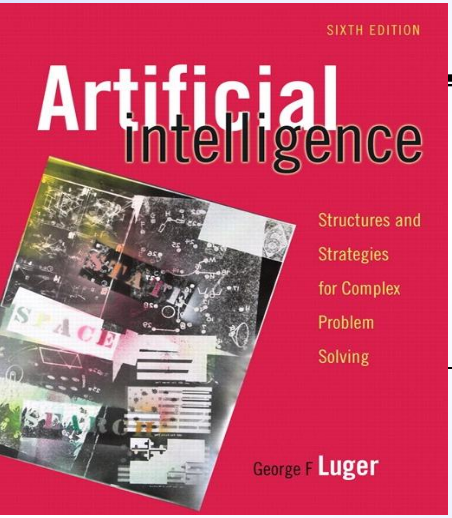
Chapter 9

Reasoning in Uncertain Situations

（不确定情况下的推理）

9 Reasoning in Uncertain Situations



George F Luger

ARTIFICIAL INTELLIGENCE *6th edition*

Structures and Strategies for Complex Problem Solving

9.0 INTRODUCTION

- As we know, we must draw useful conclusions from poorly formed and uncertain evidence using unsound inference rules.
- Human do it very successfully in daily life.

- Rule 2 :

  if

  the engine dose not turn over, and

  the lights do not come on

  then

  the problem is battery or cables.

- It is not absolute true, but useful
- Only heuristic
- The causes may be a bad starter motor and burned-out headlights.
- This is abductive（反绎, 溯因）reasoning.

- The converse（相反的事物）is true :
- Rule 2' :

  if

  the problem is battery or cables.

  then

  the engine dose not turn over, and

  the lights do not come on
- Rule 2' is absolute true, but not useful in diagnosis.

Abductive（反绎, 溯因）reasoning

- Formally, abduction states that :

  from $P \rightarrow Q$ and Q

  it is possible to infer P
- It is un-sound but useful

§1 知识的不确定性

专家系统中的不确定性表现在三个方面

- 证据或事实的不确定性
- 规则的不确定性
- 推理的不确定性

§1.1 证据的不确定性

- 证据的歧义性
- 证据的不完全性

  (1) 证据尚未收集完全

  (2) 证据的特征值不完全
- 证据的不精确性
- 证据的模糊性
- 证据的可信性
- 证据的随机性

§1.2　规则的不确定性

规则的不确定性包括：

- 构成规则前件的模式的不确定性
- 观察证据的不确定性
- 规则前件的证据组合的不确定性
- 规则本身的不确定性
- 规则结论的不确定性

在规则使用过程中，有两种典型的使用规则的不确定性

- 在推理过程中，若有多条规则可用时，则需要通过冲突消解从多条可用规则中选择一条规则激发。冲突消解策略包含有使用规则的不确定性。
- 在反向推理过程中，若有多个假设需要通过推理来验证时，先选择哪一个假设进行反向推理验证同样包含有使用规则的不确定性。

§1.3 推理的不确定性

推理的不确定性是指：由于证据的不确定性和规则的不确定性在推理过程中的动态积累和传播从而导致推理结论的不确定性。

不确定性测度的计算有以下三种基本的计算模式

1、证据组合的不确定性测度计算模式

已知证据$e_1,e_2, \ldots,e_n$的不确定性测度为$MU_1,MU_2, \ldots,MU_n$，求出逻辑组合的不确定性测度。

其中证据的逻辑组合有三种基本形式。

证据逻辑组合的不确定性测度

- **证据的合取组合的不确定性测度**

  $e_1,e_2, \ldots,e_n$的合取组合为$e_1 \wedge e_2 \wedge \ldots \wedge e_n$，n个证据合取组合的不确定性测度为$MU=f(MU_1,MU_2, \ldots,MU_n)$

- **证据的析取组合的不确定性测度**

  $e_1,e_2, \ldots,e_n$的析取组合为$e_1 \vee e_2 \vee \ldots \vee e_n$，n个证据析取组合的不确定性测度为$MU=g(MU_1,MU_2, \ldots,MU_n)$。

- **证据的否定的不确定性测度**

  证据e_i的否定为$\bar{e}_i$，证据e_i的否定的不确定性测度为$MU=n(MU_i)$。

2、并行规则的不确定性测度计算模式

已知有多条规则if e_i then h有相同的结论h，各条规则的不确定性测度为MU_i ,i=1,2, … ,n。若n条规则都被满足，那么，结论h的不确定性测度为

$MU=p(MU_1,MU_2, \ldots,MU_n)$。

这一计算模式也称为**并行法则**。并行法则给出了推理过程中有多条路径导致同一结论的情况下，由规则的不确定性而导致结论的不确定性测度的计算模式。

3、顺序(串行)规则的不确定性测度计算模式

已知两条规则if e then e'和if e' then h的规则不确定性测度分别为MU_1和MU_2，那么，规则if e then h的规则不确定性测度为$MU=s(MU_1,MU_2)$。

这一计算模式也称为顺序法则。顺序法则给出了规则不确定性在推理链中传播的计算模式。

对于一个专家系统，给定上述三种计算模式的不确定性测度计算方法，就可获得证据不同组合的不确定性测度值，并根据在推理过程中使用规则的情况，由并行法则和顺序法则最终得出结论的不确定性测度值。

在不同的专家系统中，不确定性测度的计算方法可以不同。根据不确定性测度计算方法的不同，不确定推理可以有基于概率理论的不确定推理、基于可信度理论的不确定推理和基于模糊理论的不确定推理等。

9.2.1 Stanford certainty theory

- The first assumption is to split “confidence for(支持度) ” from “confidence against（不支持度）” :
- Call MB(H|E) （可信度）the measure of belief of hypothesis H given evidence E.
- Call MD(H|E) （不可信度）the measure of disbelief of hypothesis H given evidence E.
- Now either:
  - $0 < MB(H|E) < 1$ while $MD(H|E) = 0$ or
  - $0 < MD(H|E) < 1$ while $MB(H|E) = 0$

- The measures of belief and disbelief may be tied together by :

$$CF(H|E) = MB(H|E) - MD(H|E)$$

- The combined CF（确信度因子） of the premises of a rule :

$$CF(P1 \text{ and } P2) = \min(CF(P1), CF(P2))$$

$$CF(P1 \text{ or } P2) = \max(CF(P1), CF(P2))$$

- Given the premises of a rule, the combined CF of the premises is multiplied by the CF of the rule itself to get the CF for the conclusions of the rule.

Example:

- Given a rule in the KB :
  - (P1 and P2) or P3 → R1(0.7) and R2(0.3)
- P1, P2, P3 are premises
- R1, R2 are conclusions
- 0.7 represent the expert' confidence in R1 if all the premises are known with complete certainty.
- 0.3 represent the expert' confidence in R2 if all the premises are known with complete certainty.

- If the running program has produced P1, P2, P3 :

  CF (P1) = 0.6, CF (P2) = 0.4, CF (P3) = 0.2, then :

- CF (P1(0.6) and P2(0.4)) = min(0.6, 0.4) = 0.4
- CF((0.4) or P3(0.2)) = max(0.4, 0.2) = 0.4
- The CF for R1 is 0.7 in the rule, so R1 is added to the set of case-specific knowledge with the associated CF :   (0.7) * (0.4) = 0.28
- The CF for R2 is 0.3 in the rule, so R2 is added to the set of case-specific knowledge with the associated CF :   (0.3) * (0.4) = 0.12

One further measure

- How to combine multiple CFs when two or more rules support the same result R?
- Suppose CF(R1) is the present certainty factor associated with result R and
- A previously un-used rule produces result R (again) with CF(R2) ;
- then the new CF of R is calculated as follows:

- CF(R1) + CF(R2) – (CF(R1) * CF(R2))

  when CF(R1) and CF(R2) are positive

- CF(R1) + CF(R2) + (CF(R1) * CF(R2))

  when CF(R1) and CF(R2) are negative

- $$\frac{CF(R1) + CF(R2)}{1 - \min(\,|CF(R1)|,\ |CF(R2)|\,)}$$

  Otherwise, where | X | is the absolute value of X.

Given the following rules in a “back-chaining” expert system application:

$$B \wedge not(G) \Rightarrow D(.6)$$

$$C \wedge D \Rightarrow E(.5)$$

$$A \Rightarrow C(.75)$$

$$F \Rightarrow B(.2)$$

The system can conclude the following facts (with confidences):

F(.3)   A(.8)   G(-.4)

Use the Stanford certainty factor algebra to determine **E** and its confidence.

We try to establish E through C and D. We turn first to support C through A. A has confidence .8, and supports C with confidence .75, so the confidence of C given A using the product rule, is .8*.75 = .6. To determine D, we look at the "and" of B and not G. F has confidence .3 and supports B with confidence .2, so the confidence of **B** given F is .3 * .2 = .06.

G is true with confidence -.4, thus not G is true with confidence .4. Since we are taking the "and" of **B** and **not G**, we take the MIN of its two confidences, which is .06; using the product rule, we have .06 * .6 = .036 (rounding to .04). Comparing **C** and **D**, we again take the MIN of .04 and .6, which is .04, using the product rule, we take .04* .5 = .02. So **E** is true with confidence .02.

1. 信任度与不信任度

定义1　信任度$MB(h,e)$表示证据e出现时，对结论h成立的信任程度的增加量。不信任度$MD(h,e)$表示证据e出现时，对结论h成立的不信任程度的增加量。$MB(h,e)$和$MD(h,e)$的取值范围为[0，1]。它们形式化地定义为：

$$MB(h,e)=\begin{cases}1 & P(h)=1\\ \dfrac{\max(P(h\mid e),P(h))-P(h)}{1-P(h)} & P(h)\neq 1\end{cases} \quad (1)$$

$$MD(h,e)=\begin{cases}1 & P(h)=0\\ \dfrac{\min(P(h\mid e),P(h))-P(h)}{-P(h)} & P(h)\neq 0\end{cases} \quad (2)$$

其中，$P(h)$为结论h成立的先验概率；$P(h\mid e)$为在证据e出现的条件下，结论h成立的条件概率。

$MD(h,e)$和 $MB(h,e)$的性质

性质1（互斥律）一个证据e不可能既支持又不支持某个结论h,因此有：

如果$MB(h,e)>0$，则 $MD(h,e)=0$

如果$MD(h,e)>0$，则 $MB(h,e)=0$

$MD(h,e)$ 和 $MB(h,e)$的性质

性质2 若$P(h|e)>P(h)$，表明证据e的出现增加了对结论h成立的信任程度，但是，不改变对结论h成立的不信任程度。

若$P(h|e)>P(h)$，由(1)式可见,有 $MB(h,e)>0$ ；由(2)式可见，有$MD(h,e)=0$ 。

$MD(h,e)$ 和 $MB(h,e)$ 的性质

性质3 若 $P(h|e)=P(h)$，表明证据e的出现不改变对结论h成立的信任程度，也不改变对结论h成立的不信任程度，即表明证据e与结论h之间相互独立。

若 $P(h|e)=P(h)$，由(1)式可见,有 $MB(h,e)=0$；由(2)式可见，有 $MD(h,e)=0$。即信任程度与不信任程度都不变。

$MD(h,e)$ 和 $MB(h,e)$ 的性质

性质4 若 $P(h|e) < P(h)$，表明证据e的出现增加了对结论h成立的不信任程度，但是，不改变对结论h成立的信任程度。

若 $P(h|e) < P(h)$，由(1)式可见，有

$MB(h,e) = 0$ ；由(2)式可见，有 $MD(h,e) > 0$ 。

2、可信度

在可信度不确定推理模型中，把信任度 MB 与不信任度 MD 组合成一个单一的不确定性测度，这就是**可信度**。

定义2 可信度形式化地定义为：

$$CF(h,e)=MB(h,e)-MD(h,e) \tag{3}$$

由 $CF(h,e)$、$MB(h,e)$ 和 $MD(h,e)$ 的定义(3)式、(1)式和(2)式以及 $MB(h,e)$ 与 $MD(h,e)$ 的互斥性质，可得出 $CF(h,e)$ 的计算公式为：

$$CF(h,e)=\begin{cases} 1 & 若\ P(h)=1 \\ \dfrac{P(h\mid e)-P(h)}{1-P(h)} & 若\ P(h\mid e)>P(h) \\ 0 & 若\ P(h\mid e)=P(h) \\ \dfrac{P(h\mid e)-P(h)}{P(h)} & 若\ P(h\mid e)<P(h) \\ -1 & 若\ P(h)=0 \end{cases} \tag{4}$$

由(4)式可直观地看出可信度的意义

1)若 $CF(h,e)>0$，则 $P(h|e)>P(h)$。说明证据e的出现增加了结论h为真的概率，即增加了h为真的可信度，

$CF(h,e)$ 的值越大，增加h为真的可信度就越大。若 $CF(h,e)=1$，则可推出 $P(h|e)=1$，即证据e的出现使h为真。

2)若 $CF(h,e)<0$，则 $P(h\mid e)<P(h)$。说明证据e的出现减少了结论h为真的概率，即增加了h为假的可信度，$CF(h,e)$的值越小，增加h为假的可信度就越大。若 $CF(h,e)=-1$，则可推出$P(h\mid e)=0$，即证据e的出现使h为假。

3)若 $CF(h,e)=0$，则 $P(h \mid e)=P(h)$，表示h与e独立，即证据 e的出现对h没有影响。

当已知 $P(h)$ 和 $P(h \mid e)$ 时，通过上述计算公式就可求出 $CF(h,e)$。但是，在实际应用中，获得 $P(h \mid e)$ 和 $P(h)$ 的值是比较困难的，而 $CF(h,e)$ 的值比较容易通过领域专家直接给出。

设定$CF(h,e)$的值的原则

若相应证据e能增加结论h为真的可信度，则使$CF(h,e)>0$，证据e越是支持h为真，就使$CF(h,e)$越大；反之，使$CF(h,e)<0$，证据e越是支持h为假，就使$CF(h,e)$值越小；若证据e与h无关，则使$CF(h,e)=0$。

（1）根据前提和规则的可信度求结论的可信度。

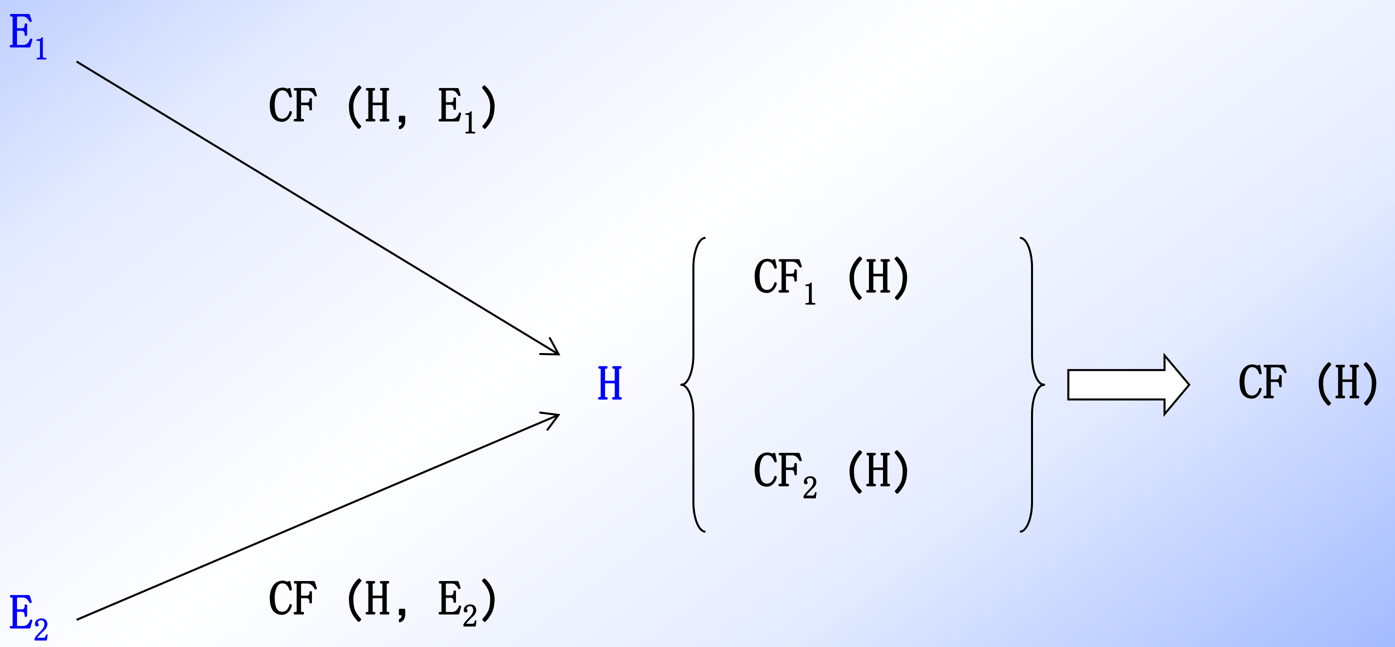
$$E \xrightarrow{CF(H,E)} H$$

CF（H）= max ｛0，CF（E）｝× CF（H，E）

即CF（E）<0时，CF（H）=0

表示对H仍是“一无所知”、“不置可否”，即“否定前件，不能保证否定后件”。

（2）使用两个独立证据和两条不同规则导出的同一结论的可信度。



由前面（1）得：

$CF_1(H) = \max\{0, CF(E_1)\} \times CF(H, E_1)$

$CF_2(H) = \max\{0, CF(E_2)\} \times CF(H, E_2)$

CF（H）定义如下：

$$CF(H)=\begin{cases} CF_1(H)+CF_2(H)-CF_1(H)\times CF_2(H) & \text{（当 } CF_1(H)\geqslant 0\text{，}CF_2(H)\geqslant 0\text{ 时）} \\ CF_1(H)+CF_2(H)+CF_1(H)\times CF_2(H) & \text{（当 } CF_1(H)<0\text{，}CF_2(H)<0\text{ 时} \\ (CF_1(H)+CF_2(H))\,/\,(1-\min(|CF_1(H)|,|CF_2(H)|)) & \text{（异号）} \end{cases}$$

(3) 合取证据的可信度

$E = E_1$ and E_2 and … and E_n

则有:

$CF(E) = \min\{CF(E_i)\}$

(4) 析取证据的可信度

$E = E_1$ or E_2 or … or E_n

则有:

$CF(E) = \max\{CF(E_i)\}$

例:

已知规则:

r_1: IF E_1 THEN H （0.9）

r_2: IF E_2 THEN H （0.7）

r_3: IF E_3 THEN H （－0.5）

r_4: IF E_4 and E_5 THEN E_1 （0.6）

已知证据:

CF（E_2）=0.8 CF（E_3）=0.2

CF（E_4）=0.6 CF（E_5）=0.7

求: CF（H）=?

解： 由 r_2：IF E_2 THEN H （0.7）和 CF（E_2）=0.8 得：

$$CF_1(H) = 0.8 \times 0.7 = 0.56$$

· 由r_3：IF E_3 THEN H （– 0.5）和 CF（E_3）=0.2 得：

$$CF_2(H) = 0.2 \times (-0.5)$$
$$= -0.1$$

· 综合CF_1（H）和CF_2（H）得：

$$CF_{1,2}(H) = (CF_1(H) + CF_2(H)) / (1 - \min(|CF_1(H)|, |CF_2(H)|))$$
$$= (0.56 - 0.1) / (1 - 0.1)$$
$$= 0.51$$

- 由r_4：IF E_4 and E_5 THEN $E_1(0.6)$和 $CF(E_4)=0.6$, $CF(E_5)=0.7$ 得：

$$CF(E_1) = 0.6\times\min\{CF(E_4), CF(E_5)\}$$
$$= 0.6\times 0.6$$
$$= 0.36$$

- 由r_1：IF E_1 THEN H （0.9）和 $CF(E_1)=0.36$ 得：

$$CF_3(H) = 0.9 \times CF(E_1)$$
$$= 0.324$$

- 综合$CF_{1,2}$（H）和CF_3（H）得：

$$CF(H) = CF_{1,2}(H) + CF_3(H) - CF_{1,2}(H)\times CF_3(H)$$
$$= 0.51 + 0.324 - 0.51\times 0.324$$
$$\approx 0.67$$

Given the following rules in a "back-chaining" expert system application:

$$B \wedge not(G) \Rightarrow D(.6)$$

$$C \wedge D \Rightarrow E(.5)$$

$$A \Rightarrow C(.75)$$

$$F \Rightarrow B(.2)$$

The system can conclude the following facts (with confidences):
F(.3) A(.8) G(-.4)
Use the Stanford certainty factor algebra to determine E and its confidence.

We try to establish E through C and D. We turn first to support C through A. A has confidence .8, and supports C with confidence .75, so the confidence of C given A using the product rule, is .8*.75 = .6. To determine D, we look at the "and" of B and not G. F has confidence .3 and supports B with confidence .2, so the confidence of B given F is .3 * .2 = .06.

G is true with confidence -.4, thus not G is true with confidence .4. Since we are taking the "and" of B and not G, we take the MIN of its two confidences, which is .06; using the product rule, we have .06 * .6 = .036 (rounding to .04). Comparing C and D, we again take the MIN of .04 and .6, which is .04, using the product rule, we take .04* .5 = .02. So E is true with confidence .02.

不精确推理基本概念

1.不确定性方法中，前提A真支持结论B为真时，CF（B，A）的取值为（ ）。

a）1　b）0　c）小于0　d）大于等于0

2.在不确定性方法中，规则的不确定性度量ＣＦ（Ｂ,Ａ）的取值范围为（ ）。

a） 0＞ＣＦ（Ｂ,Ａ）＞-１　b） 0＞ＣＦ（Ｂ,Ａ）

c） 0＜ＣＦ（Ｂ,Ａ）＜１　d）-１ ＜ＣＦ（Ｂ,Ａ） ＜ １

3. 在不确定性方法中，ＣＦ（Ｂ,Ａ）的取值为（　　）时，前提Ａ真不支持结论Ｂ真。

a） 1　b） 0　c）＜0　d）-1

4. MYCIN系统中规定，若证据A的可信度CF(A) = 0，则意味着(　)。

a) 证据不可信　b) 对证据一无所知　c) 证据可信

5.基于概率的推理中，规则E→H, 其P（H）= P（H | E），这意味着(　)。

a) E对H没有影响　b) E支持H　c) ~E支持H

1. 在基于可信度的不精确推理中，

$CF(h,e)>0$ 表示前提e（　　　）结论h的发生；

$CF(h,e)=0$ 表示前提e和结论h（　　　）；

$CF(h,e)<0$ 表示前提e（　　　）结论h的发生。

→

2. 在基于概率的不精确推理中

$P(h \mid e)>P(h)$ 表示前提e（　　　）结论h的发生；

$P(h \mid e)<P(h)$ 表示前提e（　　　）结论h的发生；

$P(h \mid e)=P(h)$ 表示前提e（　　　）结论h的发生。

9.3 The stochastic(随机) approach to uncertainty

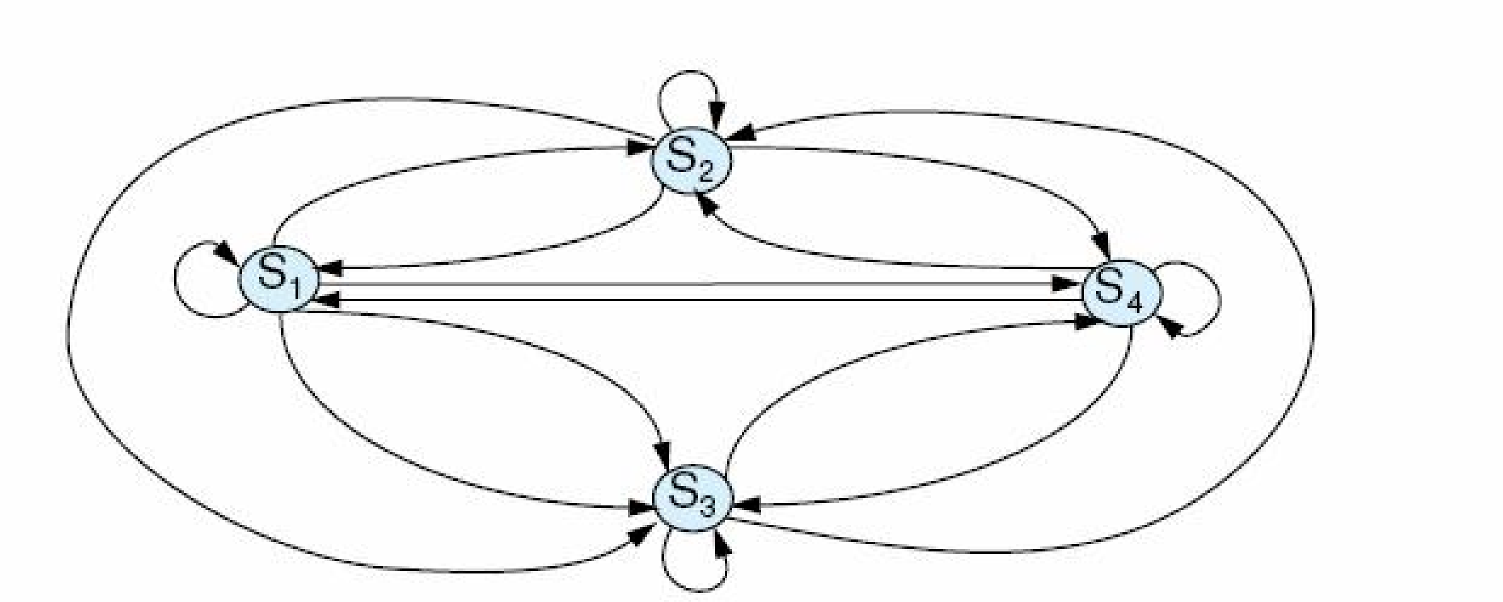
9.3.5 Markov Models: the Discrete Markov Process

- In section 5.3 we presented a probabilistic finite state machine, whose next state function was represented by a probability distribution on the current state.

- The discrete Markov process is a specialization of this approach, where the system ignores its input values.

- The system undergoes changes of state, with the possibility of remaining in the same state, at regular discrete time intervals.

A **Markov state machine** or **Markov chain** with four states, $s_1, \ldots, s_4$



- In a Markov chain, The probability of the system being in any particular state δt is :

$$p(\delta_t) = p(\delta_t \mid \delta_{t-1}, \delta_{t-2}, \delta_{t-3},\ldots)$$

- In a first-order Markov chain, the probability of the present state is a function of its direct predecessor（直接前驱）state :

- $p(\delta_t) = p(\delta_t \mid \delta_{t-1})$

DEFINITION(OBSERVABLE) MARKOV MODEL

- A **graphical model** is called an (observable) Markov model if its graph is **directed** and **the probability of arriving at any state s_t** from the set of states S at a discrete **time t** is a **function** of the **probability distributions** of its being in **previous states** of S at **previous times**. Each state s_t of S corresponds to a **physically observable situation**.

- An observable Markov model is **first-order** if the probability of **it being in** the present state s_t at any time t is a function only of **its being in** the **previous state s_{t-1}** at the time t-1, where s_t and s_{t-1} belong to the set of observable states S.

- Given $S = \{ s_1 , s_2 , \ldots\ldots , s_N \}$, we can create a set of state transition probabilities a_{ij} between any two states s_i and s_j :

$$a_{ij} = p(\sigma_t = s_j \mid \sigma_{t-1} = s_i)$$

where :

$$a_{ij} \geq 0$$

$$\Sigma a_{ij} = 1 \quad (j = 1, 2, \ldots , N)$$

Example

- Consider the weather at noon for a particular location.
- We assume this location has four different discrete states for the variable weather :

  s_1 = sunny, s_2 = cloudy,

  s_3 = fog, s_4 = precipitation（降水）
- The time intervals will be noon each consecutive（连续的）day.

$a_{ij}=$

	S1	S2	S3	S4
S1	0.4	0.3	0.2	0.1
S2	0.2	0.3	0.2	0.3
S3	0.1	0.3	0.3	0.3
S4	0.2	0.3	0.3	0.2

- **The above transition matrix determines a first-order Markov model M .**

$a_{ij}=$

	S1	S2	S3	S4
S1	0.4	0.3	0.2	0.1
S2	0.2	0.3	0.2	0.3
S3	0.1	0.3	0.3	0.3
S4	0.2	0.3	0.3	0.2

- **Suppose today is sunny (s1), we now may ask questions such as :**

1. **"what is the probability of the next five days remaining sunny ? "**
2. **"what is the probability of the next five days being sunny, sunny, cloudy, cloudy, precipitation ? "**

$a_{ij}=$

	S1	S2	S3	S4
S1	0.4	0.3	0.2	0.1
S2	0.2	0.3	0.2	0.3
S3	0.1	0.3	0.3	0.3
S4	0.2	0.3	0.3	0.2

- $O = s_1, s_1, s_1, s_2, s_2, s_4$
- $P(O \mid M)$

$= p\,(\,s_1, s_1, s_1, s_2, s_2, s_4 \mid M\,)$

$= p(s_1)\,p(s_1|s_1)\,p(s_1|s_1)\,p(s_2|s_1)\,p(s_2|s_2)\,p(s_4|s_2)$

$= 1 \times a_{11} \times a_{11} \times a_{12} \times a_{22} \times a_{24}$

$= 0.0432$

- Given today's weather, We can extend this example to determine the probability that the weather will be the same for exactly the next t days .
- i.e. , the weather remains the same until the $t + 1$ day at which time it is different.
- i.e. , from today on, the weather remains the same for $t + 1$ day.

- For any weather state s_i , and Markov model **M** ,we have the observation O :

$$O = (s_i, s_i, \ldots s_i, s_j) ,$$

where there are exactly (t + 1) s_i ,

and $s_i \neq s_j$,

then :

- $p(O \mid M) = 1 \times a_{ii}^{t} \times (1 - a_{ii})$

- **Based on this value we can calculate, within Model M, the expected number of observations of, or duration d_i within any state s_i , given that the first observation is in that state.**

- **In other words, we can calculate the average day number d_i in s_i**

the average day number d_i in s_i

$$d_i = \Sigma d \times (a_{ii})^{(d-1)} \times (1 - a_{ii}) \qquad d = 1, 2, \ldots, n$$

where n approaches ∞, so we get :

$$d_i = 1 / (1 - a_{ii})$$

9.3.6 Markov Models: Variations

- Hidden Markov Models
- Semi - Markov Models
- Markov decision processes

Given the observable Markov model of weather of Section 9.3.5:

a) Determine the probability that (exactly) the next three days will be foggy.

b) What is the probability of exactly two days of sun, then three days of precipitation, followed by one day of clouds?

Solution:

a) $O = s_3, s_3, s_3$

$P(O) = P(s_3) * P(s_3 \mid s_3) * P(s_3 \mid s_3) = 1.0 * .3 * .3 = .09$

b) $O = s_1, s_1, s_4, s_4, s_4, s_2$

$P(O) = P(s_1) * P(s_1 \mid s_1) * P(s_4 \mid s_1) * P(s_4 \mid s_4) * P(s_4 \mid s_4) * P(s_2 \mid s_4)$

$= 1.0 * .4 * .1 * .2 * .2 * .3$

$= .00048$

Exercises

2.

19.