●李树青 (南京财经大学 信息工程学院，江苏 南京 210046)

个性化信息检索技术综述*

摘 要：个性化信息检索技术是现代信息检索技术的新发展形式，也是提高现有信息检索系统的检索质量，更好满足用户要求的必然举措。本文从用户模式表达方法、个性化结果获取方法和结果呈现方法 3 个角度，对个性化信息检索技术的发展现状进行全面的分析，说明了现代个性化信息检索技术的特点，并指出未来个性化信息检索技术需要进一步解决的问题。

关键词：个性化信息检索；用户模式；信息推荐

Abstract: Personalized information retrieval technology is a new form of modern information retrieval technologies, and is also an inevitable act to improve the retrieval quality of the existing information retrieval system so as to meet the user's requirement better. This paper analyzes the development status of personalized information retrieval technology from 3 aspects, that is, the expression of user profiles, the achievement of personalized results and the presentation of the results. The paper also discusses the characteristics of the modern personalized information retrieval technologies, and points out the problems which need to be solved in future.

Keywords: personalized information retrieval; user profiles; information recommendation

1 个性化信息检索

所谓个性化信息检索是指能够为具有不同信息需求的用户提供个性化检索结果的技术，即对不同用户提交的同一种查询词语也能按照不同的用户需求而生成不同的检索结果。虽然已有的实验和研究都已经证明，个性化信息检索可以提高用户的检索体验，但是仍然有一个重要的问题需要回答，那就是个性化信息检索技术是不是一定要使用。从目前的研究进展来看，有两点值得注意。

一是从用户使用的层面上看，现阶段的 Web用户往往认为这些个性化技术并不好用，个性化功能越复杂，使用难度也越大[1]。造成这种现象的原因主要有 4个：①个性化技术往往需要详细的用户模式信息，而用户出于保护隐私的考虑，并不愿意将自己的个性化特征信息存储到Web服务器上[2]。②个性化技术往往需要复杂的运算处理，典型的搜索引擎每秒需要处理成百上千个不同的查询，一天服务的用户数量可达几百万，对众多数量的用户提供快速的个性化信息检索服务往往并不容易。③一旦用户熟悉了现有的信息检索系统访问接口，如果个性化技术使用了一些新的功能和接口，反而让用户觉得难以接受并影响他们的继续使用。④许多 Web用户的访问次数较少，检索系统所获取的该用户检索历史样本较少，此时难以进行有效的用户模式分析。所以，在设计和研究新的个性化信息检索技术时，一定要注意用户的实际使用感受，尤其对于商业搜索引擎而言更是如此。

二是从个性化信息检索技术的适用面来看，并非任何情况都需要使用个性化技术。常见的用户查询可以分为三种类型，分别是清晰查询、半模糊查询和模糊查询。其中个性化信息检索技术可以提高半模糊查询和模糊查询的检索质量，但是对于清晰查询而言，普通搜索引擎似乎更为适合[3-4]。另外，也有学者指出，在所有用户中具有相似点击情况的查询往往并不需要使用个性化信息检索技术[5]。部分学者也指出不能使用单一的个性化技术来处理各种不同的查询请求[5]。近年有学者对此问题进行了实验分析，获取了比较令人信服的结论。他主要利用查询日志信息，设计了一种大规模的个性化信息检索技术评价框架，该框架主要利用 Web用户的真实检索记录来客观评价所用个性化信息检索技术的实际效果[6]。通过分析可以看出，对于一些用户查询而言，个性化信息检索技术确实可以改善普通搜索引擎的检索效果，但是该技术只适用于点击熵 (Click Entropy) 较大的查询，而对于其他一些具有较小点击熵的查询，相应的改善效果并不理想。甚至在一些情况下，个性化信息检索技术还会产生负面的作用，所以不能对所有的查询仅仅使用同一种个性化信息检索技术。这里所谓的点击熵是指一种对查询结果点击情况差异

* 本文为江苏省教育厅“青蓝工程”基金资助项目和 2007年南京财经大学校级课题 (项目编号：B0710) 资助项目论文。

进行衡量的指标，如果所有用户都在相同查询的结果中点击相同的网页，则点击熵为 0，较少的点击熵意味着大部分用户都选择该查询结果中的相同结果网页。如对于查询 Google搜索引擎网址之类的导航性查询，几乎所有的用户都是希望直接得到它的网址，个性化信息检索技术并没有任何使用意义[5,7]。这充分说明个性化信息检索技术有着自己的适用面，应该根据不同的检索环境来选择不同的个性化技术，甚至需要进一步考虑该不该使用这些个性化信息检索技术。

2 用户模式表达

用户模式表达是获取用户个性化特征信息的关键步骤，只有准确地获取用户模式信息，才能提供准确的个性化检索信息。获取用户模式的方式有很多，按照不同的标准可以分为不同的类型。下面结合一些常见的划分标准来分析不同情况下用户模式的获取方法及其特点。

2.1 显式获取方式和隐式获取方式的比较

显式获取方式是一种出现较早的方法，基于此方式的个性化检索系统都要求用户主动提交兴趣信息，系统根据这些信息与返回结果的相似度来过滤得到最终的个性化检索结果。然而随着研究的深入，人们逐渐发现了这种方式的许多缺点。大部分检索用户并不愿意花很多额外功夫去显式地标明自己的信息需求或者通过各种各样的相关度反馈方法来不断调整查询词语[8]。而且这种主动提交的特征信息在一定程度上也充满了用户的主观意识，往往可能丢失那些反映自己信息需求并且最有价值的词语，所以实际上难以反映真实的用户兴趣。更为重要的是，随着用户访问行为的改变，用户还必须及时更新这些兴趣模式。显然，这些缺点都使得显式获取方法难以取得理想的实际运行效果。

同时，这些实验也显示出用户在开始时所访问的结果网页往往较不准确，但是通过不断的重构查询而不是花时间标明自己的检索要求，却可以逐渐增强选择网页的准确度[9]。按照这个思路，部分学者通过收集用户点击信息来对搜索结果的网页次序进行排序，以满足特定用户的偏好[10]，还有的学者结合用户点击信息进行查询扩展以提高检索准确性[11]。这就是隐式获取方法的产生背景。关于隐式获取方式的有效性问题，许多学者都给出了实验证明，如有学者通过分析用户点击信息，证明了这些隐式反馈信息可以有效地提高多数用户的搜索质量[12]。目前，隐式获取方法已经成为现在主流的用户模式获取方法。

该方法无需用户显式地提交对结果相关性的评价，事实上，一般的检索用户所提交的初始查询通常都无法获取最为满意的结果，所以用户需要通过反复调整查询词语来改善返回结果。对于更为复杂的信息需求而言，用户甚至还需多次结合浏览部分结果网页的方法来进一步判断和调整查询。显然，这些用户的行为信息都能够有效反映用户真实的信息需求，若想实现检索个性化，检索系统必须有效地获取这些隐式反馈信息。也有学者通过分析用户的所有浏览历史记录来获取更为完整的隐式反馈信息以提高搜索质量[13]。值得注意的是，在使用这些隐式方法时，一般也无需用户登录，更无需用户安装客户端代理程序，总体使用效果能够让用户感觉更为舒适。

当然，除了使用用户查询和点击结果文档信息外，有时还可以结合诸如历史查询和网页书签等内容。更复杂的方式还有获取用户阅读结果网页的时间等[14]。也有学者使用可以记录用户对网页显式评价信息和浏览行为的定制浏览器，来收集诸如浏览时间、鼠标单击情况、鼠标移动情况和鼠标滚动情况等信息，来获取用户的隐式反馈信息。实验显示，浏览时间和网页中鼠标滚动次数都和网页的查询相关性具有较强的正相关性[15]。显然，这些信息或多或少都在一定程度上能够反映通过查询词语所不能直接表达的隐式信息需求。

目前，隐式获取技术已经得到广泛深入的研究，获取的数据源类型也极大的丰富起来，常见的种类有用户上下文 (如选择浏览的网页信息等)、关键词日志、Cookie历史记录、用户提交查询和协同行为等。比较典型的系统有 iWeb等，它是一个基于用户模式的智能代理系统，通过用户提交的关键词、自由文本描述和网页样本，对用户的 Web导航、信息检索和文档过滤等操作进行个性化推荐服务，同时它还使用语义网络创建用户模式以克服简单关键词匹配可能具有的一词多义现象[16]。

2.2 用户短期访问模式和长期访问模式的比较

按照用户兴趣的持续时间可以将用户信息需求分为两类：一是短期信息需求，它通常受到用户基本信息需求、先前查询和最近返回文档的影响。二是长期信息需求，它主要受用户教育水平和职业等基本因素的影响，可以利用累积的用户查询历史和网页浏览等信息来表示，具有较强的稳定性。虽然长期信息需求会对该用户的所有短期信息需求产生一定的影响，但是对于提高当前会话过程中的检索质量而言，它的实际运用效果并不如短期信息需求[14]。一般而言，获取长期用户访问模式一般需要用户登录，否则以现有的互联网技术还不能有效准确地获取匿名用户的长期访问信息，而短期模式一般无需登录，利用简单的会话或者 Cookie等技术就可以方便地实现。

对于个性化信息检索而言，利用用户短期访问模式和长期访问模式都能为提高搜索效果提供非常重要的作用，但是这两种方式所适用的环境并不一样，不同的方式会对

不同的查询有着不同的影响。

关于两种方式的选择问题，很多学者都提出了自己的看法。如有学者指出很多用户在进行文档搜索时往往体现出一种和该用户长期信息需求并不一致的临时性信息需求[14]。显然，在这种情况下，必须要使用基于短期用户兴趣模式的个性化信息检索技术。也有学者将用户查询分为新查询和重复查询，并发现较近的搜索历史信息对改进新查询更有帮助，而全部的搜索历史信息则对重复查询有较好的改进作用[17]。这同样说明不能对所有的查询一视同仁地使用相同的方法进行个性化处理。所以，需要根据检索环境来选择使用何种访问模式，甚至可以考虑结合使用两种方式。

2.3 内容访问模式和行为访问模式的比较

所谓内容访问模式是指利用用户访问的网页信息内容或者提交的查询词语内容来获取用户访问特征，并以此进行个性化服务。而行为访问模式则主要利用用户的访问行为来获取用户访问特征，而非直接的文本内容，如利用用户是否查看具有相同 URL 的结果网页来推测用户是否具有相同的兴趣等。

由于一词多义和多词同义现象的存在，基于内容访问模式的检索方法往往会遭受所谓的词语歧义问题。这种现象往往会导致用户查询和文档的误匹配，如同义词往往会造成查全率下降，因为此时用户可能没有指定全部相关的检索同义词；再如一词多义可能会造成查准率下降，因为此时会检索出无关的网页，虽然这些网页含有相同的词语[18]。所以，准确理解网页文档的内容和它与用户真实检索意图的相关性，对于提高检索质量是非常重要的。

当个性化信息检索技术考虑了不同用户的检索行为时，相应的检索方法就可以被称为基于协作的检索方法，它的基本思想就是假设具有相似兴趣的用户往往会检索相似内容的信息。社会导航软件就是基于这种思想，它允许用户在网站中提交诸如评论、注释或者投票结果等信息，而这些信息又反过来可以为其他用户的浏览和查询提供帮助[19]。采用此类技术的著名信息检索系统有 EUREKSTER[20]，但是由于算法的复杂性，基于这种思路的商业信息检索系统并不多见。

关于对两种方法的比较评价，也有学者给出了基于比较实验的分析结论，他对 5种个性化信息检索策略进行了大规模的评价和分析，其中两种方法基于点击行为分析，3种方法基于查询内容分析，其中基于点击行为分析的个性化策略通常具有较为稳定和一致的表现，虽然该种方法只适合处理重复查询的个性化检索，而基于查询内容的个性化信息检索策略的效果则显得较不稳定，而且随着用户搜索历史的增加，个性化信息检索技术的实际运行效果将变得愈发不稳定[6]。因此，进一步研究这两种方式的不同适用条件是下一步研究需要解决的问题。

2.4 单用户模式和用户群模式的比较

所谓单用户模式是指以每个用户为单位来进行个性化信息检索服务，个性化所依据的用户访问特征因每个用户不同而不一样。而用户群模式则与此相反，它主要基于兴趣的相似度将不同用户划分为若干用户群，并以用户群来整体性地进行个性化服务。从理论上讲，前者应该具有更高的个性化精准度，但是由于受到个体样本信息不足和实时计算要求较高等因素的影响，实际运行的精确度往往并不高。而后者由于采用用户群作为个性化服务单位，所以除了能够以用户内容访问模式来进行个性化服务外，还可以利用用户群中每个用户的行为访问模式来更多地获取个性化特征，比较著名的就是基于协同过滤技术的个性化信息检索技术。如 EUREKSTER 搜索引擎就使用协同过滤技术实现了 SearchParty专有模块，该模块可以帮助用户发现和查询最为相关的一些网页[20]。它的实现方法主要是记录用户选择的查询结果网页，同时系统把这些信息在具有相同主题的用户群体间进行共享。除了使用这种协同过滤方法外，EUREKSTER搜索引擎还能支持很多的用户个性化操作，如用户如果花了很长时间浏览一些网页结果，那么如果用户再次发出相同的查询，这些先前被仔细浏览过的网页将会被排列在结果网页的前面，这样用户就无需再次进行相似的定位工作了。在界面上，系统会将那些返回结果中被用户频繁访问的网页都使用较深的颜色作为背景。即便是有些结果网页没有使用协同技术进行排序处理，这些网页也都会根据其他用户的检索历史来给出可视化的提示内容。Compass Filter系统则使用了另外一种协同过滤技术，它主要根据 Web网页文档的内在关联度来将这些文档划分为不同的 Web社区，这需要在预处理阶段来对网页的超链接关系进行分析[21]。如果用户经常访问某 Web社区中的网页，则对于他所获取的查询结果网页，如果结果网页位于该 Web社区，则这些网页应当获得一定的权重提升。也有学者综合使用协同过滤技术和传统的文档内容分析技术来提供个性化信息检索服务[22]。

3 个性化结果获取

个性化结果获取是指利用已有的用户模式信息来对用户的检索内容进行个性化处理，它是个性化操作的核心处理步骤，往往也直接影响着系统的实际运行效率。它一般可以划分为两个阶段：一是离线处理阶段，期间主要分析用户查询和访问文档的关系，并预先完成一些计算开销较大的操作；二是在线处理阶段，它利用离线处理阶段得到的信息生成用户模式，并根据用户最近的一次访问信息来

提供个性化的信息服务。

3.1 个性化结果获取的时机选择

用户在检索相关信息时，一般要经过 3个主要步骤，分别是提交查询、系统处理查询和返回查询结果。与此相应，个性化处理也可以分别应用于这 3个步骤中，形成了 3种常见的个性化结果获取方法。

1) 查询扩展。该种方式的个性化处理是利用用户模式信息来对用户的查询添加或者删除部分查询词，以更好地反映用户的个性化需求。通过添加额外的查询词来扩展短查询能够在很大程度上消除诸如同义词或者一词多义等现象产生的问题。另外，如果查询检索的结果数量太少，系统可以使用相近语义或者在统计上具有相关性的词语来替换现有查询以得到更为丰富的查询结果。其中，该方法最大的好处在于它无需改变现有的搜索技术，获取个性化结果的检索过程和获取一般的检索结果过程是一样的，用户模式只影响查询的表示形式。

2) 结合个性化信息的查询生成。该种方式的个性化处理可以提供最直接的响应，通过在检索过程中引入个性化信息来获取最终的查询结果。然而，由于个性化处理过程往往需要运行较长的时间，所以该方法并不适用于普通信息检索系统。

3) 对查询结果的个性化重排序。它可以看成是一种对现有系统的扩展，允许用户有选择地挑选指定方法来过滤查询结果。常见的实现形式有两种：一种是利用客户端软件来对搜索引擎的查询结果进行本地处理以得到个性化的信息结果，虽然这种方法可能存在性能问题，但是最大的好处在于它往往可以获取较为丰富的用户模式信息；另一种是在服务器利用用户模式信息直接过滤查询结果，有时为了能够实时探测用户当前访问模式，这种方法还需要在服务器和客户端多次进行信息传输以生成最终的个性化结果。值得注意的是，为了减少处理或者下载查询网页结果的时间，这些方法通常只会处理查询结果网页集合中前面的若干条记录，或者只通过查询结果中的文档摘要信息来进行文档信息分析等。从上述分析可以看出，查询扩展方式和对查询结果的个性化重排序是两种较为实用的方法，也是目前主要的实现方法。

3.2 客户端模式和服务器端模式的比较

服务器端通常是从服务器访问日志中获取查询历史信息或者浏览历史信息，而客户端则是获取诸如 Cookie信息或者鼠标键盘的操作行为信息等。用户查询历史无疑是了解用户信息需求的良好依据，而且诸如搜索引擎等信息检索系统可以在不干扰用户正常使用的情况下来获取这些信息[23]。利用搜索引擎服务器日志和客户端的 Cookies信息就可以识别用户和相关的点击情况，那些使用诸如 IP和最近访问时间等信息来实现的用户识别方法不仅复杂而且精确度也不高[24]。然而，不可否认的是，这些利用服务器日志获取的用户信息往往存在着精确度不高的弊端，甚至连匿名用户的准确识别都是问题。而基于客户端模式的处理技术显然没有这个问题，相反，由于它运行在客户端，所以可以获取更多的用户访问特征。同时，这种方法也无需用户主动提交信息，也不用担心隐私保护问题，而且还不会增加服务器的运行负担[25]。基于服务器端的个性化信息检索技术相当常见，大型商业信息检索系统几乎都采用这样的个性化检索策略。反之，在一些桌面检索系统或者小型检索系统中，基于客户端的个性化信息检索技术反而很常见。如 UCAIR原型系统也使用了基于客户端的个性化信息检索技术，利用从搜索引擎获取的搜索结果，结合客户端程序收集的会话查询信息和点击信息，及时更新未显示文档的排列次序[26]。该程序是一个 Web浏览器的工具栏插件，它目前只支持 IE浏览器和 Google搜索引擎，事实上也可以改进以适应其他浏览器和搜索引擎。它主要分为 3个模块：一是隐式用户建模模块，它主要用于捕获用户检索的上下文信息，如提交的查询和点击情况，并且还可以判断会话的合理长度；二是查询修改模块，它可以根据当前用户模型来重构查询；三是结果重排序模块，它可以对用户未查看的结果网页即时进行重排序，以实现个性化信息检索的效果。在具体的操作中，UCAIR关注用户的 4种操作行为，分别是提交查询、查看结果文档、点击后退按钮、点击下一页搜索结果超链。具体处理过程如下所述，首先使用一种可能的扩展查询获取搜索结果，并更新表达信息需求的用户模型向量，然后根据目前的用户模型重排序未查看的文档结果，并根据目前的用户模型重排下一页的结果网页。

有学者据此还提出了一种名称为 JITIR (Just-in-Time IR) 的信息检索工具，该系统运行在客户端，持续性地监视用户访问各种软件的行为，如在 Word中键入词语或者使用浏览器冲浪等[27]。它所采用的监视方式是隐式的，不直接干预用户的操作，但可以根据当前用户操作自动识别用户的即时信息需求，并直接将相关信息检索结果提供给用户。和 Google的 Alert技术不一样的地方在于该系统关注用户的当前活动，而且还可以根据这些活动及时更新用户模式。不过，为了获取这些信息，往往需要用户安装一些客户端代理程序，这在一定程度上增加了用户的使用复杂度，很多用户出于安全的考虑，并不愿意在客户端安装诸如浏览器插件等程序，这些都是在选择方法时需要注意的问题。

3.3 常见的个性化信息检索方法

从实现原理上看，目前的个性化信息检索方法主要有

3种，分别为基于文本内容分析的方法、基于点击流分析的方法和基于超链分析的方法。

基于文本内容分析的方法通过获取用户的查询历史和访问网页等文本信息，甚至有时还可以结合用户主动提交的，反映自身兴趣的关键词来得到个性化检索结果。从各项研究表明，单纯使用基于文本内容的分析方法存在着很多问题，样本不足和词语歧义等现象都会导致个性化效果不理想。而基于点击流分析的方法和基于超链分析的方法则使用了一些间接反映用户个性化信息需求特征的方法，往往更能有效地提供个性化的检索服务。

如由 Sun等人提出的 CubeSVD (Cube Singular Value Decomposition) 方法主要基于点击流日志分析技术，该方法非常适合 Web搜索引擎中的个性化技术实现[28]。经过用户的一段时间使用，系统可以记录下所有的点击流信息，该信息可以通过一个三元组合来表示：<user, query, visited page>，这个组合在一定程度上反映了用户的偏好。该系统首先利用提出的框架方法分析了这些元素间的相关关系；其次由于用户所提交的查询词语相对于整个查询词语集合而言非常少，所以还需要对实际运行中存在的数据稀疏问题进行处理；最终得到的输出可以表示为：<user, query, visited page, w>，其中 w为用户 user在提交查询 query时点击网页的概率。按照这个概率值，就可以将相关网页推荐给用户。通过 4 470万点击记录的评价，系统取得了比协同过滤技术和潜在语义索引技术更为精确的效果。不过，计算所需的时间开销较大，但由于主要采用离线处理方式，所以对运行效率影响不大。另外，该系统还需要定期对获取的新点击流数据进行再次处理，以保持个性化推荐的效果。

再如基于超链分析的个性化信息检索方法，它主要利用修改网页的标准 PageRank值来反映用户的个性化信息需求。如有学者提出的 PROS方法，它可以根据用户书签和经常访问的网页等信息来生成用户模式，并以此模式结合网页的超链特征来提供个性化的网页排序结果[29]。具体处理过程说明如下：用户选择的感兴趣网页被发送到 HubFinder模块，该模块收集与当前用户兴趣相关的 Hub网页，也就是那些含有较多指向高质量网页的网页；同时，该模块还使用一种定制的 HITS改进算法来分析 Web网页的结构；最后，系统利用一种被称为 HubRank的排序算法综合了 PageRank值和网页的 Hub值，并对结果进行重排序以表达用户的个性化检索特征。除此以外，基于超链分析的个性化信息检索方法还有著名的主题敏感 PageRank方法，它可以根据不同的主题来给所有网页计算不同的 PageRank值，具体的计算方法是根据这些网页在 ODP概念层次中的位置来进行。在查询期间计算查询与这些主题的相似度，并以此将这些主题敏感 PageRank值进行加权线性组合。由于大部分计算可以离线完成，所以这种方法具有较好的运行性能[30]。为了提高性能，有学者还对主题敏感 PageRank方法提出了缩放性能更高的算法[31]。也有学者对主题敏感 PageRank方法进行了扩展，当用户提交查询时，系统首先选择和用户查询最为相关的主题，并使用该主题对应的 PageRank值来对结果进行重排序[32]。其他学者还提出了利用 DNS域名定制的个性化 PageRank值，并基于 ODP概念层次开发了个性化搜索系统[33]。最近，有学者提出了根据用户已经点击过的网页的主题敏感 PageRank值，来估算用户的隐含兴趣模式[7]。这些方法仍然在个性化信息检索技术中占有重要地位。

4 结果呈现

按照结果网页的呈现内容来看，常见的方式就是直接修改呈现的结果网页内容，如在结果网页上添加反映个性化内容的超链，或者直接显示排序后的结果。与此相对，基于结果聚类的呈现技术也逐渐受到人们的关注。结果聚类本身就可以看成是一种个性化技术，它意味着用户可以根据自己的信息需求，通过导航和选择指定的类别来定制显示的结果范围。传统的信息检索系统通常使用按照相关度排序的结果列表方式来呈现搜索结果，如果用户不能十分准确地表达自己的检索需求，通常需要在列表中浏览很长时间才能找到自己所需的内容。所以，很多信息检索系统开始使用聚类方法来将结果文档划分为不同的组，以方便用户定位所需内容。在 Web应用领域，聚类处理通常有如下几个特征：一是聚类操作通常都是在获取查询结果以后才进行的，所以这个过程比较快，而且还允许用户交互式的操作；二是因为计算性能的考虑，这些聚类算法通常只使用文档摘要而不是整篇文档来进行聚类分析；三是由于聚类不要求系统事先定义类别层次目录，所以生成的聚类层次必须要方便用户导航和选择；最后是聚类结果应该提供对类别的简要说明以方便用户准确定位相关类别，即便是对于那些明显有问题的结果，系统也应该对其进行必要的聚类处理说明。比较典型的系统有 CLUSTY[34]和 KARTOO[35]等。Scatter/Gather也使用相似的方法，用户在选择聚类类别时，系统会记录这些类别信息，并对这些类别继续进行聚类，将整个 Web网络的所有网页逐渐分散成若干小的聚类结果，再反馈给用户。这种过程经过几次迭代后，聚类结果将变得很小，用户就可以直接选择所需内容[36]。

按照结果网页的呈现方法，传统的呈现模式主要是在用户进行检索访问时，系统被动地提供个性化的结果内容。而目前一种基于信息主动推荐服务的个性化结果呈现

技术逐渐受到人们的关注。系统通过收集用户对相关领域内容的反馈信息，分析具有相似兴趣模式的用户群体，通过其他相似用户的访问内容来给当前用户提供信息推荐，具体的推荐方式有很多，如通过电子邮件等[37]。这可以看成是个性化信息推荐服务在信息检索领域中新的应用。

5 结束语

个性化信息检索技术是一种非常有发展前景的信息检索技术，也是目前各种信息检索系统研究中最受人关注的一个方向，它对于提高用户的访问效果和改善用户体验以完善现有信息检索系统起着至关重要的作用。尽管相关技术和理论已经得到了长足的发展，然而个性化信息检索技术仍有很多值得研究和探讨的领域。

1) 由于用户兴趣本身是动态变化的，如何准确识别和跟踪这种不断变化的用户兴趣来有效表达用户的兴趣模式也需要进一步研究[37]。

2) 现阶段大部分算法仍然存在着缩放性问题，不适合海量信息的实时处理要求，设计和改进这些算法以提高计算性能也是亟需解决的问题等。

3) 各种个性化检索系统总体框架的实施复杂度较大，在实际的商业系统中，必须有效地解决处理海量数据可能带来的性能问题。

4) 设计界面更为友好的系统接口，以方便用户使用，它可以在不显著增加用户使用负担的情况下提供效果理想的个性化信息检索服务。□

参考文献

[1] KHOPKAR Y, SPINK A, GILES C L, et al. Search engine personalization: an exploratory study [EB/OL]. http://www.firstmonday.org/issues/issue8_7/khopkar/index.html.

[2] KOBSA A. Privacy-enhanced Web personalization [M] // BRUSILOVSKY P, KOBSA A, NEJDL W. The adaptive Web: methods and strategies of web personalization. New York: Springer-Verlag, Berlin Heidelberg, 2007.

[3] CHIRITA P A, FIRAN C, NEJDL W. Summarizing local context to personalize global Web search [C] // Proceedings of CIKM'06, 2006.

[4] CHIRITA P A, BEIDL W, PAIU R, et al. Using odp metadata to personalize search [C] // Proceedings of SIGIR'05, 2005.

[5] TEEVAN J, DUMAIS S T, HORVITZ E. Beyond the commons: Investigating the value of personalizing Web search [C] // Proceedings of PIA'05, 2005.

[6] DOU Z, SONG R., WEN J R. A large-scale evaluation and analysis of personalized search strategies [C] // Proceedings of WWW'07, 2007.

[7] LEE U, LIU Z, CHO J. Automatic identification of user goals in Web search [C] // Proceedings of WWW'05, 2005.

[8] ANICK P. Using terminological feedback for web search refinement: a log-based study [C] //Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, New York: NY, USA, ACM Press, 2003.

[9] TEEVAN J, ALVARADO C, ACKERMAN M S, et al. The perfect search engine is not enough: a study of orienteering behavior in directed search [C] // Proceedings of The SIGCHI Conference on Human factors in Computing Systems, New York: NY, USA, ACM Press, 2004.

[10] JOACHIMS T. Optimizing search engines using clickthrough data [C] //Proceedings of SIGKDD 2002, 2002.

[11] SUGIYAMA K, HATANO K, YOSHIKAWA M. Adaptive Web search based on user profile constructed without any effort from users [C] // Proceedings of WWW'04, 2004.

[12] KELLY D, TEEVAN J. Implicit feedback for inferring user preference: a bibliography [J]. SIGIR Forum, 2003, 37 (2).

[13] NUNBERG G. As google goes, so goes the nation [N]. New York Times, 2003-05.

[14] SHEN Xuehua, TAN Bin, ZHAI Chengxiang. Context-sensitive information retrieval using implicit feedback [C] // Proceedings of The 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005.

[15] CLAYPOOL M, LE P, WASEDA M, et al. Implicit interest indicators [C] // Proceedings of Intelligent User Interfaces 2001, 2001.

[16] ASNICAR F A, TASSO C. IfWeb: a prototype of user model-based intelligent agent for document filtering and navigation in the world wide web [C] // Proceedings of Workshop Adaptive Systems and User Modeling on the World Wide Web (UM97), Italy: Sardinia, 1997.

[17] TAN B, SHEN X, ZHAI C. Mining long-term search history to improve search accuracy [C] // Proceedings of KDD'06, 2006.

[18] FREYNE J, SMYTH B. An experiment in social search [M] //Adaptive Hypermedia and Adaptive Web-Based Systems, Third International Conference, Eindhoven: Springer, 2004.

[19] DIEBERGER A, DOURISH P, et al. Social navigation: techniques for building more usable systems [J]. Interactions, 2000, 7 (6).

[20] Eurekster swicki home [EB/OL]. [2008-09-01]. http://www.eurekster.com/.

[21] KRITIKOPOULOS A, SIDERI M. The compass filter: search engine result personalization using Web communities [C] // Proceeding of Intelligent Techniques for Web Personalization, Acapulco, Mexico, 2003.

[22] CLAYPOOL M, GOKHALE A, MIRANDA T, et al. Combining content-based and collaborative filters in an online newspaper [C] //ACM SIGIR Workshop on Recommender Systems - Implementation and Evaluation, ACM Press, 1999.

[23] LAWRENCE S. Context in Web search [J]. IEEE Data Eng, 2000, 23 (3).

[24] PIROLLI P L T, PITKOW J E. Distributions of surfers' paths through The World Wide Web: empirical characterizations [J]. World Wide Web, 1999 (2).

[25] Text information management group [EB/OL]. [2008-09-01]. http://sifaka.cs.uiuc.edu/ir/ucair/download.html.

[26] SHEN Xuehua, TAN Bin, ZHAI Chengxiang. Implicit user modeling for personalized search [C] //Proceedings of the 14th ACM international conference on Information and knowledge management, 2005.

[27] RHODES B J. Just-in-time information retrieval [D]. Cambridge, MA: MIT Media Laboratory, 2000.

[28] SUN J T, ZENG H J, LIU H, et al. Cubesvd: a novel approach to personalized Web search [C] // Proceedings of the 14th international conference on World Wide Web, New York: NY, USA, ACM Press, 2005.

[29] CHIRITA P A, OLMEDILLA D, NEJDL W. A personalized ranking platform for Web search [C] //3rd International Conference Adaptive Hypermedia and Adaptive Web-Based Systems. Eindhoven, The Netherlands, Springer, 2004.

[30] HAVELIWALA T H. Topic-sensitive pagerank [C] // Proceedings of the 11th international conference on World Wide Web, New York: NY, USA, ACM Press, 2002.

[31] JEH G, WIDOM J. Scaling personalized Web search [C] // Proceedings of WWW '03, Pages, 2003.

[32] QIU F, CHO J. Automatic identification of user interest for personalized search [C] //Proceedings of The 15th International Conference on World Wide Web, New York: NY, USA, ACM Press, 2006.

[33] AKTAS M S, NACAR M A, MENCZER F. Personalizing pagerank based on domain profiles [C] // Proceedings of The Sixteen WEBKDD Workshop Web Mining and Web Usage Analysis (WEBKDD'04), Seattle, Washington, 2004.

[34] Clusty the clustering search engine [EB/OL]. [2008-09-01]. http://clusty.com/.

[35] KartOO visual meta search engine [EB/OL]. [2008-09-01]. http://www.kartoo.com/.

[36] CUTTING D R, KARGER D R, PEDERSEN J O, et al. Scatter/gather: a cluster-based approach to browsing large document collections [C] // Proceedings of The 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York: NY, USA, ACM Press, 1992.

[37] 曾春，邢春晓，周立柱. 个性化服务技术综述 [J]. 软件学报，2002 (10).

作者简介：李树青，男，1976年生，讲师，博士生。

收稿日期：2008 - 12 - 17

(上接第 116页)

4 结论与展望

信息系统安全风险评估对确保信息安全具有重要意义。信息系统安全风险评估要涉及大量信息，是一个复杂的过程。有几点需要深入理解：①安全评估不是漏洞扫描。信息系统安全评估包含丰富的内容，漏洞扫描是脆弱性分析的一部分，是整个评估的重要数据来源而非全部。②安全评估不是单个的具体产品。安全评估是一个有着严格流程的体系。它是动态、发展的，而非停滞、静态的。③安全评估的结果应该可以比较。安全评估的结果必须能够相互比较才可以具有较好的参考意义，才能够保证安全评估相关研究的规范发展。④安全评估中主观因素的影响过大，会导致评估工作的随意性太大，从而不能保证评估工作的质量。

因此，对于信息系统风险评估可以在以下方面做深入研究：①组织关键信息资产的确定和安全估价，这是安全投资决策的基础。②把信息系统风险评估视为系统工程来深入研究。③基于模型的风险评估以及风险处理方法的进一步完善，引入和创建更加适应于信息安全领域的风险评估模型等[6]。④安全评估应客观化和自动化，自动化的评估将有利于降低评估成本，减少评估周期，尽快反映系统状态以利于决策，因此开发自动化的辅助评估工具十分必要。□

参考文献

[1] 田永. 信息系统的安全评估方法 [J]. 宿州学院学报，2007 (8).

[2] 张基温. 信息系统安全教程 [M]. 北京：清华大学出版社，2007：7.

[3] 罗森林. 信息系统安全与对抗技术 [M]. 北京：北京理工大学出版社，2005：8.

[4] 陆宝华，王楠. 信息系统安全原理与应用 [M]. 北京：清华大学出版社，2007：11.

[5] 陈波，于泠，肖军模. 计算机系统安全原理与技术 [M]. 北京：机械工业出版社，2006：1.

[6] 陈光匡，兴华. 信息系统安全风险评估研究 [J]. 网络安全技术与应用，2004 (7).

作者简介：宋艳，女，1974年生，博士，教授。
陈冬华，女，硕士生。

收稿日期：2008 - 12 - 15