Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Математическая логика и теория алгоритмов, осень 2021
Лекция 8: метод резолюций

Краткое содержание

Конъюнктивные нормальные формы (КНФ). Задача о выполнимости. Понимание КНФ как набора дизъюнктов. Пустой дизъюнкт. Приведение произвольной пропозициональной формулы к виду 3-КНФ, чья выполнимость эквивалентна выполнимости исходной формулы. Примеры задач, сводящихся к выполнимости КНФ: задача о 3-раскраске вершин графа, задача о расстановке ферзей на шахматной доске. Приведение КНФ к виду 3-КНФ, чья выполнимость эквивалентна выполнимости исходной КНФ. Метод резолюций. Пример использования. Доказательство корректности и полноты метода резолюций. Доказательство невозможности расстановки 3 ферзей на доске 3×3 методом резолюций. Выводы тавтологий через метод резолюций. Пример для закона контрапозиции.

В исчислении высказываний нашей целью было создание системы вывода всех тавтологий и только их (первое свойство означает полноту, второе — корректность). Теперь поставим другую задачу: создать систему опровержения всех противоречий и только их. Оказывается, такую задачу удобнее решать не для произвольных формул, а для конъюнктивных нормальных форм. Более того, для непротиворечий (т. е. выполнимых формул) получится не только проверить выполнимость, но и найти выполняющий набор.

1 Выполнимость КНФ

Напомним цепочку определений. Литералом называется переменная или отрицание переменной. Дизъюнктом называется дизъюнкция литералов. Конъюнктивной нормальной формой (КНФ) называется конъюнкция дизъюнктов. Мы будем предполагать, что в каждом дизъюнкте каждая переменная участвует не больше одного раза. Действительно, повтор того же литерала не изменит значения дизъюнкта, а если в дизъюнкте есть и переменная, и её отрицание, то этот дизъюнкт тождественно истинен и потому его наличие или отсутствие не изменит значения всей формулы. Кроме того, на КНФ удобно смотреть как на набор дизъюнктов, ведь порядок их следования не важен. Введём специальное обозначение $\perp$ для пустого дизъюнкта, т. е. не содержащего ни одного литерала. Ясно, что такой дизъюнкт будет тождественно ложным и потому КНФ, содержащая такой дизъюнкт, также будет тождественно ложной.

Мы знаем, что у каждой формулы есть эквивалентная ей КНФ. Однако такая КНФ может быть очень длинной, экспоненциальной длины от числа переменных, притом что сама формула может быть короткой. Если нас интересует лишь выполнимость, то можно получить короткую формулу, эквивалентную исходной в следующем слабом смысле:

Утверждение 1. *Для любой формулы φ существует формула φ', такая что:*

а) Множество переменных φ' включает в себя множество переменных φ;

б) φ' является 3-КНФ, т. е. КНФ, в каждом дизъюнкте которой не больше 3 литералов;

в) Длина φ' в константу раз больше длины φ;

г) Если φ невыполнима, то φ' также невыполнима;

д) Если φ выполнима, то φ' также выполнима, более того, ограничение любого выполняющего набора φ' на общие переменные даст выполняющий набор φ.

Доказательство. Идея доказательства следующая: введём новую переменную для каждой подформулы φ. Запишем формулы, означающие, что эти переменные верно вычисляются, соберём из них КНФ и добавим переменную, отвечающую за всю формулу. Например, для формулы $\neg(p \to q) \wedge (r \vee s)$ введём новые переменные t, u, v, w и потребуем одновременной истинности формул $t \leftrightarrow (p \to q)$, $u \leftrightarrow \neg t$, $v \leftrightarrow (r \vee s)$, $w \leftrightarrow (u \wedge v)$ и w. Если для данных значений p, q, r, s исходная формула была истинна, то значения всех остальных переменных можно установить по таблице истинности, и w также будет истинной. Если же исходная формула была ложной, то будет одно из двух: либо все остальные переменные вычислены верно, но $w = 0$, либо хотя бы в одной промежуточной формуле ошибка. Осталось все промежуточные формулы привести в КНФ и взять общую конъюнкцию. Например, $v \leftrightarrow (r \vee s)$ эквивалентно $(r \vee s \vee \neg v) \wedge (\neg r \vee v) \wedge (\neg s \vee v)$. Такую формулу можно получить стандартным способом, но у неё есть и явный смысл с учётом того, что $\neg A \vee B$ равносильно $A \to B$. Она эквивалентна выполнению трёх условий: $(r \to v) \wedge (s \to v) \wedge (v \to (r \vee s))$.

В общем случае рецепт точно такой же: надо завести переменные $z_1, \dots, z_n$, соответствующие как исходным переменным, так и всем нетривиальным подформулам, в том числе и φ целиком (Пусть это будет z_n). Для каждой переменной z_k, соответствующей подформуле $(z_i * z_j)$ (или $\neg z_i$), построим КНФ, эквивалентную формуле $z_k \leftrightarrow (z_i * z_j)$ (или $z_k \leftrightarrow \neg z_i$). Соберём φ' как конъюнкцию всех таких подформул и, дополнительно, переменной z_n. Проверим справедливость всех условий:

а) По построению.

б) В каждой формуле $z_k \leftrightarrow (z_i * z_j)$ (или $z_k \leftrightarrow \neg z_i$) не больше трёх переменных, значит, получится 3-КНФ.

в) Каждой операции в φ соответствует некоторая формула константной длины в φ'. Значит, длина φ' в константу раз больше длины φ.

г) Если $\varphi \equiv 0$, то при любых значениях исходных переменных если все эквиваленции выполнены, то $z_n = 0$. Значит, хотя бы один дизъюнкт не выполнен.

д) Если же при некоторых значениях исходных переменных $\varphi = 1$, то значения остальных переменных можно получить из эквиваленций, и получится $z_n = 1$, т. е. формула φ' выполнена. Напротив, если есть выполняющий набор φ', то на его ограничении на исходные переменные φ должна быть верна.

$\square$

Отметим также, что описанное преобразование φ в φ' можно выполнить алгоритмически за полиномиальное время.

Впрочем, во многих случаях задача изначально формализуется как выполнимость КНФ. Приведём два примера:

- *Задача о 3-раскраске вершин графа.* Пусть задан некоторый неориентированный граф. Ставится вопрос: можно ли его вершины раскрасить в 3 цвета, так чтобы вершины одного цвета не были соединены ребром? КНФ строится так: для каждой вершины i заводится две переменных p_i и q_i. Будем считать, что пара значений $(0,1)$ кодирует первый цвет, пара $(1,0)$ второй цвет, а пара $(1,1)$ третий цвет. Чтобы исключить вариант $(0,0)$, добавим условия $(p_i \vee q_i)$. Далее, для каждого ребра (i,j) пара (p_i,q_i) должна отличаться от пары (p_j,q_j). Это выражается такой КНФ: $(p_i \vee p_j \vee q_i \vee q_j) \wedge (p_i \vee p_j \vee \neg q_i \vee \neg q_j) \wedge (\neg p_i \vee \neg p_j \vee q_i \vee q_j) \wedge (\neg p_i \vee \neg p_j \vee \neg q_i \vee \neg q_j)$. Выполнимость конъюнкции всех таких формул эквивалентна раскрашиваемости исходного графа.

- *Задача о расстановке ферзей на шахматной доске.* Известна такая задача: можно ли расставить 8 ферзей на шахматной доске, так чтобы они не били друг друга? Мы построим формулу, любой выполняющий набор которой соответствует корректной расстановке. Для этого заведём переменные p_{ij}, истинность которых означает, что в клетке с координатами (i,j) стоит ферзь. По условию в одной строке не может быть двух ферзей, значит, в каждой должен быть ровно один. Для начала потребуем, что есть хотя бы один ферзь, записав дизъюнкты вида $(p_{i1} \vee p_{i2} \vee \dots \vee p_{i8})$. Далее, запишем условия, что в каждом столбце стоит не более одного ферзя. А именно, для каждого набора (i,j,k), где $i \neq j$, возьмём дизъюнкт $(\neg p_{ik} \vee \neg p_{jk})$. Аналогичные условия для строк можно записать отдельно, но они будут следовать из уже написанных. Также нужно написать условия для диагоналей: $(\neg p_{ij} \vee \neg p_{i+k,j+k})$ и $(\neg p_{ij} \vee \neg p_{i+k,j-k})$ при всех i, всех $j < 8$ и всех $k > 0$ (записываются только те условия, где все индексы попадают в интервал от 1 до 8). Любой выполняющий набор для такой системы задаёт расстановку ферзей. Можно рассматривать разные варианты задачи: другой размер доски, фиксированное положение некоторых ферзей (тогда добавятся дизъюнкты p_{ij}) или запрещённые клетки (тогда добавятся дизъюнкты $\neg p_{ij}$).

В обоих случаях мы получили КНФ, но не 3-КНФ: в некоторых дизъюнктах больше трёх литералов. Оказывается, выполнимость любой КНФ можно свести к выполнимости 3-КНФ более простым способом, чем выполнимость произвольной формулы.

Утверждение 2. *Пусть φ — КНФ. Тогда выполнимость φ эквивалентна выполнимости φ', образованной следующим образом: каждый дизъюнкт $(l_1 \vee l_2 \vee \dots \vee l_k)$ заменяется на такую КНФ: $(l_1 \vee x_1) \wedge (\neg x_1 \vee l_2 \vee x_2) \wedge \dots \wedge (\neg x_{k-2} \vee l_{k-1} \vee x_{k-1}) \wedge (\neg x_{k-1} \vee l_k)$, где переменные $x_1, \dots, x_{k-1}$ свои для каждого дизъюнкта.*

Доказательство. Пусть φ выполнима. Тогда при выполняющем наборе один из литералов $l_1, \dots, l_k$ истинен, например, l_j. Пусть тогда все x_i с $i < j$ равны 1, а все x_i с $i \geqslant j$ равны 0. Это сделает истинными все скобки в новой КНФ.

Пусть, напротив, выполнима φ'. Каждая из переменных x_i может сделать истинной ровно одну скобку. Значит, все они могут сделать истинными максимум $k-1$ скобку. Оставшаяся должна стать истинной за счёт l_j. А значит, и исходный дизъюнкт выполнен. $\square$

2 Резолюции

Рассуждение из последней конструкции можно формализовать так: если $(A \vee x)$ и $(B \vee \neg x)$ одновременно истинны, то и $(A \vee B)$ тоже истинно. Такое рассуждение называется *правилом резолюции* (resolution rule): $\dfrac{(A \vee x) \quad (B \vee \neg x)}{(A \vee B)}$. Сам дизъюнкт $(A \vee B)$ называется *резольвентой* (resolvent) дизъюнктов $(A \vee x)$ и $(B \vee \neg x)$. Заметим, что резольвента дизъюнктов x и $\neg x$ это пустой дизъюнкт, т. е. $\perp$. Если в A и B встречается один и тот же литерал, то второй раз его писать не обязательно. Если же в A есть y, а в B есть $\neg y$, то в резольвенте появится $y \vee \neg y$. Так как такая формула всегда истинна, никакой полезной информации из неё извлечь не получится.

Метод резолюций заключается в следующем: будем добавлять к набору дизъюнктов все возможные резольвенты. Так будем делать до тех пор, пока не будет верно одно из двух:

- Добавленным оказался пустой дизъюнкт,
- Никаких новых дизъюнктов добавить нельзя.

Теорема 3. *Метод резолюций всегда заканчивает свою работу, причём для невыполнимых КНФ выводится $\perp$ (полнота), а для выполнимых не выводится (корректность).*

Таким образом, метод резолюций позволяет проверить выполнимость формулы: достаточно добавить все возможные резольвенты и проверить, встретился ли $\perp$. Например, невыполнимая КНФ $(x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee x_2) \wedge (\neg x_2 \vee x_3) \wedge \neg x_3$ опровергается так:

1) $(x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee x_2) \vdash (x_2 \vee x_3)$
2) $(x_2 \vee x_3) \wedge (\neg x_2 \vee x_3) \vdash x_3$
3) $x_3 \wedge \neg x_3 \vdash \perp$

Доказательство. Всего существует конечное число дизъюнктов, так что в какой-то момент новые перестанут появляться, поэтому метод всегда заканчивает свою работу. Как обычно, корректность доказывается легко. Действительно, если исходная КНФ была выполнима, то она останется выполнимой после добавления любого числа резольвент. Но КНФ с $\perp$ выполнимой быть не может. Значит, для выполнимой КНФ $\perp$ не появится.

Для полноты мы докажем обратное свойство: если $\perp$ не выводится, то КНФ выполнима. Будем строить выполняющий набор рекурсивно, поддерживая следующий инвариант: после i-го шага значения $x_1 = a_1, \ldots, x_i = a_i$ выбраны так, что истинны все дизъюнкты, содержащие только переменные $x_1, \ldots, x_i$. Отсутствие $\perp$ даёт нам базу индукции: выполнены все дизъюнкты, не зависящие ни от каких переменных (т. к. таких дизъюнктов нет). Переход: пусть уже заданы значения $x_1 = a_1, \ldots, x_i = a_i$. Рассмотрим все дизъюнкты, зависящие от x_{i+1} и, возможно, каких-то из переменных $x_1, \ldots, x_i$, но не переменных $x_{i+2}, \ldots, x_n$. Некоторые из них уже истинны благодаря выбору предыдущих значений. Исключим их из рассмотрения. Если в оставшиеся входит только литерал x_{i+1}, то положим $x_{i+1} = 1$, инвариант будет соблюдён. Если в оставшиеся входит только литерал $\neg x_{i+1}$, то положим $x_{i+1} = 0$, и инвариант тоже будет соблюдён. Осталось исключить случай, когда в какой-то дизъюнкт входит x_{i+1}, а в какой-то другой $\neg x_{i+1}$. Действительно, пусть такие дизъюнкты нашлись: $D_1 = (x_{i+1} \vee D'_1)$ и $D_0 = (\neg x_{i+1} \vee D'_0)$, где D'_1 и D'_0 зависят только от $x_1, \ldots, x_i$. По предположению D'_0 и D'_1 должны быть ложны на значениях $x_1 = a_1, \ldots, x_i = a_i$, иначе мы бы исключили D_0 или D_1. Но резольвентой D_0 и D_1 будет дизъюнкт $(D'_0 \vee D'_1)$, который тоже должен быть ложен при $x_1 = a_1, \ldots, x_i = a_i$. Но тогда предположение индукции нарушено. Значит, последний случай невозможен, и потому исходная формула выполнима. $\square$

В качестве примера разберём опровержение методом резолюций возможности расстановки 3 ферзей на доске 3×3. Будем использовать изложенную ранее конструкцию, но для краткости вместо p_{ij} будем записывать просто ij, а вместо $\neg p_{ij}$ — $\overline{ij}$. Изначально формулы будут такими:

$11 \vee 12 \vee 13$	$\overline{11} \vee \overline{21}$	$\overline{12} \vee \overline{22}$	$\overline{13} \vee \overline{23}$	$\overline{12} \vee \overline{23}$	$\overline{12} \vee \overline{21}$
$21 \vee 22 \vee 23$	$\overline{11} \vee \overline{31}$	$\overline{12} \vee \overline{32}$	$\overline{13} \vee \overline{33}$	$\overline{11} \vee \overline{22}$	$\overline{13} \vee \overline{22}$
$31 \vee 32 \vee 33$	$\overline{21} \vee \overline{31}$	$\overline{22} \vee \overline{32}$	$\overline{23} \vee \overline{33}$	$\overline{11} \vee \overline{33}$	$\overline{13} \vee \overline{31}$
				$\overline{22} \vee \overline{33}$	$\overline{22} \vee \overline{31}$
				$\overline{21} \vee \overline{32}$	$\overline{23} \vee \overline{32}$

Теперь начнём выводить противоречие. Оно фактически будет повторять стандартное рассуждение с перебором вариантов. Вначале предположим, что ферзь стоит в клетке 11.

1) $11 \vee 12 \vee 13, \overline{11} \vee \overline{21} \vdash 12 \vee 13 \vee \overline{21}$

2) $11 \vee 12 \vee 13, \overline{11} \vee \overline{22} \vdash 12 \vee 13 \vee \overline{22}$

3) $11 \vee 12 \vee 13, \overline{11} \vee \overline{31} \vdash 12 \vee 13 \vee \overline{31}$

4) $11 \vee 12 \vee 13, \overline{11} \vee \overline{33} \vdash 12 \vee 13 \vee \overline{33}$

Здесь мы вывели, что тогда ферзей нет в клетках 21, 22, 31, 33.

5) $12 \vee 13 \vee \overline{21}, 21 \vee 22 \vee 23 \vdash 12 \vee 13 \vee 22 \vee 23$

6) $12 \vee 13 \vee \overline{22}, 12 \vee 13 \vee 22 \vee 23 \vdash 12 \vee 13 \vee 23$

Далее выводится, что тогда ферзь должен быть в клетке 23...

7) $12 \vee 13 \vee \overline{31}, 31 \vee 32 \vee 33 \vdash 12 \vee 13 \vee 32 \vee 33$

8) $12 \vee 13 \vee \overline{33}, 12 \vee 13 \vee 32 \vee 33 \vdash 12 \vee 13 \vee 32$

... и в клетке 32.

9) $12 \lor 13 \lor 23, \overline{23} \lor \overline{32} \vdash 12 \lor 13 \lor \overline{32}$

10) $12 \lor 13 \lor 32, 12 \lor 13 \lor \overline{32} \vdash 12 \lor 13$

Но они бьют друг друга, значит, в клетке 11 ферзя стоять не может, он должен быть в 12 или 13.

11) $12 \lor 13, \overline{13} \lor \overline{23} \vdash 12 \lor \overline{23}$

12) $12 \lor 13, \overline{13} \lor \overline{22} \vdash 12 \lor \overline{22}$

13) $12 \lor 13, \overline{13} \lor \overline{33} \vdash 12 \lor \overline{33}$

14) $12 \lor 13, \overline{13} \lor \overline{31} \vdash 12 \lor \overline{31}$

15) $12 \lor \overline{23}, 21 \lor 22 \lor 23 \vdash 12 \lor 21 \lor 22$

16) $12 \lor \overline{22}, 12 \lor 21 \lor 22 \vdash 12 \lor 21$

17) $12 \lor \overline{33}, 31 \lor 32 \lor 33 \vdash 12 \lor 31 \lor 32$

18) $12 \lor \overline{31}, 12 \lor 31 \lor 32 \vdash 12 \lor 32$

19) $12 \lor 21, \overline{21} \lor \overline{32} \vdash 12 \lor \overline{32}$

20) $12 \lor 32, 12 \lor \overline{32} \vdash 12$

Аналогично показывается, что в клетке 13 ферзя стоять также не может. Но ферзь в клетке 12 бьёт все клетки второй строки, из чего и получается итоговое противоречие.

21) $12, \overline{12} \lor \overline{22} \vdash \overline{22}$

22) $12, \overline{12} \lor \overline{23} \vdash \overline{23}$

23) $12, \overline{12} \lor \overline{21} \vdash \overline{21}$

24) $21 \lor 22 \lor 23, \overline{21} \vdash 22 \lor 23$

25) $22 \lor 23, \overline{22} \vdash 23$

26) $23, \overline{23} \vdash \perp$

3 Выводы тавтологий

Метод резолюций даёт альтернативный способ вывода тавтологий. Действительно, φ является тавтологией тогда и только тогда, когда $\neg\varphi$ является противоречием. Так что вместо того, чтобы выводить φ, можно опровергать $\neg\varphi$, приведя её предварительно к КНФ, эквивалентной по выполнимости. Для такого приведения можно использовать конструкцию из доказательства теоремы 1 для исходной формулы, только в самом конце вместо z_n написать $\neg z_n$. Из теоремы о полноте метода резолюций следует, что так можно вывести любую тавтологию. Для примера покажем, как вывести закон контрапозиции $(p \to q) \to (\neg q \to \neg p)$. Сделаем ещё одно упрощение: не будем обрабатывать отрицания отдельно, т. е. вначале преобразуем формулу к такому набору условий:

$$(r \leftrightarrow (p \to q)) \wedge (s \leftrightarrow (\neg q \to \neg p)) \wedge (t \leftrightarrow (r \to s)) \wedge \neg t.$$

Заметим, что формула $z \leftrightarrow (x \to y)$ равносильна такой КНФ:

$$(x \vee z) \wedge (\neg y \vee z) \wedge (\neg x \vee y \vee \neg z).$$

Проведя все замены, получим такой набор условий:

$$(p\vee r)\wedge(\neg q\vee r)\wedge(\neg p\vee q\vee\neg r)\wedge(\neg q\vee s)\wedge(p\vee s)\wedge(q\vee\neg p\vee\neg s)\wedge(r\vee t)\wedge(\neg s\vee t)\wedge(\neg r\vee s\vee\neg t)\wedge\neg t.$$

Теперь опровержение выводится такой последовательностью резолюций:

1) $\neg s \vee t, \neg t \vdash \neg s$
2) $r \vee t, \neg t \vdash r$
3) $p \vee s, \neg s \vdash p$
4) $\neg q \vee s, \neg s \vdash \neg q$
5) $\neg p \vee q \vee \neg r, r \vdash \neg p \vee q$
6) $\neg p \vee q, p \vdash q$
7) $q, \neg q \vdash \perp$

Заметим, что такой вывод фактически повторяет такое доказательство закона контрапозиции. Пусть формула $(p \to q) \to (\neg q \to \neg p)$ ложна. Тогда $p \to q$ истинно, а $\neg q \to \neg p$ ложно. Тогда $\neg q$ истинно, а $\neg p$ ложно. Но тогда p истинно, а q ложно, что противоречит истинности $p \to q$.