

基于灰色关联与投影算法的铁路货运量影响因素分析

梅晓玲

(郧阳师范高等专科学校数学与财经系,湖北 十堰 442000)

摘要:影响铁路货运量的因素繁多而复杂,每个因素的影响程度都不一样,影响因素的筛选与排序直接决定铁路货运量预测工作的可靠性。灰色关联法是通过数据序列曲线发展态势的相似程度来判断数据序列的联系是否紧密,投影算法是从宏观角度判断两个数据序列的接近程度。结合灰色关联法和投影算法的优点,构造数据序列排序模型,对各影响因素数据序列进行定量分析,为货运量预测的可靠性提供数据基础。实例分析铁路货运量的影响因素的影响力并进行分析排序,为影响因素的定量分析提供新的思路。

关键词:铁路货运量;影响因素;定量分析;灰色关联法;投影算法

中图分类号:TB115

文献标志码:A

引言

铁路货运是国民经济发展的支柱产业,在长距离、大宗物资运输中发挥着其他运输方式所无法替代的重要作用。近几年,对铁路货运量这一领域的研究主要侧重于货运量预测方法的构建与改进,也有部分学者逐步深入地研究铁路货运量的组合预测方法,各种预测方法都显示出不同的优缺点。在建立铁路货运量预测的预测模型时,必须考虑影响铁路货运量的各类因素,各种预测模型能够顺利进行的基石也在于影响因素的准确选择与确定,对历史影响因素进行分析,掌握其发展变化规律,以推测未来的状况。有学者在货运量预测中提到了货运量的影响因素,但深入定量分析的不多见。文献[1-3]在进行铁路货运量预测研究中,均对货运量影响因素进行了定性描述,但均未对货运量影响因素进行定量分析;文献[4]在进行铁路货运量预测研究中,对铁路货运量影响因素进行了定性分析,利用灰色关联分析法找出对货运量影响较大的因素,对影响因素进行了定

量排序,这也使得其预测结果更加可靠。

没有对影响因素的影响程度进行准确的定量分析,仅根据初步建立的指标体系对货运量进行分析预测,将会导致预测结果的可靠性不高,预测精度将减小。货运系统的复杂性决定对铁路货运量影响因素定量分析的重要性,这也逐渐引起许多学者的关注。本文将结合灰色关联分析法和投影算法各自的优点,构造影响因素数据序列的排序模型,对铁路货运量影响因素的影响因子进行定量描述,从而可以有更准确、更合理的排序,为铁路货运量的准确预测提供可靠的数据支持,可以提高预测的准确性。

1 铁路货运量影响因素的定性分析

铁路运输系统是一个受多种因素共同作用的复杂动态系统,具有不确定性、随机性和模糊性^[5]。影响铁路货运量的因素复杂而广泛,因素种类也有很多,且每个因素的影响程度都不一样,应该全面收集相关资料,进行纵向和横向的比较和分析。除国家政治及宏观调

收稿日期:2015-07-20

基金项目:教育部人文社会科学研究青年项目(14YJCZH173);湖北省教育厅重点科研项目(D20155001;D20156001);湖北省教育厅青年科研项目(Q20145001)

作者简介:梅晓玲(1980-),女,湖北十堰人,讲师,硕士,主要从事系统优化方面的研究,(E-mail)meixiaoling2005@126.com

控政策、自然灾害等难以量化因素外,结合相关文献[6-8]分析的现实要求,通过对影响因素的分析比较,明确系统哪些是主要因素和次要因素,抓住主要影响对象,对这些因素进行排序筛选,以减少预测工作量,并得到相对准确的结果。本文中选取影响铁路货运量的影响因素分别为铁路货物周转量、水路与公路货运量之和、国内生产总值、第一产业、第二产业、第三产业、进出口贸易总额、铁路行业就业人数、全国铁路营业里程9个因素,通过排序模型决定这些因素的主次顺序。虽然涉及的影响因素越多,对预测工作会越有帮助。但是,在实际的研究过程中,选取的指标过多,计算模型就会变得过于复杂,反而不能有效地解决实际问题,在一定程度上还会影响到量化分析的计算速度。所以需要影响因素进行整体分析筛选,按对铁路货运量影响因子的大小进行有效排序,筛选出更符合实际、更具有针对性的影响因素,然后构建出货运量影响因素体系。

2 基于灰色关联与投影算法的铁路货运量影响因素定量分析

灰色系统理论中的灰色关联法^[9]是一种系统分析方法,在社会和经济生活中有着广泛的应用,主要用于定量研究系统内多因素之间的复杂性特征,是各因素发展态势量化比较的有力工具。铁路货运系统作为一个复杂系统,也由若干相互关联、相互作用、相互影响的因素交互构成,因此可以考虑将灰色关联法引入到铁路货运系统的影响因素量化分析中。灰色关联法的基本原理是通过数据序列曲线发展态势的相似程度判断序列联系是否紧密,其紧密程度用关联度进行量化,曲线越接近,相应序列之间的关联度越大,反之就越小。投影算法^[10]也是分析数据序列之间关系的有效工具,该方法是将数据序列看成一个向量,则每个序列与参考序列之间均有一个夹角 θ_i ,通过各序列在参考序列上的投影值大小进行排序。因为许多数据波动较大,没有典型的分布规律,投影算法是在几何意义上从宏观角度上对两个数据序列进行大致的判断,得到两者的接近程度,虽然以投影值作为尺度仅仅能反映数据序列间宏观意义上较为粗糙的接近程度,但这种方法却能较好地把握数据序列的接近程度。两种判断思路,一是从数据序列微观的发展态势,二是从宏观投影进行分析,从而可以考虑将两者结合,构造出一种新的数据序列排序方法。

基于灰色关联与投影算法的定量化处理方法构造过程:

(1) 确定参考数列与比较数列

反映铁路货运系统行为特征的数据序列作为参考数列,一般以历年来铁路货运量实际数据为参考数据序列。影响因素参考数据序列:

$$X_0 = \{x_0(k)\} = (x_{01}, x_{02}, \dots, x_{0n}), k = 1, 2, \dots, n$$

m 个影响铁路货运系统行为的因素组成的数据序列作为比较数列,即各因素历年来的实际数据构成比较数据序列。影响因素比较数据序列:

$$X_i = \{x_i(k)\} = (x_{i1}, x_{i2}, \dots, x_{in}), k = 1, 2, \dots, n, i = 1, 2, \dots, m$$

(2) 无量纲化处理

影响铁路货运系统的因素繁杂,影响因素数据序列数值波动较大,先应对数据序列进行适当处理,将其转化为数量级别大体相近的无量纲数据。利用均值化算子进行数据预处理,得到铁路货运量影响因素初值。影响因素参考数据序列转化为:

$$Y_0 = \{y_0(k)\} = (y_{01}, y_{02}, \dots, y_{0n})$$

这里

$$y_0(k) = \frac{x_0(k)}{y_0}$$

$$\bar{y}_0 = \frac{1}{n} \sum_{k=1}^n x_0(k)$$

影响因素比较数据序列转化为:

$$Y_i = \{y_i(k)\} = (y_{i1}, y_{i2}, \dots, y_{in})$$

这里

$$y_i(k) = \frac{x_i(k)}{y_i}$$

$$\bar{y}_i = \frac{1}{n} \sum_{k=1}^n x_i(k)$$

(3) 分别求出数据序列间的关联度与投影值

各因素数据序列 Y_i 与参考数据序列 Y_0 的关联度计算公式:

$$r_{0i}(k) = \frac{\min_i \min_k |y_0(k) - y_i(k)| + \rho \max_i \max_k |y_0(k) - y_i(k)|}{y_0(k) - y_i(k) + \rho \max_i \max_k |y_0(k) - y_i(k)|}$$

得到 Y_i 与 Y_0 的关联度为:

$$R_{0i} = \frac{1}{n} \sum_{k=1}^n r_{0i}(k)$$

这里 $\min_i \min_k |y_0(k) - y_i(k)|$ 为两级最小差; $\max_i \max_k |y_0(k) - y_i(k)|$ 为两级最大差; ρ 为分辨系数, $\rho \in [0, 1]$ 。结果表明, ρ 的大小只改变关联度数值大小,不改变关联度排序,一般取 $\rho = 0.5$ 。 $R_{0i} \in (0, 1]$, R_{0i} 越大表明 X_i 对 X_0 的作用越大,若 $R_{0i} \geq R_{0j}$, $i, j = 1, 2, \dots, m$,说明因素 X_i 优于 X_j 。

将数据序列 $Y_0 = \{y_0(k)\} = (y_{01}, y_{02}, \dots, y_{0n})$, $Y_i = \{y_i(k)\} = (y_{i1}, y_{i2}, \dots, y_{in})$ 视为向量, 计算 Y_i 的行向量在 Y_0 的行向量的投影值:

$$P_{0i} = P_{Y_0}(Y_i) = \frac{\sum_{k=1}^n y_{0k} \cdot y_{ik}}{\sqrt{\sum_{k=1}^n y_{0k}^2}} (i = 1, 2, \dots, m)$$

投影值 P_{0i} 越大, 表明 X_i 对 X_0 的作用越大。

(4) 由确定的关联度与投影值构建组合排序方式

R_{0i} 和 P_{0i} 数值越大, 说明影响力度越大, 影响因素越重要。由于投影值相对于关联度值 ($R_{0i} \in (0, 11)$) 而言不具可比性, 故将 P_{0i} 值进行无量纲化处理, 转化为 $[0, 1]$ 之间的数值进行衡量, 可采取转换公式得到各因素数据序列 Y_i 在参考数列 Y_0 的投影值为:

$$\bar{P}_{0i} = \frac{P_{0i}}{\max_{i=1}^m P_{0i}}$$

由关联度 R_{0i} 与投影值 \bar{P}_{0i} 构造影响因素对参考列影响力的组合排序公式:

$$C_i = \frac{\alpha_1 \bar{P}_{0i} + \alpha_2 R_{0i}}{\alpha_1 \max_{i=1}^m (\bar{P}_{0i}) + \alpha_2 \max_{i=1}^m (R_{0i})}$$

其中, α_1, α_2 满足 $\alpha_1 + \alpha_2 = 1$, 可由实际的数据结构, 结合两种方法的优缺点, 决定对两种方法的偏好程度, 确定它们的数值, 一般取值 $\alpha_1 = \alpha_2 = \frac{1}{2}$ 。

根据 C_i 的大小对影响因素进行排序。 C_i 越大则影响因子越高, 反之则越小。

新的排序方法基于灰色关联度和投影算法, 同时反映了数据曲线形态上的微观相似性差异和数据序列之间在宏观层面的接近程度, 物理含义更加明确, 也让影响因素数据序列的影响力的排序更为准确。

3 实例分析

以文献[11]中 2001~2010 年我国铁路货运量 P_0 / 万吨及其影响因素数据作为研究对象, 即参考数列, 其中影响铁路货运量因素由铁路货物周转量 P_1 / 亿吨 km , 水路与公路货运量之和 P_2 / 万吨, 国内生产总值 P_3 / 亿元, 第一产业 P_4 / 亿元, 第二产业 P_5 / 亿元, 第三产业 P_6 / 亿元, 进出口贸易总额 P_7 / 亿元, 铁路行业就业人员 P_8 / 万人, 全国铁路营业里程 P_9 / 万 km 这 9 个元素组成 (表 1)。

灰色关联排序法和投影算法计算出的关联度 R_{0i} 与投影值 \bar{P}_{0i} 很好的体现了影响因素 P_1 到 P_9 的影响力的大小, 将两者集成得到的反映各影响因素影响力的 C_i 值

进行排序, 主次顺序依次为 $P_2, P_1, P_4, P_6, P_3, P_9, P_5, P_8, P_7$, 与文献[9]的顺序一致。这里使用灰色关联法得到的排序结果与集成方法得到的排序结果一致, 投影算法得到的国内生产总值 P_3 与全国铁路营业里程 P_9 两项影响因素的排序有所改变, 显然选择两种思想的集成方法可增加排序的准确性, 并有效地避免使用一种方法而产生偏差。

表 1 2001~2010 年全国铁路货运量及其影响因素

年	P_0	P_1	P_2	P_3	P_4
2001	167554	12910	180464	89677	14770
2002	178581	13771	192352	99215	14945
2003	193189	14694	207883	109655	15781
2004	204955	15658	220614	120333	16537
2005	224248	17247	241495	135823	17382
2006	249017	19289	268306	159878	21413
2007	269296	20726	290022	183217	22420
2008	288224	21954	310178	211924	24040
2009	314237	23797	338034	257306	28627
2010	330058	25106	355460	300670	34000
年	P_5	P_6	P_7	P_8	P_9
2001	41034	33873	29896	185.20	6.74
2002	45556	38714	39273	187.10	6.87
2003	49512	44362	42184	178.90	7.01
2004	53897	49899	51378	175.80	7.19
2005	62436	56005	70484	172.80	7.30
2006	73904	64561	95539	169.90	7.44
2007	87365	73433	116922	166.60	7.54
2008	103162	84721	140971	165.30	7.71
2009	124799	103880	166740	174.10	7.80
2010	146183	120487	179922	173.30	7.97
关联度 $R_{0i} (i = 1, 2, \dots, 9)$:					
0.9774, 0.9991, 0.6844, 0.8334, 0.6505, 0.6877, 0.5279, 0.5829, 0.6700					
投影值 $\bar{P}_{0i} (i = 1, 2, \dots, 9)$:					
0.9103, 0.8932, 0.6480, 0.8002, 0.6321, 0.6568, 0.4989, 0.5628, 0.6532					
$C_i (i = 1, 2, \dots, 9)$:					
0.9886, 0.9910, 0.7705, 0.8554, 0.6717, 0.7041, 0.5377, 0.6000, 0.6900					

4 结束语

铁路货运量预测工作能够顺利进行, 各类影响因素的分析至关重要。没有对货运量影响因素的重要性进行定量分析会影响预测结果的可靠性与准确性。灰色关联法是通过数据序列曲线发展态势的相似程度判断序列联系是否紧密, 其紧密程度用关联度进行量化。投影算法是在几何意义上从宏观角度对两个数据序列进行大致的判断, 得到两者的接近程度, 接近程度用投影值进行量化。文章结合了灰色关联分析法和投影算法的优点, 构造出新的影响因素数据序列的排序模型。通过实例利用本文方法对所选取的铁路货运量影响因素

进行定量描述,得到影响因素影响因子更为合理的排序,为铁路货运量的准确预测提供可靠的数据支持。基于灰色关联与投影算法的组合排序方法是基于数据序列的分析方法,所以对历史数据采集的准确性要求较高,数据准确与否会直接影响排序的准确性,本文只选取了较为突出的影响因素,当影响因素较多,且不稳定时,本文方法计算量会增加。

参考文献:

- [1] 张华.苏北运河货运量预测研究[D].南京:河海大学,2007.
- [2] 宋光平.铁路货运量预测方法研究[D].北京:北京交通大学,2007.
- [3] 陈实.货运量预测方法及应用研究[D].武汉:武汉理工大学,2008.
- [4] 吴晓玲.基于改进神经网络的铁路货运量预测研究[D].长沙:中南大学,2009.
- [5] 万骞.基于灰色-马尔可夫链的铁路货运量预测研究[J].铁道勘测与设计,2014,32(3):21-26.
- [6] 耿立艳,张天伟,赵鹏.基于灰色关联 LS-SVM 铁路货运量预测[J].铁道学报,2012,34(3):1-6.
- [7] 郭玉华,陈治亚,冯芬玲,等.基于经济周期的铁路货运量神经网络预测研究[J].铁道学报,2010,32(5):1-6.
- [8] 何永占.基于自组织数据挖掘的铁路客运量预测方法研究[J].铁路运输与经济,2013,35(6):28-31.
- [9] 刘思峰,谢乃明.灰色系统理论及其应用[M].4版.北京:科学出版社,2008.
- [10] 王应明.多指标决策与评价的新方法—投影法[J].系统工程与电子技术,1999,21(3):16-20.
- [11] 黄勇,徐景昊.关于铁路货运量预测研究[J].铁道运输与经济,2012,32(4):86-89.

Analysis of the Railway Freight Volume's Influence Factors Based on the Grey Correlation and Projection Algorithm

MEI Xiaoling

(Department of Maths and Finance, Yunyang Teachers' College, Shiyan 442000, China)

Abstract: The factors influencing railway freight volume are various and complicated, and the influence degree of each factor is different. The screening and ranking of influencing factors directly determine the reliability of the railway freight volume forecasting work. Gray correlation method is to determine whether the correlation of data sequences is close by the similarity degree of the data sequence curve development trend. The projection algorithm is to determine the closing degree of two data sequences from the macro perspective. Combining the advantages of grey correlation method and projection algorithm, ranking model was constructed to analyze the data sequences of the factors, and the data base for the reliability of freight volume forecasting was provided. Through instance, the influence of railway freight volume's influence factors was analyzed and analysis sorting was done, and then a new method was provided for quantitative analysis of the influence factors.

Key words: railway freight volume; the influence factors; quantitative analysis; grey correlation method; projection algorithm