

The Research on Construction of Power Grid Marketing Data Atlas Based on Data Atlas Technology

Xinbo Yin

State Grid Changzhou Electric Power Company, Changzhou Jiangsu
Email: 148312982@qq.com

Received: Jul. 22nd, 2019; accepted: Aug. 2nd, 2019; published: Aug. 9th, 2019

Abstract

In dispatching and marketing systems, the network structures of stations, lines, transformers and households are stored separately. However, due to the inconsistency of data standards and business concerns in each system, the data are isolated and inconsistent with each other, resulting in errors in the relationship between network archives and archives in a single system. It is easy to cause inaccurate grid topology. Knowledge atlas technology is developed based on large data graph database technology. It can construct power grid topology model intuitively and efficiently through semantic recognition technology and multi-data knowledge extraction technology, realize automatic identification of power grid topology, and provide strong conditions for standard management of power grid structure. This paper constructs automatic recognition model of the marketing data map, focuses on the application of Chang Zhou power supply company's marketing data map, can realize the difference analysis of the cross-system grid structure, accurately locate the cause of the data problem, and improve the efficiency of data management of the grid topology.

Keywords

Network Frame, Power Network Topology, Standard Management, Automatic Identification, Marketing System

基于数据图谱技术的电网营销数据图谱构建研究

殷新博

国网江苏省电力有限公司常州供电分公司, 江苏 常州
Email: 148312982@qq.com

收稿日期：2019年7月22日；录用日期：2019年8月2日；发布日期：2019年8月9日

摘要

调度、营销系统中分别存储了站、线、变、户的网架结构，但由于各系统数据标准和业务关注度不一致，使得数据间彼此孤立、不统一，导致单系统中网架档案本身和档案间关系错误，以及跨系统网架档案一致性和关系错误，最终造成电网拓扑不准确。知识图谱技术是基于大数据图数据库技术发展起来的，可以通过语义识别技术、多元数据知识抽取技术直观高效地构建出电网拓扑架构模型，实现电网拓扑的自动识别，为实现电网架构的标准管理提供有力条件。这篇文章从数据图谱技术出发，构建了营销数据图谱自动识别模型，重点研究了营销数据图谱应用于常州供电公司的真实数据，能够实现跨系统网架结构的差异分析，找出数据孤立节点，准确定位数据问题原因，提升网架拓扑异常数据治理效率。

关键词

网架，电网拓扑，标准管理，自动识别，营销系统

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

目前营销与生产之间未建立起联动的数据变更维护机制，随着电网用户数的增多和电网变动的频繁，数据维护的准确性无法或很难与现场保持一致，甚至一段时间后，业务系统中的数据与现场相差很大，给故障抢修、业扩现场勘查造成困难。为加快推进 95598 全网业务集中，强化营销、运检、调控专业协调、有效支撑电力故障定位、停电范围定位、95598 配电抢修指挥、供电可靠性管理、业扩报装供电方案分析、线损管理、配网规划决策、客户营配综合信息查询、准实时数据展现等业务功能，支撑客户故障智能化管理、智能化业扩报装，为营配业务信息化管理提供可视化、一体化、智能化的信息支撑，不断提高供电服务品质，有必要开展营配贯通知识图谱展示项目。

知识图谱是结构化的语义知识库，用于以符号形式描述物理世界中的概念及其相关关系[1]。知识图谱常用的表示方式是“实体 - 关系 - 实体”三元组模型，即 $G = (E, R, S)$ ，其中 E 表示知识库中实体的集合，R 表示知识库实体间关系的集合，S 表示知识库的三元组集合。三元组的形式主要包括实体 - 关系 - 实体和实体 - 属性 - 属性值两种，实体是知识图谱的基本元素，由唯一 ID 值表示，并通过属性 - 属性值形式记录实体的特征信息，不同实体之间的关联信息由关系记录[2]。知识图谱模型的研究近年来一直得到广泛关注，多数成功模型[3] [4] [5]都有一个共性：他们使用低维的嵌入式向量来表达实体和关系，不同的地方是采用何种算法预测某一三元组的可能性，比较经典的有基于 web 的方法[6]、张量神经网络算法[7]和跳跃路径方法[8]等。

目前，尽管知识图谱技术有着广泛的应用，知识图谱在电网营销方面的相关应用没有得到普及。本文从数据图谱技术出发，构建了营销数据图谱，重点研究了营销数据图谱应用，能够实现跨系统网架结构的差异分析，准确定位数据问题原因，提升网架拓扑异常数据治理效率。

2. 知识图谱关键技术

2.1. 知识抽取技术

电网资源数据来源广泛、信息丰富，但是目前全业务数据中心大数据价值无法利用、业务数据无法跨专业贯通、知识不能有效表示，研究基于生产、营销数据的知识抽取技术，实现电网两大业务系统知识抽取工作。研究多元数据抽取方法，需要分析不同类型数据抽取技术，针对生产数据，构建大量正则表达式实现知识抽取；针对营销数据，采用 POS-CBOW 关联计算模型实现基于营销业务系统的电网资源知识关系抽取。CBOW 语言模型的基本训练步骤如下：

步骤 1：在输入层，通过 M 限定输入层的上下文大小，然后在窗口中顺序读取语料库词序列 $C(t-M), C(t-M+1), \dots, C(t+M)$ ，通过哈希表得到投影层的相应词位置 $W(t-M), W(t-M+1), \dots, W(t+M)$ ，获得 $W(t)$ 词的上下 M 个词 $Context(W(t))$ 。

步骤 2：在投影层，对 $W(t)$ 的上下文 $Context(W(t))$ 做步骤 1 操作， $V(t)$ 为 $W(t)$ 上下文累加和。

$$V(t) = \sum_{t-N}^{t+N} Context(W(t)) \quad (1)$$

步骤 3：从投影层到输出层，利用词 $W(t)$ 的上下文信息，通过式(2)来生成词 $W(t)$ 的向量值，其中式(3)为词向量回归分析操作，来完成对 $W(t)$ 的判断。

$$P(W(t)|Context(W(t))) = \prod_{t-n}^{t+n} f(V(t)\theta) \quad (2)$$

$$f(V(t)\theta) = \frac{1}{1 - e^{-V(t)\theta}} \quad (3)$$

2.2. 知识图谱建模技术

基于营配贯通知识图谱建模技术主要分为两层，上层为知识图谱分析方法与文献计量方法、德尔菲法等分析方法。在知识图谱的具体应用中，分析者可进行探索式研究，在图谱绘制流程中的多个环节，通过对文献计量学网络、标准化方法、可视化映射技术等不同选择来开展多源跨图谱分析，以保证研究结果的可靠性与全面性，此外，还可将知识图谱分析方法与其他研究方法协同使用，起到互补验证的效果。下层为数据资源建模的一般流程，体现了“事实型数据 + 专用方法工具 + 专家智慧”的通用模式，事实型数据库是分析的基础，研究机构要构建多源融合的信息资源库来支持后续的分析。模型中融入了“以目标为中心”的思想，将数据分析人员、信息搜集人员、学科专家和用户等利益相关者均纳入流程之中，其中专家智慧可应用在从选题到形成报告的各个环节，用以保证模型构建的最终质量。

2.3. 知识图谱实体关联技术

基于营配贯通知识图谱实体关联技术需要实现以下几个操作：1) 数据存储，实体关联关系数据存储于图数据库中，我们称其为关联库；2) 根据用户的需求从关联库中读取数据；3) 可视化元素生成：节点和边生成相对应的可视化图形元素，采用节点控件和边控件来分别表达；4) 再采用相应的可视化布局算法，将可视化图形元素进行排版布局画在屏幕上；5) 交互操作，用户可以对可视化的节点控件和边控件进行操作，操作的结果可以通过关联库提供的接口存储到关联库中。其中，生产、营销关联库是实体关联技术的保证，也是该技术的关键点和难点。

2.4. 知识图谱智能分析技术

由于缺乏数据共享以及有效的跨业务数据的关联分析手段，在进行一些复杂业务处理时，业务人员往

往需要进行跨系统的数据查询，并且进行人工数据分析，造成大量的人力和物力开销。通过构建营配贯通知识图谱，采用基于翻译模型的知识表示学习技术对营配贯通知识图谱中的实体知识和关系进行分布式表示，从而达到生产、营销数据信息在表示上的统一，实现基于分布式表示的实体关系预测和智能分析。

3. 营销数据图谱构建

3.1. 营销电网拓扑数据关系梳理

根据营销系统中拓扑数据图谱构建相关研究元素，拟定以营销系统电网拓扑相关数据表作为收集对象，次序包括：变电站→线路→配变→用户→计量点→电能表。

依据营销电网拓扑模型梳理成果，结合营销业务应用系统、用电信息采集系统核心业务系统的物理模型，理清各物理模型的关联关系，梳理出关联的字段。

结合营销电网拓扑相关信息，梳理数据表之间的关系，如图 1 所示：

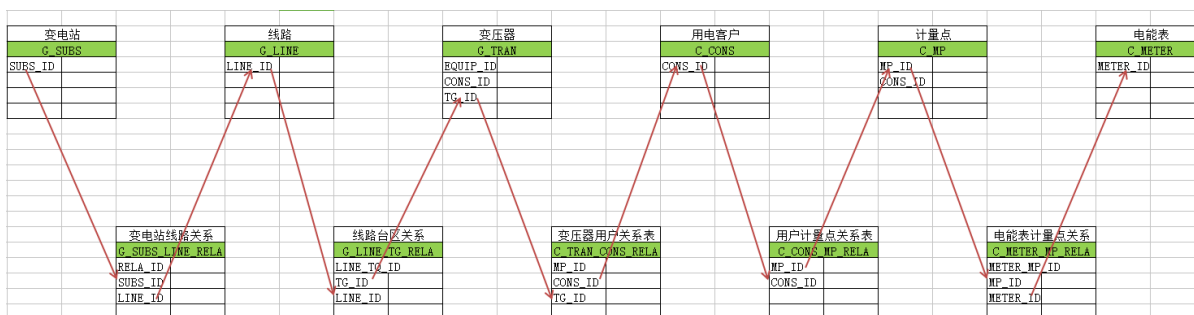


Figure 1. Association diagram of grid topology data in marketing department

图 1. 营销电网拓扑数据关系梳理

3.2. 营销电网拓扑知识图谱关系模型建立

数据图谱构建阶段分为三个阶段，分别为研究营销系统电网拓扑标准展示、研究图数据库技术在数据图谱方面的应用以及研究将站线变户电表从关系型数据模型转为图数据拓扑模型。

营销系统中分别存储了站、线、变、户的网架结构遵从国网标准(SG_CIM)，并结合网省实际业务与数据模型，构建面向企业的、可自由扩展的统一网架模型标准。在标准网架模型的基础上，完成以下工作内容：

1) 标准网架模型实例各系统相关物理表梳理

标准网架模型是营销统一标准，梳理出标准网架的对应档案物理表，从而将各系统各自的网架转化为标准网架。

2) 标准网架模型实例属性与各系统档案字段对应梳理

在网架物理表梳理的基础上，梳理出标准网架模型实例属性与各系统网架物理表字段的对应匹配情况，从而将网架档案的字段转化为标准网架模型属性。

3) 标准网架模型属性扩充

结合网省公司实际业务需求，在标准网架模型中添加需要的属性，并根据此属性从系统档案中找到对应的字段，与标准网架模型添加的属性形成关系映射，支撑标准网架模型的属性数据来源，以满足不同的业务需求。

将“站线变户电表”转化为图数据中的实体节点，将“站线变户电表”之间的连接方式表现为线条关系，用这个通用的、富有表现力的结构来建模，支持各种应用场景，如图 2 所示：

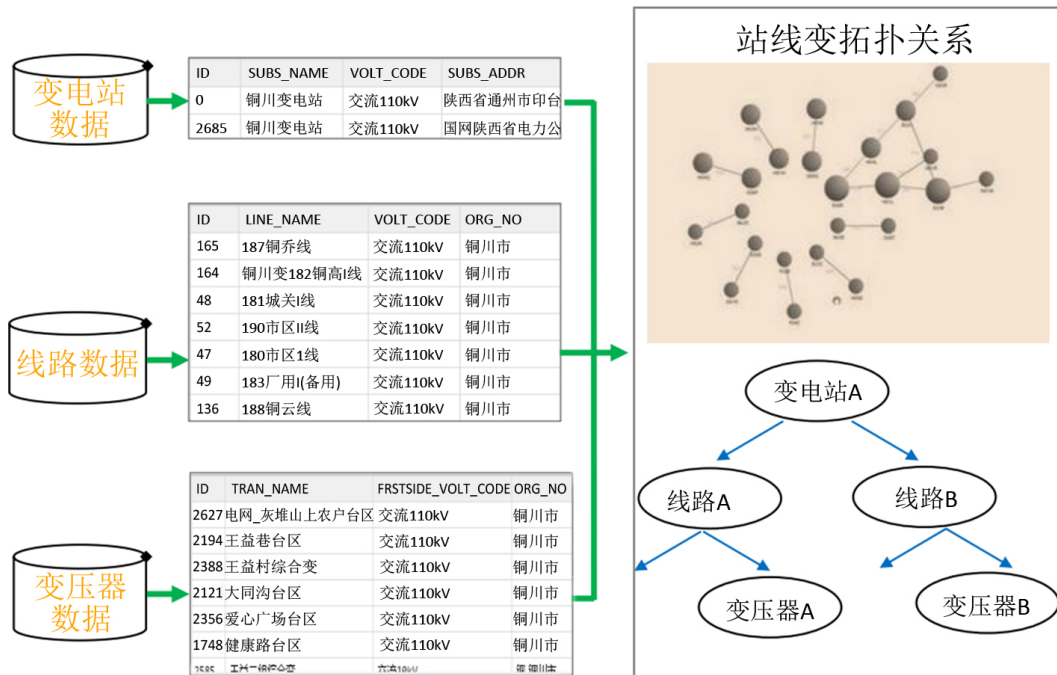


Figure 2. Association diagram of topology pattern model
图 2. 拓扑关系图谱模型图

4. 营销数据图谱应用

论文分别从图谱自动识别、拓扑节点搜索、台账信息查看、孤立节点导出、孤立点数据分类研究出发，重点针对图数据库技术在数据图谱方面的应用作了研究。

4.1. 图谱自动识别

营销数据图谱，根据各电网拓扑间的数据关系，可以自动识别并建立电网营销拓扑，实现变电站-线路-变压器-用户图谱的自动识别建立。如图 3 所示：



Figure 3. Diagram of grid topology data in marketing department
图 3. 电网营销拓扑

4.2. 拓扑节点搜索

营销数据图谱提供了节点搜索功能。搜索相应节点，将会展示该节点及与该节点相连的其他节点，同时线路可以按照 220 kV、110 V 和 10 kV 条件查询不同等级线路。如图 4 所示：

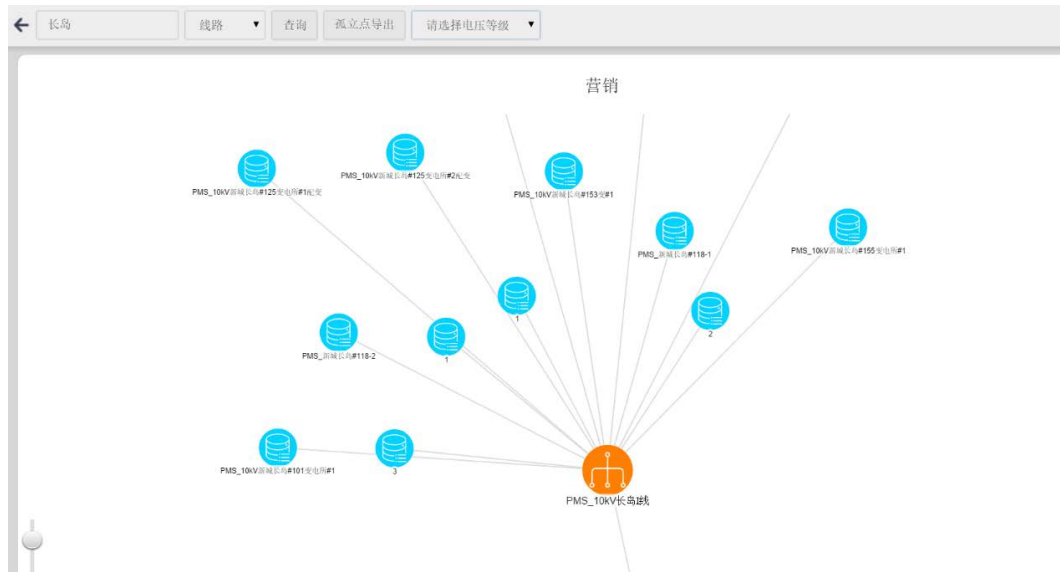


Figure 4. Diagram of node search topology
图 4. 节点搜索拓扑

4.3. 台账信息查看

图谱提供了查看每个节点的台账信息功能。右击图谱中任一节点，将会显示该节点的台账信息。如图 5 所示：

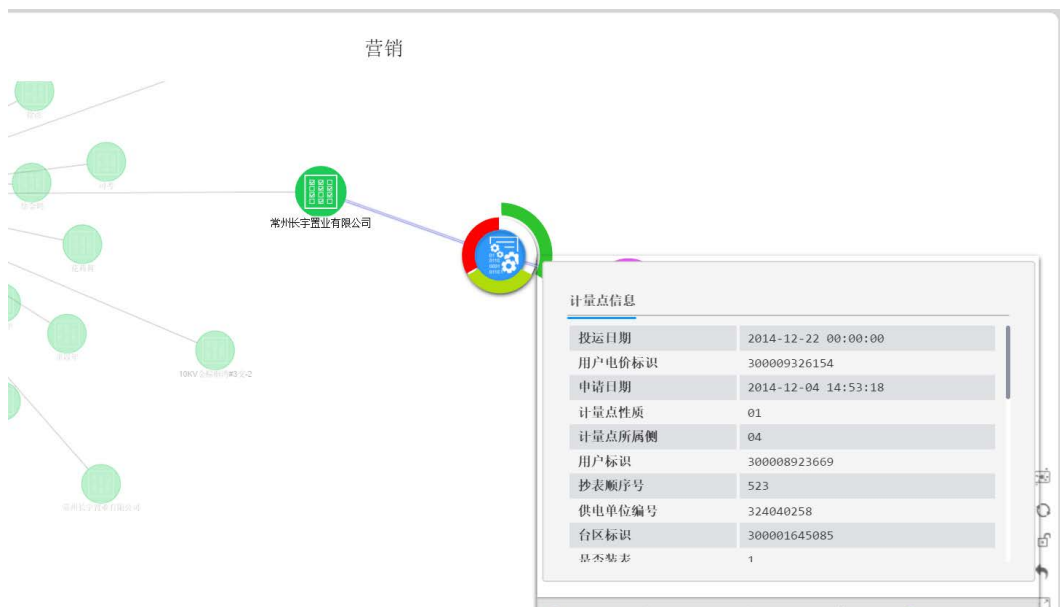


Figure 5. Diagram of information in history log
图 5. 节点台账信息

4.4. 孤立节点导出

在营销系统中，由于各节点数据间存在数据联动更新不及时，数据维护不到位的情况，导致变电站-线路-变压器-用户等节点中，存在无上下节点关联的情况，形成“孤立”节点。营销数据图谱的研究为孤立节点的识别提供的方法。点击图谱左上角的孤立点导出，可以导出连接关系孤立的节点，以表格形式导出这有助于发现目前数据中的未关联数据和异常挂接设备。如图 6 所示：



Figure 6. Diagram of isolated node recognition
图 6. 孤立节点识别

4.5. 孤立点数据分类研究

通过常州营销系统中的“站线变户” 全量数据在营销图谱工具中的处理，按照变电站、线路、变压器、用电客户四类分别进行了孤立节点的导出，对导出数据进行分类研究。

交叉比对分类明细如表 1、表 2 所示：

Table 1. Classification map of comparison results NO.1
表 1. 比对结果分类一

变电站(未关联线路的)		线路(未关联变电站且未关联变压器的)		线路(未关联变电站的)		线路(未关联变压器的)	
开闭所	18	瓷瓶、绝缘子、屏柜	38	瓷瓶、绝缘子、屏柜	38	开关	7276
110 kv	9	其他	3	其他	3	断路器	322
220 kv	3	/	/	/	/	电容器	794
500 kv	1	/	/	/	/	绝缘子、瓷瓶、屏柜、消弧线圈	144
配电所	84	/	/	/	/	110 kV	194

Continued

其他	9	/	/	/	/	220 kV	133
/	/	/	/	/	/	500 kV	15
/	/	/	/	/	/	其他	1305
合计(个)	124	合计(个)	41	合计(个)	41	合计(个)	10,183

Table 2. Classification map of comparison results NO.2

表 2. 比对结果分类二

变压器(未关联线路且未关联用户的)		变压器(未关联线路的)		变压器(未关联用户的)		用户(未关联变压器的)	
10 kV	10	10 kV	31	10 kV	10	220 kv	32
20 kV	12	20 kV	15	20 kV	12	其他	35
其他	28	其他	392	其他	73	/	/
合计(个)	50	合计(个)	438	合计(个)	95	合计(个)	67

通过变电站、线路、变压器、用电客户四类数据交叉比对梳理,共发现 11,039 个数据孤立点,属于电网网架拓扑异常数据,且现场实际数据不相符,将该部分数据交给业务部门,进行营销数据电网拓扑方面的数据维护,便于提升电网设备数据质量,确保电网设备关联数据完整性、准确性和及时性。

5. 结论

本文基于知识图谱的营配贯通数据图谱作了深入研究,可实现跨系统网架结构的差异分析,准确定位数据问题原因,提升电网网架拓扑异常数据治理效率。在电网网架设备发生变化后,系统能够实现对档案信息和网架拓扑关系自动识别,辅助用户对网架设备的及时跟踪管理,避免造成信息化管理与实际不符。未来可在故障定位、负荷预测等业务方面加强推广应用,从而提升业务部门数据管理质量,提高电力生产效益。

参考文献

- [1] 刘峤, 李杨, 段宏, 刘瑶, 秦志光. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 13-16.
- [2] 廖胜姣, 肖仙桃. 科学知识图谱应用研究概述[J]. 情报理论与实践[J]. 2009, 32(1): 122-125.
- [3] Lin, Y., Liu, Z., Sun, M., Liu, Y. and Zhu, X. (2015) Learning Entity and Relation Embeddings for Knowledge Graph Completion. *29th AAAI Conference on Artificial Intelligence*, Austin, TX, 25-30 January 2015, 2181-2187.
- [4] Neelakantan, A. (2015) Compositional Vector Space Models for Knowledge Base Inference. *Computer Science*, 1-16.
- [5] Wang, Z., Zhang, J., Feng, J. and Zheng, C. (2014) Knowledge Graph Embedding by Translating on Hyperplanes. *28th AAAI Conference on Artificial Intelligence*, Québec, 27-31 July 2014, 1112-1119.
- [6] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmman, T., Sun, S. and Zhang, W. (2014) Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 24-27 August 2014, 601-610. <https://doi.org/10.1145/2623330.2623623>
- [7] Socher, R., Chen, D., Ng, A. and Manning, C. (2013) Reasoning with Neural Tensor Networks for Knowledge Base Completion. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Lake Tahoe, NV, 5-10 December 2013, 926-934.
- [8] Lin, Y., Liu, Z., Luan, H., Sun, M., Rao, S. and Liu, S. (2015) Modeling Relation Paths for Representation Learning of Knowledge Bases. *Computer Science*, arXiv preprint arXiv: 1506.00379. <https://doi.org/10.18653/v1/D15-1082>

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2328-0514, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: aepe@hanspub.org