

Prediction and Analysis of Agricultural Products Logistics Demand in Beijing, Tianjin and Hebei Based on Big Data

Songmiao Li

Beijing Wuzi University, Beijing
Email: lisongmiao1101@163.com

Received: Jul. 27th, 2018; accepted: Aug. 9th, 2018; published: Aug. 16th, 2018

Abstract

The national strategy of Beijing-Tianjin-Hebei coordinated strategy brings opportunities for development in the logistics market. People are paying more and more attention to the safety and health of foodstuffs. At the same time, with the upgrade of people's consumption, the widespread concern about the quality and freshness of agricultural products poses a great challenge to the logistics industry. Based on the demand of agricultural products logistics, this paper establishes the index system of agricultural products logistics, designs the forecasting model based on big data and factor analysis and multiple linear regression, predicts and analyzes the logistics demand of agricultural products by using the model. It is found that the key factors influencing the logistics demand in Beijing, Tianjin and Hebei are the macro economy and freight volume, and put forward the suggestions that the Beijing-Tianjin-Hebei region should be developed according to the regional positioning, vigorously develop the logistics and communication infrastructure and provide high-quality logistics services.

Keywords

Beijing-Tianjin-Hebei Region, Agricultural Product Logistics, Big Data Forecast, SPSS

基于大数据的京津冀农产品物流需求预测分析

李松苗

北京物资学院, 北京
Email: lisongmiao1101@163.com

收稿日期: 2018年7月27日; 录用日期: 2018年8月9日; 发布日期: 2018年8月16日

摘要

京津冀协同战略等国家战略给物流市场带来发展机遇。人们对食品安全健康的关注与日俱增，同时随着人们消费升级，对农产品质量和新鲜度的广泛关注给物流业带来极大挑战。本文从农产品物流的需求出发，建立农产品物流指标体系、设计基于大数据和因子分析以及多元线性回归融合的预测模型，利用模型对农产品物流需求进行预测和结果分析，发现影响京津冀物流需求的关键因素是宏观经济和货运量，并提出京津冀地区发展应按照地区定位，并大力发展物流和通信基础设施，提供高质量的物流服务等建议。

关键词

京津冀，农产品物流，大数据预测，SPSS

Copyright © 2018 by author 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

随着京津冀协同战略等国家战略的实施，物流市场也迎来发展机遇。随着人们对食品安全健康的进一步关注，互联网 + 物流的不断融合，现有的物流产业机构和供应链已经逐渐暴露出诸多问题。京津冀农产品的发展已经不能满足消费者个性化和多样化的需求，也不符合国家降本增效的要求。因此，寻求更高效的京津冀农产品物流成为研究的必然方向。农产品物流在流通的过程中涉及的环节多，流量大，影响需求的因素多，数据繁杂，洪玉兰(2010) [1]等人针对我国农产品冷链物流流通率差，损耗大等特点，探讨了我国农产品冷链物流未来的发展方向。如何有效的提高数据的价值也成为重中之重，因此，如何在这种环境下分析影响农产品物流需求的关键因素，对其农产品物流需求进行合理的预测，对助力加快京津冀农产品物流的全面发展、保障食品安全有着重要的价值。大数据的快速发展提供了有利的研究环境。

本文研究基于大数据的京津冀农产品物流需求问题的解决方案，含三个子模块，即通过构建大数据画像，建立农产品物流指标体系、设计基于大数据和多种方法融合的模型、利用模型对农产品物流需求进行预测和结果分析并提供可行性建议。

2. 研究方法和数据选择

2.1. 研究方法

2.1.1. 大数据

农产品物流需求是个复杂的系统，包含多种影响因素，包括确定性的和不确定性的，因此想要对农产品物流的需求进行合理较准确的预测，需要找到各种影响因素。大数据有其天然的优势可以利用，大数据技术是从大量的、不完全的、有噪声的、模糊的、随机的数据中利用人工智能、机器学习、统计学等技术提取隐含在其中人们未知但又是潜在有用的信息和知识的工具，建立在大数据的分析结果上，可以进行目标的预测，发挥大数据在物流需求领域的价值。大数据画像是根据这些不规则的海量的数据来

完成对对象的详尽标签化描述, 可视化地展示出最主要的特征, 通过关键信息的抓取, 可以获取进行需求预测的指标。京津冀可以看作是一个整体, 通过对京津冀的大数据画像, 了解京津冀地区最主要的特征, 从而获取农产品物流的指标, 通过对大量数据进行分析, 不仅仅可以对影响农产品的关键因素进行分析, 还可以预测农产品物流需求的具体数量, 增加预测的全面性和准确性。

2.1.2. 因子分析方法

因子分析是主成分回归的扩展, 是一种相关分析技术。因子分析是希望能够降低变量的数目, 并在一群具有相关性的数据中, 转换为新的、彼此独立、不相关的新因子[2]。因子分析的实质是用几个潜在的但是不能观察的互不相关的随机变量, 研究很多变量之间的内部依存关系, 找到数据的基本结构去描述许多变量之间的相关关系[3]。因此, 从反映某些对象的多个变量中, 提取几个公因子, 利用这些公因子解释影响研究对象的主要因素, 同时又简化了数据结构。

2.1.3. 多元回归分析

多元线性回归模型可以分析众多因素对因变量的影响程度, 通过检验判定方程的显著性等, 并可以利用逐步回归思想进行模型的修正, 保证模型的科学性和实用性。刘宇等(2014) [4]文中对茶陵县 2001年至 2012 年的 GDP、区域消费品零售总额等指标进行分析, 构建多元线性回归模型对茶陵县全部客货周转量进行预测, 说明了线性回归模型的准确性、有效性。

综合对比分析, 确定本文选用多元线性回归方法构建模型, 通过因子分析调取主成分之后, 从 9 个指标中选取合适的因子个数作为自变量构建京津冀农产品物流需求的多元线性回归模型, 将对应的统计数据输入 SPSS 统计工具进行分析, 并根据分析结果对多元线性回归方程进行修正改进。

2.2. 数据选择

农产品物流需求量的衡量需要一个明确的指标来刻画, 对于指标选取, 由于没有统一的物流需求的衡量标准, 所以通常情况下文献对这一指标的衡量借助间接因素来度量, 研究集中在货运量、产品产量和物流总额上。邱慧等(2016) [5]指出选取货运周转量为特征因素, 人均生产总值为其相关因素, 利用灰色预测模型 GM(1, 2)模型, 对山西省未来 3 年的物流需求进行预测。李夏培(2017) [6]以北京市农产品物流需求为研究对象, 用农产品物流总额表示北京市农产品物流需求, 进行单因素分析, 预测了“十三五”时期北京市的农产品物流总额。

通过查询国家数据和相关行业大数据报告[7] [8] [9], 总结出京津冀地区的大数据画像, 如图 1 所示。可以看出, 京津冀地区的政策和经济水平发展有利于发展高水平物流的条件; 京津冀地区人口特征以及收入消费水平促进消费升级, 支撑网购的繁荣, 农产品物流的消费也会大幅提升; 物流水平的发展虽居于全国平均水平, 但还需要满足进一步的优化。

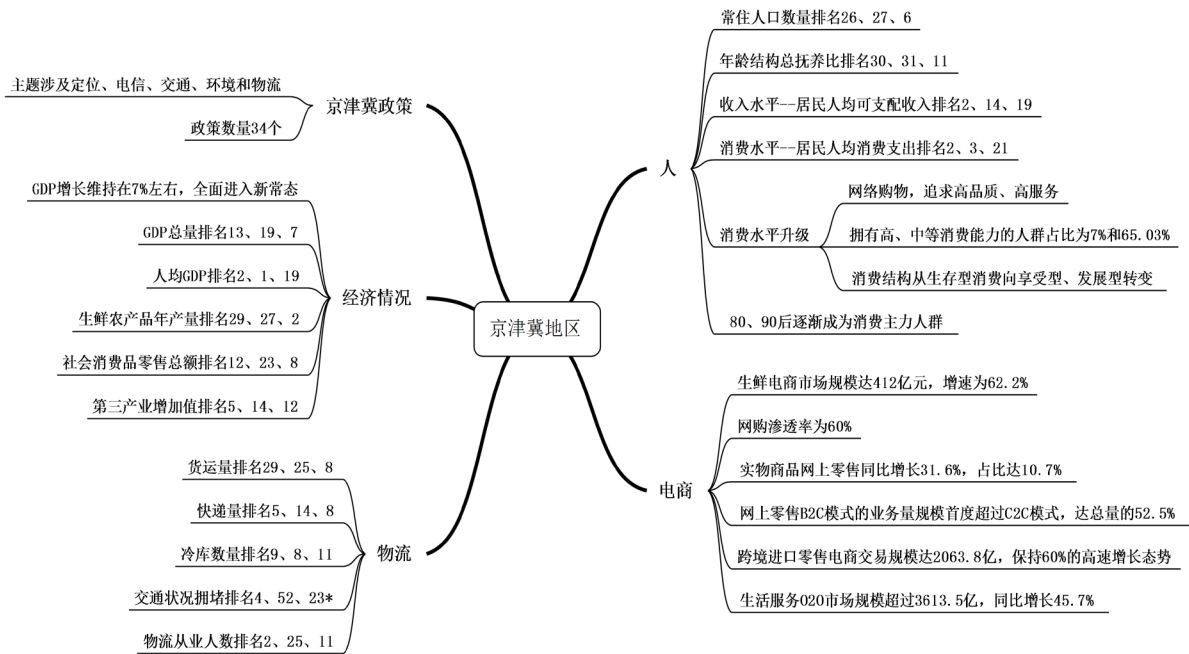
本文借鉴前人研究的成果, 根据相关学者的研究和文献资料的参考, 结合京津冀大数据画像所展示的物流发展现状和农产品物流需求影响因素的分析, 初步确定的物流需求影响因素如表 1。

3. 模型的建立和农产品物流需求预测

3.1. 模型的建立

因子分析是主成分分析的推广和发展, 把多个因子转化为少数几个综合因子的一种因子分析方法, 用观测变量来描述因子, 第 P 个因子在 i 个个案上的值可表示为:

$$F_{pi} = \sum_{j=1}^k W_{pj} X_{ji}$$



注：*表示河北省进入排名的有4个城市，此处选择最靠前的资料来源：国家统计局和相关行业报告

Figure 1. Big data portraits of the Beijing-Tianjin-Hebei region
图 1. 京津冀地区的大数据画像

Table 1. Agricultural products logistics demand influencing factor index system
表 1. 农产品物流需求影响因素指标体系

体系	指标种类	具体因素指标
农产品物流需求影响因素指标	产品供给指标	农产品的年产量(万吨) GDP(亿元)
	经济环境指标	产业结构增加值(亿元) 互联网上网人数(万人)
	人文环境指标	年末常住人口(万人) 居民消费水平(元) 社会消费品零售总额(亿元)
	物流规模	货运量(万吨) 货运周转量(亿吨/公里)

基于因子分析提取方法运行步骤如下：

1) 设观察到的随机向量为 $X = (x_1, x_2, \dots, x_p)$ ，不可观察向量为 $F = (F_1, F_2, \dots, F_p)$ ，则有：

$$\begin{cases} x_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ x_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ x_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p \end{cases}$$

其中 a_{ij} 表示因子荷载， ε_i 表示误差因子，代表公因子之外的影响因素。

2) 如果因子荷载矩阵不能对主因子进行解释时, 需要进一步做因子旋转, 使影响因子具有高负载, 使因子更加具有合理的解释。

3) 计算因子得分数学模型:

$$F_j = b_{i1}X_1 + b_{i2}X_2 + \dots + b_{in}X_n$$

得到 F_j , 可以很好的代表原始因子的信息, 将这些因子作为多元线性回归的自变量进行输入。

4) 利用 SPSS19.0 进行多元线性回归分析, 建立冷链物流需求预测的多元线性回归模型, 并对模型进行各种相关检验, 即:

$$Y = q_0 + q_1F_1 + q_2F_2 + \dots + q_pF_p + \varepsilon$$

其中 Y 为因变量, 其中 F_1, F_2, \dots, F_p 为自变量。 q_0, q_1, \dots, q_p 是 $p+1$ 个未知参数, q_0 称为回归常数, q_1, \dots, q_p 为回归系数, ε 是随机误差(残差)。

3.1.1. KMO 和 Bartlett 的检验判断是否适合因子分析

由表 2 可知, 球形度检验为 0.654, 接近 0.7, 非常适合做因子分析, 也进一步说明需要去除共线性。

3.1.2. 提取因子

表 3 为因子分析中的变量共同度。提取的 2 个特征值的情况下, 所有变量的信息丢失较少, 因子提取的效果比较好。

通过表 4 因子解释原有变量总方差的情况。可以看到, 第一个因子的特征值是 7.6, 解释原有 9 个变量总方差的 84.447%, 两个因子累计解释总变量的 96.104%。在此, 提取 2 个因子, 计算因子荷载矩阵。

Table 2. KMO and Bartlett tests

表 2. KMO 和 Bartlett 的检验

	取样足够度的 Kaiser-Meyer-Olkin 度量	0.654
	近似卡方	233.189
Bartlett 的球形度检验	df	36
	Sig.	0.000

Table 3. Communalities

表 3. 公因子方差

	初始	提取
x_1 地区生产总值(亿元)	1.000	0.976
x_2 第一产业增加值(亿元)	1.000	0.931
x_3 第三产业增加值(亿元)	1.000	0.968
x_4 居民消费水平(元)	1.000	0.960
x_5 社会消费品零售总额(亿元)	1.000	0.982
x_6 年末常住人口(万人)	1.000	0.986
x_7 货运量(万吨)	1.000	0.993
x_8 货物周转量(亿吨/公里)	1.000	0.870
x_9 互联网上网人数(万人)	1.000	0.984

提取方法: 主成份分析。

Table 4. Total variance of explained (part)
表 4. 解释的总方差(部分)

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累积 %	合计	方差的%	累积 %	合计	方差的%	累积 %
1	7.600	84.447	84.447	7.600	84.447	84.447	7.586	84.285	84.285
2	1.049	11.657	96.104	1.049	11.657	96.104	1.064	11.819	96.104
3	0.287	3.184	99.288						

提取方法：主成份分析。

表 5 显示了因子荷载矩阵，可见，各个变量中除了货运量，八个变量在第一因子的荷载很高，意味着它们与第一个因子的相关程度高，第一因子很重要，第二个因子与货运量相关性较强。因子荷载比较倾斜，考虑旋转后这两个因子的实际含义是否比较均衡。

3.1.3. 命名性解释

表 6 旋转成份矩阵为旋转后的因子荷载矩阵。采用极大方差法对因子荷载矩阵实行正交旋转以使得因子具有命名性解释。指定按第 1 个因子荷载降序的顺序输出旋转后的因子荷载，并绘制旋转后的因子荷载图，可见，第一个因子解释了除了货运量的其余变量，第二个因子解释了货运量，因子的含义比较清晰。表 7 成份得分协方差矩阵显示了两个因子的协方差矩阵。可以看到，两因子没有线性相关性，实现了因子分析的设计目标。

3.1.4. 计算因子得分

表 8 为因子得分系数矩阵。采用回归法估计因子得分系数，并输出因子得分系数，如表 8，根据表 8 用两个成分解释了原来 9 个变量，可得因子得分函数：

$$F_1 = x_1 * 0.130 + x_2 * 0.130 + x_3 * 0.128 + x_4 * 0.129 + x_5 * 0.131 \\ + x_6 * 0.129 + x_7 * 0.034 + x_8 * 0.129 + x_9 * 0.124$$

$$F_2 = -x_1 * 0.013 + x_2 * 0.077 - x_3 * 0.026 - x_4 * 0.010 + x_5 * 0.003 \\ + x_6 * 0.040 + x_7 * 0.946 + x_8 * 0.179 - x_9 * 0.125$$

3.1.5. 建立回归方程

根据因子得分表达式，计算 F_1 、 F_2 的因子得分，作为自变量，北京地区农产品物流需求量作为因变量，进行回归分析(如表 9)。

回归后的方程式 $Y = 727.789 - 0.017F_1$ ， R 方 = 0.902， $F = 83.741$ ，说明模型拟合很好，并得到了 2006~2015 年的预测值，预测值如表 10。

3.2. 农产品物流需求预测

为了预测 2016~2020 年北京地区物流需求量，需要对相关经济变量的未来变化趋势进行分析，运用 SPSS 指数平滑法对变量进行预测，预测结果如表 11。

把自变量的预测值带入因子回归模型，北京市的因子回归模型如下：

$$y = 727.789 - 0.017F_1$$

预测 2016~2020 年北京市农产品的物流需求量，预测结果如表 12。

同理可得，京津冀地区的因子回归模型如下：

Table 5. Component matrix
表 5. 成份矩阵 a

	成份	
	1	2
x_1	0.988	-0.008
x_2	0.961	0.087
x_3	0.984	-0.021
x_4	0.980	-0.004
x_5	0.991	0.010
x_6	0.992	-0.035
x_7	-0.080	0.993
x_8	0.912	0.194
x_9	0.984	-0.125

提取方法：主成份。a. 已提取了 2 个成份。

Table 6. Rotated component matrix
表 6. 旋转成份矩阵

	成份	
	1	2
x_1	0.987	-0.054
x_2	0.964	0.042
x_3	0.981	-0.067
x_4	0.978	-0.050
x_5	0.991	-0.037
x_6	0.989	-0.082
x_7	-0.033	0.996
x_8	0.921	0.150
x_9	0.977	-0.172

提取方法：主成份。旋转法：具有 Kaiser 标准化的正交旋转法。a. 旋转在 3 次迭代后收敛。

Table 7. Component score covariance matrix
表 7. 成份得分协方差矩阵

成份	1	2
1	1.000	0.000
2	0.000	1.000

提取方法：主成份。旋转法：具有 Kaiser 标准化的正交旋转法。

$$\text{天津市: } y = 545.361 + 0.035F_1 + 0.006F_2$$

$$\text{河北省: } y = 7357.864 + 0.28F_1 + 0.349F_2$$

因此，京津冀地区农产品物流需求量预测值如表 13。

Table 8. Component score coefficient matrix
表 8. 成份得分系数矩阵

	成份	
	1	2
x_1	0.130	-0.013
x_2	0.130	0.077
x_3	0.128	-0.026
x_4	0.129	-0.010
x_5	0.131	0.003
x_6	0.129	-0.040
x_7	0.034	0.946
x_8	0.129	0.179
x_9	0.124	-0.125

提取方法：主成份。旋转法：具有 Kaiser 标准化的正交旋转法。

Table 9. Regression equation variable of Beijing
表 9. 北京地区回归方程变量

农产品物流需求量	F1	F2
582.32	5843.6649	30,814.49686
596.67	6192.1337	18,291.84039
572.92	6880.57849	18,818.38096
573.04	7473.83893	18,691.45931
549.98	8518.25899	19,832.77071
547.41	9639.03864	22,478.83302
523.36	10,539.27431	23,808.02979
487.83	11,499.00674	23,315.72625
457.86	12,347.47831	23,988.59449
412.86	13,050.88561	17,736.42065

Table 10. Forecast of agricultural products logistics demand in Beijing from 2006 to 2015
表 10. 北京地区 2006~2015 年农产品物流需求预测值单位：万吨

时间	观测值	预测值	误差
2006 年	582.32	604.5552	0.0382
2007 年	596.67	595.6914	-0.0016
2008 年	572.92	581.0787	0.0142
2009 年	573.04	567.9921	-0.0088
2010 年	549.98	545.4368	-0.0083
2011 年	547.41	521.4594	-0.0474
2012 年	523.36	501.6583	-0.0415
2013 年	487.83	480.4164	-0.0152
2014 年	457.86	461.5081	0.0080
2015 年	412.86	434.4536	0.0523

Table 11. Factor F1 prediction**表 11.** 因子 F1 预测

模型		2016	2017	2018	2019	2020
F1-模型_1	预测	13,856.5553	14,666.6326	15,542.4054	16,497.7155	17,548.9937
	UCL	14,746.6184	16,484.8128	18,753.0086	21,607.2392	25,163.5685
	LCL	13,010.6149	13,014.7611	12,793.2011	12,415.9758	11,921.6910

对于每个模型，预测都在请求的预测时间段范围内的最后一个非缺失值之后开始，在所有预测值的非缺失值都可用的最后一个时间段或请求预测时间段的结束日期(以较早者为准)结束。

Table 12. Forecast of agricultural products logistics demand in Beijing from 2016 to 2020**表 12.** 北京地区 2016~2020 年农产品物流需求预测值单位：万吨

年份	2016	2017	2018	2019	2020
y	492.23	478.46	463.57	447.33	429.46

Table 13. Forecast of agricultural products logistics demand in Beijing-Tianjin-Hebei region from 2016 to 2020**表 13.** 京津冀地区 2016~2020 年农产品物流需求量预测值单位：万吨

年份	2016	2017	2018	2019	2020	合计	增长率
北京	492.23	478.46	463.57	447.33	429.46	2311.05	4.02%
天津	746.59	770.86	795.12	819.39	843.65	3975.61	24.39%
河北	12,134.91	12,453.3	12,771.68	13,090.07	13,408.46	63,858.42	13.63%
合计	13,373.73	13,702.62	14,030.37	14,356.79	14,681.57	70,145.08	42.04%

4. 预测结果与建议

4.1. 结果分析

研究京津冀农产品物流需求，首先以北京市的数据为例，选取了影响北京农产品物流发展的 9 个因素，利用因子分析方法选取了两个主成分，对北京农产品物流需求进行了因子回归模型分析，第一因子主要反映出社会经济发展水平影响，第二因子反映出物流水平影响。

分析结果显示，影响京津冀物流需求的关键因素是宏观经济和货运量，随着京津冀地区经济快速发展、人民生活水平的提高和消费的不断升级，京津冀农产品物流需求总体呈增长趋势。具体来说，到 2020 年，北京农产品的产量将会减少到 429 万吨，而河北和天津的产量将会有所增加，2020 年河北的产量相对于 2015 年增长约为 14%，天津增长幅度相对较大约为 24%，这与北京定位为政治和文化中心，而河北为辅助地区的发展定位是一致的，这将为制定京津冀农产品物流发展战略、优化物流供给系统提供重要的决策依据。

4.2. 京津冀农产品物流发展建议

4.2.1. 遵循地区定位发展经济

京津冀整体定位是“以首都为核心的世界级城市群、区域整体协同发展改革引领区、全国创新驱动经济增长新引擎、生态修复环境改善示范区”。其中三省市定位分别为，北京市“全国政治中心、文化中心、国际交往中心、科技创新中心”；天津市“全国先进制造研发基地、北方国际航运核心区、金融创新运营示范区、改革开放先行区”；河北省“全国现代商贸物流重要基地、产业转型升级试验区、新型城镇化与城乡统筹示范区、京津冀生态环境支撑区”[10]。北京、天津的发展离不开河北的支持，同样

北京、天津巨大的消费群体给河北的农产品提供需求市场。北京应深入挖掘消费者的需求特点，河北应提供高质量的农产品满足高企的需求，做好自销与供给。

4.2.2. 大力发展物流和通信基础设施

十九大报告中将物流与水利、交通、电力等并列，要求加强基础设施网络建设，结合分析结果显示，经济因子比重较大，但同时物流发展水平不高，可见，要想提高农产品物流发展水平，要从硬件入手，完善物流通道，购置新型高科技物流设备，建立达标冷库等，更好的促进农产品物流的发展。

4.2.3. 提供高质量的物流服务

消费升级的情况下，消费者的个性化和多样化需求要求物流服务及时、安全、有效的送达，提高物流服务质量才能有利于扩大农产品的销售市场。在产品质量方面，精挑农产品品类并做到以全程冷链思维保障农产品新鲜度；在物流质量方面，开展送货上门，当面验收，货到付款等服务，更多体现差异化和人性化。

参考文献

- [1] 洪玉兰, 肖其荣. 浅谈我国农产品冷链物流存在的问题及发展建议[J]. 物流科技, 2010, 33(9): 42-44.
- [2] 杨文凤. 西藏农产品物流需求研究[D]: [硕士学位论文]. 杨凌: 西北农林科技大学, 2012.
- [3] 李合龙, 李妍, 郑雪仪. SPSS 统计学实验教程[M]. 北京: 清华大学出版社, 2015.
- [4] 刘宇, 吴迎学, 党文峰. 基于多元线性回归的区域物流需求预测研究[J]. 物流技术, 2014(3): 52-54.
- [5] 邱慧, 黄解宇, 董亚兰. 基于灰色系统模型的山西省物流需求预测分析[J]. 数学的实践与认识, 2016, 46(13): 66-70.
- [6] 李夏培. 基于灰色线性组合模型的农产品物流需求预测[J]. 北京交通大学学报(社会科学版), 2017, 16(1): 120-126.
- [7] 易观智库. 2016 年主流消费人群六大标签与互联网行业变化分析[EB/OL]. <http://www.useit.com.cn/thread-11505-1-1.html>, 2016-03-03.
- [8] 高德地图. 2017 年第一季度中国主要城市交通分析报告[R]. 北京: 高德软件有限公司, 交通运输部科学研究院, 清华大学 - 戴姆勒可持续交通研究中心, 阿里云, 2017.
- [9] 智研咨询. 2017~2022 年中国生鲜电商行业市场现状分析及投资前景预测报告[R]. 北京: 智研咨询集团, 2017.
- [10] 中国新闻网. 官方明确京津冀功能定位北京担当“四个中心”[EB/OL]. <http://www.chinanews.com/gn/2015/08-23/7484049.shtml>, 2016-10-26.

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: 2163-1476, 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: orf@hanspub.org