

Road Traffic Operation Analysis and Data Mining Based on Loop Detector Big Data

Liang Wang¹, Xiaofeng Liu², Shaotang Liu¹, Hai Zhang¹, Guanghui Yan²

¹Tianjin Bureau of Traffic Management, Tianjin

²School of Automotive and Transportation, Tianjin University of Technology and Education, Tianjin

Email: 13920829956@139.com

Received: Oct. 20th, 2017; accepted: Nov. 2nd, 2017; published: Nov. 9th, 2017

Abstract

Loop detector big data was focused and its characteristics were analyzed. Then, the reasons of big data lost and abnormality were illustrated, and the data repairment method was given. Next, taking loop detector big data of Weidi Road in Tianjin City as an example, the traffic volume situation, the traffic spatial-temporal distribution and traffic congestion were analyzed. Finally, based on traffic volume data, a data mining method using traffic simulation was proposed to obtain traffic speed, and the case result showed that both the traffic congestion period and region can be acquired.

Keywords

Traffic Engineering, Traffic Control, Data Mining, Traffic Big Data, Loop Detector

基于线圈大数据的道路交通运行分析与挖掘

王亮¹, 刘晓锋², 刘少堂¹, 张海¹, 闫光辉²

¹天津市公安交通管理局, 天津

²天津职业技术师范大学, 汽车与交通学院, 天津

Email: 13920829956@139.com

收稿日期: 2017年10月20日; 录用日期: 2017年11月2日; 发布日期: 2017年11月9日

摘要

以检测线圈大数据为研究对象, 对其数据特征进行了分析, 归纳了检测线圈大数据丢失和异常的原因, 提出了数据修复的处理方法。以天津市围堤道检测线圈大数据为例, 分析了交通量、交通时空分布和交通拥堵特征。以检测线圈的流量数据为基础, 提出了基于交通仿真的车速挖掘方法, 并进行了案例分析,

案例结果表明该方法可获取交通拥堵时段和拥堵范围等信息。

关键词

交通工程, 交通控制, 数据挖掘, 交通大数据, 检测线圈

Copyright © 2017 by authors 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

近十年来,我国经历了快速的机动化进程,城市的交通拥堵现象日益普遍,与此同时,智能交通系统的建设与发展也正在大力推进。目前,我国城市道路交通信息的检测手段日益多样化,交通数据的来源呈现多元的特点,既有定点型交通数据(如检测线圈、地磁、视频、卡口等),也有移动型交通数据(如手机、浮动车、电子标签、公交卡、无人飞机等) [1] [2] [3],这些数据的综合使用,对提升交通监控和交通管理水平具有重要意义。其中,检测线圈作为一种基础性的交通流检测设备,在城市交通中,尤其是交通控制系统中得到广泛的应用,是 SCATS 和 SCOOT 自适应信号控制系统的主要数据来源。由于检测线圈可以连续工作,其采集与存储的交通数据不仅覆盖的地理范围广、体量大,同时记录了交通流的时空变化情况,当前,我国正在开展交通大数据的应用研究,在此背景下,如何有效的进行检测线圈的大数据分析与挖掘,对提高城市交通控制与管理水平具有积极的意义。

在检测线圈数据的分析与挖掘方面,弓晋丽等[4]以定点检测线圈的数据为基础(流量、占有率、速度),拟合交通参数关系,得到交通拥挤阈值,然后对交通瓶颈位置进行识别,分析交通瓶颈处的交通拥堵演化情况。皮晓亮等[5]运用环形线圈检测的交通流数据(流量、占有率、速度),使用聚类分析方法进行交通状态分类的挖掘研究。王涛等[6]采用时空滑动窗口模型、聚类分析对交通流数据进行按时段分簇,接着运用主成分分析法剔除非关键变量,最后建立分时段的、多元线性回归模型拟合交流。此外,基于统计分析、神经网络、灰色系统、贝叶斯网络等方法的交通数据挖掘研究,也日益广泛[7] [8]。但是,上述研究主要局限于某一时段、某一路段,由于缺乏交通大数据样本,其研究结果具有一定的局限性。因此,将以实际的检测线圈交通大数据为基础,开展道路运行分析与挖掘研究,揭示城市交通运行的更多特征。

2. 检测线圈大数据特征分析

大数据概念由牛津大学 Viktor Mayer 教授最早提出,他认为大数据是指不用随机分析或者抽样调查,而是采用所有数据进行分析处理的研究方法。大数据具有 4V 特征,也就是 volume (大容量), variety (多样性), value (高价值)和 velocity (高速) [9]。检测线圈大数据同样具备上述的“4V”特征,首先,检测线圈是一天 24 小时,一年 365 天工作,数据容量大;其次,采集的信息类型较为丰富,有交通量、车速、占有率、车型等信息;另外,运用检测线圈大数据,可以开展交通状态分析、交通瓶颈辨识以及交通控制系统优化等工作,数据价值较高;最后,检测线圈数据采集以后,通过通信网络快速、实时的传输到交通控制管理中心进行存储。

在我国不同的城市,由于交通控制管理的需求各有侧重,以及软硬件平台的性能各有差异,检测线圈存储的交通信息形式差异较大,上海市公路网交通监控中心所存储的环形线圈采集的交通信息种类丰富,如表 1 所示,天津市交通监控中心所存储的环形线圈交通信息相对较少,如图 1 所示。

Table 1. Loop detector data format of Shanghai City [5]**表 1.** 上海市线圈数据构成[5]

NodeID	LoopID	DataID	TimeID	XK	DK	XH	ZH	DH	TG	MT	TLG	SpeedK	SpeedH	LoopFault	...
1	1	20040201	1	16	0	1	0	0	0	0	0	22.8	20.6	0	...
1	1	20040201	2	6	0	1	0	1	0	0	0	36.6	28.7	0	...
...

Tuesday, 01 July 2014

Approach 1, Detector: 1

	00:	01:	02:	03:	04:	05:	06:	07:	08:	09:	10:	11:
:05	6	4	7	1	4	2	1	16	23	17	19	25
:10	5	6	4	0	2	2	5	21	17	16	22	31
:15	10	1	3	2	2	6	9	22	22	18	14	19
:20	9	11	3	1	3	3	4	20	22	11	19	8
:25	4	3	5	2	2	2	5	17	16	16	17	18
:30	7	3	2	4	1	5	16	30	14	12	23	19
:35	5	2	4	0	3	2	12	15	19	19	11	32
:40	6	3	2	4	1	3	11	21	19	20	22	24
:45	8	2	4	0	1	4	13	20	20	24	19	27
:50	5	3	1	1	1	5	17	27	8	18	21	20
:55	11	4	0	0	0	6	15	22	23	22	21	26
:60	5	4	2	2	0	1	15	4	11	10	18	33
Hourly												
Total	81	46	37	17	20	41	123	235	214	203	226	282

AMTotal: 1525 AMpeak 282 11:00 - 12:00

Figure 1. Storage format of loop detect data of Tianjin City**图 1.** 天津市线圈数据存储形式

表 1 中的交通基础信息有交通量、车速、占有率、分车型(大、中、小)统计, 以及一些其他的信息, 如节点编号、检测器设备编号、记录日期、记录时间段编号、大货流量、大客车速、设备工作状态等。图 1 中的交通信息相对较少, 包括记录日期、车道、每小时交通量(未分车型)、每 5 分钟间隔交通量、高峰时段等。

3. 检测线圈大数据的处理

在检测线圈的检测数据中, 经常会出现异常数据(如大量的“0”值数据)和丢失数据(如数据缺失), 其原因主要有以下几个方面:

- 1) 线圈损坏: 线圈一般埋在车道中, 受到车辆的碾压、路面变形、路面维修的影响, 导致线圈不能正常工作或检测误差巨大。
- 2) 通信网络故障: 通信故障导致采集的数据无法上传。
- 3) 交通和动力因素: 在凌晨, 不少路段的车流量极低, 导致交通参数值趋 0; 另外, 检测设备断电, 导致不能正常检测工作。

3.1. 丢失数据的处理

如果丢失的数据规模不大, 可以采用该时刻的前一时间段或后一时间段的交通参数进行平滑处理,

用平滑值进行补充；如果由于线圈损坏，导致大规模的数据缺失，可调用线圈正常工作时的数据，并考虑月变、周变、日变系数，对数据值予以修正后填补。

3.2. 异常数据的处理

异常数据与其余数据差异很大，既与线圈是否正常工作有关，也与极端的交通运行状况有关，如凌晨时分的车辆超速行驶(车速值很大)，临时交通禁行管制(流量为 0)，黄金周的大面积交通拥堵(占有率很高)。因此，对异常数据要分类处理，对于线圈未能正常工作所检测的数据，应予以剔除，按数据丢失处理；对于其他原因所检测的异常数据，应该予以保留。

4. 道路交通运行分析

以天津市河西区围堤道与贵州路、马场道、友谊路三个交叉口为例，进行线圈大数据的交通运行分析。其中，围堤道是河西区东西向重要的交通主干道，贵州路、马场道是河西区进出和平区的重要通道，友谊路则是天津的金融街，研究范围如图 2 所示，交叉口的 SCATS 系统控制形式如图 3 所示。研究区域的交通大数据来源于天津市公安交通管理局科研所的 SCATS 交通控制系统，交通数据的时间跨度为 2014 年 7 月 1 日到 2015 年 6 月 30 日，交通数据采样周期为 5 分钟，数据类型只有流量数据。

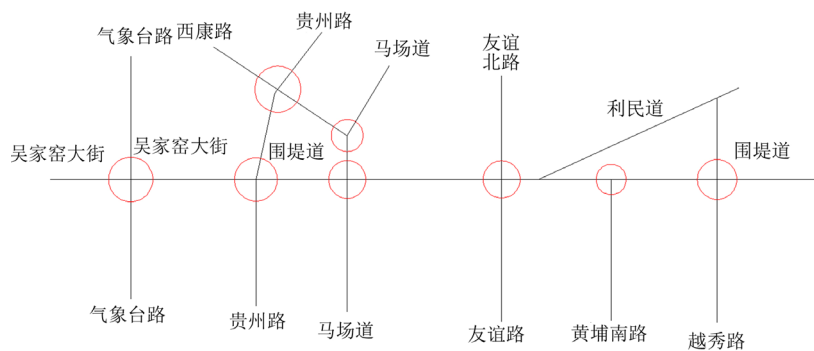


Figure 2. Research area
图 2. 研究区域

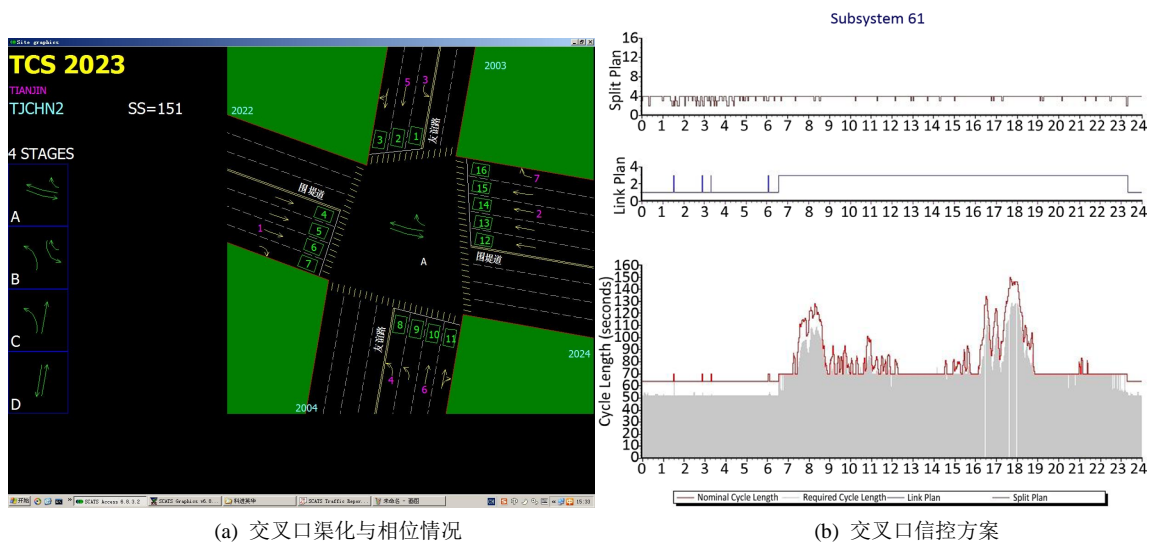


Figure 3. Scats control situation of intersection
图 3. 交叉口 SCATS 控制形式

4.1. 交通量情况

运用 Excel、Matlab 等软件工具进行交通数据的处理工作,选取的样本时间为 2015 年 5 月 4 日到 2015 年 6 月 5 日,合计 5 周,分别分析了同周内工作日平均相似系数和波动系数、不同周同一工作日相似性系数分析和波动系数,分析结果如表 2、表 3 所示。

由表 2 可知,月初(5 月 4 日到 5 月 8 日、6 月 1 日到 6 月 5 日)同周内工作日平均相似系数较大,分别为 0.8577、0.8630;月中(5 月 11 日到 5 月 15 日)、月末(5 月 25 日到 5 月 29 日)同周内工作日平均相似系数较小,分别为 0.7446、0.7299。由表 3 可知,不同周同一工作日(周二)的相似性系数最大,达到 0.8500;不同周同一工作日(周一)的相似性系数最小,其值为 0.6871;相似性系数越大,波动系数(标准差系数)越小,相似性系数越小,波动系数(标准差系数)越大。

研究区域 3 个交叉口的不同时期的交通流量情况如表 4 所示,由表可知这 3 个交叉口的交通总流量较大,最小值为 6.87 万辆,最大值为 8.40 万辆;另外,这 3 个交叉口的一整天交通总流量,在 2014 年 12 月 10 日、2015 年 3 月 11 日、2015 年 6 月 10 日 3 个时间段,呈现小比例下降趋势,这可能与天津市交通限行政策的持续实施、落实到位有一定的关系。

Table 2. Average similarity coefficient and variation coefficient of weekday in one week (2015)

表 2. 同周内工作日平均相似系数和波动系数(2015 年)

日期	平均相似系数	波动系数(标准差系数)
5 月 4 日到 5 月 8 日	0.8577	0.1341
5 月 11 日到 5 月 15 日	0.7446	0.1607
5 月 18 日到 5 月 22 日	0.8483	0.1313
5 月 25 日到 5 月 29 日	0.7299	0.1619
6 月 1 日到 6 月 5 日	0.8630	0.0946

Table 3. Average similarity coefficient and variation coefficient of the same day in different weeks (2015)

表 3. 不同周同一工作日相似性系数分析和波动系数(2015 年)

	平均相似系数	波动系数(标准差系数)
星期一	0.6871	0.2355
星期二	0.8500	0.0890
星期三	0.7708	0.1233
星期四	0.7218	0.2316
星期五	0.7979	0.1482

Table 4. Intersection traffic volume of different days (Unit: pcu/day)

表 4. 不同日期的交叉口交通流量(单位: PCU/天)

交叉口	2014 年 12 月 10 日	2015 年 3 月 11 日	2015 年 6 月 10 日
围堤道 - 贵州路	86991	87607	73759
围堤道 - 马场道	83986	74324	71907
围堤道 - 友谊路	72123	70920	68716

4.2. 交通时空分别特征

分别从时间、空间两个维度，进行交通分布特征分析，选择的交通样本为 2015 年 6 月 10 日。其中，交通时间分布特征方面，分析了交叉口同方向直行、左转车道的交通分布情况，交通空间分布特征方面，分析了围堤道断面流量以及分方向不均匀系数。

对交叉口 24 个小时的交通流量分布分析可知：同一交叉口同方向的直行车道的交通时间分布特征相似，同一交叉口同方向的左转车道的交通时间分布特征相似，这也就是说同一交叉口同方向相同功能的车道具有类似的交通时间分布特性。与之对应的却是，同一交叉口同方向不同功能的车道的交通时间分布特性差异较大。

对交叉口进行流量分析可知：围堤道不同路段的断面流量差异较大，最大断面流量为 51,769 pcu/天，断面为气象台路 - 贵州路路段；最小断面流量为 42,140 pcu/天，断面为马场道 - 友谊路路段。另外，不同路段的方向不均匀系数差别较大，如表 5 所示。

4.3. 交通拥堵特征

根据交通大数据、交叉口绿信比和道路通行能力，确定交叉口的饱和度分布情况，由于天津市交通混行现象较为突出，因此，在分析交叉口交通拥堵特征时，还充分考虑现场查勘的结果和实际驾车通行的交通拥堵主观感受，对交叉口的不同车道的交通拥堵特征进行甄别。其中，围堤道 - 贵州路交叉口的分析情况如图 4 所示，红色线段代表交通拥堵的持续时间。由分析可知，围堤道 - 贵州路西进口交通拥堵最为严重，交通拥堵从早 7 点一直持续到晚 7 点。另外，围堤道 - 贵州路东进口的交通拥堵呈现明显的波动特征，即在早 8 点、11 点、15~19 点出现 3 波交通拥堵。

Table 5. Non-uniformity coefficients of different directions of Weidi Road (Unit: pcu/day)

表 5. 围堤道各断面方向不均匀系数情况(单位: pcu/天)

	气象台路 - 贵州路	贵州路 - 马场道	马场道 - 友谊路
由东向西流量	15149	25763	23710
由西向东流量	36620	20324	18430
合计断面流量	51769	46087	42140
方向不均匀系数	0.7074	0.5590	0.5626

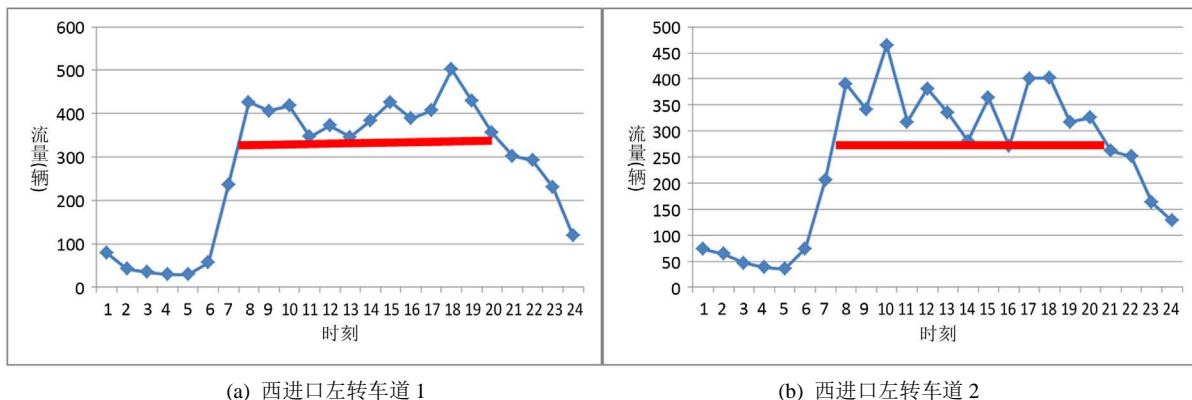


Figure 4. Congestion situation of Weidi Dao-Guizhou Road intersection

图 4. 围堤道 - 贵州路交叉口拥堵情况

5. 道路交通运行挖掘

道路交通信息的数据挖掘是对数据库中的交通信息进行知识发现, 并进一步处理、提炼新的信息, 为交通管理和出现服务。在本研究中, 由于线圈大数据的存储类型单一, 只有流量数据, 缺乏车速数据, 这难以对交通运行演化情况进行分析。本研究运用交通仿真的方法, 建立区域路网, 结合已有的交叉口渠化情况、车流分布情况以及信号控制情况, 在仿真路段中按一定间隔布设交通检测器, 车辆仿真运行以后, 软件后台记录路段定点车速, 因此, 可从交通流量中挖掘出车速分布情况。基于交通仿真的车速挖掘方法如图 5 所示。

以围堤道 - 马场道交叉口东进口路段为例, 从进口道开始, 每 100 米布设 1 个检测器, 一共布设 6 个, 对该路段的车速演化情况进行交通仿真分析, 如图 6 所示。通过 VISSIM 软件记录下路段不同交通检测器位置的车辆运行速度, 在 Matlab 中绘制基于车速的交通拥堵时空分布图, 如图 7 所示, 图中横坐标为时刻, 左侧纵坐标为离进口道停车线的距离, 右侧纵坐标为车速的分布情况。由图可知, 在 7:30 分到 8:30

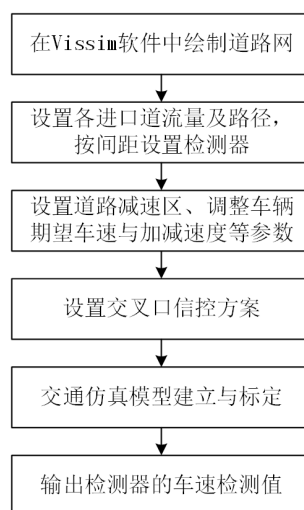


Figure 5. Flow chart of speed mining method based on traffic simulation

图 5. 基于交通仿真的车速挖掘方法流程图



Figure 6. Intersection traffic simulation

图 6. 交叉口运行仿真情况

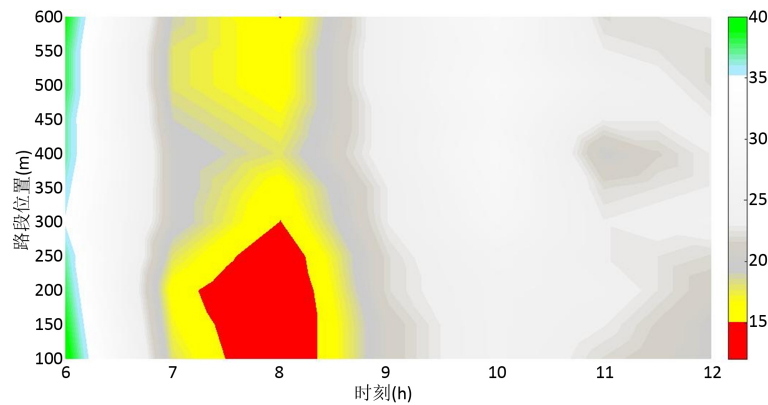


Figure 7. Speed distribution of intersection in-lanes
图 7. 交叉口进口道车速分布图

分交叉口进口道交通拥堵严重, 平均车速低于 15 km/h, 平均排队长度达到 200 余米。

6. 结论

对检测线圈大数据特征进行了分析, 总结了检测线圈数据丢失和异常的原因, 结合天津市检测线圈交通大数据进行了道路交通运行分析和挖掘, 提出了在数据有限的情况下, 基于交通仿真的车速分布挖掘方法。在下一步的工作中, 可加强多源交通信息的融合与挖掘研究, 以使交通大数据更真实的反映交通的实际运行情况, 从而为掌握交通拥堵演变提供数据支持, 为交通控制与管理决策提供技术支持。

基金项目

天津市高校科技发展基金项目(20140917); 天津市科技计划项目(智能交通与无人飞机知识科普展示)。

参考文献 (References)

- [1] 钮心毅, 丁亮, 宋小冬. 基于手机数据识别上海中心城的城市空间结构[J]. 城市规划学刊, 2014(6): 61-67.
- [2] 陈君, 杨东援. 基于 APTS 数据的公交卡乘客通勤 OD 分布估计方法[J]. 交通运输系统工程与信息, 2013, 13(4): 47-53.
- [3] Liu, X.F., Guan, Z.W., Song, Y.Q., et al. (2014) An Optimization Model of UAV Route Planning for Road Segment Surveillance. *Journal of Central South University*, **21**, 2501-2510. <https://doi.org/10.1007/s11771-014-2205-z>
- [4] 弓晋丽, 杨东援. 基于检测线圈数据的道路交通瓶颈分析[J]. 公路交通科技, 2010, 27(7): 118-122.
- [5] 皮晓亮, 王正, 韩皓, 等. 基于环形线圈检测器采集信息的交通状态分类方法应用研究[J]. 公路交通科技, 2006, 23(4): 115-119.
- [6] 王涛, 王俊峰, 罗积玉, 等. 基于时空分析的复杂交通流数据挖掘算法[J]. 四川大学学报(工程科学报), 2011, 43(5): 153-158.
- [7] 张益, 陈淑燕, 王炜. 短时交通量时间序列智能复合预测方法综述[J]. 公路交通科技, 2006, 23(8): 139-142.
- [8] Yang, M.L., Liu, Y.G. and You, Z.S. (2010) The Reliability of Travel Time Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, **11**, 162-171. <https://doi.org/10.1109/TITS.2009.2037136>
- [9] 陆化普, 孙智源, 屈闻聪. 大数据及其在城市智能交通系统中的应用综述[J]. 交通运输系统工程与信息, 2015, 15(5): 45-52.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2326-3431，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，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：ojtt@hanspub.org