文章编号：2096-1472(2019)-02-05-06
DOI:10.19644/j.cnki.issn2096-1472.2019.02.002

# 信息检索中支持结果多样化的查询性能预测

张忠敏，吴胜利
(江苏大学，江苏 镇江 212013)

**摘　要**：对支持检索结果多样化任务的查询性能预测进行了研究。分析了现有性能预测算法的不足，考虑利用不同方式衡量最终检索结果列表的多样性，并在此基础上提出同时考察查询结果的相关性性能与多样性性能的三种方法。采用TREC ClueWeb09B数据集、Web Track任务的查询集及开源的Indri搜索引擎构建实验平台并进行实验。基于Spearman、Pearson和Kendall相关系数的评价结果表明，所提出的三种方法与传统方法相比更适用于预测多样化检索结果，且在不同条件下性能稳定。

**关键词**：信息检索；查询性能预测；检索结果多样化
**中图分类号**：TP391.3　**文献标识码**：A

# Query Performance Prediction for Search Result Diversification in Information Retrieval

ZHANG Zhongmin,WU Shengli
( *Jiangsu University*,*Zhenjiang* 212013,*China*)

**Abstract:**Query performance prediction in supporting of the task of retrieval results diversification is studied.This paper analyses the shortcomings of the existing performance prediction algorithms,considers using different ways to measure the diversity of the final search results list,and proposes three methods to simultaneously examine the relevance and diversity performance of the query results.TREC's ClueWeb09B dataset and query sets for the Web Track Task,and open sourced search engine Indri are used to build the experimental platform and carry out experiments.Measured by using Spearman,Pearson and Kendall correlation coefficients,the evaluation results show that the three proposed methods are more suitable for predicting diversified retrieval results than traditional methods,and have stable performance under different conditions.

**Keywords:**information retrieval;query performance prediction;search result diversification

## 1 引言(Introduction)

查询性能预测，也称为查询困难度预测，是指在没有相关性判断信息的情况下，评估给定查询对应检索结果的性能[1-4]。查询性能预测技术可以辅助搜索引擎向用户提供更稳定的服务。从搜索引擎的角度讲，搜索引擎的管理员可以收集查询性能预测技术产生的信息，识别出当下用户检索频率高但性能较差的一类查询，增加相应的文档数量来提高该类查询的检索性能，从整体上提高搜索引擎的鲁棒性(Robust)[1]。从用户的角度讲，当用户把检索信息需求转化为查询时，存在信息丢失或输入的关键词不够具体等问题。此时查询性能预测技术可以迅速识别这些问题，借由与用户的交互辅助用户构建意义更明确的查询，提高检索有效性。

传统意义上查询性能预测指的是查询的相关性性能预测。目前依然是信息检索和WEB搜索领域的一项重要任务，已有多篇论文[1,3,5-7]对其进行了深入研究。根据是否分析查询的检索结果，通常将查询性能预测方法分为检索前(pre-retrieval)预测方法[5,8]和检索后(post-retrieval)预测方法[1,4,6,9]。

对于给定查询，检索前预测方法指在搜索引擎产生检索结果前评估查询的性能。此时可分析的信息主要有查询词项在文档集中的统计特征(例如倒转文档频率IDF)、查询短语的语言特征[10]等。虽然可以在检索开始前预测查询的性能，计算代价小且速度快，但是由于没有分析检索模型方面的特性，

基金项目：本论文得到江苏省自然科学基金(BK20171303：大数据环境下支持检索结果多样化的联邦搜索引擎技术)资助.

因而预测的有效性不高。检索后预测方法是指在搜索引擎产生检索结果后评估查询的性能。此时预测方法不仅能使用查询词项在文档集中的统计信息和查询自身的语言特征，还可以使用搜索引擎反馈的文档列表信息[11,12]，例如查询词项在返回文档集中的分布、返回文档之间的关系等。由于查询对应的返回文档集包含与检索模型算法相关的信息，检索后预测方法可分析的数据因此更多且更可靠，从而预测更加准确有效。检索后预测方法又可以大致分为三类：基于结果列表相对于整个数据集的清晰度(Clarity)[13]，基于检索结果的稳定性程度(Robustness)[14]，和分析结果列表中文档的得分分布[6,15]。

分析得分分布的预测算法在传统预测任务上结果较优。SD2[15]方法基于结果列表中文档得分的标准方差，并且根据第一篇文档的得分设定最低得分阈值来选择截断系数k，即截取的文档列表长度。Zhou等[16]提出了一种预测方法WIG(Weighted Information Gain)，该方法分析检索结果中前k篇文档的平均得分和整个文档集的平均文档得分的差值。若差值越大，说明前k篇文档的整体得分越高，即检索结果越有效。Shtok等[17]提出了另一种预测方法NQC(Normalized Query Commitment)，该方法分析了检索文档列表中靠前和靠后文档的查询漂移程度。若检索结果中文档得分的梯度越明显，则说明搜索引擎能更好地把握每篇文档与查询的相关度。也就是说漂移程度越小，检索结果的质量越高。Tao和Wu[6]提出了一种SMV(Score Magnitude and Variance)预测方法，该方法同时考虑了检索结果中文档分数的大小和方差。

检索结果多样化是信息检索领域中一项重要的任务，对于一些语义宽泛，解释多样的查询尤为必要。一般它由两个步骤实现。第一步采用传统的搜索引擎进行搜索，得到一初始的结果列表。在这一步只考虑文档的相关性。第二步对第一步的结果列表进行重排，这时既要考虑相关性，也要考虑多样性。第二步重排的过程本质上是一个双目标的优化问题。一方面需要检索与查询相关的结果，另一方面也要尽可能地覆盖查询的多个方面。事实上，若将查询$q$看成由一组子意图$A$组成，并且初始结果列表$R$中每篇文档$d$包含的子意图是$A$的一个子集，那么问题变成对于给定的文档数$\tau$找到一个子集$S$满足$S \subseteq R \wedge |S| \leq \tau$，子集$S$覆盖查询子意图的程度$|\bigcup_{d_i \in S} d_i|$取得最大值。该问题就成了一个最大化覆盖问题，它是NP难解的。

相应的，支持检索结果多样化任务的查询性能预测算法需要同时考虑查询的相关性性能和多样化性能。亦即不仅需要考虑返回结果列表中是否包含较多的相关文档，而且需要考虑它们对查询不同方面的覆盖程度。因此不同于只考虑结果列表相关性的传统性能预测方法，算法需要额外考虑结果列表文档之间的内容冗余度和新颖性。通过实验得知，现有传统的性能预测方法预测检索结果多样化性能的有效性不高，有必要专门研究支持结果多样化性能的预测算法。到目前为止，尚未见到有相关的研究。因此本文的工作有较高的创新性。

在综合考虑检索结果相关性和多样性两方面特征的基础上，本文提出三种检索后查询结果多样化性能预测算法。实验结果表明所提的算法具有较高的可用性。

## 2 检索结果多样化(Search result diversification)

由于检索结果多样化问题的复杂性高，以往研究的多样化算法大多基于贪心法。即对于一个模糊查询，从初始检索结果中迭代选择局部最优的文档，产生重排后的结果。所选文档应当最大化覆盖与该查询相关的各个方面，并且与已经选择的文档间的相似度最小。

根据是否分析查询意图，现有的多样化方法可分为隐式和显式两类。隐式多样化方法假设相似的文档会覆盖相同的子意图，因此尽可能使结果中文档之间存有较大的差异度。这类方法的特性是，不直接分析查询意图，而分析文档的相似程度。

MMR(Maximal Marginal Relevance)[18]是最早的一种隐式多样化方法。该方法分析候选文档与查询之间的相关度，并计算候选文档与已排文档之间的相似度，降低最终文档列表的冗余度。

区别于隐式多样化方法，显式多样化方法的特性是直接获取查询的子意图集合。在此基础上建立目标函数，使选择的文档最大化覆盖查询的多个子意图。Santos和Macdonald等人[19]提出了概率框架xQuAD(Explicit Query Aspect Diversification)，该框架把查询的多个方面表示为一组子查询集合。一方面评估文档与查询的相关性，另一方面考虑候选文档覆盖的子查询数目，以及这些子查询是否已被排在前面的文档覆盖。并用参数控制两者之间的平衡，使得重排后的结果兼顾相关性与内容的新颖性。

Dang和Croft等人[20]提出了PM−2显式多样化方法，该方法将子查询视作参与选举的不同党派，主要思想类似于政党选举中的Sainte−Laguë方法。该方法首先计算每个席位应被分配的子查询，然后选择出一篇针对该子查询得分最高的文档获得该席位。重复这一过程直至获得重排后的结果。

本文采用三种典型的显式多样化方法对初始结果列表进行重排，分别是xQuAD[19]、PM−2[20]和CombSUM。有效的多样化方法是我们构建实验平台和测试预测算法性能必不可少的。

## 3 多样化性能预测算法(Diversified performance prediction algorithm)

本节提出的三种多样化预测算法，均属于检索后预测算

法。在预测查询多样化性能时，分别考虑影响多样性的不同因素，分析中间检索结果列表蕴涵的信息。

### 3.1 IASUM

IASUM主要从检索结果的多样性和新颖性两个方面考察。多样性衡量结果列表是否覆盖到查询的不同子意图，新颖性则考虑结果列表是否过度偏向于某些特定子意图。对特定检索模型$M$与查询$q$，假设不同子查询的权重一致，IASUM算法如公式(1)所示。

$$IASUM(q,M)=\frac{1}{n_{subquery}}\sum_{i=1}^{k}\frac{Sum(d_i)}{i}$$
$$Sum(d_i)=\sum_{q_j=1}^{n_{subquery}}\frac{1}{60+rank_{d_i,q_j}}(1-\alpha)^{q_{jnum}} \quad (1)$$

其中，$n_{sbquery}$表示查询$q$包含的子查询总数；$k$表示文档列表截断系数，本文中设置$k$为100。

$Sun(d_i)$表示最终结果列表中位置$d_i$处的文档$i$在所有子查询检索结果中的累积得分。若文档$d_i$在子查询$q_j$所对应的结果列表中出现且$rank_{d_i,q_j}$表示其排名，可通过倒数模型$\frac{1}{60+rank_{d_i,q_j}}$将文档排名转换为有效的文档得分，否则得分为0。$\alpha$表示文档冗余的惩罚因子，本文取$\alpha=0.1$。$q_{jnum}$指子查询$q_j$的检索结果中已经被选取参与计算的文档数目。

本节提出的IASUM预测方法，在文档的相关性性能方面，利用文档排名信息度量与子查询的相关程度。而在多样化性能方面，考虑到每一个子查询结果列表均对最终结果有所贡献，IASUM利用惩罚因子$\alpha$对隶属同一子查询的文档的贡献值进行惩罚。由此可见，IASUM是一种综合考量了相关性和多样性的预测方法。

### 3.2 IASUM2

IASUM算法本质上是从检索结果列表排名信息的角度对检索结果的多样化性能进行分析。考虑到检索结果中文档的分数分布也提供了有价值的信息，因此可将IASUM算法和传统的预测算法($SD2$[15]和$WIG$[16])进行整合，得出一种组合的多样化性能预测算法IASUM2，如公式(2)所示。

$$IASUM2(q,M)=SD2(q,M)\times WIG(q,M)\times IASUM(q,M) \quad (2)$$

IASUM2方法中的$SD2$和WIG因子考虑了结果列表中文档得分的分布特征。前者基于结果列表中文档得分的标准方差，具体见公式(3)。后者基于前$k$篇文档的平均得分和整个文档集文档的平均得分间的差值，具体见公式(4)。

$$SD2(q,M)=\frac{1}{\sqrt{|q|}}\sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(score(d_i)-\mu)^2} \quad (3)$$

$$WIG(q,M)=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\sqrt{|q|}}(score(d_i)-score(D)) \quad (4)$$

其中，$\sqrt{|q|}$作为规模因子使得不同长度的查询的分数具有可比性。$score(d_i)$表示文档$d_i$在最终结果列表中的得分。$\mu=\frac{1}{k}\sum_{i=1}^{k}score(d_i)$表示结果列表前$k$个文档的平均得分。$score(D)$表示将数据集$D$作为一个文档所得的分数。

### 3.3 IASUM3

不同于仅考虑文档在不同列表中排名情况的IASUM方法和组合的IASUM2方法，IASUM3更加直接地分析利用了文档的真实得分信息(经过分数规范化，但保留了文档原始得分的特征)。如公式(5)所示。

$$IASUM3(q,M)=\frac{1}{n_{subquery}}\sum_{i=1}^{k}score(d_i)Sum(d_i)$$
$$Sum(d_i)=\sum_{q_j=1}^{n_{subquery}}score(d_i,q_j)(1-\alpha)^{q_{jnum}} \quad (5)$$

公式(5)中各变量的含义与前文基本一致，不同之处在于$Sun(d_i)$的计算方式。$score(d_i,q_j)$表示文档$d_i$在子查询$q_j$对应的结果列表中所得分数，采用0−1规范化方法处理，进而保留分数的原始分布特征。同样地，若文档未出现则设其值为0。

## 4 实验(Experiment)

### 4.1 数据集

本文实验数据来自TREC信息检索会议(Text Retrieval Conference)在Web Track任务中提供的Clueweb09−CategoryB英文数据集。Clueweb09数据集包含超过10亿个网页，规模较大。为帮助更多不同组织参与活动，TREC为其设置了规模较小的Category B集。该数据集是Clueweb09的子集，由5000万个网页构成。

实验选取来自TREC Web Track 2009—2012年采用的查询，每年50个。由于第95和第100号查询没有相关文档，于是在实验中去除了这两个查询，共计198个查询参与实验。使用查询集合中的query域作为主查询，去除停用词和叙述词后的subtopic域作为子查询。每个主查询包含3到8个子查询，以2009年的第25号查询为例，它的主查询为“Euclid”，相应的四个子查询如图1所示。

```
<topic number="25" type="ambiguous">
 <query>Euclid</query>
 <description>
  Find information on the Greek mathematician Euclid.
 </description>
 <subtopic number="1" type="inf">
  Find information on the Greek mathematician Euclid.
 </subtopic>
 <subtopic number="2" type="inf">
  I'm looking for a source for Euclid truck parts.
 </subtopic>
 <subtopic number="3" type="nav">
  Take me to the homepage for Euclid Industries.
 </subtopic>
 <subtopic number="4" type="nav">
  Take me to the homepage for the Euclid Chemical company.
 </subtopic>
</topic>
```

图1 TREC查询示例

Fig.1 A query from TREC

对于所有的查询结果，TREC的主办者聘请专家进行人工检查。判断哪些文档和查询相关，哪些和查询不相关。在多样化任务中，也判断一个文档和哪些子查询相关。

### 4.2 评价方法

对于一组查询，采用如前所述实验平台进行检索。为每

一个查询生成相应的文档序列。采用特定的性能预测算法可以预测每一个查询的困难度，然后按照预测的困难度对所有查询排序。同时对所有查询的结果按照特定的性能指标(常用的检索结果多样化性能指标有ERR-IA@20、αNDCG@20等)进行评价，并按照评价的性能值对该组查询进行排序。通过计算这两组序列的相似程度，可以评价性能预测算法的优劣。

比较两组排序序列的相似程度，可用的指标有Pearson积差相关系数、Kendall秩相关系数、Spearman秩相关系数等。这些相关系数的取值范围为[-1,1]。如两组排序完全相同，值为1。如两组排序完全相反，值为-1。如两组排序无相关性，值接近0。值越接近1，说明两组排序越相似，预测算法的性能越好。

### 4.3 实验设置及预处理

实验采用Indri开源搜索引擎中的查询项似然检索模型对所有查询及子查询进行初始检索，其中狄利克雷平滑系数$\mu=1000$，结果列表文档数N设置为1000。并使用公开的垃圾文档排名算法Waterloo spam scorer对初始检索结果进行垃圾文档过滤。Waterloo spam scorer为ClueWeb09数据集中的每篇文档赋予了一个垃圾百分数分值，分值越低，越有可能为垃圾文档。本实验中阈值设置为70%。

下一步是对初始检索结果进行多样化重排，使用xQuAD[19]、PM-2[20]和CombSUM三种多样化算法。对于四组查询，共产生12个结果(run)。最后使用Ndeval程序对这些检索结果进行评价，并根据ERR-IA@20指标选择性能最好的检索结果和对应的平衡系数$\lambda$，产生的多样化检索结果如表1所示。

表1 对初始检索结果进行多样化重排后的结果

Tab.1 Results after diversified reranking

| Diversification | RunId | $\lambda$ | ERR-IA@20 |
|---|---|---|---|
| xQuAD | indri2009xQuAD | 0.9 | 0.191 |
| | indri2010xQuAD | 0.8 | 0.241 |
| | indri2011xQuAD | 0.6 | 0.361 |
| | indri2012xQuAD | 0.7 | 0.326 |
| PM-2 | indri2009pm2 | 0.7 | 0.175 |
| | indri2010pm2 | 0.5 | 0.241 |
| | indri2011pm2 | 0.4 | 0.323 |
| | indri2012pm2 | 0 | 0.298 |
| CombSUM | indri2009combSUM | 0.5 | 0.189 |
| | indri2010combSUM | 0.4 | 0.263 |
| | indri2011combSUM | 0.4 | 0.382 |
| | indri2012combSUM | 0.4 | 0.314 |

为评估本文所提预测方法的有效性，实验选取SD2、WIG、NQC及SMV四种传统性能预测算法作为基准。各种算法中的参数根据文献[6,16,17]的推荐而设置，WIG的截断系数k设置为5，NQC的截断系数k设置为100，SMV的截断系数k设置为100。

### 4.4 多样化性能预测方法的评估分析

使用Pearson、Kendall和Spearman相关系数评价预测算法的有效性时发现，三种系数下的实验结果整体上相似。所以这里仅列出Spearman系数对应的实验结果。表2列出了传统预测算法的预测值与ERR-IA@20指标值的Spearman相关系数，表3列出了本文所提算法的预测值与ERR-IA@20指标值的Spearman相关系数。其中黑体数字表示在所有七种预测算法中的最优值。

表2 几种传统方法的spearman相关系数(ERR-IA@20)

Tab.2 Spearman correlation coefficients of several traditional methods(ERR-IA@20)

| RunId | WIG | SD2 | SMV | NQC |
|---|---|---|---|---|
| indri2009xQuAD | 0.064 | -0.077 | 0.116 | 0.181 |
| indri2010xQuAD | 0.369 | 0.256 | 0.117 | 0.244 |
| indri2011xQuAD | -0.207 | -0.185 | -0.084 | -0.075 |
| indri2012xQuAD | 0.428 | -0.123 | -0.017 | 0.103 |
| indri2009pm2 | 0.062 | 0.084 | -0.284 | -0.274 |
| indri2010pm2 | 0.422 | 0.098 | -0.213 | -0.191 |
| indri2011pm2 | 0.206 | 0.239 | -0.088 | -0.129 |
| indri2012pm2 | 0.491 | 0.285 | -0.243 | -0.239 |
| indri2009combSUM | 0.288 | 0.252 | **0.475** | 0.451 |
| indri2010combSUM | 0.429 | 0.310 | 0.043 | 0.068 |
| indri2011combSUM | -0.013 | 0.040 | 0.022 | 0.027 |
| indri2012combSUM | 0.571 | 0.487 | 0.239 | 0.219 |

表3 三种新方法的spearman相关系数(ERR-IA@20)

Tab.3 Spearman correlation coefficients of three new methods(ERR-IA@20)

| RunId | IASUM | IASUM2 | IASUM3 |
|---|---|---|---|
| indri2009xQuAD | **0.448** | 0.104 | 0.403 |
| indri2010xQuAD | 0.572 | 0.312 | **0.576** |
| indri2011xQuAD | 0.005 | -0.213 | **0.211** |
| indri2012xQuAD | **0.681** | 0.032 | 0.677 |
| indri2009pm2 | 0.437 | 0.218 | **0.465** |
| indri2010pm2 | 0.431 | 0.132 | **0.492** |
| indri2011pm2 | 0.122 | **0.355** | 0.218 |
| indri2012pm2 | 0.455 | 0.291 | **0.496** |
| indri2009combSUM | 0.369 | 0.398 | 0.369 |
| indri2010combSUM | **0.600** | 0.502 | 0.596 |
| indri2011combSUM | -0.016 | 0.022 | **0.055** |
| indri2012combSUM | 0.508 | **0.611** | 0.578 |

从表2可以发现，每一种方法的预测性都有负相关的情况出现。所以传统的基于得分分布的预测算法在查询多样化性能预测上的适用性不强。而根据表3，可以得知本文提出的三种考虑多样性的性能预测算法优于四种基准预测算法。具体来说，与作为基准的四种预测算法相比，本文提出的三种预测算法共在11个运行结果上取得了最大的Spearman系数，而前者共在一个运行结果上取得最大值。对四种基准预测算法中进行两两比较，WIG算法的有效性最高，但与本文提出的三种预测算法相比也略输一筹。在三种新方法中，IASUM3方法的有效性最高，它优于IASUM和IASUM2的主要原因是，在综合考虑相关性和多样性的基础上，它充分地利用了结果列表中文档的得分分布信息，故在实验中共有六次取得了最好成绩。

除了ERR-IA@20外，还分别使用αNDCG@20和NRBP指标值评估所得结果列表的性能和预测的排序进行比较。图2和图3分别显示了采用αNDCG@20和NRBP时的Spearman相关系数，图中列出了三种较好的预测方法(WIG、SD2和IASUM3)的性能。

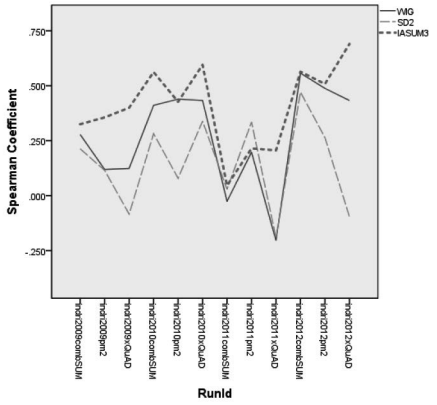


图2 WIG、SD2和IASUM3的Spearman相关系数(αNDCG@20)

Fig.2 Spearman correlation coefficients of WIG,SD2 and IASUM3(αNDCG@20)

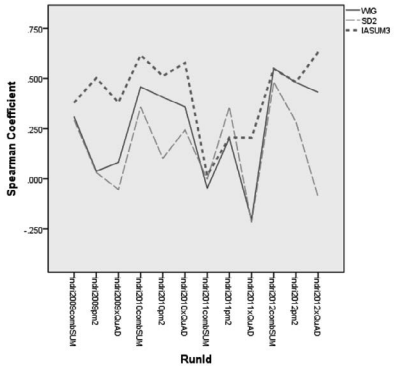


图3 WIG、SD2和IASUM3的Spearman相关系数(NRBP)

Fig.3 Spearman correlation coefficient of WIG, SD2 and IASUM3(NRBP)

总体说来，采用αNDCG@20或NRBP和采用ERR-IA@20指标效果相仿。对比图2和图3中使用不同指标测度时预测算法的有效性，可以得知在改变度量指标(αNDCG@20和NRBP)时，预测算法的Spearman系数趋势是相似的。当指标为αNDCG@20时，12个运行结果中IASUM3算法在10个运行结果上优于WIG和SD2算法，其在12个运行结果上的平均Spearman系数为0.408。当指标为NRBP时，IASUM3算法在10个运行结果上优于WIG和SD2算法，其在12个运行结果上的平均Spearman系数为0.422。另一方面，对三种指标计算所有结果的平均方差值，得到WIG和SD2的方差分别为0.226和0.199，而IASUM3的方差为0.181。这表明IASUM3算法在检索结果多样化性能的预测上，是有效且稳定的。

## 5 结论(Conclusion)

针对近些年来备受研究人员关注的查询性能预测技术，本文明确了从查询的相关性性能和多样化性能这两个方向入手研究查询性能预测，并提出了三种查询多样化性能的预测算法。为了检验提出的预测算法的有效性，在经过多样化处理的Indri运行结果上进行了查询多样化性能的预测实验。实验结果表明，和传统的预测算法相比，本文提出的预测算法IASUM和IASUM3优势明显，能有效提高多样化性能预测精度，且性能较为稳定。同时，分析文档原始得分的IASUM3算法的预测效果整体上优于分析排名得分的IASUM算法。这说明与检索结果的排名得分相比，文档得分包含了更多的文档查询相似度信息，对于预测算法而言，文档得分比排名得分具有更高的使用价值。

本文提出的查询多样化性能预测算法采用了结果列表中文档的排名与得分信息，但未考虑文档的内容，这是下一步的工作方向。我们将尝试一些基于自然语言处理的算法，以期获得更为有效的多样化查询性能预测方法。另一方面，搜索引擎在检索过程中产生了大量的搜索日志信息，并且这些日志包含了大量用户和搜索引擎的交互信息。如何有效地挖掘和分析这些数据，并用于查询性能预测，也是一项值得研究的工作。

## 参考文献(References)

[1] Zamani H,Croft W B,Culpepper J S,et al.Neural Query Performance Prediction using Weak Supervision from Multiple Signals[C].international acm sigir conference on research and development in information retrieval,2018:105-114.

[2] Mizzaro S,Mothe J,Roitero K,et al.Query Performance Prediction and Effectiveness Evaluation Without Relevance Judgments:Two Sides of the Same Coin[C].international acm sigir conference on research and development in information retrieval,2018:1233-1236.

[3] Roitman H.An Extended Query Performance Prediction Framework Utilizing Passage-Level Information[C]. international conference on the theory of information retrieval,2018:35-42.

[4] Roitman H,Erera S,Sarshalom O,et al.Enhanced Mean Retrieval Score Estimation for Query Performance Prediction[C]. international conference on the theory of information retrieval,2017:35-42.

[5] Katz G,Shtock A,Kurland O,et al.Wikipedia-based query performance prediction[C].international acm sigir conference on research and development in information retrieval,2014:1235-1238.

[6] Tao Y,Wu S.Query performance prediction by considering score magnitude and variance together[C].Acm International Conference on Conference on Information & Knowledge Management.ACM,2014:1891-1894.

[7] Raiber F,Kurland O.Query-performance prediction:setting the expectations straight[C].international acm sigir conference on research and development in information retrieval,2014:13-22.

[8] Hauff C,De Jong F M,Kelly D,et al.Query quality:user ratings and system predictions[C].international acm sigir conference on research and development in information retrieval,2010:743-744.

[9] Sondak M,Shtok A,Kurland O,et al.Estimating query representativeness for query-performance prediction[C]. international acm sigir conference on research and development in information retrieval,2013:853-856.

[10] Mothe J,Tanguy L.Linguistic features to predict query difficulty-a case study on previous TREC campaigns[C].ACM SIGIR Workshop on Predicting Query Difficulty-Methods & Applications,2005.

[11] Butman O,Shtok A,Kurland O,et al.Query-Performance Prediction Using Minimal Relevance Feedback[C].Conference on the Theory of Information Retrieval.ACM,2013.

[12] Julie A,Adrian G,Sebastien D,et al.Statistical Analysis to Establish the Importance of Information Retrieval Parameters[J].Journal of Universal Computer Science,2015,21(13):1767-1787.

[13] Cronen-Townsend S,Zhou Y,Croft W B,et al.Predicting query performance[C].international acm sigir conference on research and development in information retrieval,2002:299-306.

[14] Zhou Y,Croft W B.Ranking robustness:a novel framework to predict query performance[C].conference on information and knowledge management,2006:567-574.

[15] Cummins R,Jose J M,Oriordan C,et al.Improved query performance prediction using standard deviation[C]. international acm sigir conference on research and development in information retrieval,2011:1089-1090.

[16] Zhou Y,Croft W B.Query performance prediction in web search environments[C].international acm sigir conference on research and development in information retrieval,2007:543-550.

[17] Shtok A,Kurland O,Carmel D,et al.Predicting Query Performance by Query-Drift Estimation[J].ACM Transactions on Information Systems,2012,30(2):305-312.

[18] Carbinell J,Goldstein J.The use of MMR,diversity-based reranking for reordering documents and producing summaries[J].international acm sigir conference on research and development in information retrieval,1998,51(2):335-336.

[19] Santos R L,Macdonald C,Ounis I,et al.Exploiting query reformulations for web search result diversification[J].the web conference,2010:881-890.

[20] Dang V,Croft W B.Diversity by proportionality:an election-based approach to search result diversification[C].international acm sigir conference on research and development in information retrieval,2012:65-74.

## 作者简介：

张忠敏(1994-)，女，硕士生.研究领域：信息检索.

吴胜利(1963-)，男，博士，教授.研究领域：数据库与信息系统，信息检索.