基于云计算的新型 B2B 平台研究与构建

熊刚[1]，范东[1]，姚彦洁[1]，侯家琛[1]，刘胜[1]，章华明[2]

[1]中国科学院自动化研究所，复杂系统智能控制与管理国家重点实验室，北京市智能化技术与系统工程技术研究中心，北京，100190

[2]江苏中科智能工程有限公司，苏州工业园区唯亭镇，215000，苏州

摘要：随着互联网规模和互联网数据规模的不断增加，基于互联网的 B2B 商业份额渐渐以其快捷性慢慢占领传统商业行为份额的市场。同时随着近年来云计算(Cloud Computing)技术的兴起，使得高效处理和分析海量商业客户数据成为可能，包括海量客户资料、海量商品信息等，因此构建基于云计算的新型 B2B 平台逐渐开始吸引研究者的研究兴趣。本文提出了基于云计算的新颖 B2B 平台框架设计和实现方案，并通过在真实业务数据集上进行验证，显示了本文所提框架能有效提高传统 B2B 平台的业务响应效率，能有效降低业务响应时间，表明了在设计中采纳复杂网络技术和云计算技术后，能较好提高传统 B2B 平台的业务逻辑维护效率。

关　键　词：云计算；B2B；效率提升；海量数据；复杂网络

中图法分类号：TP301.6

A New B2B Platform based on Cloud Computing

Gang Xiong[1], Dong Fan[1], YanJie Yao[1], JiaCheng Hou[1], Sheng Liu,

Huaming Zhang[2]

[1] State Key Laboratory of Intelligent Control and Management of Complex Systems, Beijing Engineering Research Center for Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190

[2] Jiangsu China Sciences Intelligent Engineering Co. Ltd. Suzhou Industry Park, WeiTing Town, 215000, Suzhou

ABSTRACT: With the fast development of Internet and its data scale, B2B (Business to Business), whose speed and high availability advantage is based on Internet, is eroding more and more market share of traditional business. In recent years, the new data processing technologies, such as cloud computing, assure the enhancement of computer's computing capacity; it has become possible for researchers to process and analyze the massive business data, including large scale customer data, and large scale commodity information. A new B2B platform frame, based on cloud computing with less running time and better response efficiency, is proposed here to promote transaction handle efficiency. By adoption of complex network theory and cloud computing technology, the new platform has been evaluated to decrease transaction handle time, and make more benefits.

Key words: cloud computing; B2B; efficient promotion; mass data; complex network

1 引 言

电子商务(Electronic Commerce)是指借用通过电子手段来实现商务目的。而根据百度百科[1]中对电子商务的定义："在全球各地广泛的商业贸易活动中，在因特网开放的网络环境下，基于浏览器/服务器应用方式，买卖双方不谋面地进行各种商贸活动，实现消费者的网上购物、商户之间的网上交易和在线电子支付以及各种商务活动、交易活动、金融活动和相关的综合服务活动的一种新型的商业运营模式。"此定义表明电子商务是随着计算机技术和互联网技术的广泛应用而发展起来的一种崭新的商业运营模式。因此，从某种意义上说，计算机技术推动了电子商务的发展，从技术上决定了电子商务发展的深度和广度。

一方面，电子商务的发展在很大的程度上依赖于网络计算平台的开发利用，特别是网络化软件的发展；另一方面，电子商务的发展又为网络化软件的发展提供了一个很好的实际应用平台，为网络化软件的进一步发展指明方向，提出很多迫切需要解决的关键科学问题。也就是说，网络化软件和电子商务相互促进，共同协调发展。

本文在研究已有网络化软件包之间的关系上，用新兴的复杂网络分析技术统计网络化 B2B 电子商务软件的特征，分析及演化和耦合特征，针对已有 B2B 电子商务软件维护工作量大、软件组件间紧耦合等缺点，采用复杂网络技术[2]和云计算技术[3]对其改进，并通过在真实数据上进行实验，发现其有效提高了软件维护效率和缩短了业务响应时间。

本文结构如下：第 2 节介绍了电子商务发展的历史的最新研究进展，结合复杂网络技术对其面临的挑战和新的研究方向进行了分析；第 3 节针对已有电子商务软件平台的缺点，用复杂网络技术和云计算技术对电子商务软件平台进行了改进，提出了基于云计算的新型 B2B 电子商务平台；第 4 节通过统计已有电子商务软件平台中软件包的固有统计特征，给出了新型平台在真实数据上的软件耦合改进实验结果和软件效率提升实验结果；第 5 节总结全文并给出了展望。

2 相关研究

2.1 电子商务发展历史和面临的挑战

电子商务始于上个世纪 60 年代，而自上个世纪 90 年代以来随着互技术的发展而得到迅猛发展。一般将电子商务分为以下七种基本模式：

1） 企业对企业（Business-to-Business，简称 B2B)
2） 企业对消费者(Business-to-Customer，简称 B2C)
3） 消费者对消费者(Customer-to-Customer，简称 C2C)
4） 企业对职业经理人或市场 (Business-to-Manager or Marketing，简称 B2M)
5） 职业经理人对消费者(Manager-to- Customer，简称 M2C)
6） 商业机构对行政机构 (Business-to-Administration，简称 B2A)或 Business-to-Government，简称 B2G)
7） 消费者对行政机构(Customer－to-Administration,简称 C2A)或(Customer－to- Government，简称 C2G)

从过去电子商务软件的发展历史而言，传统电子商务软件的设计基本上都是依赖顶层设计的思想，但是由于当前互联网时代信息资源的快速变化以及全球经济一体化所带来的前所未有挑战，使得过去按照顶层

设计思路构建的电子商务软件在维护和软件效率上很难适应目前网络化软件的变化需求。

2.2 云计算技术对电子商务软件的影响

依据美国国家标准与技术学会信息技术实验室所给的云计算的定义，云计算是一种能够通过网络以便捷的、按需的方式获取计算资源（网络、服务器、存储、应用和服务）的模式，这些资源来自一个共享的、可配置的资源池，能够快速提供和释放，并只需要很少的管理工作或很少的与服务供应商的交互。

依据国内相关学者[4]对云计算给出的相关定义：云计算是并行计算(Parallel Computing)、分布式计算(Distributed Computing)和网格计算(Grid Computing)的发展，或者说是这些计算机科学概念的商业实现。云计算基础架构本质：资源虚拟化的并行计算。其基本特征是：

1）系统建立在大规模的廉价服务器集群之上；
2）通过基础设施与上层应用程序的协同构建以达到最大效率利用硬件资源的目的；
3）通过软件的方法容忍多个节点的错误。

通过云计算对这三个方面的基本特征，达到了分布式系统两个方面的目标，即系统的可扩展性和可靠性。

从这个角度而言，云计算不是新技术，而是新的资源利用模式。我们可以把云计算看作是前代计算模式(比如网格计算、透明计算、效用计算)的阶段升级。云计算的实质是一种基础架构管理的方法论，是把大量的计算资源组成 IT 资源池，用于动态创建高度虚拟化的资源提供用户使用。在云计算环境下，所有的计算资源，比如说服务器能够动态地从硬件基础架构上增减，以适应工作任务的需求。所以，云计算最真实完整的基础架构是一个由多种产品和服务集成起来的端到端的解决方案。云计算基础架构的本质是通过整合、共享和动态的硬件设备供应来实现 IT 投资的利用率最大化。因此，计算云的基础架构的本质可以看成是：并行计算+资源虚拟化。

2.3 网络化软件及其统计特征

网络化软件[4]（Networked Software）是指在复杂网络环境中运行的结构和行为可以动态演化的软件。网络化软件的典型特征主要包括：

1）用户和开发者的数量都很大：以互联网为基础的应用没有规模和边界限制，软件的开发大众化；

2）软件模块数量庞大，爆炸式增长，需求变化大：开放的开发平台、旺盛的用户需求、盈利的商业模式造就了庞大的数量并呈爆炸式增长，使其开发者需要能开发细粒度、可配置、可组合的按需服务满足用户个性化和多样化的需求。

3）各模块间耦合松散，对互操作要求高：网络化软件各模块间要求粒度小，并且可以通过耦合和解耦合快速满足用户个性化的互操作需要。

目前常见的网络化软件的典型的应用形态主要包括 SAAS（Software as a Service）和在线查询获取、离线使用的网络化软件如 IPhone App Store。

目前国内外学者通过对网络化软件的研究，如 Linux 内核和 Tomcat 的研究，发现网络化软件明显存在无标度特性，属于复杂网络中的小世界网络[5][6][7]。Linux 内核中是函数抽象为节点，函数之间的调用关系抽象为边。Tomcat 软件包中是类抽象为点，类与类之间的继承或聚合关系抽象为边。

因此如何应用基于复杂系统的网络化软件的建模与分析方法来对电子商务建模，开发新的高效、迅捷的网络化电子商务软件是目前电子商务领域的研究热点问题。

3 基于云计算的新型 B2B 平台构建

3.1 基于 PASS 的电子商务 B2B 新型平台

云计算是一种计算模式，代表了在某种程度上共享资源进行设计、开发、部署、运行应用，以及资源的可扩展收缩和对应用连续性的支持。目前大部分的云计算基础构架是由用户终端向数据中心传送可申请服务的网络，以及建立在服务器上的不同层次的虚拟化技术组成的，使人们可以在任何有提供网络基础设施的地方使用这些服务。因此，云的基本概念是通过网络将庞大的计算处理程序自动分拆成无数个较小的子程序，再由多部服务器所组成的庞大系统搜索、计算分析之后将处理结果回传给用户终端。在具体的电子商务应用中，可以利用云计算技术将电子商务软件的相关业务逻辑组件抽象为细粒度的原子组件，放置在 PASS 平台即服务层。通过这项技术，远程的电子商务软件服务供应商可以在数秒之内，达成处理数以千万计甚至亿计的信息，达到和“超级电脑”同样强大效能的网络服务。

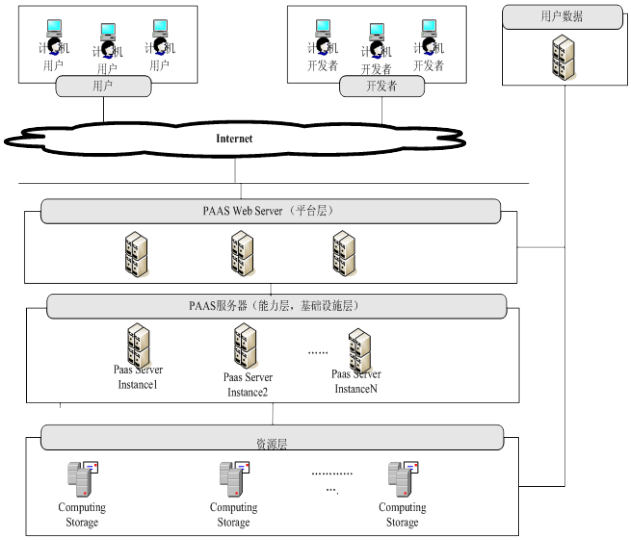

图 1 目标系统逻辑框架

在具体的项目研究中，为了克服过去网络化的电子商务软件维护工作量大，软件组件间紧耦合不利于升级和修改的缺点，我们引入了云计算技术对其进行改进，将电子商务软件的各个业务逻辑功能组件放置在云计算的 PASS（Platform as a Service）层，由开发者开发各组件间耦合度低的电子商务软件组件，再由使用者在 SAAS 层进行调用，并借助工作流引擎等先进技术，使得使用者可以依据自身的具体电子商务逻辑构建相应的电子商务软件，方便了用户的使用，其目标系统的具体框架如图 1 所示.

将图 1 中云计算的三层框架进行细化，得到图 2 所示的基于云计算的新型 B2B 平台框架：

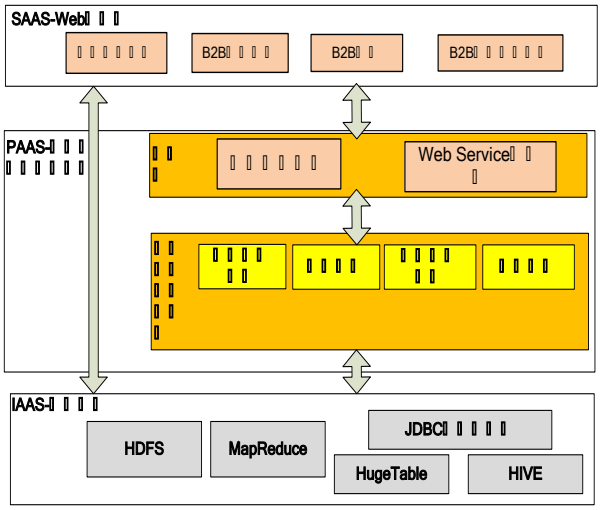

图 2 基于云计算的新型 B2B 平台框架

图 2 所示系统的总体架构包含三层，与用户交互的 web 服务层（SAAS 层），接收用户功能并通过工作流引擎对各个原子电子商务组件进行组合的电子商务逻辑功能层（PAAS 层）和提供基础云计算设施的 IAAS 层。其中 web 服务层通过 SOAP 消息和 PASS 层电子商务原子业务逻辑组件进行通信，IAAS 层支持对海量电子商务数据的并行化处理及海量客户信息搜索操作，下面将分层分别进行详细介绍。

Web 服务层：主要通过工作流引擎对用户提供已经封装好的电子商务交易操作，如电子商鉴权操作、B2B 海量企业客户数据搜索、B2B 交易处理和 B2B 账户资金管理等。通过 SAAS 的方式进行构建，并提供租户和用户的两种方式供使用者选择，建设灵活的企业电子商务交易平台。

电子商务逻辑功能层：主要完成具体原子电子商务逻辑的构建，如数据搜索引擎的相关原子组件、加密算法的相关原子组件、用户权限管理的相关原子组件和交易处理的相关原子组件等。并对 Web 服务层提供通过命令行接口方式和 web Service 接口方式，可以灵活的方便普通用户和管理员级别的超级用户对该层的电子商务原子业务逻辑组件进行调用。

云平台层：通过 HDFS 等开源云计算模型组件提供海量数据的存储和海量数据的分布式计算，数据通过文件的形式进行存储，便于进行海量电子商务客户数据的快速搜索和及时处理。云计算技术提高了后台海量数据的处理速度和规模，提高了后台算法的运行效率。

3.2 软件耦合 Betweenness 改进

复杂网络可抽象为一个由结点以及结点之间的边组成的图（Graph）结构：$G=(V,e)$，其中 V 是网络中的节点集合，e 是边的集合。节点的个数 $N=|V|$，边的条数 $L=|e|$。在研究复杂网络时，一般区分为有向图、无向图、无权图、有权图分别进行对待。在复杂网络中，如果不考虑结点之间边的方向性，则图 G 被称为无向图；反之，如果考虑结点间边的方向性，图 G 则被称为有向图。

最常见的对复杂网络进行分类的方式是依据节点的度分布特性进行分类，按照这种方式，将复杂网络分为随机图模型[8]、小世界模型[9]、无标度模型[10]和大统一混合模型[11]。Erdos 与 Rényi 于 1960 年提出的 ER 随机图模型理论（ER 模型），指出随机图属于指数网络，并且该指数网络中的节点是同质的，它们的度大致相同，绝大部节点的度都位于网络节点平均度附近，网络节点度分布随度数的增加呈指数衰减，网络中不存在度数特别大的节点。Watt 与 Strogatz 在 1998 年提出 Watt Strogatz 小世界网络模型(WS 模型)。随机图模型与小世界网络模型的主要区别是：前者的群聚系数小，而后者的群聚系数大。目前国内外研究界把具有较小平均路径长度和较大群聚系数的网络统称为小世界网络。Barabási 和 Albert 在

1999 年首先提出无标度网络模型(称为 BA 模型或 BA 网络)。无标度网络中节点是异质的，其节点度服从幂律分布。大部分节点只与少数几个其它节点连接，但网络中存在为数不多的度数特别大的节点，称为中心节点或者 hub 节点，这些中心节点对无标度网络的特性起着主导和支配作用。小世界网络和无标度网络是 1998 年以来的新发现，具有普遍性。和谐统一的混合择优模型（HUHPM）和大统一混合模型（LUHNM）由方锦清等人在 2007 年提出，基本抓住了真实世界中两个基本连接方式的混合特点，即随机性和确定性两大类连接方式，通过引进一个总混和比，把一些现有的模型的特性包括在内，具有一定的普适性。大统一混合网络模型是随机图模型、WS 模型、BA 模型的进一步拓广。

在网络化电子商务软件中，类或函数一般被抽象为复杂网络的点，类之间的关系或者函数之间的调用关系被抽象为边，依据已有的研究表明，网络化电子商务软件属于无标度网络，同时还有研究表明，网络化电子商务软件中存在 5%左右的中心类或者函数，它们控制着整个软件逻辑的 95%以上[12]。

在复杂网络中标定某个重要节点时，常常使用相关的中心性维度来定义和衡量该节点的重要性，一般使用最常见的三种节点中心性维度：节点中介中心性、节点度中心性以及节点接近中心性，我们在实验中采用了节点中介中心性。

如果在复杂网络组成的图 G 中，对于 $\forall v_i, v_j \in V$，都存在至少一条路径使之相连接，那么称 G 为连通图。由于结点 v_i 与 v_j 之间可能存在多条路径，其中长度最短的一条称之为连接 v_i 和 v_j 的最短路径。假设 $\sigma(v_i, v_j)$ 是节点 v_i 和 v_j 之间最短路径的数目，而 $\sigma(v_i, v_j | v_k)$ 是节点 v_i 和 v_j 之间经过顶点 v_k 的最短路径条数，当 $v_k \in \{v_i, v_j\}$ 时，则 $\sigma(v_i, v_j | v_k)$ =0；当 $v_i = v_j$ 时，$\sigma(v_i, v_j)$ =1。那么图 G 中节点 v_k 的中介中心性 $B(v_k)$ 定义如下[13]：

$$B(v_k) = \sum_{v_i, v_j \in V} \frac{\sigma(v_i, v_j | v_k)}{\sigma(v_i, v_j)} \quad (1)$$

从上面的定义可以看出，节点-中介中心性是图 G 中通过节点 v_k 的最短路径条数与整个图中最短路径条数的比例。通过节点 v_k 的最短路径条数越多，则它的节点-中介中心性越大，而图中的最短路径又常常是信息快速传导的途径，因此节点-中介中心性反映了节点对网络中信息传导的承载程度。在网络化电子商务软件网络中，节点-中介中心性越大的节点往往是不同类或者函数之间信息沟通的“桥梁”，地位非常重要。

我们在对实际的电子商务系统进行改进时，由于该系统基于 Java 进行创建，因此我们首先统计了该软件 5 个历史版本中类之间的继承或聚合关系，将关系抽象为边，统计结果如下：

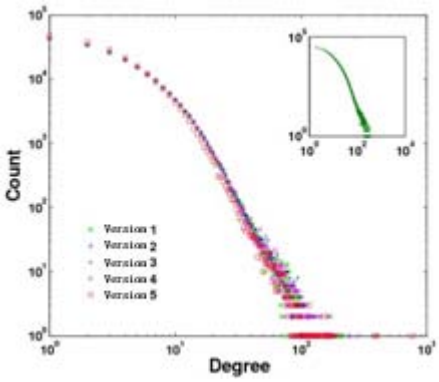

图 3 原有电子商务软件度分布统计

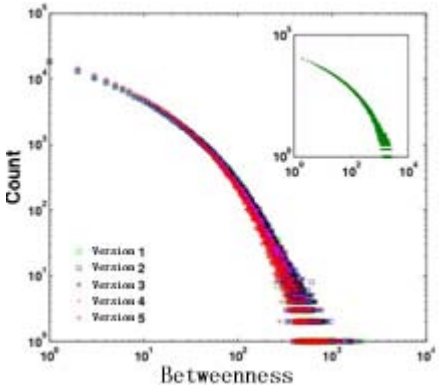

图 4 原有电子商务软件中介性分布统计

经过统计和分析图 3、4 可知[14]：该电子商务软件原有软件包中的类的度分布和中介性分布都符合 Power Law 分布，并且经过统计可知度和中介性数值最大的 5%节点所占的数值占了整个软件包的度数值和中介性数值的 80%以上，它们在整个软件包中占据着重要的地位[15]。因为在软件耦合的改进中，我们将这些前 5%的类节点尽量拆分为业务逻辑单一或者类方法依赖性低的类，将软件包松耦合化，提高了软件包的可复用度，降低了业务逻辑变动时代码的修改维护时间。

4 基于云计算的新型 B2B 平台验证

4.1 软件系统业务响应时间实验

为了对基于云计算的新型 B2B 平台的性能进行验证，在图 3 和图 4 所示的真实电子商务软件系统上测试了业务响应时间。真实系统前端展现层采用 Flex 结合 Java 技术的 Web 页面进行实现，电子商务逻辑功能层采用 Tomcat6.0 作为承载 Web 应用和 WebService 服务的容器。云平台层采用 Hadoop-0.19.0.core.jar 中的 API 实现，后台数据存储搭建在 20 个节点组成的集群上，每个节点内存 16GB，硬盘容量 12TB，CPU 主频是 4*2GHz ，部署 RedHat 6 操作系统。

然后我们将图 3 和图 4 中的目标系统划分为核心子系统和外围辅助子系统，其中核心子系统由 5 个子系统构成，外围辅助子系统由 4 个子系统构成，为了评估目标系统通过引入复杂网络和云计算技术前后系统性能的变化，引入了“业务平均响应时间”（AVG Transaction Response Time）来衡量系统构建前后性能的提升，实验结果请参见图 5 和 6。为了避免随机误差，每个实验数据是该业务操作在 10 次执行下的平均值。每个子系统的柱状图中左边的是改造前的性能数据（Previous），右边的是改造后的性能数据（Current）[16]。

$$\text{Promotion Rate} = \left(\frac{\text{AVG Pre Response Time-AVG Cur Response Time}}{\text{AVG Pre Response Time}} \right) \tag{2}$$

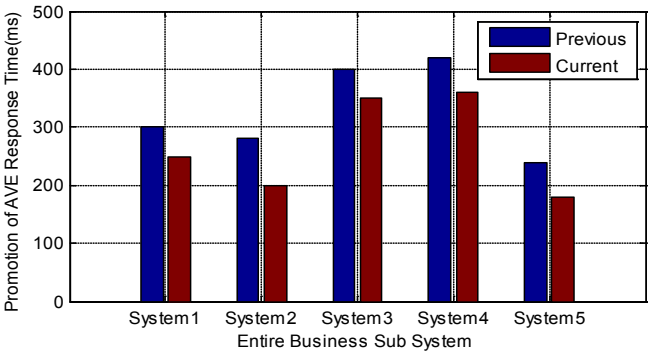

图 5 五个核心子系统的性能提升

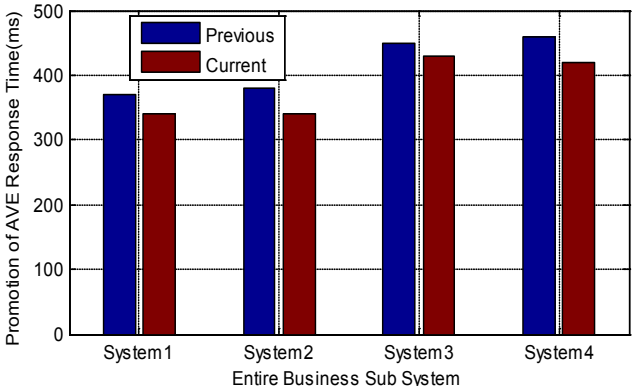

图 6 四个外围辅助子系统的性能提升

经分析图 5 和 6 可知：由于复杂网络技术和云计算技术的引入，五个核心子系统“业务平均响应时间”的提升在 13%和 30%

之间。四个外围辅助子系统“业务平均处理时间”的提升在 5%和 11%之间，经比较可知，外围辅助子系统的性能提升幅度低于核心子系统的性能提升幅度，新型平台对核心子系统的影响更大[17]。

4.2 软件系统代码维护工作量实验

通过统计原有电子商务软件原有软件包中的类的度分布和中介性分布的 Power Law 分布特性，将度和中介性数值最大的 5%类进行重构，将其业务逻辑进行拆分和分解为耦合度低的子类，测试了改造前后当某个业务发生改变，核心代码改变时对两者的逻辑代码维护量的影响，如图 7 所示：

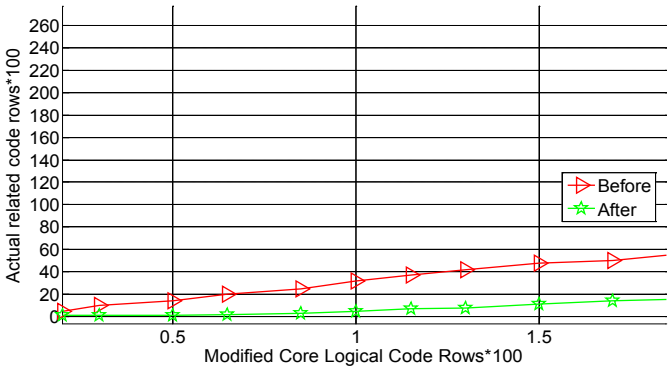

图 7　目标系统逻辑代码维护量对比

通过分析图 7 可知，经过软件耦合度 Betweenness 的改进，目标电子商务软件系统的核心逻辑代码发生改变时（即图 7 中的 X 轴），其引起的外围相关逻辑代码该变量（即图 7 中的 Y 轴）大幅度减小，减小幅度至原来的 1/4 至 1//5，证实了经过软件类节点的 Betweenness 改进后，使得软件整体业务逻辑代码耦合度降低，从软件工程角度而言，软件代码之间的耦合度降低，软件代码的复用度增加，对软件的改造趋于良性[18]。

5　结束语

本文已有网络化软件包之间的关系的基础上，用新兴的复杂网络分析技术和云计算技术构建了基于云计算平台的新型 B2B 平台，相对于已有电子商务软件的贡献如下：通过使用复杂网络技术统计和分析网络化 B2B 电子商务软件的特征，分析及演化和耦合特征，针对已有 B2B 电子商务软件维护工作量大、软件组件间紧耦合等缺点，采用复杂网络技术和云计算技术对其改进，并通过在真实业务数据上进行实验，发现其有效提高了软件维护效率和缩短了业务响应时间。

后续研究将围绕以下工作展开：（1）进一步研究电子商务软件的复杂动力系统特性，并将其用于改进软件效率和优化软件体系结构；（2）将基于云计算的新型 B2B 平台在更大规模的互联网业务数据上进行测试，以期改进软件架构并提高软件效率。

致 谢

本文得到国家自然基金项目 70890084, 60921061, 90920305; 中国科学院项目 2F09N05, 2F09N06, 2F10E08, 2F10E10; 江苏中科智能工程有限公司企业院士工作站项目 2F11D01 等的资助。留学回国人员科研启动基金项目 4T05Y03 等的资助。

参 考 文 献

1. http://en.wikipedia.org/wiki/Electronic_Commerce
2. Jure Leskovec, Kevin J. Lang, Michael Mahoney. Empirical comparison of algorithms for network community detection[C]. In WWW ’10: Proceedings of the 19th International Conference on World Wide Web (2010), pp. 631-640.
3. J. Dean, S. Ghemawat. Mapreduce: Simplified Data Processing on Large Clusters [C]. OSDI’04, 2004, pp. 137-150.

4. 周进.复杂网络的建模及其控制.博士学位论文. 武汉大学，武汉，20009.
5. 何克清，马于涛，刘靖，李兵，彭蓉. 软件网络.北京:科学出版社，2008.
6. 刘靖. 基于复杂网络的软件结构分析及优化研究. 博士学位论文，武汉大学，武汉，2007.
7. 李德毅，张海粟. 超出图灵机的互联网计算. 中国计算机学会通讯，2009，12(5):8-16.
8. Erdös, Paul, Rényi, Alfréd.On the Evolution of Random Graphs [J], Publ. Math. Inst. Hung. Acad. Sci., Ser. A 5, 17-61 (1960), 05C80.
9. D.J.Watts and S.H.Strogatz. Collective Dynamics of “Small World” Networks [J]. Nature, 1998, 393 (4): 440-442.
10. A.L.Bababasi and Albertr. Emergence of Scaling in Random Networks [J]. Science, 1999, 286: 5092512.
11. 方锦清，汪小帆，郑志刚. 非线性网络的动力学复杂性研究.物理学进展. 2009,29(1):1-74.
12. 王红春. 网络化软件多粒度动态特性分析. 博士学位论文，武汉大学，武汉，2010.
13. C. Kiss, A. Scholz, M. Bichler. Evaluating Centrality Measures in Large Call Graphs[C].Presented at IEEE Joint Conference on E-Commerce Technologyand Enterprise Computing, E-Commerce and E-Services, 2006 in San Francisco, USA.
14. 陈康，郑纬民.云计算:系统实例与研究现状. 软件学报. 2009 年 05 期.pp:1337-1348
15. 徐志伟，廖华明，余海燕，查礼.网络计算系统的分类研究.计算机学报. 2008 年 09 期. pp:1509-1515
16. 冯登国，张敏，张妍，徐震. 云计算安全研究. 软件学报.Journal of Software.2011 年 01 期.pp:1-13
17. J.W. Han, M. Kamber. Data Mining: Concepts and Technique [M]. The Morgan Kaufmann Series in Data Management Systems 2nd Edition, Morgan Kaufmann Publishers ,2006
18. B.Bollobas. Modern Graph Theory [M]. New York: Springer, 1998.