目录

[第四章 基于时间树的搜索经验提取及查询推荐方法研究 2](#_Toc487465470)

[4.1 数据观察及问题提出 2](#_Toc487465471)

[4.1.1 数据观察 2](#_Toc487465472)

[4.1.2 问题提出 2](#_Toc487465473)

[4.2 基于时间树的搜索经验提取 2](#_Toc487465474)

[4.2.1 基于时间树的子任务划分方法 2](#_Toc487465475)

[4.2.2 子任务划分的评价方法 2](#_Toc487465476)

[4.2.3 对比实验 3](#_Toc487465477)

[4.2.4 结论 3](#_Toc487465478)

[4.3 基于时间树的查询推荐方法研究 3](#_Toc487465479)

[4.3.1 基于时间树的查询推荐方法 3](#_Toc487465480)

[4.3.2 对比实验 3](#_Toc487465481)

[4.3.3 结论 3](#_Toc487465482)

[4.4 结论 3](#_Toc487465483)

# 第四章 基于时间树的搜索经验提取及查询推荐方法研究

第三章中我们证明了时间树中蕴含着用户的搜索经验。本章将针对搜索经验中的几种研究如何提取这些搜索经验，以及如何利用这些搜索经验帮助用户更好地进行探索式搜索。

## 4.1 数据观察及问题提出

### 4.1.1 数据观察

1 提出粒度定义

2 根据粒度提出子任务

3 根据子任务定义点击切题性

4 根据切题性找出规律，提出根据位置进行子任务划分的方法。

首先我们从用户的搜索日志数据中提取出用户进行的所有查询，发现这些查询词在语义上覆盖的主题范围有大有小。例如……。基于这一观察，我们提出探索式搜索过程中查询粒度的概念，将用户在进行探索式搜索过程中进行的查询划分为三个粒度—粗粒度，中粒度以及细粒度。我们定义粗粒度查询为涵盖用户所有查询主题的查询，定义查询主题非常具体的查询为细粒度查询，将主题范围介于两者之间的查询定义为中粒度查询。

有了各级粒度的定义，我们可以将用户进行的所有查询划分为三种粒度。并在时间树上对用户进行查询的粒度进行观察，发现用户进行的查询从整体上呈下降的趋势。说明用户在完成一个任务的过程中，粒度是由粗到细的。但是从局部上看，我们发现用户在时间树上的某些位置进行查询的粒度也会有所起伏。依照前面用户在完成一个任务过程中粒度由粗到细的理论，我们提出假设，用户从粒度发生起伏的位置开始之后的某一个阶段内，是在进行某一个子任务。即用户进行复杂探索式搜索过程中，会按照子任务进行搜索。

在这一假设的基础上我们继续进行观察，发现用户在查询后进行的点击，有与源查询的主题是否一致之分。于是我们提出点击切题性的概念。并且发现用户的点击切题性有三种情况出现。大部分切题，大部分不切题，以及切题不切题数量相差不大，我们对其中的前两种进行分析。发现大部分不切题的点击，从源查询所处子任务的角度考虑，这些点击大概率仍然是属于其源查询的子任务的。因此我们考虑是否时间树的结构与用户对子任务的划分有一定联系。并且时间树与线性日志相比，本身在结构上就有蕴含因果以及主题的优势。于是我们考虑是否能够从时间树结构的角度出发，对用户进行的复杂搜索任务进行子任务划分。因此我们提出了基于时间树结构的子任务划分方法。

### 4.1.2 问题提出

## 4.2 基于时间树的搜索经验提取

### 4.2.1 基于时间树的子任务划分方法

### 4.2.2 子任务划分的评价方法

### 4.2.3 对比实验

### 4.2.4 结论

## 4.3 基于时间树的查询推荐方法研究

### 4.3.1 基于时间树的查询推荐方法

### 4.3.2 对比实验

### 4.3.3 结论

## 4.4 结论