王士同 陈剑夫 编著

问题求解的人工智能神经网络方法

气象出版社

问题求解的人工智能、神经网络方法

王士同　陈剑夫　编著

气象出版社

（京）新登字046号

内 容 简 介

本书是问题求解技术方面的一本专著。问题求解中的启发式搜索技术是人工智能的核心技术，是当代人工智能的基本支柱之一；基于神经网络计算原理的问题求解是当前神经网络应用的一个重要的方面，为组合优化问题求解提供了一个全新方法。著者根据国内外大量最新资料及自己的科研成果，从人工智能、神经网络角度介绍了四类问题求解技术。

书中的许多新概念、新方法、新成果将使读者耳目一新。本书适宜于应用数学、自动控制、计算机、人工智能以及神经网络方面的科技人员阅读。

问题求解的人工智能、神经网络方法

王士同　陈剑夫　编著

责任编辑：吴向东　终审：周诗健

封面设计：严瑜仲责任技编：席大光　责任校对：吴向东

*

气象出版社出版

（北京西郊白石桥路46号　邮编100081）

北京昌平环球印刷厂印刷

新华书店总店科技发行所发行　全国各地新华书店经销

*

开本：787×1092　1/32　印张：10.625　字数：234千字

1995年1月第一版　　1995年1月第一次印刷

印数：1-1800

ISBN 7-5029-1807—8/O·0031

定价：9.50元

前　　言

人工智能自1956年创立以来，已取得了许多重大成就。可以说，人工智能开创了智能化的新纪元。问题求解是人工智能的核心课题。目前，人工智能学术界在问题求解技术研究方面存在两大流派：符号主义流派；连接主义流派。符号主义思想认为：人类尽管自身的能力和知识有限，但却能运用自如地解决那些看来难以胜任的问题，并作出正确的决策，其秘密在于，人们通常不是采用系统的、精确的方法去追求问题的最佳解，而是通过逐步尝试的方法，达到有限的合理的目标，即找到最佳或较满意的解，也就是运用启发式搜索原理。启发式搜索原理是人工智能的核心原理之一，其特点是把启发式知识与搜索相结合，即用搜索弥补知识的不足，反过来，启发式知识越多，则搜索量越少。目前启发式搜索技术已有很多著名成果，正逐步走向成熟。连接主义思想从人工智能角度来看则是以连接主义为基础的所谓大脑模式的人工智能，它运用神经网络的计算原理来求解优化问题。基本神经网络的问题求解技术是一种无知识的随机搜索技术，它利用神经元的巨量并行和神经元的动力学演变来快速求解优化问题。基于神经网络的问题求解技术为问题求解理论提供了一种全新的方法。自1985年Hopfield教授开创了这一问题求解技术以来，已取得了引人注目的成就，目前这方面的研究正方兴未艾。

作者认为，在问题求解理论方面，尽管符号主义和连接主义在表面上看来是对立的，但它们实际上只是从不同的侧

面或不同的层次上揭示了问题求解的本质，提供了问题求解的方法，它们有可能相互统一和结合。因此，将启发式搜索技术与基于神经网络的问题求解技术相结合应该是一个极有前途的问题求解理论之研究方向。

本书是问题求解技术方面的一本专著。作者根据国内外最新研究成果以及自己的科研成果，从人工智能和神经网络角度介绍了问题求解的四大类求解技术：数学函数式启发式搜索技术；统计启发式搜索技术；模糊启发式搜索技术；以及基于神经网络计算原理的问题求解技术。同时还介绍了并行式和学习式问题求解技术以及这四大类求解技术的应用实例。

本书共分九章。第一章简要介绍统计推断理论和模糊数学基本理论。第二章简要介绍问题求解的基本内容以及四大类问题求解技术。第三、四章研究数学函数式启发式搜索技术，它们实质上就是利用数学函数式表示的启发信息的估计精度来引导问题求解的搜索过程。第五章研究统计启发式搜索技术，它们是基于统计推断理论，利用问题求解空间不同部分之间的统计量之差异来引导问题求解的搜索过程，第六、七章研究模糊启发式搜索技术。模糊启发式搜索技术根据模糊集理论来表示不精确且模糊的启发信息，并进而进行独特的模糊启发式搜索。第八章研究基于并行求解机制和学习理论的启发式搜索技术。第九章主要以Hopfield神经网络为计算模型来研究基于神经网络计算原理的问题求解技术，同时亦介绍了以模拟退火法和Kohenen 模型为计算模型的神经计算的问题求解技术。

本书第一、五两章，由陈剑夫同志编写，其余章节均由王士同编写，并由王士同统编全书。本书由夏振华、杨国庆

两位教授审阅。他们仔细审阅了全稿，并提出了许多宝贵意见；高群同志抄写了本书的部分书稿；本书的出版得到了国家自然科学基金委员会的大力支持，笔者在此一并致谢！

由于编著者学识疏浅，书中错漏之处恳请读者批评指正。

作者

994. 4

目　　录

第一章　统计推断和模糊数学基础

§1.1　概 率 基 础

在实际问题中，所测量到的数据都是带有误差的数据，故我们在测量数据时，还需要测量误差。由于所测量的数据是有分布的，所以人们可以把它视为一个随机变量，而把测量到的数据看作该随机变量所取的值。从这个意义上看，概率分布与随机变量是分不开的。

设X为离散型随机变量，可能的值为$x_1,x_2,\cdots$，若

$$P(X=x_i)=p_i,\ i=1,2,3,\cdots$$

其中p_i已知，且$\Sigma p_i=1$，则称上式为随机变量X的概率分布。

对任一实数x，考虑事件“$X\leqslant x$”的概率$P(X\leqslant x)$。若X是离散型随机变量，则有

$$P(X\leqslant x)=\sum_{x_i\leqslant x}P(X=x_i)$$

不难看到，$P(X\leqslant x)$是x的函数，我们称它为随机变量X的分布函数，记为$F(x)=P(X\leqslant x)$

若X是连续型随机变量，则它的分布函数也可以用$P(X\leqslant x)$来描述，即

$$F(x)=P(X\leqslant x)=\int_{-\infty}^{x}f(x)\mathrm{d}x$$

其中右端的被积函数$f(x)$称为连续型随机变量X的概率密度函数。

分布函数$F(x)$和概率密度函数$f(x)$具有如下性质：①$F(x)$是x的不减函数；②$0 \leqslant F(x) \leqslant 1, F(-\infty)=0, F(+\infty)=1$；③$F(x)$至多有可列个间断点；④$f(x)=F'(x)$；⑤$f(x) \geqslant 0$；⑥$\int_{-\infty}^{\infty} f(x)\mathrm{d}x=1$

期望和方差是关于随机变量的极为重要的概念。

设X是离散型随机变量，其概率分布为

$$(X=x_i)=p_i, \quad i=1,2,3,\cdots$$

若级数$\sum x_i p_i$绝对收敛，则称$\sum x_i p_i$是随机变量X的期望，记作$E(X)$，即$E(X)=\sum\limits_i x_i p_i$。

设X为连续型随机变量，其概率密度函数为$f(x)$。若积分$\int_{-\infty}^{\infty} xf(x)\mathrm{d}x$绝对收敛，则称这个积分值为随机变量$X$的期望，记作$E(X)$，即$E(X)=\int_{-\infty}^{\infty} xf(x)\mathrm{d}x$。

期望E具有如下的重要性质。设X,Y是随机变量，a为常数，则：①$E(a)=a$；②$E(a \cdot X)=a \cdot E(X)$；③$E(X \pm Y)=E(X) \pm E(Y)$；④当$X,Y$相互独立时，$E(X \cdot Y)=E(X) \cdot E(Y)$。

随机变量X离差平方的期望，称为随机变量X的方差，记作$D(X)$，即$D(X)=E(X-E(x))^2$。

方差具有下列重要性质。设X，Y是随机变量。a,b为常数，则：①$D(a)=0$；②$D(a \cdot X)=a^2 \cdot D(X)$；③$D(X \pm a)=D(X)$；④当$X$，$Y$相互独立时，$D(aX \pm bY)=a^2D(X)+b^2D(Y)$；⑤$D(X)=E(X^2)-E^2(X)$。

随机变量的标准差定义为随机变量X方差的算术平方根。

随机变量的一个重要分布是正态分布。若随机变量X的概率密度函数$f(x)$为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$$

$$-\infty<x<\infty$$

则称随机变量X服从正态分布，记作$X\sim N(\mu,\sigma^2)$。正态分布的两个参数μ和σ具有特别意义，即

$$E(X)=\mu;\ D(X)=\sigma^2$$

下面介绍重要的大数定律。

定理1.1 设随机变量$X_1,X_2,\cdots$是独立同分布的（常简记成 i.i.d., independent identically distribution），$E(X_1)=\mu<\infty$，则有

$$①弱大数定律：\overline{X}_n\overset{定义}{=\!=}\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}E(X_1)=\mu\ (n\to\infty)$$

亦即$\forall\varepsilon>0$，有 $P\{|\overline{X}_n-\mu|\leqslant\varepsilon\}\to1$。

②强大数定律：$\lim\limits_{n\to\infty}\overline{X}_n=\mu$

即 $$P\{\lim_{n\to\infty}\overline{X}_n=\mu\}=1.$$

接着介绍中心极限定理。

定理1.2 设$x_1,x_2,\cdots,x_n,\cdots$是i.i.d.，令$a=E(x_1),b=D(x_1)$，$B_n=\sqrt{n}b$，且$b>0$，则有

$$P\left\{\frac{1}{B_n}\sum_{k=1}^{n}(x_k-a)<x\right\}$$

$$\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-t^2/2)dt.$$

§1.2 序贯统计推断

统计推断是数理统计学科中的重要内容之一。序贯统计推断是通过对样本的抽样观察，对关于样本的统计假设进行分析和推断。在序贯统计推断中，样本容量不是事先固定的，它到底是多少，要在抽样过程中决定，它与到目前为止所得到的信息有关。采用序贯抽样，往往可以减少抽样次数，还可以解决一些在固定样本容量的情况下解决不了的问题。序贯统计推断理论是 Wald 在第二次世界大战期间为了战争的需要而发展起来的。本节介绍Wald的序贯概率比检验法SPRT (sequential probablity ratio test) 和较新的均值渐近序贯固定宽度可信度估计法ASM (asymptotically efficient sequential fixed-width confidence estimation of the mean)。在本书所要介绍的统计启发式搜索技术中要应用SPRT和ASM统计推断法。

1.2.1 SPRT统计推断法

假设 $x_1, x_2, \cdots, x_n \cdots$ 是独立分布的随机变量序列(i.i.d.)，$f(x;\mu)$是它的分布密度函数。两个简单假设 H_0: $\mu=\mu_0$，H_1: $\mu=\mu_1$，$\mu_1 \neq \mu_0$。因此给定n个观察值，可以组成如下的和式

$$S_n=\sum_1^n \ln\frac{f(x_i;\mu_1)}{f(x_i;\mu_0)},\quad n\geqslant 1$$

根据停止规则，当

$$-b<S_n<a$$

继续进行取样。如果$S_R\leqslant -b$，接受假设H_0，取样在第k个

观察停止。如果$S_R \geqslant a$，接受假设H_1。其中a和b是预先给定常数，$0<a<b<\infty$。

SPRT 法已很完善，下面我们简单介绍以后要用到的定理。

定理1.3 若假设H_0，H_1为真，则停止变量k取有限值的概率为 1 。

定理1.4 若$P_\mu(|Z|>0)>0$，则$P_\mu(R>n)\leqslant e^{-cn}$ $(c>0)$，其中R是SPRT的停止随机变量，$Z \xlongequal{\triangle} \ln\dfrac{f(x;\mu_1)}{f(x;\mu_0)}$.

定理1.5 给定显著水平(α,β)，令

$$A=\frac{1-\beta}{\alpha},\quad B=\frac{\beta}{1-\alpha},$$

$$Z_i \xlongequal{\triangle} \ln\frac{f(x_i;\mu_1)}{f(x_i;\mu_0)},$$

$$Z \xlongequal{\triangle} \ln\frac{f(x;\mu_1)}{f(x;\mu_0)}.$$

若$E_{\hat{\mu}_i}(|Z|)<\infty$，$E_{\hat{\mu}_i}(Z)\neq 0$ $(i=0,1)$，那么SPRT的停止变量的均值（即平均子样容量）是

$$E_{\hat{\mu}_i}(R)\approx\frac{\alpha\ln\dfrac{1-\beta}{\alpha}+(1-\alpha)\ln\dfrac{\beta}{1-\alpha}}{E_{\hat{\mu}_i}(Z)}$$

若 $$f(x;\mu)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

那么 $$Z \xlongequal{\triangle} \ln\frac{f(x;\mu_1)}{f(x;\mu_0)}$$

$$=\frac{1}{2\sigma^2}[2x(\mu_1-\mu_0)+(\mu_0^2-\mu_1^2)]$$

$$S_n \stackrel{\triangle}{=} \sum_{i=1}^{n} Z_i$$

$$= \frac{1}{2\sigma^2}\left[2(\mu_1 - \mu_0)\sum_{i}^{n} x_i + n(\mu_0^2 - \mu_1^2)\right]$$

故若随机变量是正态分布的，则SPRT的停止规则如下：

$$\begin{cases} 若\sum_{i}^{n} x_i \geqslant \dfrac{\sigma^2 g_1}{\mu_1 - \mu_0} + \dfrac{n}{2}(\mu_1 + \mu_0)，拒绝假设H_0 \\ 若\sum_{i}^{n} x_i \leqslant \dfrac{\sigma^2 g_2}{\mu_1 - \mu_0} + \dfrac{n}{2}(\mu_1 + \mu_0)，接受假设H_0 \\ 否则继续观察x_{n+1} \end{cases}$$

其中　$g_1 = \ln\dfrac{1-\beta}{\alpha}$，$g_2 = \ln\dfrac{\beta}{1-\alpha}$。

犯第Ⅰ类弃真错误的概率 $P_1 \leqslant \alpha$，犯第Ⅱ类纳伪错误的概率 $P_2 \leqslant \beta$。

1.2.2　ASM法

设$x_1, x_2, \cdots, x_n, \cdots$是i.i.d.随机变量序列，其共同的分布密度函数为F，$F \in \mathscr{R}$。其中$\mathscr{R}$是具有有限四阶矩的分布函数集合。给定$\delta > 0$和$r(0 < r < 1)$，我们用下式定义停止变量$R(\delta)$，$R(\delta)$是满足下式的最小整数。

$$R \geqslant \frac{a^2}{\delta^2}\left\{\frac{1}{R}\left(1 + \sum_{1}^{R}(x_i - \overline{X}_R)^2\right)\right\}$$

其中

$$\overline{X}_R = \frac{1}{R}\sum_{1}^{R} x_i，$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt，$$

$$a=\phi^{-1}\left(\frac{1+r}{2}\right).$$

令μ是$\{x_i\}$的均值，则有下面定理成立。

定理1.6 基于上述定义有下面性质成立。

性质(一)：$\forall F\in\mathscr{R}$，$\mu\in(\overline{X}_{R(\delta)}-\delta,\ \overline{X}_{R(\delta)}+\delta)$的概率大于$r$。其中$(\overline{X}_{R(\delta)}-\delta,\ \overline{X}_{R(\delta)}+\delta)$是固定宽度信度区间，以后用$I(\overline{X}_{R(\delta)},\ \delta)$表示，$\mu$是$\{x_k\}$的均值。

性质(二)：若$\delta_1<\delta_2$，则有$R(\delta_2)<R(\delta_1)$a.s.且$\lim\limits_{\delta\to 0}R(\delta)=\infty$a.s.，其中a.s.表示几乎处处成立。

性质(三)：$\forall F\in\mathscr{R}$，有

$$\lim_{\delta\to 0}\frac{\delta^2E(R(\delta))}{a^2\sigma^2(F)}=1$$

其中$\sigma^2(F)$是F的方差，$E(R(\delta))$是$R(\delta)$的均值。

性质(四)：$P(R(\delta)>n)\sim O(\exp(-cn^2))$，$c>0$且是常数。令$\alpha=1-r$，由于$a^2\sim O(|\ln(1-r)|)$，于是由性质(三)得

$$E(R(\delta))\sim O(\delta^2|\ln\alpha|)$$

§1.3 模糊集理论

一些事物的全体叫做一个普通集合，有时常称做集合。这些事物中的每一个个体都称为这个集合的元素，普通集合是一种边界明确的集合，一个元素与一个集合之间只有完全属于和完全不属于两种关系，不存在中间状态。一个集合A可以用其特征函数来表示。这个函数μ_A定义于论域U上，但只取0，1值，即

$$\mu_A(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases}$$

但是，现实世界中存在着许多边界不明确的分类。例如，“远远大于1的实数”就是论域（实数轴）上的一个没有明确边界的分类。例如，我们不能肯定5这个数是不是远大于1。显然说5这个数对于“远大于1的实数”这个分类的隶属程度是0.2，比肯定地说5属于或不属于这个分类要合理得多。这一修改意味着把普通集合特征函数的值域从$\{0,1\}$扩展到区间$[0,1]$之中，因此产生了模糊集这个新概念。

定义1.1 论域U中的一个模糊集合A由一个隶属函数$\mu_A(x)$：$U\to[0,1]$所表征，隶属函数把区间$[0,1]$中的一个数$\mu_A(x)$与U中的每一个元素x对应起来，说明x对A的隶属程度。

例1.1 令论域U是区间$[0,100]$，U的元素x代表人的年龄。这时，老年人的概念可表达为U的一个模糊集合A，其隶属函数可定义为

$$\mu_A(x)=\begin{cases}0 & 0\leqslant x\leqslant 50\\ \left[1+\left(\dfrac{x-50}{5}\right)^{-2}\right]^{-1} & 50\leqslant x\leqslant 100\end{cases}$$

从这个例子可以看到，年老这个词的意义可由年龄的论域中的一个模糊集合来表示。

就论域的类型而言，模糊集有下列两种表示法：

①设论域U为有限域，令$U=\{x_1,x_2,\cdots,x_n\}$，U上的任一模糊集A，其隶属函数$\mu_A(x_i)$，$i=1,2,\cdots,n$，则此时A可表示成

$$\begin{aligned}A&=\mu_A(x_1)/x_1+\mu_A(x_2)/x_2+\cdots+\mu_A(x_n)/x_n\\&=\sum_{i=1}^{n}\mu_A(x_i)\Big/x_i\end{aligned}$$