

文章编号:1005-0523(2020)02-0089-08

基于熵权法的农村公路线网规模组合预测

王雪梅¹, 郭 旷¹, 秦连群², 鲍雨婕³, 周希灿⁴

(1. 同济大学浙江学院交通运输工程系, 浙江 嘉兴 314000; 2. 浙江省安吉县交通运输局, 浙江 湖州 313300;
3. 上海理工大学管理学院, 上海 200093; 4. 宁波诺丁汉大学土木工程系, 宁波 315100)

摘要:农村公路线网规模预测是“四好农村路”规划的主要内容之一。鉴于单一预测方法的局限性,基于熵权法理论,对国土系数法、人口经济密度法、连通度法合理分配权重,提出组合预测方法;以浙江省湖州市安吉县的农村公路为实例,对安吉县历年农村公路网的规模进行拟合,分别验证三种单一预测方法的合理性及其有效性;通过绝对平均误差、绝对平均百分比误差、均方根误差三种误差指标,构建评价矩阵,计算得出三种单一预测方法的权重系数分别为0.115 4, 0.240 3, 0.644 3,验证了熵权法能够客观地分配权重系数;对未来特征年农村公路线网规模进行预测,表明组合预测方法可以有效降低单一方法预测的风险。

关键词:农村公路;线网规模;组合预测;熵权法

中图分类号:U412.1

文献标志码:A

DOI:10.16749/j.cnki.jecjtu.2020.02.012

农村公路是由县道、乡道和村道组成的^[1]。农村公路线网规模主要与公路网的区域面积、人口分布、社会经济发展水平、交通需求和主要节点分布有关。农村公路线网布局规划的基本控制量是线网规模。因此,农村公路线网规模应该有一个适度的值,使之与人口、经济和社会发展水平相协调,从量和度上反映农村公路的发展水平。

Bate 和 Granger^[2]在 1969 年提出了组合预测理论,将多种预测模型合理组合,并且最大化地提高了多种预测样本信息的利用率,使得这比单一模型更系统、更全面。所以,组合预测理论成为了一个重要的研究方向。例如,文献[3-4]使用层次分析法来分配权重,在公路货运周转量和电力负荷预测方面取得了较好的效果;文献[5-7]通过熵权法分配权重,在干线公路通车里程、公路货运量和光伏输出功率的预测方面得到了较好的效果;文献[8]利用遗传算法分配权重,在预测交通流中得到了较好的效果;文献[9]则提出了 4 种基于相关性的组合预测方法,并验证了其方法的有效性;文献[10]引入指数加权移动平均模型对不同时段的电力负荷数据进行权重分配,改进后的电力负荷预测模型获得了较高的精确度。

文献[11-13]采用了国土系数法、综合系数法、人口经济密度法、连通度法、弹性系数法、公路周转量分析法等方法对农村公路线网总规模进行了预测。本文通过对上述预测方法的对比分析,拟选取国土系数法、人口经济密度法、连通度法三个方法对农村公路线网规模进行预测,这三种预测方法均属于回归分析方法,选取与线网规模相关的区域面积、人口、经济、交通节点等因素,拟合自变量与因变量的相关关系,并对各自模型中的系数构建线性和对数回归方程,从而实现农村公路线网总规模的估计和预测。同时,选取绝对平均误差(MAE, mean absolute error)、绝对平均百分比误差(MAPE, mean absolute percent error)、均方根误差(RMSE, root mean square error)三种误差指标全面评价预测效果。再基于熵权法理论,计算单一方法的权重系数,以此构建组合预测模型,从而得到更加科学、客观的预测结果。

收稿日期:2019-10-18

基金项目:安吉县“四好农村路”中长期发展规划项目(0219002)

作者简介:王雪梅(1988—),女,讲师,研究方向为交通运输规划与管理。

1 预测方法

1.1 单一预测方法

1.1.1 国土系数法

国土系数法是指在综合考虑公路网所处区域的人口、面积、经济发展水平基础上,计算得到这个区域内合理的理论道路长度。该理论认为“道路长度与人口和面积的平方根及其经济指标成正比”,见式(1)

$$L=K\sqrt{P\times A} \quad (1)$$

式中: L 为道路长度,km; P 为人口数,万人; A 为国土面积(平方公里); $K=f(\overline{GDP})$: K 是人均 \overline{GDP} 的函数, \overline{GDP} 单位为元, K 单位为公里/(万人 \times 平方公里)^{0.5}。

1.1.2 人口经济密度法

人口经济密度法是指在综合考虑所处地区的人口密度和经济强度基础之上,根据所预测的道路面积密度,从而得出公路网的总规模。人口密度是人口数量与土地面积的比值,经济密度是单位土地面积的国内生产总值的产出。可以看出,道路面积密度与经济密度、人口密度之间有一定的相关关联度,见式(2)

$$L=K\times A \quad (2)$$

式中: L 为道路长度,km; A 为国土面积(平方公里); $K=f(\overline{GDP})$:单位国土面积人口,GDP产值的函数, $\frac{(\text{万人}\times\text{百万美元})^{\frac{1}{2}}}{\text{平方公里}}$ 。

1.1.3 连通度法

连通度法也可以称为节点模型法,是体现所处公路网中各个节点之间依靠公路交通来连接的连通强度,反映了规划区域内各交通节点之间的连通状况,体现其通达深度。网络几何学是连通度法的理论依据,通过分析网络几何形状结构来构建合理规模的公路网连通度模型,见式(3)

$$L=C\times\delta\times\sqrt{n\times A} \quad (3)$$

式中: L 为公路网规模,km; n 为区域内的节点数; A 为区域面积(平方公里); δ 为线网非直线系数; C 为公路网连通度。

通过模型可以看出,在确定所处区域面积及其节点数的情况下,路网的连通度和变形系数是影响公路网规模的主要变量。其中,路网的变形系数受到路线所处区域的地形分布影响,若是区域内的地势起伏较大,山地数量多,与之相对应的路线也会有较大的弯曲程度,则该路网的变形系数也较大;反之,则较小。路网连通度的值越高,代表各节点间的连接程度也越高,相应的路网规模越完善。对于不同的路网形式有不同的连通度计算方法。

1.2 组合预测方法

1.2.1 组合预测的基本原理

组合预测是一种选取适当的权重,对多种不同的预测方法所得的结果进行加权平均后的预测方法。在某次预测中,对预测对象 y 假设有 m 种预测方法,将这 m 种方法组合成最终的一个预测方法,从而得到对 y 的预测结果,即 $y=Y(y_1,y_2,y_3,\dots,y_m)$ 。如果各种方法的权重为 $W=[w_1,w_2,w_3,\dots,w_m]^T$,满足 $\sum_{i=1}^m w_i=1$,则组合预测模型可表示为

$$Y(y_1,y_2,y_3,\dots,y_m)=\sum_{i=1}^m w_i y_i \quad (4)$$

式中: y_i 为第 i 个预测方法的预测值; w_i 为第 i 种预测方法的权重。

确定各个预测方法在组合预测中的权重分配是组合预测的核心,权重系数优化标准的不同形成了组合预测方法的不同。

1.2.2 基于熵权法的组合预测方法

在信息论中,熵表示不确定性的量度,指标的信息熵越小,则该指标提供的信息量越多,其作用在综合评价中也越大,相应的权重占得越高。熵权值代表评价指标在竞争意义上的激烈程度,具有较强的客观性,是确定组合预测模型组合权重系数并提高组合预测模型精度的有效方法之一^[14]。

文献[15-16]分别对熵的基本原理进行了理论研究。当系统可能处于几种不同状态,每种状态出现的概率为 $p_i(i=1,2,\dots,m)$ 时,该系统的熵定义为

$$E=-\sum_{i=1}^m p_i \ln p_i \tag{5}$$

m 种单项预测方法, n 项误差评价指标,形成的原始指标矩阵 $R=(r_{ij})_{m \times n}$, 对于某个指标 r_j , 有

$$E_j=-\sum_{i=1}^m p_{ij} \ln p_{ij} \tag{6}$$

其中

$$p_{ij}=r_{ij} / \sum_{i=1}^m r_{ij}, j=1,2,\dots,n \tag{7}$$

本文基于熵权法理论,根据每个单一预测模型在系统中的信息熵,为每个单一预测方法分配权重,由此构建一个组合预测模型,达到客观、准确的对农村公路线网规模进行预测的目的。

1.3 误差指标

本文为了对各个预测方法得到的结果更好的评价、分析、比较,选取 3 个相关评价指标

1) 绝对平均误差

$$MAE=\frac{1}{N} \sum_{t=1}^N |y_t-\hat{y}_t| \tag{8}$$

2) 绝对平均百分比误差

$$MAPE=\frac{1}{N} \sum_{t=1}^N \left| \frac{y_t-\hat{y}_t}{y_t} \right| \tag{9}$$

3) 均方根误差

$$RMSE=\sqrt{\frac{1}{N} \sum_{t=1}^N (y_t-\hat{y}_t)^2} \tag{10}$$

式中: N 为预测样本数; y_t 为在 t 时刻所得到的测量值; \hat{y}_t 为在 t 时刻所得到的预测值。

2 实例分析

本文以浙江省湖州市安吉县农村公路线网规模预测为例,安吉县域面积 1 886 km², 选取 2014—2018 年农村公路通车里程,GDP,人口以及人均 GDP 数据,如表 1 所示。

表 1 安吉县 2014—2018 年农村公路通车里程、GDP、人口以及人均 GDP 数据

Tab.1 Data of rural road mileage, GDP, population and per capita GDP of Anji county from 2014 to 2018

项目	2014 年	2015 年	2016 年	2017 年	2018 年
通车里程/km	1 890	1 942	1 968.3	1 982.6	2 008.59
GDP/亿元	285.06	303.35	324.87	360.31	404.32
人口/万人	46.38	46.41	46.61	46.85	47.07
人均 GDP/元	61 591	65 379	69 848	77 097	86 099

2.1 单一预测模型构建

2.1.1 国土系数法

模型标定方法步骤,如下:

- ① 计算 $\frac{L}{\sqrt{P \times A}} = K$;
- ② 标定 $K=f(\overline{AGDP})$;
- ③ 整合模型。

依据式(1)反推 K 值,采用线性和对数模型对国土系数法 K 值进行拟合。分析得到的数学模型如下,拟合 K 值见图 1 和表 2。

1) 线性模型

$K=0.000\ 01AGDP+5.716, R^2=0.803$ 显著相关。

2) 对数模型

$K=0.916\ln(AGDP)+3.647, R^2=0.834$ 显著相关。

2.1.2 人口经济密度法

模型标定方法:

- ① 计算 $K=\frac{L}{A}$;
- ② 计算 $\overline{GDP} = \sqrt{GDP \times P} / A$;
- ③ 标定 $K=f(\overline{GDP})$;
- ④ 整合模型。

依据式(2)反推 K 值,采用线性和对数模型对人口经济密度法 K 值进行拟合。分析得到的数学模型如下,拟合 K 值见图 2 和表 2:

1) 线性模型

$K=4.575APGDP+0.735, R^2=0.864$ 显著相关。

2) 对数模型

$K=0.308\ln(APGDP)+1.877, R^2=0.878$ 显著相关。

2.1.3 连通度法

综合考虑连通度模型中山区以及平原地形所产生的影响,确定非直线系数 $\delta=1.3, A=1\ 886\ \text{km}^2, n=185$ 个,依据式(3)反推 C 值。本文采用线性和对数模型对连通度 C 值进行拟合。根据分析得到的数学模型如下,见图 3 和表 2。

1) 线性模型

$K=0.036X+2.441, R^2=0.946$ 显著相关。

2) 对数模型

$K=0.092\ 1\ln X+2.462, R^2=0.992$ 显著相关。

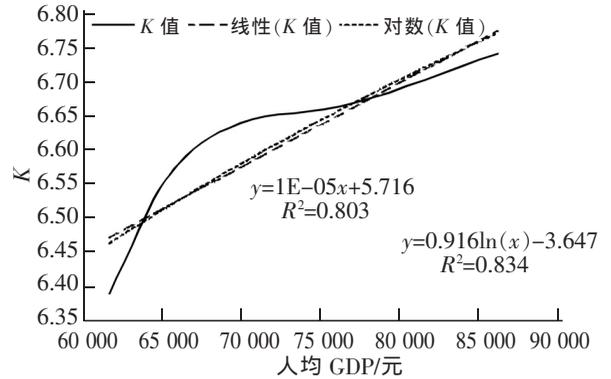


图 1 国土系数法拟合 K 值

Fig.1 Fitting K value of the land coefficient method

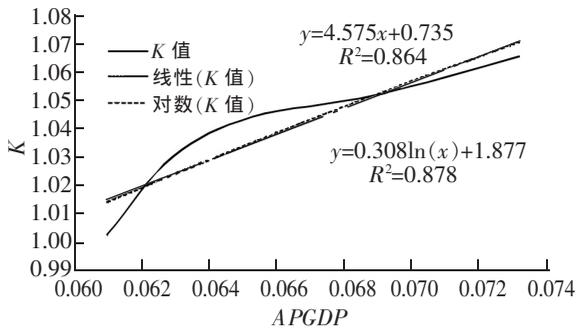


图 2 人口经济密度法拟合 K 值

Fig.2 Fitting K value of the population and economic density method

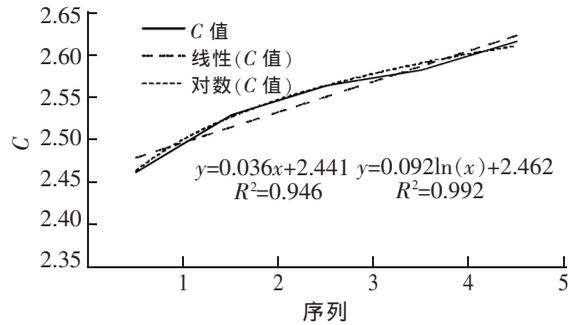


图 3 连通度法拟合 C 值

Fig.3 Fitting C value of the connectedness method

通过分析经济社会发展对公路线网连通状况的需求,本路网介于方格网状与三角网状之间,其临界状态的连通度计算方法如下,参考上述路网连通临界状态的分析,得到此区域的连通度值 C 在 2~3.22 之间为合理的。

1) 对于结构是方格形的路网,如图 4 所示。取边长 $s=1$,则 $N=(n+1)(n+1), L=2n(n+1), A=n^2$,可得

$$C = \frac{L}{\sqrt{AN}} = \frac{2n(n+1)}{\sqrt{n^2(n+1)^2}} = 2$$

2) 对于结构是正三角形的路网,如图 5 所示。取边长 $s=1$,则 $N=(n+1)(n+1), L=2n(n+1)+n^2, A = \frac{\sqrt{3}}{2} n^2$ 可得

$$C = \frac{L}{\sqrt{AN}} = \frac{2n(n+1)+n^2}{\sqrt{\frac{\sqrt{3}}{2} n^2(n+1)^2}} = \sqrt{\frac{\sqrt{3}}{2}} \left(2 + \frac{n}{n+1} \right)$$

对 $n \rightarrow +\infty$ 取极限,得到 $C=3.22$ 。

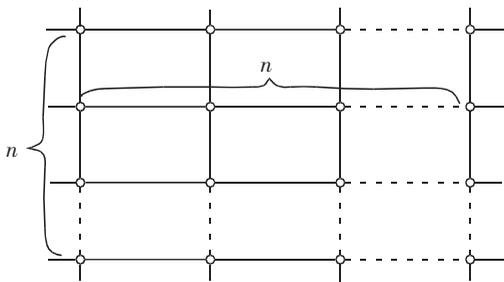


图 4 方格形结构路网

Fig.4 Grid structure road network

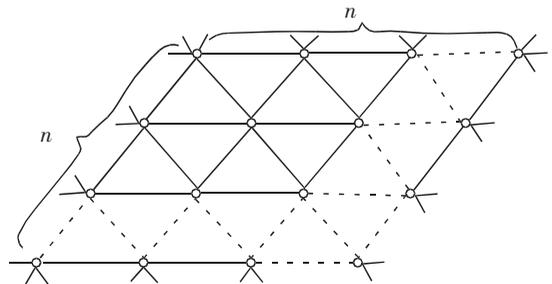


图 5 正三角形结构路网

Fig.5 Equilateral triangle structure road network

2.2 拟合结果及误差评价

依据表 2 所示的国土系数法 K 值、人口经济密度法 K 值、连通度法 C 值的拟合结果,分别带入式(1)~式(3),得到如表 3 所示的三种单一预测方法对应的农村公路线网规模拟合值,采用 $MAE, MAPE, RMSE$ 三个误差指标衡量预测方法的拟合精度,见表 3,三种单一预测方法精度均较高,说明本文所采用的方法均适用于农村公路线网规模的预测,其中连通度法的拟合精度最高。

表2 三种单一预测方法的系数拟合结果
Tab.2 Coefficient fitting results of three single forecasting methods

预测方法	模型	2014年	2015年	2016年	2017年	2018年
国土系数法 K 值	线性模型	6.332	6.370	6.414	6.487	6.577
	对数模型	6.455	6.510	6.570	6.661	6.762
	平均值	6.393	6.440	6.492	6.574	6.669
人口经济密度法 K 值	线性模型	1.014	1.023	1.033	1.050	1.070
	对数模型	1.015	1.025	1.036	1.053	1.071
	平均值	1.015	1.024	1.035	1.052	1.071
连通度法 C 值	线性模型	2.477	2.513	2.549	2.585	2.621
	对数模型	2.462	2.526	2.563	2.590	2.610
	平均值	2.470	2.519	2.556	2.587	2.616

表3 三种单一预测方法和组合预测方法拟合结果及误差指标评价

Tab.3 Fitting results and error index evaluation of three single forecasting methods and combination forecasting method

预测方法	误差指标			拟合里程/km				
	MAE	MAPE	RMSE	2014	2015	2016	2017	2018
国土系数法	26.22	1.33%	30.04	1890.9	1905.29	1924.99	1954.1	1987.1
人口经济密度法	12.44	0.64%	14.55	1913.6	1931.2	1951.8	1983.3	2019.14
连通度法	4.7	0.24%	5.33	1896.3	1934.6	1962.8	1986.7	2008.4
组合预测法	6.90	0.36%	8.82	1899.8	1930.4	1955.8	1982.1	2008.5

2.3 组合预测模型构建

用三个误差指标来评价三种预测方法,进而得到评价矩阵为

$$Y = \begin{bmatrix} 26.22 & 1.33\% & 30.04 \\ 12.44 & 0.64\% & 14.55 \\ 4.7 & 0.24\% & 5.33 \end{bmatrix}$$

具体计算熵权法的权重步骤,如下:

① 数据处理。用样本最小值除以其他值: $X_{ij} = \frac{\min y_{ij}}{y_{ij}}$;

② 计算贴近度 $x_{ij} = \frac{X_{ij}}{\sum_{i=1}^m \sum_{j=1}^n X_{ij}}$;

③ 计算熵值。根据熵的定义,各预测方法权重的不确定性由误差指标 j 通过下列条件熵来表示: $E_j =$

$$\sum_{i=1}^m \frac{x_{ij}}{x_j} \ln \frac{x_{ij}}{x_j}; \text{再对 } E_j \text{ 进行归一化处理得到评价指标 } j \text{ 的重要性熵值: } e(x_j) = \frac{x_{ij}}{\ln m} E_j;$$

④ 计算指标的客观权重 $\theta_j = \frac{1 - e(x_j)}{\sum_{j=1}^n [1 - e(x_j)]}$;

⑤ 计算各预测方法的权重 $\omega_j = \frac{\sum_{i=1}^m (\theta_j X_{ij})}{\sum_{j=1}^n \sum_{i=1}^m (\theta_j X_{ij})}$ 。

根据上述步骤得到,国土系数法(Y_1)、人口经济密度法(Y_2)和连通度法(Y_3)的权重系数分别为 $\{0.1154, 0.2403, 0.6443\}$,其中连通度法的权重系数最高,说明其预测值更接近实际情况。依据公式(4),得到组合预测模型公式为

$$Y=0.1154Y_1+0.2403Y_2+0.6443Y_3$$

利用上述公式,可求得组合预测模型的拟合值,误差指标见表 3,与实际值对比如图 6 所示。虽然组合预测模型的拟合精度较连通度法低些,但精度相较国土系数法和人口经济密度法都有提高。一方面,本文选取节点数量和非直线系数较合理,确定 C 值较精确,连通度法获得较好的拟合精度;另一方面,后续对未来特征年农村公路线网规模预测中发现(见表 5),连通度法计算出的预测值过于保守,与历年来公路实际里程的涨幅不匹配。所以,通过实例证明,本文基于熵权值理论,提出的组合预测方法能有效避免单一预测方法的局限性,降低单一预测方法的风险。

2.4 预测结果

依据 2.1 和 2.2 中构建的模型,计算得到如表 4 所示的国土系数法 K 值、人口经济密度法 K 值、连通度法 C 值的预测结果,分别代入式(1)~式(3),得到如表 5 所示的三种单一预测方法对应的农村公路线网规模预测值,代入式(4),计算得到如表 5 所示的组合预测值。不难发现,人口经济密度法预测的线网规模较大,增长速度偏快,国土系数法预测的线网规模适中,但增长速度过快,连通度法预测的线网规模较小,且增长速度过慢。组合预测方法规避了三种单一预测方法的风险,获得较好的预测结果和增长速度,符合安吉县农村公路线网规模发展趋势。

表 4 三种单一预测方法的系数预测结果

Tab.4 Coefficient prediction results of three single forecasting methods

预测方法	2020 年	2025 年	2030 年	2035 年
国土系数法 K 值	6.808	7.243	7.839	8.504
人口经济密度法 K 值	1.107	1.189	1.282	1.391
连通度法 C 值	2.667	2.782	2.888	2.990

表 5 三种单一预测方法和组合预测方法预测结果

Tab.5 Forecasting results of three single forecasting methods and combination forecasting method

年份	国土系数法		人口经济密度法		连通度法		组合预测方法	
	预测结果 /km	年平均增长率/%						
2020	2 035.4	—	2 087.4	—	2 048.0	—	2 056.0	—
2025	2 185.9	1.48	2 241.6	1.48	2 136.1	0.86	2 167.2	1.08
2030	2 388.3	1.85	2 418.2	1.58	2 217.5	0.76	2 285.5	1.09
2035	2 616.2	1.91	2 622.8	1.69	2 295.8	0.71	2 411.3	1.10

3 总结

在农村公路的布局规划中,其前提和首要工作是合理地确定农村公路网发展规模。本文在充分掌握安吉县现有公路线网建设状况以及预测理论方法的基础上,为避免单一预测方法的局限,基于熵权法理论,提

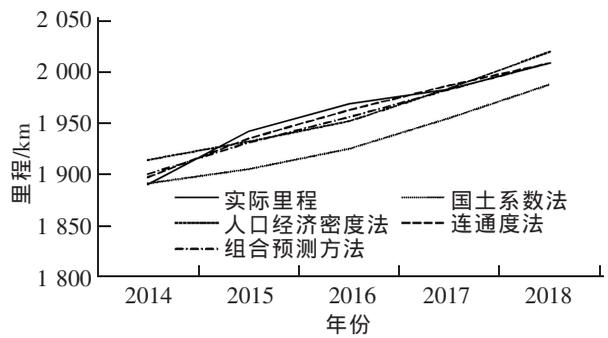


图 6 三种单一预测方法和组合预测方法拟合值与实际值对比

Fig.6 Comparison between fitting values and actual values of three single forecasting methods and combination forecasting method

出组合预测方法。计算结果得出,国土系数法、人口经济密度法和连通度法的权重系数分别为 0.115 4,0.240 3,0.644 3,其中连通度法的权重系数最高,这与连通度法的误差指标最低相符,也验证了熵权法能够客观地分配权重系数。研究结果表明,国土系数法、人口经济密度法、连通度法和组合预测方法均适用于农村公路线网规模,拟合精度均较高。组合预测方法预测结果更合理,降低了单一预测方法的风险,符合农村公路线网规模发展趋势,可为各地“四好农村路”的建设提供切实可行的依据,使之更好地与区域社会经济平衡、协调、有序的发展。

参考文献:

- [1] 王肖峰,吴楼,王雪梅,等.安吉县农村公路发展规划研究[J].公路,2019,64(2):185-190.
- [2] BATES J M,GRANGER C W J. The combination of forecasts[J]. Journal of the Operational Research Society,1969,20(4):451-468.
- [3] 寇毅.组合预测模型在公路货运周转量预测中的运用[J].中国水运,2007,7(11):216-217.
- [4] 王颖,何子君,杨玉新,等.基于层次分析法的中长期电力负荷组合预测[J].河北电力技术,2011,30(2):32-34.
- [5] 赵山春,牛惠民.基于熵的干线公路通车里程组合预测[J].交通科技与经济,2010,12(1):25-27.
- [6] 詹斌,吕腊梅,黄馨.基于熵权法的公路货运量组合预测[J].物流技术,2016,35(6):54-57.
- [7] 杨锡运,刘欢,张彬等.基于熵权法的光伏输出功率组合预测模型[J].太阳能学报,2014,35(5)744-749.
- [8] 犹勇.组合预测方法在城市交通流量预测中的运用[J].重庆工商大学学报(自然科学版),2007,24(4):122-125.
- [9] 王应明.基于相关性的组合预测方法研究[J].预测,2002,21(2):58-62.
- [10] 李颖玥,王勋,康琛,等.基于指数加权移动平均多维组合模型的电力负荷预测[J].华东交通大学学报,2019,36(5):102-108.
- [11] 李玉梅.河南省农村公路网规划与评价[D].西安:长安大学,2006:76-81.
- [12] 王凯.农村公路网的规划方法与应用研究[D].深圳:深圳大学,2016:14-17.
- [13] 王菲,陈景雅.设区市区县道公路网规模研究[J].华东交通大学学报,2015,32(1):54-58.
- [14] 任玉珑,刘焕,望玉丽,等.基于熵权法和支持向量机的中长期电力负荷预测[J].统计与决策,2009(14):46-48.
- [15] 陈华友.组合预测方法有效性理论及其应用[M].北京:科技出版社,2008:55-73.
- [16] 熊崇俊.基于熵的组合预测法研究[J].科技进步与对策,2006(4):68-69.

Combination Forecasting of Rural Road Network Scale Based on Entropy Weight Method

Wang Xuemei¹, Guo Kuang¹, Qin Lianqun², Bao Yujie³, Zhou Xican⁴

(1. Department of Transportation Engineering, Tongji Zhejiang College, Jiaxing 314000, China; 2. Anji Transportation Bureau, Huzhou 313300, China; 3. School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China; 4. Department of Civil Engineering, University of Nottingham Ningbo China, Ningbo 315100, China)

Abstract: The rural road network scale prediction is one of the main contents of "Four-merit rural roads" planning. In view of the limitations for a single forecasting method, this paper proposes a combined forecasting method based on the theory of entropy weight, which reasonably allocates weight of the land coefficient method, the density of population and economic method, and connectivity method. Taking the rural road in Anji county as an example, it fits the rural road network scale and verifies the validity and rationality of the three single forecasting methods. Through three kinds of error indexes of absolute average error, absolute average percentage error and root mean square error, it constructs evaluation matrix, and respectively calculates the weight coefficients of the three single forecasting methods, namely 0.115 4, 0.240 3 and 0.644 3, which verify that entropy weight method can distribute weight coefficients objectively. It shows that the combination forecasting method can reduce the forecasting risk of a single model, which predicts the rural road network scale in future years.

Key words: rural road; network scale; combination forecasting; entropy weight method