

基于本体的物流行业知识库的研究*

饶国政¹, 冯志勇¹, 张 丽², 聂旭飞¹

(1. 天津大学 计算机科学与技术学院, 天津 300072; 2. 天津科技大学 经济与管理学院, 天津 300222)

摘 要: 构建了基于本体的物流行业信息化知识库。应用语义 Web 和 Web 服务技术研究物流知识资源, 设计了物流领域本体, 实现了文本摘要、规则抽取、本体整合等关键功能, 解决物流系统中语义和重用问题, 提高了物流企业的知识共享和协作能力。这不仅为物流行业信息化奠定了坚实基础, 而且为知识库的复用创造了条件。

关键词: 本体; 物流; 知识库; 知识复用; 面向服务的架构

中图分类号: TP316 文献标志码: A 文章编号: 1001-3695(2008)09-2734-04

Study on ontology-based logistics industry repository

RAO Guo-zheng¹, FENG Zhi-yong¹, ZHANG Li², NIE Xu-fei¹

(1. School of Computer Science & Technology, Tianjin University, Tianjin 300072, China; 2. School of Economy & Management, Tianjin University of Science & Technology, Tianjin 300222, China)

Abstract: This paper presented ontology-based logistics industry repository. Logistics knowledge resources were discussed with semantic Web and Web services technology. The problem of semantic and reuse in logistics information system was resolved by implementation of designing logistics domain ontology, text summary, rules extraction and ontology merge. The knowledge sharing ability of logistics enterprises is improved, which is the foundation of logistics industry informationization and the conditions of reusing repository.

Key words: ontology; logistics; repository; knowledge reuse; service oriented architecture

0 引言

随着经济全球化和信息技术的迅猛发展,全球产业结构的调整不断深入,社会生产、物资流通、商品交易及其管理方式正发生深刻的变革。继物质资源市场、人力资源之后的“第三利润源”——现代物流正在世界范围内迅速发展,成为现代经济中最活跃的增长因素。

在我国,随着经济的快速发展和现代物流理念的逐渐普及,物流产业发展迅速,已经作为国民经济中的一个新兴的产业部门,成为 21 世纪我国的主要支柱产业和国民经济新的增长点。据文献[1]显示,社会物流总额(指一定时期内实现社会物流的全部物品的总价值,在一定意义上也就是采购货物的总价值)从 2000 年开始进入快速增长期,2000—2006 年,全国社会物流总额由 17.1 万亿元人民币上升至 59.7 万亿元人民币,7 年间增长 2.5 倍,年均增幅 19.6%,大大高于同期 10.7% 的 GDP 年均增长率。随着物流产业的发展,对物流信息化提出了更高、更新的要求。由于物流行业的复杂性、动态性和分散性,以及管理模式的多变性,当前物流企业的信息化水平还难以达到企业的需求,这直接影响了物流中心与用户的沟通和协作,阻碍了物流服务质量的提高。就整体来看,国内的物流信息化状况依然存在很多问题,还不能达到国际化的要求,从而制约了我国物流产业的发展。面对这种现状,迫切需要针对物流信息化存在的问题,结合现代物流的发展趋势和发展要

求,分析物流产业对信息系统的具体要求,研究适合中国国情的物流信息化关键性技术。

行业信息化知识库,是对行业核心应用软件起支撑作用的行业信息化经验积累的集合,也可以说是行业经验的沉淀;是行业信息化建设的关键知识、规律及诀窍等知识的汇聚;是知识复用和软件开发的基础。马希文教授曾指出“软件是人类知识的浓缩和结晶”,知识开发将从软件开发中分离出来,知识模块将从软件平台中分离出来。陆汝钤院士也曾在 2004 年智能科学技术基础理论重大问题研讨会上提出:“知识产业要从软件产业中分离出来,我们将从软件复用走向知识复用……”因此,行业应用知识资源的开发利用水平很大程度上代表了行业信息化建设的成功与否,以及在行业信息化建设中所占有的举足轻重的地位。在行业信息化中,无论是技术、管理,还是数据,沉淀下来都是知识资源,尤其是以行业应用知识体现的行业标准、行业模型、行业构件,更是宝贵的知识资源。充分开发利用行业应用知识资源,构建先进的行业信息化知识库,对分布在物流行业中的零散知识进行有效管理,挖掘蕴藏在行业中的宝贵的知识财富,并在物流行业中进行充分的复用,是大势所趋。

目前很多学者建立了各种不同的知识库系统^[2~5],大多作为专家系统的一部分。但是物流行业的知识库比较少,且只有少量的国外研究者将知识库应用于物流中的某些环节,如文献[6]给出了一个利用知识库制定物流战略规划框架;文献

[7] 基于知识库实现了一个物料自动编号的物流信息系统; 文献[8]通过 EDI 和工作流为美国国防部提供了一个军事物流(后勤)解决方案。但是这些系统难以实现知识的共享和交换。这种情况造成了用大量精力建立起来的基于领域知识的系统应用面狭窄, 不能充分发挥效率。因此设计良好的共享重用体系不仅可以减少知识库的冗余, 保证知识库系统的简洁性, 而且还可以提高知识系统的性能。

1 本体与知识库

知识表示是研究用机器表示知识的可行性、有效性的一般方法; 是一种数据结构与控制结构的统一体, 既考虑知识的存储又考虑知识的使用。知识表示可看成是一组描述事物的约定, 即把人类知识表示成机器能处理的数据结构。知识表示方法通常可以分为基于符号的表示方法和基于连接机制的表示方法。前者主要是面向逻辑知识的表示, 如一阶谓词逻辑表示法、产生式表示法、框架表示法以及脚本表示法等; 后者则主要是面向形象知识的表达, 如语义网络表示法、过程表示法、Petri 网表示法以及面向对象的表示法等。在实际的知识表示过程中, 传统知识表示方法在知识表示、传递和共享过程中存在先天性的缺陷。其一方面, 所表达的知识无法保证在传递和共享过程中知识理解的惟一性与无二义性; 另一方面, 在大量的原子性知识的环境中对复杂知识的表达与推理可能会产生组合爆炸。为了解决上述知识表示方法中存在的问题, 在知识工程领域需要引入本体的概念^[9]。

本体(ontology)是一个哲学概念, 它是“共享概念化的形式的、明确的规范”^[10]和“某些方面概念化的明确解释或表示”^[11]。

从知识表示角度考虑, 本体和知识库有类似之处, 都是对一个具体或是抽象领域中包含的知识的定义、表示和组织。根据 Gloria 定义^[12]知识库是知识系统的知识模块, 它包含特定领域抽象或特定的知识, 这些知识以机器可读的格式表达。在本体概念被提出来后, 理想的领域知识库是应该建立在领域本体的基础之上, 并根据领域本体中的全部或部分概念生成系统所需的知识库。而且本体为人们描述目标世界提供了一组通用词汇, 而这种通用的词汇正是实现知识系统化的基础。通用词汇和知识的系统化有利于实现知识的标准化。

传统的知识管理方法往往只是从语法上对知识进行描述, 并不对其意义进行描述, 常采用基于关键字的检索方法(字符匹配), 查全率和查准率都较低。物流领域本体提供了领域相关的概念及其关系描述, 因此可以对具体知识进行标注, 形成含有丰富语义信息的知识库, 进而可以利用这些语义信息进行检索, 查全率和查准率都较高。此外由于基于本体的知识管理具有语义层、自动化、支持推理等优点, 不仅可以清晰地描述领域知识库中的概念及其关系, 还可以实现领域知识的共享和重用, 十分有利于领域知识库的管理和维护, 大大深化了知识管理的内涵, 促进知识的共享和复用。

国内外已有许多研究者将本体引入知识库中, 如文献[13]从半结构化的数据中采用语义模式创建疾病知识库; 文献[14]基于本体构建了医学知识库; 文献[15]基于 ontology 探讨了领域知识库构建的分层模型, 并将构建的知识库应用于采用 C++ 和 ASP 实现的基于 Web 的学习支持系统中; 文献[16]结合中国的法律法规给出了一个法律领域的本体模型

DOLegal, 介绍了法律知识库的设计思路及架构。

上述这些应用领域相对而言问题比较单一, 知识的变化较慢。但像物流这样包含采购、检测、包装、运输、仓储、配送等众多环节, 包含的知识跨越了多个领域, 而且错综复杂, 随时间变化速度较快的行业应用却非常少见。

2 SOA 与 Web 服务

伴随着企业需求的不断变化, 一种被誉为下一代 Web 服务的技术架构, 由于它是以实现服务资源的反复重用、服务资源之间的通信、协调, 从而消除服务资源的多样性、层次性、复杂性带来的困难为目标, 再一次引起业内关注, 这就是面向服务的架构(service oriented architecture, SOA)。它是一种完全低耦合的分布应用集成方法^[17], 将成为占有绝对优势的软件工程实践方法。SOA 中主要有三种角色(图 1): a) 服务使用者。它是一个应用程序、一个软件模块或需要一个服务的另一个服务。它发起对注册中心的服务的查询, 通过传输绑定服务, 并执行服务功能, 服务使用者根据接口契约来执行服务。b) 服务提供者。它是一个可通过网络寻址的实体, 接收和执行来自使用者的请求, 并将自己的服务和接口契约发布到服务注册中心, 以便服务使用者可以发现和访问该服务。c) 服务注册中心。它是服务发现的支持者, 包含一个可用服务的存储库, 并允许感兴趣的服务使用者查找服务提供者接口。

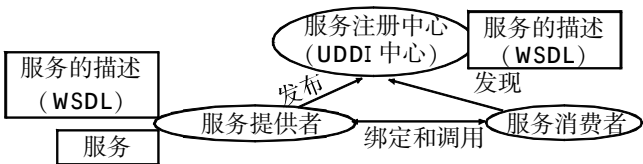


图 1 SOA 中的角色

SOA 是一种新的分布应用集成示范。Web 服务是 SOA 的一个实例, 基本的 Web 服务规范包括 Web 服务描述语言(WSDL)、简单对象访问协议(SOAP)、通用描述、发现与集成(UDDI)、Web 服务安全(WSSecurity)、Web 服务事务(WS Coordination、WS Atomic Transaction)等规范。Web 服务体系使用 WSDL 描述服务, 使用 UDDI 来发布和查找服务, 使用 SOAP 协议来执行服务调用^[18,19]。WSDL 就是描述 XML Web 服务的标准 XML 格式, 由 Ariba、Intel、IBM 和微软等开发商提出。它用一种与具体语言无关的以抽象方式定义了给定 Web 服务收发的有关操作和消息。

Web 服务的规范的标准化和松耦合的特性使得我们可以把物流行业知识库包装成一个用 Web 服务来实现的 SOA 构件, 这样知识库不仅重用性好而且可维护性大大提高。

3 基于本体的物流行业知识库

知识库的构建, 是一个耗时长、代价高的工作。首先需要建立一套科学的基于物流行业知识的可扩展资源分类方法, 对成千上万的异构知识源进行梳理; 然后就是研究这种大型知识库的知识表示和推理, 以及这些知识如何在物流行业的众多企业中得到广泛的复用。

如图 2 所示, 基于本体的物流行业信息化知识库体系结构采用分层模式, 自底向上分别为三层: a) 行业知识资源层, 在该层根据行业的实际情况, 将行业知识资源分为不同的类型。b) 知识库层(第二层)。在该层总体上采用 Web 服务标准, 将知识库中的经过语义标注的知识库包装成一个个 Web 服务,

每个 Web 服务可以作为一个独立的整体被复用,从而方便行业的各种不同层次的应用需求。c) 应用层(第三层)。该层主要是为不同的使用者提供 API、Web 服务和控制接口等提供各种各样的应用接口,方便不同的用户查询,从而更好地为不同的使用者服务。

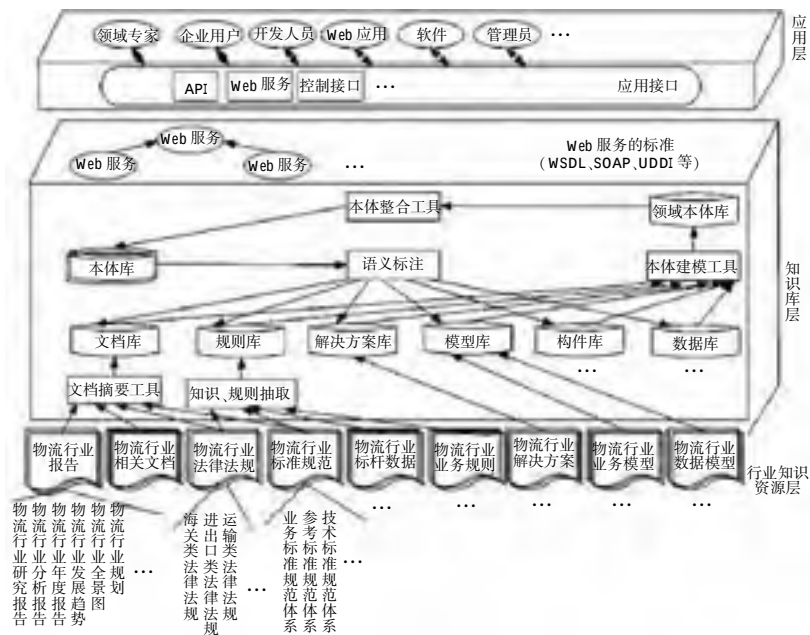


图 2 基于本体的物流行业知识库的体系结构

1) 行业知识资源梳理 物流行业隐藏着大量丰富的知识,不能得到很好的应用,行业知识资源可以分为物流行业报告、物流行业法律法规、物流行业标准规范、物流行业标杆数据、物流行业业务规则、物流行业解决方案、物流行业业务模型、物流行业数据模型、物流行业领域构件和物流行业数据库、物流行业其他相关文档等。其中物流行业报告又包含数以百计的像物流行业研究报告、物流行业分析报告、物流行业年度报告、物流行业发展趋势、物流行业全景图和物流行业规划等。物流行业又包括上百个像海关通关类、进出口类和运输类等类型的法律法规,更是有数以百计的行业标准规范、物流行业标杆数据,数以千计的物流行业业务模型和物流行业数据模型、物流行业领域构件等,以及数以万计的物流行业解决方案、物流行业其他相关文档。这些成千上万的知识源异构地分布在行业中,因此要建立完善的分类系统对行业知识资源进行有效的梳理,通过像文本摘要工具和知识、规则抽取方法等,将这些知识以结构化和半结构化的形式存放到相应的知识库中,如文档库、规则库、解决方案库、模型库和构件库等。

2) 文本摘要 该摘要工具能用来对海量的文档进行摘要处理,从而为用户产生物流相关文档源的精简表示,使用户可以方便省时地了解文本源的中心意思,从而快速地获得有价值的信息。如图 3 所示是网上的一篇与物流相关的文档,经过该工具的自动摘要生成如图 4 所示的能表达该文档的缩略版,并最终能将该文档的缩略版转换为 XML 文档,存储在相应的库中(如文档库)。

3) 知识、规则抽取 OWL(Web ontology language) 是一种本体语言,可以用于语义知识的表示。其主要作用是将领域内的概念分类,通过约束、布尔组合等来描述概念自身的语义含义,利用属性来表达概念与概念之间复杂的依赖。所以通过本体可以得到一个类的实例所具有的属性,以及类与类之间的关系。但 OWL 所拥有的描述能力主要来自于以类别为基础的关联性的推理,若知识不是用类别的方式表达时,OWL 便难以表示。例如: If ... Then... 的关系是很普通的逻辑判断,但是 OWL 描述以规则为基础的知识的的能力并不足。在物流领域

中,有很多物流规则需要用类似的 If... Then... 来表达,所以有必要在 OWL 概念知识库基础上制定一个规则库,从而更好地表达物流领域知识。

如在配货过程中,要遵循的配货规则为:易产生水分的货物不能和怕水、怕潮的货物一起运输;易燃烧的货物和易爆炸的货物不能一起运输;有毒、有放射性、有腐蚀性等的危险品要分别单独运输,不能和其他货物一起运输。当把货物都配好之后,便可以为每批货物找合适的运输工具进行运输。需要遵循的规则为:怕潮、怕水的货物要用篷车运输;液态货物要用油罐车或管道运输;危险品要用专用车辆运输等规则。



图 3 网上一篇与物流相关的文档



图 4 生成的摘要

以上规则语义表达明确,便于理解,并且支持将规则放在规则库中供程序调用,这就避免了在业务规则发生变化时更改源程序。以下是规则库中描述上述规则的片段:

```
# rule file
@ prefix LO: http://www.owl-ontologies.com/logistics.owl .
[ hygroscopicRule: (? x LO: isHygroscopic ? b) (? y LO: isWatery ? b) (? b rdf: type LO: True)
  → (? x LO: cannotTransportWith ? y) ]
[ explosiveRule: (? x LO: isExplosive ? b) (? y LO: isFlammable ? b) (? b rdf: type LO: True)
  → (? x LO: cannotTransportWith ? y) ]
[ poisonousRule: (? x rdf: type LO: PoisonousCargo) (? y isPoisonous ? b) (? b rdf: type LO: False)
  → (? x LO: cannotTransportWith ? y) ]
[ dangerousTransportRule: (? x rdf: type LO: DangerousCargo)
  → (? x LO: isTranportedBy LO: SpecialVehicle) ]
```

规则库独立于应用程序,所以规则可以动态地增加、删除、修改,而不会对应用程序产生影响。

4) 领域本体库的构建 领域本体的开发使用的是 Protégé 3.1, 利用从各个库中的结构化和半结构化的数据完成本体的建模,构建物流领域的本体库,并采用德国汉堡大学计算机科学系所开发的描述逻辑推理机 RACER 对本体进行推理。

图 5 所示是物流本体模型的整体结构,依据该模型目前实现的顶层本体及其派生的领域本体模型中有 171 个类和 300 个属性,是对物流领域建立的一个核心的本体模型,涉及到货物、仓库、运输工具、物流模式、公司等几个物流子领域。

5) 本体整合 对于物流行业这样一个较大的领域来说,建立一个大而全的本体几乎是不可能,不同的人和组织可能根据自己的需要构建不同的物流领域本体。由于不同的人们和组织可能具有不同的世界观,这些不同的世界观甚至有可能是冲突的,这就必然导致本体具有分布和异构的属性,并且本体本身并不能协调这种异构性,这种异构性为相同或不同领域的知识系统之间的互操作性带来了严重的问题。本体整合是解决本体自身语义异构性的一种重要手段。图 6 所示的是基于 Jena 开发的本体整合工具 OntMerge, 它采用实体内外部特征分析相结合的方式来计算实体间的相似度,并以此为依据进行本

体整合。针对实体内部特征中缺少语义理解的问题,在使用字符串比较方法的基础上引入字典对概念的语义进行识别。同时,该工具综合语法和语义两个方面得到的结果作为内部特征相似度,并通过字典技术克服本体层中同义词、近义词的问题;其次,针对实体的外部特征,采用启发式规则计算本体中实体间的结构相似度,进而可以判断两个实体在本体中的地位是否相似。通过综合使用这两种方式可以有效地识别出待整合本体中相似的部分,提高了本体整合的准确性。

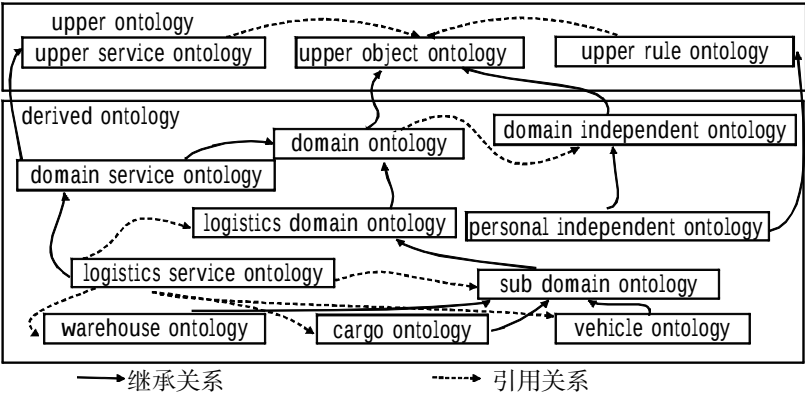


图 5 物流本体模型整体结构

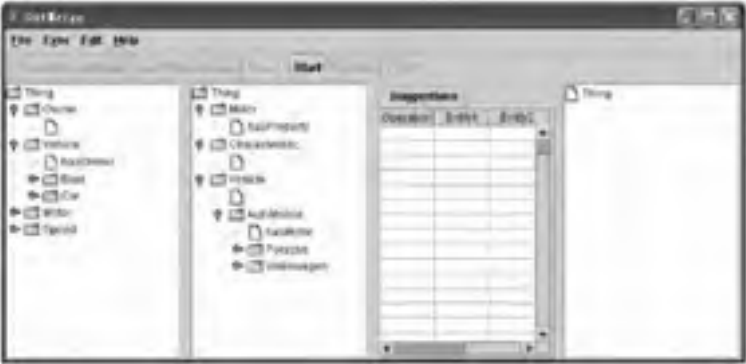


图 6 本体整合工具 OntMerge

6) 知识的语义标注 语义标注就是利用领域本体对原始数据作标记,使其具有语义信息,是使机器理解所处理数据的一种手段。语义标注的过程就是创建机器可理解的知识的过 程。知识提供了领域(如物流领域)内部不同主体(人、机器、软件系统等)之间进行交流(对话、互操作、共享等)的一种语义基础。在语义 Web 中,这种知识主要是面向机器理解的,以本体形式表达的。物流行业的知识库大多是非结构化和半结构化的,非常不适用于查询与复用。本文利用本体库通过语义标注工具^[20]对物流行业成千上万的知识源进行语义标注,就是利用本体对那些结构化和非结构化的知识进行本体实例化,从而为这些知识添加了语义信息,在一定程度上让机器能理解知识,为知识库的复用创造更有利的条件。

根据不同的用途进行语义标注,需要明确要标注的对象。从本体的角度来说,就是明确将本体中的哪些成分与现实世界建立联系,即将哪些知识标注为本体的实例,因此会涉及标注的粒度问题。本体标注的粒度可根据需要由粗到细分为: a) 标注概念,即建立概念和个体间的联系,一个个体是本体中某个概念的实例,那么它就会继承该概念的属性。 b) 标注属性,即给出实例对应的属性值,这样有利于进行推理和查询。 c) 标注关系,即指出实例间的关系,本体建模时仅会指出关系的定义域和值域,在进行关系标注时需要根据相应的规则和条件实例化这种关系。 d) 标注实例,如一个实例可能会出现有不同的 URI,这就需要通过语义标注来标志出不同的 URI。

7) 应用接口 该层主要是为不同的使用者提供 API、Web 服务和控制接口等提供各种各样的应用接口,主要是采用 WSDL 将相应的经过标注的知识库包装成 Web service,并将相应

的知识库平台的组件发布为可供调用的 API;同时提供控制接口,为用户提供个性化的服务,方便不同的用户查询和应用,从而更好地为不同的使用者服务。

4 结束语

物流行业知识库可以大大提高物流知识的准确程度和使用效率;尽可能地复用知识,有利于行业信息化的推进。而 ontology 形式表达了领域知识的基本概念、属性、处理方法和内在关系,在知识表示和推理方面很有优势,是构建知识库的理想技术。本文在分析物流信息化的需求和现状之后,指出了知识库在物流行业中的作用和地位;在研究了知识库与物流知识库相关理论和技术的基础上,建立了一种结合本体和 SOA 技术的物流知识库模型,并实现了系统的关键功能。但是目前所做的工作还不彻底,有许多问题值得进一步研究,进一步对知识库系统的完备性进行验证,系统的细节还需进一步的研究,也需要在实际应用中得到更好的完善。

参考文献:

[1] 国家发展改革委员会,国家统计局,中国物流与采购联合会. 2006 年全国物流运行情况通报[EB/OL] . (2007- 05-10) . http://www.ndrc.gov.cn/zjgx/t20070315_121530.htm.

[2] WHITE C C III. A survey on the integration of decision analysis and expert systems for decision support[J] . IEEE Trans on Systems, Man and Cybernetics, 1990, 20(2) : 358-364.

[3] RAPHAEL M J, DELOACH S A. A knowledge base for knowledge-based multiagent system construction[C] //Proc of IEEE National Aerospace and Electronics Conference. Dayton: IEEE, 2000: 383-390.

[4] GUO J, LUQI. A survey of software reuse repositories[C] //Proc of the 7th IEEE International Conference and Workshop on the Engineering of Computer Based Systems. Edinburgh: IEEE, 2000: 92-100.

[5] DOCTOR G, RAMACHANDRAN S. enabling knowledge sharing with an institutional repository[C] // Proc of the 7th IEEE International Conference on Advanced Learning Technologies. Southampton: IEEE, 2007: 432-433.

[6] QI Xu. A framework for logistics strategy planning based on knowledge syntheses[C] //Proc of International Conference on Computational Intelligence and Security. Guangzhou: IEEE, 2006: 933-936.

[7] HSU Pang-yen, MA Shao-ping, SUN Fu-chun, et al. Implementation of a knowledge-based system for automatic part numbering in a logistics information system[C] //Proc of IEEE International Conference on Service Operations and Logistics, and Informatics. Shanghai: IEEE, 2006: 1024-1029.

[8] GIANNOTTI E, NEGRON R, ANTZOULIS T. An integrated information system approach for performance based logistics[C] //Proc of IEEE Systems Readiness Technology Conference. Anaheim: IEEE, 2006: 336-340.

[9] 袁磊,张浩,陈静,等. 基于本体化知识模型的知识库构建模式研究[J] . 计算机工程与应用, 2006, 42(30) : 65-68, 104.

[10] Description of OIL[EB/OL] . [2007-11- 22] . <http://www.ontoknowledge.org/oil/>.

[11] OntoEdit[EB/OL] . [2007-11- 22] . <http://www.ontoknowledge.org/tools/ontoedit.shtml>.

[12] Z NIGA G L. Ontology: Its transformation from philosophy to information systems[C] //Proc of the 2nd International Conference on Formal Ontology in Information Systems. New York: ACM Press, 2000: 187-197.

(下转第 2740 页)

第 i 个指标的第 j 个数据的实际值; x_{imax} 为第 i 个指标所有数据中的最大值(如对第 2 个指标, $x_{imax} = 2\ 977.61$)。对于最大值的归一化结果, 考虑到 SVR 能够获得较好的收敛性, 其取值为 0.999 9。结果如表 1 所示。将 1978—1995 年的样本作为训练样本, 1996—2003 年的样本作为测试样本。

表 1 上海市物流需求规模及经济指标统计数据

年 份	指 标						
	x_1 第 一产业产 值 /亿元	x_2 第 二产业产 值 /亿元	x_3 第 三产业产 值 /亿元	x_4 区 域 零 售 总 额 / 亿 元	x_5 区 域 外 贸 总 额 / 亿 美 元	x_6 居 民 消 费 水 平 /元	y 货 物 运 输 量 / 万 吨
1978	0.121 9	0.070 9	0.017 1	0.024 4	0.027 5	0.029 7	
1979	0.126 2	0.074 3	0.018 1	0.030 8	0.035 3	0.034 4	0.270 1
1980	0.111 9	0.079 3	0.022 1	0.036 2	0.041	0.037 2	0.276
1981	0.117 2	0.082 1	0.023 5	0.04	0.037 8	0.039 3	0.277 5
1982	0.147 4	0.083 7	0.025	0.040 4	0.035 4	0.038 7	0.291 3
1983	0.149 8	0.085 7	0.027 9	0.045 3	0.037 7	0.041 4	0.297 4
1984	0.191 2	0.092 5	0.033	0.055 7	0.04	0.048 8	0.318 4
1985	0.216 4	0.109 4	0.040 9	0.078 1	0.047 1	0.069 3	0.333 9
1986	0.218 1	0.112 8	0.045 4	0.088 7	0.047 4	0.08	0.367 3
1987	0.239 3	0.122 4	0.053 6	0.101 4	0.054 6	0.087 3	0.375 2
1988	0.303 1	0.145 4	0.063 1	0.133 2	0.065 9	0.113	0.383 3
1989	0.328 2	0.156 6	0.067 4	0.149 2	0.071 4	0.062 4	0.381
1990	0.361 1	0.162 1	0.081	0.150 4	0.067 6	0.135 1	0.368 8
1991	0.369 6	0.185 2	0.103 8	0.172 1	0.073 2	0.162 8	0.379 6
1992	0.378 4	0.227 5	0.135 3	0.209 3	0.088 8	0.191 2	0.407 4
1993	0.423 3	0.302 4	0.192 5	0.281 2	0.115 9	0.279 9	0.417 2
1994	0.538 3	0.383 9	0.262 1	0.347 1	0.144 4	0.359 4	0.393 7
1995	0.683 3	0.473 5	0.333	0.436 9	0.173 2	0.451 5	0.379 7
1996	0.793 0	0.531 5	0.431 4	0.523	0.202 6	0.520 8	0.631 1
1997	0.839 7	0.589 2	0.514 1	0.596 8	0.225 4	0.585 1	0.632 7
1998	0.869 6	0.620 4	0.592 2	0.662 5	0.285 3	0.619	0.636 7
1999	0.886 2	0.656 2	0.672 3	0.716 3	0.351 4	0.694 7	0.666 6
2000	0.921 7	0.726 7	0.774 2	0.775 7	0.498	0.776 6	0.719
2001	0.947 2	0.791 1	0.843 3	0.838 3	0.554 3	0.845	0.747 7
2002	0.977 5	0.861 3	0.925 9	0.916 6	0.661 4	0.961 5	0.780 3
2003	0.999 9	0.999 9	0.999 9	0.999 9	0.999 9	0.999 9	0.841 2
2004							0.999 9

3.3 仿真结果及分析

针对上述的“影响因素—区域物流需求”SVR 预测模型, 在 MATLAB 6.5 环境下进行训练与测试的仿真实验。SVR 的核函数和参数采用折交叉验证方法进行优化选择, 为 $K(x_i, x_j) = ((x_i \times x_j) + 1)^2$, $\varepsilon = 0.12$, $C = 0.05$ 。测试样本的仿真结果如表 2 所示。总体上说, 预测结果能满足预测要求, 误差比较小, 拟合程度较高。

本文阐述的“影响因素—区域物流需求”SVR 预测模型, 是建立在对区域物流需求与其影响因素(主要是区域经济) 之间内在关系研究的基础上提出的。该模型在一定程度上反映了区域物流需求与其影响因素之间的复杂映射关系, 它不同于以往简单的物流需求线性回归或时间序列预测模型。

表 2 上海市物流需求规模预测结果

年 份	真 实 值	预 测 值	误 差	误 差 率 /%
1997	45 938	40 515.07	- 5 422.93	- 11.804 9
1998	46 230	44 249	- 1 981	- 4.285 1
1999	48 398	47 208.33	- 1 189.67	- 2.458 1
2000	52 206	50 460.39	- 1 745.61	- 3.343 7
2001	54 049	55 232.57	1 183.57	2.189 8
2002	56 652	59 512.87	2 860.87	5.049 9
2003	61 073	65 187.73	4 114.73	6.737 4
2004	72 605	72 330.12	- 274.88	- 0.378 6

4 结束语

本文将 SVR 方法应用于区域物流需求预测研究, 从区域经济等影响因素指标与区域物流需求之间的内在关系的角度, 建立了“影响因素—区域物流需求”SVR 预测模型, 采用 ε -insensitive 损失函数, 选用多项式核函数, 其实际结果有相当理想的精度, 能满足预测要求。“影响因素—区域物流需求”SVR 预测模型有很强的学习性、自适应性, 且收敛快、准确性高, 说明该模型是可信的。但对核函数的选择, 参数和损失函数的选取没有确定的方法, 主要靠使用者的经验, 难以找到全局最优解, 只能找到满意解, 对此问题还需要进一步研究。

参考文献:

[1] ARRIDO R A, MAHMASSANI H S. Forecasting freight transportation demand with the space-time multinomial probity model[J]. Trans on Research Part B, 2000, 34(5) : 403-418.

[2] FITE J, TAYLOR G, USHER J, *et al.* Forecasting freight demand using economic indices[J]. International Journal of Physical Distribution & Logistics Management, 2001, 31(4) : 299.

[3] 张拥军, 叶怀珍, 任民. 神经网络模型预测运输货运量[J]. 西南交通大学学报, 1999, 34(5) : 602-605.

[4] 王隆基, 张仲鹏, 孙晓霞. 基于 BP 神经网络的物流预测方法[J]. 起重运输机械, 2005(5) : 30-32.

[5] 汪宇瀚. 预测物流需求的一元线性回归分析法[J]. 商场现代化, 2006(34) : 136.

[6] 陈森, 周峰. 基于灰色系统理论的物流需求预测模型[J]. 统计与决策, 2006(3) : 59-60.

[7] 初良勇, 田质广, 谢新连. 组合预测模型在物流需求预测中的应用[J]. 大连海事大学学报, 2004, 30(4) : 43-46.

[8] 陈黎. 我国区域物流发展预测[J]. 统计与决策, 2006(12) : 127-129.

[9] 后锐, 张毕西. 基于 MLP 神经网络的区域物流需求预测方法及其应用[J]. 系统工程理论与实践, 2005(12) : 43-47.

[10] VAPNIK V. An overview of statistical learning theory[J]. IEEE Trans on NN, 1999, 10(3) : 988-999.

[11] CRISTIANINI N, SHAWE-TAYLOR J. 支持向量机导论[M]. 李国正, 王猛, 曾华军, 译. 北京: 电子工业出版社, 2004.

[12] 夏国恩, 曾绍华, 金炜东. 支持向量回归机在铁路客运量时间序列预测中的应用[J]. 计算机应用研究, 2006, 23(10) : 180-182.

(上接第 2737 页)

[13] IU Xiao-li, WU Guo-qing, YANG Min. Building knowledge base from semi-structured data[C] //Proc of International Conference on Machine Learning and Cybemetics. Hong Kong: IEEE, 2007: 839-843.

[14] XU Bin-feng, LUO Xiao-gang, PENG Cheng-lin, *et al.* Based on ontology: construction and application of medical knowledge base[C] //Proc of IEEE/ICME International Conference on Complex Medical Engineering. Beijing: IEEE, 2007: 1586-1589.

[15] 范黎林, 王晓东, 屈喜龙. 基于 Ontology 知识库系统建模[J]. 计算机应用研究, 2006, 23(5) : 134-136.

[16] 何庆, 汤庸, 黄永钊. 基于本体的法律知识库的研究与实现[J]. 计算机科学, 2007, 34(2) : 175-177.

[17] NEWCOMER E, LOMOW G. Understanding SOA with Web services [M]. [S. l.] : Addison-Wesley Professional, 2004.

[18] McINTOSH, ROGER L. Open-source tools for distributed device control within a service-oriented architecture[J]. Journal of the Association for Laboratory Automation, 2004, 9(6) : 404-410.

[19] KNIGHT, W. Security: built-in or bolted-on to the SOA? [J]. Infosecurity Today, 2005, 2(2) : 38-40.

[20] 廖述梅. 基于本体的语义标注原型评述[J]. 计算机工程与科学, 2006, 28(9) : 123-125, 128.