Формальные доказательства

- Формальные доказательства — большой раздел математической логики
- ⋆ Формальное доказательство — это конечная последовательность синтаксически корректных формул, составленная по правилам, определяемым системой доказательств
- Система доказательств состоит из
  - правил вывода (получения новых формул из имеющихся в доказательстве)
  - аксиом (формул, которые можно включать в доказательство без ограничений)
- Цель формального доказательства — из набора формул-условий получить требуемую формулу-заключение

  Пример:
  - формулы — слова над алфавитом $\{S, (,)\}$
  - аксиома — слово S
  - правила вывода: любой символ S в формуле можно заменить на SS, (S) или $()$
  - вывод формулы $(()())$ из пустого набора условий:

    S
    (S)
    (SS)
    $(()S)$
    $(()())$

  - ⋆ формула над $\{(,)\}$ выводится из пустого набора условий
  - ⇔ она является правильной расстановкой скобок
- ⋆ Такие системы доказательств называются формальными грамматиками

Доказательство теорем

- Как выглядит теорема?
  - даны условия $F_1, \ldots, F_k$ и гипотеза G
  - доказать, что из условий следует гипотеза
  - ⋆ т.е. что формула $(F_1 \wedge F_2 \wedge \cdots \wedge F_k) \to G$ — тавтология
  - ⋆ если формула $X \to Y$ — тавтология, то формула Y называется следствием X
- Простейший случай: $F_1, \ldots, F_k, G$ — булевы формулы
  - ⋆ могут быть более сложные формулы (с предикатами, кванторами и т. д.)
  - ⋆ даже для булевых формул проверка «в лоб» очень трудоемка: таблицу значений формулы из m литералов и n переменных можно вычислить за время $\Theta(m \cdot 2^n)$
- Доказательство от противного:
  - доказать, что $\overline{(F_1 \wedge F_2 \wedge \cdots \wedge F_k) \to G}$ — противоречие
  - эквивалентная формула: $F_1 \wedge F_2 \wedge \cdots \wedge F_k \wedge \bar{G}$
  - ⋆ если каждую из формул $F_1, \ldots, F_k, \bar{G}$ заменить на эквивалентную КНФ, общая формула станет КНФ
- Задача: дана КНФ, является ли она противоречием?

★ Наблюдение: Y — следствие $X \Leftrightarrow X$ эквивалентна $X \wedge Y$

★ Следствие: 0 — следствие $X \Leftrightarrow X$ — противоречие

- Стратегия доказательства: получить 0 как следствие исходной формулы
- Метод резолюций — система доказательств, реализующая эту стратегию

Лемма о следствии

Лемма

Для любых булевых формул X, Y, Z формула $Y \vee Z$ — следствие формулы $(X \vee Y) \wedge (\bar{X} \vee Z)$.

Доказательство:

- пусть $F_{|\vec{b}}$ обозначает результат подстановки набора значений $\vec{b}$ в формулу F
- пусть $\vec{b}$ — произвольный набор, такой что $((X \vee Y) \wedge (\bar{X} \vee Z))_{|\vec{b}} = 1$

$\Rightarrow$ $(X \vee Y)_{|\vec{b}} = 1$, $(\bar{X} \vee Z)_{|\vec{b}} = 1$

- если $X_{|\vec{b}} = 1$, то $\bar{X}_{|\vec{b}} = 0 \Rightarrow Z_{|\vec{b}} = 1$
- если $X_{|\vec{b}} = 0$, то $Y_{|\vec{b}} = 1$

$\Rightarrow$ $(Y \vee Z)_{|\vec{b}} = 1$

$\Rightarrow$ $((X \vee Y) \wedge (\bar{X} \vee Z)) \to (Y \vee Z)$ — тавтология $\square$

Метод резолюций

Метод резолюций:

- формулы, которыми оперирует метод — это клозы (элементарные дизъюнкции)
- клоз рассматривается как множество литералов
  - порядок литералов не важен, повторяющиеся литералы стираются
- единственное правило вывода — правило резолюций:
  - если есть клозы вида $x \vee C$ и $\bar{x} \vee D$ (x — переменная), дописать клоз $C \vee D$
  - ⋆ клоз, содержащий пару литералов $\{y, \bar{y}\}$, не дописывается
  - если C и D — пустые множества литералов, дописывается пустой клоз $\square$
- аксиом нет
- условия — все клозы КНФ, поданной на вход метода
- цель — получить пустой клоз

Теорема о полноте

Теорема о полноте метода резолюций

КНФ $F = C_1 \wedge \cdots \wedge C_k$ является противоречием $\Leftrightarrow$ существует доказательство методом резолюций с условиями $C_1, \ldots, C_k$ и заключением $\square$.

Доказательство достаточности:

- рассмотрим доказательство методом резолюций с заключением $\square$
- каждая формула является либо условием, либо получено по правилу резолюций из каких-то предыдущих формул
  - а значит, является следствием конъюнкции этих формул согласно лемме
- отношение «быть следствием» транзитивно
- любая формула вида $C_{i_1} \wedge \cdots \wedge C_{i_j}$ является следствием F

$\Rightarrow$ любая формула в доказательстве является следствием F

$\star$ пустой клоз является следствием формулы $x \wedge \bar{x}$, а значит, задает константу 0

$\Rightarrow$ 0 — следствие F $\Rightarrow$ F — противоречие $\square$

Комментарий:

$\star$ мы доказали корректность метода: если существует доказательство, содержащее пустой клоз, то заданная КНФ действительно является противоречием

$\star$ обратная импликация доказывает полноту метода: если КНФ — противоречие, то это можно доказать методом резолюций

Доказательство необходимости

- Проведем индукцию по числу n переменных в F
- База индукции: $n = 1$
  - F — противоречие $\Rightarrow$ F содержит клозы x и $\bar{x}$

  $\Rightarrow$ по правилу резолюций из x и $\bar{x}$ выводится пустой клоз
- Шаг индукции:
  - пусть $F = F(x_1, \ldots, x_n)$, $S = \{C_1, \ldots, C_k\}$
  - считаем, что клоз не может содержать одновременно x_i и $\bar{x}_i$
    - если такой клоз есть, он задает константу 1 и может быть удален из F
  - построим два множества клозов, S^+ и S^-:
  - $S^+ = \{C \in S \mid \text{ в } C \text{ нет переменной } x_n\} \cup \{C \mid (C \vee x_n) \in S\}$
  - $S^- = \{C \in S \mid \text{ в } C \text{ нет переменной } x_n\} \cup \{C \mid (C \vee \bar{x}_n) \in S\}$
  - ⋆ докажем, что КНФ $F^+ = \bigwedge_{C \in S^+} C$ является противоречием:
  - пусть существует набор значений $b_1, \ldots, b_{n-1}$ такой, что $F^+_{|b_1, \ldots, b_{n-1}} = 1$
  - рассмотрим значения всех клозов из множества S на наборе $b_1, \ldots, b_{n-1}, 0$:
    - если клоз C не содержит переменную x_n, то $C_{|b_1, \ldots, b_{n-1}, 0} = C_{|b_1, \ldots, b_{n-1}} = 1$
    - если клоз имеет вид $C \vee x_n$, то $(C \vee x_n)_{|b_1, \ldots, b_{n-1}, 0} = C_{|b_1, \ldots, b_{n-1}} = 1$
    - клоз вида $C \vee \bar{x}_n$ превращается в 1 за счет значения $b_n = 0$

  $\Rightarrow$ $F_{|b_1, \ldots, b_{n-1}, 0} = 1$, что невозможно, так как F — противоречие
  - ⋆ аналогично, $F^- = \bigwedge_{C \in S^-} C$ является противоречием
    - к гипотетическому набору, выполняющему F^-, надо добавить $b_n = 1$
  - ⋆ по предположению индукции, из каждого из множеств S^+, S^- можно вывести пустой клоз

Шаг индукции — окончание

- Рассмотрим вывод пустого клоза из множества S^+
  - если в выводе участвовали только клозы из S, то из S выводим пустой клоз
  - пусть в выводе участвовал хотя бы один клоз $C \in S^+ \setminus S$; тогда $(C \vee x_n) \in S$

  $\Rightarrow$ построим вывод из S, заменив в выводе из S^+ каждый клоз из $S^+ \setminus S$ на соответствующий клоз из S

  $\Rightarrow$ во всех следствиях из таких клозов добавится литерал x_n

  $\Rightarrow$ из S выводится клоз x_n

$\star$ аналогично, из вывода пустого клоза из S^- получим вывод клоза $\bar{x}_n$ из S

$\Rightarrow$ из клозов x_n и $\bar{x}_n$ получим пустой клоз $\square$

Комментарий:

- $\star$ искать доказательства методом резолюций может компьютер
  - существуют различные стратегии оптимизации поиска вывода
- $\star$ на более общем варианте метода резолюций (для формул логики первого порядка) основан язык Пролог

$\bigstar$ Если формула F не является противоречием, то метод резолюций заканчивает работу, когда не может вывести больше ни одного нового клоза

- по построенным клозам можно восстановить набор значений, выполняющий F

Пример доказательства методом резолюций

Вася всегда приходит на совещание, если босс его позвал. Если босс хочет видеть Васю, он зовет его на совещание. Если босс не хочет видеть Васю и не зовет его на совещание, то Васю скоро уволят. Вася не пришел на совещание. Докажите, что его скоро уволят.

- Запишем теорему, которую надо доказать:
  - $\star$ $\Big((invite \to attend) \wedge (see \to invite) \wedge ((\overline{see} \wedge \overline{invite}) \to fire) \wedge (\overline{attend})\Big) \to fire$
- Отрицание теоремы:
  - $\star$ $(invite \to attend) \wedge (see \to invite) \wedge ((\overline{see} \wedge \overline{invite}) \to fire) \wedge (\overline{attend}) \wedge \overline{fire}$
- КНФ отрицания теоремы и множество клозов:
  - $\star$ $(\overline{invite} \vee attend) \wedge (\overline{see} \vee invite) \wedge (see \vee invite \vee fire) \wedge (\overline{attend}) \wedge (\overline{fire})$
  - $S = \{\overline{invite} \vee attend,\ \overline{see} \vee invite,\ see \vee invite \vee fire,\ \overline{attend},\ \overline{fire}\}$

Доказательство:

1.	$see \vee invite \vee fire$	условие
2.	$\overline{fire}$	условие
3.	$see \vee invite$	по правилу резолюций из 1,2
4.	$\overline{see} \vee invite$	условие
5.	$invite$	по правилу резолюций из 3,4; $invite \vee invite = invite$
6.	$\overline{invite} \vee attend$	условие
7.	$attend$	по правилу резолюций из 5,6
8.	$\overline{attend}$	условие
9.	$\square$	по правилу резолюций из 7,8 жалко Васю. . .