

# Application of Machine Learning in the Prediction of Compressive Strength of Concrete

Weidong Rao

Yunnan University of Finance and Economics, Kunming Yunnan  
Email: 625952318@qq.com

Received: Feb. 14<sup>th</sup>, 2017; accepted: Mar. 3<sup>rd</sup>, 2017; published: Mar. 6<sup>th</sup>, 2017

---

## Abstract

In this paper, the data of compressive strength of concrete are modeled by decision tree, boosting, random forest, artificial neural network and support vector machine methods. Ten-fold cross-validation is adopted to assess the performance of these methods in terms of the prediction accuracy. It is seen that the Random Forest method has the best performance in general.

## Keywords

Machine Learning, Prediction, Compressive Strength

---

# 机器学习在混凝土抗压强度预测中的应用

饶炜东

云南财经大学, 云南 昆明  
Email: 625952318@qq.com

收稿日期: 2017年2月14日; 录用日期: 2017年3月3日; 发布日期: 2017年3月6日

---

## 摘要

本文对混凝土的抗压强度的数据采用决策树、Boosting、随机森林、人工神经网络、支持向量机这五种方法进行建模, 采用十折交叉验证评价预测精度。发现随机森林法具有较好的预测效果。

## 关键词

机器学习, 预测, 抗压强度

Copyright © 2017 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

随着科学技术的发展, 混凝土越来越广泛地应用到各个工程建设中, 并且根据使用功效的差别, 可以配置不同类型的混凝土。然而随着掺合料、粗骨料、细骨料等比例的变化, 混凝土的抗压强度也受到极大的影响[1] [2] [3] [4]。混凝土的抗压性能是设计混凝土配合比的一个重要参考指标, 然而配合比与抗压性能之间并不是简单的线性关系, 而是极其复杂的非线性关系。近年来, 为了更快地预测某种配合比混凝土的抗压强度, 许多试验和研究方法被开发出来。机器学习的各种方法也被用来预测各种类型混凝土的强度, 例如神经网络[5], 支持向量机[6], 高斯过程[7], 相关向量机[8]等。本文使用一组实际数据, 建立各种机器学习方法的混凝土抗压强度的预测模型, 并通过交叉验证的方法来验证各种机器学习方法的可靠性。

## 2. 实验数据

本文研究的数据来源于台湾重华大学信息管理系的叶怡成教授, 数据下载地址为, <http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>。该数据总观测值为 1030 个, 没有任何缺失数据, 其中共有 9 个分量, 可细分为输入变量 8 个, 分别为: 水泥, 矿渣, 粉煤灰, 水, 超减水剂, 粗骨料, 细骨料和养护龄期; 输出变量 1 个, 混凝土抗压强度。该数据是没有经过放缩处理的原始数据, 表 1 中列出了数据的最大值和最小值。

**Table 1.** Concrete data  
**表 1.** 混凝土数据

	最大值	最小值
水泥(Kg/m <sup>3</sup> )	540	102
矿渣(Kg/m <sup>3</sup> )	359.4	0
粉煤灰(Kg/m <sup>3</sup> )	200.1	0
水(Kg/m <sup>3</sup> )	247	121.8
超减水剂(Kg/m <sup>3</sup> )	32.2	0
粗集料(Kg/m <sup>3</sup> )	1145	801
细骨料(Kg/m <sup>3</sup> )	992.6	594
养护龄期(天)	365	1
抗压强度(Mpa)	2.33	82.6

通过对数据的初步了解，可以看出，这些数据可以把混凝土的抗压强度作为因变量，其他变量作为自变量，建立一定的回归模型，并做出相应的预测。然而该数据过于复杂，传统的方法并不能获得一个较好的结果，于是考虑使用机器学习的方法。

### 3. 机器学习

机器学习是计算机科学的子领域，它使计算机能够在没有明确编程的情况下学习[9]。不同于传统方法的模型驱动，机器学习是数据驱动，它探索构建了从数据中学习并对数据进行预测的算法[10]。

传统的统计模型大多对数据的具体分布有一定的要求或者是假定，然而在现实中，真实的数据可能不满足那些假定或要求，这个时候再用传统的方法，它们的某些优良的性质将不可能得到，甚至某些结论都是错误的。这个时候选择机器学习的方法会比较合理，机器学习法不需要对数据分布做任何假定，并且产生的结果也可以用交叉验证的方法来进行评价。

#### 3.1. 决策树回归

决策树分类是通过一定的规则对数据样本进行分类的过程，当数据结构为连续形变量时，称为决策树回归[11]。该方法模型是将数据样本不断地进行划分，根据不同的条件，数据将划分到不同的分支结构下面，并且这种分支结构能够保证数据对这一条件依赖的差别达到最大，从而最后将所有数据划分到各种不同的分支下[12]。这样得出的每一次划分，得出的偏差比其他划分方法都小，从而对于全部划分，能够保证最小偏差。使用 R 软件的程序包 *rpart.plot* 进行决策树回归，得出的决策树如图 1。

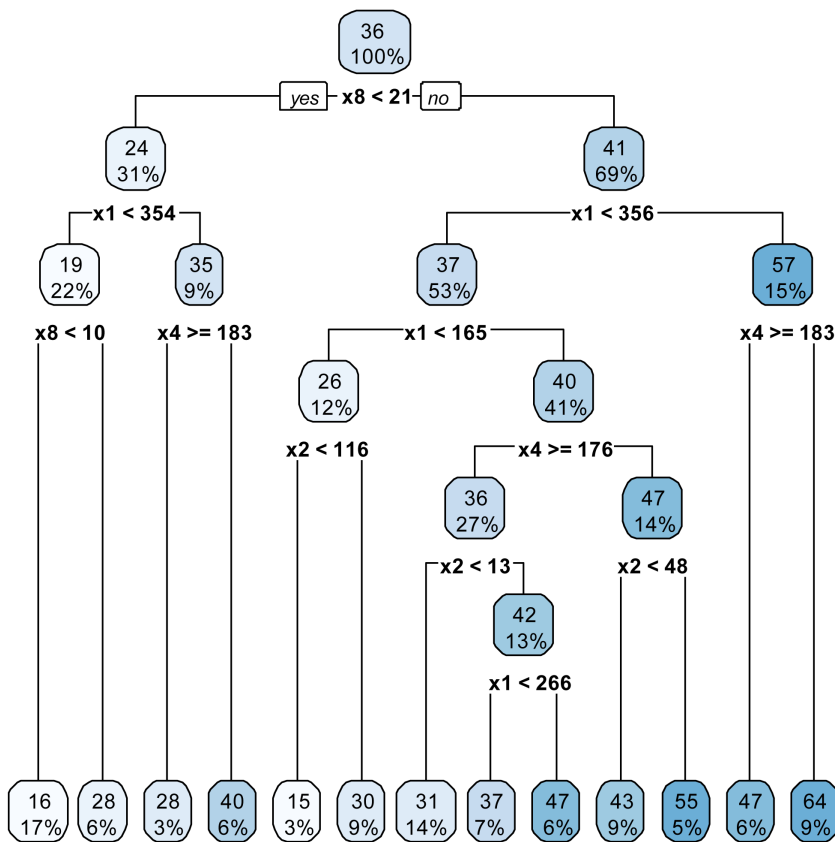


Figure 1. Decision tree

图 1. 决策树

### 3.2. Boosting 回归

Kearns 和 Valiant 提出了强学习器和弱学习器的概念，并且证明出两者可以等价，即弱学习器可以提升为强学习器，这就是 Boosting 算法的由来[13]。后来，Boosting 算法被用于回归分析，以训练误差较大的回归算法做弱学习器，通过 Boosting 算法将误差较大的回归算法，如决策树回归，当作弱学习器，从而构造出训练误差较小的强学习器。通过弱学习器提升为强学习器的方法各有不同，从而有各种不同的 Boosting 方法。本文采用的是基于模型的 Boosting 算法，并且弱学习器选择决策树回归。在 R 软件中可以使用程序包 *mboost* 进行 Boosting 回归。

### 3.3. 随机森林

随机森林由 Breiman 提出，是决策树的一种组合方法[14]。该方法的基本原理是[15]：

- 1) 将所有观测值当作 bootstrap 抽样的样本，并且得出一组抽样样本；
- 2) 随机选择抽样样本的部分数据建立决策树，并且对数据随机划分；
- 3) 重复 1)和 2)过程  $m$  次，获得  $m$  个决策树，构成随机森林模型；
- 4) 数据预测时，每个决策树对因变量进行预测，得出  $m$  个预测值，最后取出现次数最多的数值作为该因变量的预测值。

该方法的优点在于随机性的引入增强了抗噪声的能力，并且能够有效地避免过拟合的情况，对复杂的数据有较强的适应能力。在 R 软件中可以使用程序包 *randomForest* 进行随机森林回归。

### 3.4. 人工神经网络回归

人工神经网络是由大量的节点构成，其相关原理是，将上层节点的值加权后传递给下一层，依次传递到最终输出节点，再根据输出节点的误差大小情况给前面节点层一个激励或者抑制的信号，从而改变权重，最后经过反复传递，达到输出误差在某个范围内。某一节点的加权过程可以表示为[16]

$$y = f\left(\sum_k w_k x_k + \theta_k\right)$$

其中  $y$  表示该节点层的输出值， $f$  为激活函数(activation function)， $w_k$ 、 $x_k$  和  $\theta_k$  分别代表该节点层第  $k$  个节点的权重、值和阈值。使用 R 软件的程序包 *neuralnet* 进行人工神经网络回归，权值衰减选取(0.1, 0.5)，隐藏节点取(4, 5, 6)，经过 6 次试验，选取样本内的预测相对误差值最小的方案，权值衰减取 0.5，隐藏层节点个数为 4。

### 3.5. 支持向量机回归

在一般的线性回归问题中，假设拟合函数为  $f(x) = \omega^T x + b$ ，支持向量机的目的是最小化  $\omega^T \omega / 2 + C \sum (\xi_i + \xi_i^*)$ ，使得

$$|y_i - f(x_i)| \leq \varepsilon + \xi_i$$

其中， $\varepsilon$  表示误差， $C$  是对误差的惩罚， $\xi_i$  和  $\xi_i^*$  指松弛因子。该二次规划问题可以转化为其对偶问题，

$$\max g(a, a^*) = -\frac{1}{2} \sum (a_i - a_i^*) (a_j - a_j^*) x_i x_j + \sum y_i (a_i - a_i^*) - \varepsilon \sum a_i + a_i^*$$

在非线性问题中，支持向量机的基本思想是应用非线性映射  $\Phi(x)$ ，将数据样本  $x$  投影到某个高维的特征空间，然后在该特征空间内进行线性回归，最后将结果返回到原始空间[17]，公式如下：

**Table 2.** Results of cross validation error  
**表 2.** 交叉验证误差结果

机器学习	均值	最小值	最大值
决策树	0.257	0.216	0.333
Boosting	0.207	0.176	0.237
<b>随机森林</b>	<b>0.145</b>	<b>0.107</b>	<b>0.186</b>
人工神经网络	0.227	0.164	0.295
支持向量机	0.170	0.134	0.242

$$f(x, \omega) = \omega \Phi(x) + b$$

其中， $\omega$  和  $b$  分别表示权重和阈值。

在高维特征空间中数据样本的内积运算  $\langle \Phi(x_i), \Phi(x_j) \rangle$  较为复杂，因此考虑使用核函数  $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$  近似，核函数能够使得向量在映射前后的内积相等，核函数可以选多项式函数，径向基函数以及 Sigmoid 函数等[18] [19]。在 R 软件中可以使用程序包 *e1071* 进行支持向量机回归，该程序包使用的是径向基函数。

#### 4. 交叉验证

交叉验证，是一种的统计学方法，先将样本数据分割成  $n$  份，依次取出一份作为测试集，剩余部分作为训练集，用机器学习方法建立模型后，将测试集带入模型，从而得出预测值的相对误差，依次进行  $n$  次以后，可以得出  $n$  个预测误差，取其均值作为该方法的交叉验证的误差，最后通过比较误差来判断各种建模方法的可靠性。本文中对所有机器学习方法采用的都是 10 折交叉验证(10-fold cross validation)，使用 R 软件随机将数据分成 10 份，并且后面的计算全部使用这些已经分好的集合。

本文总共使用了决策树、Boosting、随机森林、人工神经网络、支持向量机这五种方法来对该混凝土的抗压强度数据进行回归分析，并且对这五种方法都进行了十折交叉验证，通过对比十折交叉验证的结果，可以看出五种方法对混凝土抗压强度预测的可靠性。接下来看看机器学习法得出预测值的相对误差的均值、最小值和最大值，的结果如表 2 所示。

可以看出使用随机森林回归得出的十折交叉验证的误差均值最小为 0.145，因此可以考虑使用随机森林建立抗压强度的模型。

#### 5. 结论

影响混凝土抗压强度的因素跟多，为了建模具有复杂结构的数据，引入机器学习的方法。通过各种主流的机器学习方法对一组实际数据建立相关模型，随后使用交叉验证的方法，检测各种机器学习方法的可靠性，发现误差均值最小的为随机森林法。

#### 参考文献 (References)

- [1] 胡立志, 刘士清, 宋正林, 周建文. 粉煤灰混凝土抗压强度和成熟度之间的规律研究[J]. 混凝土, 2014(8): 61-63.
- [2] 张林春, 张爱莲, 邓宏卫. 矿物掺合料对粉煤灰陶粒混凝土抗压强度的影响[J]. 混凝土, 2010(6): 83-85.
- [3] 夏冰华, 刘远才, 徐国林. 废玻璃混凝土抗压性能的试验研究[J]. 混凝土, 2013(7): 26-29.
- [4] 肖建庄, 范玉辉, 林壮斌. 再生细骨料混凝土抗压强度试验[J]. 建筑科学与工程学报, 2011, 28(4): 26-29.

- [5] 胡明玉, 唐明述. 神经网络在高强粉煤灰混凝土强度预测及优化设计中的应用[J]. 混凝土, 2001(1): 13-17.
- [6] 马金鱼. GA-LSSVM 模型在粉煤灰混凝土强度预测中的应用[J]. 宁夏工程技术, 2009, 8(4): 325-327.
- [7] 张研, 苏国韶, 燕柳斌. 高强混凝土强度预测的高斯过程机器学习模型[J]. 混凝土, 2011(11): 18-20.
- [8] 李仲欣, 韦灼彬, 沈锦林. 相关向量机在珊瑚混凝土抗压强度预测中的应用[J]. 混凝土, 2016(7): 1-6.
- [9] Simon, P. (2013) *Too Big to Ignore: The Business Case for Big Data*. Wiley.
- [10] Kohavi, R. and Provost, F. (1998) Glossary of Terms. *Machine Learning*, **30**, 271-274.
- [11] Quinlan, J.R. (1986) Induction of Decision Trees. *Machine Learning*, **1**, 81-106.  
<https://doi.org/10.1007/BF00116251>
- [12] 唐华松, 姚耀文. 数据挖掘中决策树算法的探讨[J]. 计算机应用研究, 2001, 18(8): 18-19.
- [13] Kearns, M.J. and Valiant, L.G. (1996) Cryptographic Limitations on Learning Boolean Formulae and Finite Automata. *Journal of the ACM*, **41**, 67-95.
- [14] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [15] 宋捷, 吴喜之. 一种新的 Boosting 回归树方法[J]. 统计与信息论坛, 2010, 25(5): 9-13.
- [16] 王鸿斌, 张立毅, 胡志军. 人工神经网络理论及其应用[J]. 山西电子技术, 2006(2): 41-43.
- [17] 王定成, 方廷健, 高理富, 等. 支持向量机回归在线建模及应用[J]. 控制与决策, 2003, 18(1): 89-91.
- [18] Xu, L., Krzyzak, A. and Suen, C.Y. (1992) Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. *IEEE Transactions on Cybernetics*, **22**, 418-435. <https://doi.org/10.1109/21.155943>
- [19] 李昌华, 杨兵, 谢维信. 手绘图形结构的识别方法研究[J]. 西安电子科技大学学报: 社会科学版, 2000, 27(7): 98-102.

**期刊投稿者将享受如下服务:**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [sa@hanspub.org](mailto:sa@hanspub.org)