

Research on Knowledge Graph of Power Material Supplier Evaluation Based on Multi Data Sources

Hengliang Qu¹, Kan Xu², Na Jin³, Jinli Zhang²

¹State Grid Shanghai Electric Power Company Pudong Power Supply Company, Shanghai

²State Grid Shanghai Electric Power Company Electric Power Research Institute, Shanghai

³State Grid Shanghai Electric Power Company Jinshan Power Supply Company, Shanghai
Email: honeylily0629@163.com

Received: Mar. 30th, 2020; accepted: Apr. 14th, 2020; published: Apr. 21st, 2020

Abstract

In this paper, the key technology of knowledge graph is studied, and a process of building knowledge graph of power material supplier evaluation based on multi data source fusion is proposed. The top-down knowledge graph is used for ontology construction, knowledge extraction, knowledge representation, knowledge fusion, and finally knowledge storage to realize the knowledge graph of supplier evaluation in different information systems, which is convenient for query and visual display. It can summarize the supplier evaluation data from multiple dimensions and get comprehensive evaluation, which provides a solid data basis for the formulation of precise manufacturing supervision strategy for suppliers.

Keywords

Knowledge Graph, Supplier Evaluation, Multiple Data Sources

基于多数据源的电力物资供应商评价知识图谱研究

瞿恒亮¹, 许侃², 金娜³, 张金丽²

¹国网上海市电力公司浦东供电公司, 上海

²国网上海市电力公司电力科学研究院, 上海

³国网上海市电力公司金山供电公司, 上海

Email: honeylily0629@163.com

收稿日期: 2020年3月30日; 录用日期: 2020年4月14日; 发布日期: 2020年4月21日

摘要

本文对知识图谱关键技术进行研究,提出一种多数据源融合的电力物资供应商评价知识图谱构建流程。采用自顶向下知识图谱进行本体构建,再进行知识抽取、知识表示、知识融合,最后进行知识存储。实现对不同信息系统内的供应商评价知识图谱,方便查询和可视化展示。实现从多个维度对供应商评价数据进行汇总,得到综合评价,为对供应商制定精准监造策略提供了坚实的数据依据。

关键词

知识图谱, 供应商评价, 多数据源

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

电力物资供应商评价知识图谱是指在物资全寿命质量信息收集整理的基础上,以供应商为维度,对质量信息进行归纳分析,实现对供应商多个维度的评价,并将分析结果进行可视化展示。国网公司采购的电力物资种类多达上万种,每一种物资的供应商也有多家,供应商也多达上万家。国网公司三大物资供应链九大业务,从计划、招标、合同签订、履约、仓储、应急、废旧、质量监督和供应商管理,每个环节都存在对供应商进行评价的数据。这些数据分散于不同的信息系统中,数据的来源多种多样,数据的格式分为三种:结构化数据,来源于 ERP 系统、电商平台 ECP、辅助工具系统的数据库;半结构化数据,来源于 XML 文件、JSON 文件等;非结构化数据,主要是各种 Word 文件、Excel 文件、PDF 文件、图片、视频、音频等格式文件。

每年国网公司物资采购进行招投标,供应商的历史评价数据对供应商进行客观评价有很重要的参考价值。由于这些数据散落在 ERP、ECP 等各个系统中,目前没有进行数据集成和综合分析,且部分数据在系统中是以文本形式存在,标准化、结构化程度较差,很难对供应商数据进行全面展示和利用。当对供应商进行评估时,导致定量指标的可操作性不强,容易变成人为主观打分,影响指标的合理性和客观性。

近年来,知识图谱作为一种新的知识表示方法和数据管理模式,在自然语言处理、问题回答、信息检索等领域有着重要的应用。本文研究提出一种多数据源融合的电力物资供应商评价知识图谱构建流程,并对关键技术进行研究,采用自顶向下知识图谱进行本体构建,再进行知识抽取、知识表示、知识融合,最后进行知识存储。实现对不同信息系统内的供应商评价进行知识融合,方便查询和可视化展示,实现从多个维度对供应商进行综合评价,为对供应商实行精准监造提供了坚实的数据依据。

2. 电力物资供应商评价知识图谱

2.1. 供应商评价知识图谱构建方法

知识图谱(Knowledge Graph)的概念由谷歌 2012 年正式提出,旨在实现更智能的搜索引擎,并且于 2013 年以后开始在学术界和业界普及。知识图谱本质上是一种叫做语义网络(semantic network)的知识库,

即具有有向图结构的一个知识库，其中图的结点代表实体(entity)或者概念(concept)，而图的边代表实体/概念之间的各种语义关系，比如说两个实体之间的相似关系[1]。知识图谱的基本单位，便是“实体(Entity) - 关系(Relationship) - 实体(Entity)”构成的三元组，这也是知识图谱的核心。

常用的知识图谱技术架构主要采用自底向上的构建技术[2]，如图 1 所示。其包括四大部分：数据采集、信息抽取、知识融合和知识加工。

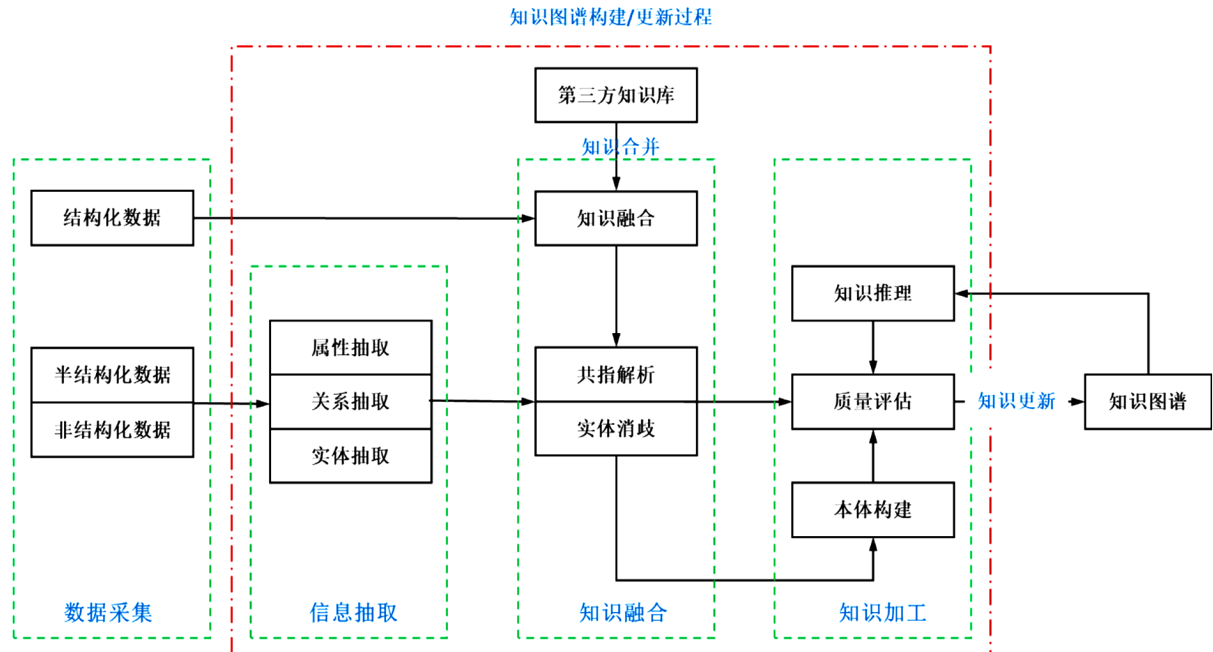


Figure 1. Technical framework of knowledge graph
图 1. 知识图谱技术架构

构建知识图谱是一个迭代更新的过程，根据知识获取的逻辑，每一轮迭代包含三个阶段：

信息抽取：从各种类型的数据源中提取出实体、属性以及实体间的相互关系，在此基础上形成本体化的知识表达。信息抽取是一种自动化地从半结构化和无结构数据中抽取实体、关系以及实体属性等结构化信息的技术。涉及的关键技术包括：实体抽取、关系抽取和属性抽取。

知识融合：在获得新知识之后，需要对其进行整合，以消除矛盾和歧义，比如某些实体可能有多种表达，某个特定称谓也许对应于多个不同的实体等。知识融合包括两部分内容：实体链接和知识合并。

知识加工：对于经过融合的新知识，需要经过质量评估之后(部分需要人工参与甄别)，才能将合格的部分加入到知识库中，以确保知识库的质量，知识加工主要包括三方面内容：本体构建、知识推理和质量评估。

2.2. 供应商评价知识图谱构建

知识图谱主要有自顶向下与自底向上两种构建方式。自顶向下指的是先为知识图谱定义好本体与数据模式，再将实体加入到知识库，该构建方式需要利用一些现有的结构化知识库作为其基础知识库。由于电力物资供应商评价的结构化数据较好，数据逻辑清晰，结构简单，因此本文采用自顶向下知识图谱进行本体构建，再进行知识抽取、知识表示、知识融合，最后进行知识存储[3]。知识图谱构建基本流程，如图 2 所示。

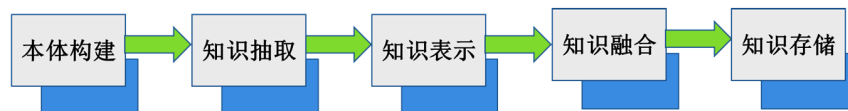


Figure 2. Basic process of knowledge graph construction

图 2. 知识图谱构建基本流程

本文多数据源融合的知识图谱构建过程，如图 3 所示：

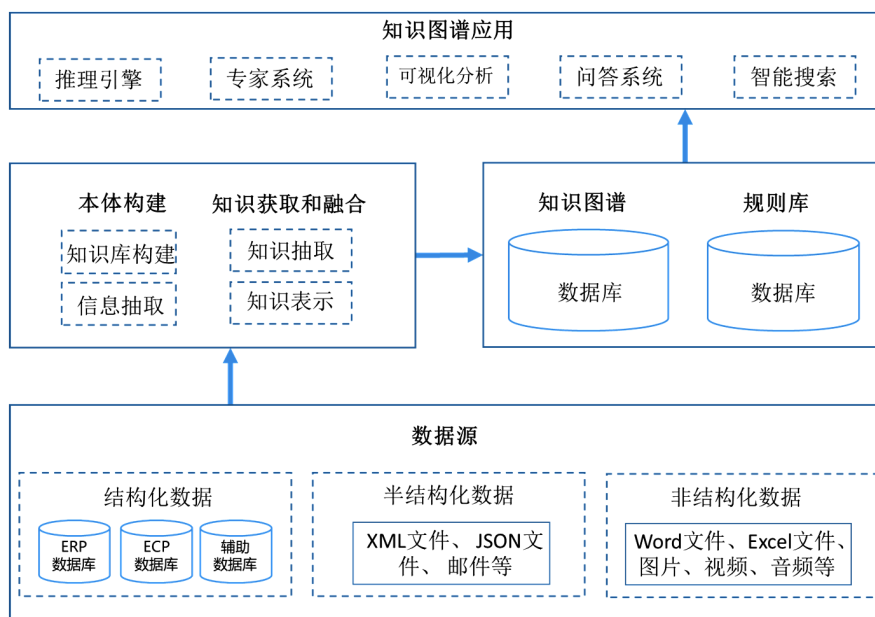


Figure 3. The construction process of knowledge graph based on multiple data sources

图 3. 基于多数据源的知识图谱构建过程

数据源

用于构建知识图谱的本体库数据源可以来源于结构化数据、半结构化数据和非结构化数据，以及现有的一些通用知识图谱库等。国网公司供应商评价数据用于构建本体库的数据源如下：

- 1) 结构化数据。主要来源于关系数据库，如：ERP 系统、电商平台 ECP 系统、智慧供应链辅助系统。
- 2) 半结构化数据。主要来源于：XML 文件、JSON 文件等。
- 3) 非结构化数据。数据主要来源于文本报告资料的数据，如 WORD、EXCEL、TXT、PDF、PPT，各种格式的图片、视频等。

对于结构化数据，本文应用 ETL 工具对数据进行抽取、转换、清洗、加载到知识本体库中；对于半结构化数据，通过工具转换成结构化数据，再通过 ETL 工具进行数据处理；对于非结构化数据，通过人工将能转化的数据转换成结构化数据，同时保留原始文件的链接关系，能直接打开和浏览原文件。最后通过 ETL 工具实现将多数据源数据加载到知识库本体库中。

本体构建

本体(ontology)是对概念进行建模的规范，是描述客观世界的抽象模型，以形式化方式对概念及其之间的联系给出明确的定义。本体定义了知识图谱中的数据模式，因而，本体构建研究的成果能在很大程度上辅助知识图谱的构建[4]。针对不同的应用领域和不同的需求，本体构建的方法也有所不同。本文从多种数据源中构建相应的知识本体库，然后通过映射成全局本体库。

知识图谱分为数据层和模式层，数据层是由一系列事实数据构成，而模式层则用来从数据中构建实体、属性、关系，是知识图谱的核心。本体层实际就是模式层(schema层)，schema采用关系型数据库。高质量、标准化的schema能有效降低领域数据之间对接的成本。构建schema层的方法采用自上而下的构建：本体层的最顶层也就是最抽象的一层向下逐渐细化、逐渐构建。本体层的最底层往往细化到一个概念。

Supplier (Entity)-First level index (Relationship)-Manufacturing (Entity);

即：供应商(实体) - 一级指标(关系) - 生产制造(实体);

Manufacturing (Entity)-Two level index (Relationship)-Production process (Entity);

即：生产制造(实体) - 二级指标(关系) - 生产工艺(实体);

以此类推，层层自顶向下构建供应商评价本体知识库。

知识获取和融合

多数据源存在分散、异构、自治的形式存在，另外还具有冗余、噪音、不确定、非完备的特点，清洗并不能解决这些问题，因此从这些知识出发，通常需要融合和验证的步骤，将不同源、不同结构的数据融合成统一的知识图谱，以保证知识的一致性。

知识融合：将不同数据源获取的知识进行融合构建数据之间的关联。包括实体对齐、属性对齐、冲突消解、规范化等。在获得新知识之后，需要对其进行整合，以消除矛盾和歧义，比如某些实体可能有多种表达，某个特定称谓也许对应于多个不同的实体等。通过实体链接、关系链接、属性链接以及属性值链接等方法进行知识融合。

知识加工：对于经过融合的新知识，需要经过质量评估之后，特别是抽取非结构化数据需要人工参与甄别，才能将合格的部分加入到知识库中，以确保知识库的质量。例如将上面供应商本体库与生产制造本体库进行知识抽取和知识融合，其知识表示形式如下所示：

Supplier (Entity)-First level index (Relationship)-Manufacturing (Entity)-Two level index (Relationship)-Production process (Entity);

即：供应商(实体) - 一级指标(关系) - 生产制造(实体) - 二级指标(关系) - 生产工艺(实体)。

知识图谱

首先对关系型知识图谱数据库MySQL进行初始化，创建数据库表、视图、存储过程等。将“实体-关系-实体”三元组，以二叉树或森林的数据结构保存到知识图谱数据库，以便进行各种应用。

规则库

规则(Rule)就是描述某领域内知识的产生式规则的集合。本文规则库主要存储供应商监造策略措施，即根据供应商每项评价指标对应设置相应的监造措施。监造策略可以是一条措施，也可以包含多条措施。

推理引擎

本文知识库由知识图谱库和规则库构成，是问题求解所需要的领域知识的集合。推理引擎主要包含三个部分：模式匹配器(Pattern Matcher)、议程(Agenda)和执行引擎(Execution Engine)。推理引擎是实施问题求解的核心执行机构，它实际上是对知识进行解释的程序，根据知识的语义，对按一定策略找到的知识进行解释执行，并输出结果。

专家系统

知识库是问题求解所需要的领域知识的集合，包括基本事实、规则和其他有关信息。知识库是专家系统的核心组成部分。构建由知识图谱库与规则库组成的知识库，按照知识图谱构建方法，建立供应商评价知识图谱。通过推理引擎算法，实现对知识图谱库和监造措施库智能匹配，从而输出基于供应商评价的监造策略。

3. 供应商评价知识图谱库的应用

供应商评价指标体系是国网公司对供应商实行差异化监造策略的最重要的依据，由于供应商绩效评价数据散落在 ERP、ECP、辅助系统等各个信息系统中，没有进行数据集成和综合分析，且部分数据在系统中是以文本形式存在，标准化、结构化程度较差。因此，对供应商实行精准监造一直处于探索阶段。现在通过建立基于多数据源的电力物资供应商评价知识图谱，很容易获得供应商所有评价数据，为对供应商实行精准监造提供了坚实的数据依据。

3.1. 供应商评价知识图谱

根据《国家电网有限公司物资质量监督管理办法》和《国家电网有限公司供应商绩效评价管理细则》，本文以供应商绩效评价指标为例，说明供应商评价知识图谱在监造策略中的应用。电力物资供应商是监造的对象，也是监造策略的重点实施对象。电力物资供应商评价指标体系包括五个方面指标进行评价：供应商的资质能力、绩效评价、履约评价、供需依存度、供应商负面评价。具体到每一个指标又包括多个细分的评价指标。以供应商绩效评价指标为例子，依据业务需求报告内容，建立供应商绩效评价模型。供应商绩效评价指标包括：生产制造阶段质量信息分析、安装调试阶段质量信息分析和运行维护阶段质量信息分析。供应商绩效指标评价体系，见表 1。

Table 1. Supplier performance index evaluation system

表 1. 供应商绩效指标评价体系

序号	供应商名称	一级指标	权重	二级指标	权重	三级指标	权重	专家评分
1				生产制造阶段	40%	原材料	10%	5
						组部件	20%	5
						生产工艺	30%	4.5
						出厂试验	15%	5
						抽检	25%	4.8
2	XX 供应商评价	绩效评价	40%	运输交接阶段	30%	包装发运	30%	4.6
						供应商基建评价	20%	5
3				安装调试阶段	20%	到货验收	50%	4.6
						设备安装	70%	5
4				运行维护阶段	10%	交接试验	30%	5
						供应商运行阶段评价	60%	4.6
						故障发生率	40%	4.5

根据上面建立的供应商知识图谱数据库，通过自底向上方法，很容易获取并汇总某类物资供应商各个指标专家评分数据，然后按照统计模型用指标权重乘以评分，算出带权重指标评分数据，最后将各个指标评分数据累加就得到供应商综合评价数据。

监造策略库示例，见表 2。

供应商评价知识图谱专家系统通过推理引擎进行算法，以供应商评价为核心对供应商知识图谱库和监造策略库(规则库)进行智能匹配。最终自动输出供应商监造策略，为供应商制定精准监造策略提供智能化工具。

Table 2. Supervision strategy library
表 2. 监造策略库

供应商绩效评价	电压等级	物料类别	监造策略
优秀(4.5~5 分)	110 KV	电抗器	1、减少检查。
	220 KV	电抗器	1、减少检查。
良好(4~4.5 分)	110 KV	电抗器	1、增加 R 点。
	220 KV	电抗器	1、增加 W 点。
合格(3~4 分)	110 KV	电抗器	1、增加 H 点。 2、增加生产环境专项检查。
	220 KV	电抗器	1、增加 H 点。 2、原材料组部件厂内抽检。
问题(1~3 分)	110 KV	电抗器	1、增加 H 点。 2、增加飞行检查。
	220 KV	电抗器	1、增加 H 点。 2、增加出厂试验项目。
风险(0 分)	110 KV	电抗器	1、组织专家专项评审。 2、建立定期召开协调会机制。
	220 KV	电抗器	1、组织专家专项评审。 2、建立定期召开协调会机制。

智能搜索及数据可视化分析

基于知识图谱的智能搜索是一种基于长尾的搜索，搜索引擎以知识卡片的形式将搜索结果展现出来。例如，按物资类别搜索“110 KV 变压器”，按供应商名称搜索“XX 供应商”，知识图谱系统根据用户的查询请求，经过查询式分析后的标准查询语句进入知识库检索引擎，引擎会在知识库中检索相应的实体以及与其在类别、关系、相关性等方面匹配度较高的实体。通过对知识库的深层挖掘与提炼后，引擎将给出具有重要性排序的完整知识体系，并以直观的方式展现给用户。供应商评价知识图谱数据可视化分析，如图 4 所示。

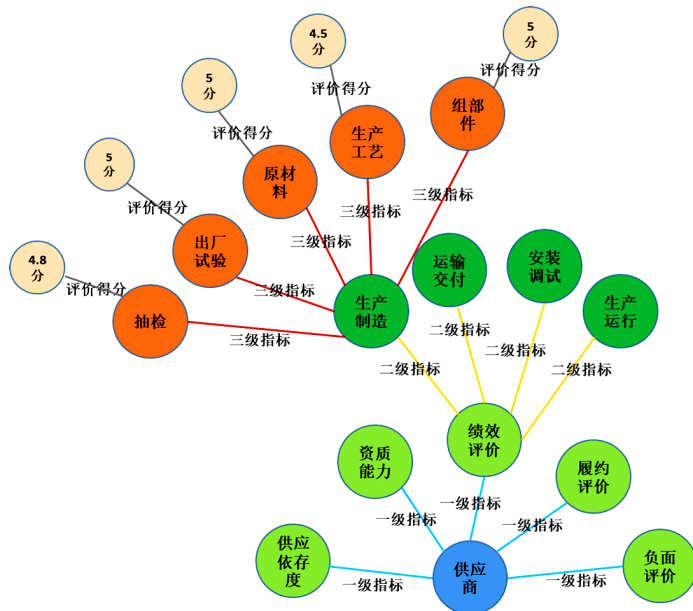


Figure 4. Visual analysis of supplier evaluation knowledge graph
图 4. 供应商评价知识图谱可视化分析

3.2. 效果比较

本文针对 12 个供应商, 采取不同方式制定供应商监造策略。一种是通过人工专家根据供应商评价手工制定监造策略; 另一种是通过供应商知识图谱系统, 通过专家系统自动输出供应商监造策略。从四个维度进行比较: 监造策略制定花费时间、监造策略完备性、监造策略客观性、人力资源。应用效果比较, 见表 3。

Table 3. Application effect comparison

表 3. 应用效果比较

比较对象	监造策略制定花费时间	监造策略完备性	监造策略客观性	人力资源
行业专家(人工)	3 个工作日	有缺失、遗漏	凭个人经验	需要有行业经验的专家
供应商知识图谱系统	5 分钟	较完备	较客观	普通技术人员

人工专家制定监造策略与系统制定监造策略效果比较:

1) 监造策略制定花费时间: 行业专家需要花费 3 个工作日搜集、整理相关资料, 才能完成一个供应商的监造策略制定; 供应商知识图谱系统只需要 5 分钟就能完成输出监造策略制定工作。极大地提高了工作效率。

2) 监造策略完备性: 人工完成资料搜集和整理, 存在缺失和遗漏的情况, 造成减少一些监造环节; 系统只要输入的供应商资料是完整的, 就不存在这种情况, 完备性较好。

3) 监造策略客观性: 专家毕竟都是人, 即使针对很客观的事物, 每个人都有自己的偏好, 对同一件事物的看法不完全一样, 存在差异性; 而知识图谱系统只根据供应商每项评分自动进行监造策略匹配, 不存在个人偏见, 因此客观性较好。

4) 人力资源: 供应商监造策略制定, 需要工作人员具有一定工作经验积累, 才能较客观地制定监造策略, 任何时候行业专家都是一种稀缺资源。而知识图谱系统对个人的经验依赖较少, 一般技术人员也能很好完成。

4. 结束语

本文提出一种多数据源融合的电力物资供应商评价知识图谱构建流程, 并对关键技术进行研究。采用自顶向下知识图谱进行本体构建, 再进行知识抽取、知识表示、知识融合, 最后进行知识存储。实现对不同信息系统内的供应商评价知识图谱, 方便查询和可视化展示。实现从多个维度对供应商进行汇总数据, 获得综合评价, 为对供应商制定精准监造策略提供了坚实的数据依据。通过将供应商知识图谱数据库与供应商监造策略数据库结合起来, 实现供应商评价与监造策略的智能匹配, 自动输出精准监造策略, 实现对供应商精准监造。

参考文献

- [1] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
- [2] 徐增林, 盛泳潘, 贺丽荣, 王雅芳. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 590-600.
- [3] 黄兆玮. 旅游知识图谱构建研究[D]: [硕士学位论文]. 桂林: 桂林电子科技大学, 2018: 6-7.
- [4] 胡芳槐. 基于多种数据源的中文知识图谱构建方法研究[D]: [博士学位论文]. 上海: 华东理工大学, 2015.