中文信息检索中的相关反馈

Relevance Feedback in Information Retrieval

战学刚 林鸿飞 姚天顺

(东北大学计算机系 沈阳 110006)

Abstract Relevance feedback is a query modification technique used in information retrieval to produce improved result. Over the past 30 years, various methods have been proposed to formulate improved queries by relevance feedback. This paper introduces a query formulation method and shows that this method is better than classic query formulation methods.

Keywords Information retrieval, Query formulation, Relevance feedback

一、引言

对于基于统计的信息检索系统,影响其性能的主要环节有:1. 特征项的选择、2. 权重的计算方法、3. 查询的表示形式、4. 查询的调整(修改)、5. 相似度的计算方法。当系统确定了其索引形式和相似度的计算方法后,系统性能的提高,主要依赖于查询的构造和调整方法,本文主要介绍如何通过相关反馈来修改查询,以提高检索系统的性能。

在信息检索系统中精确地构造用户查询是非常困难的。在绝大多数系统中,索引和检索过程对于用户来说是不透明的,所以,用户很难根据自己的信息需求构造精确的查询。为了缓解这一困难,提高系统性能,人们提出了相关反馈技术,所谓相关反馈,指的是系统根据初始用户查询,检索到一组样本文献,然后,根据用户在样本文献中的选择,构造出改进的查询,并据此再次进行检索。把新的检索结果作为新的样本文献,这一过程可以循环进行。随着样本文献数目的递增,我们最终可望获得较为精确的查询,并据此得出较好的检索结果。

相关反馈技术的发展已经有30多年的历史,在信息检索的研究和发展过程中,人们针对布尔模型、向量空间模型和概率模型对相关反馈技术进行了广泛的研究,实验评估结果清楚表明,相关反馈技术是一种非常有效的提高系统性能的方法,但是,在传统的信息反馈方法中,也存在着严重的缺点。例如,在概率模型中应用相关反馈时,文献只能被表示成二元向量形式,尽管有人曾经尝试取消这一限制。另一方面,特征项的权值可以很容易地加入到向量空间模型中,但是,在基于向量的相关反馈模型中,并不存在确定某些参数的系统化方法。本节后面的描述便是这方面的一个实例。

Maron 和 kuhns 在1960年指出[1],与原始查询相近的特征项可以加入到查询中,以便检索到更多的相关文献。在早期的 Smart 系统中,也曾做过许多关于相关反馈的实验[2]。1965年 Rocchio 发表了他在查询修改方面的实验结果[2],他将特征项重新加权和查询扩展结合起来,并基于向量空间模型定义了查询修改方法:

$$Q_1 = Q_0 + \frac{1}{n_1}\sum_{i=1}^{n_1} R_i - \frac{1}{n_2}\sum_{i=1}^{n_2} S_i$$

其中:Q_0=原始查询的特征向量;R_i=相关文献 i 的特征向量;S_i=不相关文献 i 的特征向量;n_1=相关文献数;n_2=不相关文献数,因此 Q_1是原始查询的特征向量、相关文献 i 的特征向量和不相关文献 i 的特征向量的向量和,实验表明,对于向量空间模型,Ide 提出的 dec-hi 方法是较好的通用相关反馈技术[3]。dec-hi 方法所采用的查询修改方法如下:

$$Q_1 = Q_0 - \sum_{i=1}^{n_1} R_i - S$$

其中:Q_0=原始查询的特征向量;R_i=相关文献 i 的特征向量;S=最不相关文献的特征向量。

Rocchio 和 Ide 所提出的方法,都是下面公式的特殊形式:

$$Q_1 = Q_0 + \beta\sum_{i=1}^{n_1} R_i - \gamma\sum_{i=1}^{n_2} S_i$$

其中,β 和 γ 是可以由系统设计人员调整的参数。然而,迄今为止,并不存在确定这些参数的系统化方法。Rocchio 和 Ide 只是给出了两种具体的参数形式。

Wong 等人在80年代末期,提出了一种基于用户对某些已检索到的文献对的优先判定关系的适应性查询构造方法[4-6]。它对于文献的表示形式没有限制(文献既可以用二元表示形式,也可以用加权表示形

式1。这种方法的显著特点是，在给定用户对样本文献的优先判定关系时，可以用迭代算法生成新的查询，而无须引用特定的公式或参数。这种迭代算法是否收敛取决于用户对样本文献的优先判定关系结构。

上述各种查询构造方法都是针对英文检索系统的，由于英文单词有词形变化，在英文信息检索系统中，通过词干提取(Stemming)把同根词聚合成一个特征项来表示；而中文词汇的提取则是通过分词过程来完成的。因此，中、英文检索系统在特征项的构成方面是有差异的。我们利用一个基于向量空间模型的检索系统 Infolite 对几种主要的相关反馈方法进行了对比和实验。结果表明，适应性查询构造方法优于其它方法。

二、适应性查询构造方法

对于给定的文献集 **D**，用户优先关系是一个 **D** 上的二元关系$\prec\cdot$定义为：

$\mathbf{d}\prec\cdot\mathbf{d}'$当且仅当用户认为$\mathbf{d}'$优先于 d。

我们假设每个 $d\in D$ 都是由一个 n 维向量表示的，$\mathbf{d}=(d_1,d_2,\cdots,d_n)$。从相关排序的角度来看，我们的主要目标是找到一个文献向量集 **D** 上的保序函数 f，使得当$\mathbf{d}'$优先于 **d** 时，有：

$$\mathbf{d}\prec\cdot\mathbf{d}'\rightarrow f(\mathbf{d})<f(\mathbf{d}'), \tag{1}$$

满足(1)的函数保证了较为优先的文献在文献的相关排序中排列在较前的位置。换句话说，条件(1)蕴涵：

$$f(\mathbf{d})\geqslant f(\mathbf{d}')\rightarrow\neg(\mathbf{d}\prec\cdot\mathbf{d}'). \tag{2}$$

我们称上述的函数 f 为可接受的相关排序函数。f 被定义为：

$$f(\mathbf{d})=\sum_{i=1}^{n}w_i d_i=\mathbf{q}\cdot\mathbf{d} \tag{3}$$

其中，$\mathbf{q}=(w_1,w_2,\cdots,w_n)$是一个 n 维向量。优先关系$\prec\cdot$被称为**弱线性**的，如果存在上述的线性函数 f，并且对于任意的$\mathbf{d},\mathbf{d}'\in\mathbf{D}$，有：

$$\mathbf{d}\prec\cdot\mathbf{d}'\rightarrow f(\mathbf{d})<f(\mathbf{d}')=\mathbf{q}\cdot\mathbf{d}<\mathbf{q}\cdot\mathbf{d}' \tag{4}$$

其中 **q** 称为查询向量。

对于每个满足优先关系$\mathbf{d}\prec\cdot\mathbf{d}'$的文献对$\mathbf{d},\mathbf{d}'\in\mathbf{D}$，我们可以定义一个差向量$\mathbf{b}=\mathbf{d}'-\mathbf{d}$。

设：

$$\mathbf{B}=\{\mathbf{b}=\mathbf{d}'-\mathbf{d}\mid\mathbf{d},\mathbf{d}'\in\mathbf{D}\text{ 且 }\mathbf{d}\prec\cdot\mathbf{d}'\}。 \tag{5}$$

根据定义，如果优先关系是弱线性的，$\mathbf{d}\prec\cdot\mathbf{d}'$蕴涵$\mathbf{q}\cdot\mathbf{d}<\mathbf{q}\cdot\mathbf{d}'$，即$\mathbf{q}\cdot\mathbf{b}>0$。这说明为每个满足$\mathbf{d}\prec\cdot\mathbf{d}'$的文献对找到满足(4)的查询向量 **q**，等价于求解下列的不等式：

$$\text{对于所有的 }\mathbf{b}\in\mathbf{B},\mathbf{q}\cdot\mathbf{b}>0 \tag{6}$$

每个差向量$\mathbf{b}\in\mathbf{B}$定义了一个超平面，$\mathbf{q}\cdot\mathbf{b}=0$，并以 **b** 作为法向量。该平面穿过原点，将 n 维空间分为两部分，其中在法向量一侧的部分称为超平面的正向。显然，对于超平面$\mathbf{q}\cdot\mathbf{b}=0$正向的任意向量 **q** 都满足$\mathbf{q}\cdot\mathbf{b}>0$，如图1所示。对于任意两个差向量$\mathbf{b},\mathbf{b}'\in\mathbf{B}$和任意查询向量 **q**，位于其对应超平面正向重叠区域的任意向量 **q** 都满足$\mathbf{q}\cdot\mathbf{b}>0$，且$\mathbf{q}\cdot\mathbf{b}'>0$。一般地，任何位于所有由 **B** 中的向量所定义的超平面正向重叠区域的任意向量 **q** 都满足不等式(6)。这一区域成为解区域。解区域中的任何向量均称为解向量。实际上，解区域$\mathbf{C_D}$是一个棱锥形区域，即，

1)对于任意 $k>0,\mathbf{q}\in\mathbf{C_D}\rightarrow k\mathbf{q}\in\mathbf{C_D}$，

2)$\mathbf{q},\mathbf{q}'\in\mathbf{C_D}\rightarrow(\mathbf{q}+\mathbf{q}')\in\mathbf{C_D}$。

注意，由于 $k>0$，所以$\mathbf{C_D}$不是子向量空间。

下面我们将给出一个从样本文献集 **S**(**S** 是 **D** 的子集)中求出解向量的学习算法[5]，在本文以后的描述中，**B** 表示差向量的集合，其定义如下：

$$\mathbf{B}=\{\mathbf{b}=\mathbf{d}'-\mathbf{d}\mid\mathbf{d},\mathbf{d}'\in\mathbf{S}\text{ 且 }\mathbf{d}\prec\cdot\mathbf{d}'\} \tag{7}$$

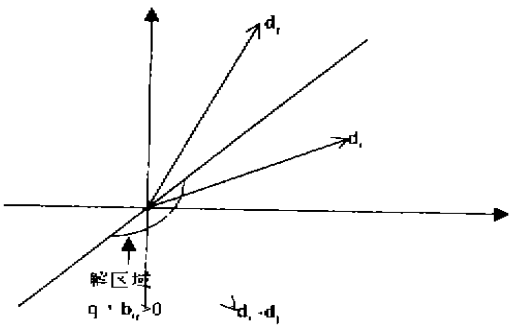

图1　解区域图示

算法1

1. 选定初始查询向量 $\mathbf{q}_0$，并置 $k=0$。

2. 设 $\mathbf{q}_k$ 是第 $k+1$ 次迭代的查询向量，从差向量集合 **B** 中，选出下列的差向量子集，

$$\Gamma(\mathbf{q}_k)=\{\mathbf{b}=\mathbf{d}'-\mathbf{d}\mid\mathbf{d},\mathbf{d}'\in\mathbf{S},\mathbf{d}\prec\cdot\mathbf{d}',\text{且 }\mathbf{q}_k\cdot\mathbf{b}\leqslant 0\} \tag{8}$$

如果$\Gamma(\mathbf{q}_k)=\Phi$，则停止迭代，$\mathbf{q}_k$ 即是一个解向量。

3. 置：$\mathbf{q}_{k+1}=\mathbf{q}_k+\sum_{\mathbf{b}\in\Gamma(\mathbf{q}_k)}\mathbf{b}$。

4. 置 $k=k+1$；转第2步。

以上过程可能在解向量中产生负分量。在实际应用中，负分量是不允许的，在某些相关反馈算法中，它们通常被用0值代替。但用0值代替负值可能使一个解向量变为非解向量。因此，我们允许负分量存在，事实上，在概率模型中也使用负分量。

下面我们用一个实例来说明上述算法。

例：设有一个文献向量集$\mathbf{D}=\{\mathbf{d}_1,\mathbf{d}_2,\mathbf{d}_3,\mathbf{d}_4\}$，其中：

$\mathbf{d}_1=(0.1,1.0)$，$\mathbf{d}_2=(0.7,1.0)$，

$\mathbf{d}_3=(1.0,0.2)$，$\mathbf{d}_4=(1.0,0.8)$

并假设在 D 上定义了优先关系：

$\mathbf{d}_1<\cdot\mathbf{d}_2$，$\mathbf{d}_1<\cdot\mathbf{d}_3$，$\mathbf{d}_2<\cdot\mathbf{d}_3$，$\mathbf{d}_1<\cdot\mathbf{d}$

基于这一优先关系，根据公式(7)我们可以得到如下的差向量集合：

$\mathbf{b}_{11}=\mathbf{d}_1$，$\mathbf{d}_1=(0.7,1.0)-(0.0,1.0)=(0.7,0.0)$

$\mathbf{b}_{12}=\mathbf{d}_2-\mathbf{d}_1=(0.7,1.0)-(1.0,0.0)=(-0.3,1.0)$

$\mathbf{b}_{31}=\mathbf{d}_3-\mathbf{d}=(1.0,0.2)-(0.0,1.0)=(1.0,-0.8)$

$\mathbf{b}_{24}=\mathbf{d}_4-\mathbf{d}_1=(1.0,0.2)-(0.1,1.0)=(0.9,0.2)$

表1给出了根据3个不同的初始查询向量构造新的查询向量的详细步骤，从而得到了3个解向量。

表1　三个初始查询向量的查询构造过程

代次数	$\mathbf{q}_0=(0.3,1.0)$	$\mathbf{q}_0=(1.0,0.1)$	$\mathbf{q}_0=(1.0,1.0)$
1	$\Gamma(\mathbf{q}_0)=\{\mathbf{b}_{11}\}$ $\mathbf{q}_1=\mathbf{q}_0+\mathbf{b}_{11}=(1.3,0.2)$	$\Gamma(\mathbf{q}_0)=\{\mathbf{b}_{21}\}$ $\mathbf{q}_1=\mathbf{q}_0+\mathbf{b}_{21}=(0.7,1.1)$	$\Gamma(\mathbf{q}_0)=\mathbf{B}$ $\mathbf{q}_1=\mathbf{q}_0+\Sigma\mathbf{b}=(1.4,0.4)$
2	$\Gamma(\mathbf{q}_1)=\{\mathbf{b}_{24}\}$ $\mathbf{q}_2=\mathbf{q}_1+\mathbf{b}_{24}=(1.0,1.4)$	$\Gamma(\mathbf{q}_1)=\{\mathbf{b}_{11}\}$ $\mathbf{q}_2=\mathbf{q}_1-\mathbf{b}_{11}=(1.7,0.3)$	$\Gamma(\mathbf{q}_1)=\{\mathbf{b}_{24}\}$ $\mathbf{q}_2=\mathbf{q}_1+\mathbf{b}_{24}=(2.1,1.4)$
3	$\Gamma(\mathbf{q}_2)=\Phi$ 解 $(1.0,1.2)$	$\Gamma(\mathbf{q}_2)=\{\mathbf{b}_{24}\}$ $\mathbf{q}_3=\mathbf{q}_2+\mathbf{b}_{24}=(1.4,1.3)$	$\Gamma(\mathbf{q}_2)=\{\mathbf{b}_{11}\}$ $\mathbf{q}_3=\mathbf{q}_2+\mathbf{b}_{11}=(2.1,0.6)$
4		$\Gamma(\mathbf{q}_3)=\Phi$ 解 $(1.4,1.3)$	$\Gamma(\mathbf{q}_3)=\{\mathbf{b}_{24}\}$ $\mathbf{q}_4=\mathbf{q}_3+\mathbf{b}_{24}=(1.8,1.6)$
5			$\Gamma(\mathbf{q}_4)=\Phi$ 解 $(1.8,1.6)$

下面给出了上表中第一种情形的详细说明。初始向量 $\mathbf{q}_0$ 位于(0.3,1.0)。容易验证，$\mathbf{q}_0$ 满足条件 $\mathbf{q}_0\cdot\mathbf{d}_1<\mathbf{q}_0\cdot\mathbf{d}_2$，$\mathbf{q}_0\cdot\mathbf{d}_4<\mathbf{q}_0\cdot\mathbf{d}_2$，及 $\mathbf{q}_0\cdot\mathbf{d}_4<\mathbf{q}_0\cdot\mathbf{d}_3$。而 $\mathbf{q}_0$ 却显然不满足 $\mathbf{q}_0\cdot\mathbf{d}_1<\mathbf{q}_0\cdot\mathbf{d}_3$。从几何上看，这意味着与 $\mathbf{d}_3$ 相比，$\mathbf{q}_0$ 更接近于 $\mathbf{d}_1$。通过 $\mathbf{q}_1=\mathbf{q}_0+\mathbf{b}_{31}=\mathbf{q}_0+(\mathbf{d}_3-\mathbf{d}_1)=(1.3,0.2)$ 变换之后，$\mathbf{q}_1$ 已接近于 $\mathbf{d}_3$。此时，$\mathbf{q}_1$ 满足条件 $\mathbf{d}_1<\cdot\mathbf{d}_1$，却不满足 $\mathbf{d}_4<\cdot\mathbf{d}_2$。下一步的变换 $\mathbf{q}_2=\mathbf{q}_1-\mathbf{b}_{24}=\mathbf{q}_1+(\mathbf{d}_2-\mathbf{d}_4)$ 所生成的向量 $\mathbf{q}_2$ 位于解区域，即 $\mathbf{q}_2$ 满足所有由用户所定义的优先关系：

$\mathbf{q}_2\cdot\mathbf{d}_1<\mathbf{q}_2\cdot\mathbf{d}_2$，　$\mathbf{q}_2\cdot\mathbf{d}_1<\mathbf{q}_2\cdot\mathbf{d}_3$，

$\mathbf{q}_2\cdot\mathbf{d}_4<\mathbf{q}_2\cdot\mathbf{d}_3$　　$\mathbf{q}_2\cdot\mathbf{d}_4<\mathbf{q}_2\cdot\mathbf{d}_2$。

在此需要着重指出的是，如果我们从空向量开始求解，则在第一次迭代时所得到的向量即对应于 Rocchio 的最优查询[2]，这说明 Rocchio 的最优查询并不一定是满足所有优先关系的解向量。

还应指出的是，上述迭代算法并不一定收敛，即不一定在有限步骤内结束。Wong 等人证明了如果用户优先关系是弱线性的，则无论怎样选取初始向量，算法1都将在有限步骤内计算解向量[5]。

三、实验结果

我们利用 Infolite 检索系统对几种主要的相关反馈方法进行了对比和实验。Infolite 是一个基于向量空间模型的实验系统。Infolite 的主要资源包括 ①基于向量空间模型的检索引擎；②分词词典(含72,000条词目，其中名词21,931条)；③特征项表(含4,096项，覆盖分词词典中15,000条名词词目)。其中，特征项表是根据《同义词词林》和《中国分类主题词表》从分词词典中抽取出来的。

测试语料包括《人民日报》电子版、新加坡《联合早报》网络版、其它网上新闻(含大陆、港、台的中文新闻出版物)。

评价检索系统的两个主要指标是查准率(precision)和查全率(recall)。当用户将自己的查询提交给系统后，整个文献库在逻辑上划分为四部分：检索到的相关文献、检索到的不相关文献、未检索到的相关文献、未检索到的不相关文献。

文献库中包含与用户需求相关的和不相关的文献，系统根据用户提交的查询语句进行检索，所检索到的文献集既包含相关文献，也包含不相关文献。查准率和查全率的定义如下

$$\text{查准率}=\frac{\text{检索到的相关文献数}}{\text{检索到的全部文献数}}$$

$$\text{查全率}=\frac{\text{检索到的相关文献数}}{\text{文献库中全部相关文献数}}$$

查准率从一个方面描述了检索系统的查询开销。如果某次查询的查准率是85%，则15%的文献是不相关文献。查全率是检索系统查找用户所需信息能力的标志，对于一个具体的检索系统，其查准率随查全率的增加而减少，所以我们通常用平均查准率来评估检索系统。

表2给出了几种相关反馈方法在 Infolite 上的部分测试结果(平均查准率及相对于初始查询结果的性能提高)。

从表2可以看出，在三种相关反馈方法中，适应性查询构造方法的性能最好。此外，针对《联合早报》的平均查准率最低，其主要原因是我们的词典下完全适用于新加坡的用语，对于《联合早报》中出现的部分词汇无法识别，因此无法确定其对应的特征项。

表2 几种相关反馈方法在 Infolite 上的部分测试结果

	初始检索	Rocchio	Ide dec-hi	Adaptive
人民日报(3200篇)	0.3547	0.4922 +38.79%	0.5933 +67.26%	0.6215 +75.21%
联合早报(1200篇)	0.1984	0.2857 +44.00%	0.3066 +54.53%	0.2913 +46.77%
其它报刊(1600篇)	0.2159	0.3517 +62.87%	0.3804 +76.19%	0.4136 +91.57%
平均(6000篇)	0.2563	0.3763 +48.82%	0.4268 +66.52%	0.4421 +72.49%

结论 多年的研究和实验结果表明，相关反馈技术是一种非常有效的提高系统性能的方法。本文通过三种方法的简要介绍和实验比较，证明对于中文信息检索系统，适应性查询构造方法优于 dec-hi 和 Rocchio 方法。建议在实际系统中采用这种方法。

为了保证适应性查询构造算法的收敛性，应确保优先关系是弱线性的。这一点，我们可以通过一段检测程序来判断优先关系是否是弱线性的。

参 考 文 献

1 Maron M E, Kuhns J L. On Relevance, Probabilistic Indexing and Information Retrieval. Association for Computing Machinery, 1960, 7(3): 216～244

2 Rocchio J J Jr. Relevance Feedback in Information Retrieval. In Salton G, ed. The SMART Retrieval System-Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice-Hall, 1971: 313～323

3 Ide E. New Experiments in Relevance Feedback. In Salton G, ed. The SMART Retrieval System-Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice Hall, 1971: 337～354

4 Bollmann P, Wong S K M. Adaptive Linear Information Retrieval Models. In: Proc of the Tenth ACM SIGIR Conf on Research and Development in Information Retrieval. New Orleans, 1987: 157～163

5 Wong S K M, Yao Y Y. Query Formulation in Linear Retrieval Models. J of the American Society for Information Science, 1990, 41: 334～341

6 Wong S K M, Yao Y Y, Bollmann P. Linear Structure in Information Retrieval. In: Proc of the Eleventh ACM SIGIR Conf on Research and Development in Information Retrieval. Grenoble, France, 1988: 219～232

(上接第23页)

(6)用户浏览器自动请求上述图像，并在请求过程中将上一次接收到的 Cookie 发送给 ad.example.com。从 ad.example.com 的响应包含一个新的 Cookie。

关心隐私的人们担心：

(1)用户在第3步从一个他自己都不知道已访问过(一次未核实的事务)的站点 ad.example.com 收到了一个第三方 Cookie，

(2)第6步中在第二次图像请求里将第一次的 Cookie 返回给了 ad.example.com。

如果随着每一次图像请求都有一个 Referer 头标发送给 ad.example.com，那么 ad.example.com 就会根据用户访问过哪些站点积累出用户的概况。这里是 http://www.example1.com/home.html 和 http://www.example2.com/home.html。诸如此类的广告站点可以选择一些广告吸引用户，孤立地看获取用户概况的过程其危害程度相对说来并不严重，但若用户概况涉及的并不仅仅是一个体而是一实实在在的人时，其隐私性就显现出来了。比如用户在 www.example1.com 进行某种注册时，就会出现隐私被侵犯的情况。

RFC2109通过要求用户 agent 拒绝来自未核实事务响应的 Cookie 来限制 Cookie 可能的危害作用。RFC2109进一步规定用户可以配置 agent 接收缺省值为 not 的 Cookie。如此选择缺省值与业务模型依赖于 Cookie 的广告网络公司有关，这些公司在规范基本完成(1996年7月)到 RFC 出台(1997年2月)期间已经得到蓬勃发展。

七、S-HTTP

S-HTTP(Secure HTTP)技术是在电子商务活动中发展起来的。为了推广电子商务，美国成立了商业 Internet 联盟。在 Internet 上开展商务活动所必须解决的首要难题就是安全性问题，为此该联盟组织开发了在 HTTP 基础上扩充并增加安全功能的 S-HTTP。S-HTTP 可以采用多种方式封装信息。它的封装包括加密、签名和基于 MAC(Message Authentication Code)的认证。同时一个消息可以被反复封装加密。S-HTTP 还定义了头标以便来进行密钥传输，认证传输，并支持多种加密协议。

结束语 Web 已是 Internet 的主流业务，随着越来越多的个人、公司、网络以及国家连入 Internet，Web 的用户会变得更加庞大，其安全性就更显重要。本文主要介绍了作为 Web 基础的 HTTP 安全机制，它包括：HTTP/1.0基本认证机制及其缺陷；HTTP/1.1摘要访问控制机制；代理认证；如何保护用户的隐私；Cookie 对隐私、安全的影响以及用于电子商务的 S-HTTP。值得指出的是近些年 IETF 非常关注安全问题，目前 Internet 的安全体系结构日趋成熟，一个安全的 Internet(包括 Web)指日可待。

参 考 文 献

1 Krishnamurthy B, et al. Key differences between HTTP/1.0 and HTTP/1.1. Computer Networks, 1999, 31(11/16): 1737～1751

2 Stallings W. Data and Computer Communications. Fifth edition. 北京：清华大学出版社，Prentice Hall，1997

3 Tanenbaum A S. Computer Networks. Third edition. 北京：清华大学出版社，Prentice Hall，1997

4 裴有福. Web 技术大全. 北京：中国水利水电出版社，1998

5 蒋继洪，黄月江. 计算机系统、数据库系统和通信网络的安全与保密. 成都：电子科技大学出版社，1995