ЛИКБЕЗ
Лекция 3:
Пропозициональная логика.
(Исчисление высказываний.)

Дмитрий Ицыксон

ПОМИ РАН

7 октября 2007

План

1. Формулы исчисления высказываний
2. Булевы функции и их представление в КНФ и ДНФ
3. Интерпретации, выполнимые, противоречивые формулы, тавтологии
4. Полиномиальные фрагменты: 2-SAT и хорновские формулы
5. Пропозициональные системы доказательств
6. Резолюционная система доказательств
7. Метод резолюций и “divide and conquer” алгоритмы
8. **NP** vs co**NP**
9. Булевы схемы и сложность булевых функций
10. Предикатные формулы, интерпретации и модели

Литература

1. Н.К. Верещагин, А. Шень. Языки и исчисления.
2. М.В. Дмитриева. Методы решения задач искусственного интеллекта. Автоматическое доказательство теорем.

Язык пропозициональных формул

Γ — бесконечное множество пропозициональных переменных.
$\Gamma = \{x_1, x_2, x_3, \dots\}$.

- Пропозициональная переменная является пропозициональной формулой
- Если A — пропозициональная формула, то (A) — тоже пропозициональная формула.
- Если A — пропозициональная формула, то $\neg A$ — тоже пропозициональная формула.
- Если A, B — пропозициональные формулы, то $A \wedge B$ — тоже пропозициональная формула.
- Если A, B — пропозициональные формулы, то $A \vee B$ — тоже пропозициональная формула.
- Если A, B — пропозициональные формулы, то $A \to B$ — тоже пропозициональная формула.

Примеры пропозициональных формул

- x_1;
- $x_1 \wedge \neg x_1$;
- $x_1 \vee \neg x_1$;
- $(x_1 \vee x_2) \to x_3$;
- $(x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_3) \wedge (x_3 \vee x_2)$;
- $(x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_3) \vee (x_3 \wedge x_2)$;

Таблицы истинности

$\neg$	
a	$\neg a$
0	1
1	0

$\vee$		
a	b	$a \vee b$
0	0	0
0	1	1
1	0	1
1	1	1

$\wedge$		
a	b	$a \wedge b$
0	0	0
0	1	0
1	0	0
1	1	1

$\rightarrow$		
a	b	$a \rightarrow b$
0	0	1
0	1	1
1	0	0
1	1	1

Интерпретация

Пусть φ — пропозициональная формула с переменными $x_1, x_2, \ldots, x_n$.
Интерпретацией формулы φ называется отображение $\sigma : \{x_1, x_2, \ldots, x_n\} \to \{0, 1\}$.
Значение формулы $I_\sigma(\varphi)$ при заданной интерпретации определяется индуктивно по построению формулы:

- $I_\sigma(x_i) = \sigma(x_i)$;
- $I_\sigma((A)) = I_\sigma(A)$;
- $I_\sigma(\neg A) = \neg I_\sigma(A)$;
- $I_\sigma(A \wedge B) = I_\sigma(A) \wedge I_\sigma(B)$;
- $I_\sigma(A \vee B) = I_\sigma(A) \vee I_\sigma(B)$;
- $I_\sigma(A \to B) = I_\sigma(A) \to I_\sigma(B)$.

Булевы функции

Определение. Булевой функцией мы называем функцию из $\{0,1\}^n$ в $\{0,1\}$.
Замечание. Каждая пропозициональная формула от n переменных задает булеву функцию из $\{0,1\}^n$ в $\{0,1\}$.
Примеры.

- Parity (четность): $f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n \bmod 2$
- Majority (большинство):
$$f(x_1, x_2, \ldots, x_n) = \begin{cases} 1, & \text{если } x_1 + x_2 + \cdots + x_n \geq \dfrac{n}{2} \\ 0, & \text{если } x_1 + x_2 + \cdots + x_n < \dfrac{n}{2} \end{cases}$$
- $f(x_1, x_2, \ldots, x_n) = \begin{cases} 1, & \text{если } \overline{x_1 x_2 \ldots x_n} \ \text{ — простое число} \\ 0, & \text{иначе} \end{cases}$

КНФ и ДНФ

Определение. Литералом называется пропозициональная переменная или ее отрицание: x_i, $\neg x_i$.

Определение. Дизъюнктом или клозом называется дизъюнкция нескольких (возможно одного) литералов: $(x_1 \vee \neg x_2 \vee x_3)$.

Определение. Формулой в КНФ называется конъюнкция нескольких (возможно одного) дизъюнктов:
$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_3) \wedge \neg x_1$.

Определение. Конъюнктом или мономом называется конъюнкция нескольких (возможно одного) литералов:
$(x_1 \wedge \neg x_2 \wedge x_3)$.

Определение. Формулой в ДНФ называется дизъюнкция нескольких (возможно одного) конъюнктов:
$(x_1 \wedge \neg x_2 \wedge x_3) \vee (\neg x_1 \wedge \neg x_3) \vee \neg x_1$.

Представление булевых функций формулами

Теорема. Любую булеву функцию можно записать в виде пропозициональной формулы в КНФ и ДНФ.

Иллюстрация.

$f(x_1, x_2, x_3)$			
x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0

Формула в ДНФ:

$(\neg x_1 \wedge \neg x_2 \wedge \neg x_3) \vee$
$(\neg x_1 \wedge \neg x_2 \wedge x_3) \vee$
$(x_1 \wedge \neg x_2 \wedge \neg x_3)$

Формула в КНФ:

$(x_1 \vee \neg x_2 \vee x_3) \wedge$
$(x_1 \vee \neg x_2 \vee \neg x_3) \wedge$
$(\neg x_1 \vee x_2 \vee \neg x_3) \wedge$
$(\neg x_1 \vee \neg x_2 \vee x_3) \wedge$
$(\neg x_1 \vee \neg x_2 \vee \neg x_3)$

Формулы де Моргана

- значение $\neg(A \vee B)$ совпадает со значением $(\neg A \wedge \neg B)$
- значение $\neg(A \wedge B)$ совпадает со значением $(\neg A \vee \neg B)$

Следствие. Отрицание формулы в ДНФ “есть” формула в КНФ

Доказательство.

$\neg((l_{1,1} \wedge l_{1,2} \wedge \cdots \wedge l_{1,n_1}) \vee (l_{2,1} \wedge l_{2,2} \wedge \cdots \wedge l_{2,n_2}) \vee \ldots$
$\vee(l_{k,1} \wedge l_{k,2} \wedge \cdots \wedge l_{k,n_k}))$

можно переписать в виде

$(\neg l_{1,1} \vee \neg l_{1,2} \vee \cdots \vee \neg l_{1,n_1}) \wedge (\neg l_{2,1} \vee \neg l_{2,2} \wedge \cdots \vee \neg l_{2,n_2}) \wedge \ldots$
$\wedge(\neg l_{k,1} \vee \neg l_{k,2} \vee \cdots \vee \neg l_{k,n_k})$

Выполнимость, общезначимость, противоречивость

Определение. Пропозициональная формула называется выполнимой , если существует такая интерпретация, при которой значение формулы равняется 1.

Определение. Пропозициональная формула называется невыполнимой (или противоречивой) , если при всех интерпретациях значение формулы равняется 0.

Определение. Пропозициональная формула называется общезначимой (или тавтологией) , если при всех интерпретациях значение формулы равняется 1.

Определение. Пропозициональная формула называется необщезначимой , если при существует интерпретация, при которой значение формулы равняется 0.

О сложности...

- Задача определения по формуле в КНФ, выполнима ли она, является сложной (**NP**-трудной).
- Проверить формулу в КНФ, является ли она тавтологией, очень просто. Достаточно проверить, что в каждом дизъюнкте есть пара: переменная и ее отрицание.
- Задача определения по формуле в ДНФ, выполнима ли она, является простой: достаточно проверить, есть ли в ней конъюнкт, в который одновременно не входит переменная и ее отрицание.
- Проверить формулу в ДНФ, является ли она тавтологией, сложно (**NP**-трудно).

Полиномиальные фрагменты

Определение. Формула в КНФ называется **хорновской**, если в каждй дизъюнкт не более одной переменной входит без отрицания. Пример:
$(\neg x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2) \wedge (\neg x_2 \vee \neg x_4) \wedge (x_3)$.

Теорема. Выполнимость хорновской формулы можно проверить за полиномиальное время.

Доказательство. Вход: хорновская формулу φ

1. Пусть U — дизъюнкты, состоящие из одного литерала.
2. Если $x \in U$ и $\neg x \in U$, то формула невыполнима.
3. Подставить в φ правильное значение для всех литералов из U.
4. Если U не пусто, вернуться к шагу 1.
5. После подстановок хорновская формула осталась хорновской. Значит, теперь все дизъюнкты содержат более, чем 1 переменную и хотя бы одну с отрицанием.
6. Подставим 0 вместо всех переменных.

Полиномиальные фрагменты: 2-SAT

Определение. 2-SAT - это язык выполнимых формул в 2-КНФ.
Например: $(x_1 \vee \neg x_2) \wedge (x_3 \vee \neg x_2) \wedge (x_1)$
Теорема. Распознать язык $2 - SAT$ можно за полиномиальное (линейное) время.
Доказательство.

- Избавимся от всех одноэлементных дизъюнктов подстановкой в них правильного значения.
- Сопоставим формуле с n переменными $x_1, x_2, \ldots, x_n$ ориентированный граф с $2n$ вершинами: $x_1, x_2, \ldots, x_n, \neg x_1, \neg x_2, \ldots, \neg x_n$.
- Если формула содержит дизъюнкт $(a \vee b)$, то соединим ребром: $(\neg a, b)$ и $(\neg b, a)$
  Смысл: Если есть ребро (x, y) и мы присвоили значение $x := 1$, то мы обязаны присвоить значение $y := 1$.

2-SAT: полиномиальный алгоритм

- Граф обладает свойством кососимметричности: если есть путь из a в b, то есть путь из $\neg b$ в $\neg a$.
- Если в графе есть путь из в $\neg a$ и из $\neg a$ в a, то формула невыполнима (не присвоить значение переменной a). Покажем, что если таких a нет, то формула выполнима.
- Индукция по числу дизъюнктов. База: 1 дизъюнкт.
- Пусть x такой литерал, что нет пути из x в $\neg x$. Пусть V_x — множество вершин, достижимых из x ($x \in V_x$).
- Подставим всем литералам из V_x значение 1.
- Получившаяся формула не содержит одноэлементных дизъюкнтов.
- Конфликтов не было: если из x был путь в y и в $\neg y$, то из y был путь в $\neg x$, т.е. из x был путь в $\neg x$, противоречие.

Пропозициональные системы доказательств

- Мы доказываем, что формула является тавтологией. (Иногда, что ее отрицание является противоречием).
- Доказательство — это строка.
- Доказательство можно легко проверить (за полиномиальное время). Найти доказательство, возможно, сложно.
- Корректность: если формула имеет доказательство, то она тавтология.
- Полнота: все тавтологии имеют доказательства.

Гильбертовская система H

11 схем аксиом:

1. $A \to (B \to A)$
2. $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$
3. $(A \wedge B) \to A$
4. $(A \wedge B) \to B$
5. $A \to (B \to (A \wedge B))$
6. $A \to (A \vee B)$
7. $B \to (A \vee B)$
8. $(A \to C) \to ((B \to C) \to (A \vee B \to C))$
9. $\neg A \to (A \to B)$
10. $(A \to B) \to ((A \to \neg B) \to \neg A)$
11. $A \vee \neg A$

Правило вывода (modus ponens):

$$\frac{A; (A \to B)}{B}$$

Резолюционная система

- Предназначена для доказательства противоречивости формул в КНФ.
  Что делать, чтобы можно было бы доказывать тавтологичность всех формул?
- Приведем формулу в ДНФ эквивалентными преобразованиями.
- Формула в ДНФ - тавтология $\iff$ ее отрицание - противоречие в КНФ.

Приведение формулы в ДНФ

- Избавляемся от импликации: $A \to B$ заменяем на $\neg A \vee B$.
- По правилам де Моргана проносим отрицания до переменных
- Пользуясь дистрибутивностью $A \wedge (C \vee D) = A \wedge C \vee A \wedge D$ раскрываем скобки.
- Упрощаем, выкидывая невыполнимые конъюнкты

Пример. $((A \vee C) \to B) \wedge (A \vee C)$
избавляемся от импликации: $(\neg(A \vee C) \vee B) \wedge (A \vee C)$
проносим отрицания: $((\neg A \wedge \neg C) \vee B) \wedge (A \vee C)$
раскрываем скобки:
$(\neg A \wedge \neg C \wedge A) \vee (\neg A \wedge \neg C \wedge C) \vee (B \wedge A) \vee (B \wedge C)$
упрощаем: $(B \wedge A) \vee (B \wedge C)$

С помощью метода резолюций мы показываем противоречивость формулы в КНФ. Выводим противоречие из множества дизъюнктов. Шаг за шагом выводим новые дизъюнкты из уже имеющихся по правилу:

$$\frac{(A \vee x); (B \vee \neg x)}{(A \vee B)},$$

где A и B не содержат конфликтующих литералов.
Корректность: если выполнимы $(A \vee x)$ и $(B \vee \neg x)$, то выполним и $(A \vee B)$
Доказательство: вывод пустого дизъюнкта $\square$

Пример

$$(x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_2) \wedge (x_2 \vee \neg x_3) \wedge (x_3 \vee \neg x_2) \wedge (x_1 \vee x_3) \wedge (\neg x_1 \vee \neg x_3)$$

1. $\frac{(x_1 \vee \neg x_2);(x_2 \vee \neg x_3)}{(x_1 \vee \neg x_3)}$
2. $\frac{(x_1 \vee \neg x_3);(x_1 \vee x_3)}{(x_1)}$
3. $\frac{(x_1);(\neg x_1 \vee x_2)}{(x_2)}$
4. $\frac{(x_2);(\neg x_2 \vee x_3)}{(x_3)}$
5. $\frac{(x_3);(\neg x_1 \vee \neg x_3)}{(\neg x_1)}$
6. $\frac{(x_1);(\neg x_1)}{\square}$

Полнота метода резолюций

Теорема. Из множества дизъюнктов любой противоречивой формулы в КНФ можно по правилам резолюции вывести пустой дизъюнкт $\square$.

Доказательство. Доказываем индукцией по числу переменных.

- Если переменная одна и формула противоречива, то в ней есть дизъюнкты (x) и $(\neg x)$. Применяем $\frac{(x);(\neg x)}{\square}$.
- Переход. Выберем переменную x. Сделаем подстановку $x := 1$ во все дизъюнкции. Выполненные дизъюнкции выкинем.
- Подставленная формула тоже противоречива. По индукционному предложению, существует вывод пустой дизъюнкции $\square$.
- Когда мы вернем литерал $\neg x$ во все дизъюнкции, получится либо корректный вывод противоречия $\square$, либо вывод $(\neg x)$.
- Аналогично, делая подстановку $x := 0$, получим либо вывод $\square$, либо вывод x, остается применить правило $\frac{x;\neg x}{\square}$.

Алгоритмы “divide and conquer”

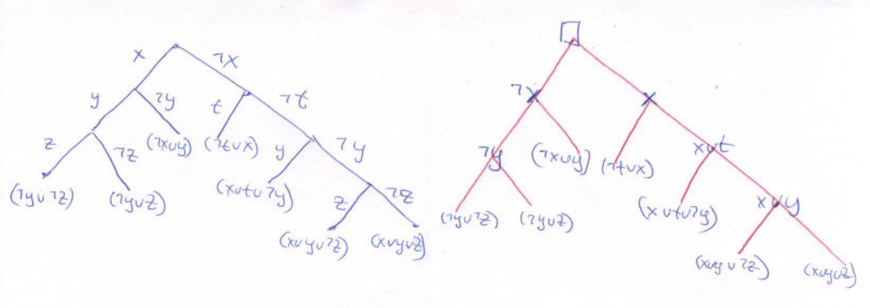
Алгоритм для проверки формулы на выполнимость. Схема алгоритма:

Вход: Формула F в КНФ

1. Если F содержит пустой дизъюнкт, то выдать “’невыполнима’
2. Если в F нет дизъюнктов, то выдать “выполнима”
3. Выбираем каким-либо образом литерал x
4. Рекурсивно вызываем алгоритм для формулы $F[x := 0]$, если ответ “выполнима”, то выдать “выполнима”
5. Рекурсивно вызываем алгоритм для формулы $F[x := 1]$, если ответ “выполнима”, то выдать “выполнима”

Резолюции vs “divide and conquer”

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



Определение. Язык $L \in$ co-**NP** тогда и только тогда, когда $\overline{L} \in$ **NP**. ($\overline{L} = \Sigma^* \setminus L$)
Лемма. $\overline{SAT}$ co-**NP** полный.
Доказательство. $L \in$ co-**NP** $\iff$ $\overline{L} \in$ **NP**. Значит, $\overline{L}$ сводится к SAT, т.е. существует такая полиномиально вычислимая функция f, что $x \in \overline{L} \iff f(x) \in \mathrm{SAT}$. Это можно переписать $x \in L \iff f(x) \in \overline{SAT}$.
Замечание. Совпадают ли классы **NP** и co-**NP** — это открытый вопрос.

NP vs co-NP

Лемма. Если SAT $\in$ co-**NP**, то **NP**=co-**NP**.
Доказательство. SAT $\in$ co-**NP** $\iff$ $\overline{SAT} \in$ **NP** $\iff$ существует полиномиально ограниченное и полиномиально проверяемое отношение $R \subset \Sigma^* \times \Sigma^*$, что $\overline{SAT} = \{x | \exists y : (x, y) \in R\}$.
Пусть $L \in$ **NP**, L сводится к SAT. Т.е., существует такая полиномиально вычислимая функция f, что $x \in L \iff f(x) \in$ SAT (дополнительно нужно потребовать $|f(x)| \geq |x|$).
Следовательно, $x \in \overline{L} \iff f(x) \in \overline{SAT}$.
$\overline{L} = \{x | \exists y (f(x), y) \in R\} \in$ **NP**. Значит, $L \in$ co-**NP**. Т.е. **NP** $\subset$ co-**NP**.
Пусть $L \in$ co-**NP**, L сводится к $\overline{SAT}$. Т.е., существует такая полиномиально вычислимая функция f, что $x \in L \iff f(x) \in \overline{SAT}$. $L = \{x | \exists y (f(x), y) \in R\} \in$ **NP**. Значит, co-**NP** $\subset$ **NP**.
Следствие. co-**NP** = **NP** $\iff$ SAT $\in$ co-**NP** $\iff$ $\overline{SAT} \in$ **NP**

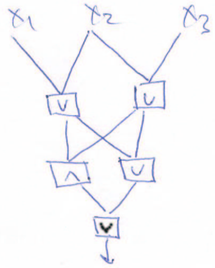
Системы доказательств и **NP** vs co-**NP**

$\overline{SAT} \in \mathbf{NP}$ означает, что для каждой невыполнимой формулы существует полиномиальное по размеру и полиномиально проверяемое доказательство.
Значит, если есть система доказательств, в которой все формулы из $\overline{SAT}$ имеют короткие (полиномиальные) опровержения, то co-$\mathbf{NP} = \mathbf{NP}$.
Для резолюционной системы существуют формулы, которые имеют экспоненциальные резолюционные опровержения.

Булевы схемы

Булева схема - это

- Ориентированный граф без циклов
- Ровно одна вершина, из которой не выходит ребер (выход)
- n вершин в которые не входят ребра
- Все остальные вершины помечены логическими связками $\vee, \wedge, \neg$. (арность связки должна равняться числу исходящих ребер)



Схемная сложность булевых функций

- Размер схемы — это количество вершин в графе, задающей схему.
- Схемная сложность функции — это минимальный размер схемы, вычисляющей эту функцию.
- Схему размера k можно записать с помощью $O(k \log k)$ битов.
- Количество схем размера $2^{\frac{n}{2}}$ не превосходит $2^{O(\frac{n}{2}2^{n/2})}$.
- Количество булевых функций от n переменных равняется 2^{2^n}.
- Значит, существует булева функция, сложность которой не менее $2^{\frac{n}{2}}$.
- Открытый вопрос: построить функцию большой схемной сложностью.

Язык предикатных формул

Γ — бесконечное множество предметных переменных.
$\Gamma = \{x_1, x_2, x_3, \ldots\}$.
$\mathfrak{F} = \{f_1^{(i_1)}, f_2^{(i_2)} ...\}$ — множество функциональных символов с указанием их арности, $i_k \geq 0$ (в этом множестве есть бесконечное число функциональных символов любой арности).
Определение. Термы:

- Предметная переменная $x \in \Gamma$ — терм.
- Если $f^{(i)} \in \mathfrak{F}$, а $t_1, t_2, \ldots, t_i$ — термы, то $f^{(i)}(t_1, t_2, \ldots, t_i)$ — терм.

Пример.

- $f^{(0)}()$ — терм;
- $f^{(2)}(x, y)$ — терм;
- $f^{(2)}(g^{(1)}(x), h^{(3)}(x, y, g^{(1)}(x)))$ — терм.

Язык предикатных формул

$\mathfrak{P} = \{p_1^{(i_1)}, p_2^{(i_2)} \ldots\}$ — множество предикатных символов с указанием их арности, $i_k \geq 0$ (в этом множестве есть бесконечное число предикатных символов любой арности).

Определение. Атомарной формулой называется строчка вида $p^{(i)}(t_1, t_2, \ldots, t_i)$, где $p^{(i)} \in \mathfrak{P}$, а $t_1, t_2, \ldots, t_i$ — термы.

Определение. Предикатная формула

- Если A — атомарная формула, то A — предикатная формула.
- Если A, B — предикатные формулы, то $(A), \neg A, A \vee B, A \wedge B, A \to B$ — предикатные формулы.
- Если A — предикатная формула, $x \in \Gamma$, то $\forall x A$ и $\exists x A$ являются предикатными формулами.

Примеры предикатных формул

- p(f(x)) — свободная переменная x;
- $p_1(f_1(x)) \vee p_2()$ — свободная переменная x;
- $\forall x \exists y (p_1(z) \vee p_1(x))$ — свободная переменная z;
- $\forall x (p_1(f(x))) \to \exists y p_1(y)$ — замкнутая формула;
- $\forall y (p_1(x, y) \vee \exists z p_2(f(x, y), g(x)))$ — свободная переменная x.

Определение. Переменная называется свободной, если она не входит в область действия квантора по этой переменной. Формула без свободных переменных называется замкнутой.

Интерпретация

Пусть φ — предикатная формула со свободными переменными $x_1, x_2, \ldots, x_k$, функциональными символами $f_1^{(i_1)}, f_2^{(i_2)}, \ldots, f_m^{(i_m)}$ и предикатными символами $p_1^{(j_1)}, p_2^{(j_2)}, \ldots p_n^{(j_n)}$.
Интерпретацией формулы φ называется множество M, заданные на нем отображения $f^{(i_s)} : M^{i_s} \to M$ и предикаты $p^{(j_r)} : M^{j_r} \to \{0, 1\}$, каждой переменной x_l сопоставлен элемент M.
Для каждой такой интерпретации можно посчитать значение формулы.
Моделью называется интерпретация, в которой значение формулы равняется 1.

Примеры

- $\forall x(p(x) \to q(x))$.
  В интерпретации $M = \mathbb{Z}$, $p(x) = x \,\vdots\, 4, q(x) = x \,\vdots\, 2$ значение формулы 1.
  В интерпретации $M = \mathbb{Z}$, $p(x) = x \,\vdots\, 3, q(x) = x \,\vdots\, 2$ значение формулы 0.
- $\forall x(p(f(x))) \to \forall x p(x)$ В интерпретации $M = \mathbb{Z}$, $p(x) = x \,\vdots\, 2, f(x) = 2x$ значение формулы 0.
- $(p() \vee q()) \wedge (p() \vee \neg q())$ — пропозициональные формулы — это частный случай предикатных. Если в предикатных формулах содержатся только нульместные предикаты и нет кванторов, предметных переменных и функциональных символов.