逻辑智能体

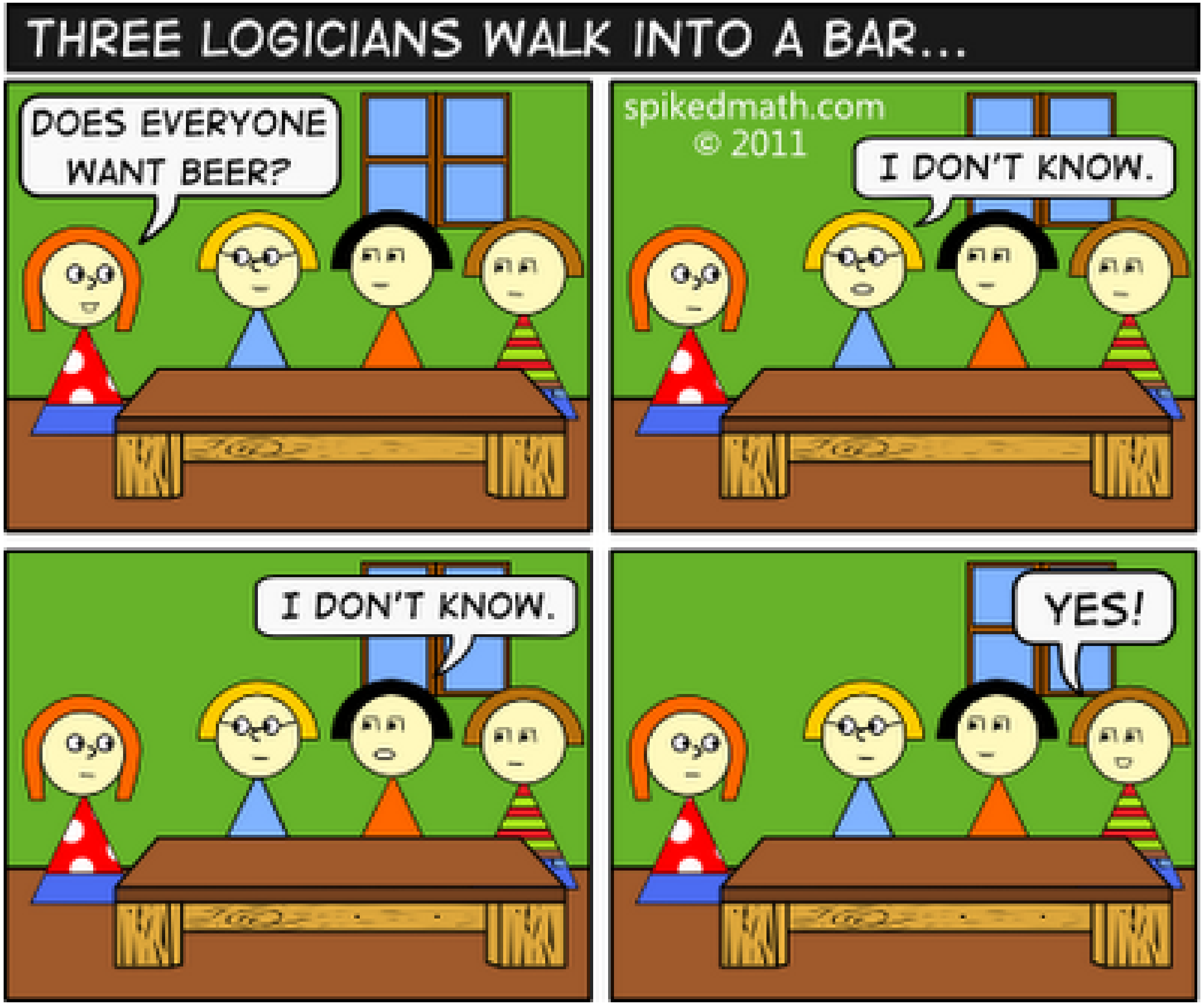
第7章

提纲

1. 基于知识的智能体
2. 基于知识智能体实例
3. 知识的逻辑表示和推理
4. 命题逻辑：一种简单的逻辑
5. 命题逻辑定理证明

一、基于知识的智能体

-----问题引出


THREE LOGICIANS WALK INTO A BAR...
DOES EVERYONE WANT BEER?
spikedmath.com © 2011
I DON'T KNOW.
I DON'T KNOW.
YES!

思考题:

最后一个人为什么回答“**yes**”？

一个故事

- 你的室友回到了寝室，他/她身上湿透了
- 你知道以下几件事：
  - 你的室友身上湿了
  - 如果你的室友身上湿了，要么是因为下雨，要么是洒水车，
  - 如果你的室友是被洒水车弄湿了，那么洒水车一定正在运行
  - 如果你的室友是被雨淋湿了，那么他/她一定没带伞
  - 伞不在伞架上
  - 如果伞不在伞架上，那么要么是你拿了伞，要么是你室友拿了伞
  - 你没有拿伞

思考题：

1.你能推断出洒水车是否在运行吗？

2.人工智能能推断出洒水车是否在运行吗？

故事的知识库

- 室友湿了
- 室友湿了=> (室友是被雨淋湿的 OR 室友是被洒水车淋湿的)
- 室友是被洒水车淋湿的=>洒水车在运行
- 室友是被雨淋湿的=>NOT(室友拿了伞)
- 伞不见了
- 伞不见了=> (你拿了伞 OR 室友拿了伞)
- NOT(你拿了伞)

基于知识的智能体

Inference Engine ← Domain-independent Algorithm

Knowledge Base ← Domain-Specific Content

- 知识库 =由形式语言组成的句子集合
- 构建智能体（或其他系统）的**声明性**方法：
  - 告诉它所需要知道的一切
- 然后他能够问自己要做什么——回答应该是基于知识库的
- 智能体可以在知识级别被查看

  例如，他们知道什么（无论是如何实现的）
- 或者在如何实现的级别
  - 例如，知识库中的数据结构和算法

一个简单的基于知识的智能体

```
function KB-AGENT(percept) returns an action
    static: KB, a knowledge base
            t, a counter, initially 0, indicating time

    TELL(KB, MAKE-PERCEPT-SENTENCE(percept, t))
    action ← ASK(KB, MAKE-ACTION-QUERY(t))
    TELL(KB, MAKE-ACTION-SENTENCE(action, t))
    t ← t + 1
    return action
```

- 智能体应该能够:
  - 表示状态、行为等等
  - 融入新的感知
  - 更新对周围世界的内部表示
  - 推断周围世界的隐藏属性
  - 推断出正确的行为

思考题：

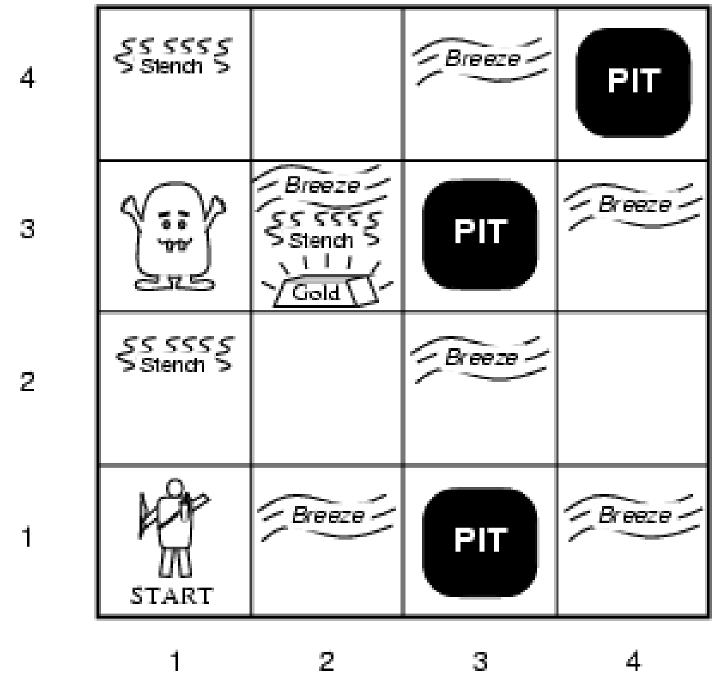
一个智能体应该具备哪些能力？

二、基于知识智能体实例

-----怪兽世界

怪兽世界的PEAS描述

- Performance measure
  - 黄金 +1000, 死亡 -1000
  - 每步 -1, 使用箭 -10
- Environment
  - 与怪兽相邻的方块是臭的
  - 无底黑洞旁边的广场有微风
  - 亮光和金子在同一个方块中
  - 如果你面对怪兽射击，会杀死它
  - 射击会消耗唯一的箭
  - 如果和金子在同一个方块中，抓取可以获得金子
  - 如果和金子在同一个方块中，放手可以扔掉金子
- Sensors: 臭味发射器，风扇，发光器，撞击器，尖叫声
- Actuators: 向左转，向右转，向前，抓取，放手，射击



怪兽世界的特点

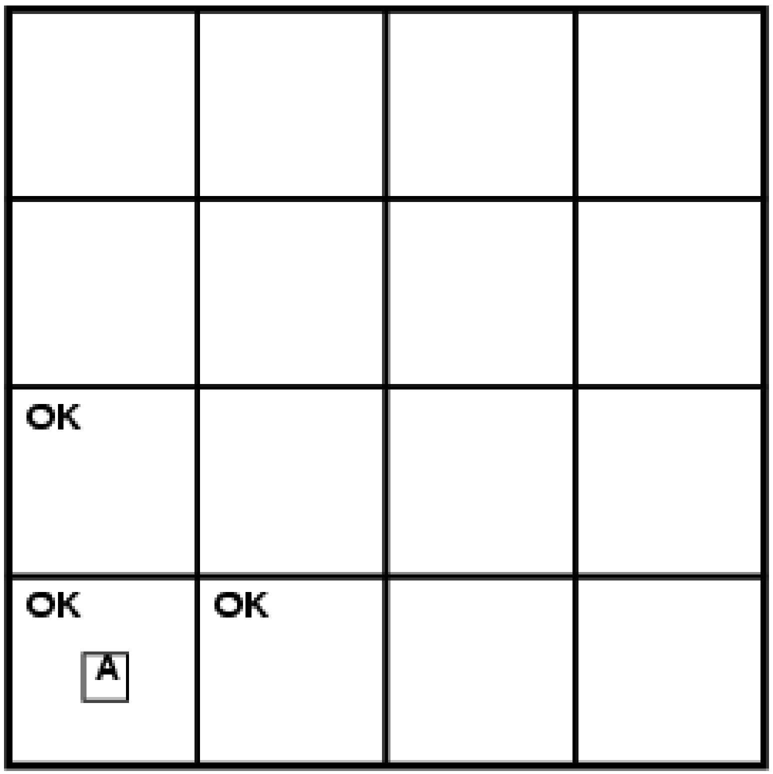
- 完全可观察性？ 完全特定的结果
- 片段式的？ 否 – 在行动上是顺序的
- 静止？ 是 – 怪兽和无底黑洞不会移动
- 离散性？ 是
- 单智能体? 是 – 怪兽本质上是一个自然体

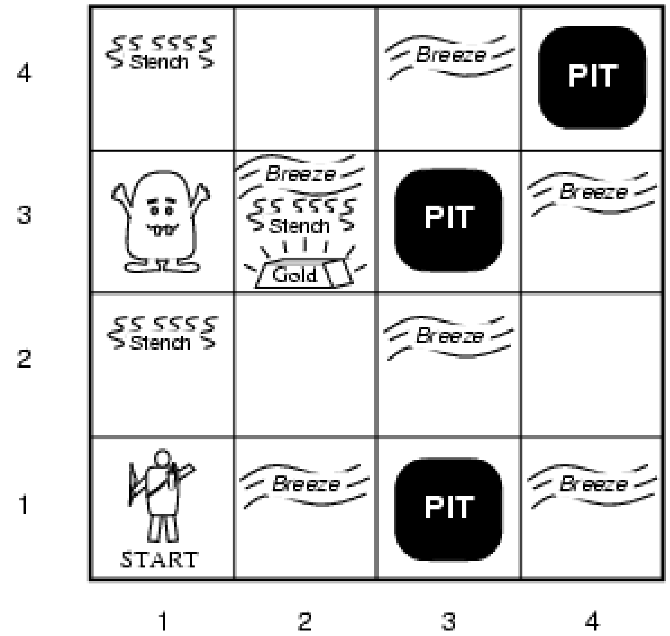
思考题：

1.智能体所在的怪兽世界环境复杂吗？

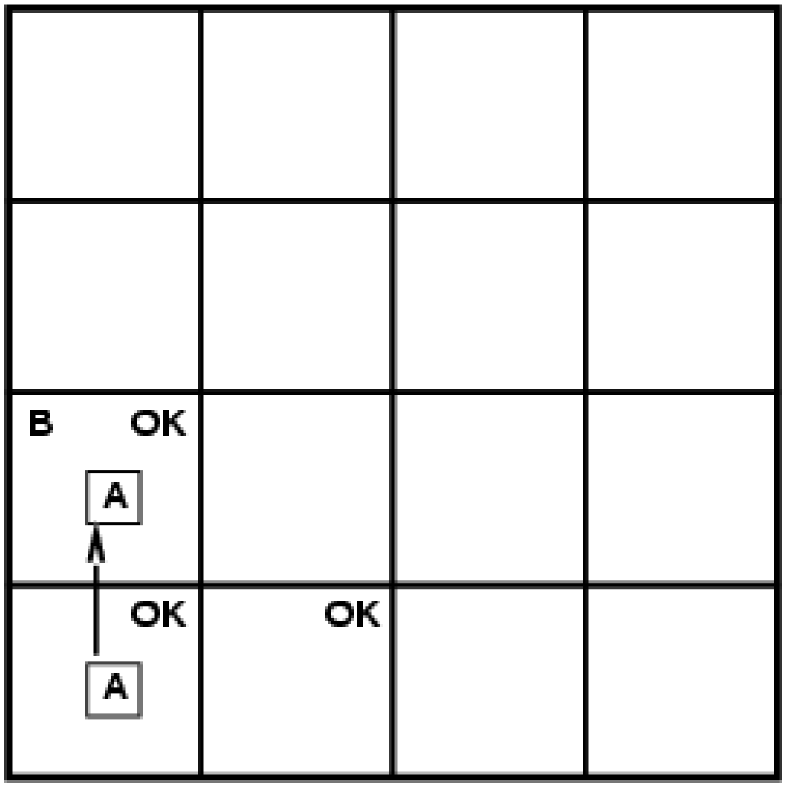
2. 智能体如何获取金子？

探索一个怪兽世界

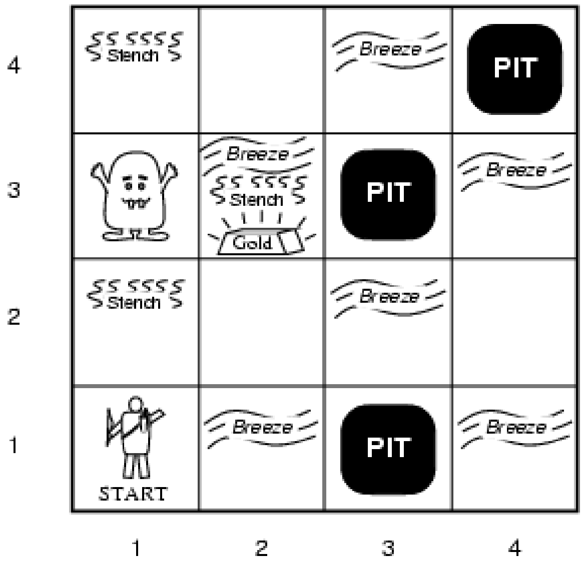

OK
OK
A
OK


Stench
Breeze
PIT
Breeze
Stench
Gold
PIT
Breeze
Stench
Breeze
START
Breeze
PIT
Breeze
1
2
3
4

探索一个怪兽世界

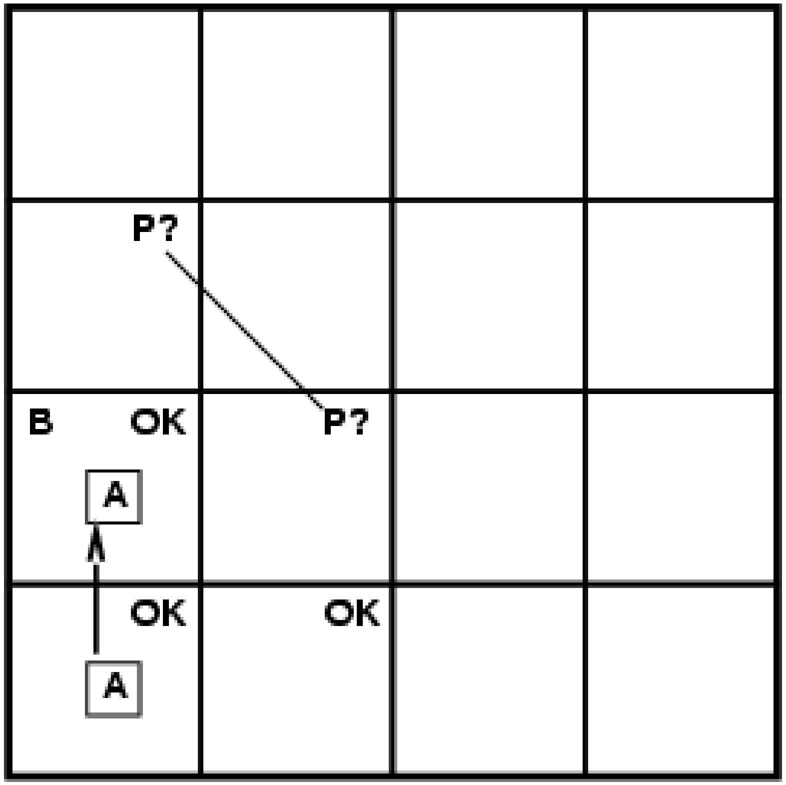


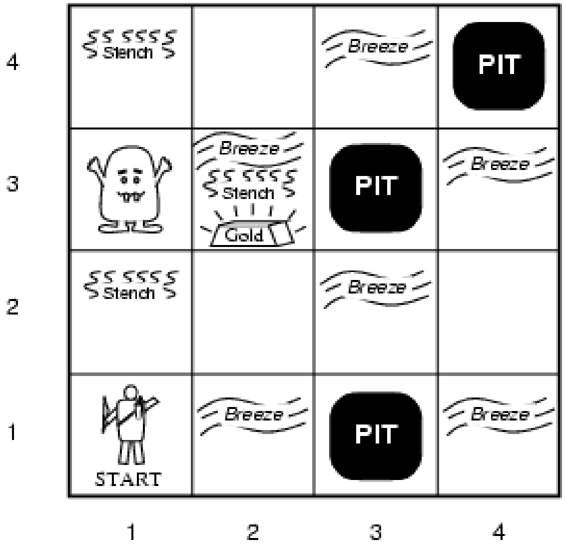
B
OK
A
OK
OK
A



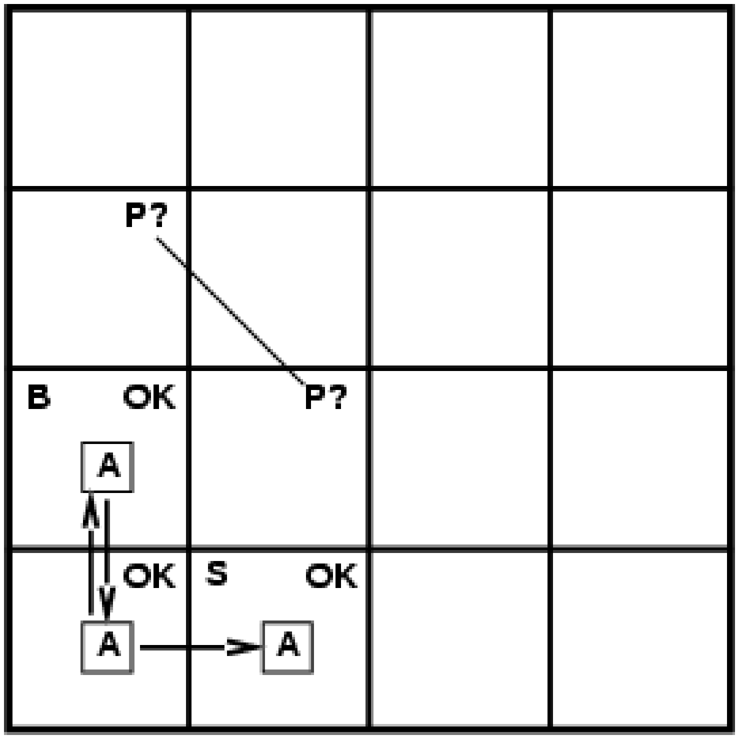
4
3
2
1
Stench
Breeze
PIT
Breeze
Stench
Gold
PIT
Breeze
Stench
Breeze
START
Breeze
PIT
Breeze
1
2
3
4

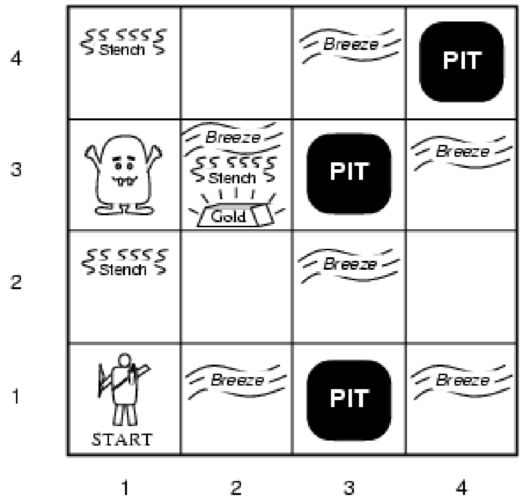
探索一个怪兽世界


P?
B
OK
P?
A
OK
OK
A

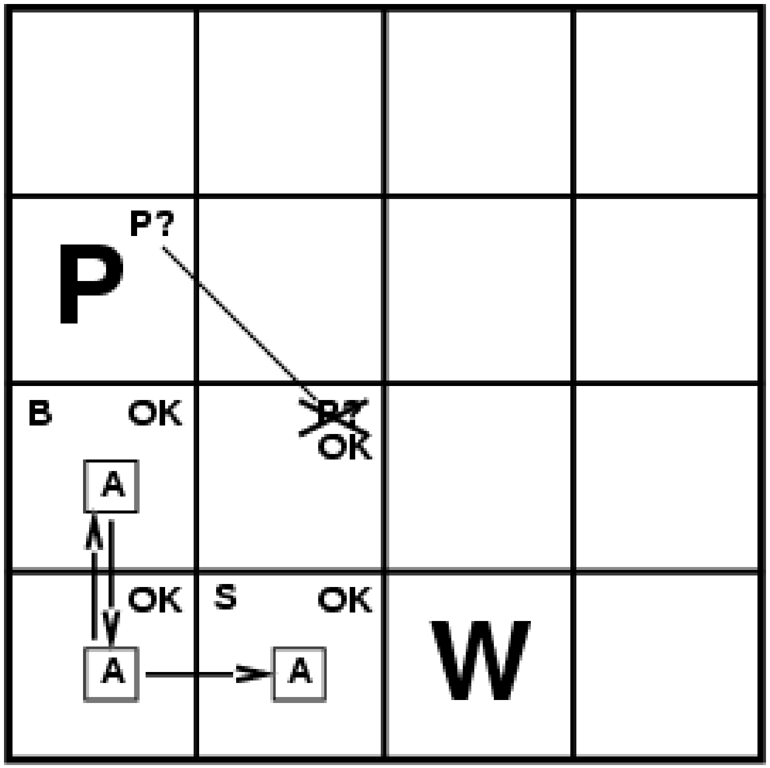

4
Stench
Breeze
PIT
3
Breeze
Stench
Gold
PIT
Breeze
2
Stench
Breeze
1
START
Breeze
PIT
Breeze
1
2
3
4

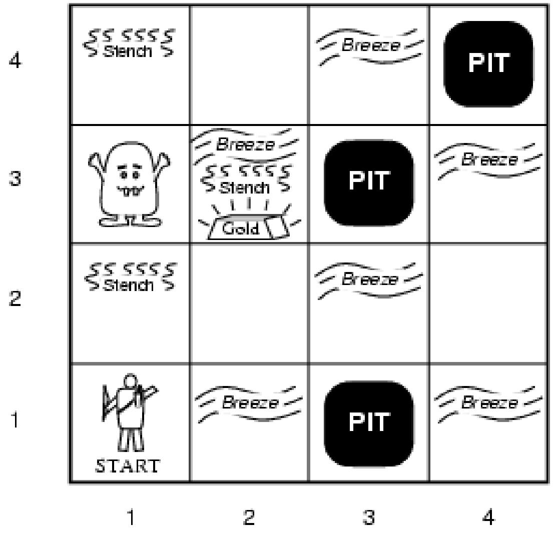
探索一个怪兽世界


P?
B
OK
P?
A
OK
S
OK
A
A

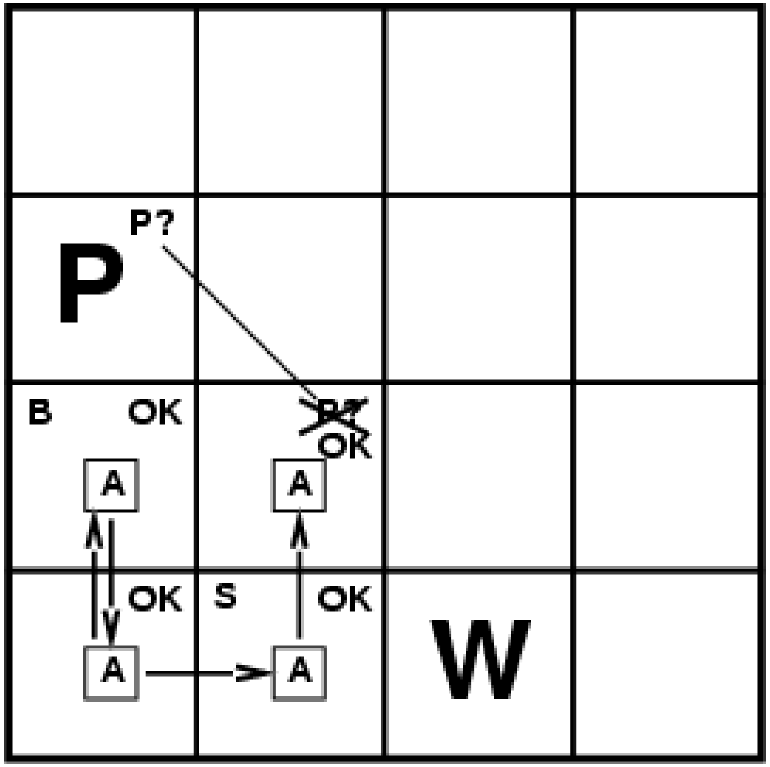

Stench
Breeze
PIT
Breeze
Stench
Gold
PIT
Breeze
Stench
Breeze
START
Breeze
PIT
Breeze
4
3
2
1
1
2
3
4

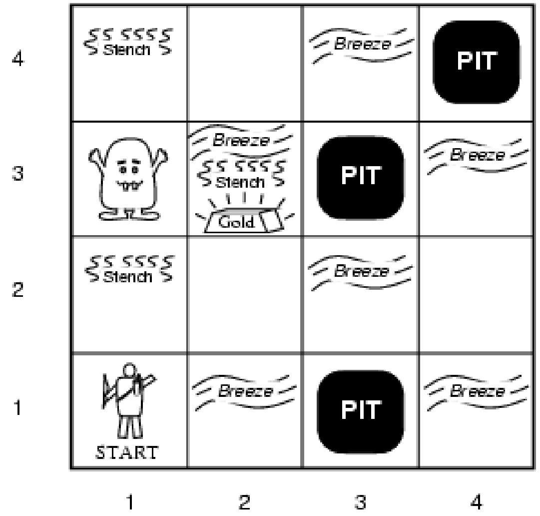
探索一个怪兽世界


P?
P
B
OK
P?
OK
A
OK
S
OK
A
A
W

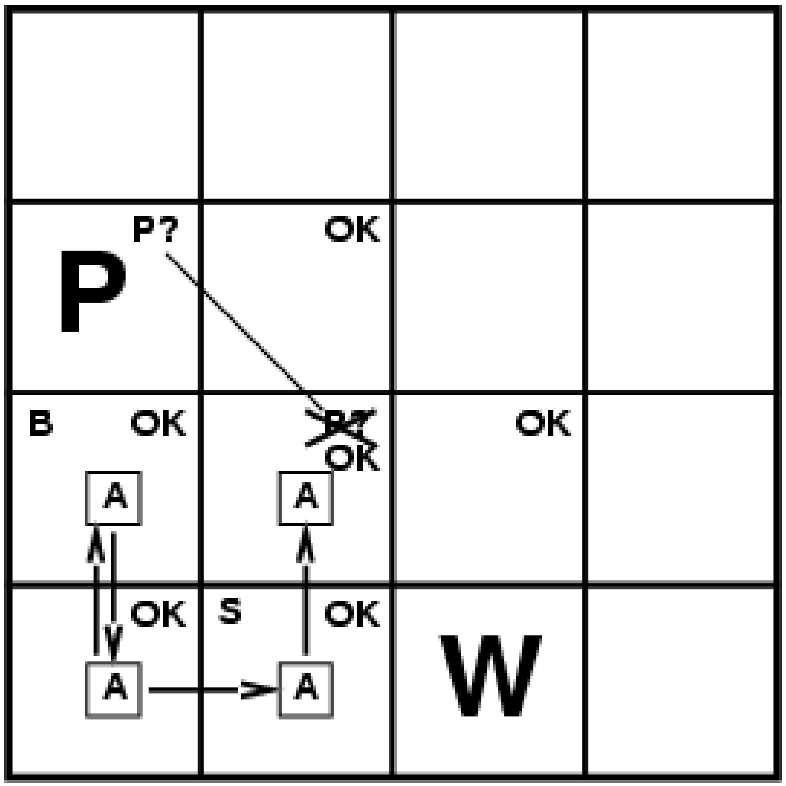

4
3
2
1
Stench
Breeze
PIT
Breeze
Stench
Gold
PIT
Breeze
Stench
Breeze
START
Breeze
PIT
Breeze
1
2
3
4

探索一个怪兽世界

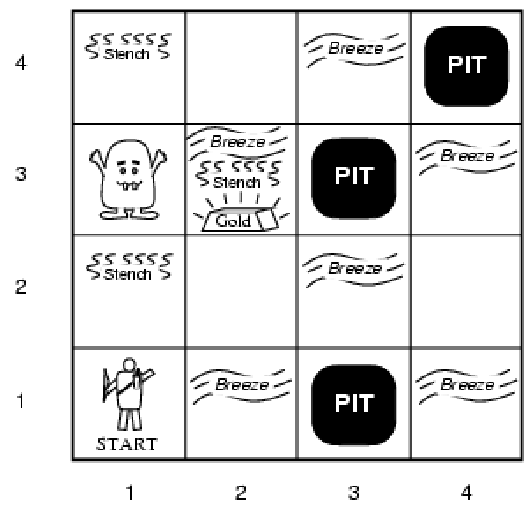

P?
P
B
OK
OK
A
A
OK
S
OK
A
A
W


Stench
Breeze
PIT
Breeze
Stench
Gold
PIT
Breeze
Stench
Breeze
START
Breeze
PIT
Breeze
1
2
3
4

探索一个怪兽世界

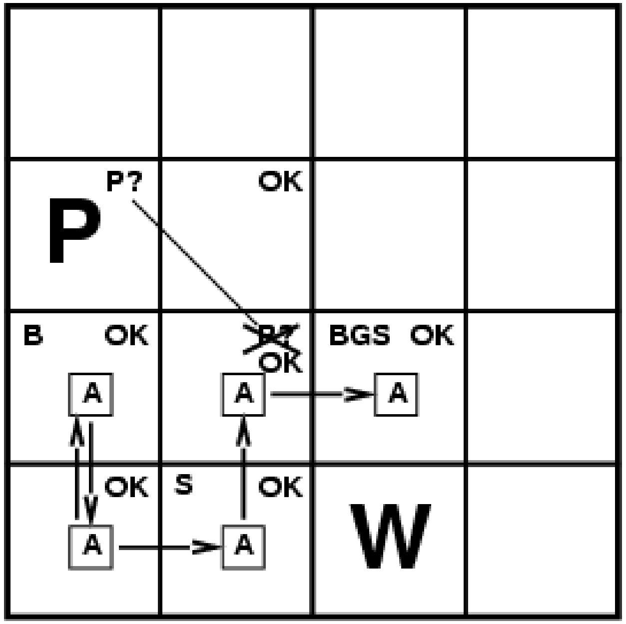


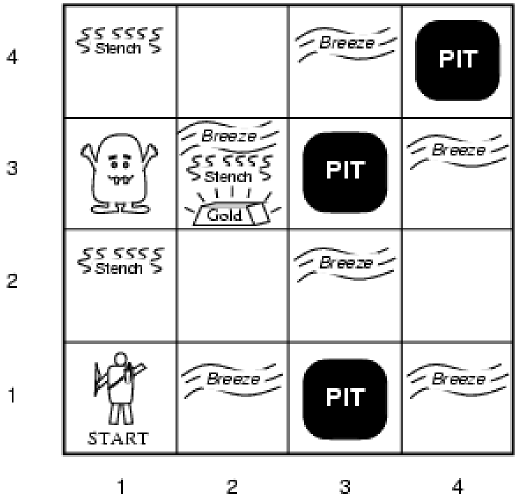
P?
P
OK
B
OK
OK
OK
A
A
OK
S
OK
A
A
W



4
Stench
Breeze
PIT
3
Breeze
Stench
Gold
PIT
Breeze
2
Stench
Breeze
1
START
Breeze
PIT
Breeze
1
2
3
4

探索一个怪兽世界


P?
OK
P
B
OK
~~P?~~
OK
BGS OK
A
A
A
OK
S
OK
A
A
W


Stench
Breeze
PIT
Breeze
Stench
Gold
PIT
Breeze
Stench
Breeze
START
Breeze
PIT
Breeze
4
3
2
1
1
2
3
4

思考题：

如何构建可以表示智能知识并推理出金子所在位置的智能体？

三、知识的逻辑表示和推理

什么是逻辑?

- **逻辑** 是一个用来处理事实的形式化系统，目的是得到正确的结论
  - *“用来区分真和假的工具” – Averroes (12世纪.)*
- **语法:** 构造知识库有效句子的规则
  - 例如, $x + 2 \geq y$ 是有效的算术语句, $\geq x2y +$ 就不是
- **语义:** 句子的“含义”, 或者逻辑语句和真实世界之间的关系
  - 具体而言,语义决定了句子的真假
  - 例如, $x + 2 \geq y$ 为真时可以取$x = 5$ 且$y = 7$

逻辑推理

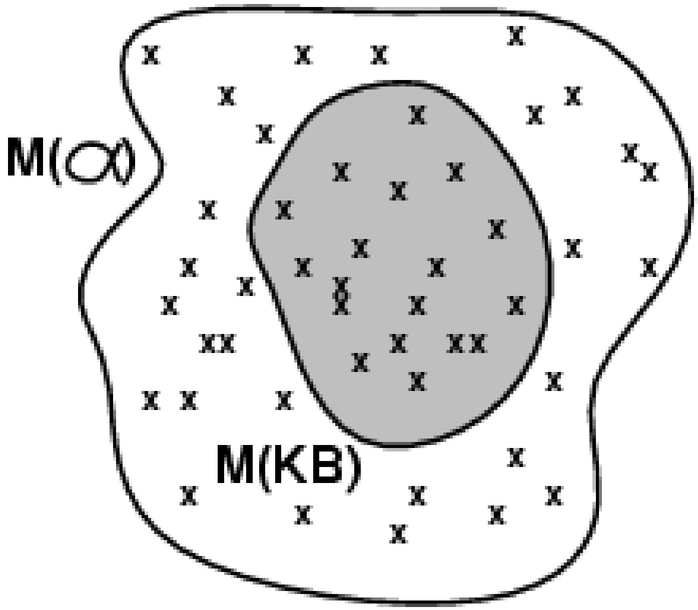
- 蕴含 意味着一样东西是从另一样东西中派生出来的:

$$KB \models \alpha$$

- 知识库KB蕴含句子α 当且仅当α 在所有KB为真的世界里为真
  - 例如，包含“巨人赢了”和“红军赢了”的知识库KB蕴含了“要么巨人赢了要么红军赢了”
  -
  - 例如, x+y = 4 蕴含 4 = x+y
  -
  - 蕴含是基于语义的句子（例如，语法）之间的关系

模型

- 逻辑学家通常根据模型来思考，这些模型是可以评估真假的形式化的结构世界
- 如果α 在m中为真，我们说*m*是句子α 的模型
- *M(α)* 是所有α的模型的集合
- 那么 KB $\models$ α iff $M(KB) \subseteq M(\alpha)$
  - 例如， *KB* = 巨人队赢了and 红队赢了；α = 巨人队赢了

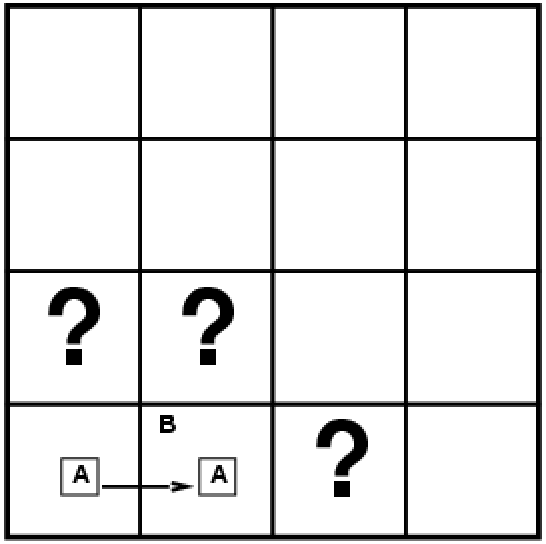


思考题:

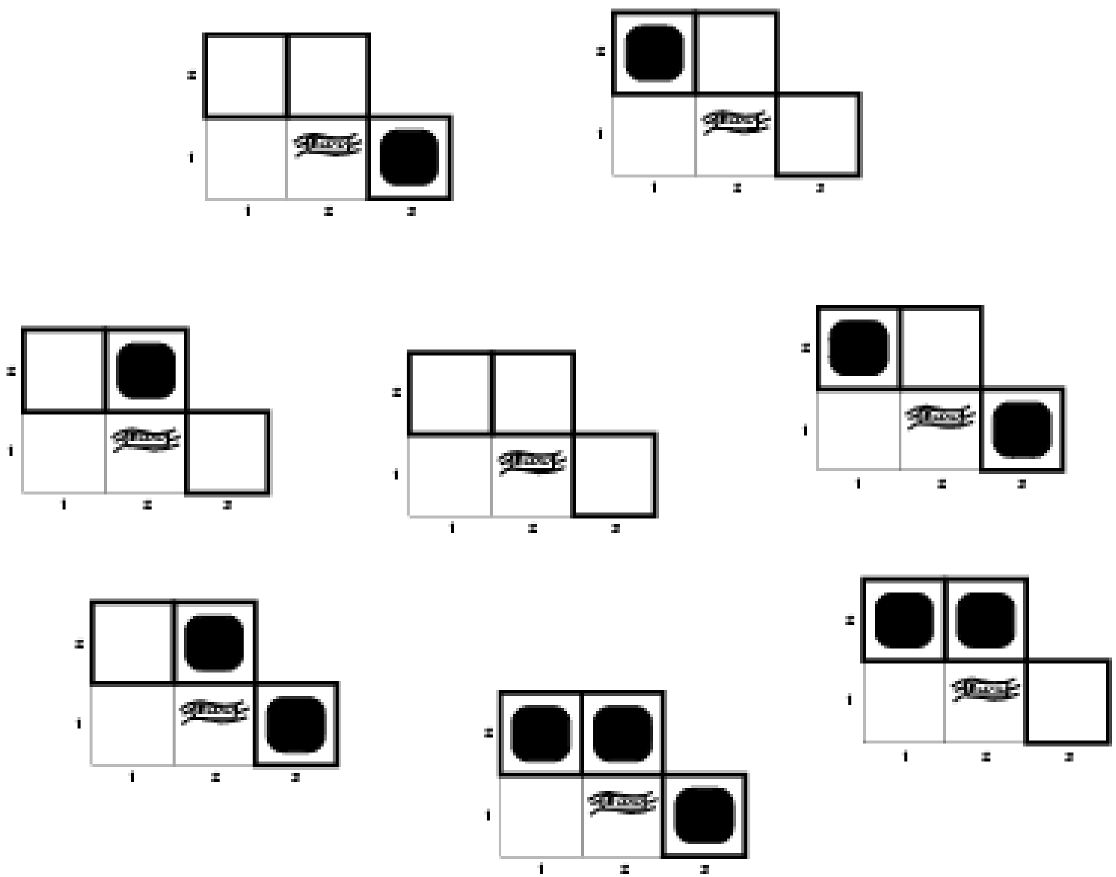
语句X=0 和语句XY=0 中KB是哪个？ α是哪个？谁蕴含了谁？

在怪兽世界中的蕴含规则

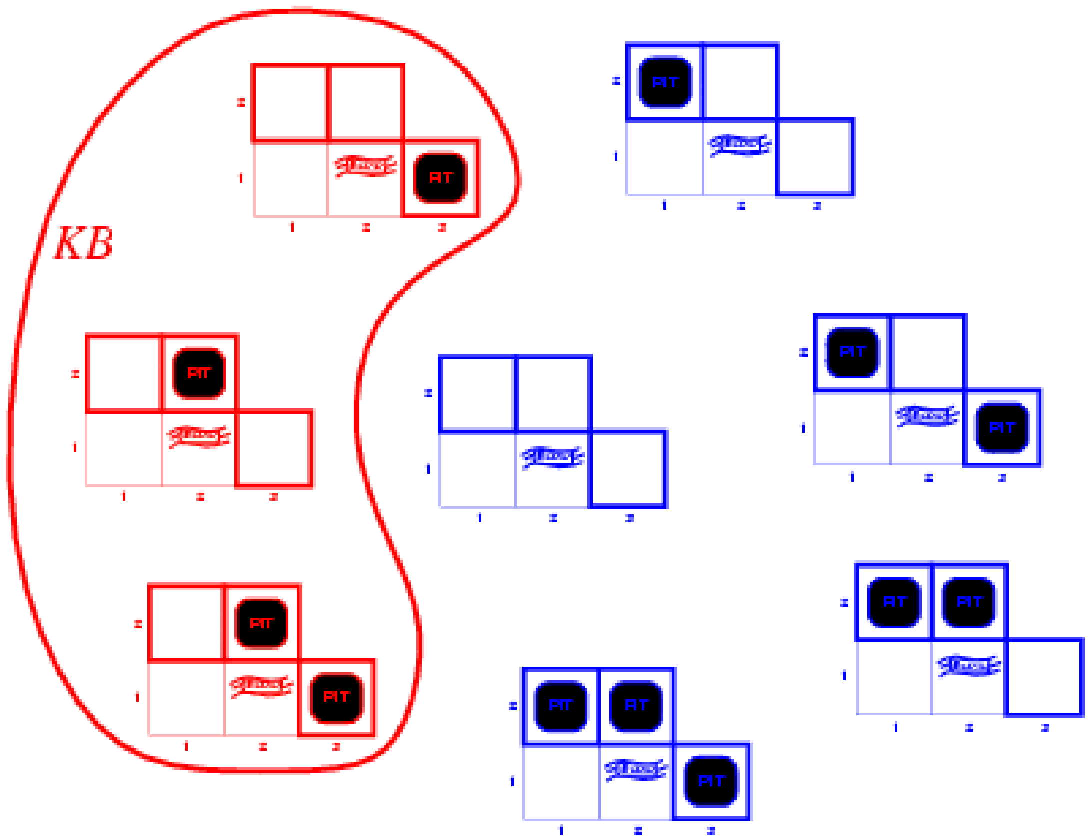
- 在检测到[1,1]中没有东西后，向右移动, [2,1]中有微风
- 假定只有无底黑洞，来考虑知识库KB可能的模型
- 布尔选择⇒ 8 个可能的模型



怪兽世界

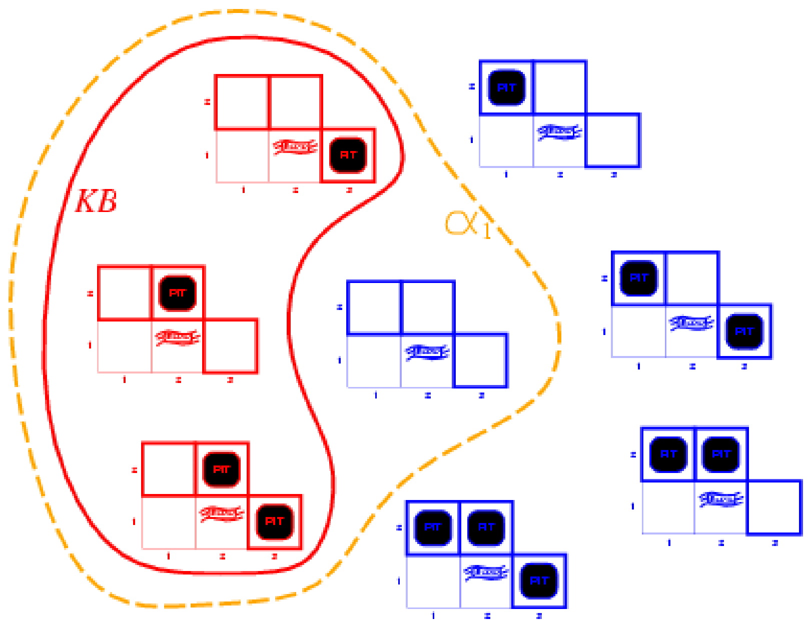


怪兽世界的模型



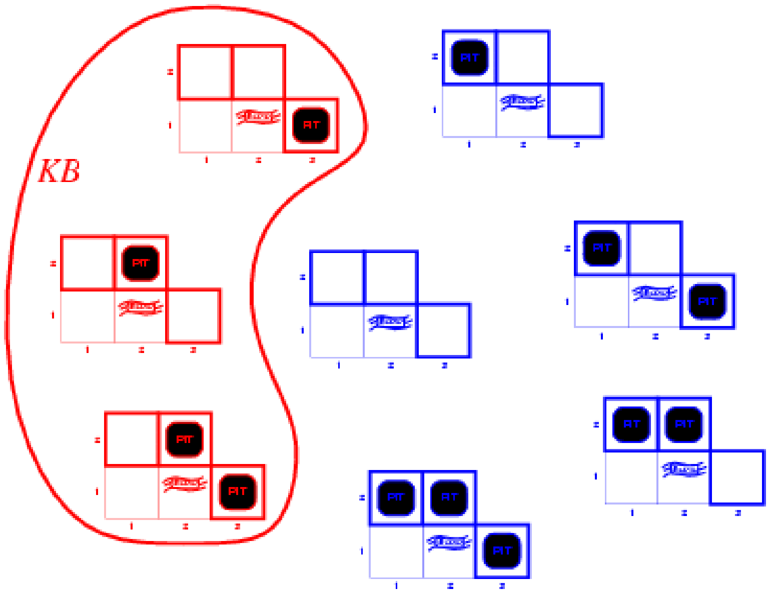
- KB = 怪兽世界的规则+ 观察

怪兽世界的模型



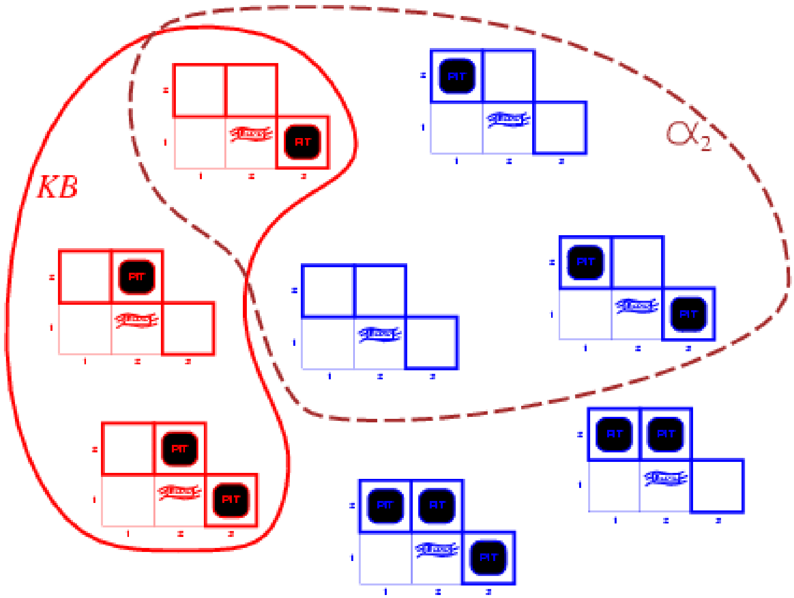
- *KB* =怪兽世界的规则 + 观察
- α_1 = “[1,2] 是安全的”, *KB* $\models \alpha_1$, 从模型验证中被证明

怪兽世界的模型



- *KB* =怪兽世界的规则+ 观察

怪兽世界的模型



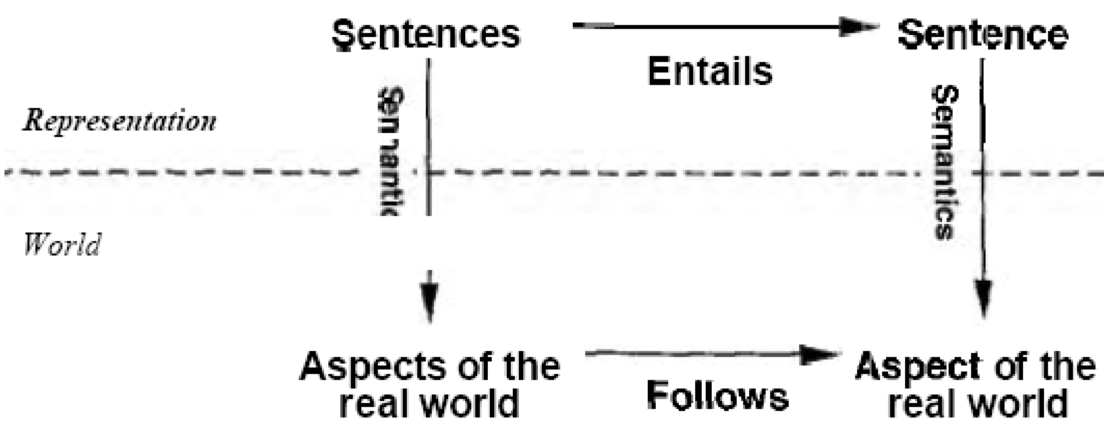
- KB = 怪兽世界的规则 + 观察
- α_2 = “[2,2] 是安全的", $KB \not\models \alpha_2$
-

推理算法的性质

- 仅导出所需句子的推理算法被称为**好的推理或真值保留**
- 如果一个推理算法可以推倒出任何它所蕴含的句子，那么它是**完备**的
- **如果在现实世界中KB为真，那么通过好的推理过程从KB派生的任何句子*Alpha*在现实世界中也为真。**

- 逻辑智能体必须解决的最后一个问题是逻辑推理过程与智能体所存在的真实环境之间的联系（如果有的话）。

- 传感器和学习



思考题：

KB = 怪兽世界的规则 + 观察

α_3 = “[1,3] 中有怪兽", $KB \models \alpha_3$

模仿上面讲课内容，把KB $\models \alpha_3$为真世界表出来

四、命题逻辑：一种简单的逻辑

命题

命题：能判断真假的陈述句。这种陈述句的判断只有两种可能：一种是 正确的判断，一种是错误的判断。称判断为正确的命题的真值（或 值）为真，称判断为错误的命题的真值（或值）为假。

因此又可称命题是具有唯一真值的陈述句或判断结果惟一的陈述句

命题的真值: 判断的结果

真值的取值: 真与假　　二者取一

真命题: 真值为真的命题

假命题: 真值为假的命题

注意：感叹句、祈使句、疑问句都不是命题

陈述句中的悖论以及判断结果不惟一确定的也不是命题

思考题

例 下列句子中那些是命题？

(1) 是无理数.

(2) $2 + 5 = 8$.

(3) $x + 5 > 3$.

(4) 你有铅笔吗？

(5) 这只兔子跑得真快呀！

(6) 请不要讲话！

(7) 我正在说谎话.

命题的分类

简单命题(原子命题)：

简单陈述句构成的命题

复合命题：

由简单命题用联结词联结而成的命题

简单命题符号化

用大写英文字母 $P, Q, R, \ldots, P_i, Q_i, R_i\ (i \geqslant 1)$ 表示简单命题，将表示命题的符号放在该命题的前面，称为命题符号化。

用“1”表示真，用“0”表示假

对简单命题而言，它的真值是确定的，因而又称为命题常项或命题常元。

例如，令

P: π是有理数，则 P 的真值为 0

Q：$2 + 5 = 7$，则 Q 的真值为 1

命题逻辑：语法

- 原子语句:
  - 一个 *命题逻辑符号* 表示一个真或假的状态
- 取非:
  - 如果 P 是一个语句,那么 $\neg P$ 也是一个语句
- 析取:
  - 如果 P 和Q 都是语句,那么 $P \vee Q$ 也是语句
- 合取:
  - 如果 P 和Q 都是语句,那么 $P \wedge Q$也是语句
- 蕴含:
  - 如果 P 和Q 都是语句,那么$P \Rightarrow Q$也是语句
- 双条件:
  - 如果 P 和Q 都是语句,那么$P \Leftrightarrow Q$也是语句

- $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ 被称作为*逻辑连接符号*

命题逻辑：语法

$$
\begin{aligned}
Sentence &\rightarrow AtomicSentence \mid ComplexSentence \\
AtomicSentence &\rightarrow \textbf{True} \mid \textbf{False} \mid \text{Symbol} \\
Symbol &\rightarrow P \mid \mathbf{Q} \mid \mathbf{R} \mid \ldots \\
ComplexSentence &\rightarrow \neg\, Sentence \\
&\mid\ (\, Sentence \land Sentence\,) \\
&\mid\ (\, Sentence \lor Sentence\,) \\
&\mid\ (\, Sentence \Rightarrow Sentence\,) \\
&\mid\ (\, Sentence \Leftrightarrow Sentence\,)
\end{aligned}
$$

Figure 7.7 A BNF (Backus–Naur Form) grammar of sentences in propositional logic.

$\neg, \land, \lor, \Rightarrow, \Leftrightarrow$ 从最高到最低

命题逻辑：语义

每个模型为每个命题符号指定真/假

例如.	$P_{1,2}$	$P_{2,2}$	$P_{3,1}$
	假	真	假

使用这些符号，可以自动枚举8种可能的模型。

用于评估模型m的真值的规则如下：

$\neg S$ 为真	iff S 为假		
$S_1 \wedge S_2$ 为真	iff S_1 为真	and	S_2 为真
$S_1 \vee S_2$ 为真	iff S_1 为真	or	S_2 为真
$S_1 \Rightarrow S_2$ 为真	iff S_1 为假	or	S_2 为真
$S_1 \Rightarrow S_2$ 为假	iff S_1 为真	and	S_2 为假
$S_1 \Leftrightarrow S_2$ 为真	iff $S_1 \Rightarrow S_2$ 为真	and	$S_2 \Rightarrow S_1$ 为真

用简单的推理过程评价任意一个句子, 例如,

$$\neg P_{1,2} \wedge (P_{2,2} \vee P_{3,1}) = \mathit{true} \wedge (\mathit{true} \vee \mathit{false}) = \mathit{true} \wedge \mathit{true} = \mathit{true}$$

连接词的真值表

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
$false$	$false$	$true$	$false$	$false$	$true$	$true$
$false$	$true$	$true$	$false$	$true$	$true$	$false$
$true$	$false$	$false$	$false$	$true$	$false$	$false$
$true$	$true$	$false$	$true$	$true$	$true$	$true$

思考题:

求下列复合命题的真值

(1) $2+2=4$ 当且仅当 $3+3=6$.

(2) $2+2=4$ 当且仅当 3 是偶数.

(3) $2+2=4$ 当且仅当 太阳从东方升起.

(4) 函数 $f(x)$ 在x_0 可导的充要条件是它在 x_0 连续.

构建怪兽世界的知识库

如果[i, j]中有无底黑洞，令$P_{i,j}$ 为真.

如果 [i, j]中有风，令$B_{i,j}$ 为真.

$\neg P_{1,1}$

$\neg B_{1,1}$

$B_{2,1}$

- “无底黑洞导致相邻的方块中有风”

$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

$B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

推理的真值表

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	KB	α_1
$false$	$false$	$false$	$false$	$false$	$false$	$false$	$false$	$true$
$false$	$false$	$false$	$false$	$false$	$false$	$true$	$false$	$true$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$false$	$true$	$false$	$false$	$false$	$false$	$false$	$false$	$true$
$false$	$true$	$false$	$false$	$false$	$false$	$true$	$\underline{true}$	$\underline{true}$
$false$	$true$	$false$	$false$	$false$	$true$	$false$	$\underline{true}$	$\underline{true}$
$false$	$true$	$false$	$false$	$false$	$true$	$true$	$\underline{true}$	$\underline{true}$
$false$	$true$	$false$	$false$	$true$	$false$	$false$	$false$	$true$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$true$	$true$	$true$	$true$	$true$	$true$	$true$	$false$	$false$

通过枚举推理

- 对所有模型来说，深度优先的枚举是好的且完备的

```
function TT-ENTAILS?(KB, α) returns true or false
  symbols ← a list of the proposition symbols in KB and α
  return TT-CHECK-ALL(KB, α, symbols, [])

function TT-CHECK-ALL(KB, α, symbols, model) returns true or false
  if EMPTY?(symbols) then
     if PL-TRUE?(KB, model) then return PL-TRUE?(α, model)
     else return true
  else do
     P ← FIRST(symbols); rest ← REST(symbols)
     return TT-CHECK-ALL(KB, α, rest, EXTEND(P, true, model) and
              TT-CHECK-ALL(KB, α, rest, EXTEND(P, false, model)
```

- 对于n个符号来说, 时间复杂度为$O(2^n)$, 空间复杂度为$O(n)$

思考题:

枚举推理方法的复杂度是指数增长的，是NP问题，同学们思考下有无其他更好的方法?

五、命题逻辑定理证明

逻辑等价性

- 两个语句是逻辑等价的当且仅当下列谓词逻辑在相同模型中为真: **α ≡ ß iff α ╞ β and β ╞ α**

$(\alpha \wedge \beta)$	$\equiv$	$(\beta \wedge \alpha)$	commutativity of $\wedge$
$(\alpha \vee \beta)$	$\equiv$	$(\beta \vee \alpha)$	commutativity of $\vee$
$((\alpha \wedge \beta) \wedge \gamma)$	$\equiv$	$(\alpha \wedge (\beta \wedge \gamma))$	associativity of $\wedge$
$((\alpha \vee \beta) \vee \gamma)$	$\equiv$	$(\alpha \vee (\beta \vee \gamma))$	associativity of $\vee$
$\neg(\neg\alpha)$	$\equiv$	α	double-negation elimination
$(\alpha \Rightarrow \beta)$	$\equiv$	$(\neg\beta \Rightarrow \neg\alpha)$	contraposition
$(\alpha \Rightarrow \beta)$	$\equiv$	$(\neg\alpha \vee \beta)$	implication elimination
$(\alpha \Leftrightarrow \beta)$	$\equiv$	$((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$	biconditional elimination
$\neg(\alpha \wedge \beta)$	$\equiv$	$(\neg\alpha \vee \neg\beta)$	de Morgan
$\neg(\alpha \vee \beta)$	$\equiv$	$(\neg\alpha \wedge \neg\beta)$	de Morgan
$(\alpha \wedge (\beta \vee \gamma))$	$\equiv$	$((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$	distributivity of $\wedge$ over $\vee$
$(\alpha \vee (\beta \wedge \gamma))$	$\equiv$	$((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$	distributivity of $\vee$ over $\wedge$

有效性和可满足性

如果一个语句在所有的模型中为真，那么这个语句是有效的

例如, $True, A \lor \neg A, \quad A \Rightarrow A, \quad (A \land (A \Rightarrow B)) \Rightarrow B$

有效性是通过推论定理与推理相关联的：

$KB \models \alpha$ if and only if ($KB \Rightarrow \alpha$) is valid

如果一个语句在一些模型中为真，那么这个语句是可满足的

例如, $A \lor B, \quad C$

如果一个语句在任何模型中都不为真，那么这个语句是不可满足的

例如，$A \land \neg A$

可满足性通过以下规则与推理相关联：

$KB \models \alpha$ if and only if ($KB \land \neg \alpha$) is unsatisfiable

$\alpha \models \beta$ if and only if the sentence ($\alpha \land \neg\beta$) is unsatisfiable

证明方法

- 证明方法被分为两类（粗略的）：
  - 推理规则的应用
    - 从旧的语句中合法的（真值保留的）生成新的语句
    - 证明 = 可以将推理规则用作是标准搜索算法中的运算符的一系列推理规则应用程序
    - 通常需要将句子转换成形式语言的形式
  - 模型检查
    - 真值表枚举（始终是*n*的指数次）
    - 改进的后向追踪，例如，**Davis--Putnam-Logemann-Loveland (DPLL)**
    - 在模型空间中的启发式搜索（真值保留的但是不完备的）

      例如, **min-conflicts-like hill-climbing** 算法

推理规则

- **Modus Ponens:**

$$\frac{\alpha \Rightarrow \beta, \qquad \alpha}{\beta}$$

- **And-Elimination:**

$$\frac{\alpha \wedge \beta}{\alpha}$$

- **其他规则:**

$$\frac{\alpha \Leftrightarrow \beta}{(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)} \quad \text{and} \quad \frac{(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)}{\alpha \Leftrightarrow \beta}$$

前面的推导是推理规则的一系列应用，称为**证明**。

找到证明就像找到**搜索问题**的解决方案一样。

搜索证明是**枚举模型**的替代方法。

$$R_6: \quad (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$$

Then we apply And-Elimination to R_6 to obtain

$$R_7: \quad ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}).$$

Logical equivalence for contrapositives gives

$$R_8: \quad (\neg B_{1,1} \Rightarrow \neg(P_{1,2} \vee P_{2,1})).$$

Now we can apply Modus Ponens with R_8 and the percept R_4 (i.e., $\neg B_{1,1}$), to obtain

$$R_9: \quad \neg(P_{1,2} \vee P_{2,1}).$$

Finally, we apply De Morgan's rule, giving the conclusion

$$R_{10}: \quad \neg P_{1,2} \wedge \neg P_{2,1}.$$

That is, neither [1,2] nor [2,1] contains a pit.

思考题:

用上面讲的推理规则判断下面推理是否正确

若明天是星期一或星期三，我就有课．若有课，今天必备课．我今天下午没备课．所以，明天不是星期一和星期三.

归结

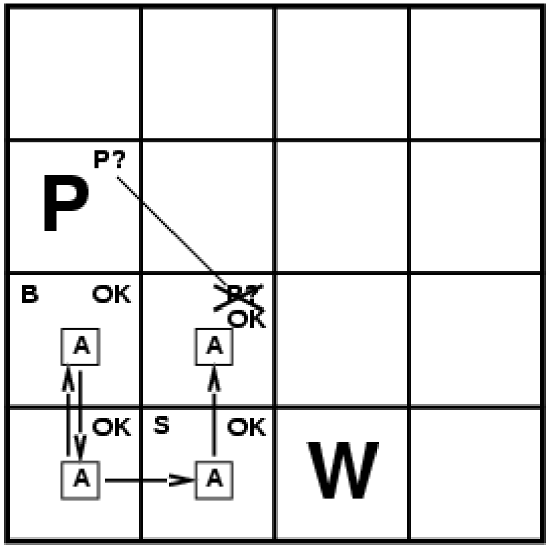
- 归结 推理规则 (for CNF):

$$\frac{l_1 \vee \ldots \vee l_k, \qquad\qquad m_1 \vee \ldots \vee m_n}{l_1 \vee \ldots \vee l_{i-1} \vee l_{i+1} \vee \ldots \vee l_k \vee m_1 \vee \ldots \vee m_{j-1} \vee m_{j+1} \vee \ldots \vee m_n}$$

where l_i and m_j are complementary literals.

E.g., $$\frac{P_{1,3} \vee P_{2,2}, \qquad\qquad \neg P_{2,2}}{P_{1,3}}$$

- 对于命题逻辑来说，归结是真值保留且完备的。



转换为CNF

联合范式Conjunctive Normal Form (CNF)

析取范式的合取范式

例如, $(A \lor \neg B) \land (B \lor \neg C \lor \neg D)$

$B_{1,1} \Leftrightarrow (P_{1,2} \lor P_{2,1})$

1. 消除⇔, 用$(\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)$代替$\alpha \Leftrightarrow \beta$.

$(B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})) \land ((P_{1,2} \lor P_{2,1}) \Rightarrow B_{1,1})$

2. 消除⇒,用$\neg\alpha \lor \beta$代替$\alpha \Rightarrow \beta$.

$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg(P_{1,2} \lor P_{2,1}) \lor B_{1,1})$

3.用de Morgan‘s 规则和双重否定将¬ 移进去:

$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land ((\neg P_{1,2} \land \neg P_{2,1}) \lor B_{1,1})$

4. 应用分配律(∧ 和∨) 和结合律:

$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1})$

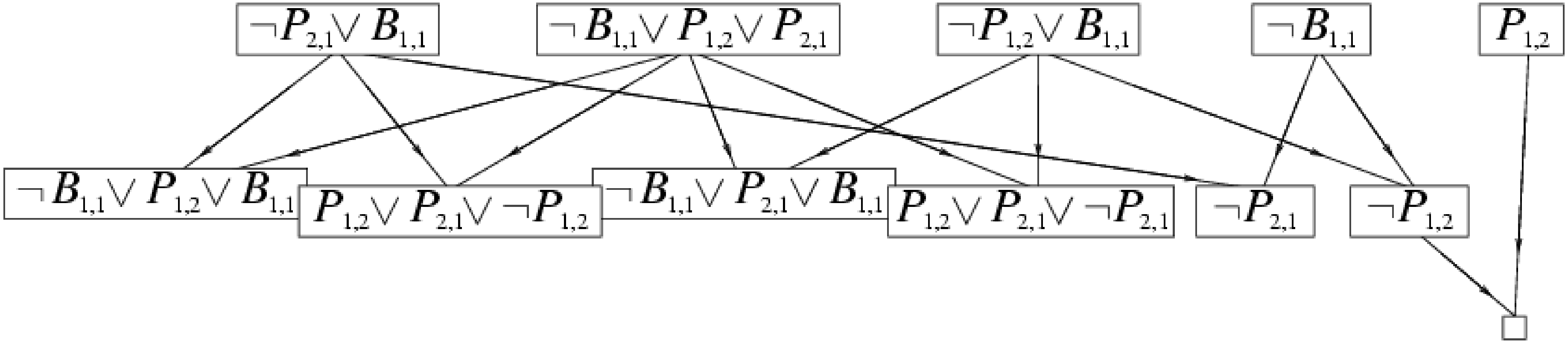
归结算法

- 用反证法证明, 例如, 假设$KB \wedge \neg\alpha$ 是不可满足的

```
function PL-RESOLUTION(KB, α) returns true or false
    clauses ← the set of clauses in the CNF representation of KB ∧ ¬α
    new ← { }
    loop do
        for each Ci, Cj in clauses do
            resolvents ← PL-RESOLVE(Ci, Cj)
            if resolvents contains the empty clause then return true
            new ← new ∪ resolvents
        if new ⊆ clauses then return false
        clauses ← clauses ∪ new
```

归结的例子

- $KB = (B_{1,1} \Leftrightarrow (P_{1,2} \lor P_{2,1})) \land \neg B_{1,1}$
- $\neg P_{1,2}$



If a set of clauses is unsatisfiable, then the resolution closure of those clauses contains the empty clause.

思考题：

用归结原理证明：

已知： $\mathrm{KB}=p\to(q\vee r)\wedge s\to\neg q\wedge p\wedge s$

$\alpha=r\vee t$

证明： $\mathrm{KB}\models\alpha$

前向链接和反向链接

- Horn 形式(严格的)
  - KB = 连接 Horn 子句

- Horn子句: 析取式中最多有一个子句为正
  - 例如, $\neg L_{1,1} \vee \neg \text{Breeze} \vee B_{1,1}$, $L_{1,1} \wedge \text{Breeze} \Rightarrow B_{1,1}$
  - 命题符号, (符号的合取) ⇒ 符号
  - 例如, $C \wedge (B \Rightarrow A) \wedge (C \wedge D \Rightarrow B)$,

- Modus Ponens (Horn形式): 对Horn KBs是完备的

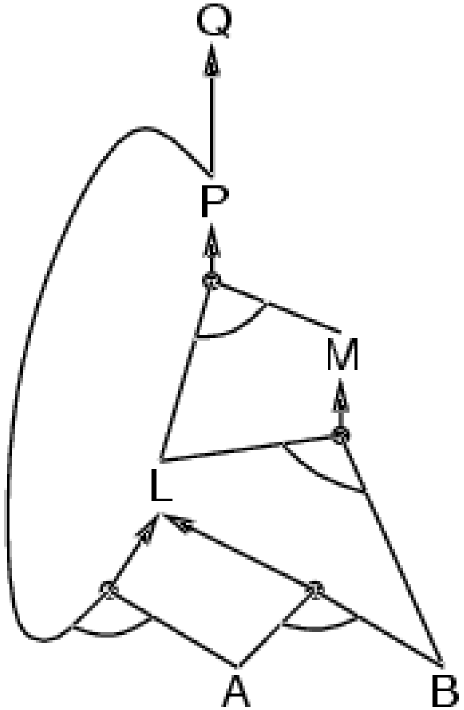
$$\frac{\alpha_1, \ldots, \alpha_n, \qquad \alpha_1 \wedge \ldots \wedge \alpha_n \Rightarrow \beta}{\beta}$$

- 可以使用前向链接 or 后向链接.
- 这些算法都非常的自然，并且可以在线性时间内运行

前向连接

- 想法:触发任何原型在*KB*中被满足的规则
  - 在*KB*中增加结论，直到需要查询的信息被找到

$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$



前向链接算法

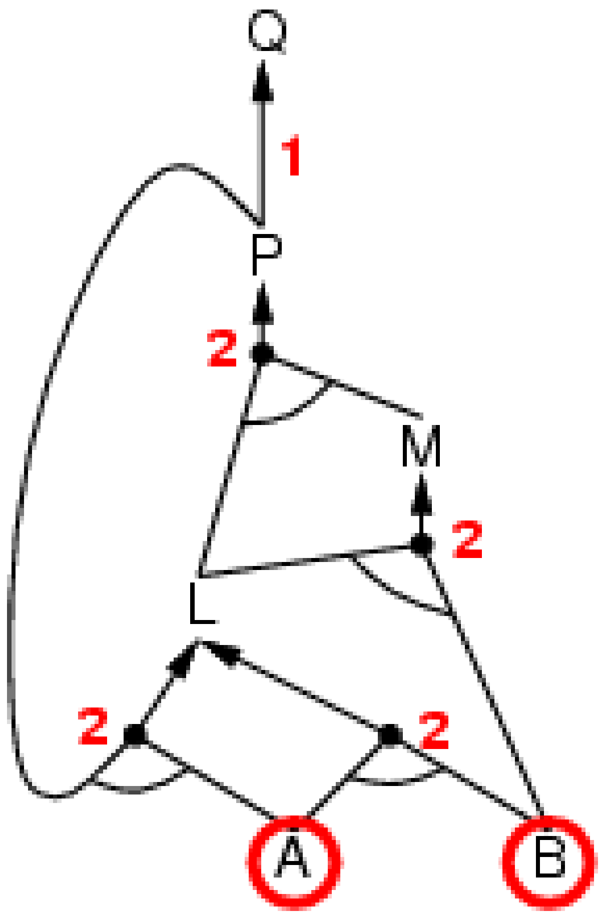
```
function PL-FC-ENTAILS?(KB, q) returns true or false
  local variables: count, a table, indexed by clause, initially the number of premises
                   inferred, a table, indexed by symbol, each entry initially false
                   agenda, a list of symbols, initially the symbols known to be true

  while agenda is not empty do
      p ← POP(agenda)
      unless inferred[p] do
          inferred[p] ← true
          for each Horn clause c in whose premise p appears do
              decrement count[c]
              if count[c] = 0 then do
                  if HEAD[c] = q then return true
                  PUSH(HEAD[c], agenda)
  return false
```

- 前向链接对于Horn KB来说是真值保留且完备的

前向链接的例子

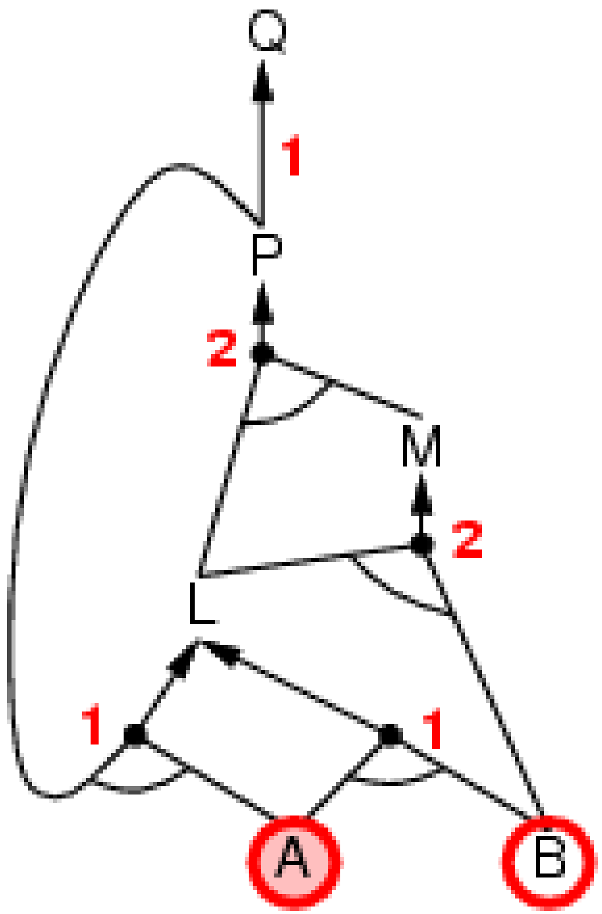
$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$



A
B

前向链接的例子

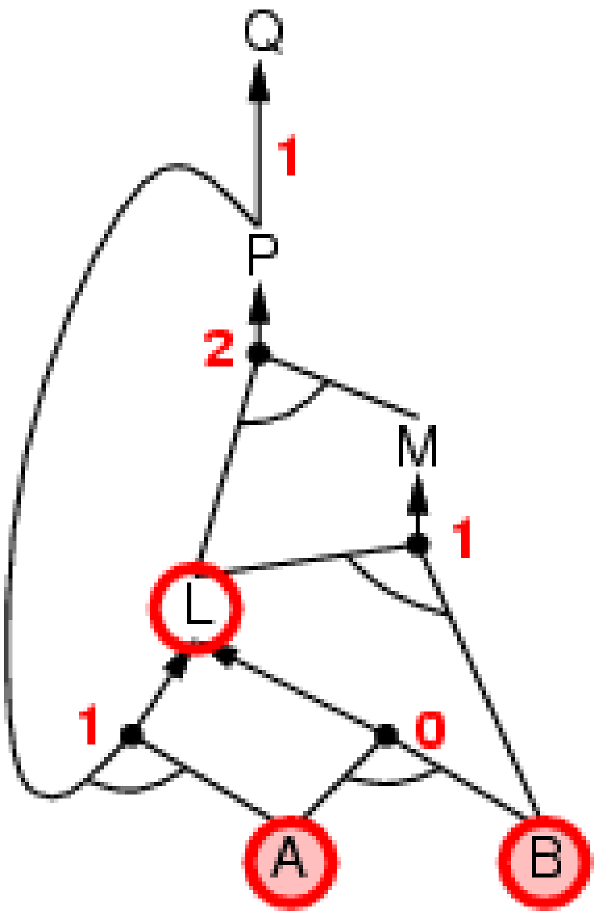
$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$



A ∧ ?P => ?L

前向链接的例子

$P \Rightarrow Q$
$L \wedge M \Rightarrow P$
$B \wedge L \Rightarrow M$
$A \wedge P \Rightarrow L$
$A \wedge B \Rightarrow L$
A
B



A ∧ B => L

前向链接的例子

$P \Rightarrow Q$

$L \wedge M \Rightarrow P$

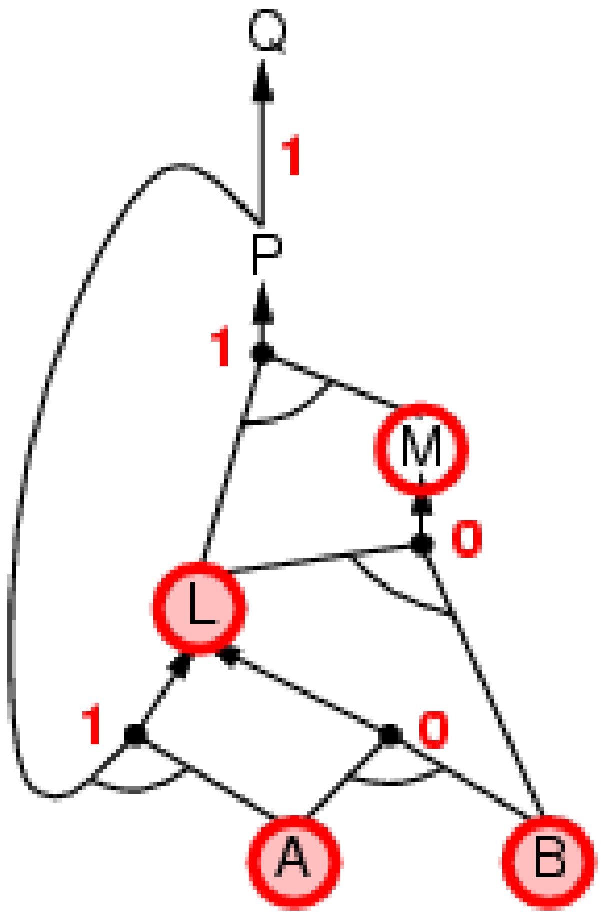
$B \wedge L \Rightarrow M$

$A \wedge P \Rightarrow L$

$A \wedge B \Rightarrow L$

A

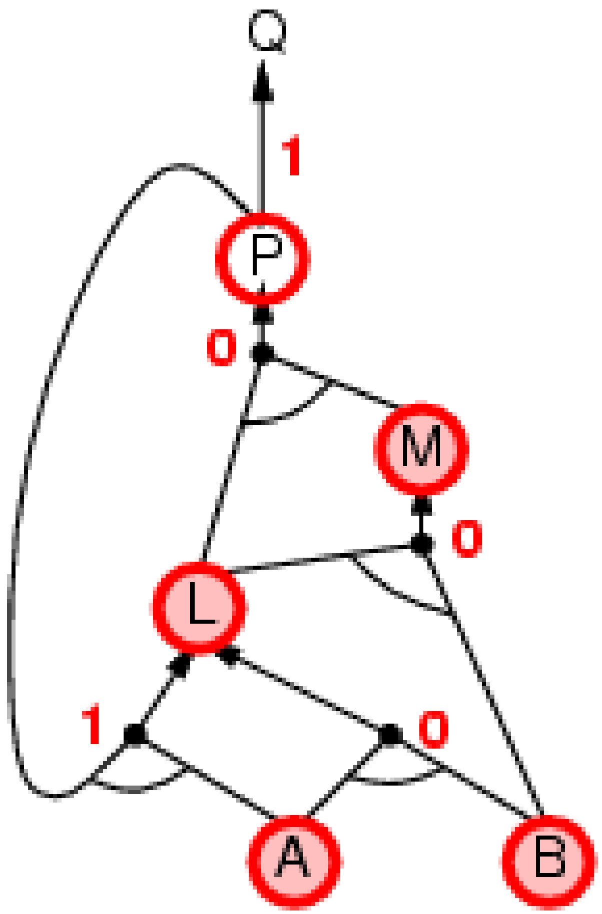
B



L ∧ B => M

前向链接的例子

$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$



M ∧ L => P

前向链接的例子

$P \Rightarrow Q$

$L \wedge M \Rightarrow P$

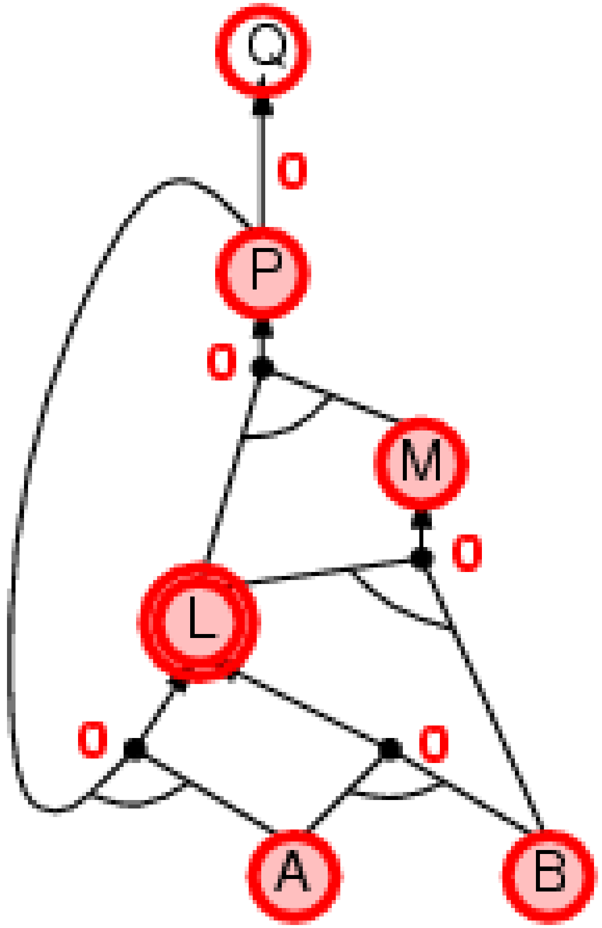
$B \wedge L \Rightarrow M$

$A \wedge P \Rightarrow L$

$A \wedge B \Rightarrow L$

A

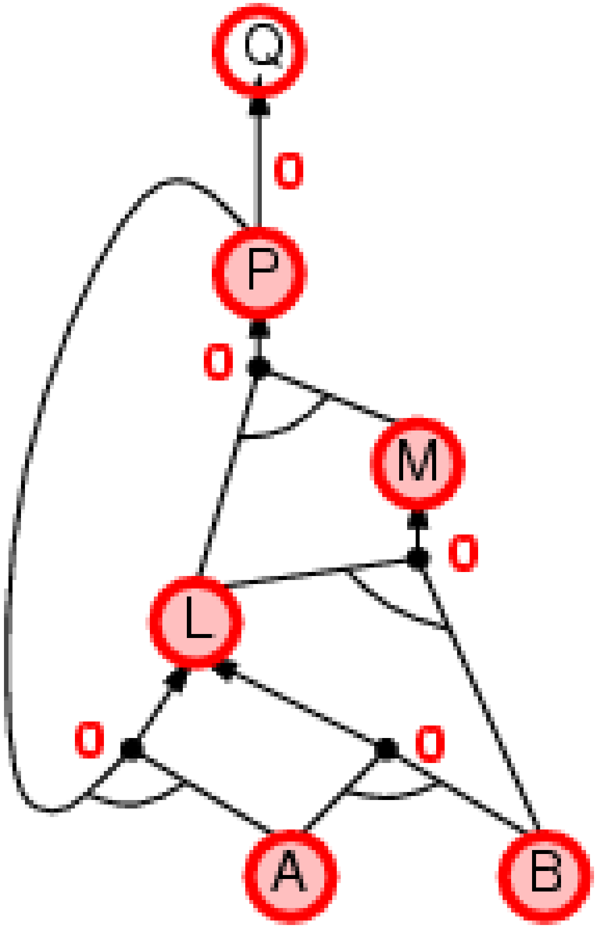
B



P => Q

前向链接的例子

$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$



A ∧ P => L

前向链接的例子

$P \Rightarrow Q$

$L \wedge M \Rightarrow P$

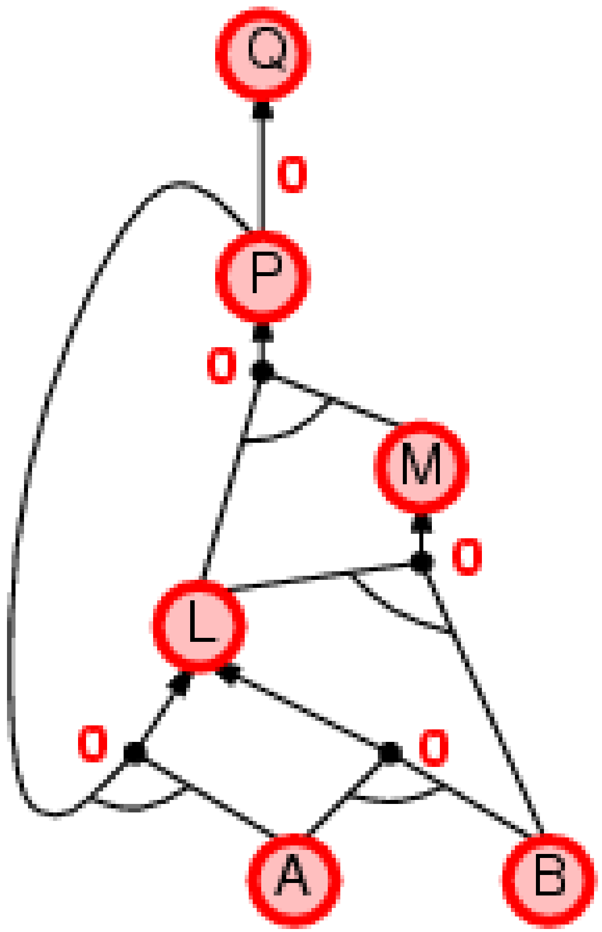
$B \wedge L \Rightarrow M$

$A \wedge P \Rightarrow L$

$A \wedge B \Rightarrow L$

A

B



P => Q

完备性的证明

FC派生出KB所蕴含的每个原子语句

1. FC到达一个没有新的原子子句可派生的不动点
2. 假设最终的状态为模型m，对每个符号分配真/假值
3. 在m中，原始KB 中的每个子句都为真

   $\boldsymbol{a_1 \wedge \ldots \wedge a_k \Rightarrow b}$

4. *因此m 是KB的一个模型*
5. 如果$KB \models q$, q 在KB 的每个模型中都为真，包括m

后向链接

想法: 从需要查询的q中进行反向链接：

利用BC 证明q,

检查q*是否是已知的*, 或者

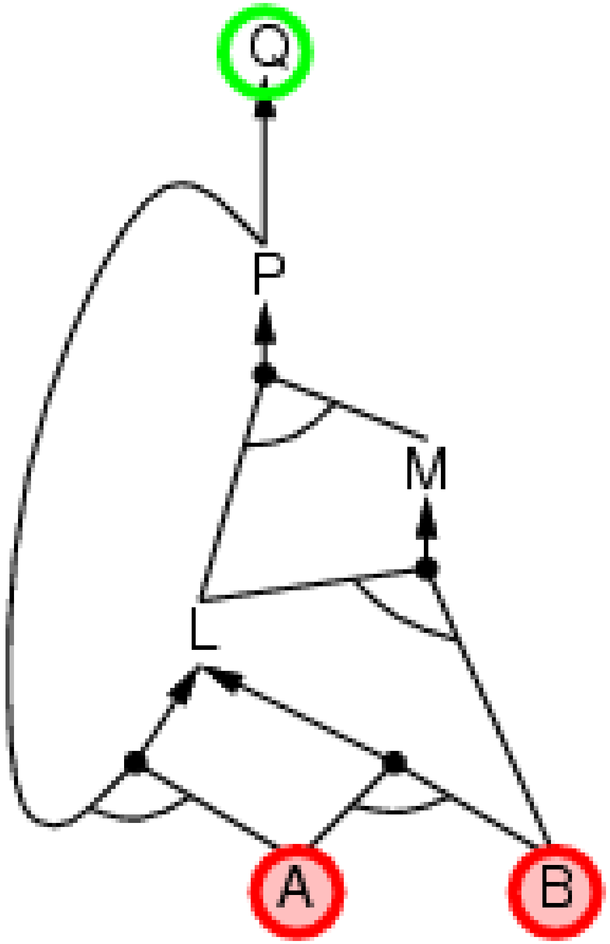
由BC证明某些规则的前提都能推断出q

避免循环: 检查新的子目标是否已经在目标堆栈上

避免重复工作: 检查是否是新的子目标

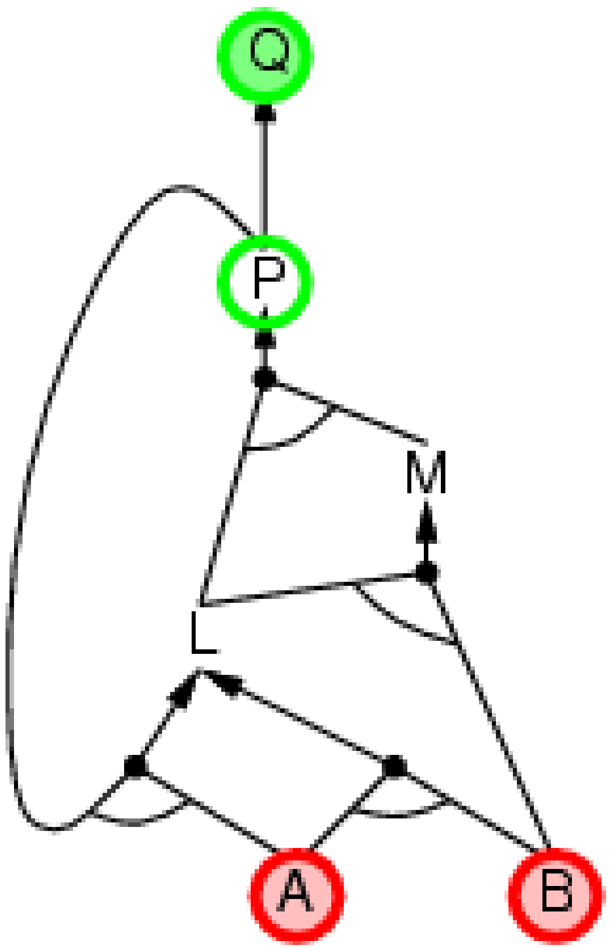
1. 已经被证明为真, 或者
2. 已经被证明为假

后向链接的例子



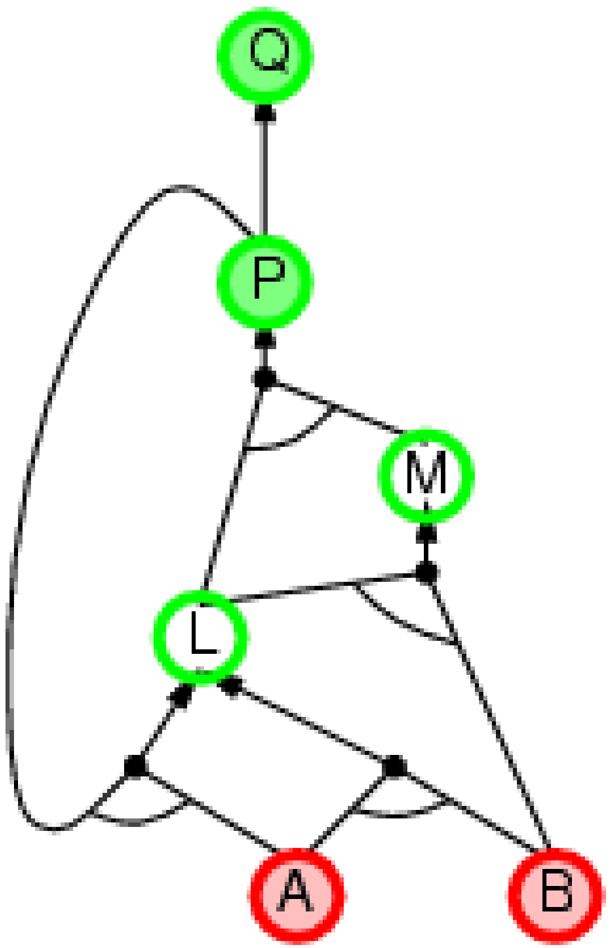
A
B
Q?

后向链接的例子



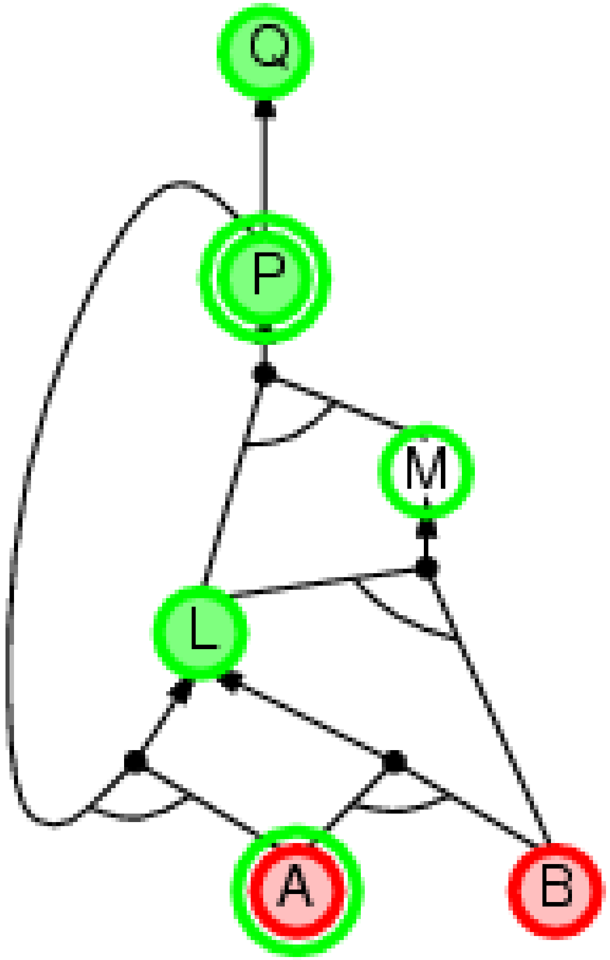
A
B
Q?<=P?

后向链接的例子



A
B
Q?<=P?
L? ∧ M? => P?

后向链接的例子



A

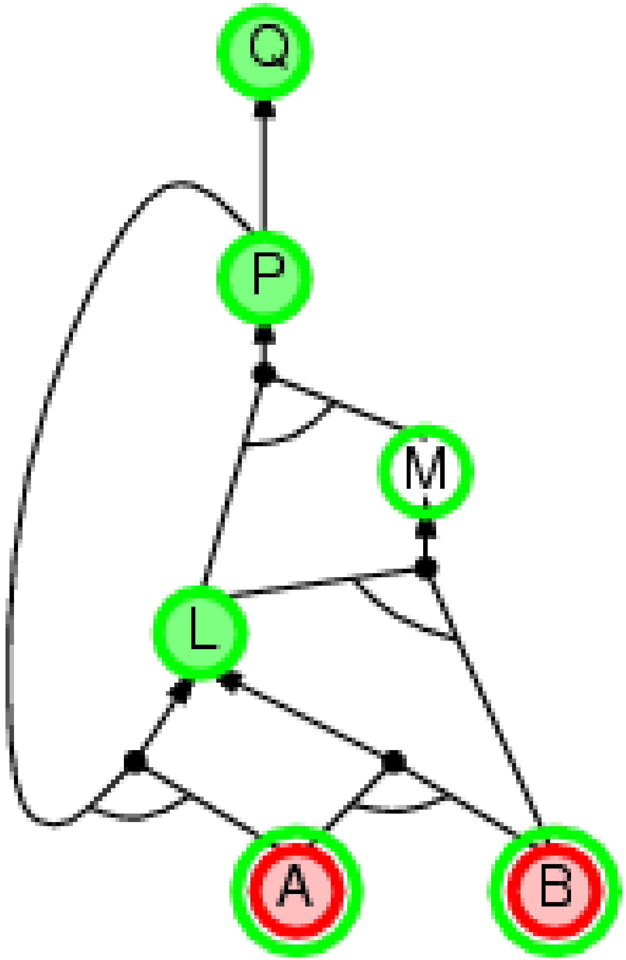
B

Q?<=P?

L? $\wedge$ ***M? => P?***

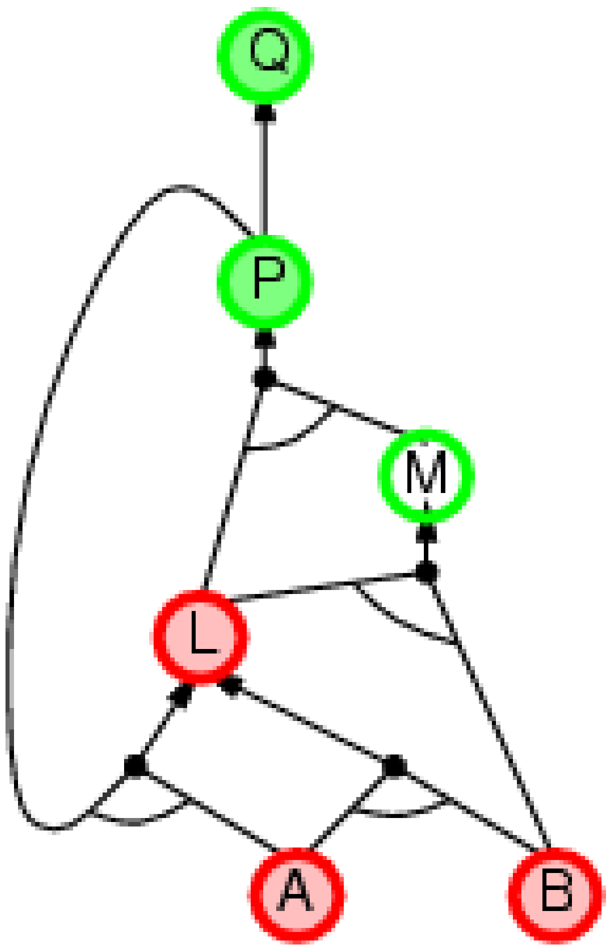
P? $\wedge$ ***A => L?***

后向链接的例子



A
B
Q?<=P?
L? ∧ M? => P?
P? ∧ A => L?
A ∧ B => L

后向链接的例子



A

B

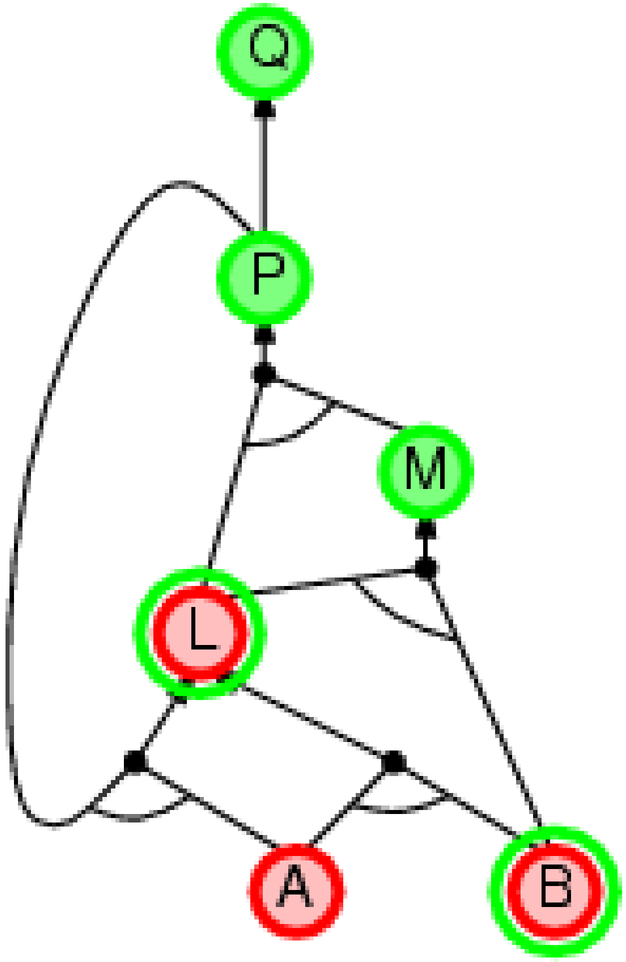
Q?<=P?

L? ∧ M? => P?

P? ∧ A => L?

A ∧ B => L

后向链接的例子



A

B

Q?<=P?

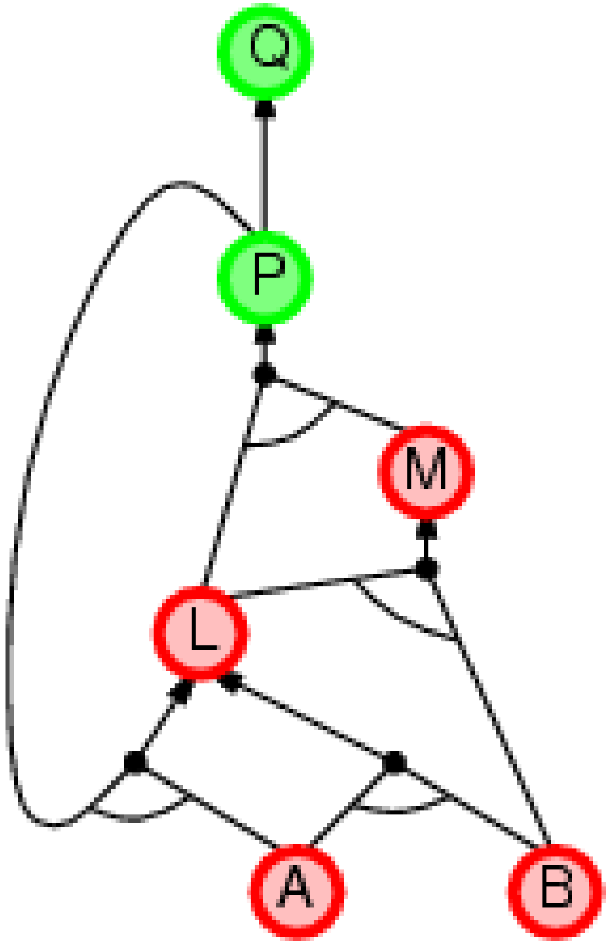
L? ∧ M? => P?

P? ∧ A => L?

A ∧ B => L

L ∧ B => M

后向链接的例子



A

B

Q?<=P?

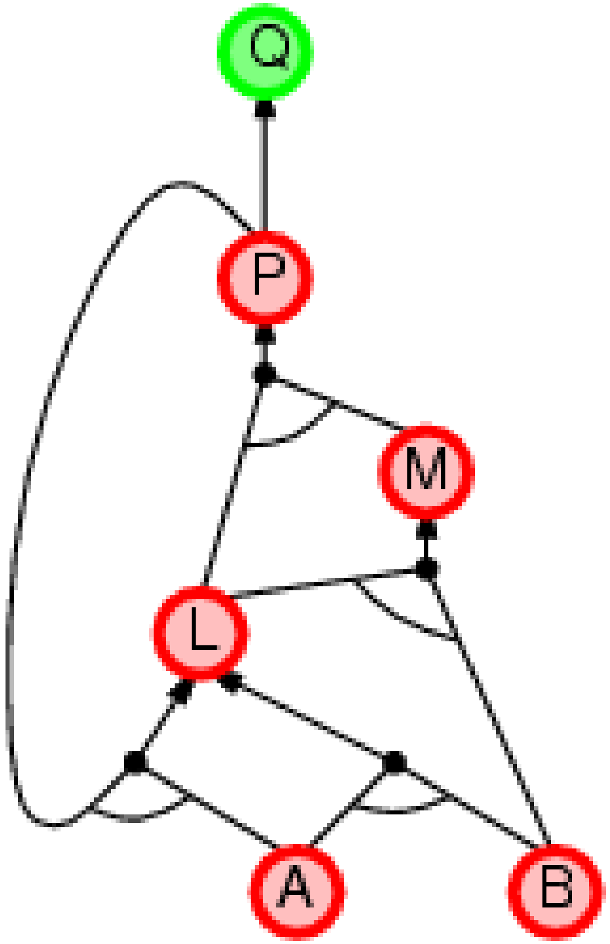
L ∧ M => P?

P? ∧ A => L?

A ∧ B => L

L ∧ B => M

后向链接的例子



A

B

Q?<=P

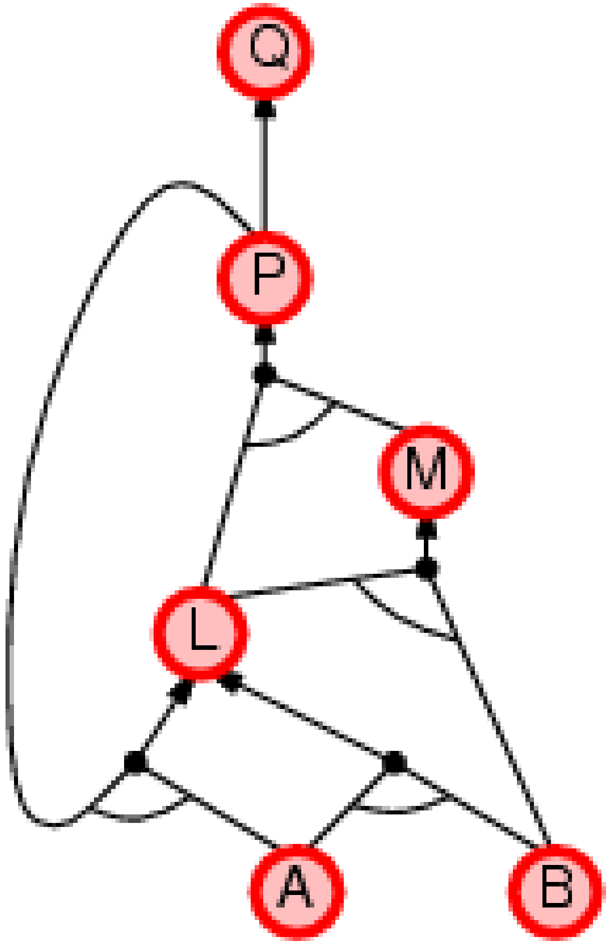
L ∧ M => P

P? ∧ A => L?

A ∧ B => L

L ∧ B => M

后向链接的例子



A
B
Q<=P
L ∧ M => P
P ∧ A => L
A ∧ B => L
L ∧ B => M

思考题

前向链接与后向链接之间的差异是什么？

前向链接vs. 后向链接

- **FC** 是数据驱动的, 自动的, 无意识的过程,
  - 例如, 目标识别, 路径决策
- 可能会做许多与目标相关的工作

- **BC** 是目标驱动的, 正确解决问题,
  - 例如, 我的钥匙在哪里？ 我如何才能上博士课程?

- BC的复杂度要远远小于KB的线性大小

总结

- 逻辑智能体在知识库中应用推理来得到新的信息并做决策
- 逻辑的基本概念:
  - 语法：语句的形式化结构
  - 语义：模型中为真的语句
  - 蕴含：给定一个语句的真假时另一个语句必然的真假
  - 推理：从其他语句中推倒出的语句
  - 真值保留：推倒能生成仅为真的语句
  - 完备性：推倒能生成所有蕴含的语句
- 怪兽世界需要具备表达部分和否定信息的能力，以及通过例子推理的能力等.
- 归结对于命题逻辑来说是完备的
  对于Horn词来说前向链接，后向链接是线性的，完备的