

并行数据库的语义查询优化研究: 基于 Agent^①

许新华¹, 谌 颀², 李 春²

1. 南京信息职业技术学院, 南京 210046; 2. 天河学院, 广州 510540

摘要: 并行数据库技术中的查询处理及优化方法, 都各有优劣; 作者提出了一种基于 Multi-Agent 技术的语义查询模型(SQMAS), 并以此模型为基础建立了一种基于 Agent 的并行数据库语义查询方法; 同时为了保证系统组内、组间 Agent 之间的高效通信, 采用了树型拓扑结构(TTMAS)的通信模型, 系统内各 Agent 使用通信原语高效通信、协作, 且满足 Agent 间的通信路由最优, 从而保证了 SQMAS 的查询效率.

关键词: 并行数据库; 查询优化; 代理; 语义查询与代理系统; 树型拓扑多代理系统

中图分类号: TP181; TP311

文献标识码: A

当前并行数据库系统的研究基本围绕着关系数据库进行, 其查询处理及优化方法, 都各有优劣; 而语义查询是一种完全不同于传统的基于查询树^[1]的查询技术, 它利用关系上定义的各种完整性约束条件, 通过修改, 把给定的查询变换为更有效的等价查询.

1 语义查询优化方法

语义查询优化方法不是简单摒弃传统的查询优化方法, 它通过查找与给定查询有关的完整性约束条件, 修改给定的查询为更有效的等价查询, 从而获得更有效的查询执行策略.

2 语义查询 MAS 系统的结构模型

本语义查询 MAS(Semantic Query MAS, 以下简称 SQMAS)采用多 Agent 技术, 其中的各 Agent 既有思考型、反应性, 也有混合型结构. SQMAS 包括以下 Agent 组和 Agent:

1) 查询监测组(QueryDetectGroup), 成员有: 查询接收 Agent(QueryReceive-Agent); 特征获取 Agent(FeaturesGet-Agent); 知识形成 Agent(KnowledgeShape-Agent).

2) 内核组(KernelGroup), 成员有: 任务规划 Agent(TaskPlan-Agent); 任务管理 Agent(TaskManage-Agent).

3) 知识库组和规则集组(KDB&RuleSetGroup), 成员有: 知识库 Agent(KDB-Agent); 规则集 Agent(RuleSet-Agent).

4) 输出组(ExportGroup), 成员有: 约束条件输出 Agent(ConditionExport-Agent).

此外, 还有如下的 Agent: 命名服务 Agent(NameServer-Agent); 辅助通信 Agent(AssCommu-Agent, 如 GGA、MASGA).

系统中的各种 Agent 作为一个相对独立的主体存在于一个复杂的环境中, 对外界(如并行数据库环境)来说, 它主要执行两个过程: 查询接收 Agent(QueryReceive-Agent)接收外界信号, 约束条件输出 Agent(ConditionExport-Agent)做出反馈. 语义查询 MAS 系统的结构如图 1.

① 收稿日期: 2006-12-14

作者简介: 许新华(1968-), 男, 湖北孝感人, 副教授, 主要从事计算机网络, 数据库的研究.

系统内核(System Kernel)主要包括处理器与知识库和行为规则集,一般由知识系统、问题求解程序、知识描述知识(元知识)、系统状态知识组成,反映 Agent 对内部问题和领域问题的处理能力.

2.1 命名服务 Agent 和辅助通信 Agent

命名服务 Agent(NameServer-Agent):该 Agent 拥有一个全局性的共享知识交换区,负责 MAS 全局的系统服务,管理 MAS 的 Agent 注册表中所有 Agent 的信息,主要的操作对象是与 Agent 管理和维护有关的 Agent 注册信息表.

辅助通信 Agent(AssCommu-Agent):包括各组内单元通信门户(GGA)、系统级通信门户(MASGA),其主要功能是帮助 Agent 在通信时进行规范化信息的生成或分解.

2.2 查询监测组(QueryDetectGroup)和约束条件输出组(ExportGroup)

二者主要负责控制 Agent 的行为,它最一般的过程是监测、接收信息、交由处理器进行信息处理、产生并执行一个任务流系列和约束条件输出.二者被看作是 Agent 联系周围环境如并行数据库系统的桥梁,体现了 MAS 的交互和社交能力.

1)查询监测组(QueryDetectGroup):负责监测、接收并处理并行数据库的查询条件(WHERE 子句),并最终形成知识级信息输出供内核组(KernelGroup)使用,其组内结构如图 2 所示.包括①查收 Agent(QueryReceiver-Agent);②特征 Agent(FeaturesGet-Agent);③知识形成 Agent(KnowledgeShape-Agent);④知识库和规则 Agent(KDB&RuleSet-Agent).

2)约束条件输出组(ExportGroup):该组作为与并行数据库环境交互的末端,主要是把从任务管理 Agent(内核组,KernelGroup)得到的约束条件通过约束条件输出 Agent(ConditionExportAgent)反馈给并行数据库环境从而完成语义查询优化的任务.

2.3 内核组(KernelGroup)

内核是系统的底层结构,包括以下组成部分:

1)内核 Group:MAS 系统的核心,主要由任务规划 Agent(TaskPlan-Agent)、任务管理 Agent(TaskManage-Agent)和规则集 Agent(RuleSet-Agent)构成,其结构如图 3 所示.

2)知识库和行为规则集 Group:知识库和行为规则集包括元知识、运行状态知识和行为规则集.

3 语义查询 MAS 的通讯模型设计

本系统由查询监测 Group、内核 Group、知识库和行为规则集 Group、输出 Group 组成,其中每个 GROUP 在所属的 MAS 里是唯一标识的,每个 GROUP 中包含至少一个 Agent,而每个 Agent 在所属的 MAS 内也是被唯一标识的. Agent 必须具有通信的能力才能执行任务,本 MAS 系统采用人工智能领域中经典模型—黑板模型结构(black board architecture, BBA)^[2]描述多 Agent 系统之间的通信方法.本系统需要解决 3 类通信,即 GROUP 内的单播、多播和广播通信;不同 GROUP 间的单播、组播、广播和选播通信;单个 Agent 向 MAS 的广播.

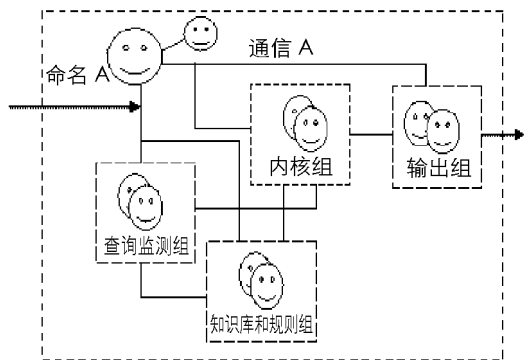


图 1 语义查询 MAS(SQMAS) 系统的结构

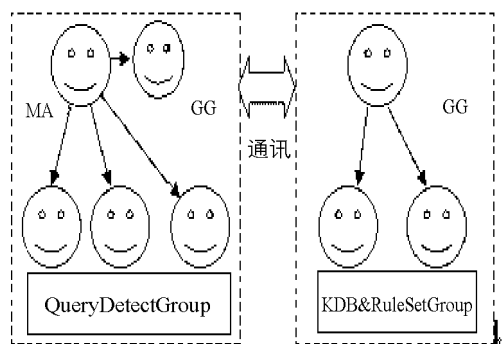


图 2 查询监测组、知识库组与规则集组的组成及通信

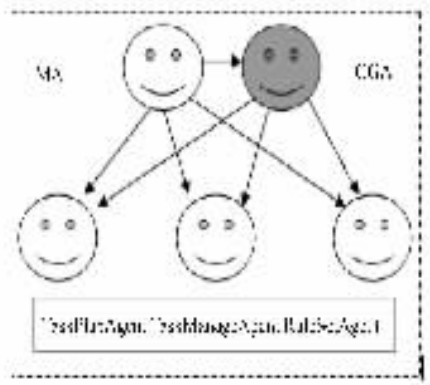


图 3 内核组(KernelGroup)的组成

3.1 GROUP 内的单播、多播和广播通信

下面给出这些 Agent 的详细说明: 黑板模型结构是 Agent 通信模型的基础, 它可以为 Agent 通信模型提供 Agent 的并发控制、Agent 的交互、Agent 的实时控制等服务, 黑板模型结构中不同知识源替换为 Agent, 并且加入 1 个单元(组)门户 Agent(group gateway agent, GGA), 可得到 Agent 通信单元模型(CU), 如图 4 所示^[3].

同一个组内各 Agent 间发送消息的通信原语^[4]为:

Send_Message(Sender, Receiver, Message)

1) 当 Agent 进行单播通信时, 通信原语为:

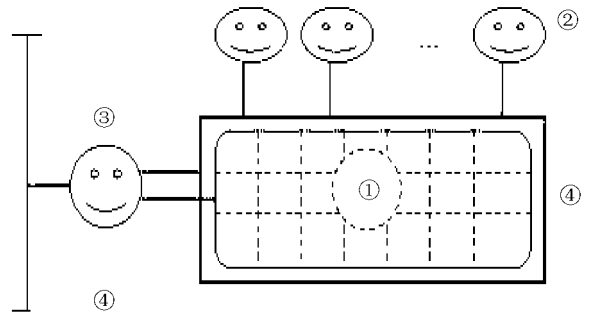
Send_Mesaage(Agent A₁, Agent A₂, Message)

2) 当 Agent 进行多播通信时, 通信原语为:

Send_Message(Agent A₁, Multicast(A), Message)

3) 当 Agent 进行广播通信时, 通信原语为:

Send_Message(Agent A₁, Group(default), Message)



①共享知识交换区 ②普通 Agent
③单元门户 Agent ④通信总线

图 4 Agent 通信单元模型

3.2 不同 GROUP 间的单播、组播、广播和选播通信

MAS 内对于 Agent 的组织不仅要满足一个智能系统需要的恰当的层次结构, 还应满足 Agent 间的通信路由最优, 求解问题的过程实质上是把问题分解和细化的过程, 因此, 本系统构造了一个树形拓扑的多 Agent 系统(tree topology multi-agent system, TTMAS).

TTMAS 通信模型如图 5 所示^[3]. 整个通信模型呈树状拓扑结构, TTMAS 中根节点是命名服务 Agent(NameServer-Agent). 该 Agent 拥有一个全局性的共享知识交换区, 它为不同 MAS 间的通信起纽带作用; 树中的其他节点是区域门户 Agent(Commnication Area Gateway Agent, CAGA), 它实现通信区域(CA)间的通信路由以及消息队列的控制. 在 TTMAS 中每个节点 Agent 都拥有自己区域的共享知识交换区, 并且都能起到信路由的作用, 所以, 该通信模型能够实现系统全局的通信服务.

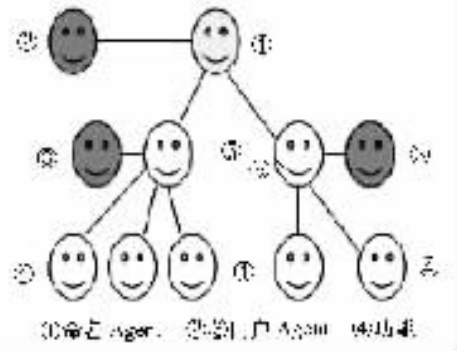


图 5 TTMAS 通信模型图

系统的不同组中各 Agent 间发送消息的通信原语^[4]为 Send_Message(Sender, Receiver, Message)

1) 当不同组间 Agent 进行单播通信时, 通信原语为 Send_Message(Agent A₁, Group(B); : Agent A₁, Message), 其中: Group(B)是 MAS 内名称为 B 的通信单元; 符号(:;)为通信单元内的域控制符;

2) 当不同组间 Agent 进行多播通信时, 通信原语为 Send_Message(Agent A₁, Multicast(B), Message), 其中: Multicast(B)={B₂, B₃, ..., B_k}; B_i(i=2, 3, ..., k)为 Group B 中任意 Agent;

3) 当 Agent 进行不同组间的广播通信时, 通信原语为 Send_Message(Agent A₁, Group(B), Message), 其中, Group(B)是名称为 B 的通信单元或者组;

4) 当 Agent 进行选播通信时, 通信原语为 Send_Message(Agent A₁, Group(B); Any, Message), 其中, Group(B); : Any 为 Group B 中功能相似的任意一个 Agent.

3.2 单个 Agent 向 MAS 的广播

当 Agent 对 MAS 进行广播时, 通信原语为 Send_Message(Agent A₁, ALL, Message), 其中 ALL 为 MAS 内所有 Agent.

4 语义查询优化的理论分析

试验用的数据库包括以下 3 张表:

1) 借书证登记表 CARDS(CARDNO, NAME, DEPARTMENT, PROFESSION), 其中的属性分别表

示借书证号、借书人姓名、单位、职业；

2) 图书登记表 BOOKS(BOOKNO, CATALOG, BOOKNAME, KIND, AUTHOR, PRICE, POSITION), 各属性人别表示图书编号、分类、书名、种类、作者、价格、摆放位置；

3) 借书登记表 BORROWS(CARDNO, BOOKNO, DATE), 表示某人某时借了某书。

假设有如下几个完整性约束^[5]：

规则 R1: BOOKNAME='数据库系统基础' → CATALOG='TP', 表示若书名为‘数据库系统基础’, 则该书的分类为‘TP’。

规则 R2: PROFESSION='teacher' ↔ CARDNO < 100000, 表示教师的借书证编号小于 100000, 借书证号小于 100000 的人是教师。

规则 R3: KIND='foreign' → PROFESSION='teacher', 表示只有教师才能借外文书。

使用语义查询, 改写查询条件, 分以下几种情况:

(1) 逻辑化简

当查询条件中的某个谓词与某一完整性规则相矛盾时, 可将该谓词替换为 FALSE。

设有 Q1: 查询借书证号在 100000 到 200000 之间的教师名单、借书证编号。

```
SELECT NAME, CARDNO FROM CARDS
```

```
WHERE PROFESSION='teacher' AND CARDNO BETWEEN 100000 AND 200000;
```

由规则 2 可知, 教师的借书证编号小于 100000, 故上式 WHERE 子句的逻辑值为假(FALSE), 即 Q1 有一个等价查询 Q1':

```
SELECT NAME, CARDNO FROM CARDS WHERE . F.
```

返回结果为空, 所需时间几乎为零。

(2) 引入索引^[5]

引入索引, 即在搜索条件中增加新的具有可用索引特别是唯一性索引、聚集索引的谓词。

设有查询 Q2: 查找“数据库系统基础”这本书的摆放位置。

```
SELECT POSITION FROM BOOKS WHERE BOOKNAME='数据库系统基础';
```

利用规则 R1, 我们可以得到等价的查询 Q2':

```
SELECT POSITION FROM BOOKS WHERE CATALOG='TP' AND BOOKNAME='数据库系统基础';
```

设在属性 CATALOG 上建有索引, 而在 BOOKNAME 上未建立索引。利用索引执行查询时, Q2' 和 Q2 的执行效率的差别是明显的。

(3) 消除部分悬浮元组^[6]

所谓“悬浮”元组, 是指查询结果中不存在来自该元组与其它元组连接的成份。

设有查询 Q3: 查找借外文书的人员的姓名和单位。

```
SELECT NAME, DEPARTMENT FROM CARDS, BOOKS, BORROWS
```

```
WHERE CARDS. CARDNO = BORROWS. CARDNO AND BOOKS. BOOKNO = BORROWS. BOOKNO AND KIND='foreign';
```

利用规则 R3, 则可以得到新的查询 Q3':

```
SELECT NAME, DEPARTMENT FROM CARDS, BOOKS, BORROWS
```

```
WHERE PROFESSION='teacher' AND CARDS. CARDNO = BORROWS. CARDNO AND KIND='foreign' AND BOOKS. BOOKNO = BORROWS. BOOKNO;
```

通过消除 CARDS 中的部分元组, 能够减少中间结果。

(4) 消除连接中多余的关系

当一个关系仅参与连接, 且它的属性上无选择条件、参与连接运算的属性被连接谓词中另一个属性所引用, 该关系是可以消除的。设有查询 Q4: 查询教师借阅书籍的总数。

```
SELECT COUNT(BOOKNO) FROM CARDS, BOKROWS WHERE CARDS. CARDNO = BORROWS. CARDNO;
```

其等价查询 Q_4' 如下:

```
SELECT BOUNT(BOOKNO)FROM BORROWS WHERE CARDNO<100000;
```

(5) 模拟实验

二次模拟实验的主要差别在于图书登记表 BOOKS 容量, 第一次有 2 万个元组, 第二次有 5 万个元组, 实验的结果见表 1.

理论分析和模拟对比试验表明, 利用 SQMAS 模型设计的语义查询优化算法能够显著地提高查询效率, 对于中间结果可能出现 100000 条元组的查询操作而言, 采用语义查询, 整体运行时间可以节约 30% 以上; 并且可以推测, 关系元组数目越大, 查询效率越高.

表 1 二次 SQMAS 模拟实验数据

第一次实验	费时/s	$T(Q')/T(Q)/\%$	第二次实验	费时/s	$T(Q')/T(Q)/\%$
Q_1	5.1		Q_1	10.1	
Q_1'	0.01	0.2	Q_1'	0.01	0.1
Q_2	4.1		Q_2	12.4	
Q_2'	2.8	68.3	Q_2'	7.9	63.7
Q_3	10.8		Q_3	32.3	
Q_3'	8.2	75.9	Q_3'	17.2	53.3
Q_4	2.1		Q_4	4.7	
Q_4'	1.5	71.4	Q_4'	2.9	61.7

参考文献:

- [1] 李建中. 并行数据操作算法和查询优化技术[J]. 软件学报, 1994, 5(10): 13-25.
- [2] Yilmaz C, Daytona B F L, Orlando F L. A dynamic knowledge exchange tool for intelligent agents. The Third CLIPS Conference at NASA Johnson Space Center, September, 1994: 96-103.
- [3] 王 斌, 张尧学, 陈松乔. 一种基于黑板模型的多 Agent 系统通信方法[J]. 小型微型计算机系统, 2002, 23(11), 1355-1358.
- [4] 王 斌, 张尧学, 陈松乔. Agent 通信模型[J]. 中南工业大学学报, 2002, 33(2): 209-213.
- [5] 许向阳, 冯玉才, 王元珍. 语义完整性在查询优化中的应用[J]. 小型微型计算机系统, 1996, 17(10): 35-39.
- [6] 许向阳. 查询语义优化[J]. 微型机与应用, 1998, (3): 53-57.

Agent-based Query Optimization in Parallel Database

XU Xin-hua¹, CHEN Hang², LI Chun²

1. Nanjing College of Information Technology, Nanjing 210046; 2. Tianhe College, Guangzhou, 510540

Abstract: The query processing and optimization methods in parallel database technology all have their merits and shortcomings; then the authors proposed one kind semantic query model based on Multi-Agent technology (SQMAS), and taking this model as the foundation, they have established one kind parallel database semantics query method based on the Agent, and to ensure the high efficiency of communications between the Agents of the system's groups, they have also adopted a communication model by Tree Topology Multi-Agent System(TTMAS), every Agent in this system communicates and collaborates at high efficiency by using communication primitive, thereby ensuring the query efficiency in SQMAS.

Key words: the parallel database; query optimization; agent; SQMAS; TTMAS