Abstract Interpretation

枫聆

2021 年 4 月 13 日

目录

Motivation

单调分析框架的诞生让做程序分析的人可以构造一个精准的，数学形式化的分析. 在这个框架下，我们需要一个 lattice，每个 instruction 相关联的 transfer functuons 和一个初始的状态. 几乎先有的程序分析手法都可以总结到这个单调分析框架上.

那么抽象解释可以理解为在单调分析框架上又迈出了一步. 我们经常的设计分析思路：开始从一个简单的分析开始，分析结果的正确性可以很容易得到证明. 这里的分析是指 collecting semantics，即从程序的语义里面收集我们想要的信息. 最开始的分析一般来说它的可行性或者可计算性是比较难处理的或者说根本不可能处理，因为有可能我们关注的 property 所在 lattice 对应的 underlying set 的基数是比较大的，或者不满足一些良好的性质例如 ACC. 因此我们尝试使用近似计算的方法，把 property 所在的 lattice 变小，这个过程可能是一个迭代的过程，直到我们最终可以计算为止或者达到一个理想的效果. 而抽象解释就是提供这样一种系统的方法来帮助我们.

实际分析 domain 太复杂，也许这个实际分析 domain 满足一些不错的性质比如 ACC 等等，但是还是掩盖不了它过于复杂. 所以我们尝试使用某个抽象的 domain 来代替分析，这个抽象的 domain 要比实际分析的 domain 更小，我们分析起来更加得心应手，但是问题是如何构造这个抽象的 domain? 它分析出来的结果是否

可以作为真的 truth? 它的结果能在多大程度上说明一些问题? 后期我们是否可以考虑优化抽象的 domain 来让结果更好一点?

Correctness Relations

Definition 2.1. (Correctness relations). A relation $R \subseteq V \times L$ is said to be a correctness relation iff it satifies the following two conditions:

1. $\forall v, l_1, l_2,\ (v\ \mathcal{R}\ l_1)$ and $(l_1 \leq l_2) \rightarrow (v\ \mathcal{R}\ l_2)$;
2. $\forall v, \forall L' \subseteq L,\ (\forall l \in L',\ (v\ \mathcal{R}\ l)) \rightarrow v\ \mathcal{R}\ (\bigwedge L')$.

这里 V 表示 concrete values，L 表示 abstract values 构成的 lattice. $v\ \mathcal{R}\ l_1$ 表示 l_1 是 v 的一个 approximation.

用自然语言来描述就是 (1) 若 v 对应某个 l_1，那么对于 l_1 的一个 upper approximation l_2，有 $v \in l_2$. (2) 若 v 同时对应多个 abstract values，这里应该是一个 and 的关系，那么 v 可以对应它们的一个 greatest lower bound.

Lemma 2.2. If $\mathcal{R} \subseteq V \times L$ is a correctness relation, then

$$v\ \mathcal{R}\ \top$$
$$(v\ \mathcal{R}\ l_1) \text{ and } (v\ \mathcal{R}\ l_2) \rightarrow v\ \mathcal{R}\ (l_1 \wedge l_2)$$
$$(v\ \mathcal{R}\ l_1) \text{ or } (v\ \mathcal{R}\ l_2) \rightarrow v\ \mathcal{R}\ (l_1 \vee l_2)$$

前面简单的描述了一下 correctness relation 操作含义，但是 correctness relation 中依然模糊是它里面的 lattice 到底在刻画一个怎样东西? 我们的分析中为什么要引入 lattice? 为了让这个 lattice is reasonable，我们来具体定义这个 lattice 的 meet 和 join 操作的内在含义.

Definition 2.3. we need the meet and the join operator in order to combine abstract values:

1. If a value is described by both l_1 and l_2, by combine these two properties, we obtain the more precise information $l_1 \wedge l_2$.
2. If a value is described by either l_1 or l_2, the most precise info that we can infer is $l_1 \vee l_2$.

在上面定义的 operations 的内在含义下，smaller 代表更精准，bigger 代表更安全，更安全就是指考虑的更全面，不会丢掉信息而造成一些潜在的问题. $\wedge$ 等价于逻辑连词的 and，$\vee$ 等价于逻辑连词的 or. 在实际分析的过程中 $\wedge$ 和 $\vee$ 这两个 operations 就要针对我们关注的 properties 来具体定义，但是它的最本质内在含义是每次都是尽可能在不丢失精确度的可能下，去尽可能的提高结果本身的精确度.

Annotation 2.4. 如何取证明一个分析的正确性 To prove the correctness of the analysis, it is su±cient to prove

1. The initial property (abstract state) l_0 is a correct approximation of the initial value (concrete state) $v_0 : v_0 \ \mathcal{R} \ l_0$.

2. Each transition preserves the correctness relation

$$\forall v_1, v_2, l_1, l_2, \ (v_1 \rightsquigarrow v_2) \text{ and } (v_1 \ \mathcal{R} \ l_1) \text{ and } (f_L(l_1) = l_2) \rightarrow v_2 \ \mathcal{R} \ l_2.$$

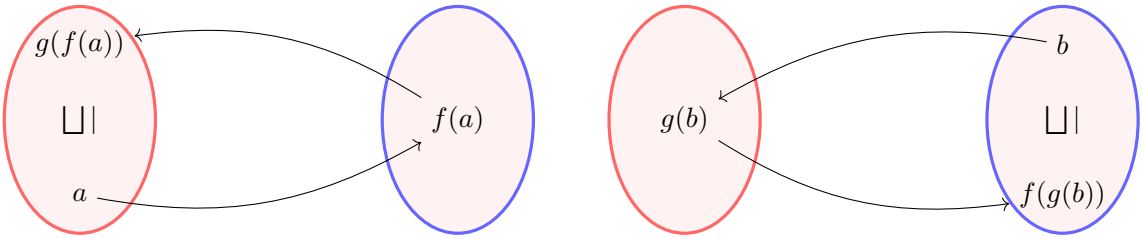
用自然语言来描述就是你首先要保证初始状态下 correctness relation 的存在，而后在状态传递的过程中这个 correctness relation 依然是保持的，这个过程就关系到两个传递函数 concrete value transfer function 和 abstract value transfer function.

Galois connections

关于 galois connection 我们通常可以看到两个定义，下面我来说明两个定义是等价，也就是说可以从任意一个推出另外一个.

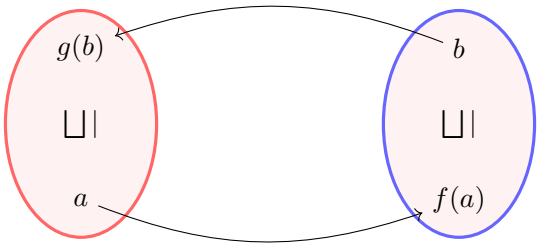
Definition 3.1. nlab 上的定义更贴近 adjunction 的味道 Given posets A and B, a Galois connection between A and B is a pair of order-preversing functions $f\colon A\to B$ and $g\colon B\to A$ such that $a \le g(f(a))$ and $b \ge f(g(b))$ for all $a \in A$, $b \in B$.

注意这里的 order-preversing，最原始的定义用的是 order-reversing，导致我在这里弄出了一些矛盾.



Proposition 3.2. Given posets A and B, a pair of order-preversing functions $f\colon A\to B$ and $g\colon B\to A$ is a Galois connection between A and B if and only if, for all $a \in A$, $b \in B$, we have

$$f(a) \le b \text{ if and only if } a \le g(b).$$



证明. ($\Rightarrow$). 前提 (f,g) 是一个 galois connection，给定 $a \in A$, $b \in B$. 若 $f(a) \le b$, 两边同时 apply g, 有 $g(f(a)) \le g(b)$，同时有 $a \le g(f(a))$. 那么 $a \le g(b)$. 若 $a \le g(b)$，同理可以得到 $f(a) \le b$.

($\Leftarrow$). 前提 (f,g) 满足 $f(a) \le b$ if and only if $a \le g(b)$. 我们直接取 $b = f(a)$, 那么 $f(a) \le f(b)$ 当且仅当 $a \le g(f(a))$. 同理直接取 $a = g(b)$，那么 $g(b) \le g(b)$ 当且仅当 $f(g(a)) \le b$. □

关于 adjunction 的东西 $fg \to id$ 和 $gf \to id$, 这个箭头是一个 natural transform, 至于更细的东西要去看看 category theory 了! PAAA 上说 f 和 g 互为 “weak inverse”，看起来也是比较形象啊！所以有下面的一个命题.

Annotation 3.3. 现在我们尝试把 galois connection 放到抽象解释范畴上，让 lattice A 表示我们原本一个 analysis domain，lattice B 表示一个更抽象的 analysis domain 用来加快我们的分析或者让我们的分析可计算，那么这里 $f\colon A\to B$ 称为abstraction function，$g\colon B\to A$ 称为concretization function.

这里的 f 和 g 都是单调函数意味，原本的实际值之间的关系在 abstract domain 上依然保持，反过来亦然，同时 galois connection 强化的两个条件

1. $a \leq g(f(a))$ 表示抽象过程是可能会丢失精度的，但是依然是正确的. 我的理解是把具体的值通过映射放到抽象域里面做运算得到的结果，再映射回来可能比在具体域里面做运算得到的结果要稍微差一点，但是正确性是可以保证的.

2. $b \geq f(g(b)$ 表示具体化的过程不会丢失精度. 本来抽象值，放到具体域里面操作一遍，再映回来是不会丢精度，这也是可以很自然想到的.

如果更细致去刻画一下我自己的 annotation 就是

1. $x, y \in A$
$$x \vee y \leq g(f(x) \vee f(y)).$$

2. $x', y' \in B$
$$x' \wedge y' \geq f(g(x) \wedge g(y)).$$

为此我们需要去分别证明 $f(x \vee y) = f(x) \vee f(y)$ 和 $g(x \wedge y) = g(x) \wedge g(y)$.

Proposition 3.4. Galois connection 引发的 semilattice homomorphism For a Galois connection (f, g) of join semlattice A and B, f preserves finite join:

1. $f(\bot_A) = \bot_B$;
2. $f(x \vee y) = f(x) \vee f(y)$.

And similarly

1. $g(\top_B) = \top_A$;
2. $g(x \wedge y) = f(x) \wedge f(y)$.

啧啧，没想到啊 galois connection 竟然弄了一个 semilattice homomorphism 出来，突然想找一下 characterization of semilattice homomorphism.

证明. (1) 由于 B 上 $\bot_B$ 的性质，有 $f(\bot_A) \geq \bot_B$. 反过来由于 A 上的 $\bot_A$ 性质，有 $g(\bot_B) \geq \bot_A$，再用一下 galois connection 的性质，有 $f(\bot_A) \leq \bot_B$. 综上两边夹，所以 $f(\bot_A) = \bot_B$.

(2) 由于 f 是 monotone，有 $f(x) \leq f(x \vee y)$ 和 $f(x) \leq f(x \vee y)$，所以 $f(x) \vee f(y) \leq f(x \vee y)$ 是一个 upper bound. 最关键是确界证明 $f(x \vee y) \leq f(x) \vee f(y)$. 由于 galois connection，有 $x \leq g(f(x))$，再由 g 是 monotone，有 $x \leq g(f(x) \vee f(y))$. 同理也有 $y \leq g(f(x') \vee f(y'))$，那么

$$x \vee y \leq g(f(x) \vee f(y))$$

再用一下 galois connection，就有 $f(x \vee y) \leq f(x) \vee f(y)$. 综上两边夹，所以 $f(x \vee y) = f(x) \vee f(y)$. 值得关注是 f 和 g 都只是一个 semilattice homomorphism，而且是两种不同操作. 其实我们做分析的时候，也只是使用一个 semilattice，至于是 meet 还是 join 和原本分析过程的方向是有关的. □

Proposition 3.5. weak inverse For galois connection (f, g), we have the equations

$$f \circ g \circ f = f$$
$$g \circ f \circ g = g.$$

这个命题内在是在说明你把分析结果用 f 和 g 进行迭代是不会改变精度的!

证明. (1).

$$\begin{aligned} a \leq g(f(a)) \Rightarrow \quad & f(a) \leq f \circ (g \circ f(a)) \quad f \circ (g \circ f) \\ f(a) \in B \Rightarrow \quad & (f \circ g) \circ f(a) \leq f(a) \quad (f \circ g) \circ f, \end{aligned}$$

所以 $f \circ g \circ f = f$. 同理可证第二个式子. □

New Analysis: Correctness Relation and Transfer Function

这章讲如何把 analysis function 通过 galois connection 在 analysis domains 之间传递.

Definition 3.6. The correctness relation $\mathcal{S} \subseteq V \times M$ for the new analysis is defined as follows

$$v \ \mathcal{S} \ m \iff v \ \mathcal{R} \ g(m).$$

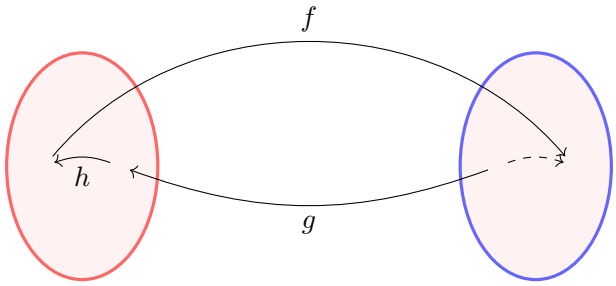
其中的 M 表示新的 analysis domain 或者说是 abstract domain，它还是一个 lattice.这里是在说 abstract domain 上确定的 relations 它同样是在原本的 concrete domain 里面是保持的.

Definition 3.7. 单调分析框架的定义 A specification of an analysis in the extended monotone framework is given by

- an analysis domain D that is semilattice with ordering $\leq$.
- monotone analysis functions $h_n \colon D \to D$ where n is node of program graph.
- an initial element $d_0 \in D$;

Definition 3.8. Galois connection 诱导出来的 analysis function A specification of an analysis in the extended monotone framework and a Galois connection (f, g) from D to the D' gives rise to the induced analysis given by:

- an analysis domain D' that is semilattice with ordering $\leq'$.
- monotone analysis functions $f \circ h_n \circ g \colon D' \to D'$ where n is node of program graph.
- an initial element $f(d_0) \in D$;



搞半天 galois connection 可以构造这样 monotone analysis functions.

在实际中我们不会直接通过上面诱导出来的 analysis functions，因为构造起来太过于繁琐，常常会用一个 h' 来安全的替换它的使用

Proposition 3.9. Assume that we have an analysis specification h'_n and d'_0 over D' statifying

- $f \circ h_n \circ g\ (d') \leq h'_n(d')$;
- $f(d_0) \leq d'_0$.

Furthermore, assume that

- $AA\colon Q \to D'$ (assignment analysis) sloves constraints obtained from the program graph and the analysis specification h' and d'_0 over D'.

Then we also have that

- $AA\colon Q \to D'$ sloves the constraints obtained from the analysis specification $f \circ h_n \circ g$ and $f(d_0)$ over D', and
- $g \circ AA\colon Q \to D'$ solves the constraints obtained from the analysis specification h_n and d_0 over D.

按照条件构造一个新的 analysis function 得到的结果是包含原来的 solution 在里面的，然后再把它还原到本身的 analysis domain 上, 只不过造成 over-approximations 了.

Inducing along the Concretisation Function

Definition 3.10. We shall say that the sequence $(x_1\nabla\cdots\nabla x_n)_n$ eventaully stabilises whenever there is a number N such that $x_1\nabla\cdots\nabla x_n = x_1\nabla\cdots\nabla x_n\nabla x_{n+1}$ for all $n > N$.

这个 ∇ 表示 widening，这个序列第 n 个元素是 $x_1\nabla\cdots\nabla x_n$，最终会趋于稳定.

Definition 3.11. An operator $\nabla\colon D\times D\to D$ is a strong widening whenever

- $x_1\nabla x_2 \geq x_1\vee x_2$ holds for all $x_1, x_2\in D$ and
- the sequence $(x_1\nabla\cdots\nabla x_n)_n$ eventually stabilises for all choices of sequence $x_1, x_2, \cdots$.

widening 弄了一个 upper bound 出来，这个 upper bound 不需要是确界.

Proposition 3.12. 任意次 widening 操作 upper bound 的性质依然保留 If ∇ is a strong widening then

$$x_1\nabla\cdots\nabla x_n \leq x_1\nabla\cdots\nabla x_n\nabla x_{n+1}$$

for all $n > 0$.

证明. 当 $n = 1$ 时

$$x_1 \leq x_1\nabla x_2.$$

这是显然地，假设对任意的 $n = k$ 有 $x_1\nabla\cdots\nabla x_k \leq x_1\nabla\cdots\nabla x_k\nabla x_{k+1}$ 成立，那么当 $n = k+1$ 时

$$\begin{aligned}(x_1\nabla\cdots\nabla x_k\nabla x_{k+1})\nabla x_{k+2} &\geq (x_1\nabla\cdots\nabla x_k\nabla x_{k+1})\vee x_{k+2}\\ &\geq x_1\nabla\cdots\nabla x_k\nabla x_{k+1}.\end{aligned}$$

所以原式在任意 $n > 0$ 时成立. □

Proposition 3.13. If D satisfies the ACC then the join operation $\vee$ is a strong widening.

证明. 这太显然了，简直 trivial，ACC 在这里保证了任意非空集合都有最大元素，那么它们的 join 肯定不会超过它，也就是 eventually stabilises. □

Definition 3.14. widening operation 的构造 Given galois connection pair (f, g) of A between B, we defined widening operation as follows

$$x\nabla y = g(f(x)\vee f(y))$$

where $x, y\in A$.

Lemma 3.15.

$$x_1\nabla\cdots\nabla x_n = g(f(x_1)\vee\cdots\vee f(x_n)).$$

证明. 用归纳法来证明，当 $n=2$ 时做为 base 是显然成立的，假设 $n=k$ 时成立，那么当 $n=k+1$ 时, 有

$$\begin{aligned}(x_1\nabla\cdots\nabla x_k)\nabla x_{k+1} &= g(f(x_1\nabla\cdots\nabla x_k)\vee f(x_{k+1}))\\ &= g(f(\underline{g(f(x_1)\vee\cdots\vee f(x_n))})\vee f(x_{n+1}))\\ &= g(f(\underline{g(f(x_1))}\vee\cdots\vee \underline{g(f(x_n))})\vee f(x_{n+1}))\\ &= g(\underline{f(g(f(x_1)))}\vee\cdots\vee \underline{f(g(f(x_n)))}\vee f(x_{n+1}))\end{aligned}$$

前面我们证明过 $f\circ g\circ f=f$, 所以最后有 $g(f(x_1)\vee\cdots\vee f(x_{k+1}))$. □

Proposition 3.16. strong widening operation Given galois connection pair (f,g) of A between B and B statifies the ACC then ∇ defined above is a strong widening.

证明. (1) 根据 galois connection reduce 出来的 semilattice homomorphism，有 $f(x\vee y)=f(x)\vee f(y)$，再根据 galois connection 的 definition，有

$$x\vee y\leq g(f(x\vee y))=x\nabla y.$$

(2) 由于 B 是满足 ACC 的，所以存在 $f(x_1)\vee\cdots\vee f(x_n)\leq f(x_m)$，根据前面的 lemma 即 $x_1\nabla\cdots\nabla x_n\leq g(f(x_m))$，那么只要 $n>m-1$ 就有 $x_1\nabla\cdots\nabla x_n=x_1\nabla\cdots\nabla x_n\nabla x_{n+1}$ 成立. 所以 $(x_1\nabla\cdots\nabla x_n)_n$ eventually stabilises. □

Proposition 3.17. 一般地 strong widening 构造

$$x\nabla y=g(f(x)\nabla' f(y)),$$

if $nabla'$ is a strong widening on B, then ∇ is a strong widening on A.