

## 目 录

|                   |      |
|-------------------|------|
| 第一章 计算机情报检索概述     | 1—1  |
| 1.1 情报与社会的发展      | 1—1  |
| 1.2 情报检索与文献检索     | 1—2  |
| 1.3 文献情报检索系统的基本功能 | 1—3  |
| 1.4 文献情报检索系统的基本原理 | 1—4  |
| 1.4.1 文献与文献标识     | 1—4  |
| 1.4.2 文献——语词矩阵    | 1—6  |
| 1.4.3 三种基本的文献检索方式 | 1—7  |
| 1.5 联机情报检索        | 1—10 |
| 1.6 关于本课程的说明      | 1—11 |
| 第二章 基于倒排档的检索系统    | 2—1  |
| 2.1 倒排档检索技术发展简史   | 2—1  |
| 2.2 布尔逻辑          | 2—5  |
| 2.3 典型的文档结构       | 2—7  |
| 2.4 检索过程          | 2—11 |
| 2.5 检索式的逻辑运算      | 2—12 |
| 2.5.1 运算顺序的正确控制   | 2—13 |
| 2.5.2 集合的逻辑运算     | 2—17 |

|                  |      |
|------------------|------|
| 2.6 倒排档检索机制的加强   | 2-19 |
| 2.6.1 邻接         | 2-19 |
| 2.6.2 截词         | 2-21 |
| 2.6.3 范围检索       | 2-21 |
| 2.6.4 加权         | 2-21 |
| 2.7 商业性检索系统介绍    | 2-22 |
| 2.7.1 DIALOG 系统  | 2-23 |
| 2.7.2 STAIRS 系统  | 2-24 |
| 2.7.3 MEDLARS 系统 | 2-29 |

|                      |      |
|----------------------|------|
| 第三章 文献情报检索的数据结构和检索技术 | 3-1  |
| 3.1 情报检索中的数据结构       | 3-1  |
| 3.1.1 逻辑结构与物理结构      | 3-1  |
| 3.1.2 线性表            | 3-5  |
| 3.1.3 树              | 3-8  |
| 3.1.4 图              | 3-13 |
| 3.2 查找技术             | 3-14 |
| 3.2.1 顺序查找           | 3-15 |
| 3.2.2 基于索引的方法        | 3-17 |
| 3.2.2.1 二分法查找        | 3-18 |
| 3.2.2.2 分块查找法        | 3-20 |
| 3.2.2.3 索引顺序法        | 3-23 |
| 3.2.2.4 B-树          | 3-28 |
| 3.2.3 基于 Hash 的查找方法  | 3-29 |
| 3.2.3.1 碰撞问题及其解决     | 3-30 |

|                    |      |
|--------------------|------|
| 3.2.3.2 截词检索       | 3—34 |
| 3.2.3.3 Hash法与情报检索 | 3—36 |

|                      |      |
|----------------------|------|
| 第四章 检索效果及其改善         | 4—1  |
| 4.1 检索效果及其测量指标       | 4—1  |
| 4.2 影响检索效果的主要因素      | 4—5  |
| 4.2.1 情报提问对情报需求的表达程度 | 4—6  |
| 4.2.2 数据库的选择和比较      | 4—8  |
| 4.2.3 检索途径的选择        | 4—9  |
| 4.2.4 检索词的选择与调节      | 4—9  |
| 4.2.5 检索式的结构         | 4—11 |
| 4.3 提高检索效果的反馈调整方法    | 4—12 |
| 4.3.1 反馈调整在检索过程中的作用  | 4—12 |
| 4.3.2 调节检索策略的若干方法    | 4—15 |

|                |      |
|----------------|------|
| 第五章 自动标引       | 5—1  |
| 5.1 自动标引和人工标引  | 5—1  |
| 5.2 西文自动标引方案简介 | 5—3  |
| 5.2.1 词频统计原理   | 5—3  |
| 5.2.2 逆文献频率法   | 5—6  |
| 5.2.3 信号——噪音法  | 5—7  |
| 5.2.4 词辨别值法    | 5—10 |
| 5.2.5 词短语的构造   | 5—16 |
| 5.3 自动标引中的词表   | 5—18 |

张庆国

|                              |      |
|------------------------------|------|
| 第六章 聚类检索                     | 6—1  |
| 6.1 问题的提出                    | 6—1  |
| 6.2 SMART 系统                 | 6—1  |
| 6.2.1 文献的向量表示和匹配度计算          | 6—2  |
| 6.2.2 聚类文件的生成和 SMART 系统的文档结构 | 6—4  |
| 6.2.3 提问式的反馈调整               | 6—10 |
| 6.2.4 动态文献空间                 | 6—14 |
| 6.2.5 聚类检索和分类检索的区别           | 6—15 |
| 6.3 倒排检索和聚类检索的结合             | 6—16 |
| 6.3.1 SIRE 系统                | 6—16 |
| 6.3.2 加权的布尔检索                | 6—20 |

|                         |      |
|-------------------------|------|
| 第七章 检索效果的改善(续)          | 7—1  |
| 7.1 文献——语词矩阵的若干推论       | 7—1  |
| 7.1.1 词联接矩阵             | 7—1  |
| 7.1.2 词结合矩阵和改良型文献——语词矩阵 | 7—2  |
| 7.2 与词结合矩阵相关的权和提问向量     | 7—4  |
| 7.3 通过结合反馈进行的提问自动修正     | 7—8  |
| 7.4 检索策略的最优化            | 7—11 |

|                |     |
|----------------|-----|
| 第八章 数据检索系统     | 8—1 |
| 8.1 概论         | 8—1 |
| 8.2 数据库管理系统的结构 | 8—4 |
| 8.2.1 信息项的结构   | 8—4 |
| 8.2.2 关系数据库模式  | 8—8 |

|                      |          |
|----------------------|----------|
| 8.2.3 层次数据库模式        | 8—13     |
| 8.2.4 网络数据库模式        | 8—18     |
| 8.3 查询和查询语言          | 8—19     |
| 8.3.1 分步法            | 8—21     |
| 8.3.2 “菜单”方法         | 8—22     |
| 8.3.3 表查询法           | 8—23     |
| 8.3.4 例举查询           | 8—24     |
| <br>第九章 事实检索         | <br>9—1  |
| 9.1 事实检索和自然语言处理      | 9—2      |
| 9.2 自然语言处理的句法分析系统    | 9—3      |
| 9.2.1 自然语言的处理层次      | 9—3      |
| 9.2.2 短语结构语法         | 9—4      |
| 9.2.3 转换语法           | 9—9      |
| 9.2.4 扩充转换网络语法       | 9—12     |
| 9.3 知识的表示            | 9—18     |
| 9.4 目前水平上的事实检索系统     | 9—23     |
| <br>第十章 情报信息的存贮和输入输出 | <br>10—1 |
| 10.1 数据标识的代码化        | 10—1     |
| 10.2 数据库的存贮载体        | 10—1     |
| 10.2.1 磁带数据库         | 10—5     |
| 10.2.2 磁盘数据库         | 10—7     |
| 10.2.3 其他存贮设备        | 10—8     |
| 10.3 情报资料的输入手段       |          |

|             |           |       |
|-------------|-----------|-------|
| 10. 3. 1    | 键到纸介质方式   | 10—8  |
| 10. 3. 2    | 键到磁介质方式   | 10—9  |
| 10. 3. 3    | 联机终端输入方式  | 10—10 |
| 10. 3. 4    | 全自动字符识别方式 | 10—11 |
| 10. 3. 4. 1 | 光学字符识别法   | 10—11 |
| 10. 3. 4. 2 | 光学标记阅读装置  | 10—14 |
| 10. 4       | 情报资料的输出手段 | 10—15 |
| 结 语         |           | 10—17 |

## 第七章 检索效果的改善(续)

在第四章中,我们讨论了检索效果及其一些定性的改善方法。在那之后,我们又几次涉及了检索效果的改善问题。本章我们将再介绍一些改善检索效果的定量的技术方法。

### 7.1 文献—语词矩阵的若干推论

#### 7.1.1 词联接矩阵

设检索系统中有文献—语词矩阵

$$D = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{pmatrix} \quad (7.1)$$

其中  $a_{ij}$  表示第  $j$  个标引词对第  $i$  篇文献的重要程度,即权。 $a_{ij}$  的一种有用的取值方法为取离散值 0 和 1。

我们给出矩阵  $D$  的转置  $D^T$

$$D^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{N1} \\ a_{12} & a_{22} & \cdots & a_{N2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1M} & a_{2M} & \cdots & a_{NM} \end{pmatrix} \quad (7.2)$$

再令

$$K = D^T D = (k_{ij})_{MM} \quad (7.3)$$

则  $K$  是  $M$  行  $M$  列的一个方阵。我们在 5.3 中已经引出了这个矩阵,并且把它称为词联接矩阵。

## 7. 1. 2 词结合矩阵和改良型文献—语词矩阵

词联接矩阵用来确定同一文献的标引词对，而这样的标引词对中的词无须在词义上相关。事实上，意义相近的词一般不全是同一文献的标引词，不然的话，则表明标引中有冗余的成分了。

列检索提问式的人可能使用与文献标引不同，但意义相近的词。为了满足这种要求，我们引入一个词的结合矩阵  $T$ ，矩阵  $T$  的元素  $t_{ij}$  表示在什么程度上适用于第  $i$  个标引词的文献同时也适用于第  $j$  个标引词。

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1M} \\ t_{21} & t_{22} & \cdots & t_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ t_{M1} & t_{M2} & \cdots & t_{MM} \end{pmatrix} \quad (7.4)$$

例如，五个标引词

波、速度、海、洋、风

的词结合矩阵可表示如下：

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 1 \end{pmatrix}$$

这个词结合矩阵表明，第三个标引词“海”和第四个标引词“洋”被认为在标引文献的意义上是等价的 ( $t_{34} = t_{43} = 1$ )； $t_{51} = \frac{1}{2}$  表明用第五个标引词“风”标引的文献中有一些但不是全部也适用于第一个词“波”； $t_{15} = 0$  表示用“波”标引的文献被认为并不一定与“风”相关。



需要指出的是，词结合矩阵不是对称的，如本例中的  $t_{51} \neq$

$t_{15}$ 。

引入词结合矩阵后，我们可以导出改良型的文献—语词矩阵

$$DT = (d_{ij})_{NM} \quad (7.5)$$

DT 中的元素

$$\begin{aligned} d_{ij} &= a_{1j}t_{1j} + a_{12}t_{2j} + \dots + a_{1M}t_{Mj} \\ &= \sum_{k=1}^M a_{ik}t_{kj} \end{aligned} \quad (7.6)$$

我们考察  $d_{ij}$  的意义。在定义  $d_{ij}$  的式 (7.6) 中，第一项中的  $a_{1j}$  表示文献 1 与第 1 个标引词的相关程度， $t_{1j}$  表示第 1 个标引词与第 j 个标引词的相近程度，因此， $a_{1j}t_{1j}$  表示以第一个标引词作为中介时第 1 篇文献与第 j 个标引词的相关程度。其它各项分别是以第二个词、第三个词、...、第 M 个词作为中介的。于是我们看到， $d_{ij}$  表示在考虑了系统的所有词间关系后，第 1 篇文献和第 j 个标引词的相关程度，或权。

例如，用上例中的五个标引词标引了六篇文献。文献—语词矩

阵 D 如下

$$D = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

其改良型文献—语词矩阵

$$DT = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ \frac{1}{2} & 1 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & 1 & 1 & 1 \\ \frac{1}{2} & 0 & 1 & 1 & 1 \end{pmatrix}$$

· 我们通过第五篇文献来比较  $D$  和  $DT$ 。在  $D$  中,  $a_{54}=0$ , 这表明第五篇文献与第四个标引词“洋”并没有直接标引关系。但在词结合矩阵  $T$  中我们看到, “洋”和第三个标引词“海”在标引文献的意义上是等价的 ( $t_{34}=1$ )。而“海”又用来标引了这第五篇文献。因此第四个标引词“洋”也是适用于第五篇文献的 ( $a_{54}=1$ )。又虽然  $a_{51}=0$ , 但  $a_{55}=1$ ,  $t_{51}=\frac{1}{2}$ , 所以  $a_{51}=\frac{1}{2}$ , 即标引词“风”把它对“波”的关系传递给了第五篇文献。 $a_{51}=\frac{1}{2}$  表示第五篇文献在较小的程度上也适用于“波”。

## 7.2 与词结合矩阵相关的权和响应向量

设提问向量

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_M \end{pmatrix} \quad (7.6)$$

其中  $q_j$  是用户请求中第  $j$  个词所带的权。 $Q$  是  $M$  维空间中的一个向量,  $q_j$  是它的分量。

在文献数据库中, 每篇文献均用一文献向量描述。对于以式 (7.6) 的形式规定的请求, 第 1 篇文献的匹配度值可用公式

$$r_1 = \sum_{j=1}^M a_{1j} q_j \quad (7.7)$$

来计算。当  $l$  取所有值时，我们得到一个对提问 (7.6) 的响应向量

$$R = DQ = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix} \quad (7.8)$$

响应向量中的非零分量表示用一个或多个标引词标引的文献。而这些标引词在请求向量中对应非零权。

如果用词结合矩阵  $T$  表示标引词之间的联系, 那么文献—语词矩阵  $D$  应该用其改良型的  $DT$  代替, 式 (7.8) 成为

$$R = DTQ \quad (7.9)$$

词结合矩阵  $T$  和提问向量  $Q$  的乘积

[illegible]

是一个向量。我们通过它的一个分量来看这个向量的意义。

TQ 的一个分量是

$$t_{11}q_1 + t_{12}q_2 + \dots + t_{1M}q_M$$

与原提问向量中的相应分量  $q_i$  比较, 有

$$t_1, a_1 + t_{1,2} a_2 + \dots + t_{1,M} a_M$$

$$= t_{11}q_1 + (t_{12}q_2 + \dots + t_{1, i-1}q_{i-1} + t_{1, i+1}q_{i+1} + \dots + t_{1M}q_M)$$

$$= q_1 + (t_{12}q_2 + \dots + t_{1, i-1}q_{i-1} + t_{1, i+1}q_{i+1} + \dots + t_{1M}q_M)$$

$$> q_1$$

当上式中的等号成立时，说明  $t_{1j} = 0 (i \neq j)$ ，因而第 1 个标引词与其标引词无关；若第 1 个标引词与其它标引词相关，则至少有一个  $t_{1j} \neq 0 (i \neq j)$ ，这时上式中的不等号成立。于是，TQ 中的各分量是原提问向量 Q 中各相应分量经过词结合矩阵 T 修正后的值。

我们再用 M 表示屏蔽运算，它把小于某给定阈值（如 1）的所有分量都化为零，并用 1 取代其它的所有分量。那么 MTQ 就是一个向量  $Q^{(1)}$ ，其中值为 1 的分量表示被看作是和提问向量 Q 的分量等价或非常相关的标引词。这样的标引词被称为第一代的词。

例如，设提问向量 Q 和词结合矩阵 T 如下：

$$Q = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 1 \end{pmatrix}$$

于是有

$$TQ = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ \frac{1}{2} \end{pmatrix}$$

指定  $M$  的屏蔽级为 1, 则有

$$Q^{(1)} = MTQ = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

其中第一代词为与分量  $q_1^{(1)}$ ,  $q_3^{(1)}$  和  $q_4^{(1)}$  相应的标引词。

仿照从  $Q$  中导出  $Q^{(1)}$  的方式, 可以从  $Q^{(1)}$  中导出

$$Q^{(2)} = MTQ^{(1)} = (MT)^2 Q$$

$$Q^{(3)} = (MT)^3 Q$$

.....

等等, 与  $Q^{(2)}$ ,  $Q^{(3)}$ , ... 中非零分量对应的词分别称为第二代词。

第三代词, ... 等等。如果用  $S_1$  表示第 1 代词的集合, 则显然有

$$S_0 \subset S_1 \subset S_2 \subset \dots$$

其中  $S_0$  表示与原提问向量  $Q$  中的非零分量对应的词。它们可称作第零代词, 即提问词。

提问词, 第一代词, 第二代词及更多代词的全体与向量

$$Q + MTQ + (MT)^2 Q + \dots \quad (7.10)$$

的非零分量相对应。上式也可写成以下形式

$$(I - MT)^{-1} Q \quad (7.11)$$

其中  $I$  代表单位矩阵

第  $1 + l$  代词的个数等于或大于第  $l$  代词的个数。多出来的这些词是我们通过词结合矩阵扩展出来的。对这些扩展出来的词应该加以限制, 而不能把它们与上一代甚至上几代的词相混看待。限制是通过加权实现的, 即连续几代的词的权递减。前面的序列 (7.9)

可改写为

$$Q + \lambda MTQ + (\lambda MT)^2 Q + \dots \quad (7.12)$$

或

$$(I - \lambda MT)^{-1} Q \quad (7.13)$$

我们把

$$(I - \lambda MT)^{-1}$$

称为扩展因子。使用扩展因子后，提问向量由  $Q$  变成  $(I - \lambda MT)^{-1} Q$ 。

于是响应向量的形式 (7.8) 变成

$$R = D(I - \lambda MT)^{-1} Q$$

### 7.3 通过结合反馈进行的提问自动修正

我们考虑一个检索输出的  $n$  篇文献的文献—语句矩阵  $D$  和它的转置  $D^T$ 。

$$D = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

$$D^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix}$$

其中  $m$  表示这  $n$  篇输出文献涉及的标引词个数。在矩阵  $D^T$ ，每一行表示一个标引词与这  $n$  篇输出文献的各自适用程度。

响应向量

$$R = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

其中分量  $r_1$  表示第 1 篇输出文献与用户提问的匹配程度。因此，向量  $D^T R$  的第 1 个分量

$$\sum_{t=1}^n a_{t1} r_t = a_{11} r_1 + a_{21} r_2 + \dots + a_{n1} r_n \quad (7.15)$$

表示就这  $n$  篇输出文献而言，第 1 个标引词与用户提问的适用程度。这个值应该被加进原提问  $Q$  中相应的分量值  $q_1$  中去。即

$$q'_1 = q_1 + \sum_{t=1}^n a_{t1} r_t \quad (7.16)$$

当  $i$  取遍所有  $m$  个标引词时，我们便有修改后的提问向量

$$Q' = Q + D^T R = \begin{pmatrix} q_1 + \sum_{t=1}^n a_{t1} r_t \\ q_2 + \sum_{t=1}^n a_{t2} r_t \\ \dots\dots\dots \\ q_m + \sum_{t=1}^n a_{tm} r_t \end{pmatrix} \quad (7.17)$$

这个新的提问向量不仅取决于原来的请求  $Q$ ，而且还取决于所产生的输出  $R$ 。修正后的请求取决于输出的程度可以通过引入参数  $\lambda$  控制。即令

$$Q' = Q + \lambda D^T R \quad (7.18)$$

下图是根据上面的反馈形式处理向量  $Q$  的情况：

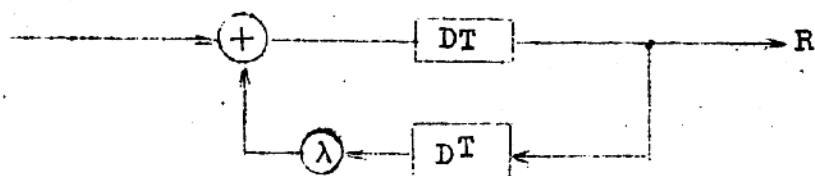


图 7.1 通过输出反馈自动调节用户请求

用修改后的提问向量  $Q + \lambda D^T R$  代替  $Q$  之后, 式 (7. 9) 便成为

$$R = DT(Q + \lambda D^T R) \quad (7. 19)$$

解出  $R$

$$\begin{aligned} R &= (I - \lambda DTD^T)^{-1} DTQ \\ &= [I + \lambda DTD^T + (\lambda DTD^T)^2 + (\lambda DTD^T)^3 + \dots] DTQ \end{aligned} \quad (7. 20)$$

将上式代入式 (7. 18), 则我们得到反馈检索的提问形式

$$\begin{aligned} Q' &= Q + \lambda D^T R \\ &= Q + \lambda D^T [I + \lambda DTD^T + (\lambda DTD^T)^2 + \dots] DTQ \\ &= Q + \lambda D^T DTQ + \lambda^2 D^T DTD^T DTQ + \dots \\ &= [I + \lambda D^T DT + (\lambda D^T DT)^2 + \dots] Q \\ &= (I - \lambda D^T DT)^{-1} Q \\ &= (I - \lambda KT)^{-1} Q \end{aligned} \quad (7. 21)$$

式中  $K$  是词联接矩阵

检索的实际过程如下: 设检索从一个请求向量  $Q$  开始, 并指定需要一个给定值数量的命中文献, 如  $n_0$  篇。开始令  $\lambda = 0$ , 按式 (7. 21) 给出的提问向量检索。每次检索后, 得到一响应向量, 其中的非零分量表示输出的命中文献。如果检出文献少于  $n_0$  篇, 或感觉到查全率过低, 则给定一个较小的正数  $\lambda$ , 并按式 (7. 21) 重新确定提问向量并再次检索。如检索结果仍小于  $n_0$  篇或查全率还是低, 则逐渐增大  $\lambda$  的值反复进行上述过程, 直至输出的文献数达到用户要求的  $n_0$  篇, 或用户对查全率基本满意为止。上述增大  $\lambda$  值的过程可以是自动的。

每次反馈时, 提问向量可能增加一些非零分量, 这意味着下次检索将增加一些检索词, 即扩大了请求。我们注意到, 扩大请求的



方式是由内部标引词的词间关系而不是由用户的兴趣控制的。如果用户参预控制扩大请求的过程, 则效果会更好一些。用户参预控制的方法是: 每次  $\lambda$  值变化后, 打印输出与新的提问向量中非零分量相应的词, 以及由式 (7.15) 确定的权。这样, 用户便能知道提问向量  $Q$  的扩展情况。如果他认为新增加的词与原提问是无关的, 那么可删掉这个词。

在本节的讨论中我们可以看到, 提问的修正方法并非直接出于优化查全率和查准率的企图。它受数据库特性的影响要大于受用户要求的影响。在下一节中, 我们将介绍一种直接从提高查全率和查准率的角度考虑的检索提问式修正方法。

## 7.4 检索策略的最优化

设用户的提问向量为

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix}$$

我们已有响应向量

$$R = DQ = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

其中

$$r_i = \sum_{j=1}^m a_{ij} q_j \quad (i=1, 2, \dots, n)$$

我们称  $r_i$  为文献  $i$  的响应函数。