

基于信息熵的西部区域交通运输结构分析及其灰色预测

王 辉, 钱勇生*, 曾俊伟, 广晓平

(兰州交通大学 交通运输学院, 甘肃 兰州 730070)

摘 要: 随着新型城镇化的建设, 西部区域综合交通运输网络有了很大的改善, 对其交通运输结构的发展程度进行研究已显得尤为重要. 运用信息熵理论和灰色预测模型, 对西部区域交通运输结构的演变趋势进行分析与预测, 并以甘肃省为例进行实证分析. 结果表明: 由于近几年公路运输的快速发展, 甘肃省运输结构的信息熵呈逐年下降趋势, 各运输方式运量发展不均衡; 对比 GM(1,1) 模型与灰色马尔科夫模型可得, 后者更适用于西部区域交通运输结构的演变预测. 研究结果可为西部区域交通运输结构的完善提供相应的参考.

关键词: 综合运输; 交通运输结构; 信息熵; 灰色预测; 马尔科夫模型

1 引言

随着“一带一路”战略的推进, 西部地区城镇化的建设已成为优化全国国土空间开发格局和提高全国城镇化质量的关键途径. 这势必会对作为先导和支撑的交通运输提出新的时代要求. 对西部区域交通运输结构的演变趋势进行研究, 可以掌握西部区域交通运输结构的发展现状以及各运输类型运量间的差异程度, 并发现其运输结构完善过程中存在的缺陷, 有助于政府部门制定相应的发展规划.

目前, 国内外对综合交通与社会发展间的互动关系研究的较多, 随着综合运输体系的构建, 许多学者又把注意力转移到了区域交通运输结构优化配置方面. 董敏, 郭子坚等^[1]运用系统动力学研究了综合运输体系与社会经济发展间的互动关系, 并进行了实证分析, 验证了所构建模型的可靠性; 唐建桥, 王慈光等^[2]构建了区域交通运输结构优化模型与算法, 通过对珠三角运输结构的分析, 找到其运输结构的最佳调整路径; 于世军等^[3]采用信息熵理论分析了区域交通运输结构的演变过程, 并运用主成分分析与灰色关联相结合的方法, 从众多因子中筛选出了对江苏省交通运输结构演变影响较大的因子; 铁钰等^[4]采用马尔科夫模型对中国女子竞走成绩进行了预测, 结果表明该模型对中国女子竞走成绩的预测具有较高的精度和实用价值; 周庆元^[5]分别应用 GM(1,1) 模型和灰色马尔科夫模型对江苏省历年粮食产量进行预测, 对比发现灰色马尔科夫模型的预测效果更好, 与实际值间的误差更小. 此外, 信息熵理论与灰色预测还分别在土地利用^[6-8]、火灾及交通事故^[9-10]、客流量及空气质量等(见文[11-13])领域的应用中取得了良好的效果. 然而, 国内外对区域交通运输结构的演变过

收稿日期: 2017-05-19

资助项目: 国家社科基金青年项目 (14CJY052); 国家社科基金项目 (15BJY037); 甘肃省自然科学基金项目 (1606RJZA017,1508RJZA112)

* 通信作者

程研究不足,且缺少对其未来可能变化趋势的预测.本文运用信息熵理论来衡量各类运输方式运量间的均衡发展程度,并对运输结构的演变趋势进行预测,最后以甘肃省为例进行实证分析,为西部区域运输结构的完善提供理论支撑.

2 研究方法

2.1 信息熵

信息熵对复杂系统的有序程度可进行定量描述,区域交通运输结构信息熵可以反映出某个时间段运输结构的变化过程,以及各种运输方式协同运作程度.假定西部区域交通运输方式可分为 N 个类型(只考虑公路、铁路、航空三种运输方式,水运和管道运输不做考虑),且运输总量为 A ,则每种运输方式间的运量分别为 A_1, A_2, \dots, A_n ,那么 $A = A_1 + A_2 + \dots + A_n = \sum A_i, (i = 1, 2, \dots, n)$.

根据 Shannon 提出的信息熵,西部区域交通运输结构信息熵 H 可以定义为:

$$H = - \sum P_i \log P_i \quad (1)$$

其中, $P_i = A_i/A = A_i/\sum A_i, (i = 1, 2, \dots, n)$,显然 $\sum P_i = 1$.

式中 H 为交通运输结构的信息熵,当然 $H \geq 0$.

信息熵大小反映的是交通运输结构的均衡程度,信息熵越高,则说明交通运输的职能类型数越多,且各职能类型的运量相差越小,运量分布越均衡.当且仅当 $A_1 = A_2 = A_3 = \dots = A_n$ 时, $P_1 = P_2 = P_3 = \dots = P_n = P_e = \frac{1}{n}$,此时信息熵取到最大值 $H_m = \log n$.

然而,交通运输方式的类型 N 随着社会的发展可能会发生变化,此时交通运输结构的信息熵将会缺乏可比性,故有必要引入交通运输结构均衡度和优势度.依据前面的公式,西部区域交通运输结构的均衡度 J 可定义如下:

$$J = H/H_m = - \sum P_i \log P_i / \log N \quad (2)$$

可以看出,西部区域交通运输结构的均衡度表示的是西部区域交通运输结构发展的均衡程度,其值是实际信息熵与最大信息熵之间的比值. J 值越大,意味着交通运输结构的均衡性就越强.当 $J = 0$ 时,西部区域交通运输结构的发展处于极不均衡态.然而,当 $J = 1$ 时,西部区域交通运输结构的发展达到理想的平衡态.

相应地优势度 I 表达式为:

$$I = 1 - J \quad (3)$$

优势度反映区域内一种或者几种运输方式支配该区域交通运输结构的程度,与均衡度的意义相反.

2.2 灰色 GM(1,1) 模型

1) 构造数据序列

构造原始数据序列: $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$. 对原始时间数据序列进行累加,记为: $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$,其中, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), (k = 1, 2, \dots, n)$.

2) 紧邻均值等权序列

对 $X^{(1)}$ 作紧邻均值等权生成序列 $Z^{(1)}$ 记为:

$$Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)) \quad (4)$$

其中, $z^{(1)}(1) = x^{(1)}(1), z^{(1)}(k) = \frac{1}{2}x^{(1)}(k) + \frac{1}{2}x^{(1)}(k-1), k = 2, 3, \dots, n.$

3) 建立灰色 GM(1,1) 模型

采用一阶单变量微分方程对生成数据序列 $X^{(1)}$ 进行拟合建模, 得到灰色预测模型, 其白化形式为:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b$$

其中, a 为发展参数, b 为灰色作用量, 及参数列 $\hat{a} = (a, b)^T$, 借助最小二乘法求出

$$\hat{a} = (B^T B)^{-1} B^T Y_n \quad (5)$$

式中,

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & 1 \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix}$$

4) 求解与还原

求出白化方程的解, 即时间响应序列为:

$$\bar{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, (k = 1, 2, \dots, n-1) \quad (6)$$

按 $\bar{x}^{(0)}(k+1) = \bar{x}^{(1)}(k+1) - \bar{x}^{(1)}(k)$ 累减生成还原值序列:

$$\bar{x}^{(0)} = (\bar{x}^{(0)}(1), \bar{x}^{(0)}(2), \dots, \bar{x}^{(0)}(n)) \quad (7)$$

根据模型中的发展系数 $-a$, 可方便地判别各个事物的发展趋势. a 的大小表明该事物发展速度的快慢.

2.3 马尔科夫模型

针对 GM(1,1) 模型预测中的不足之处, 运用马尔科夫模型对灰色预测结果进行修正.

马尔科夫模型的具体运算步骤:

1) 构建 GM(1,1) 模型, 计算得到相应的灰色预测值

$$\bar{x}^{(0)} = (\bar{x}^{(0)}(1), \bar{x}^{(0)}(2), \dots, \bar{x}^{(0)}(n)). \quad (8)$$

2) 计算预测值与实际值之间的相对误差

$$Q = \frac{\bar{x}^{(0)}(k)}{x^{(0)}(k)} \times 100\% \quad (9)$$

3) 根据相对误差的大小, 划分状态区间

依据误差范围的集中程度, 划分适当的状态区间. $\otimes_i = [\otimes_{1i}, \otimes_{2i}]$, 其中 \otimes_{1i} 是状态区间相对误差的最小限制, \otimes_{2i} 是其最大限制.

4) 构建状态转移矩阵

状态转移可能性的大小是用状态转移概率来描述的, 定义状态转移概率:

$$P_{ij}^{(k)} = \frac{M_{ij}^{(k)}}{M_i} \quad (10)$$

其中, $M_{ij}^{(k)}$ 是由状态 E_i 经过 k 步到状态 E_j 出现的次数, M_i 是 E_i 出现的次数. n 为划分的状态数目, 所构建 $n \times n$ 的状态转移概率矩阵如下:

$$P^{(k)} = \begin{bmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \dots & P_{1n}^{(k)} \\ P_{21}^{(k)} & P_{22}^{(k)} & \dots & P_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ P_{n1}^{(k)} & P_{n2}^{(k)} & \dots & P_{nn}^{(k)} \end{bmatrix}$$

状态转移矩阵的作用是通过现在的初始状态确定未来状态的可能转向.

5) 对灰色预测值进行修正

由马尔科夫状态转移概率矩阵, 可以确定预测值的相对误差出现的状态区间 $[\otimes_{1i}, \otimes_{2i}]$, 取相对误差状态区间的中间值对灰色预测值进行修正, 得到修正后的预测值 y , 具体公式为:

$$y = \frac{\bar{x}^{(0)}(k)}{1 \pm \frac{1}{2}(\otimes_{1i} + \otimes_{2i})}. \tag{11}$$

3 实例分析

3.1 甘肃省交通运输结构演变分析

依据《甘肃发展年鉴 (2015)》整理得到甘肃省 2005-2014 年交通运输结构数据, 并根据上述的公式计算甘肃省货运结构的信息熵、均衡度与优势度, 如表 1 所示.

表 1 甘肃省货运结构的信息熵、均衡度及优势度

年份	货运总量 (万吨)	铁路货运 量 (万吨)	公路货运 量 (万吨)	民航货运 量 (万吨)	信息熵 (H)	均衡度 (J)	优势度 (I)
2005	25795.08	3274	22520	1.08	0.3810	0.3460	0.6532
2006	27460.32	3633	23826	1.32	0.3913	0.3561	0.6439
2007	29452.33	4126	25325	1.33	0.4056	0.3692	0.6308
2008	22714.44	4512	18201	1.44	0.4992	0.4544	0.5456
2009	25459.33	4646	20812	1.33	0.4757	0.4330	0.5670
2010	28977.13	4926	24050	1.13	0.4563	0.4153	0.5847
2011	34146.22	5355	28790	1.22	0.4348	0.3957	0.6043
2012	45831.34	6313	39517	1.34	0.4012	0.3652	0.6348
2013	51471.55	6394	45072	5.55	0.3763	0.3426	0.6574
2014	57235.96	6450	50780	5.96	0.3532	0.3215	0.6785

利用表 1 中的数据绘制甘肃省货运结构信息熵、均衡度与优势度的变化趋势, 如图 1 所示.

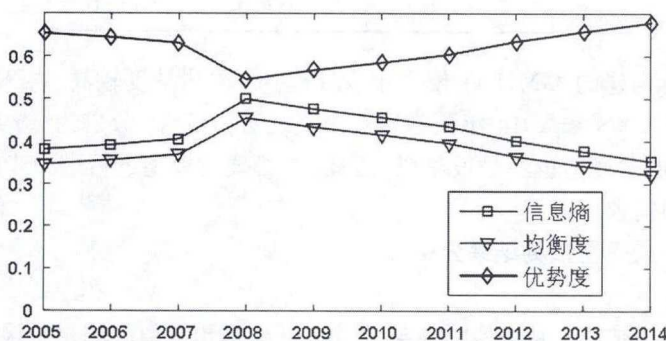


图 1 甘肃省货运结构信息熵、均衡度以及优势度的演变趋势

按照甘肃省货运结构的信息熵、均衡度和优势度的变化趋势, 甘肃省货运结构近十年的变化能被划分为两个阶段. 2005-2008年是甘肃省货运结构明显上升的阶段, 这个时期各种运输方式货运量的变化相对平稳, 都在保持稳定的增加. 而2009-2014年是甘肃省货运结构明显的下降阶段, 此阶段各运输方式间的货运量比重相差较大, 公路货运量和航空货运量在迅速增加, 其运量在五年间翻了好几倍, 而铁路运量的增幅相对较小.

近几年, 政府加大了对甘肃省公路交通的建设力度, 公路运输发展的非常快, 而铁路运输并未发生大的变化. 数据表明, 2005年甘肃省公路里程仅仅4.13万公里, 而截止到2014年, 公路里程达到14.01万公里, 相比2005年翻了近4倍, 而铁路里程仅是由原来的0.23万公里增加到了0.38万公里, 加上公路运输灵活便捷的优点, 使得公路运输量大大增加, 以致于各运输方式运量的差距越来越大, 即信息熵和均衡度不断下降, 相应的优势度却不断上升.

3.2 甘肃省交通运输结构演变趋势预测

3.2.1 GM(1,1) 模型预测结果分析

根据GM(1,1)模型原理, 利用MATLAB编程, 得到甘肃省货运结构信息熵的GM(1,1)预测模型:

$$\hat{x}^{(1)}(k+1) = -26.0419e^{-0.0175k} + 26.4229, k = 1, 2, \dots, n-1.$$

相应的GM(1,1)模型预测值如表2所示.

表2 GM(1,1) 模型预测结果

年份	信息熵实际值	GM(1,1) 模型	
		预测值	相对误差
2005	0.3810	0.3810	
2006	0.3913	0.4518	0.1545
2007	0.4056	0.4439	0.0944
2008	0.4992	0.4362	-0.1263
2009	0.4757	0.4286	-0.0990
2010	0.4563	0.4211	-0.0771
2011	0.4348	0.4138	-0.0483
2012	0.4012	0.4066	0.0135
2013	0.3763	0.3995	0.0617
2014	0.3532	0.3926	0.1115

由表2知, 所构建的GM(1,1)模型中2012年的预测精度较高, 相对误差绝对值仅为1.35%, 然而2006、2008和2014年预测相对误差绝对值较大, 分别为15.45%, 12.63%以及11.15%, 其他年份的预测值也不是很理想. 因此, 有必要对模型进行进一步的改进和优化, 使其能符合实际运用需求.

3.2.2 灰色马尔科夫模型预测结果分析

1) 状态划分

按照甘肃省货运结构信息熵实际值与GM(1,1)模型预测值间的相对误差, 可将其划分为如表3所示的四种状态.

表 3 货运结构信息熵状态划分

状态	相对误差区间
E1	(-0.2, -0.1]
E2	(-0.1, 0]
E3	(0, 0.1]
E4	(0.1, 0.2]

2) 构建状态转移矩阵并修正预测值

各年份预测值所在区间状态如表 4 所示.

表 4 各年份预测值所在区间状态

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
相对误差		0.1545	0.0944	-0.1263	-0.0990	-0.0771	-0.0483	0.0135	0.0617	0.1115
所处区间		E4	E3	E1	E2	E2	E2	E3	E3	E4

根据状态转移概率矩阵的确定方法, 得到:

$$P^{(1)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 2/3 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}, P^{(2)} = \begin{bmatrix} 0 & 0.667 & 0.333 & 0 \\ 0.111 & 0.444 & 0.333 & 0.111 \\ 0.111 & 0.333 & 0.444 & 0.111 \\ 0.333 & 0 & 0.333 & 0.333 \end{bmatrix}.$$

进一步对预测值修正, 得到如表 5 所示结果.

表 5 GM(1,1) 模型预测结果与灰色马尔科夫模型预测结果对比

年份	信息熵实际值	GM(1,1) 模型		灰色马尔科夫	
		预测值	相对误差	预测值	相对误差
2005	0.3810	0.3810		0.3810	
2006	0.3913	0.4518	0.1545	0.3929	0.0041
2007	0.4056	0.4439	0.0944	0.4228	0.0424
2008	0.4992	0.4362	-0.1263	0.5132	0.0280
2009	0.4757	0.4286	-0.0990	0.4512	-0.0515
2010	0.4563	0.4211	-0.0771	0.4433	-0.0285
2011	0.4348	0.4138	-0.0483	0.4356	0.0018
2012	0.4012	0.4066	0.0135	0.3872	-0.0349
2013	0.3763	0.3995	0.0617	0.3805	0.0112
2014	0.3532	0.3926	0.1115	0.3414	-0.0334

由于 2014 年处于第四种状态, 经过一年的转换, 确定 2015 年甘肃省货运结构信息熵最有可能处于第三种状态, 取第三种状态区间的中间值对灰色预测值进行修正, 得到 2015 年甘肃省货运结构信息熵马尔科夫修正值为 0.3674. 由以上转移矩阵 $P^{(2)}$ 确定 2016 年甘肃省货运结构信息熵状态处于第三种状态, 经过计算得到修正值为 0.3610. 同理, 可求出 2017-2020

年的马尔科夫修正值如表 6 所示.

表 6 甘肃省 2015-2020 年货运结构信息熵预测值

年份	2015	2016	2017	2018	2019	2020
预测值	0.3674	0.3610	0.3548	0.3486	0.3425	0.3366

一般情况,灰色模型预测的结果是一条平滑的直线,只体现了信息熵的一种变化趋势,而不能反映其波动性,因而导致在货运结构预测中对波动较大的年份预测误差较大.而灰色马尔科夫修正值与甘肃省货运结构信息熵实际值间的相对误差较小,预测精度比 GM(1,1) 模型高,适用于西部区域运输结构演变趋势的预测.依据以上的预测结果分析可得,甘肃省货运结构的信息熵和均衡度还会持续下降,但幅度会不断减小.

4 结语

1) 本文运用信息熵理论对甘肃省 2005-2014 年交通运输结构的演变趋势进行分析,结果表明:近几年,甘肃省公路运输的快速发展使得各运输方式的运量相差越来越大,运输结构的信息熵和均衡度持续下降,而优势度正不断增加.

2) 首次将灰色马尔科夫模型应用到区域交通运输结构演变趋势的预测上,结果表明:该模型在预测波动性序列中比单一的 GM(1,1) 模型精度要高,实用性较强,具有明显的优势;对预测结果分析可得,在未来几年中,甘肃省运输结构的信息熵还会持续下降,但下降幅度会逐渐减小,建议甘肃省应注重交通运输结构的优化和调整,使各运输类型达到均衡发展态势.本文的尝试无疑对西部区域交通运输结构的建设和完善具有较好的借鉴价值.

参考文献

- [1] 董敏,郭子坚,唐国磊,张茜.综合运输体系与区域经济互动影响的系统动力学模型[J].大连海事大学学报,2010,36(4):64-67.
- [2] 唐建桥,王慈光,左大杰.区域运输结构优化路径模型及算法[J].数学的实践与认识,2013,43(131):85-91.
- [3] 于世军,刘英舜,梁先登.区域交通运输结构信息熵演变及动力分析[J].交通运输系统工程与信息,2011,11(4):91-96.
- [4] 铁钰,郑超,代长义,徐颖华.基于灰色马尔科夫模型的中国女子竞走成绩的预测[J].数学的实践与认识,2015,45(3):187-194.
- [5] 周庆元.基于灰色马尔科夫模型的粮食产量预测方法[J].统计与决策,2012(17):64-66.
- [6] 王晓娇,陈英,齐鹏,张仁陟.基于信息熵的张掖市土地利用结构分析及其灰色预测[J].干旱区研究,2011,28(1):92-97.
- [7] 谢汀,刘爱宁,高雪松,邓良基等.基于信息熵和灰色关联的成都市建设用地结构时空变化及驱动力分析[J].农业现代化研究,2015,36(1):118-125.
- [8] Huan H Y, Tan Q M. The forecast of cultivate land quantity based on Grey-Markov Model: A case study of Jiangsu Province[J]. Grey Systems, 2011(1): 276 - 280.
- [9] Mao Z L, Sun J H. Application of grey-markov model in forecasting fire accidents[J]. Procedia Engineering, 2011, 11(11): 314-318.
- [10] Chen K M, Xie L F, Xiang W S. Traffic accidents prediction using improved Grey-Markov Model[J].

Advanced Materials Research, 2012, 378-379: 222-225.

- [11] 杨琦, 杨云峰, 冯忠祥, 赵现伟. 基于灰色理论和马尔科夫模型的城市公交客运量预测方法 [J]. 中国公路学报, 2013, 26(6): 169-175.
- [12] Li J, Liu W J. Population forecasting in China based on the Grey-Markov Model[J]. International Conference on Information Management, 2011, 3(2): 133-136.
- [13] 杨锦伟, 孙宝磊. 基于灰色马尔科夫模型的平顶山市空气污染物浓度预测 [J]. 数学的实践与认识, 2014, 44(2): 64-70.

Study on the Transportation Structure of Western Region Base on Information Entropy and Gray Prediction

WANG Hui, QIAN Yong-sheng, ZENG Jun-wei, GUANG Xiao-ping

(School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: With the construction of new urbanization, the comprehensive transportation network of western region has been greatly improved, so the study of the development degree of transportation structure is important. The information entropy and gray prediction model are used to study the trend of the transport structure of western region, and Gansu Province is took as an example to do empirical analysis. The results show that the information entropy of transportation structure in Gansu province is decreasing year by year, and the development among the different transportation modes is unbalanced because of the rapid development of highway transportation in recent years. In addition, the prediction results of gray Markov model is more reasonable than the GM(1,1) model in the prediction of the evolution trend of transportation structure. The results of this study can provide a reference for the improvement of regional transportation structure.

Keywords: comprehensive transportation; transportation structure; information entropy; gray prediction; Markov model