Математическая логика

Лектор:

Подымов Владислав Васильевич

e-mail:
valdus@yandex.ru

2016, весенний семестр

Лекция 2

Логика высказываний:
синтаксис, семантика,
выполнимость, общезначимость

Метод семантических таблиц

SAT

Логика высказываний: вступление

Что такое логика высказываний?

В логике высказываний на основе

- истинности “простых” высказываний и
- “примитивных” причинно-следственных связей

анализируется истинность более “сложных” высказываний

Например, можно рассмотреть такое предложение:

Если будете прогуливать лекции,

то ничего хорошего из этого не выйдет

Логика высказываний: вступление

Что такое логика высказываний?

В логике высказываний на основе

- истинности “простых” высказываний и
- “примитивных” причинно-следственных связей

анализируется истинность более “сложных” высказываний

Например, можно рассмотреть такое предложение:

$$\underbrace{\textbf{Если будете прогуливать лекции, то ничего хорошего из этого не выйдет}}_{A}$$

Можно ли сказать, что это “простое высказывание”?

Логика высказываний: вступление

Что такое логика высказываний?

В логике высказываний на основе

- истинности “простых” высказываний и
- “примитивных” причинно-следственных связей

анализируется истинность более “сложных” высказываний

Например, можно рассмотреть такое предложение:

Если будете прогуливать лекции,

то ничего хорошего из этого не выйдет

$$\rightarrow$$

Здесь есть причинно-следственная связь ...

Логика высказываний: вступление

Что такое логика высказываний?

В логике высказываний на основе

- истинности “простых” высказываний и
- “примитивных” причинно-следственных связей

анализируется истинность более “сложных” высказываний

Например, можно рассмотреть такое предложение:

Если будете прогуливать лекции,

то ничего хорошего из этого не выйдет

$$\mathtt{A} \rightarrow \quad \mathtt{B}$$

... между двумя простыми высказываниями

Логика высказываний: вступление

Что такое логика высказываний?

В логике высказываний на основе

- истинности “простых” высказываний и
- “примитивных” причинно-следственных связей

анализируется истинность более “сложных” высказываний

Например, можно рассмотреть такое предложение:

Если будете прогуливать лекции,

то ничего хорошего из этого не выйдет

$$A \rightarrow \neg B$$

А одно из высказываний можно сделать ещё проще

Логика высказываний: вступление

Что такое логика высказываний?

В логике высказываний на основе

- истинности “простых” высказываний и
- “примитивных” причинно-следственных связей

анализируется истинность более “сложных” высказываний

Например, можно рассмотреть такое предложение:

Если будете прогуливать лекции,

то ничего хорошего из этого не выйдет

$$\boxed{\mathtt{A} \rightarrow \neg\, \mathtt{B}}$$

Логика высказываний: вступление

Что получилось после формализации высказывания?

Что-то очень похожее на **формулу булевой алгебры**,
но не совсем:

Какой смысл имеет построенное высказывание?

Булева алгебра:
значение формулы — это булева функция

A	B	$A \to \neg B$
0	0	1
0	1	1
1	0	1
1	1	0

Логика высказываний: ?

Логика высказываний: вступление

Для языка логики высказываний будут описаны

алфавит:	символы, используемые в языке
синтаксис:	правила, по которым из символов строятся высказывания языка (**формулы**)
семантика:	значение этих высказываний

Затем будут

- строго сформулированы интересующие нас свойства формул (выполнимость, невыполнимость, общезначимость)
- описаны **логические** средства проверки этих свойств (метод семантических таблиц)
- показаны примеры сведения решения практических задач к проверке сформулированных свойств
- показаны методы проверки этих свойств, используемые на практике (DPLL)

Алфавит, синтаксис

Алфавит

Пропозициональные переменные Var

Логические связки $\&, \vee, \to, \neg$

Скобки (,)

Синтаксис[1,2]

Что такое формула: (это индуктивное определение)

- x — атомарная формула, или атом ($x \in \mathrm{Var}$)
- Составные формулы: (φ, ψ — формулы)

$$(\varphi \,\&\, \psi) \qquad (\varphi \vee \psi) \qquad (\varphi \to \psi) \qquad (\neg\varphi)$$

- других формул нет

(этот пункт индуктивного определения будет опускаться)

Приоритет связок: $\neg$, потом $\&$, потом $\vee$, потом $\to$

Скобки можно опускать согласно приоритету

---

[1] $\sigma\upsilon\nu\tau\alpha\xi\iota\sigma$ (древнегреческий) — составление, построение, порядок

[2] Ожегов. Толковый словарь: Раздел грамматики — наука о законах соединения слов и о строении предложений

Семантика

Семантика[1] — это значение, смысл

Как определить значение формулы?

- Значения атомов задаются интерпретацией
$$\mathcal{I} : \mathsf{Var} \to \{\mathbf{t}, \mathbf{f}\}$$
$$(\mathbf{t} = \mathbf{true},\ \mathbf{f} = \mathbf{false})$$
- Значение $\mathcal{I}(\varphi)$ составной формулы φ в интерпретации $\mathcal{I}$ определяется так:
$$\begin{array}{lcl}
\mathcal{I}(\varphi \,\&\, \psi) = \mathbf{t} & \Leftrightarrow & \mathcal{I}(\varphi) = \mathbf{t} \text{ и } \mathcal{I}(\psi) = \mathbf{t} \\
\mathcal{I}(\varphi \vee \psi) = \mathbf{t} & \Leftrightarrow & \mathcal{I}(\varphi) = \mathbf{t} \text{ или } \mathcal{I}(\psi) = \mathbf{t} \\
\mathcal{I}(\varphi \to \psi) = \mathbf{t} & \Leftrightarrow & \mathcal{I}(\varphi) = \mathbf{f} \text{ или } \mathcal{I}(\psi) = \mathbf{t} \\
\mathcal{I}(\neg\varphi) = \mathbf{t} & \Leftrightarrow & \mathcal{I}(\varphi) = \mathbf{f}
\end{array}$$

В логике принято использовать немного другие обозначения:
$$\begin{array}{lcl}
\mathcal{I} \models \varphi & \Leftrightarrow & \mathcal{I}(\varphi) = \mathbf{t} \\
\mathcal{I} \not\models \varphi & \Leftrightarrow & \mathcal{I}(\varphi) = \mathbf{f}
\end{array}$$

---

[1] Ожегов, Шведова. Толковый словарь русского языка

Семантика

Пример

$$\mathsf{Var} = \{\mathtt{A}, \mathtt{B}, \ldots\} \qquad \varphi : \mathtt{A} \to \neg\mathtt{B} \qquad \mathcal{I}(\mathtt{A}) = \mathbf{f}, \mathcal{I}(\mathtt{B}) = \mathbf{t}$$

Имеет место следующее:

$\mathcal{I} \not\models \neg\mathtt{B}$ (так как $\mathcal{I}(\mathtt{B}) = \mathbf{t}$)

$\mathcal{I} \models \mathtt{A} \to \neg\mathtt{B}$ (так как $\mathcal{I}(\mathtt{A}) = \mathbf{f}$)

Содержательное пояснение

$\mathtt{A}$: высказывание “Я прогуливаю лекции”

$\mathtt{B}$: высказывание “Из этого выйдет что-то хорошее”

$\mathcal{I}$: мир, в котором я живу

- $\mathcal{I}(\mathtt{A}) = \mathbf{f}$: я прилежно хожу на лекции
- $\mathcal{I}(\mathtt{B}) = \mathbf{t}$: из этого выйдет что-то хорошее

$\mathcal{I} \models \mathtt{A} \to \neg\mathtt{B}$: тот, кто сказал “Если я прогуливаю лекции, то из этого не выйдет ничего хорошего”, **прав**

Семантика

Пример

$\mathsf{Var} = \{\mathtt{A}, \mathtt{B}, \ldots\}$ $\qquad \varphi : \mathtt{A} \to \neg\mathtt{B}$ $\qquad \mathcal{I}(\mathtt{A}) = \mathbf{f}, \mathcal{I}(\mathtt{B}) = \mathbf{f}$

Имеет место следующее:

$\mathcal{I} \not\models \neg\mathtt{B}$ (так как $\mathcal{I}(\mathtt{B}) = \mathbf{f}$)

$\mathcal{I} \models \mathtt{A} \to \neg\mathtt{B}$ (так как $\mathcal{I}(\mathtt{A}) = \mathbf{f}$)

Содержательное пояснение

$\mathtt{A}$: высказывание "Я прогуливаю лекции"

$\mathtt{B}$: высказывание "Из этого выйдет что-то хорошее"

$\mathcal{I}$: мир, в котором я живу

- $\mathcal{I}(\mathtt{A}) = \mathbf{f}$: я прилежно хожу на лекции
- $\mathcal{I}(\mathtt{B}) = \mathbf{f}$: из этого не выйдет ничего хорошего

$\mathcal{I} \models \mathtt{A} \to \neg\mathtt{B}$: тот, кто сказал "Если я прогуливаю лекции, то из этого не выйдет ничего хорошего", **прав**

Семантика

Пример

$\mathsf{Var} = \{\mathtt{A}, \mathtt{B}, \dots\}$ $\qquad \varphi : \mathtt{A} \to \neg\mathtt{B}$ $\qquad \mathcal{I}(\mathtt{A}) = \mathbf{t}, \mathcal{I}(\mathtt{B}) = \mathbf{t}$

Имеет место следующее:

$\mathcal{I} \not\models \neg\mathtt{B}$ (так как $\mathcal{I}(\mathtt{B}) = \mathbf{t}$)

$\mathcal{I} \not\models \mathtt{A} \to \neg\mathtt{B}$ (так как $\mathcal{I}(\mathtt{A}) = \mathbf{t}$ *и* $\mathcal{I}(\neg\mathtt{B}) = \mathbf{f}$)

Содержательное пояснение

$\mathtt{A}$: высказывание "Я прогуливаю лекции"

$\mathtt{B}$: высказывание "Из этого выйдет что-то хорошее"

$\mathcal{I}$: мир, в котором я живу

- $\mathcal{I}(\mathtt{A}) = \mathbf{t}$: я прогуливаю лекции
- $\mathcal{I}(\mathtt{B}) = \mathbf{t}$: из этого выйдет что-то хорошее

$\mathcal{I} \not\models \mathtt{A} \to \neg\mathtt{B}$: тот, кто сказал "Если я прогуливаю лекции, то из этого не выйдет ничего хорошего", **неправ**

Выполнимость и общезначимость

Какие же свойства формул исследуются в логике высказываний?

- Формула φ выполнима ($\| \models \varphi$[1]), если существует интерпретация $\mathcal{I}$, такая что $\mathcal{I} \models \varphi$
- Формула φ невыполнима ($\| \not\models \varphi$), если для любой интерпретации $\mathcal{I}$ верно $\mathcal{I} \not\models \varphi$
- Формула φ общезначима ($\models \varphi$), если для любой интерпретации $\mathcal{I}$ верно $\mathcal{I} \models \varphi$

Встречались ли вам эти свойства формул раньше?

Логика высказываний	*Булева алгебра*
φ выполнима	φ выполнима
φ невыполнима	$\varphi \equiv 0$
φ общезначима	$\varphi \equiv 1$

[1] Это необщеупотребимое обозначение, его придумал я

Выполнимость и общезначимость

Как связаны между собой выполнимость, общезначимость и невыполнимость?

$$
\begin{tikzcd}
\text{формула } \varphi \text{ выполнима} \arrow[d, leftrightarrow, "\varphi = \psi" , "\text{выдать противоположный ответ}"'] \\
\text{формула } \psi \text{ невыполнима} \arrow[d, bend left, "\chi = \neg\psi"] \\
\text{формула } \chi \text{ общезначима} \arrow[u, bend left, "\psi = \neg\chi"]
\end{tikzcd}
$$

Проверка каждого из этих свойств может быть легко **сведена** к проверке каждого другого свойства

Метод семантических таблиц

$$
\begin{tikzcd}
\text{формула } \varphi \text{ выполнима} \arrow[d, leftrightarrow, "\begin{array}{l}\varphi = \psi \\ \text{выдать противоположный ответ}\end{array}"] \\
\text{формула } \psi \text{ невыполнима} \arrow[d, bend left, "\chi = \neg\psi"] \\
\text{формула } \chi \text{ общезначима} \arrow[u, bend left, "\psi = \neg\chi"]
\end{tikzcd}
$$

Далее рассматриваем задачу

проверки общезначимости (булевых) формул

Метод семантических таблиц

Как проверить общезначимость формулы φ?

Булева алгебра: вычислить столбец значений реализуемой функции и проверить, содержит ли он хотя бы один 0

В терминах логики это перебор всех интерпретаций и вычисление значения φ в каждой из них

Можно ли сократить этот перебор?

Да, и это основное назначение метода семантических таблиц:

- предположим, что формула φ необщезначима, и попробуем построить интерпретацию $\mathcal{I}$, такую что $\mathcal{I} \not\models \varphi$
- на каждом шаге построения имеем формулы, предполагаемые истинными и ложными в $\mathcal{I}$, и из готовых предположений получаем новые с "более простыми" формулами
- если **все** предположения привели к противоречивым требованиям к $\mathcal{I}$, то формула признаётся общезначимой

Метод семантических таблиц

Семантическая таблица (логики высказываний) — это пара множеств формул: $T = \langle\ \Gamma \mid \Delta\ \rangle$

Таблица T выполнима, если существует интерпретация $\mathcal{I}$, такая что для любой формулы $\varphi \in \Gamma$ верно $\mathcal{I} \models \varphi$, а для любой формулы $\psi \in \Delta$ верно $\mathcal{I} \not\models \psi$

закрыта, если $\Gamma \cap \Delta \neq \emptyset$

атомарна, если все формулы из Γ и Δ атомарны

Содержательное пояснение:

Таблица T — это предположение о том, что формулы из Γ истинны, а формулы из Δ ложны

Выполнимая таблица — это предположение, верное хотя бы в одной интерпретации

Закрытая таблица — это очевидно неверное предположение

Атомарная незакрытая таблица — это явное описание интерпретаций $\mathcal{I}$, в которых предположение верно

Метод семантических таблиц

Для краткости будем опускать фигурные скобки в записи множеств семантической таблицы

Утверждение

$\models \varphi \Leftrightarrow$ **семантическая таблица** $\langle\ \emptyset \mid \varphi\ \rangle$ **невыполнима**

Утверждение

Любая закрытая таблица невыполнима

Утверждение

Любая незакрытая атомарная таблица выполнима

Доказательство.

Самостоятельно (это очень просто)

Метод семантических таблиц

Чтобы доказать общезначимость φ, достаточно разработать правила преобразования таблиц, позволяющие свести неявное противоречие (невыполнимую таблицу) к явному противоречию (набору закрытых таблиц)

Доказательства такого вида называются логическим выводом

Вывод, в котором участвуют семантические таблицы, принято называть табличным, а правила преобразования семантических таблиц — правилами табличного вывода

Как же выглядят эти правила?

Метод семантических таблиц

Правила табличного вывода имеют следующий вид:

$$(*)\colon\ \frac{T_0}{T_1}, \qquad (**)\colon\ \frac{T_0}{T_1,\ T_2},$$

где T_0, T_1, T_2 — семантические таблицы

Правила прочитываются так:

(*): **таблица T_0 выполнима тогда и только тогда, когда выполнима таблица T_1**

(**): **таблица T_0 выполнима тогда и только тогда, когда выполнима хотя бы одна из таблиц T_1, T_2**

Таблицы T_1, T_2 в (**) — это альтернативы

Метод семантических таблиц

Правила табличного вывода для логики высказываний:

$$L\& \quad \frac{\langle\ \Gamma,\ \varphi\,\&\,\psi \mid \Delta\ \rangle}{\langle\ \Gamma,\ \varphi,\ \psi \mid \Delta\ \rangle} \qquad R\& \quad \frac{\langle\ \Gamma \mid \Delta,\ \varphi\,\&\,\psi\ \rangle}{\langle\ \Gamma \mid \Delta,\ \varphi\ \rangle, \langle\ \Gamma \mid \Delta,\ \psi\ \rangle}$$

$$L\vee \quad \frac{\langle\ \Gamma,\ \varphi \vee \psi \mid \Delta\ \rangle}{\langle\ \Gamma,\ \varphi \mid \Delta\ \rangle, \langle\ \Gamma,\ \psi \mid \Delta\ \rangle} \qquad R\vee \quad \frac{\langle\ \Gamma \mid \Delta,\ \varphi \vee \psi\ \rangle}{\langle\ \Gamma \mid \Delta,\ \varphi,\ \psi\ \rangle}$$

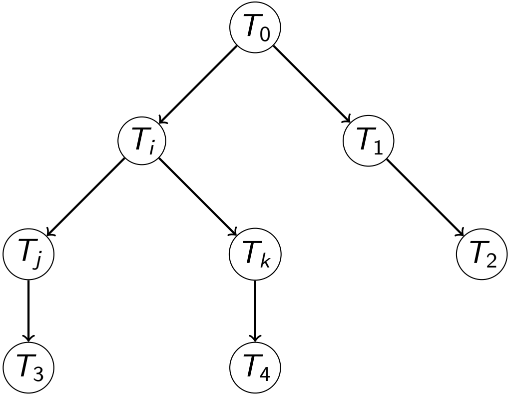
$$L\!\to \quad \frac{\langle\ \Gamma,\ \varphi \to \psi \mid \Delta\ \rangle}{\langle\ \Gamma,\ \psi \mid \Delta\ \rangle, \langle\ \Gamma \mid \Delta,\ \varphi\ \rangle} \qquad R\!\to \quad \frac{\langle\ \Gamma \mid \Delta,\ \varphi \to \psi\ \rangle}{\langle\ \Gamma,\ \varphi \mid \Delta,\ \psi\ \rangle}$$

$$L\neg \quad \frac{\langle\ \Gamma,\ \neg\varphi \mid \Delta\ \rangle}{\langle\ \Gamma \mid \Delta,\ \varphi\ \rangle} \qquad R\neg \quad \frac{\langle\ \Gamma \mid \Delta,\ \neg\varphi\ \rangle}{\langle\ \Gamma,\ \varphi \mid \Delta\ \rangle}$$

Метод семантических таблиц

Табличный вывод для таблицы T_0 — это корневое дерево, такое что:

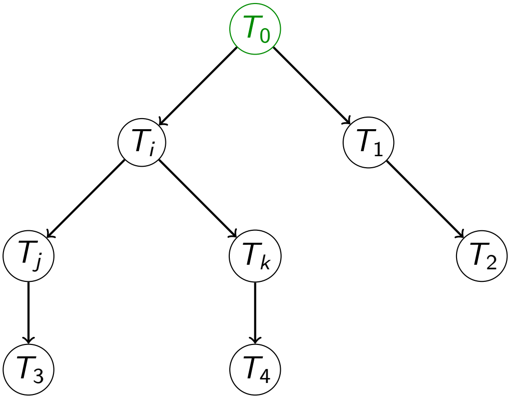
1. его вершинам приписаны семантические таблицы



Метод семантических таблиц

Табличный вывод для таблицы T_0 — это корневое дерево, такое что:

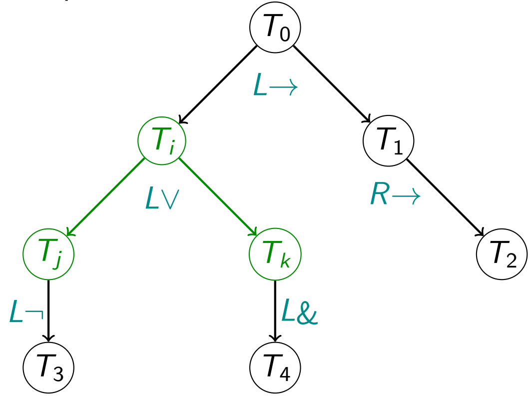
2. его корню приписана таблица T_0



Метод семантических таблиц

Табличный вывод для таблицы T_0 — это корневое дерево, такое что:

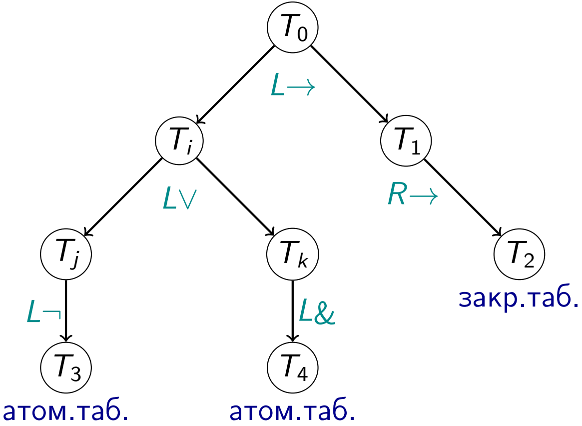
3. из T_i исходят дуги в T_j (и T_k) $\quad\Leftrightarrow$
   $\dfrac{T_i}{T_j, (T_k)}$ — правило табличного вывода



Метод семантических таблиц

Табличный вывод для таблицы T_0 — это корневое дерево, такое что:

4. метки его листьев — закрытые или атомарные таблицы



Метод семантических таблиц

Табличный вывод успешен, если он конечен и **все** его листья помечены закрытыми таблицами

Зачем нужен успешный табличный вывод?

Он показывает, что таблица, для которой он построен, невыполнима

Если он построен для таблицы $\langle\ \emptyset \mid \varphi\ \rangle$, то $\models \varphi$

Утверждение
Любой табличный вывод в логике высказываний конечен

Доказательство.
Глубина вывода для таблицы $\langle\ \Gamma \mid \Delta\ \rangle$ не превосходит $N + 1$, где N — суммарное число связок в формулах из Γ, Δ ▼

Метод семантических таблиц

Примеры табличного вывода

$$\langle \ \mid \mathtt{A} \to \mathtt{A} \vee \mathtt{B} \ \rangle$$

$$\big\downarrow R{\to}$$

$$\langle \ \mathtt{A} \mid \mathtt{A} \vee \mathtt{B} \ \rangle$$

$$\big\downarrow R\&$$

$$\langle \ \mathtt{A} \mid \mathtt{A},\ \mathtt{B} \ \rangle$$

закрытая таблица

Вывод успешен: $\models \mathtt{A} \to \mathtt{A} \vee \mathtt{B}$

$$\langle \ \mid \mathtt{A} \vee \mathtt{B} \to \mathtt{A} \ \rangle$$

$$\big\downarrow R{\to}$$

$$\langle \ \mathtt{A} \vee \mathtt{B} \mid \mathtt{A} \ \rangle$$

$$\swarrow \quad L\vee \quad \searrow$$

$$\langle \ \mathtt{A} \mid \mathtt{A} \ \rangle \qquad\qquad \langle \ \mathtt{B} \mid \mathtt{A} \ \rangle$$

незакрытая атомарная таблица

Вывод неуспешен: $\not\models \mathtt{A} \vee \mathtt{B} \to \mathtt{A}$

Метод семантических таблиц

Лемма корректности правил вывода

Каково бы ни было правило табличного вывода $L\&$, $R\&$, $L\vee$, $R\vee$, $L\to$, $R\to$, $L\neg$, $R\neg$

$$\frac{T_0}{T_1,\ (T_2)},$$

таблица T_0 выполнима тогда и только тогда, когда выполнима таблица T_1 (или выполнима таблица T_2)

Доказательство.

Подробно остановимся только на правиле $L\to$:

$$\frac{\langle\ \Gamma,\ \varphi \to \psi \mid \Delta\ \rangle}{\langle\ \Gamma,\ \psi \mid \Delta\ \rangle, \langle\ \Gamma \mid \Delta,\ \varphi\ \rangle}$$

Пусть верхняя таблица выполнима: существует интерпретация $\mathcal{I}$, такая что $\mathcal{I} \models \chi'$ для $\chi' \in \Gamma$, $\mathcal{I} \not\models \chi''$ для $\chi'' \in \Delta$, и $\mathcal{I} \models \varphi \to \psi$

Тогда верно хотя бы одно из двух: $\mathcal{I} \not\models \varphi$, $\mathcal{I} \models \psi$ — а значит, хотя бы одна из нижних таблиц выполнима

Рассуждения в обратную сторону аналогичны ▼

Метод семантических таблиц

Теорема о табличном выводе в логике высказываний

Пусть D — табличный вывод для семантической таблицы T. Тогда T невыполнима в том и только в том случае, если D успешен

Доказательство.

Следует из леммы корректности правил табличного вывода, конечности табличного вывода и невыполнимости закрытых таблиц ▼

Следствие

$\models \varphi \Leftrightarrow$ **для таблицы** $\langle\ \emptyset \mid \varphi\ \rangle$ **существует успешный табличный вывод**

Следствие

$\models \varphi \Leftrightarrow$ **все выводы для таблицы** $\langle\ \emptyset \mid \varphi\ \rangle$ **успешны**

SAT

формула φ выполнима

$\varphi = \psi$

выдать противоположный ответ

формула ψ невыполнима

$\psi = \neg\chi$ $\chi = \neg\psi$

формула χ общезначима

Теперь рассматриваем задачу

проверки выполнимости (булевых) формул

SATisfiability

(для экономии места будем опускать “&”, а вместо $\neg \mathtt{x}$ писать $\overline{\mathtt{x}}$)

(и еще будем писать 0/1 вместо **true**/**false** для значений переменных)

SAT: приложения

Зачем столько внимания задаче SAT?

Что "практически" полезного можно решить с её помощью?

Приведём несколько примеров и областей и показать, как задачи из этих областей сводятся к SAT

Но сначала несколько обозначений:

- если $\widetilde{\mathrm{x}}^{\mathrm{n}}$ — в точности все переменные формулы φ, то $\mathcal{I}_\varphi : \widetilde{\mathrm{x}}^{\mathrm{n}} \to \{0, 1\}$ — интерпретация, в которой выполняется формула φ:
$$\mathcal{I}_\varphi \models \varphi$$
- $\exists \mathbf{x}\ \varphi$ — это формула $\varphi_0 \vee \varphi_1$, где φ_i получается из φ подстановкой значения i на место переменной x
- $\varphi \leftrightarrow \psi$ — это сокращение для формулы $\varphi \to \psi \mathbin{\&} \psi \to \varphi$

SAT: приложения

Пример

Женя, Аня, Катя и Петя решили собраться вечером и поучить логику за чашкой кофе

Неожиданно выяснилось, что

- Женя по будням много работает, а по выходным отсыпается, так что может встречаться только по понедельникам, средам и четвергам
- у Ани по пятницам спецсеминар
- Катя по средам вечером занимается плаванием
- а Петя по вторникам ходит на бокс, а каждый четверг пьёт пиво со школьными друзьями

Когда же им собираться?

$$\varphi : (\texttt{Mon} \vee \texttt{Wed} \vee \texttt{Thu})\ \overline{\texttt{Fri}}\ \overline{\texttt{Wed}}\ \left(\overline{\texttt{Tue}}\ \overline{\texttt{Thu}}\right)$$

Формула φ выполнима, и единственная интерпретация $\mathcal{I}_\varphi$:

$$\mathcal{I}_\varphi(\texttt{Mon}) = 1,\ \mathcal{I}_\varphi(\texttt{Tue}) = \cdots = \mathcal{I}_\varphi(\texttt{Sun}) = 0$$

Значит, из всей недели только понедельник подходит для логики

SAT: приложения

Пример: *проверка достижимости вершины в графе*

$$i \longrightarrow v \longrightarrow f$$

Достижима ли вершина f из i?

Закодируем вершины:

$$i \sim 00, \quad v \sim 01, \quad f \sim 10$$

Перепишем задачу как набор формул:

- вершины i и f: $I(\mathtt{x}_1, \mathtt{x}_2) = \overline{\mathtt{x}_1}\ \overline{\mathtt{x}_2}$, $F(\mathtt{x}_1, \mathtt{x}_2) = \mathtt{x}_1\ \overline{\mathtt{x}_2}$
- описание дуг графа:
  $R(\mathtt{x}_1, \mathtt{x}_2, \mathtt{x}_1{}', \mathtt{x}_2{}') = \overline{\mathtt{x}_1}\ \overline{\mathtt{x}_2}\ \overline{\mathtt{x}_1{}'}\ \mathtt{x}_2{}' \vee \overline{\mathtt{x}_1}\ \mathtt{x}_2\ \mathtt{x}_1{}'\ \overline{\mathtt{x}_2{}'}$

$$\varphi : I(x_1^0, x_2^0)\ \&\ R(x_1^0, x_2^0, x_1^1, x_2^1)\ \&\ R(x_1^1, x_2^1, x_1^2, x_2^2)$$
$$\&(F(x_1^0, x_2^0) \vee F(x_1^1, x_2^1) \vee F(x_1^2, x_2^2))$$

$|\!\models \varphi \Leftrightarrow f$ достижима из i в данном графе, а интерпретацией $\mathcal{I}_\varphi$ описывается путь из i в f

SAT: приложения

Пример: *Bounded Model Checking* (*коротко и “на пальцах”*)

Рассмотрим программу π (*или несколько взаимодействующих программ*), работающую над конечными данными, принимающими значения из области D

Можно ли с помощью SAT убедиться, что заданные “плохие” состояния данных никогда не будут получены программой при недолгой (в пределах k шагов) работе?

Добавим к данным значение счётчика команд: $D' = D \times PC$

Объявим элементы D' вершинами графа и соединим так, как данные преобразуются выполнением инструкций программы

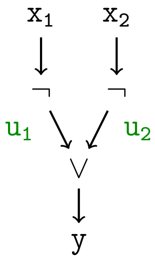
Начальное и “плохие” состояния данных — это вершины графа

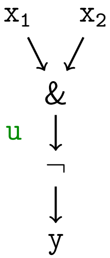
Плохие состояния данных могут быть получены $\Leftrightarrow$ хотя бы одна плохая вершина достижима из начальной в графе

SAT: приложения

Пример: *проектирование схем* (*проверка эквивалентности*)

Рассмотрим две схемы из функциональных элементов:





Как проверить, реализуют ли эти схемы одну и ту же функцию?

Например, так:

$$\exists u_1 \, \exists u_2 \, ((y \leftrightarrow u_1 \vee u_2) \,\&\, (x_1 \leftrightarrow \neg u_1) \,\&\, (x_2 \leftrightarrow \neg u_2)) \oplus$$
$$\exists u \, ((y \leftrightarrow \neg u) \,\&\, (u \leftrightarrow x_1 \,\&\, x_2))$$
$$\stackrel{?}{\equiv} 0$$

SAT: теория и практика решения

Насколько эффективно можно решать задачу SAT?

Есть два взгляда на этот вопрос:

- Теоретический
  - Каковы наилучшие оценки сложности решающего алгоритма?
  - Где находится задача в иерархии классов сложности?
- Практический
  - Можно ли написать программу, которая будет решать задачу за разумное время?

SAT — яркий пример задачи, для которой возможности эффективного решения сильно различаются с теоретической и практической точек зрения

SAT: теория и практика решения

Теоретический взгляд

Теорема Кука

SAT — NP-полная задача

Насколько это плохо для практического решения?

- Задачу SAT можно решить[1]
- Время работы решающего алгоритма неразумно велико[2]

Небольшая оговорка

Когда говорят о задаче SAT, часто имеют в виду проверку выполнимости формул, представленных в конъюнктивной нормальной форме (КНФ)

Такое допущение не ограничивает общности задачи: можно *быстро* привести произвольную формулу к КНФ

---

[1] за полиномиальное время на недетерминированной машине Тьюринга

[2] если $P \neq NP$, то время работы неполиномиально; но если **вдруг** окажется, что $P = NP$, то её могут научиться решать быстро

SAT: теория и практика решения

Практический взгляд

Два основных “класса” **эффективных** решающих алгоритмов:

DPLL[1], локальный поиск

Десятки программных средств

(SAT-решателей, или SAT-солверов):

MiniSAT, zChaff, RSat, Lingeling, WinSAT, ...

Решение для **миллионов** переменных и множителей в формулах, возникающих на практике

Конференции, соревнования, использование в индустрии и науке

И при этом

никто не понимает, почему оно работает

---

[1] **D**avis–**P**utnam–**L**ogemann–**L**oveland

SAT: локальный поиск

Схема работы алгоритмов локального поиска:

1. выбрать начальный набор значений переменных
2. если на полученном наборе КНФ истинна, то завершить работу
3. иначе выбрать переменную, изменить её значение и перейти к (2)
4. по необходимости выбрать новый набор значений переменных или досрочно завершить работу

- Выбор осуществляется с привлечением случайности
- цель “блужданий” по наборам — получить как можно больше истинных множителей
- если КНФ невыполнима или подходящий выполняющий набор не найден, то ответ не определён

SAT: DPLL

Схема работы DPLL-алгоритмов для КНФ C:

```
DPLL(cnf C):
  if(C == "true") return true
  if(C == "false") return false
  pick X: X is variable of C
  pick b: b = 0 or b = 1
  if(DPLL(C{X=b})) return true
  return DPLL(C{X=!b})
```

Дополнительные оптимизации:

- свёртка констант: если в множителе КНФ осталось одно слагаемое $\mathtt{X}$ (или $\overline{\mathtt{X}}$), то немедленно подставить значение $\mathtt{X} = 1$ (или $\mathtt{X} = 0$)
- деполяризация: если переменная $\mathtt{X}$ входит в C только без отрицания (или только с отрицанием), то немедленно подставить значение $\mathtt{X} = 1$ (или $\mathtt{X} = 0$)

SAT: DPLL

Пример

$$(x_1 \vee x_2 \vee x_3)(x_1 \vee \overline{x_2} \vee x_3)(\overline{x_1} \vee \overline{x_2})(\overline{x_2} \vee \overline{x_3})$$

$$\xrightarrow{x_2 = 1}$$

$$(x_1 \vee x_3)\,\overline{x_1}\,\overline{x_3}$$

Пояснения:

Выбираем переменную x_2 и значение 1

SAT: DPLL

Пример

$$\begin{tikzcd}[row sep=large]
(x_1 \vee x_2 \vee x_3)\,(x_1 \vee \overline{x_2} \vee x_3)\,(\overline{x_1} \vee \overline{x_2})\,(\overline{x_2} \vee \overline{x_3}) \arrow[d, "x_2 = 1"'] \\
(x_1 \vee x_3)\,\overline{x_1}\,\overline{x_3} \arrow[d, "x_1 = x_3 = 0"'] \\
\textbf{false}
\end{tikzcd}$$

Пояснения:

Свёртка констант для x_1, x_3

SAT: DPLL

Пример

$$\begin{tikzcd}
(x_1 \lor x_2 \lor x_3)\,(x_1 \lor \overline{x_2} \lor x_3)\,(\overline{x_1} \lor \overline{x_2})\,(\overline{x_2} \lor \overline{x_3}) \arrow[d, "x_2 = 1"'] \\
(x_1 \lor x_3)\,\overline{x_1}\,\overline{x_3} \arrow[d, "x_1 = x_3 = 0"'] \arrow[u, bend right] \\
\textbf{false} \arrow[u, bend right]
\end{tikzcd}$$

Пояснения:

Появился ложный множитель: откат

SAT: DPLL

Пример

$$
\begin{tikzcd}
 & (x_1 \lor x_2 \lor x_3)\,(x_1 \lor \overline{x_2} \lor x_3)\,(\overline{x_1} \lor \overline{x_2})\,(\overline{x_2} \lor \overline{x_3}) \arrow[dl, "x_2 = 1"'] \arrow[dr, "x_2 = 0"] & \\
(x_1 \lor x_3)\,\overline{x_1}\;\overline{x_3} \arrow[d, "x_1 = x_3 = 0"'] \arrow[ur, bend right] & & x_1 \lor x_3 \\
\mathbf{false} \arrow[u, bend right] & &
\end{tikzcd}
$$

Пояснения:

Переходим к оставшемуся значению 0 для x_2

SAT: DPLL

Пример

$$
\begin{tikzcd}
 & (x_1 \vee x_2 \vee x_3)(x_1 \vee \overline{x_2} \vee x_3)(\overline{x_1} \vee \overline{x_2})(\overline{x_2} \vee \overline{x_3}) \arrow[dl, "x_2 = 1"'] \arrow[dr, "x_2 = 0"] & \\
(x_1 \vee x_3)\,\overline{x_1}\,\overline{x_3} \arrow[d, "x_1 = x_3 = 0"'] \arrow[ur, bend right] & & x_1 \vee x_3 \arrow[d, "x_1 = x_3 = 1"] \\
\textbf{false} \arrow[u, bend right] & & \textbf{true}
\end{tikzcd}
$$

Пояснения:

Деполяризация для x_1 и x_3

SAT: DPLL

Пример

$$
\begin{tikzcd}[column sep=small]
& (x_1 \vee x_2 \vee x_3)\,(x_1 \vee \overline{x_2} \vee x_3)\,(\overline{x_1} \vee \overline{x_2})\,(\overline{x_2} \vee \overline{x_3}) \arrow[dl, "x_2 = 1"'] \arrow[dr, "x_2 = 0"] & \\
(x_1 \vee x_3)\,\overline{x_1}\,\overline{x_3} \arrow[d, "x_1 = x_3 = 0"'] & & x_1 \vee x_3 \arrow[d, "x_1 = x_3 = 1"] \\
\textbf{false} & & \textbf{true}
\end{tikzcd}
$$

Пояснения:

СТОП: формула выполнима

Конец лекции 2