

基于 Multi-agents 系统的分布式数据挖掘^{*}

庄艳 陈继明 徐丹 潘金贵

(南京大学计算机软件新技术国家重点实验室 南京 210093)

摘要 计算机网络的发展以及海量数据的分布式存储,滋生了分布式数据挖掘(DDM)这一新的数据挖掘方式。本文针对多 agent 系统下的分布式数据挖掘进行了初步的研究,对 agent 方法用于 DDM 的优势、基于 agents 的分布式数据挖掘的问题,以及典型的基于 agent 的分布式数据挖掘系统和该领域的进一步研究方向作了一个概要的综述。

关键词 数据挖掘,分布式数据挖掘,基于多 agent 系统的分布式挖掘

Distributed Data Mining Based on Multi-agent System

ZHUANG Yan CHEN Ji-Ming XU Dan PAN Jin-Gui

(State Key Lab for Novel Software Technology, Nanjing University, Nanjing 210093)

Abstract The development of network and the storage of huge data in a distributed way bring on the distributed data mining (DDM). The article gives a primary study focus on the Distributed Data Mining Based on Multi-agent system. We summarize the advantages of agents for DDM, problems in the agent-based system for distributed data mining, and some representative agent-based Distributed Data Mining systems, at last, the future work of the area.

Keywords Data mining, Distributed data mining, Data mining based on multi-agent system

数据挖掘是用于在大规模数据集中获取感兴趣知识的过程。传统的数据挖掘系统^[1]设计大多是面向于集中式数据集进行的。然而随着网络的发展,大块的数据分散位于不同地理位置的主机上。而对于这些分布式数据的挖掘就导致了分布式数据挖掘的产生并受到越来越多的关注,并且网络有限的带宽和计算资源也推动了分布式数据挖掘的研究和发展。在分布式数据挖掘中,数据集是由若干个局部数据库组成,它们通过计算机与网络相连而形成。因而可以看出,分布式数据挖掘可以在本地和全局两个层面上进行,其简单结构如图 1 所示^[2]。

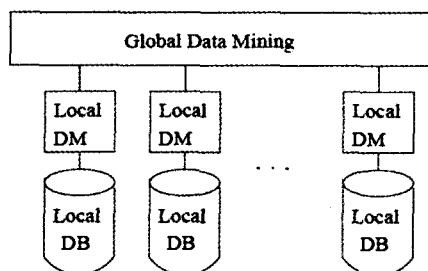


图 1 分布式数据挖掘结构图

文[3,4]提出分布式数据挖掘希望在各自的站点上得到部分结果,然后把该结果发送到某些站点来进行聚合,最终得到全局的结果。

智能 agent 是为了完成其目标,能够灵活自治其行为的个体。利用 Agents 的灵活性、自治性、协作性、自适应等特性,MAS(Multi-Agent System)广泛地应用到了分布式环境中。根据订购的基本思想,文[7]提出以 MAS 作为基本平

台,构架了 Open Agent Architecture (OAA 框架)。将每个用户功能可以抽象为一个 agent,可以向系统提供服务,也能向系统的其他 agents 申请服务。而系统中使用 facilitator agent 来管理任务的分派。因而从提供方和请求方来说,它们之间是透明的,它们之间的消息传递就是类似于一种 blackboard 方式。提供方只需要把所能提供的服务公布在“facilitator”上,而请求方亦只需把申请公布于此。请求和服务的匹配则由 facilitator 来完成。又如文[8]提出利用 MAS 来进行电子商务行为,它使得系统能够同时考虑到用户的兴趣和其用于浏览网站的设备。文[9]提出利用 MAS 来进行 3D 虚拟社区的设计。

由此可见,Agent 在分布式环境中的应用已经得到了发展,并在解决分布式问题上产生了一些实际的效果。本文主要讨论如何采用 agent 技术来实现分布式数据挖掘^[10,11],包括分布式数据挖掘的特点和它面临的问题;多 agent 技术在数据挖掘领域的研究,主要提出其应用优势、主要问题、解决方案以及当前该领域的应用前景。

1 分布式数据挖掘的实用性

(1) 数据存放的分散性

分散在各个物理站点的数据本身就具有分散特性。如果将它们集中存放于某个站点,使用集中式的方式进行数据挖掘,可以想象到数据量和通信开销都是相当大的。而且从其他站点传送来的数据由于受到安全和自治性限制,根本不可能发送到中心站点。而在分布式方式下,可以让数据在本地进行处理,之后对部分结果进行综合来获得最终结果。

(2) 克服集中方式下资源受限的瓶颈

^{*} 本课题得到国家自然科学基金项目(60473113)、国家自然科学基金重点项目(60533080)资助。庄艳 硕士研究生,主要研究领域为分布式虚拟环境、Agent 技术;陈继明 博士研究生,主要研究领域为 XML、分布式虚拟环境;徐丹 硕士研究生,主要研究领域为分布式虚拟环境、Agent 技术;潘金贵 教授,博士生导师,主要研究领域为多媒体信息处理、多媒体远程教育系统。

随着网络数据量的不断增大,对这些数据进行集中式分析,会耗费巨大的计算资源,甚至超出某个站点的计算能力,从而无法对这些数据进行全体的分析。而在分布式方式下,丰富的计算资源为其提供了很好的扩展空间。

(3)超越单一的数据挖掘方法的限制^[12]

由于单一的数据挖掘方法无法满足要求,而如果重新设计针对于该数据源的方法又会增大开销,因此在现有集中式方法下,采用协作的方法来融合多种挖掘方法进行数据分析更加有效。能够把集中式下有效成熟的数据挖掘算法,应用于分布式场景下,这时只需要考虑它们上层的协作和综合问题,不需要再重新设计算法,故很好地实现了方法的可重用性。

2 分布式数据挖掘的关键问题^[13]

(1)同类数据与不同类数据:集中式数据挖掘下,大部分工作都是面对的同类数据,也就是在同一个平台上运行的相同的数据库管理系统,并使用同样的模式进行管理。若是不同类数据,则需要将局部管理模式集成为全局管理模式,而后再进行处理。否则,很可能导致属性之间的矛盾。

(2)数据复制:本地数据库中的数据根据需要,可能被复制到其它站点上。这样就会引起数据一致性问题。在本地的修改无法被相关站点及时得知,从而引起了挖掘结果的偏差。

(3)通信开销:集中式数据挖掘中,算法需要考虑的只是I/O和CPU的时间开销。而在分布式环境下,还需要考虑通信开销。而这一开销是受限于带宽和在网络上发送的信息数量的,因此需要为分布式数据挖掘重新建立不同于集中式的开销模型。

(4)结果整合:是分布式数据挖掘中关键的一步。对于分布式数据挖掘来说,总的目标是在各个站点上完成分析得到部分结果,而后再基于部分结果进行整合,得到整个系统的全局结果。对于数据集合在各个站点上的分析,则可以利用现有的集中式的数据挖掘方法进行。而最后基于部分结果的整合,不同于简单的集中式方法。例如在本地感兴趣的样品在全局范围内可能失去兴趣,因而在分布式数据挖掘中,需要从所有站点上收集它们的局部兴趣并进行全局兴趣度验证来得到全局的结果。

(5)动态环境下的多变性:传统的数据挖掘方法视数据为静态的,在一个拥有充足数据的环境中执行。而在现实中,尤其是分布式环境下,数据是变化的,执行环境也是动态的,因而在这种情况下,数据挖掘很难正确地传递时序的结果和与时序相关的结果。

3 Multi-agents 技术在分布式数据挖掘中的研究

3.1 多 Agents 技术在分布式数据挖掘中的优势

多 Agents 技术应用于分布式数据挖掘中有如下几个优势。

(1)Agent 的自治特性:对应于数据源的自治特性,agent 可以按照本地访问限制和安全策略来进行对本地数据的访问,进行不同数据源上信息的协作,从而加强了私有信息的保护。

(2)Agent 的主动性:限制了用户监管和对运行数据挖掘过程的干预。用户只需要在任务初始阶段给予 agents 设定的目标和方法,在运行过程中,由 agents 来进行对任务执行过程的调整。

(3)Agent 的自适应性:动态地选择数据源和数据的搜集。分布式环境或者现实环境的一个重要问题就是环境的变化,导致了数据源发生变化。在此情况下,agents 可以根据预先设定的规范——期望的数据量、类型以及质量等等,来进行数据源的搜索和选择。这样的 agents 也可以成为在动态环境下为静态数据挖掘方法搜索数据源的基本工具,从而使得静态的方法能够对动态数据进行分析。

(4)Agent 的协作特性:扩充传统的数据挖掘方法,使得它们能够适应分布式环境下的大块数据;实现多种技术合作的数据挖掘工作;整合来自于各个局部站点的结果,获得全局结果。

3.2 关键问题

对于利用 agents 来进行数据挖掘,在本地 agents 所进行的操作就是局部数据的挖掘工作,是一种静态的集中的数据挖掘方式,因而可以使用传统的集中式数据挖掘方法进行。问题的关键就是在多个 agents 获得了局部结果之后如何进行整体结果的整合,即多个 agents 在数据挖掘过程中的协作问题。

(1)制定局部目标即选取各地的数据模型:必须基于样本给予每个 agent 挖掘的目标。如果采用集中式的方式来进行对所有数据源的挖掘,则根据数据源特点选取整体数据模型即可。然而,现在需要对分散在各个站点的数据源进行,因而出现了目标的分散,需要对各个站点上的数据建立合适的模型,使得在得到部分结果之后,能够构成较好的整体模型。例如,在局部站点上 agents 数据挖掘后感兴趣的内容,在全局来看,却是兴趣度比较低的。因而在挖掘之前,基于全局数据来对 agents 制定目标是获得较好的挖掘结果的基础。

(2)如何协作与知识整合:在挖掘过程中,不同数据源的 agents 如何利用各自的本地信息,进行协作,从而得到较为准确的全局结果。也就分成了是在学习之前、之中还是之后进行协作的三种方式。协作时发布信息的多少,协作方的选取,也影响了结果的整合方式。同时也涉及到了通信量等问题。这一部分涉及到了全局结果和局部结果生成之间的关系。

(3)减少通信量和保证精度:局部站点的兴趣度和全局的兴趣度不同,在一般情况下,我们期望得到全局的感兴趣结果。局部站点的 agents 将其结果发送给相关站点进行全局结果的分析,即使这种分布式数据挖掘的方式已经一定程度上减少了数据源在网络上的流动,然而过多无用的局部结果也会导致传输量的增多,因而在制定好目标的基础上,需要对局部站点上所得到的结果进行进一步精简来减少局部无用结论的传递,也可以通过挖掘过程中的协作来舍弃一些多余结论。

3.3 解决方案

由 agents 在各个数据源进行工作,得到基于本地数据源的结果。基于这些局部结果来进行知识的整合,从而得到最终结果。现提出以下几种知识整合方式^[14]。

(1)使用专门的管理 agents 来负责整合所有站点的部分结果。局部结果是 agents 基于对其本地数据源的考察而得出的,因而在整合时需要用其他数据源的数据来考察是否这些局部结果能够满足全局要求。而这就需要从其他站点传来大量的数据用于考察,也可以使用移动 agents 携带结论到各个站点进行验证,然而,这些方法都会导致网络中通信量的加大。

(2)组整合。将 agents 分组来整合它们的局部结果。这

是一种分块的方法,使得结果的验证在相关的小范围内进行,筛选掉一部分,而后再在组和组之间形成更高一层的组,进行结论的验证。这样,随着层次的升高,其结果也就更趋向于全局结果。而这其中,由于多个组同时进行验证,可以减少一部分无用的局部结果。但这种方式,又需要考虑到 agents 之间即数据源之间的分组标准,需要将相关的数据源分到一组进行挖掘。如果将相关性较小的分到一组,很可能会筛减掉全局感兴趣的知识。

(3)共享结论,在本地进行整合。agents 在本地数据源进行工作时,利用其它站点已生成的知识,来产生本地数据源的有效信息。例如站点 A 上的 agents 形成了“车”这种概念。那么在 B 站点上,agents 碰到了“公交车”这个概念,就可以利用 A 站点的概念来进行解释。从而不需要 B 站点重新对“车”来进行描述。这种方式在每个节点都会产生全局结论,而不需要采取另外的方法来获取全局结果。问题在于,每个站点都形成全局结论是没有必要的,造成了重复工作。

综合上面的几种方式,现给出一种综合的知识整合方式:基于共享结论的理念,agents 之间是一种知识的共享模式。利用管理 agents 来记录已有的知识和其发布的站点。当某个站点的 agents 本地挖掘时需要用到知识时,先去查看是否已有站点提供,如果有,则直接从发布站点上获得结论,进而进行下一步工作;若没有,则在本地进行挖掘。如果中间出现管理 agents 存储的结论过多的问题,则可以采用分块的方法,将相关的数据源的挖掘分到一组,提供管理 agent 进行知识的记录,其他 agents 则是在各个数据源上进行挖掘工作。同时利用管理 agents 在各个块之间的信息交流来实现各个块之间的知识共享。根据相关性进行的分块,则是希望减少管理 agents 之间的通信,形成较好的全局结果。具体的图解如图 2 所示。



图 2 基于知识共享的不同站点 agents 的知识整合模型

图 2 所示模型是一种 blackboard-based MAS,因而数据挖掘考虑的问题可以集中在分块和数据挖掘技术的选择和目标制定上,而分布式的协作和通信则可以经由这种 MAS 构架的研究来进一步发展。同时,OAA^[15]的系统构架也是基于发布订购模式下的服务共享方式,也可以用于这种系统的基构架。

4 Multi-agents 用于分布式数据挖掘的现状

本节将介绍当前几种基于 agents 的数据挖掘系统,说明它们如何利用 MAS 来解决分布式数据挖掘的问题。

4.1 基于异构站点的集合式数据挖掘系统——BODHI^[16]

BODHI 是基于 CDM 思想构建的系统,CDM(Collective

Data Mining)提出一种构架,在这种构架下对多个异构站点进行分布式数据分析,在保证全局结果正确的情况下,减少网络的通信量。

基本思想:任何方法都可以被表示成由多个基本方法组成的集合。学习的目标即是产生一个方法,它是由多个基本方法组成的多项式,而这一方法逼近于所给定的数据集合。对于给定的一个基本方法集合,则需要给出多项式中的系数(即表现各个基本方法对表达该数据集合的作用)。

CDM 描述:①从整体上根据数据模型的类型来选择合适的正交数据模型,即按照 CDM 所期望的方式选择合适的基本方法集合;②在每个本地站点上使用基本模型得到基本正交系数;③将本地模型进行合并(也即是多项式相加的结果),将模型转换为用户需要的规范模型表示输出。

BODHI 实现:在 BODHI 系统中提供了消息交换和运行时环境即 agent 工作站,用于移动 agent 在各个站点的执行。挖掘的工作被委派到本地的 agent 工作站。移动 agents 携带它们的状态、数据和知识在各个 agent 工作站之间移动。一个核心 facilitator 则负责系统初始化和数据挖掘工作的协作问题,以及 agents 之间的交流和控制。系统利用本地的 facilitator 来进行多 agents 之间的协作。从 facilitator 与用户的交互获取挖掘目标,而后分派给 agents 来进行协作执行。当 agents 需要向其他数据源进行操作时,就带着自身状态和已获得知识,移动到其他站点,继续工作。而 agent 工作站则给了 agents 运行的一个运行时平台,用于进行 agents 与站点的交流以及多站点之间的交流。其系统结构如图 3^[16]。

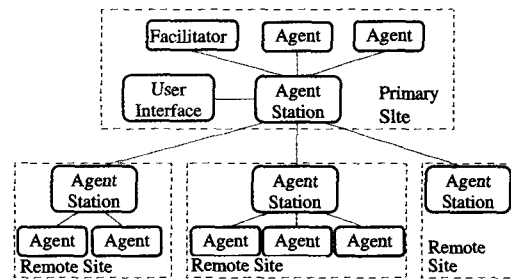


图 3 BODHI 系统的框架图

评价:最后结果的生成方法减少了大量的通信数据。而且本地模型的处理,将有助于全局模型的导出。利用 agent 工作站的方式屏蔽了底层与站点的交互细节;利用 facilitator 进行协作的管理,而不需要全局的 agents 管理;agents 在移动到其他站点工作时,携带有已获得的知识;用户可以通过建立和删除 agents 来实现数据挖掘方法的添加和删除。然而,需要基于整体数据集合给出正交模型,这一决策的制定直接关系到最后结果是否准确。而且这种模型的表述能力,能否满足需要也未能给出证明。

4.2 可扩展的分布式数据挖掘系统——PADMA (Parallel Data Mining Agents)^[17,20]

针对问题:这种分布式数据挖掘系统是基于 agents 信息处理构架和并行计算两个方面。目标是为某个应用背景提供灵活的并行的基于 agents 的数据挖掘服务。

系统构架由三部分组成:数据挖掘 agents;用于协调 agents 之间协作的 facilitator;用户接口。数据挖掘 agents,用于针对于某个感兴趣领域对数据进行挖掘操作。在 PADMA 的描述中是使用了 text 的分析和分类;Facilitator,数据挖

掘 agents 之间的并行和共享信息是通过此来实现的。同时也负责用户和数据挖掘 agents 之间的信息交互; PADMA 提供的用户接口是基于 Web 的用户图形接口。用于为用户显示 agents 获取的信息以及用户对数据挖掘的需求。用户向 facilitator 发送挖掘请求, 而后 facilitator 将此请求广播到各个 agents 上; 由 agents 从本地数据中挖掘符合于请求的信息返回给 facilitator; 最后, facilitator 将收集到的信息返回给用户。

基于 MAS 的并行数据挖掘方法: 对于用户提出的 query, facilitator 广播到各个站点的 agents 进行处理; agents 根据 query 的要求在本地构成聚类, 而后将结果发送给 facilitator; facilitator 根据 agents 结果进行进一步聚类, 从而形成全局结果, 返回给用户。各个 agents 则是在本地进行挖掘, agents 之间不进行通信, 它们只是根据用户的 query 需要对本地的数据进行聚类, 而后将结果返回给 facilitator 进行全局的聚类。具体结构如图 4 所示^[17]。

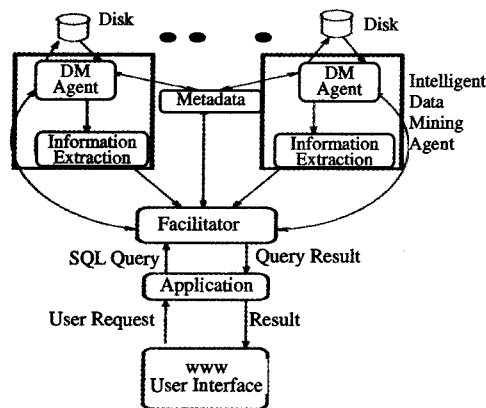


图 4 Architecture of PADMA

注: 其中的 metadata 是在本地无法聚类的数据, 然而又是与 query 相关的, 因而期望转交到 facilitator 上进行进一步站点之间的聚类。

评价: 减少了全局的控制, 不需要为各个站点建立各自的模型。各个站点都是基于同一个模型来生成结果, 因而在全局得到结果的时候, 只需要基于这些部分结果再进一步聚类就可以了。减少了 agents 之间的通信, agents 的工作集中在了本地数据挖掘上, 减少了网络的通信量; 多 agents 同时在多个站点进行数据聚类, 提高了并行效率; 利用 facilitator 来进行最后结果的进一步聚类从而得到了全局结果, 即减少了数据源的传递, 而使用聚合局部结果的方式来得到全局结果; 最重要的是由于多个 agents 之间是独立的, 因而其系统扩展性比较好, 新的数据源的加入只需要添加 agents 用于其上服务, 然后将结果通知 facilitator 即可。但由于各个站点上的数据源可能具有相关性, 而不是仅仅本地的相关度比较大, 因而在进行挖掘的时候, 这些相关数据集中在一起得到的聚类结果优于分别在各个站点进行聚类后, 再在 facilitator 上进行综合的结果。这种情况, 在该系统下就无法得到较好的解决。系统认为直接对数据源的聚类只是发生在本地, 站点之间的聚类则需要由本地的聚类结果再去聚合。

4.3 中间学习的分布式数据挖掘系统——JAM^[18]

核心思想: 将携带各种数据挖掘方法的 agents 分散到各个站点进行挖掘, 利用 Meta-Learning 方式来将结果连接成全局结果。

JAM 构架: 由多个数据站点 (Datasites) 组成。每个数据

站点由本地数据库、learning agent、meta-learning agent、本地用户配置文件和图形用户接口组成。具体的描述如下: ①本地 learning agents 在本地数据库上进行操作, 计算数据站点的本地分类。不同的 agents 具有各自的数据挖掘技术, 通过添加和删除 agents 来实现对方法的增减。②Meta-learning agent 是可以移动与各个站点之间, 用于综合各个站点的分类并与自己的分类进行合并。这些行为都是在站点间同时独立发生的。③数据站点的管理者即可以通过用户配置文件来对本地的操作进行控制。包括参数、所能使用的本地数据库名、安全机制、选择的本地 learning agents 等。而且管理者也可以通过用户接口来控制动态的学习过程和 agents 之间的信息交换。

工作方式描述: 按照最终结果的生成, JAM 系统采取了两种方式, 即集中式合并得到全局结果以及分散式在本地形成全局的结果。过程如下:

(1) 各个 agents 在本地站点上利用一定的数据挖掘算法, 获得本地分类结果。

(2) (集中式) 各个站点将结果发送到中心站点进行整体分类结果的计算; (分散式) 各个站点将其本地分类结果发送到系统各个其他站点。

(3) (集中式) 基于 DB* (选取的各个节点的典型数据, 能够表达系统中所有数据源的特征数据集) 来利用学习算法对获得的各地的分类结果计算, 得到全局结果。 (分散式) 各个站点获得来自于其他站点的所有局部结果, 而后选取本地数据库中的部分数据 T, 利用学习算法对所有的局部结果进行计算, 得到本站点的对于全局结果的理解。

(4) (集中式) 将在中心站点获得的全局结果发送到各个站点, 作为数据特征结果, 来对该站点接收的数据进行过滤; (分散式) 各个站点基于本地获得的全局结果对进入的数据进行过滤。

评价: 利用文[19]提出的 meta-learning agents, 来实现集中式整合结果和分散整合结果, 灵活性很强。就这两种方式来说, 集中式更利于从全局整体来看待系统数据, 从全局上获得系统结果; 而对于分散式来说, 其在获取全局结果的同时, 更侧重于本地具体情况的影响, 因而更有利于全局结果在本地实际数据源的应用。同时, meta-learning agents, 使得本地分类时还考虑其他站点的分类结果对其的影响, 从而可以很好地利用各个站点间数据的关联性, 能够协同工作, 减少无关局部结果的产生。而对于本系统来说, 如何确定 DB* 和 T, 直接关系到结果, 同时考虑其他站点对本站点分类影响时, 需要选取恰当的站点进行协同 (数据相关性与本站点不大的站点, 对本地的结果影响不大)。

结束语 在利用 MAS 来进行分布式挖掘中, 主要问题还是在于减少网络的通信量以及集中式的数据处理量。上面的几个系统都采用了相应的方式来进行考察, 然而依旧还存在着各自的问题。主要的发展方向在以下几点:

(1) 分布式数据挖掘如何进行整体知识的认识, 从而外派给各个子节点以挖掘任务 (即根据期望挖掘的数据源类型、特点等信息选取合适的表示方式, 用以外派给各个子节点具体的挖掘子任务)。

(2) 多个数据源之间的数据关联性影响到了是否需要将它们协同考虑, 而协同时采用何种策略 (例如数据源之间有无主次之分)。

(3)在利用 agents 进行数据挖掘时采用何种协作方式来获得较好的全局结果。

(4)局部节点的知识是否都能用于全局结果的构成,如何消除其中全部不感兴趣因素(这也就是如何给各个节点以挖掘目标的问题)。

致谢 本文的形成得益于周志华教授的《数据挖掘》课程的学习以及他的指导。

参 考 文 献

- Han J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000
- Fu Yongjian. Distributed Data Mining: An Overview [B]. IEEE TCDP newsletter, Spring, 2001
- Park B H, Kargupta H. Distributed Data Mining: Algorithms, Systems and Applications. Data Mining Handbook, 2002
- Provost F. Distributed Data Mining: Scaling up and Beyond. In: Kargupta H, Chan P, eds. Advances in Distributed and Parallel Knowledge Discovery AAAI Press / The MIT Press, 2000
- Nwana H S. Software Agents: An Overview. Knowledge Engineering Review, 1996, 11
- Jennings N R, Wooldinge M J. Agent Technology-Foundation, Application and Markets. Springer, 1997
- Martin D L, Cheyer A J, Moran D B. The Open Agent Architecture. A Framework for Building Distributed Software Systems [J]. Applied Artificial Intelligence: An International Journal, 1999, 13 (1-3): 91~128
- de Meo P, Rosaci D, Sarne G M L, et al. An XML-based Adaptive Multi-agent System for Handling E-commerce Activities. ICWS-Europe 2003, LNCS 2853, Berlin. Springer-Verlag, Heidelberg, 2003. 152~166
- Huang Z, Eliens A, Visser C. 3D Agent-based Virtual Communities. In: Wagner W, Beitler M, eds. Proc. Int Web3D Symposium, ACM Press, 2002. 137~144
- CHEN Gang. A distributed data mining system based on agent. Computer Engineering, 2001, 27(9): 65~68(in Chinese)
- Klusck M, Lodi S, Moro G. Agent-based Distributed Data Mining: The KDEC Scheme. In: AgentLink, volume 2586 of LNCS. Springer, 2003
- Zhang Z, Zhang C, Zhang S. An Agent-Based Hybrid Framework for Database Mining. Applied Artificial Intelligence, 2003, 17(5-6): 383~398
- Davies W H E, Edwards P. Agent-based Knowledge Discovery [C]. In: AAAI Spring Symposium on Information Gathering, 1995
- Davies W, Edwards P. Distributed learning: an agent-based approach to data-mining. R. Scotland, AB9 2UE: University of Aberdeen, 1997
- Bailey S M, Grossman R L, Sivakumar H, et al. Papyrus: A System for Data Mining over Local and Wide Area Clusters and Super-Clusters. [Technical report]. University of Illinois at Chicago
- Kargupta H, Park B, Hershberger D, et al. Collective Data Mining: A New Perspective Toward Distributed Data Mining. In: Advances in Distributed and Parallel Knowledge Discovery, AAAI/MIT Press, 2000. 131~178
- Kargupta H, Hamzaoglu I, Stafford B. Scalable distributed data mining using an agent based architecture [A]. In: Proc. of KDD97 [C]. Menlo Park, CA: AAAI Press, 1997. 211~214
- Stolfo S J, Prodromidis A L, Tselepis S, et al. JAM: Java agents for meta-learning over distributed databases. In: Hecherman D, Mannila H, eds. Proc. Third International Conference on Knowledge Discovery and Data Mining (KDD-97), Newport Beach, California, USA, AAAI Press, 1997. 74~81
- Prodromidis A L, Chan P K, Stolfo S J. Meta-learning in Distributed Data Mining Systems: Issues and Approaches. In: Advances in Distributed and Parallel Knowledge Discovery. Chapter 3. 2000
- Rana O F, Walker D W, Li Mao-zhen, et al. PaDDMAS: Parallel and distributed data mining application suite [A]. In: Proc. 14th Int Conf. on Parallel and Distributed Processing Symposium [C]. Cancun Mexico: IEEE, 2000. 387~392
- Wagner G. Toward Agent-Oriented Information Systems. [Technical report]. Institute for Information, University of Leipzig, March 1999
- McDonald J T, Talbert M L, Deloach S A. Heterogeneous Database Integration Using Agent-oriented Information Systems
- 马振宇, 黄启春. 支持分布应用的可组态管理信息系统体系结构. 大连理工大学学报, 2003, 43(S1): 118~122
- 石双元, 陈琼, 吴新明. 基于构件的信息系统开发框架. 计算机工程与科学, 2004, 26(10): 106~109
- 张明宝, 夏安邦. 基于面向服务体系架构的敏捷虚拟企业信息系统框架. 计算机集成制造系统, 2004, 10(8): 985~990
- 王颖, 吴荣泉, 黄美锋, 等. 一个面向服务的 EAI 框架. 计算机工程, 2006, 32(1): 279~281
- Yu Ming-Hui, Fei Qi, Chen Xue-Guang. A Role Interactive Model Based on Method of information System Function Partition. Science & Technology Progress and Policy, 2004, 9: 105~107
- Ren Ming-Lun, Zhu Wei-Dong, Yang Shan-Lin. Component Based Architecture of Information Systems. Mini-Micro Systems, 2004, 25(7): 1159~1163
- Feng C, Qianxiang W, et al. An architecture-based approach for component-oriented development. In: Computer Software and Applications Conference, 2002. COMPSAC 2002. Proceedings. 26th Annual International, Aug. 2002. 450~455
- Succi G, Pedrycz W, etc. Package-oriented software engineering: a generic architecture. IT Professional, 2001, 3(2): 29~36
- Kirk D, Roper M, Wood M. Defining the problems of framework reuse. In: Computer Software and Applications Conference, 2002. COMPSAC 2002. Proceedings. 26th Annual International, Aug. 2002. 623~626
- Chaudet C, Greenwood R M, et al. Architecture-driven software engineering: specifying, generating, and evolving component-based software systems. Software (IEEE), 2000, 147(6): 203~214
- Fontaine A, Truong Anh-Thu, Manley T. A Survey of Strategies for Object Persistence. CSCI 5802-Spring, 2006
- Chai Yue-ting, Zhang Xiao-dong, Dong Jin. Research on the reconstructivity of agile supply chain management system. Journal of Tsinghua University, 2000, 40(3): 68~71
- Nikolaïdou M, Alexopoulou N, Tsadimas A, et al. A Consistent Framework for Enterprise Information System Engineering, 2006. <http://www.hua.gr/Pythagoras/Website/Docs/PDF/A Consistent Framework for Enterprise Information System Engineering.pdf>
- Zachman A J. A Framework for Information Systems Architecture. IBM Systems Journal, 1999, 31(3): 445~470
- Wang Zhongjie, Xu Xiaofei, Zhan Dechen. A Survey of Business Component Identification Methods and Related Techniques. International Journal of Information Technology, 2005, 2(4): 229~238
- PENG Yan-hong. Study on Enterprise Information Systems Architecture. Journal of System s Science and System s Engineering, 2000, 19(1): 77~86
- 张彬, 李国辉, 郭军. 一种基于分层平面的工作流组织模型的研究与实现. 小型微型计算机系统, 2004, 25(7): 1253~1255

(上接第 135 页)