1 Автоматизированный вывод

1.1 Проблема автоматизированного вывода

Зафиксируем некоторую сигнатуру σ для всех сущностей, которые будут рассмотрены в рамках этой лекции.

Задача I

Дано множество предложений Φ и предложение ϕ, определить, верно ли

$$\Phi \models \phi$$

Очевидно, что если рассматривать эту задачу семантически, то не существует алгоритма её решения. Но используя теорему об адекватности, она может быть сведена к выводимости ϕ из Φ - задаче которая может быть решена алгоритмически.

Задача II

Дано множество предложений Φ и предложение ϕ, определить, верно ли

$$\Phi \vdash \phi$$

Итак, необходимо алгоритмически определить, верно ли, что $\Phi \vdash \phi$. Предположим, что Φ конечно.

Сокращение 1

Первоначальная задача может быть переформулирована как ($n \geq 0$):

$$\triangleright \{\phi_1, \ldots, \phi_n\} \vdash \phi_0$$

Рассматривая конъюнкцию предпосылок и вводя импликацию, задача может быть сведена к вопросу о выводимости одного предложения.

Сокращение 2

Первоначальная задача может быть переформулирована как:

$$\triangleright \psi$$

если мы введем $\psi = \bigwedge_{i=1}^{n} \phi_i \to \phi_0$

1.2 Принципиальное ограничение

Первоначальная задача была сведена к вопросу, верно ли, что $\rhd\phi$. Но при таком подходе возникает принципиальное ограничение.

Теорема (о неполноте, Гедель)

Если предложение ϕ достаточно сильное (а именно включает арифметику Пеано), вопрос, верно ли, что $\rhd\phi$ не разрешим, т.е. не существует алгоритма, отвечающего на этот вопрос.

Теорема (вычислимая перечислимость)

Для любого предложения ϕ существует алгоритм, который либо скажет, что $\rhd\phi$, либо никогда не остановится.

Принимая во внимание вторую теорему, Практический подход к автоматизированному рассуждению показывает, насколько эффективно мы можем построить упомянутый алгоритм.

1.3 Сведение к противоречивости

Первоначальная проблема была сведена к доказуемости одной формулы ϕ. Но этот вопрос, в свою очередь, может быть сведён к вопросу о *противоречивости* некоторого множества. Действительно:

Сведение к противоречивости

Для любых предложений ϕ_i, $\rhd\{\phi_1, \dots, \phi_n\} \vdash \phi_0$ эквивалентно утверждению о том, что множество предложений $\{\phi_1, \dots, \phi_n, \neg\phi_0\}$ противоречиво.

Доказательство

Слева направо:

$$\frac{\phi_1, \dots, \phi_n \vdash \phi_0}{\phi_1, \dots, \phi_n, \neg\phi_0 \vdash \bot}$$

Обратное включение:

$$\frac{\phi_1, \dots, \phi_n, \neg\phi_0 \vdash \bot}{\phi_1, \dots, \phi_n \vdash \phi_0}$$

Таким образом, вопрос о логическом следствии может быть сведён к вопросу о противоречивости.

2 Сколемизация/Гербрендизация

2.1 Нормальные формы Сколема и Гербранда

Формула ϕ называется $\forall$-формулой, тогда и только тогда, когда ϕ не содержит кванторов $\exists$. Аналогично для $\exists$-формул: ϕ называется $\exists$-формулой, тогда и только тогда, когда ϕ не содержит кванторов $\forall$.

Нормальная форма Сколема

Начнем с некоторой формулы $\phi = Q_1 x_1 \dots Q_n x_n \psi$, находящейся в пренексной нормальной форме. Определим **Нормальную форму Сколема** (или **Сколемизацию**) $Sk(\phi)$ формулы ϕ следующим образом.

- если ϕ является $\forall$-формулой, то $Sk(\phi) = \phi$
- в противном случае $\phi = \forall x_1 \dots \forall x_n \exists y \psi(\bar{x}, y)$ и для некоторого *нового n-местного функционального символа* f возьмём

$$Sk(\phi) = Sk(\forall x_1 \dots \forall x_n \psi(\bar{x}, f(\bar{x})))$$

пример:

$$Sk(\forall x \exists y \forall z (p(x,y) \to q(f(y), z) \land y = h(z))) =$$
$$\forall x \forall z (p(x, g(x)) \to q(f(g(x)), z) \land g(x) = h(z))$$

Нормальная форма Гербранда

Начнем с некоторой формулы $\phi = Q_1 x_1 \dots Q_n x_n \psi$, находящейся в пренексной нормальной форме. Определим **Нормальную форму Гербранда** (или **Гербрендизацию** $Hb(\phi)$ формулы ϕ следующим образом.

- если ϕ является $\exists$-формулой, то $Hb(\phi) = \phi$
- в противном случае $\phi = \exists x_1 \dots \exists x_n \forall y \psi(\bar{x}, y)$ и для некоторого *нового n-местного функционального символа* f возьмём

$$Sk(\phi) = Sk(\exists x_1 \dots \exists x_n \psi(\bar{x}, f(\bar{x})))$$

Отметим, что во втором пункте этого определения, параметр n *может быть равен* 0. В таком случае символ f является нулярным, т.е. это новая константа c, и в этом случае: $\phi = \forall y \psi(y)$ и $Sk(\phi) = \psi(c)$.

2.2 Свойства сколемизации и гербрендизации

Отметим, что $\phi \not\sim Sk(\phi)$ и $\phi \not\sim Hb(\phi)$. Но, тем не менее, существуют теоремы:

Теорема (о сколемизации)

Для любой формулы ϕ верно следующее: ϕ выполнима $\Leftrightarrow$ $Sk(\phi)$ также выполнима.

Доказательство

Индукция по количеству кванторов $\exists$ n. Если $n = 0$, то $Sk(\phi) = \phi$ и доказывать нечего. Шаг индукции. Предположим, что $\forall x_1 \dots \forall x_n \psi(\bar{x}, f(\bar{x}))$ выполнима. Тогда понятно, что $\forall x_1 \dots \forall x_n \exists y \psi(\bar{x}, y)$ будет также выполнима, потому что значение $y = f(\bar{x})$ - следствие такой переменной y. Обратное включение, если $\forall x_1 \dots \forall x_n \exists y \psi(\bar{x}, y)$ является выполнимой в некоторой модели $\mathcal{M}$, то для любого кортежа $\bar{a} \in M$ (состоящего из значений x_i) существует такой элемент b (значение y) что $\mathcal{M} \models \psi(\bar{a}, b)$. Можно определить означивание f, каждому $\bar{a}$ сопоставляя b, и в таком случае $\mathcal{M} \models \psi(\bar{a}, f(\bar{b}))$, следовательно, $Sk(\phi)$ выполнима.

Теорема (о гербрендизации)

Для любой формулы ϕ верно следующее: $\models \phi \Leftrightarrow \models Hb(\phi)$

Доказательство

Эта теорема может быть сведена к предыдущей, если заметить, что:

- ϕ тождественно истинна $\Leftrightarrow$ $\neg\phi$ невыполнима
- $Hb(\phi) \sim \neg Sk(\neg\phi)$

Тогда $\models \phi \Leftrightarrow \neg\phi$ невыполнима $\Leftrightarrow Sk(\neg\phi)$ невыполнима $\Leftrightarrow \models \neg Sk(\neg\phi) \Leftrightarrow \models Hb(\phi)$.

Таким образом, сколемизация и гербрендизация помогают нам избавиться от кванторов в вопросах тождественной истинности/выполнимости формул.

3 Унификация

3.1 Подстановки

Определение

Подстановка θ - это конечное отображение из некоторого множества переменных V на множество термов $T(\sigma)$. Таким образом, $\theta : V \to T(\sigma)$. Также для определения подстановки будут использоваться следующие обозначения:

$$\theta = \{x_1/t_1, \dots, x_n/t_n\}$$

при условии, что $\theta(x_i) = t_i$ для всех $1 \leq i \leq n$. Отметим, что по определению мы предполагаем, что $x_i \neq t_i$ для всех i.

Определение (композиция подстановок)

Для любых двух подстановок $\theta_1 = \{x_i/t_i | i \leq n\}$ и $\theta_2 = \{y_j/s_j | j \leq m\}$ существует подстановка $\theta_1 \circ \theta_2$ - **композиция** θ_1 и θ_2, определённая следующим образом:

$$\{x_i/\theta_2(t_i) | i \leq n, x_i \neq \theta_2(t_i)\} \cup \{y_j/s_j | j \leq m, y_j \notin \{x_1, \dots, x_n\}\}$$

3.2 Свойства композиции

Предложение

Для любых двух подстановок $\theta_1 = \{x_i/t_i | i \leq n\}$ и $\theta_2 = \{y_j/s_j | j \leq m\}$ и для любого выражения e верно следующее $(\theta_1 \circ \theta_2)(e) = \theta_2(\theta_1(e))$

Доказательство

Прежде всего отметим, что достаточно доказать это утверждение только для переменных, т.е. $e = v$ - некоторая переменная, потому что подстановка действует на выражения локально (на вхождения переменных). Тогда существует два множества переменных: $X = \{x_1, \dots, x_n\}$ и $Y = \{y_1, \dots, y_m\}$, возможны 4 случая, в зависимости от того, $v \in X$, и аналогично для $v \in Y$. Например, если $v \notin X$ и $v \notin Y$, то $v = (\theta_1 \circ \theta_2)(v) = \theta_2(v) = \theta_2(\theta_1(v))$. Остальные случаи рассматриваются по определению $\circ$.

Следствие

Операция $\circ$ ассоциативна: $(\theta_1 \circ \theta_2) \circ \theta_3 = \theta_1 \circ (\theta_2 \circ \theta_3)$

3.3 Унификаторы

Определение (унификатор)

Дано множество выражений E и подстановка θ, θ называется **унификатором** E тогда и только тогда, когда для любых $e_1, e_2 \in E$, $\theta(e_1) = \theta(e_2)$.

Унификатор может как существовать, так и нет. Кроме того, в абсолютном смысле он не единственен, потому что, комбинируя унификатор с заменой переменных, можно получить другой унификатор.

Определение

Даны две подстановки θ_1 и θ_2, Подстановка θ_1 называется **более общий** чем θ_2, тогда и только тогда, когда существует такая подстановка δ, что $\theta_2 \subseteq \theta_1 \circ \delta$. Это отношение обозначается как $\theta_1 \geq \theta_2$.

пример: $\{x_1/f(y_1), x_2/y_2\}$ более общая чем $\{x_1/f(g(z_1, z_2)), x_2/h(z_1)\}$, потому что

$$\{x_1/f(g(z_1, z_2)), x_2/h(z_1)\} \subseteq \{x_1/f(y_1), x_2/y_2\} \circ \{y_1/g(z_1, z_2), y_2/h(z_1)\}$$

3.4 наиболее общие унификаторы

Определение (наиболее общий унификатор)

Дано множество выражений E и подстановка θ, θ называется **наиболее общим унификатором** тогда и только тогда, когда

- θ является унификатором для E,
- для любого другого унификатора θ' выражения E верно, что $\theta \geq \theta'$, т.е. θ более общий чем θ'.

Отметим, что наиболее общий унификатор не единственен (если он существует), потому что унификатор, полученный в результате комбинации любого наиболее общего унификатора θ с заменой переменных, также является наиболее общим.

Алгоритм унификации

Существует эффективный алгоритм, вычисляющий некоторый наиболее общий унификатор для любого множества выражений E или определяющий, что наиболее общего унификатора для E не существует.

4 Хорновские дизъюнкты

4.1 Хорновские дизъюнкты

Определение (литерал)

Литерал - это атомарная формула или её отрицание.

Определение (Хорновский дизъюнкт)

Хорновский дизъюнкт - это дизъюнкция литералов, т.е. это элементарная дизъюнкция. Существует специальный Хорновский дизъюнкт - пустой дизъюнкт, обозначаемый как $\square$ и означающий ложную формулу $\perp$.

Примеры:

- $p(x, f(y)) \vee \neg q(x, x) \vee p(h(x, y), h(y, f(x)))$
- $p(x, f(y), z) \vee q(x) \vee p(h(x, y, z), h(y, f(x)), x) \vee \neg q(f(y))$
- $\neg p(f(f(f(x)))) \vee p(x)$

Соглашение. Поскольку Хорновские дизъюнкты - это просто дизъюнкция, далее будем считать, что Хорновский дизъюнкт - это *множество* всех литералов, входящих в его состав:

$$h = l_1 \vee \ldots \vee l_k = \{l_1, \ldots, l_k\}$$

4.2 Правило резолюции

Определение (резолюция)

Даны два Хорновских дизъюнкта h_1 и h_2, если существует два литерала $p(\bar{t}) \in h_1$ и $\neg p(\bar{s}) \in h_2$ таких, что кортежи термов $\bar{t}$ и $\bar{s}$ унифицируемы, применимо правило **резолюции** вывода, и если θ - наиболее общий унификатор $\bar{t}$ и $\bar{s}$, то:

$$\frac{h_1 \qquad h_2}{\theta((h_1 \cup h_2) \setminus \{p(\bar{t}), \neg p(\bar{s})\})}(Res)$$

Теорема (корректность резолюции)

Если для некоторых Хорновских дизъюнктов h_1, h_2 и h_0 верно, что

$$\frac{h_1 \quad h_2}{h_0}(Res)$$

и в некоторой структуре $\mathcal{M} \models h_1 \wedge h_2$ (означающей, что $\mathcal{M} \models \forall \bar{x} h_i(\bar{x})$ для $i \in \{1, 2\}$), то $\mathcal{M} \models h_0$.

Доказательство

Предположим, что существуют такие литералы, что $h_1 = \{p(\bar{t})\} \cup h_1'$, $h_2 = \{\neg p(\bar{s})\} \cup h_2'$ и кортежи термов $\bar{t}$ и $\bar{s}$ унифицируемы при помощи наиболее общего унификатора θ, таким образом,

$$\frac{h_1 \qquad h_2}{\theta(h_1' \cup h_2')}(Res)$$

Также предположим, что $\mathcal{M} \models h_1 \wedge h_2$ но $\mathcal{M} \not\models (h_1 \cup h_2) \setminus \{p(\bar{t}), \neg p(\bar{s})\}$. Тогда существует такой кортеж $\bar{a} \in M$, что $\mathcal{M} \models \neg\theta(h_1' \cup h_2')(\bar{a})$. Поскольку h_i' - дизъюнкции, все литералы из $\theta(h_i')(\bar{a})$ ложны в $\mathcal{M}$. Но так как h_1 и h_2 тождественно истинны на $\mathcal{M}$, $\mathcal{M} \models \theta(p(\bar{t}))(\bar{a})$ и $\mathcal{M} \models \theta(\neg p(\bar{s}))(\bar{a})$. Но $\theta(p(\bar{t}) = p(\bar{q}) = \theta(p(\bar{s}))$, следовательно, получаем

$$\mathcal{M} \models p(\bar{q})(\bar{a}) \wedge \neg p(\bar{q})(\bar{a})$$

это противоречие завершает доказательство. $\square$

Следствие (корректности резолюции)

Для любого множества Хорновских дизъюнктов H, если по правилу резолюции из H может быть выведен пустой Хорновский дизъюнкт $\square$, то H противоречиво.

Доказательство

Индукцией по высоте дерева вывода можно показать, что в исчислении резолюций для любого дерева вывода, если Хорновский дизъюнкт h выводим из H, то из истинности всех H следует истинность h. Следовательно, если $h = \square$, то H не может быть истинным т.е. он является противоречивым.

Теорема (полнота резолюции)

Для любого множества Хорновских дизъюнктов H, если H противоречив, то по правилу резолюции из H может быть выведен пустой дизъюнкт $\square$.

4.3 Общий алгоритм опровержения

Напомним исходную задачу: определить истинность импликации

$$\phi_1, \ldots, \phi_n \models \phi_0$$

Эта задача была сведена к вопросу о невыполнимости (противоречивости)

$$\phi_1 \wedge \ldots \wedge \phi_n \wedge \neg\phi_0$$

Сколемизацией этот вопрос был сведён к противоречивости бескванторной формулы ψ, которая может быть преобразована в КНФ. Все элементарные дизъюнкции в этой КНФ будут соответствовать некоторому Хорновскому дизъюнкту h_i, составляющему множество Хорновских дизъюнктов H. Итак, наконец, первоначальная задача была сведена к противоречивости H, которая может быть определена методом резолюции.