]$, и мы обобщим функцию $\mathrm{OMB}(X)$ на все эти порядки.

Определение 2. *Для $X \subseteq [n]$ булева функция $\mathrm{OMB}_i(X)$, где $i \in [D+1]$, принимает значение 1 тогда и только тогда, когда максимальный элемент множества X в порядке $<_i$ расположен на нечетной позиции*

[3]В действительности, важно лишь, чтобы множество из $\lfloor n/(D+1) \rfloor$ наибольших элементов в каждом из этих порядков не пересекалось с аналогичным множеством для других порядков. Мы предпочитаем фиксировать конкретное простое семейство порядков с таким свойством.

в этом порядке. (Если X пусто, полагаем $\mathrm{OMB}_i(X) = 0$.*)*

Порядок на множестве $[n] \times \ldots \times [n]$ (где декартово произведение берется d раз), который мы теперь построим, будет по существу лексикографическим. Однако, есть существенные отличия. В лексикографическом порядке все компоненты заданных кортежей длины d из $[n]^d$ сравниваются в фиксированном порядке. Мы же будем сравнивать k-ые компоненты заданных кортежей в одном из порядков $<_1, \ldots, <_{D+1}$ и в каком именно, зависит от сравниваемых кортежей и от k. Более точно, это зависит от порядкового номера $k-1$-ой компоненты заданных кортежей в порядке, который мы использовали для сравнения их $k-1$-ых компонент.

Перейдем к точному определению порядка на кортежах. Первые компоненты α_1 и β_1 двух заданных кортежей длины d, $(\alpha_1, \ldots, \alpha_d)$ и $(\beta_1, \ldots, \beta_d)$, сравниваются в порядке $<_1$. Если они не равны, то мы уже сравнили кортежи: больше тот кортеж, чья первая компонента больше. В противном случае, мы сравниваем вторые компоненты кортежей. При этом мы пользуемся следующим рекурсивным правилом для выбора следующего порядка.

Предположим, что порядок $<_{i_k}$ (для сравнения k-ых компонент кортежа) уже определен и оказалось, что $\alpha_k = \beta_k$. Порядок для сравнения $(k+1)$-ых компонент определяется порядковым номером α_k (который совпадает с β_k по предположению) в порядке $<_{i_k}$. А именно,

$$i_{k+1} \equiv \left\lceil \frac{\mathrm{num}_{i_k}(\alpha_k)}{2} \right\rceil (mod(D+1)).$$

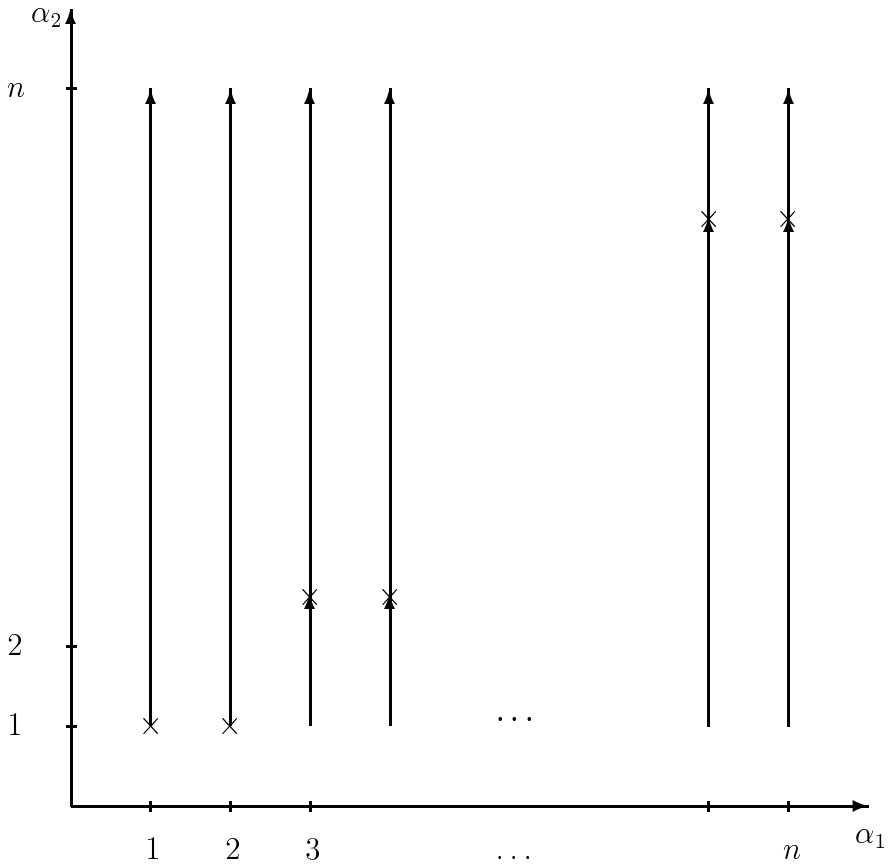
Другими словами, при увеличении α_k от минимума к максимуму (в порядке $<_{i_k}$), число i_{k+1} пробегает следующие значения:

$$1, 1, 2, 2, \ldots, (D+1), (D+1), 1, 1, \ldots, (D+1), (D+1), \ldots. \qquad (1)$$

Важно, что для всякого кортежа α порядки, используемые для сравнения компонент α с компонентами другого кортежа β зависят лишь от α и от номера компоненты (но не зависят от β).

Легко проверить, что определенное нами бинарное отношение иррефлексивно и транзитивно. Таким образом, мы определили отношение строгого линейного порядка.

Это отношение порядка несложно устроено в случае $d = 2$. Будем изображать пары (α_1, α_2) точками на плоскости (смотри рисунок).



Случай $d = 2$.

Первая компонента α_1 соответствует горизонтальной оси, вторая компонента α_2 – вертикальной. Пары сравниваются в соответствии с двумя правилами:

Правило 1. Пара, расположенная левее, меньше пары, расположенной правее.

Правило 2. В каждом столбце минимальная пара помечена знаком "$\times$". (В первых двух столбцах минимальные пары расположены в нижнем ряду. В двух последующих столбцах они расположены в $\frac{n}{D+1}$ ряду снизу, в следующих двух столбцах – в $\frac{2n}{D+1}$ ряду снизу, и так далее.) Стрелки указывают направление возрастания пар: Над минимальной парой, чем выше пара, тем она больше. То же самое верно и для пар ниже минимальной. Наконец, самая нижняя пара больше самой верхней.

В случае $d > 2$ построенное отношение порядка будет не таким простым. Однако, поскольку мы будем рассуждать по индукции, мы всегда

будем рассматривать только две соседние компоненты кортежа. Таким образом, мы будем находиться в ситуации, весьма схожей со случаем $d = 2$. Единственное различие будет состоять в том, что порядок на первой координате может быть отличным от $<_1$.

Для определения нашей функции нам потребуется еще одно обозначение. Пусть $\alpha = (\alpha_1, \ldots, \alpha_d)$ – кортеж длины d из множества $[n]^d$ и $k \in [d]$. Напомним, что мы сопоставили с кортежем $\alpha = (\alpha_1, \ldots, \alpha_d)$ определенные порядки $<_{i_1}, \ldots, <_{i_d}$ (порядок $<_{i_k}$ – это порядок, в котором k-ая компонента α сравнивается с k-ыми компонентами других кортежей). Положим $\mathrm{odd}_k(\alpha) = 1$ если $\mathrm{num}_{i_k}(\alpha_k)$ нечетно и $\mathrm{odd}_k(\alpha) = 0$ в противном случае.

Для $X \subseteq [n]$ обозначим через $\max_i(X)$ порядковый номер максимального элемента X в порядке $<_i$, то есть $\max_i(X) = \max_{t \in X} \mathrm{num}_i(t)$.

Теперь мы можем определить нашу функцию. Ее значение на кортеже $(X_1, \ldots, X_d)$ зависит только от максимального кортежа из множества $X_1 \times \ldots \times X_d$.

Определение 3. *Пусть $X = (X_1, \ldots, X_d)$ – кортеж непустых подмножеств $[n]$. Пусть $\alpha = (\alpha_1, \ldots, \alpha_d)$ – максимальный кортеж в $X_1 \times \ldots \times X_d$ (в определенном выше порядке на множестве $[n]^d$). Мы определяем нашу функцию по следующей формуле:*

$$\mathrm{OMB}_d^D(X) = \mathrm{odd}_1(\alpha) \oplus \mathrm{odd}_2(\alpha) \oplus \ldots \oplus \mathrm{odd}_d(\alpha).$$

Если хотя бы одно из множеств X_i пусто, положим $\mathrm{OMB}_d^D(X) = 0$.

3 Представление OMB_d^D в виде ДНФ

В этом разделе мы изучим вопросы, связанные с представлением функции $\mathrm{OMB}_d^D(X)$ в виде ДНФ.

Сначала выпишем явно некоторое представление функции $\mathrm{OMB}_d^D(X)$ в виде ДНФ. Для кортежа $\alpha = (\alpha_1, \ldots, \alpha_d)$ из множества $[n]^d$ введем обозначение $\mathrm{odd}(\alpha) = \mathrm{odd}_1(\alpha) \oplus \mathrm{odd}_2(\alpha) \oplus \ldots \oplus \mathrm{odd}_d(\alpha)$. Тогда, нетрудно убедиться, что

$$\mathrm{OMB}_d^D(X) = \bigvee_{\alpha, \mathrm{odd}(\alpha)=1} \left(\left(\bigwedge_{k=1}^{d} x_{\alpha_k}^k \right) \wedge \left(\bigwedge_{k=1}^{d} \left(\bigwedge_{j:\ \alpha_k <_{i_k} j} \overline{x}_j^k \right) \right) \right). \qquad (2)$$

Действительно, в правой части записана дизъюнкция по всем кортежам α, таким что $\mathrm{odd}(\alpha) = 1$, конъюнктов, выражающих, что кортеж α является максимальным в $X_1 \times \ldots \times X_d$. Для того чтобы кортеж был максимальным достаточно выполнения двух условий. Во-первых, все переменные кортежа должны быть истинны, а во-вторых, в каждой координате k, все переменные большие α_k, в соответствующем этой компоненте порядке, должны быть ложны. Это и записано в каждом конъюнкте.

Заметим, что каждый конъюнкт в правой части (2) имеет размер не больше $d + dn = d(n+1)$, а всего конъюнктов столько же, сколько кортежей α с $\mathrm{odd}(\alpha) = 1$, то есть не более n^d.

Отсюда сразу получаем следующее утверждение.

Лемма 1. *При постоянном d функция $\mathrm{OMB}_d^D(X)$ представима в виде* ДНФ *полиномиального размера.*

Этот результат нельзя обобщить на произвольные d. То есть, для d, растущих с ростом n, функция $\mathrm{OMB}_d^D(X)$ уже не представима в виде ДНФ полиномиального размера. Для того чтобы доказать это нам потребуется новое определение. Оно необходимо только в этом разделе и не будет использоваться в других частях работы. Для простоты изложения, мы формулируем это определение и следующую лемму только для функции $\mathrm{OMB}_d^D(X)$. Хотя аналогичное определение осмысленно и лемма верна для любой функции.

Определение 4. *Множество $A \subseteq \{(X_1, \ldots, X_d) \mid \forall i, X_i \subseteq [n]\}$ называется трудным множеством для функции $\mathrm{OMB}_d^D(X)$, если выполнены два условия:*

1. *для всякого набора $X = (X_1, \ldots, X_d) \in A$ верно $\mathrm{OMB}_d^D(X) = 1$;*
2. *для любых различных $X^1, X^2 \in A$, $X^1 = (X_1^1, \ldots, X_d^1), X^2 = (X_1^2, \ldots, X_d^2)$ существует $X^0 = (X_1^0, \ldots, X_d^0)$, такой что $X_k^1 \cap X_k^2 \subseteq X_k^0 \subseteq X_k^1 \cup X_k^2$ для всякого k, и $\mathrm{OMB}_d^D(X^0) = 0$.*

Лемма 2. *Если для функции $\mathrm{OMB}_d^D(X)$ существует трудное множество A размера k, то всякое представление $\mathrm{OMB}_d^D(X)$ в виде* ДНФ *содержит не менее k конъюнктов.*

Доказательство. Пусть $\phi(X)$ – ДНФ, представляющая $\mathrm{OMB}_d^D(X)$. Тогда для любого $X = (X_1, \ldots, X_d) \in A$ существует истиный на X конъюнкт из $\phi(X)$. Для доказательства леммы достаточно убедиться, что

никакие два разных элемента $X^1 = (X^1_1, \ldots, X^1_d), X^2 = (X^2_1, \ldots, X^2_d) \in A$ не могут быть истины на одном и том же конъюнкте.

Предположим, что это не так, и X^1, X^2 истины на одном и том же конъюнкте ϕ. Тогда нетрудно понять, что этот конъюнкт не содержит переменных, на которых X^1 и X^2 отличаются, а именно, он не содержит переменных $X^1_k \triangle X^2_k$ для всякого k. Но тогда этот конъюнкт будет истинен и на всяком наборе $X^0 = (X^0_1, \ldots, X^0_d)$, таком что $X^1_k \cap X^2_k \subseteq X^0_k \subseteq X^1_k \cup X^2_k$ для всякого k. Однако, это противоречит второму пункту Определения 4. □

Теперь мы можем доказать, что для больших d функция $\mathrm{OMB}^D_d(X)$ не представима в виде ДНФ полиномиального размера.

Лемма 3. *Всякая* ДНФ, *представляющая функцию* $\mathrm{OMB}^D_d(X)$, *содержит не менее* $\left\lfloor \frac{n}{2} \right\rfloor^d$ *конъюнктов.*

Доказательство. Для всякого кортежа $\alpha = (\alpha_1, \ldots, \alpha_d)$ из множества $[n]^d$ определим кортеж $X^\alpha = (X^\alpha_1, \ldots, X^\alpha_d)$, где X^α_k – множество всех элементов $[n]$, не превышающих α_k в порядке, соответствующем k-й компоненте α.

По предыдущей лемме нам достаточно показать, что существует трудное множество A для функции $\mathrm{OMB}^D_d(X)$ размера $\left\lfloor \frac{n}{2} \right\rfloor^d$. Построить такое множество не сложно. Рассмотрим произвольный кортеж $\alpha = (\alpha_1, \ldots, \alpha_d)$ из множества $[n]^d$, такой что, если d – нечетно, то $\mathrm{odd}_k(\alpha) = 1$, для всякого k, а если d – четно, то $\mathrm{odd}_1(\alpha) = 0$ и $\mathrm{odd}_k(\alpha) = 1$, для всякого $k > 1$ (напомним, что d фиксировано в формулировке леммы). Таких кортежей как раз не менее $\left\lfloor \frac{n}{2} \right\rfloor^d$. Рассмотрим кортежи $X^\alpha = (X^\alpha_1, \ldots, X^\alpha_d)$, соответствующие всем таким α. Они и составят трудное множество A.

Осталось проверить, что A действительно является трудным множеством. Первый пункт определения следует непосредственно из построения. Для проверки второго, рассмотрим два набора $X^\alpha = (X^\alpha_1, \ldots, X^\alpha_d)$, $X^\beta = (X^\beta_1, \ldots, X^\beta_d) \in A$. Рассмотрим первую координату k, в которой кортежи α и β различаются. Поскольку на всех предыдущих координатах α и β совпадают, то порядки, в которых сравниваются их k-ые компоненты совпадают. А значит наборы X^α и X^β также совпадают в первых $k-1$ координатах, а в координате k либо $X^\alpha_k \subseteq X^\beta_k$, либо $X^\beta_k \subseteq X^\alpha_k$. Без ограничения общности, пусть $X^\alpha_k \subseteq X^\beta_k$. Если $\mathrm{odd}_k(\alpha) = \mathrm{odd}_k(\beta) = 1$ (это происходит, если либо d нечетно, либо d четно и $k > 1$), обозначим через

$\gamma_k \in [n]$ элемент, следующий за α_k в порядке $<_{i_k}$, где $<_{i_k}$ – порядок на k-ой компоненте кортежей α и β. Для γ_k верно, что $\alpha_k <_{i_k} \gamma_k <_{i_k} \beta_k$, причем γ_k расположено на четной позиции в порядке $<_{i_k}$, и порядок, который γ_k задает на $k+1$-ой компоненте такой же как и у α_k (последнее легко увидеть из рекурсивного правила (1) определения порядков на компонентах). Обозначим через γ кортеж $(\alpha_1, \ldots, \gamma_k, \ldots, \alpha_d)$. Тогда в качестве набора X^0 из второго пункта Определения 4 можно взять набор X^γ, так как $\mathrm{OMB}_d^D(X^\gamma) = 0$ из-за того, что γ_k стоит на четной позиции в k-ой компоненте, а остальные компоненты максимального кортежа X^γ такие же, как у α. Если же $\mathrm{odd}_k(\alpha) = \mathrm{odd}_k(\beta) = 0$ (это происходит, если d четно и $k = 1$), то обозначим через $\gamma_k \in [n]$ элемент, предшествующий β_k в порядке $<_{i_k}$. Вновь, $\alpha_k <_{i_k} \gamma_k <_{i_k} \beta_k$, γ_k расположено на нечетной позиции и задает тот же порядок на следующей компоненте, что и β_k. Обозначим через γ кортеж $(\beta_1, \ldots, \gamma_k, \ldots, \beta_d)$. В качестве X^0 можно взять набор X^γ. □

4 Основной результат

Лемма 4. *Функция* $\mathrm{OMB}_d^D(X)$ *вычислима персептроном степени* d.

Доказательство. По определению, кортеж $\alpha = (\alpha_1, \ldots, \alpha_d)$ принадлежит множеству $X_1 \times \ldots \times X_d$ тогда и только тогда, когда конъюнкция всех переменных $x^1_{\alpha_1}, \ldots, x^d_{\alpha_d}$ принимает значение 1. Обозначим эту конъюнкцию через $f_\wedge(\alpha)$.

Выпишем все кортежи из $[n]^d$ в порядке возрастания: $t_1, t_2, \ldots, t_{n^d}$. Тогда функция $\mathrm{OMB}_d^D(X)$ равна знаку суммы

$$\sum_j (-1)^{\mathrm{odd}_1(t_j)+\ldots+\mathrm{odd}_d(t_j)} \cdot 2^j \cdot f_\wedge(t_j).$$

Действительно, для всякого $t_j \notin X_1 \times \ldots \times X_d$ верно $f_\wedge(t_j) = 0$. Так что, только t_j, лежащие в $X_1 \times \ldots \times X_d$, дают вклад в указанную сумму. А коэффициенты 2^j выбраны так, что вклад наибольшего кортежа превышает вклад всех остальных кортежей. □

Следующая теорема является основным результатом работы.

Теорема 1. *Функция* $\mathrm{OMB}_d^D(X)$ *не вычислима никаким персептроном степени не выше* D *с общим весом* w *если верно*

$$D^5 \log n < \delta n, \tag{3}$$

$$w < 2^{\left(\frac{\varepsilon n}{D^4}\right)^d}. \tag{4}$$

Здесь ε, δ *– фиксированные положительные числа.*

Доказательство. Предположим, что персептрон p степени не выше D с общим весом w вычисляет нашу функцию и выполняется неравенство (3). Мы докажем, что неравенство (4) ложно.

В дальнейшем мы будем подставлять в p вместо переменных только 0 и 1, так что без ограничения общности можно предполагать, что p – мультилинейный многочлен (то есть, его степень по каждой переменной не превышает 1). Кроме того, мы будем отождествлять его входные переменные с кортежами длины d подмножеств $[n]$ и будем писать $p(X_1, \ldots, X_d)$, где $X_i \subseteq [n]$.

Заметим, что можно предполагать, что для всякого $(X_1, \ldots, X_d)$ верно $p(X_1, \ldots, X_d) \neq 0$. Действительно, если это не так, можно вместо многочлена p рассмотреть многочлен $2p - 1$, который всегда принимает нечетные значения, а значит, не обращается в 0. Это не меняет степень персептрона и увеличивает его общий вес не более чем в 3 раза, что, очевидно, не существенно для нашего результата.

Значение многочлена p при заданных булевых значениях переменных всегда не превосходит общего веса персептрона. Следовательно, для доказательства теоремы достаточно найти $X_1, \ldots, X_d$, такие что

$$|p(X_1, \ldots, X_d)| \geqslant 2^{\left(\frac{\varepsilon n}{D^4}\right)^d}.$$

Сначала мы опишем неформально построение таких $X_1, \ldots, X_d$ (Раздел 4.1), а затем перейдем к формальному доказательству. В конце работы мы сформулируем результаты, следующие из Теоремы 1.

4.1 План доказательства

Мы построим последовательность из $\Omega\left((\frac{\varepsilon n}{D^4})^d\right)$ кортежей длины d подмножеств $[n]$. При этом, значение p на почти всех кортежах этой последовательности будет превышать по абсолютной величине как минимум в

два раза значение p на предыдущем кортеже. Для небольшого числа кортежей в этой последовательности значение p все же может уменьшиться по сравнению со значением p на предыдущем кортеже, однако совокупное уменьшение будет несоизмеримо меньше совокупного увеличения. Поскольку абсолютная величина p на первом кортеже последовательности будет не меньше 1, значение p на последнем кортеже последовательности будет экспоненциально зависеть от длины последовательности, что и даст требуемый результат.

Как мы получим кортеж, на котором значение $|p|$ будет в два раза больше, чем на предыдущем кортеже? Для этого зафиксируем все множества $X_1, \ldots, X_{d-1}$ (и оставим свободными переменные X_d). Функция $\mathrm{OMB}_d^D(X_1, \ldots, X_d)$ при этом по существу превратится в функцию $\mathrm{OMB}_i(X_d)$, для некоторого $i \in [D+1]$, а полином p станет полиномом в переменных $x_1^d, \ldots, x_n^d$ степени не выше D. После этого, остается воспользоваться леммой Бейгеля из [3] (Лемма 5 ниже), которая утверждает, что если $\max_i(X_d)$ не очень близок к n, то можно изменить множество X_d, так что абсолютное значение многочлена, представляющего $\mathrm{OMB}_i(X_d)$, увеличивается не менее чем в 2 раза и при этом $\max_i(X_d)$ увеличивается не сильно. В этой лемме существенно, чтобы в определении $\mathrm{OMB}(X_d)$ и в определении $\max(X_d)$ использовался один и тот же порядок на $[n]$. Применяя лемму Бейгеля к порядку $<_i$, мы увеличиваем значение $|p|$ не менее чем в 2 раза.

Рассмотрим случай $d = 2$ более подробно. Начнем со входных множеств $X_1^0 = \{1\}, X_2^0 = \{1\}$. Сначала зафиксируем первую координату X_1^0. Из результата Бейгеля (Лемма 5 ниже) следует, что если максимальный элемент множества X_2^0 достаточно маленький (как в нашем случае), то можно $\Omega(n)$ раз сменить значение X_2, так что каждое изменение будет удваивать величину $|p(X_1^0, X_2)|$. Таким образом, существует X_2^1, такое что

$$|p(X_1^0, X_2^1)| \geqslant 2^{\Omega(n)} |p(X_1^0, X_2^0)|.$$

Теперь мы фиксируем вторую координату X_2^1 и меняем X_1 так, что (а) значение $|p|$ уменьшается не сильно и (б) в новом порядке на второй компоненте кортежа максимальный элемент множества X_2^1 вновь становится маленьким. Таким образом, мы снова можем применить Лемму 5 и получить X_2^2, такое что

$$|p(X_1^1, X_2^2)| \geqslant 2^{\Omega(n)} |p(X_1^1, X_2^1)|.$$

Мы покажем, что эту процедуру можно повторить $\Omega(n)$ раз. Так что, в конце мы получим множества $X_1^{\Omega(n)}, X_2^{\Omega(n)}$, такие что

$$|p(X_1^{\Omega(n)}, X_2^{\Omega(n)})| \geqslant 2^{\Omega(n^2)}|p(X_1^0, X_2^0)| \geqslant 2^{\Omega(n^2)}.$$

Наш анализ покажет, что константа в $\Omega(n^2)$ окажется порядка $\Omega(D^{-8})$ и, таким образом, мы получим $w \geqslant |p(X_1^{\Omega(n)}, X_2^{\Omega(n)})| \geqslant 2^{\Omega(n^2/D^8)}$.

Как изменяются рассуждения в случае $d = 3$? Мы сначала фиксируем первую координату $X_1^0 = \{1\}$. Функция $\mathrm{OMB}_d^D(X_1^0, X_2, X_3)$ полностью совпадает с уже рассмотренной нами функцией от двух переменных, а именно с функцией $\mathrm{OMB}_d^D(X_2, X_3)$. Так что, мы можем воспользоваться неравенством, доказанным нами для случая $d = 2$. После этого мы фиксируем вторую и третью координаты и изменяем первую так, чтобы (а) значение $|p|$ уменьшилось не сильно и (б) мы могли снова применить неравенство для случая $d = 2$ ко второй и третьей компонентам. Мы докажем, что эту процедуру можно будет повторить $2^{\Omega(n)}$ раз.

Таким образом, мы будем вести доказательство индукцией по d (называемом k в ключевой Лемме 12).

Перейдем к формальному доказательству.

4.2 Увеличение значения $|p|$

Сначала сформулируем необходимую нам лемму, по существу доказанную в [3]. Напомним, что мы отождествляем наборы булевых значений переменных $x_1, \ldots, x_n$ и подмножества X множества $[n]$ и используем для функций и многочленов обозначение $f(X)$ вместо $f(x_1, \ldots, x_n)$.

Лемма 5 ([3]). *Существует положительная константа γ, такая что верно следующее. Пусть $f(X)$ – многочлен с целыми коэффициентами степени не выше D от $m = \gamma D^2$ переменных. Предположим, что для всякого непустого $M \subseteq [m]$ верно, что $sgn(f(M)) \neq sgn(f(\emptyset))$. Тогда существует $M \subseteq [m]$, такое что*

$$|f(M)| \geqslant 2|f(\emptyset)|.$$

Эта лемма позволяет нам увеличивать $|p|$ за счет увеличения размеров множеств $X_1, \ldots, X_d$.

4.3 Уменьшение размера входа

В этом разделе нам потребуется операция симметризации многочленов от нескольких переменных. Пусть $f(x)$ – многочлен от m переменных $x = (x_1, \ldots, x_m)$. Пусть σ – некоторая перестановка на множестве $[m]$, обозначим через $\sigma(x) = (x_{\sigma(1)}, \ldots, x_{\sigma(m)})$. Обозначим через Σ_m множество всех перестановок на множестве $[m]$. Тогда симметризацией $f^{sym}(x)$ многочлена $f(x)$ назовем многочлен

$$f^{sym}(x) = \frac{\sum_{\sigma \in \Sigma_m} f(\sigma(x))}{m!}.$$

Следующая классическая лемма из [13] позволяет сводить анализ многочленов от нескольких переменных к анализу многочленов от одной переменной.

Лемма 6 ([13]). *Пусть $f(x)$ – мультилинейный многочлен степени D от m переменных $x = (x_1, \ldots, x_m)$. Тогда существует многочлен $g(y)$ от одной переменной степени не выше D, такой что для всякого $x \in \{0,1\}^m$ верно $f^{sym}(x) = g(\sum_k x_k)$.*

Теперь мы можем доказать основную лемму этого раздела. Она позволит нам уменьшать размер X_i, при этом уменьшая $|p(X)|$ не слишком сильно.

Лемма 7. *Пусть $f(X)$ – многочлен степени не выше D от m переменных. Тогда существует набор значений переменных $X^0 \subseteq [m]$, такой что $|X^0| \leqslant D$ и*

$$|f(X^0)| > \frac{|f([m])|}{(D+1)m^D}.$$

Доказательство. Предположим, что для всякого подмножества $X^0 \subseteq [m]$, такого что $|X^0| \leqslant D$, верно $|f(X^0)| \leqslant \frac{|f([m])|}{(D+1)m^D}$.

Рассмотрим симметризацию $f^{sym}(X)$ нашего многочлена $f(X)$. Заметим, что $f^{sym}([m]) = f([m])$ и $|f^{sym}(X_0)| \leqslant \max_{|X|=|X_0|} |f(X)|$. Тогда по Лемме 6 существует многочлен $g(y)$ степени не выше D, где $y \in \mathbb{R}$, такой что

1. $|g(y)| \leqslant \frac{|f([m])|}{(D+1)m^D}$ при $y = 0, 1, \ldots, D$;
2. $|g(m)| = |f([m])|$.

Поскольку многочлен степени D однозначно определяется своими значениями в $D+1$ точке, мы можем явно выписать многочлен $g(y)$ через его значения $g(0), g(1), \ldots, g(D)$, как интерполяционный многочлен Лагранжа:

$$g(y) = g(0)\frac{(y-1)(y-2)\ldots(y-D)}{-1(-2)\ldots(-D)} + g(1)\frac{y(y-2)\ldots(y-D)}{1(-1)\ldots(1-D)} + \ldots + g(D)\frac{y(y-1)\ldots(y-D+1)}{D(D-1)\ldots 1}.$$

Теперь оценим величину $|g(m)|$. Каждый член в сумме, выписанной выше, при $y = m$ по модулю меньше чем $m^D \cdot \max_{0 \leqslant i \leqslant D} |g(i)|$. Следовательно,

$$|g(m)| < (D+1)m^D \cdot \max_{0 \leqslant i \leqslant D} |g(i)| \leqslant (D+1)m^D \frac{|f([m])|}{(D+1)m^D} = |g(m)|.$$

Мы получили противоречие, и лемма доказана. □

4.4 Доказательство основного результата

С помощью лемм из предыдущих разделов мы теперь докажем серию лемм о функции $\mathrm{OMB}_d^D(X)$.

Сначала покажем, как применять Лемму 5 к функции $\mathrm{OMB}_i(X)$.

Лемма 8. *Пусть дана функция* $\mathrm{OMB}_i(X)$, *где* $X \subseteq [n]$ *и* $i \in [D+1]$, *и пусть многочлен* $f(X)$ *степени не выше* D *представляет* $\mathrm{OMB}_i(X)$ *своим знаком. Пусть набор значений входных переменных* X^0 *таков, что* $max_i(X^0) < n - 2\gamma D^2$, *где* γ *– константа из Леммы 5. Тогда, существует набор значений входных переменных* X^1, *такой что* $\max_i(X^1) < \max_i(X^0) + 2\gamma D^2$, *и*

$$|f(X^1)| \geqslant 2|f(X^0)|.$$

Доказательство. Мы хотим воспользоваться Леммой 5. Для этого нам нужно найти подмножество $B \subseteq [n]$ из γD^2 переменных, которые мы оставим свободными, остальные переменные мы зафиксируем. При этом мы хотим, чтобы получившийся из f многочлен удовлетворял условиям Леммы 5.

Пусть s – порядковый номер $\max_i(X^0)$ в порядке $<_i$. Тогда в качестве B возьмем множество, состоящее из элементов с порядковыми номерами $s+1, s+3, \ldots, s+2\gamma D^2-1$ в порядке $<_i$. По условию леммы, $max_i(X^0) < n - 2\gamma D^2$, а значит множество B определено корректно.

Все элементы множества B больше $\max_i(X^0)$ в порядке $<_i$, и при этом позиция всякого элемента множества B имеет другую четность, по сравнению с $\max_i(X^0)$. Следовательно, для всякого непустого $M \subseteq B$ по определению $\mathrm{OMB}_i(X)$ имеем $\mathrm{OMB}(X^0 \cup M) \neq \mathrm{OMB}(X^0)$. Таким образом, для многочлена $f(X^0 \cup M)$ от переменных $M \subseteq B$ выполняются условия Леммы 5, и существует $M \subseteq B$, такое что $|f(X^0 \cup M)| \geqslant 2|f(X^0)|$. Осталось заметить, что номер максимального элемента нового множества $X^1 = X^0 \cup M$ в порядке $<_i$ не превышает $s + 2\gamma D^2 - 1$, а значит $\max_i(X^1) < \max_i(X^0) + 2\gamma D^2$. □

Напомним, что $p(X_1, \ldots, X_d)$ – многочлен представляющий функцию $\mathrm{OMB}_d^D(X_1, \ldots, X_d)$ своим знаком.

Теперь мы докажем лемму, аналогичную Лемме 8, для функции $\mathrm{OMB}_d^D(X)$.

Лемма 9. *Пусть $X_1^0, \ldots, X_d^0$ – кортеж подмножеств $[n]$. Пусть $\alpha = (\alpha_1, \ldots, \alpha_d)$ – максимальный кортеж в множестве $X_1^0 \times \ldots \times X_d^0$. Пусть $<_i$ – порядок, используемый для сравнения на d-ой компоненте α. Предположим, что $\mathrm{num}_i(\alpha_d) \leqslant n - 2\gamma D^2$, где γ – константа из Леммы 5. Тогда, существует X_d^1, такой что $\max_i(X_d^1) < \max_i(X_d^0) + 2\gamma D^2$, и*

$$|p(X_1^0, \ldots, X_{d-1}^0, X_d^1)| \geqslant 2|p(X_1^0, \ldots, X_{d-1}^0, X_d^0)|.$$

Замечание 1. *В действительности, аналогичное утверждение можно доказать для произвольной координаты, а не только для координаты d, но нам такое утверждение не потребуется, так что для простоты изложения мы рассмотрим только случай координаты d.*

Доказательство Леммы 9. Заметим, что функция $\mathrm{OMB}_d^D(X_1^0, \ldots, X_{d-1}^0, X)$ от переменных $X \subseteq [n]$ либо совпадает с $\mathrm{OMB}_i(X)$, либо совпадает с ее отрицанием. Следовательно, либо многочлен $f(X) = p(X_1^0, \ldots, X_{d-1}^0, X)$, либо многочлен $f(X) = -p(X_1^0, \ldots, X_{d-1}^0, X)$ представляет функцию $\mathrm{OMB}_i(X)$ своим знаком. Применяя Лемму 8 к многочлену $f(X)$, мы получаем требуемый результат. □

Теперь мы покажем, как применять Лемму 7 к функции $\mathrm{OMB}_d^D(X)$.

Лемма 10. *Пусть $X_1^0, \ldots, X_d^0$ – кортеж подмножеств $[n]$, и пусть $k \in [d]$ – некоторая координата кортежа. Тогда, существует множество X_k^1, такое что $|X_k^1| \leqslant D$ и*

$$|p(X_1^0, \ldots, X_k^1, \ldots, X_d^0)| \geqslant 2^{-(D+1)\log n} |p(X_1^0, \ldots, X_k^0, \ldots, X_d^0)|.$$

Доказательство. Рассмотрим многочлен

$$f(X) = p(X_1^0, \ldots, X_{k-1}^0, X, X_{k+1}^0, \ldots, X_d^0),$$

где $X \subseteq X_k^0$. Применив к этому многочлену Лемму 7, получаем множество $X_k^1 \subseteq X_k^0$, такое что $|X_k^1| \leqslant D$ и

$$|f(X_k^1)| \geqslant \frac{|f(X_k^0)|}{(D+1)|X_k^0|^D} \geqslant \frac{|f(X_k^0)|}{n \cdot n^D} \geqslant 2^{-(D+1)\log n} |f(X_k^0)|,$$

где второе равенство следует из того, что $D+1 \leqslant n$ (по неравенству (3)) и $|X_k^0| \leqslant n$. □

Теперь мы покажем, как менять порядок на заданной компоненте максимального элемента входного множества.

Лемма 11. *Пусть $X_1^0, \ldots, X_d^0$ – кортеж подмножеств $[n]$. Пусть $\alpha = (\alpha_1, \ldots, \alpha_d)$ – максимальный кортеж в множестве $X_1^0 \times \ldots \times X_d^0$. Пусть $<_i$ – порядок, используемый для сравнения k-ой компоненты α. Предположим, что $\mathrm{num}_i(\alpha_k) \leqslant n - 2\gamma D^2(D+1)$ и $|X_{k+1}^0| \leqslant D$. Тогда существует множество X_k^1, такое что выполняются следующие условия:*

1. $\max_i(X_k^1) < \max_i(X_k^0) + 2\gamma D^2(D+1)$.
2. *Пусть l – номер порядка, который задает $\max_i(X_k^1)$ на $k+1$-ой компоненте кортежа. Тогда $\max_l(X_{k+1}^0) \leqslant \frac{Dn}{D+1}$.*
3. $|p(X_1^0, \ldots, X_k^1, \ldots, X_d^0)| \geqslant |p(X_1^0, \ldots, X_k^0, \ldots, X_d^0)|$.

Доказательство. Поскольку размер множества X_{k+1}^0 не превышает D, существует порядок $<_l$, такой что порядковый номер максимального элемента X_{k+1}^0 в этом порядке не превышает $nD/(D+1)$. Действительно, последнее означает, что X_{k+1}^0 не пересекается с множеством из $n/(D+1)$ самых больших элементов в этом порядке. А наши порядки подобраны

таким образом, что множества из $n/(D+1)$ максимальных элементов попарно различны для всех $D+1$ порядков. А значит, никакое D-элементное множество не может пересекать каждое из них.

Рассмотрим многочлен

$$f(X) = p(X_1^0, \ldots, X, X_{k+1}^0, \ldots, X_d^0),$$

где $X \subseteq [n]$. Добавлением к X_k^0 лишь одного элемента из $2(D+1)$, следующих за максимальным элементом X_k^0 в порядке $<_i$, мы можем изменить порядок на $k+1$-ой компоненте, так чтобы новый порядок был порядком $<_l$ из предыдущего абзаца. Однако такое изменение X_k^0, вообще говоря, может сильно уменьшить $p(X_1^0, \ldots, X_d^0)$, чего мы хотим избежать.

Так что, нам снова потребуется воспользоваться Леммой 5. Нам нужно определить множество B из γD^2 переменных, которые мы оставим свободными. Это множество нужно определять аккуратно, так как мы хотим, чтобы после замены входного множества новый порядок на $k+1$-ой компоненте был $<_l$.

Обозначим через U множество из $2\gamma D^2(D+1)$ чисел, следующих за максимальным элементом X_k^0 в порядке $<_i$:

$$U = \{x | x \in [n],\ \mathrm{num}_i(\alpha_k) + 1 \leqslant \mathrm{num}_i(x) \leqslant \mathrm{num}_i(\alpha_k) + 2\gamma D^2(D+1)\}.$$

Заметим, что правило (1) для определения порядка на $k+1$-ой компоненте периодично с периодом $2(D+1)$. В каждом периоде всякий порядок встречается дважды: один раз на четной позиции и один раз на нечетной. Следовательно, U содержит ровно $2\gamma D^2$ элементов, задающих порядок $<_l$ на $k+1$-ой компоненте, обозначим через U_l множество соответствующих номеров. Ровно половина чисел в U_l находится на четных позициях. Обозначим через B_0 множество этих номеров, а через $B_1 = U_l \setminus B_0$.

Мы утверждаем, что условия Леммы 5 выполняются либо для B равного B_0, либо для B равного B_1. Действительно, все элементы в U_l больше чем $\max_i(X_k^0)$. Следовательно, для всякого непустого $M \subseteq U_i$ верно $\max_i(X_k^0 \cup M) = \max_i(M)$. Далее, все компоненты максимального элемента в множестве $X_1^0 \times \cdots \times (X_k^0 \cup M) \times \cdots \times X_d^0$, кроме k-ой, не зависят от M. Действительно, его первые $k-1$ компонент совпадают с соответствующими компонентами максимального элемента множества $X_1^0 \times \cdots \times X_k^0 \times \cdots \times X_d^0$, а все элементы U_l задают одинаковый порядок на $k+1$-ой компоненте. Отсюда следует, что функция $\mathrm{OMB}_d^D(X_1^0, \ldots, X_k^0 \cup M, \ldots, X_d^0)$, как функция от переменных M равна либо $\mathrm{OMB}_i(X_k^0 \cup M)$,

либо ее отрицанию. Без ограничения общности предположим, что она равна $\mathrm{OMB}_i(X_k^0 \cup M)$. Тогда для всякого непустого $M \subseteq U_l$ знак $f(X_k^0 \cup M) = p(X_1^0, \ldots, X_k^0 \cup M, X_{k+1}^0, \ldots, X_d^0)$ совпадает со знаком $\mathrm{OMB}_i(X_k^0 \cup M)$. Функция $\mathrm{OMB}_i(X_k^0 \cup M)$ зависит лишь от четности номера максимального элемента M. Следовательно, знак $f(X_k^0 \cup M)$ постоянен для всякого непустого $M \subseteq B_0$ и то же самое верно для подмножеств B_1. Кроме того, для подмножеств B_1 знак $f(X_k^0 \cup M)$ другой, чем для подмножеств B_0. Положим $B = B_0$, если знак $f(X_k^0 \cup M)$ не равен знаку $f(X_k^0)$ для непустых $M \subseteq B_0$ и положим $B = B_1$ в противном случае.

Теперь мы можем применить Лемму 5 и получить $M \subseteq B$, такое что $|f(X_k^0 \cup M)| \geqslant 2|f(X_k^0)|$. Взяв $X_k^1 = X_k^0 \cup M$, мы завершаем доказательство леммы (мы получаем немного больше чем требуется, так как значение $|p|$ не просто не уменьшается, а даже удваивается). □

С помощью трех доказанных лемм о функции $\mathrm{OMB}_d^D(X)$ мы можем теперь доказать ключевую лемму.

Лемма 12. *Пусть $(X_1^0, \ldots, X_d^0)$ – кортеж подмножеств $[n]$. Пусть $\alpha = (\alpha_1, \ldots, \alpha_d)$ – максимальный кортеж в множестве $X_1^0 \times \ldots \times X_d^0$. Пусть $<_i$ – порядок, используемый для сравнения k-ой компоненты α. Предположим, что $\mathrm{num}_i(\alpha_k) \leqslant \frac{Dn}{D+1}$. Тогда существуют $X_k', X_{k+1}', \ldots, X_d'$, такие что*

$$|p(X_1^0, \ldots, X_k', \ldots, X_d')| \geqslant 2^{\left(\frac{\varepsilon n}{D^4}\right)^{d-k+1}} |p(X_1^0, \ldots, X_k^0, \ldots, X_d^0)|.$$

Здесь ϵ – некоторая положительная константа.

Доказательство. Доказательство проходит индукцией по убывающему k.

База индукции $k = d$. В этом случае неравенство получается посредством $\varepsilon n / D^3$ применений Леммы 9. Поскольку вначале $\max_i(X_d^0) \leqslant \frac{Dn}{D+1}$, а каждое применение Леммы 9 увеличивает $\max_i(X_d^0)$ не более чем на $2\gamma D^2$, мы можем сделать $n/(D+1)2\gamma D^2$ повторений. В результате мы получаем X_d', такое что $|f(X_d')| \geqslant 2^{\Omega(n/D^3)}|f(X_0^d)|$ (даже с меньшей степенью 3 в D^3, чем требуемое 4).

Шаг индукции. Для доказательства леммы в случае $k < d$ мы многократно повторяем следующие три шага.

Шаг 1. Применим к $k+1$-ой координате Лемму 10. В результате мы получаем множество $\widetilde{X}^0_{k+1}$ с не более чем D элементами и

$$|p(X^0_1, \ldots, \widetilde{X}^0_{k+1}, \ldots, X^0_d)| \geqslant 2^{-(D+1)\log n} |p(X^0_1, \ldots, X^0_k, X^0_{k+1}, \ldots, X^0_d)|.$$

Шаг 2. Применим Лемму 11 к входным множествам $X^0_1, \ldots, X^0_k, \widetilde{X}^0_{k+1}, \ldots, X^0_d$ и компоненте k. В результате мы получаем множество X^1_k, такое что:

1. $\max_i(X^1_k) < \max_i(X^0_k) + 2\gamma D^2(D+1)$.
2. Пусть l – порядок, который задает $\max_i(X^1_k)$ на $k+1$-ой компоненте кортежа. Тогда $\max_l(\widetilde{X}^0_{k+1}) \leqslant \frac{Dn}{D+1}$.
3. $|p(X^0_1, \ldots, X^1_k, \widetilde{X}^0_{k+1}, \ldots, X^0_d)| \geqslant |p(X^0_1, \ldots, X^0_k, \widetilde{X}^0_{k+1}, \ldots, X^0_d)|$.

Шаг 3. Заметим, что после Шага 2 кортеж $(X^0_1, \ldots, X^1_k, \widetilde{X}^0_{k+1}, \ldots, X^0_d)$ удовлетворяет предположению индукции для $k+1$. Пользуясь предположением индукции, мы получаем $X^1_{k+1}, X^1_{k+2}, \ldots, X^1_d$, такие что

$$|p(X^0_1, \ldots, X^1_k, \ldots, X^1_d)| \geqslant 2^{\left(\frac{\varepsilon n}{D^4}\right)^{d-k}} |p(X^0_1, \ldots, X^1_k, X^0_{k+1} \ldots, X^0_d)|. \qquad (5)$$

Как первые k компонент максимального элемента $X_1 \times \cdots \times X_d$ изменились после Шагов 1, 2 и 3? Первые $k-1$ компонента остались незатронутыми, а k-ая компонента увеличилась на не более чем $2\gamma D^2(D+1)$ (на втором шаге) в порядке $<_i$. В начале, по условию леммы, порядковый номер k-ой компоненты был меньше, чем $\frac{nD}{D+1}$. Таким образом, мы можем повторять эти три шага $\frac{n}{2\gamma D^2(D+1)^2}$ раз. Выберем ε так, что $\frac{n}{2\gamma D^2(D+1)^2} \geqslant \frac{2\varepsilon n}{D^4}$. Тогда, число повторений будет не меньше $\frac{2\varepsilon n}{D^4}$.

В течении каждого повторения $|p|$ сначала уменьшается в не более чем $2^{(D+1)\log n}$ раз на Шаге 1, а затем увеличивается в $2^{(\frac{\varepsilon n}{D^4})^{d-k}}$ раз на Шаге 3. Мы можем предполагать, что постоянная δ в неравенстве (3) достаточно мала, чтобы экспонента в первом множителе была меньше, чем $\varepsilon n/2D^4$. Тогда, после каждого повторения величина $|p|$ увеличивается в

$$2^{(\frac{\varepsilon n}{D^4})^{d-k} - \frac{\varepsilon n}{2D^4}} \geqslant 2^{(\frac{\varepsilon n}{D^4})^{d-k}/2}.$$

раз. Возводя этот множитель в степень, равную числу повторений, мы получаем необходимый множитель $2^{(\frac{\varepsilon n}{D^4})^{d-k+1}}$. □

Теперь мы можем завершить доказательство теоремы. Положим в Лемме 12 $X_1^0 = \ldots = X_d^0 = \{1\}$ и $k = 1$. Условия леммы выполняются, и следовательно, существуют $X_1', \ldots, X_d'$, такие что

$$|p(X_1', \ldots, X_d')| \geqslant 2^{\left(\frac{\varepsilon n}{D^4}\right)^d} |p(\{1\}, \ldots, \{1\})| \geqslant 2^{\left(\frac{\varepsilon n}{D^4}\right)^d}.$$

□

Выбирая параметры в Теореме 1, мы получаем следующие результаты.

Теорема 2. *Пусть $D(n) = n^{o(1)}$. Тогда для всякого $\epsilon > 0$ функция $\mathrm{OMB}_d^{D(n)}(X)$ вычислима персептроном степени d, но невычислим никаким персептроном степени не выше $D(n)$ с общим весом $2^{n^{d(1-\epsilon)}}$.*

Теорема 3. *Пусть d и D – константы. Тогда функция $\mathrm{OMB}_d^D(X)$ вычислима персептроном степени d, и требует общего веса $2^{\Omega(n^d)}$ при вычислении персептронами степени не выше D.*

Сформулируем также следствие Теоремы 1 и Леммы 2, касающееся функций, представимых в виде полиномиальных ДНФ.

Теорема 4. *Для всякого постоянного d и $D \leqslant \frac{\varepsilon n^{1/5}}{\log n}$, где ε – некоторая положительная константа, существует (явно заданная) функция от n переменных, представимая в виде полиномиальной* ДНФ*, и требующая общего веса $2^{\Omega\left(n^d/D^{4d}\right)}$ при вычислении персептронами степени не выше D.*

Благодарности.

Автор выражает благодарность Николаю Константиновичу Верещагину и Гарри Бурману за привлечение внимания автора к изучавшейся задаче.

Список литературы

[1] E. Allender. Circuit complexity before the dawn of the new Millennium. *Proc. 16th Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, 1180, 1-18, 1996.

[2] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135-148, 1994.

[3] R. Beigel. Perceptrons, PP and the polynomial hierarchy. *Computational Complexity*, 4:339-349, 1994.

[4] R. Beigel, N. Reingold, and D. A. Spielman. PP Is closed under intersection. *Journal of Computer and System Sciences*, 50(2):191-202, 1995.

[5] A. Chandra, L. Stockmeyer, and U. Vishkin. Constant depth reducibility. *SIAM Journal on Computing*, 13:423-439, 1984.

[6] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H. Simon. Relations between communication complexity, linear arrangements, and computational complexity. *Proc. of the 21st Conf. on Foundations of Software Technology and Theoretical Computer Science (FST TCS)*, 2001, pages 171–182.

[7] M. Goldmann, J. Håstad, and A. A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277-300, 1992.

[8] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *Journal on Computer and System Science*, 46:129-154, 1993.

[9] J. Håstad. Almost optimal lower bounds for small depth circuits. In Randomness and Computation, Advances in Computing Research, Vol 5, ed. S. Micali, 1989, JAI Press Inc, pp 143-170.

[10] J. Håstad. On the size of weights for threshold gates. *SIAM Journal on Discrete Mathematics*, 7(3):484-492, 1994

[11] A. R. Klivans, and R. A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *Journal of Computer and System Sciences* 68(2):303-318, 2004.

[12] A. R. Klivans, and A. A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2-3):97–114, 2007.

[13] M. L. Minsky and S. A. Papert. *Perceptrons.* MIT Press, Cambridge, MA, 1968.

[14] S. Muroga. *Threshold logic and its applications.* Wiley-Interscience, 1971.

[15] J. Myhill and W. H. Kautz. On the size of weights required for linear-input switching functions. *IRE Trans. on Electronic Computers*, EC10(2):288-290, 1961.

[16] N. Nisan. The communication complexity of threshold gates. *Combinatorics, Paul Erdös is Eighty (Vol. 1)*, D Miklós, V. T. Sós, and T. Szönyi, Eds., Jason Bolyai Math. Society: Budapest, Hungary, 1993, 301-315.

[17] R. O'Donnell, and R. A. Servedio. New degree bounds for polynomial threshold functions. *35th Annual Symposium on Theory of Computing (STOC)*, 2003, pages 325–334.

[18] V. V. Podolskii. A uniform lower bound on weights of perceptrons. Computer Science - Theory and Applications *Proc. 3rd Int. Sympos. on Computer Science in Russia, CSR 2008. Moscow, Russia. June 7-12, 2008*, Lecture Notes in Computer Science. V. 5010. Berlin: Springer, 2008. P. 261-272.

[19] A. Razborov. Lower bounds on the size of bounded-depth networks over a complete basis with logical addition. *Matematicheskie Zametki*, Vol. 41, No 4, 1987, pages 598-607. English translation in Mathematical Notes of the Academy of Sci. of the USSR, 41(4):333-338, 1987.

[20] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. *Proc. of the 39th Symposium on Theory of Computing (STOC)*, 2007, pages 294–301.

[21] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. *Proc. of the 40th Symposium on Theory of Computing (STOC)*, 2008, pages 85–94.

[22] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings, 19th ACM Symposium of Theory of Computing*, pages 77-82, 1987.

[23] A. Yao. Separating the polynomial-time hierarchy by oracles. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1-10, 1985.

[24] В. В. Подольский. Перцептроны с большим весом. *Проблемы передачи информации*, 45(1):51-59, 2009.