云知道：基于云计算的大规模社会智能推理模型*

胡浩宇，王寅峰，王卓立，狄盛

[1](香港大学 计算机系，香港)

Sensible Cloud: Cloud based Large-scale Social Intelligence Reasoning

Haoyu Hu[1+], Yinfeng Wang[1], Cho-Li Wang[1], Sheng Di[1]

[1](Department of Computer Science, the University of Hong Kong, Pokfulam Road, Hong Kong SAR, China)

+ Corresponding author: Phn +852-28578449, Fax +86-**-****-****, E-mail: hyhu@cs.hku.hk

Abstract: With increasing usage of pervasive computing and more and more social relations among users on social websites, it is crucial to guide users to make wise decisions based on their experiences such that they could adapt to the unfamiliar environments as quickly as possible. This paper presents the design of BetterLife 2.0 framework, which facilitates implementation of large-scale social intelligence applications in cloud environment. Based on case-based reasoning, this framework makes use of HaDoop File System (HDFS) and HBase to implement an efficient management and analysis on case-reservoir. By transplanting such a case-based reasoning engine jCOLIBR2 as well as its algorithms to CNGrid platform, our system can provide users with real-time query services. Moreover, the Betterlife2.0 users could submit their query requests at any time and from anywhere through mobile devices. Experimental results show that such a Cloud-based model could improve the efficiency up to 5 times compared to traditional solutions, and it can also support millions of user requests to get high scalability.

Key words: CNGrid; Cloud Computing; Social Intelligence; Case based Reasoning

摘 要: 随着普适计算应用的日益推广，越来越多的用户通过社交网站建立关联，因此如何在此基础上有效地共享用户经验并引导用户做出明智的决策是辅助用户迅速适应陌生环境和高效处理问题的关键。设计了一种基于云计算的支持大规模社会智能推理的框架 :Betterlife2.0。该框架通过基于案例的推理方法(Cased-based Reasoning)采用 Hadoop 分布式文件系统（HDFS），Hadoop 数据库系统（HBase）等技术，实现高效的案例库管理与分析。通过将案例推理引擎 jCOLIBR2 及其相应算法移植到中国国家网格(CNGrid)，该系统能利用网格资源为用户提供实时的社会智能搜索服务。 Betterlife2.0 用户还可在任意时刻任何地点通过手机等移动终

* This research is supported by China 863 grant 2006AA01A111 and Hong Kong RGC grant HKU7176/06E. (863 国家网格计划香港节点); Hong Kong UGC Special Equipment Grant (SEG HKU09) (香港教资会专用设备建设基金)

作者简介: 胡浩宇(1984－),男,湖北武穴人,学位(或目前学历),职称,主要研究领域为内容传递网络,云计算;王寅峰(1974－),男,博士,职称,主要研究领域为高维案例推理模型及算法; 王卓立(1961－), 男,博士,副教授,博导,主要研究领域为网格计算,云计算,普适计算等; 狄盛(1981－), 男,山东济南人, 博士生,主要研究领域为自治模型,资源调度优化算法和云计算.

端访问 CNGrid 网格平台的网格服务。 测试结果表明该云计算模式的计算效率利用案例推理方法比传统方法提高近 5 倍，并可以支持数以百万计的用户推理请求，具有良好的可扩展性。

关键词: 中国国家网格; 云计算; 社会智能; 基于案例的推理

1 引言

Web 2.0 的广泛应用使得 Internet 从一系列网站转变到一个为最终用户提供复杂在线应用的服务平台。以用户为中心的社交网站能够极大的促进网络上人与人之间的信息交换和协同合作，例如，Facebook [1], Yelp [2], Amazon[3]与 Netfix [4]将用户从独立的小圈子融入到大规模社交网络的交际模式已经获得了巨大成功。社会智能（Social Intelligence）作为一种理解各种情境下人的行为以及在此基础上引导适当行为的计算模式，涉及一系列的知识、用户经验和如何解释群体与个体信息以及如何处理陌生环境的问题。社会智能的最终目标是实现和谐人际关系的智能推理以及“以人为本”的行为指导。

随着计算机技术的深入发展，从 GPS 定位技术，RFID 标签到移动终端，普适计算已经渗透到社会中的每一个角落，基于情境感知的移动终端应用已经在社交网络中深受欢迎。譬如，基于用户轨迹挖掘的智能位置服务可以提供基于多人轨迹数据的大众旅游推荐，通过汇集大量用户的轨迹数据还可以实现个性化朋友和地点推荐。虽然使用大规模的位置信息与大众化旅游推荐可以找出经典的旅游线路，但无法实现 “以用户为中心”的自适应智能分析效果。事实上，不同用户有不同的偏好，相同用户在不同的时间地点以及环境中也会做出差别迥异的选择。因此，针对用户喜好进行量身订做的推荐系统不仅能丰富人们之间的沟通与选择，还可以更加有效的加强个人与环境的互动和适应能力。人们在不知不觉中使用服务时将会留下大量的个人与环境间的行为记录和反馈，如用户在就餐后通过手机登陆 Yelp 对餐馆的口味进行评价，而这些信息对其他 Yelp 用户具有重要的参考价值。

随着人们越来越乐于共享信息，在线社区、社交网站等能收集到越来越多的用户数据和经验。这些数据携带了群体间复杂的行为信息，并能改变系统对行为或组织的理解。通常人们愿意吸收他人经验的程度取决于相互间的关系，而这种关系将最终影响人们的决策。传统的社会智能理念主要集中在如何识别出交互中的规则、名称与模式，如协议，策略等，以便引导出符合当前社会情形的行为。尽管这种认知方法在语言学与人工智能中起了一定的作用，然而，将其应用到人际关系分析时还会遇到很大的限制和挑战。

具体而言，面对海量信息，用户很难准确的判断哪些是有用的信息，当他们需要某些参考信息时，必须人工的进行查找，如果要获取相对可靠和准确的信息则需付出更多的努力。不难看出，在一个社会网络中，n 个用户可达 n^2 种交互关系，将这种超大计算复杂度的数据分析工作交给高性能计算机利用多节点并行执行是保证系统响应时间的有效手段。因此，CNGrid 计算平台所特有的快速处理海量信息的能力可以提高知识发现与信息识别的效率，为云计算模型所需提供的超大规模的存储与计算服务（即支持“任何时候、任何地点”均可用的普适计算应用）构成了良好的基础。

基于案例的推理技术（Case-Based Reasong，简称 CBR）采用的方法是根据已有案例和经验找出相似问题的解决策略。该方法已经在情景感知的推荐系统中得到了广泛应用，特别适用于在不清楚领域知识的情境下处理有参考案例的复杂决策问题。而且，基于相似性分析的数据检索是通过比对分析算法解析历史的问题与解决方案，帮助用户以尽可能少的代价做出准确的决策。当需要处理不断累积的案例库与历史信息时，CNGrid 平台下的云计算数据处理必须具备良好的可扩展性以面对成指数增长的案例数据。

总而言之，Betterlife2.0 系统的目标是利用 CNGrid 的计算优势提供可扩展的推理框架以实现对个人日常生活中订制推荐服务的优化处理。它同时基于 MapReduce 技术 [5]，采用案例推理算法和社会网络分析策略对海量数据进行大规模分析和处理，为移动终端用户提供更快更好的智能在线决策服务。本文后续内容包括相关工作分析，Betterlife2.0 系统设计详述，并通过一个典型场景对系统的性能进行分析，最后对本研究工作

进行总结与展望。

2 Betterlife 2.0 体系结构

Betterlife 2.0 系统结构如图 1 所示，从上倒下包括三层：1) CNGrid 应用接口；2) 云计算层；3) CNGrid 基础设施。用户的智能搜索请求首先通过移动设备或门户网站提交到任何一个 CNGrid 网格社区的服务器，该服务器利用 CNGrid 基础设施和其他服务器进行通信，并将请求在多个社区的不同集群服务器间共享，最终利用云计算层中的基于 MapReduce 的案例推理模型（CBR）完成信息的查询和聚合：即通过 MapReduce 算法的 Map 模块将请求提交到各个 CNGrid 服务器节点，然后使用 Reduce 模块处理从各个服务器返回的结果。

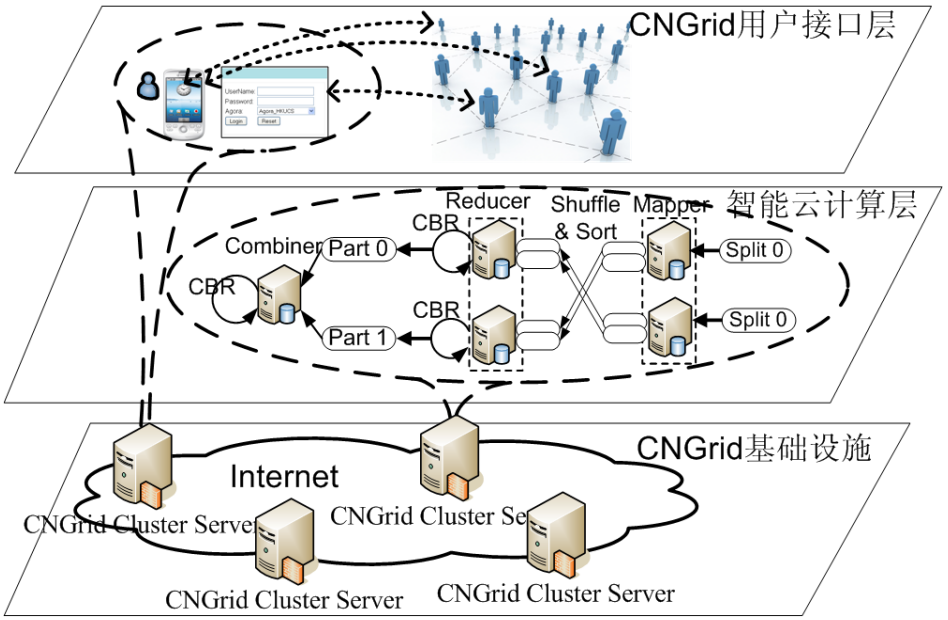


图 1：Betterlife 2.0 体系结构

具体而言，CNGrid 应用接口层为用户提供方便的分布式访问接口，负责用户请求的分发和结果聚合。系统提供两种客户端的访问接入：Web 客户端——可看做扩展的社交网站以方便用户数据生成概览；移动客户端（例如手机终端）——方便用户随时随地上载并修改个人情景信息，发起请求或接收系统相应的推荐信息。云计算层采用 Hadoop 的分布式文件系统（HDFS） [6]，存储以案例和社会网络信息组成的应用系统数据，包括社会关系拓扑结构和社会关系接近程度。云计算层可以避免由于记录急速增长的用户在线交互行为（即大量 I/O 操作）而造成响应时间的显著延迟。因此，原来由于复杂推理所形成的影响响应时间的瓶颈将转变为：如何优化 MapReduce 功能与计算节点的负载均衡以保证访问服务质量（Quality of Service，简称 QoS）。我们的系统扩展实现了 jCOLIBRI2 [7]的 CBR 引擎（云计算层的核心模块），并通过计算案例间的相似度以检索最相似的案例，最终使用 HBase [8]方法将新增案例进行存储。CNGrid 基础设施层通过整合广域分布的集群计算机，利用部署在集群上的 CNGrid Web 服务接口实现 CBR 引擎的协同操作。本文主要介绍云计算层中的基于 MapReduce 的案例推理模型和社会网络分析算法，CNGrid 应用接口和 CNGrid 基础设施可参阅相关文献 [9]。

3 CBR 智能云计算

3.1 基于MapReduce的案例推理模型

在将通用的 CBR 系统 jCOLIBRI2 移植到 Hadoop 的 MapReduce 框架时，首先必须扩展 jCOLIBRI2 以保证案例可以存储在 Hadoop 的 HDFS 系统上。一个典型的 CBR 过程包括：

- 检索：给定目标问题，找出与其匹配的案例以简化求解的过程。通常一个案例中包括对问题的描述，解决方案和求解的过程。
- 重用：将检索出的解决方案映射到目标问题上，一般需要修改已有的方案来适应新的情况。
- 修订：完成解决方案映射与修改后，测试该方案是否可行还是需要继续完善修改。
- 维护：将成功适用的方案与问题作为经验存储在案例库中，使案例库的覆盖度随系统的使用逐渐增大，判断效果愈来愈好。

Betterlife 2.0 的推理引擎与 jCOLIBRI2 的区别如图 2 所示。jCOLIBRI2 的推理需要先将案例库装载到内存中，再开始 CBR 的处理过程。这种方式存在如下两个问题：

(1) 计算能力依赖于单一服务器，当案例库规模非常大时将消耗大量的处理时间。

(2) 单一服务器的内存受限，基于 JVM 的处理过程在大量请求时不能保证系统的性能。

Betterlife 2.0 的推理引擎使用基于 MapReduce 的并行推理模式，其可扩展性与易维护性都有很大的提高。解决方案的修订与维护主要由用户完成，例如用户可以通过手机反馈对系统推荐方案的评价。因为 Hadoop 能够很好的存储大量的案例，Betterlife 2.0 的推理引擎将主要关注于保障响应及时性的要求。

Betterlife 2.0 的推理引擎在 HDFS 上的工作流如图 2 所示。

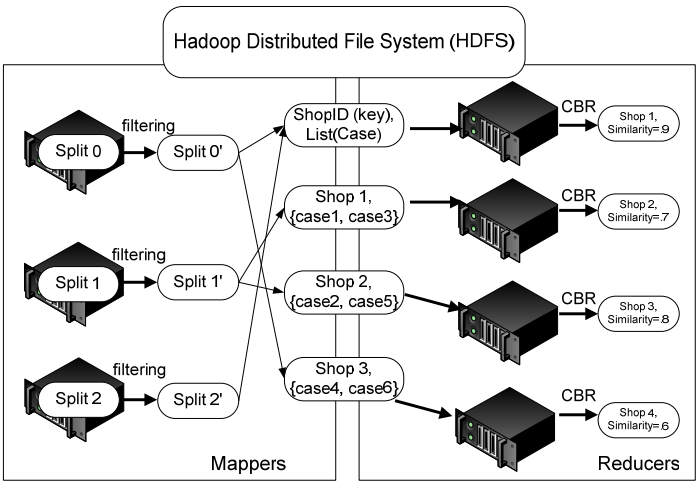


图 2：Betterlife 2.0 的推理引擎工作流

Map 功能包括检索与过滤这两个部分（以下以用户查询商品信息为例进行描述）。

在进行相似性案例检索时，每一个 mapper 均从本地读取数据以避免大量的数据传输影响 MapReduce 的操作。当某个节点失效时，HDFS 通过在其它节点上复制数据重新运行 Map 功能。在检索完成后，临时输出结果以一系列字符串的形式表示。过滤阶段对这一系列数据再进行解释，抽取 UserID, 产品咨讯(如条形码)等信息。无关的数据（如不匹配的）和过期的数据将被过滤掉。过滤后的数据将用来直接计算相似度或附加上对应的社会关系接近程度。这一系列数据将被写入中间输出结果作为 Reducer 的输入。在典型购物应用中，用户只关心“最好”的那一个商店，所以我们将 ShopID 作为 key 而 Mapper 中的其它数据作为 value。

每个 Reducer 将收到一个作为 key 的 ShopID 与一组有相同 ShopID 的案例。这些案例与请求案例之间的相似性度量采用最近邻居算法(*kNN*)，相似度函数定义如下：

$$Similarity(N,P)=\sum_{i=1}^{n}Sim(N_i,P_i)*W_i$$

$$Distance(N,P)=1-Similarity(N,P)$$

$$\sum_{i=1}^{n}W_i=1$$

其中 N_i 是新请求案例 N 的第 i 个属性的值，P_i 是已有案例 P 的第 i 个属性的值，n 是案例中相关属性的个数。$Sim(N_i, P_i)$是针对第 i 个属性的本地相似值。*Similarity(N, P)*是通过计算所有属性的加权平均值来度量两个案例间的全局相似性，根据已知的领域知识设定权重，每个属性的权重取值在[0,1]之间。距离函数*Distance(N, P)*是计算新请求案例 N 与已有案例 P 之间考虑在所有属性上的整体相似性。在本文的应用中，我们定义了如下四种度量相似性的函数：

位置相似：基于 GPS 的位置相似性由公式(1)计算，*MaxDistance* 是在一定范围内预先定义的两点间最大距离，当 $Distance(N_{gps}, P_{gps}) > MaxDistance$ 时，相似性为 0。

时间相似：我们假设个体的行为模式虽然每天都相似，但在具体的一天当中发生的时段变化很大，其相似性由由公式(2)计算，$Diff(N_t, P_t)$计算在一天中以分钟为粒度的相对间隔。

社会关系相似：在评价案例中个体之间的接近程度时，个体间的关系越接近则相互影响力也越大，如公式(3)所示。由于个体间可能处于多个社会群体中互相之间存在多种互连的关系，需要找出其中连接权重最大的关系以计算相似性(相互影响力)。

价格相似：案例中价格的相似性由 McSheery 的“Less is better”公式计算，与新案例的请求价格无关。为使 Reducer 的返回结果简洁，我们设定过滤掉相似度小于 0.8 的案例。

3.2 智能社会网络分析算法

一个合理的推荐系统需要了解个体的社会关系，当所有的案例按照用户分组后，他们之间的关系就构成了一个带权图如图 3 所示。我们使用广度优先(BFS)的搜索策略以保证取得全局最优解，使用 MapReduce 根据关系距离公式找出 *k* 个最近节点(案例)。

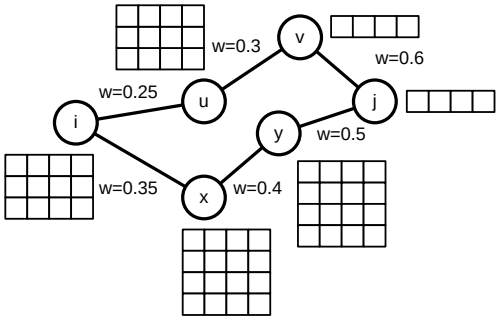


图 3：社会关系近似度广播图

案例 *C* 在某个节点上的第 *i* 次关系接近程度计算如下：

$$\begin{aligned} Distance(N,P) &= 1-W_{gps}*Sim(N_{gps},P_{gps}) \\ &- W_t*Sim(N_t,P_t) \\ &- W_{price}*Sim(N_{price},P_{price}) \\ &- W_{social}*\max_{p(i,j)\in P(i,j)}\prod_{e(u,v)\in p(i,j)}w(u,v) \end{aligned}$$

$$Distance(C,Q)_{i+1}=c-W_{social}*w_1*w_2*\ldots*w_i*w_{i+1}$$

$$Distance(C,Q)_{i+1} = c-(c-Distance(C,Q)_i)*w_{i+1}$$

算法 1 和算法 2 分别表示用 Map 与 Reducer 的处理过程，其中所有节点在开始时设置为白色(WHITE)，当运行到某个节点时，该节点标记为灰色(GRAY)，处理过的节点置为黑色(BLACK)。

算法 1: SocialNetworkMapper (key *k*, Node *n*)

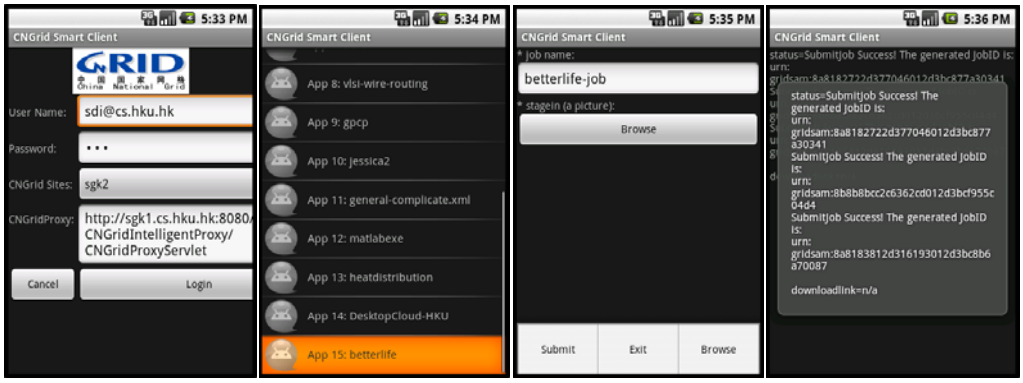
```
1.  if n.color == GRAY then
2.      for all edge e of Node n do
3.          vnode.distance ← 新节点 (e.ToID);
4.          vnode.color ← GRAY;
5.          word ← vnode.Id;
6.          输出 <word, vnode>;
7.          n.color ← BLACK;
8.      end for
9.  end if
10. word ← n.Id;
11. Emit < word, n>
```

算法 2: SocialNetworkReducer (Key *k*, Iterator *V*)

```
1.  distance ← MAX;
2.  color ← WHITE;
3.  edges ← NULL;
4.  for all Node u ∈ V do
5.      if u.edges.size > 0 then
6.          edges ← u.edges;
7.      end if
8.      if u.distance < distance then
9.          distance ← u.distance; /*保存最小距离*/
10.     end if
11. end for
12. Node n ← 新节点 (k);
13. n.distance ←distance;
14. n.edges ← edges;
15. 输出 <k, n>;
16. if color == GRAY then
17.   reporter.incrCounter(counters.MOREGRAY, 1);
18. end if
```

4 移动网格接入

我们将整个 Hadoop 的应用封装成网格系统后面的应用，同时制作了一个 Android 平台上的手机客户端，从而让用户能方便的随时随地提交 Context 的信息。通过调用在 CNGrid 网格社区服务器上部署的应用代理，对用户提交的请求进行分布式的调度，实现不同社区服务器资源的共享和联合查询。该设计充分利用了网格系统的资源优势和移动客户端的方便性。下图显示了 CNGrid BetterLife 应用的手机客户端界面。



5 性能测试与分析

CPU	2 x Intel Quad-Core E5540 Xeon CPU, 2.53GHz, 8MB cache
Memory	16GB DDR3 memory, 1066MHz, dual ranked UDIMMs
Storage	2 x 250GB 7.2K RPM SATA hard disks, running in RAID-1
Network Interface	Broadcom 5709 dual-port
OS	Fedora 11

图 1 实验环境节点配置

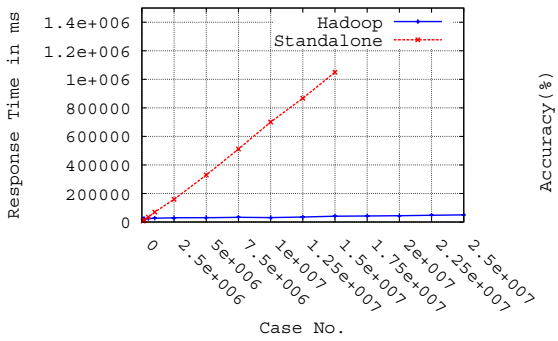


图 2 Case 数目对查询时间的影响

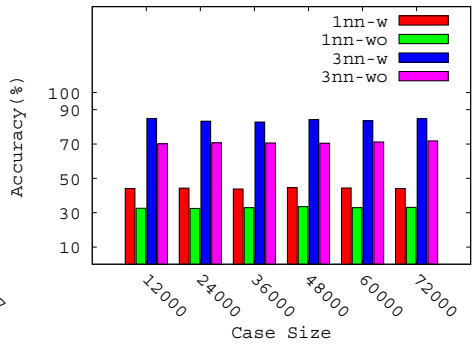


图 3 不同 k 和 社会信息下的准确率比较

我们实现了整个 BetterLife 2.0 架构并通过一个基于位置信息的价格比较购物应用来衡量系统的查询响应时间和基于社会信息的搜索准确度。该应用能让移动手机用户拍下所需购买物品的条形码和图片，然后提交搜索，让网格系统调用后台的 Hadoop 程序进行大规模的比较匹配，从而能找到附近商店的该物品，同时又推荐考虑和用户有相关兴趣的其他用户，从而返回更相关的购物推荐。这个应用被部署在 HTC Magic Android 手机上。该实验旨在比较独立的机器和 Hadoop 框架下的 CBR 推理的性能，并比较有和没有社会网络关系的分析时 CBR 推理的准确型。

我们用合成数据集在各种案例库的大小下进行压力和响应时间测试。对于性能，我们测量的时间不包括 Android 的客户端和服务器之间应用程序的接口查询/结果通讯时间。对于系统评估的准确性，我们使用 10 倍交叉验证方法用于三样本数据集加权 KNN 算法。在实验中，设定 K = 1 和 3。对于 k = 3，该解决方案出现在最好的 3 个全局最相似的案件被认为是正确的。我们的系统里已注册 103 个账户，并模拟着 103 个用户

的行为活动，包括对产品的评论和订阅，加入团体以构成一个小虚拟社区，从而形成历史案例。测试结合了香港 7 - Eleven 便利店的真实地点和该 103 个用户的社会网络拓扑结构。为获得足够多的具有不同 Context 的用户案例，测试采用了预先定义的规则（如位置集群，产品集型产业集群，集群时间）对用户的行为进行模拟。对于每一组数据，随机选择其中 1%的案例进行手工验证，验证解决办法的合理性。 Hadoop 的集群由多达 16 个计算（Slave）节点操作和一个 Master 节点。

(A) 查询响应时间

首先比较独立执行情况和采用 5 个节点的 Hadoop 集群的执行效率。该测量集中分析推理引擎的执行速度。图 2 比较了 jCOLIBRI2 推理引擎（独立机）和基于 Hadoop 的 BetterLife 2.0 响应时间（毫秒）。案例个数为 2500K，单机需要 159 秒的响应时间，而 Hadoop 则只需 29 秒，执行效率提高了 5 倍。当继续增加案例规模时，单机甚至不能完成推理工作。特别是，我们的测试显示主内存为 16GB 的机器也只能容纳 15000K 个案例。这主要是由受限的 JVM 堆大小造成的。然而，在 Hadoop 框架的支持下，推理引擎不但可以运行，而且可以支持 25000K 的案例，且其响应时间几乎只有 50 秒。

HDFS 的试验结果是正面的。在从 2 个 Slave 节点启用一个 400MB 数据写的情况下，HDFS 的需要 23 秒才能完成。平均 I / O 速率约为 17.4MB / s。读操作也显示了类似的结果。另一项测试，在 15 个节点写 3GB 数据，它仍然只需要 23 秒才能完成，平均 I / O 速率是 130 MB / s，这表明 Hadoop 是在大型数据 I / O 和硬件资源的增加下几乎可以线性提高性能。另外应当指出，MapReduce 有一个启动延迟的调用发生。这个问题可以通过在线 MapReduce 方案部分解决。另一种 CBR 的性能提高方法是应用特定领域的索引，这样不会扫描所有在数据集的案件。 Hadoop 可以用来进行脱机操作的索引工作，这是不要求用户参与的工作流程。

(B) 社会网络分析效果

图 3 显示了查询的准确度提高与社会网络信息的关系，采用 K-NN 邻近算法(K = 1 和 k = 3)。当 k = 3，在这两种情况下的精度都是令人满意的（至少 70％）。对于这两种 K = 1 和 k = 3 时，考虑到社会关系（标记为由 1nn-w 和 3nn-w）后，结果精度提高多达 10％。这是由于在数据合成时，我们模仿一些虚假用户在网站提供的产品评级。由于这些用户一般与其他用户社会亲近率低，他们的 Case 不太可能被应采用和检索，即使考虑到他们的产品信息的价格是低廉的。这反映了有些人打算通过垃圾信息来促进和推广自己的产品。对于测试的准确性，我们还发现，用 1-NN 的准确性检索方法并不如预期的高。 总之，我们已经显示了 BetterLife 2.0 准确性的显着改善。与单机 jCOLIBRI2 框架相比， BetterLife 2.0 可以支持可伸缩的推理引擎，而响应时间是在可接受的水平（考虑案例库的大小是 25000K 大），实际的处理查询的时间是在 30 秒内，这是相对短的。如果启动成本可以进一步缩短，并利用其他 Online MapReduce 的解决方案，结果会更理想。

6 相关工作

6.1 基于案例分析的大规模推荐系统

一般的推荐系统大致可分为四类：基于个人偏好的推荐、基于相似用户的推荐、基于物品的推荐以及混合式推荐。典型的协同过滤算法在基于物品的推荐中被大多数的大型商务网站(eBay, Amazon, Netflix 等)所采用。通过其它用户对物品的评价预测该物品对相似用户的效用。A.Das [10] 提出了一个基于协同过滤算法与 MapReduce 确保性能的实时 Google 新闻推荐引擎。Zhao [11] 在 Hadoop 上实现了一个面向用户的协同过滤算法以解决可扩展性问题。Zhang [12] 提出了从个人及大众用户轨迹数据中进行数据挖掘，根据用户喜好做个性化推荐以定制移动服务。传统的 CBR 推荐系统在面对大量案例处理时往往出现性能严重下降的问题，采用分治策略的 MapReduce 方法能够很好的支持大规模数据处理，可被用来有效的扩展 CBR 推荐系统。

6.2 智能社会网络分析算法

相关研究表明通常网络中的相似性是由网络中人们的影响力或交互作用所产生的。Leskovec 讨论了信息的 cascade 现象，个体由于大众的影响而采纳新的活动或思想。在一些极端情况下，如音乐下载网站中的排行榜常决定了其它成员的行为。

传统的确定社会网络中节点接近程度或关系重要性的方法是采样与调研，但是大规模社会网络中的结构化特征很难从小规模的网络中推断出来。Berscheid [13] 提出了 Relationship Closeness Inventory (RCI) 度量个体与其社会组织的关系相互依赖的程度，通过计算共享的时间，活动的差异程度与影响力的强弱推断接近的程度。Aron [14] 认为接近程度可以通过一个完整的认知过程，即描述一个个体处在另一个个体概念中的程度，使用 Inclusion of Other in Self scale (IOS) 的 7 个温氏图表示关系从完全无关到完全重叠的情况。

但是这两个理论没有说明社会关系接近程度的传递性问题。由于接近程度通常反映了两者间的互利关系，如何有效地度量社会关系接近程度以保障有意义的交互行为是大规模社会智能研究中的一个重要问题。Computational Social Science [15] 在收集社会网络信息的规模，深度与广度方面对评价社会关系接近程度提出了很高的计算要求。H. Karloff [16] 论文表明 MapReduce 对运行于 Parallel Random Access Machine 上的算法性能提升有很大的优势。Tang [17] 采用 MapReduce 对社会影响在大规模网络中的传递的扩展性进行了分析。我们采用广度优先的单源最短路径 [18] 方法来计算社会关系的接近程度。

不同其它研究关注于在特定领域中提升推荐算法的准确度，Betterlife 2.0 致力于提供一种易维护、易扩展、真实有效的通用推荐服务机制以满足人们在日常生活中各种各样的推荐请求。而且，Betterlife 2.0 还考虑了历史数据的可信度，通过评估个体在社会网络中的关系给出可信建议。一些推荐系统在处理海量信息时性能会大幅下降，Betterlife 2.0 则采用云计算上的基于案例推理的模式能很好的处理可扩展性问题。

7 总结

本文设计了一种基于网格平台和云计算模式的支持大规模社会智能推理的框架：Betterlife2.0。该框架通过基于案例的推理方法采用 Hadoop 分布式文件系统（HDFS）技术和 CNGrid 计算平台，实现高效的案例库管理与分析。本文采用的社会网络分析算法利用 Mapreduce 实现对用户社会关系在 Hadoop 文件系统中的实时处理。相对于传统的推理引擎，本文优化的算法能够显著的降低用户请求的处理时间。此外，通过支持移动终端访问 CNGrid 的网格服务，Betterlife2.0 拥有更好的灵活性和可扩展性。未来的工作包括基于 Betterlife2.0 框架设计更多的交互式应用，例如交通流量监控、旅店推荐等，也会整合更丰富的上下文信息以优化系统的精度。

References:

[1] Facebook website: http://www.facebook.com/

[2] Social networking based searching web site: http://www.yelp.com/

[3] Amazon (US-based multinational electronic commerce): http://www.amazon.com/

[4] TV Shows & Movies Online System: http://www.netflix.com/

[5] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun. ACM, 2002. 51(1): 107-113

[6] HDFS Project: http://hadoop.apache.org/hdfs/

[7] Recio-Garc´ıa J, Bridge D, D´ıaz-Agudo B, and Gonz´alez-Calero P. CBR for CBR: A Case-Based Template Recommender System for Building Case-Based Systems. In: ECCBR08, ser. Lecture Notes in Computer Science, K.-D. Althoff, R. Bergmann, M. Minor, and A. Hanft, Eds., 5239(1), 2008: 459–473.

[8] HBase Project: http://hbase.apache.org/

[9] CNGrid Project: http://en.wikipedia.org/wiki/CNGrid

[10] Das A, Datar M, Garg A, and Rajaram S. Google news personalization:scalable online collaborative filtering. In: International World Wide Web Conference (WWW), 2007, 271-280.

[11] Zhao Z and Shang M. User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop. In: International Workshop on Knowledge Discovery and Data Mining. IEEE Computer Society. 2010, 478–481.

[12] Zheng V W, Cao B, Zheng Y and Xie X and Yang Q. Collaborative Filtering Meets Mobile Recommendation: A User-Centered Approach, Journal of AAAI, 2010: 236-241.

[13] Berscheid M. Measuring closeness: The relationship closeness inventory (rci) revisited. In The handbook of closeness and intimacy, LAWRENCE ERLBAUM ASSOCIATES LTD. Erlbaum, 2004. 81–101.

[14] Aron A, Aron E, and Smollan E. Inclusion of other in the self scale and the structure of interpersonal closeness. Journal of Personality and Social Psychology, 1992, 63(4):596–612.

[15] Lazer D, et al. Computational Social Science. Science, 2009, 323(Feb):721–723.

[16] Karloff H. A Model of Computation for MapReduce. Time, 2010, 938–948.

[17] Tang J, Sun J, Wang C, and Yang Z. Social influence analysis in large-scale networks. International Conference on Knowledge Discovery and Data Mining, 2009, 807–816.

[18] ElGindy H and Pang C Y. Two Parallel Algorithms for Shortest Path Problems. In: Proc 1980 Conf on Parallel Processing, Computing Inform. Science, University of Pennsylvania. 1980, 3(Sept):244–253.