

Prediction of compressive strength of concrete containing fly ash using data mining techniques

Francisco F.Martins ¹ and Aires Camões ²

¹ Associate Professor, C-TAC/UM, Territory, Environment and Construction Research Centre, University of Minho, Guimarães, Portugal. Email: ffm@civil.uminho.pt

² Assistant Professor, C-TAC/UM, Territory, Environment and Construction Research Centre, University of Minho, Guimarães, Portugal. Email: aires@civil.uminho.pt.

Number of words: 5864

Number of tables: 5

Number of figures: 6

Abstract

The concrete compressive strength is the most used mechanical property in the design of concrete structures. Therefore, the use of rational models to its prediction, to simulate the effects of its different constituents and its properties can play an important role in the achievement of the safety-economy required. Models to forecast the concrete compressive strength have already been presented before by some researchers. However, the comparison of different rational models and the application of models to predict the importance of the different constituents in the concrete behaviour have not yet been approached. Therefore, developing these models will be necessary namely to take into account the quality, i.e. the activity, of the most used mineral addition in concrete: fly ash. This study compared different Data Mining techniques to predict the compressive strength of fly ash concrete along time. The presented models are able to learn the complex relationships between several variables like the uniaxial compressive strength, the different concrete compounds and its mix design, the different properties of the fly ash used and the relative influence of its.

Keywords: concrete strength, fly ash, data mining, artificial neural networks, support vector machines

1. Introduction

The concrete compressive strength is the most used mechanical property in the design of concrete structures. The prediction of concrete strength before its placement allows engineers to improve planning and quality control. Moreover, a well defined concrete strength prediction can save time, accelerating the overall construction, mainly in large constructions with a lot of concreting stages. However, environmental concerning and the need of higher performance concretes have lead to more complex mixes than the traditional ones. Nowadays the recycled industrial wastes available leads to the quasi-systematic use of mineral addition in concrete, namely fly ash from electrical power plants. Therefore, the factors that affect the concrete compressive strength have increased in number and complexity leading to non linear relationships. These relationships are usually empirical ones or based on correlation analysis due to the fact that until today there is no general law describing the phenomena and explaining this hugely complex system. In this context the traditional methods based on generalization of previous experience are not enough accurate to provide satisfactory relationships. That's why intelligent models with the capability of learning with examples have been applied in the prediction of the concrete compressive strength (Gupta, 2007, Kim *et al.*, 2002, Lai and Serra, 1997, Saridemir, 2009 and Topçu and Saridemir, 2008).

In this study the forecasting of concrete strength was carried on using Data Mining (DM) techniques taking into account the properties of the most used mineral addition in concrete: fly ash (FA). These techniques are powerful intelligent tools that learn with examples and experiences and were applied successfully to predict concrete strength by other authors (Gupta, 2007, Kim *et al.*, 2002, Lai and Serra, 1997, Saridemir 2009 and Topçu and Saridemir, 2008). However, these predictions didn't take into

account the quality of the FA used. In fact, to the best of our knowledge, the physical and the chemical parameters of the FA have never been used before in these predictions.

In this research work one compared the predictive capacity of several DM techniques to forecast the concrete compressive strength taking into account physical and chemical parameters of the FA used in order to quantify its activity and its influence on the compressive strength. Therefore the main characteristics of the FA, usually determined and currently available, were considered as input variables in the DM models.

This paper begins to present a brief description of FA properties effect on concrete strength. Then it is presented a definition of the data mining techniques and its application on prediction of concrete behaviour. It is also explained how to evaluate the different algorithms of DM. Finally the results, discussion and conclusions are presented.

2. Influence of FA characteristics on concrete compressive strength

The incorporation of FA in concrete is not new, began in the 1930s in the USA (ACI Committee 232, 1996), and the effects of FA in concrete performance are sufficiently understood and well documented (Wesche, 1991, Malhotra and Ramezani pour, 1994, ACI Committee 232, 1996, Joshi and Lohtia, 1997 and Camões, 2002). Therefore, in this section we only intend to address the main aspects directly related to the present work, namely the effect of physical and chemical characteristics of FA on concrete performance, mainly on compressive strength.

Unless physical and chemical properties of FA can vary considerably depending on their origin and even between different supplies from the same Power Station some properties can be generalized.

2.1 Physical properties

FA particles have spherical shape, are essentially vitreous (80 %), and have a high fineness. Certain FA has also irregular or angular particles. Most of the particles have a diameter of less than 1 μm and 150 μm , and can be thinner or thicker than cement ones. The average diameter of FA particles is usually similar to cement ones, between 7 μm and 12 μm (Joshi and Lohtia, 1997). The Blaine specific surface ranges frequently from 250 m^2/kg to 550 m^2/kg (Alonso and Wesche, 1991).

Broadly, the physical characteristics of FA have an appreciable variation with respect to its origin. According to Malhotra and Ramezani-pour, 1994, the FA source is not related to its fineness or its specific surface. The authors consider that there is a very slight correlation between thinness, as measured by percentage retained on sieve of 45 μm and Blaine specific surface.

FA particles larger than 125 μm are very porous. The occurrence of these particles is associated with large amounts of carbon. This unburned material is responsible for the high specific surface typically found in FA. As a result, high levels of carbon imply greater demand for water in concrete containing FA with high loss on ignition (LOI). The carbon content also affects the strength of concrete actions to freeze-thaw: the higher the carbon content of FA, the lower the resistance of concrete (Alonso and Wesche, 1991).

According to these authors, FA should have a particle size and specific surface similar to or lower than the cement to avoid variations in physical properties of concrete particularly in workability. This recommendation is linked to the presence of carbon in FA, avoiding high amounts of particles larger than 125 μm , endowed with high porosity and with greater concentration of carbon particles.

It is commonly accepted that more fineness leads to greater pozzolanic activity. Like most chemical reactions occur more rapidly with increasing fineness of the particles it is expected that the pozzolanic activity of fly ash must be dependent on the area available for reaction (Jalali, 1991, Neville, 1995).

Furthermore, the spherical shape of FA particles is particularly advantageous from the point of view of the water demand and high specific surface indicates that the material exhibits high reactivity with calcium hydroxide (Neville, 1995).

In this context, it is supposed that the higher specific surfaces and the lower LOI the better the activity of FA on concrete will be.

2.2 Chemical properties

The chemical constituents of the majority of FA particles are compounds and crystals of silica, SiO_2 , alumina, Al_2O_3 , ferric oxide, Fe_2O_3 , and lime, CaO . In substantially lower amounts there are other components such as MgO , Na_2O , K_2O , SO_3 , MnO and TiO_2 . FA also contains carbon particles not consumed in the combustion (Alonso and Wesche, 1991, Malhotra and Ramezaniapour, 1994 and ACI Committee 232, 1996). The alluded four main components of FA record appreciable change, and may have understood values, according to the ACI Committee 232, 1996, between the following: SiO_2 – 35 % to 60 %; Al_2O_3 – 10 % to 30 %; Fe_2O_3 – 4 % to 20 %; CaO – 1 % to 35 %.

The FA pozzolanic activity is closely related to the SiO_2 content, as it is the amorphous silica that will combined with the free lime and water giving resulting in the formation of additional quantities of calcium silicate hydrates. According to Alonso and Wesche, 1991, FA with SiO_2 content less than 35 % are virtually inactive as pozzolans and should not be used in concrete. However Halstead, 1986, Mehta, 1985 and Joshi and Lohtia, 1997, have a different opinion and consider that in terms of chemical composition, with the exception of calcium content, the variation of FA constituents does not affect significantly the cementitious or pozzolanic properties. Alonso and Wesche, 1991, also indicate that FA with high content of lime (15 % to 40 %) can have hydraulic and binder properties and its inclusion in concrete should be avoided.

The role played by the oxides is not consensual. ACI Committee 232, 1996, refers the failure of studies aimed at establishing relations between the amount of oxides - SiO_2 , Al_2O_3 , Fe_2O_3 - and the performance of the FA. Although the content of SiO_2 appear to be related to the pozzolanic activity, reduced levels of this product do not imply negative effects on the behavior and characteristics of fresh or hardened concrete (Camões, 2002). The establishment of minimum values for the total amount of oxides ($\text{SiO}_2 + \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$) is criticized by several authors (Swamy, 1993, ACI Committee 262, 1996 and Joshi and Lohtia, 1997). While it may be understandable attempt to ensure the presence of sufficient glassy constituents (according to Mindess, 1994, the higher the glass phase is the better FA will be), one should not confuse the reactivity of the glassy phase with increasing resistance, because the development of mechanical characteristics is always achieved by the combined effect of several other factors, such as the fineness and the properties of cement. Malhotra and Ramezani pour, 1994, mention the existence of a good correlation between the amount of $\text{SiO}_2 + \text{Al}_2\text{O}_3$ and

the pozzolanic activity at long term. The same authors also reported that the amount of Fe_2O_3 presented in almost FA occurs in the form of hematite and magnetite non reactive. These aspects should be the reason why researchers have been obtained low correlations between the pozzolanic activity index (measured by the ratio between the compressive strength) and the amount of $\text{SiO}_2 + \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$.

The sulphates, SO_3 , may affect the optimum FA content as it can influence its mechanical properties and setting time. A maximum level must be considered to avoid an excess of SO_3 in hardened concrete that can contribute significantly to a worse performance particularly when the concrete is subjected to sulphates attack (Jalali, 1991).

The carbon in FA is the result of incomplete combustion of coal and organic additives used in the process of ash collection. In general, the carbon is not measured directly but using the determination of LOI content. LOI includes, besides the free carbon, losses of combined water and carbon dioxide from the hydrates and carbonates present in the FA. However LOI is assumed, without committing a great error, as approximately equal to the carbon content.

The contribution of the amount of carbon in FA is decisive in the water demand of pastes, mortars and concretes. The total water needed to obtain a paste with the same consistency is greater the higher the carbon content. The carbon contained in FA has high porosity and large specific surface, being able to absorb not only significant amounts of water as well chemical admixtures dissolved during the mixing of concrete, including superplasticizers, air entraining agents or setting time retardants.

According to Alonso and Wesche, 1991, in general, the lower the carbon content the better the FA for use in mortar and concrete. ACI Committee 363, 1992, states that

when used in high strength concretes the desirable LOI should be less than 3 %, although higher values are considered in standards, enabling its acceptance. Day, 1995, states that the carbon content should not exceed 8 %, much lower levels are preferred. However, Malhotra and Ramezaniapour, 1994, concluded that the carbon does not significantly affect the pozzolanic activity index, determined by the relationship between the compressive strength.

3. Data Mining

3.1 Definition and applications to the concrete behaviour

The process of Knowledge Discovery in Databases encompasses three main stages: pre-processing, Data Mining and pos-processing. In the Data Mining stage an algorithm is applied for extraction of patterns from the data (Fayyad *et al.*, 1996). Generally, in all Data Mining techniques there are a set of training examples and a set of testing examples. The algorithms learn with the training examples and are tested with the testing examples. During the learning process the several parameters of the algorithms are adjusted to optimize the results. To assess the accuracy of the algorithms some metrics can be used most of them based on errors. After the validation of the algorithms they can be used as models for forecasting the values of the output variables.

Many authors have been succeeded with the application of intelligent tools. Several researchers (Kasperkiewicz *et al.*, 1995, Lai and Serra, 1997, Ni and Wang, 2000, and Kim and Kim, 2002) used neural networks to predict 28-day compressive strength of concrete for different mixes. Topçu and Saridemir, 2008, developed neural networks and fuzzy logics for predicting the 7, 28 and 90 days compressive strength of concretes containing high-lime and low-lime fly ashes. Zarandi *et al.*, 2008, developed

fuzzy polynomial neural networks (FPNN) to predict the 28-day compressive strength of concrete. They constructed six different FPNN architectures and used experimental data of 458 different concrete mixes collected from three distinct sources. They were well succeeded in predicting the compressive strength of concrete mixes. Saridemir, 2009, based on several experimental results gathered from the literature used neural networks to predict the compressive strength of concretes containing metakaolin and silica fume for different concrete mixes and several curing days. Özcan *et al.*, 2009, carried on experimental works and applied artificial neural network and fuzzy logic models to predict the compressive strength of silica fume concrete. They obtained good results using several concrete mixes at five different ages. Yang *et al.*, 2003, applied neural networks to predict compressive strength, slump value and mix portions of concrete. The networks were trained using standard tables of two companies.

Ruan Xiang, 2009, and Zhitao *et al.*, 2008, applied support vector machine (SVM) to predict concrete carbonation. Chen *et al.*, 2009, have shown that the SVM has a good performance for estimating the exposed temperature of fire-damaged concrete structures. Gupta, 2007, applied the SVM to predict the compressive strength of high performance concrete.

3.2 DM algorithms

The DM algorithms used in this study were Regression Trees (RT), Multiple Regressions (MR), Artificial Neural Networks (ANN), Support Vector Machines (SVM) and k-Nearest Neighbours (k-NN).

The Decision Trees (Quinlan, 1986) have an inverted tree structure composed of nodes and descendent branches. The result of a test performed at each node indicates the

branch to continue the process. This process is repeated until the final decision can be drawn and a class is attributed to the register. The regression trees are a particular case of the decision trees where the classes are replaced by values (Figure 1).

The multiple regressions are similar to simple regressions but with several independent variables instead of one independent variable.

The artificial neural networks are based on the architecture of the human brain. They are based on process units named neurons that have a large number of interconnections that allow the communication between them. Each neuron has a set of input and output connections that have an associate weight. The level of activation of a neuron is determined by an activation function (Haykin, 1999 and Cortez, 2010). The neuron receives signal from the input connections and calculates a new activation value which is send through the output connections. The value is a result of the calculation of the value of the neuron activation using the activation function that has as input argument the value of the weighted sum of the input values. In this study it was used the multilayer perceptron configuration (Haykin, 1999) composed by an input layer, a hidden layer and an output layer (Figure 2).

The Support Vector Machines were developed by Cortes and Vapnik, 1995, for binary classification. The basic idea was to separate the dataset in two classes or categories. To do so, a hyperplane in a multidimensional space separates the examples in sets of the same category. The optimal separating hyperplane between the two classes by maximizing the margin between the closest points of the two classes (Meyer, 2010). The points lying in the boundaries are called support vectors and the optimal separating hiperplane is at the middle of the margin. The points lying on the wrong side are weighted down to reduce there influence (Meyer, 2010). When a linear separator cannot

be found there is a transformation via kernel techniques to a higher dimensional space (Figure 3) (Meyer, 2010). The Radial Basis Function kernel was adopted in this study (Cortez, 2010):

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right), \quad \gamma > 0 \quad (1)$$

The performance is affected by a penalty parameters, C, the width of the ϵ -insensitive zone and the kernel parameter, γ (Cortez, 2010).

The k-Nearest Neighbour (Hechenbichler and Schliep, 2004) is a simple supervised learning algorithm that can be used in classification and regression problems. In classification problems an instance query is classified according to its neighbours' classes (Figure 4). The class in majority among the nearest neighbours is attributed to the query instance. In regression problems the property value for the instance query is obtained as the average of the weighted values of the k nearest neighbours. This implies the calculation of the distance between the target and its neighbours in the multidimensional space. Generally, the weights are attributed according to the distance. The closest neighbours are more weighted than the more distant neighbours.

The training process of this study used 2/3 of the total dataset and included the optimization of the parameters involved in the different techniques (H, γ and k). It was used a grid search for the number of hidden nodes H [0, 2, 4, 6, ..., 20], the parameter of the kernel γ [2^{-15} , 2^{-13} , ..., 2^3] and the number of nearest neighbours k [2,3,4,...,12]. To access the predictive performance of the models, the 5-fold cross-validation (Efron, 1993) was used, where the data was divided into 5 partitions of equal size.

The testing process used the remaining part of the total dataset and the best parameters (H, γ and k) of the training phase.

To evaluate the performance of the regression models the coefficient of determination (R^2), the mean absolute deviation (MAD) and the root mean squared error (RMSE) were used. The last two metrics are given by:

$$MAD = \frac{1}{N} \times \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

where N denotes the number of examples, y_i the real value and \hat{y}_i the value estimated by the model.

The complexity of the models imposes not only the analysis of the metrics but also careful interpretation of the results. To help this interpretation it is also important to evaluate the relative importance of the input parameters in the model. Therefore, it is necessary to carry on a sensibility analysis (Kewley *et al.*, 2000). This analysis is applied after the training phase and allows analysing the response of the model when the input parameters are changed from its minimum to its maximum value. During the process each parameter is changed while the others remain with their middle values. When the parameter is relevant it is obtained a high variance at the model output. A higher variance means a greater importance.

4. Materials and concrete compositions

The data used in this study was collected from Kim *et al.*, 1992, Naproux, 1994, and Camões, 2002, where the used mix-designs as other details, namely concerning the production and fresh behaviour of concrete, can be seen.

All the compositions were produced with cement, sand as fine aggregate, coarse aggregate and water. Air entraining agents were not used and some mixtures were made including FA and superplasticizer.

Naproux, 1994, has used 7 different types of FA because it was his intention to study the effect on concrete properties derived from reducing FA particle size. So he used brut FA as received and enhanced ones by grinding or sieving.

Observing Table 1 one can see the chemical composition of the FAs used in the different concrete compositions.

All the compressive strength results were measured on cylindrical specimens with a slenderness ratio of 2.

5. Data mining used data

As already mentioned the compressive strength results were gathered from Kim *et al.*, 1992, and Naproux, 1994, and compiled by Waller *et al.*, 1997. Results from Camões, 2002, were also used. Kim *et al.*, 1992, have performed 26 different concrete compositions and have tested 24 at 6 ages and 2 at 3 ages. Naproux, 1994, have performed 13 concrete compositions and have tested all of them at 3 different ages. Camões, 2002, have performed 11 compositions and have tested all at 6 distinct ages.

The database analysed here was composed of 255 records being 150 extracted from Kim *et al.*, 1992, 39 from Naproux, 1994, and 64 from Camões, 2002. Instead of using only one DM model, one intends to use several models to compare their performances and to include the quantification of the importance of the different components of the concrete mixture, namely the chemical and physical characteristics of FA.

In order to simplified the model one intends minimize the amount of input parameters as possible. Therefore, among the known physical and chemical characteristics of FA described in Table 1, one as chosen for DM input parameters only those known to affect the compressive strength of concrete as mentioned in 2. Furthermore, the content of SO_3 was not taking into account because there is a lack of information on this parameter in Kim *et al.*, 1992.

The data were separated in two sets. The training set composed of 169 records and the testing set composed of 84 records. The input parameters are: FA replacement ratio (% of total cementitious material, TCM, i.e. $TCM = C + FA + SF$); FA characteristics (SiO_2 , Al_2O_3 , Fe_2O_3 , CaO, LOI, in % and Blaine, in m^2/kg); silica fume replacement ratio (SF - % of total cementitious material), total cementitious material (TCM), water/TCM ratio (W/TCM); fine aggregate (ssa), coarse aggregate (ca), high rate water reducing agent (HRWRA - % of solid content related to TCM) and age of samples (Age). The output parameter is the compressive strength of concrete (f_c). Tables 2 and 3 present some statistical data of the parameters used in the training and testing databases, respectively. The coefficient of variation of the parameters for training and testing set is quite similar. This means that the variability is similar in both datasets. The SF has the higher variability while the coarse aggregate (ca) has the lower variability.

6. Results and discussion

With the built database the predictive models were trained to forecast the concrete compressive strength. Tables 4 show the errors and the coefficient of determination

obtained in the training phase. It can be seen that the ANN and the SVM models have the best predictive capacity with lower errors and higher coefficient of determination.

However, it is also important to see if the importance given by the models to the input variables is according to the experience. The importances for all the models are presented in Table 5.

As it was earlier explained the models are based on different algorithms. Therefore, the importance attributed to the input parameters differs from model to model.

It seems that the RT model should not be considered because of the null importance attributed to the majority of the input variables. The MR model also seems to be not applicable since it concentrates almost half of the total importance on only two parameters (SiO_2 and Al_2O_3) and does not consider a reasonable importance to the age (only 3 % of importance).

Considering the three remaining models, ANN model gave a non expectable and an apparently exaggerated importance to CaO content (31 %), only the SVM and k-NN attributes reasonable importance to age and only the SVM model attributes a significant importance to W/TCM ratio (17 %). Furthermore, the k-NN model attribute null influence to two parameters (SiO_2 and SF) and only the SVM model attributes a significant importance to FA content.

Thus, analyzing the results and comparing them with the experience, i.e. with what is expected, it seems that the SVM was the best model for the tested data, demonstrating sensitivity to parameters known to affect the compressive strength.

It is also important to note that, according to the SVM model, the influence of FA characteristic parameters in the compressive strength of concrete proved to be

marginal. Even the LOI or the Blaine reveals little influence (0.8 % and 1.1 % respectively).

Figures 5 and 6 shows the comparison between the measured and predicted concrete compressive strengths for the SVM model using the training and the testing set, respectively. These figures confirm the good predictive capacity of the SVM model.

7. Conclusions

The data mining techniques have the capacity of learning with examples. In this study several data mining techniques were use to predict the 3, 7, 28, 56, 90 and 180 days compressive strength values of concrete mixes containing several percentages of cement replaced by FA. The training phase indicates the ANN and the SVM as the best predictive capacity models. However the importance attributed by the ANN model to the input parameters is not according to the experience. On the contrary, the SVM model has not this shortcoming, because it demonstrates sensitivity to parameters known to affect the compressive strength. This sensitivity associated to predictions very close to the experimental results makes the SVM model the best one.

According to the SVM model, the influence of FA physical and chemical characteristics in the analysed data concerning compressive strength of concrete seems to be marginal.

Based on the analysed data the compressive strength values of concretes containing fly ash can be accurately predicted using SVM without spending much time.

8. Acknowledgments

C-TAC/UM is a research unit (UI4047) of the “Fundação para a Ciência e Tecnologia” (FCT) currently under partial financial support of the strategic project Pest-OE/ECI/UI4047/2011.

References

- ACI Committee 232 (1996). Use of fly ash in concrete. ACI committee report 232.2R-96. ACI, April, pp. 34.
- ACI Committee 363. (1992). State-of-the-art report on high strength concrete. ACI committee report 363R-92. ACI, September, pp. 55.
- Alonso, J. L. & Wesche, K. (1991). Characterization of fly ash. Fly ash in concrete: properties and performance, Report of technical committee 67-FAB – use of fly ash in building (K. Wesche ed.) RILEM, E&FN SPON, pp. 3-23.
- Camões, A. (2002). High performance concrete incorporating fly ash, PhD Thesis, University of Minho, pp. 456.
- Chen, B.-T., Chang, T.-P., Shih, J.-Y. & Wang, J.-J. (2009). Estimation of exposed temperature for fire-damaged concrete using support vector machine. *Computational Materials Science* 44 (3), 913-920.
- Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*. Kluwer Academic Publishers. 20(3): 273-297.
- Cortez, P. (2010). Data mining with neural networks and support vector machines using the R/rminer Tool. *Proceedings of 10th Industrial Conference on Data Mining, Advances in Data Mining, Applications and theoretical aspects.* (P. Perner (ed.)). *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 572-583.

- Day, Ken W. (1995). Concrete mix design, quality control and specification. E&FN SPON, London, pp. 350.
- Efron, B. and Tibshirani, R. (1993). In An Introduction to the Bootstrap. Chapman & Hall.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery: an overview. In Advances in Knowledge Discovery and Data Mining (Fayyad et al. eds.). AAAI Press / The MIT Press, Cambridge MA, 471-493.
- Gupta, S. M. (2007). Support vector machines based modelling of concrete strength. Engineering and Technology, World Academy of Science 36, 305-311.
- Halstead, W. (1986). Use of fly ash in concrete. Transportation research board publication n° 127, Washington, pp. 66.
- Haykin, S. (1999). Neural networks - a comprehensive foundation. 2nd ed. New Jersey: Prentice-Hall.
- Hechenbichler, K. & Schliep, K. (2004). Weighted k-nearest-neighbor techniques and ordinal classification. Discussion Paper 399, SFB 386, Ludwig-Maximilians University Munich. Available from: <http://epub.ub.uni-muenchen.de/archive/00001769/01/paper_399.pdf>. [Accessed 7 August 2010]
- Jalali, S. (1991). A study of factors affecting the mechanical behaviour of lime-fly ash mixtures. Coventry Polytechnic, PhD Theses, February, pp. 411.
- Joshi, R. C. & Lohtia, R. P. (1997). Fly ash in concrete: production, properties and uses, Advances in Concrete Technology (V. M. Malhotra ed.), Volume 2, Gordon and Breach Science Publishers, Ottawa, Ontario, Canada, pp. 269.
- Kasperkiewicz, J., Racz, J. & Dubrawski, A. (1995). HPC strength prediction using artificial neural network. Journal of Computing in Civil Engineering 9(4), 279-284.

- Kewley, R., Embrechts, M. & Brenemam, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks* 11(3), 668-679.
- Kim, J. I. & Kim, D. K. (2002). Application of neural networks for estimation of concrete strength. *KSCE Journal of Civil Engineering* 6 (4), 429-438.
- Kim, J.K., Park, Y.D., Sung, K.Y. & Lee, S.G. (1992). The production of high-strength fly ash in Korea. 4th CANMET/ACI International Conference on Fly Ash, Silica Fume, Slag and Natural Pozzolans in Concrete (V.M. Malhotra ed.), Supplementary papers, Istanbul. May.
- Lai, S. & Serra, M. (1997). Concrete strength prediction by means of neural network. *Construction and Building Materials* 11 (2), 93-98.
- Malhotra, V. M., & Ramezaniapur, A. A. (1994). Fly ash in concrete. 2nd edition, CANMET – Canada Centre for Mineral and Energy Technology, pp. 307.
- Mehta, P. K. (1985). Influence of fly ash characteristics of the strength of Portland fly ash mixtures. *Cement and concrete research*, Elsevier Science Ltd, Volume 15, Number 4, pp. 669-674.
- Meyer, D. (2010). Support Vector Machines. The interface to libsvm in package e1071. Available from: <http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>. [Accessed 7 August 2010].
- Mindess, S. (1994). Materials selection, proportioning and quality control. High performance concretes and applications (S.P. Shah and S.H. Ahmad ed.), Edward Arnold, London, 1-25.

- Naproux, P. (1994). Les microcendres (cendres volantes traitées) et leur emploi dans les bétons hydrauliques. PhD Thesis. Institut National de Sciences Appliquées de Toulouse, France.
- Neville, A. M. (1995). Properties of concrete, Fourth edition, Longman Group Limited, England, pp. 844;
- Ni, H.-G. & Wang, J.-Z. (2000). Prediction of compressive strength of concrete by neural networks. *Cement and Concrete Research* 30, 1245-1250.
- Özcan, F., Atiş, C. D., Karahan, O., Uncuoğlu, E. & Tanyildizi, H. (2009). Comparison of artificial neural network and fuzzy logic models for predicting of long-term compressive strength of silica fume concrete. *Advances in Engineering Software* 40, 856-863.
- Quinlan, J. (1986). Induction of Decision Trees. *Machine Learning*, Kluwer Academic Publishers 1, 81-106.
- R Development Core Team. (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3. Available from: <<http://www.R-project.org>>. [Accessed 7 August 2010].
- Ruan Xiang (2009). Prediction of concrete carbonation depth based on support vector regression. *Third International Symposium on Intelligent Information Technology Application*, iita, vol. 3, 172-175.
- Saridemir, M. (2009). Prediction of compressive strength of concretes containing metakaolin and silica fume by artificial neural networks. *Advances in Engineering Software* 40, 350-355.

- Swamy, R. (1993). Fly ash and slag: standards and specifications – help or hindrance? *Materials and structures / Matériaux et constructions* Vol. 26, 600-613.
- Topçu, I. B. & Saridemir, M. (2008). Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic. *Computation Materials Science* 41, 305-311.
- Waller, V., Naproux, P. & de Larrard, F. (1997). Contribution des fumées de silice et des cendres volantes silico-alumineuses à la résistance en compression du béton. Quantification. *Bulletin des Laboratoires des Ponts et Chaussées*. 208. Mars-Avril – réf. 4098. 53-65.
- Wesche, K. (1991). Fly ash in concrete: properties and performance. Report of technical committee 67-FAB – use of fly ash in building (K. Wesche ed.) RILEM, E&FN SPON, 3-23.
- Yang, D.-S., Park, S.-K. & Lee, J.-H. (2003). A prediction on mix proportion factor and strength of concrete using neural network. *KSCE Journal of Civil Engineering* 7 (5), 525-536.
- Zarandi, M. H. F., Türksen, I. B., Sobhani, J. & Ramezani pour, A. A. (2008). Fuzzy polynomial neural networks for approximation of the compressive strength of concrete. *Applied Soft Computing* 8, 488-498.
- Zhitao, L., Hongming, H. & Shengli, Z. (2008). Research on support vector machine's prediction of concrete carbonization. *International Seminar on Business and Information Management, isbim*, vol. 1, 319-322.

Table captions:

Table 1: FA chemical composition (%) and Blaine specific surface (m^2/kg).

Table 2: Statistical of the input and the output parameters (training set).

Table 3: Statistical of the input and the output parameters (testing set).

Table 4: Mean values of the metrics obtained in the training phase.

Table 5: Importance of the input variables in the evaluation of f_c (%).

Table 1: FA chemical composition (%) and Blaine specific surface (m²/kg).

	Kim et al.	Naproux							Camões
S _i O ₂	55.1	56.7	54.5	55.1	55.8	56.6	57.3	58.1	50.3
Al ₂ O ₃	34.9	27.1	28.2	28.1	27.9	27.3	27.0	26.5	26.8
Fe ₂ O ₃	3.7	5.8	5.9	5.5	5.1	5.8	5.8	6.1	6.3
CaO	3.6	2.9	3.2	3.0	2.8	2.9	2.9	3.0	5.4
MgO	1.2	0.6	0.9	0.8	0.8	0.7	0.6	0.6	1.6
K ₂ O	0.7	1.9	2.3	2.1	2.0	1.8	1.6	1.5	1.9
Na ₂ O	0.4	0.1	0.3	0.2	0.2	0.1	0.1	0.0	0.6
SO ₃	-	0.6	0.3	0.4	0.5	0.6	0.7	0.8	0.6
LOI	6.8	2.5	3.5	3.2	2.9	2.5	2.1	1.8	7.4
Blaine	332	680	667	569	450	310	232	158	388

Table 2: Statistical of the input and the output parameters (training set).

Parameters	Min.	Mean	Max.	Standard Deviation	Coefficient of variation (%)
Inputs					
FA	0	89.12	360	83.43	93.61
S ₁ O ₂	0	51.39	58.10	11.70	22.76
Al ₂ O ₃	0	30.41	34.90	7.76	25.52
Fe ₂ O ₃	0	4.404	6.320	1.53	34.78
CaO	0	3.825	5.425	1.21	31.63
LOI	0	6.203	7.440	1.95	31.38
Blaine	0	341.8	680	103.55	30.29
SF	0	0.5207	44	4.77	916.49
TCM (kg/m ³)	271	460.2	600	91.67	19.92
W/TCM	0.25	0.3979	0.7343	0.11	28.86
ssa (kg/m ³)	223	622.6	788	143.91	23.11
ca (kg/m ³)	936	1057	1256	91.01	8.61
HRWRA	0	0.6674	5	0.99	148.91
Age (days)	3	88.12	365	107.82	122.35
Output					
f _c (MPa)	7.043	47.604	96.8	19.69	41.37

Table 3: Statistical of the input and the output parameters (testing set).

Parameters	Min.	Mean	Max.	Standard Deviation	Coefficient of variation (%)
Inputs					
FA	0	86.8	360	83.93	96.69
SiO ₂	0	51.4	58.1	11.77	22.90
Al ₂ O ₃	0	30.43	34.9	7.80	25.64
Fe ₂ O ₃	0	4.393	6.320	1.53	34.91
CaO	0	3.816	5.425	1.21	31.74
LOI	0	6.196	7.440	1.96	31.57
Blaine	0	341.6	680	104.11	30.48
SF	0	0.5238	44	4.80	916.52
TCM (kg/m ³)	271	461.1	600	92.03	19.96
W/TCM	0.25	0.3973	0.7343	0.12	29.14
ssa (kg/m ³)	223	623.1	788	144.77	23.24
ca (kg/m ³)	936	1056	1256	90.96	8.62
HRWRA	0	0.6663	5	1.00	150.20
Age (days)	3	84.82	365	109.26	128.81
Output					
f _c (MPa)	7.20	47.54	93.60	20.47	43.07

Table 4: Mean values of the metrics obtained in the training phase.

	RT	MR	ANN	SVM	k-NN
MAD	7.486	7.026	4.380	5.610	7.722
RMSE	9.354	8.418	7.414	7.435	9.861
R ²	0.776	0.817	0.867	0.858	0.750

Table 5: Importance of the input variables in the evaluation of f_c (%).

	RT	MR	ANN	SVM	k-NN
FA	0	0.33	1.51	19.49	2.68
SiO ₂	0	22.21	5.19	1.06	0
Al ₂ O ₃	0	31.24	10.64	1.82	16.31
Fe ₂ O ₃	0	0.92	5.07	1.93	4.55
CaO	0	9.41	31.33	1.01	1.17
LOI	0	11.56	4.82	0.79	17.78
Blaine	0	0.7	5.98	1.09	13.98
SF	0	0.09	2.65	0.98	0
TCM	19.92	9.93	1.22	17.16	6.07
W/TCM	0	0	3.58	16.98	9.73
SSA	0	6.15	1.71	0.37	1.49
CA	9.73	1.68	4.59	1.35	6.12
HRWRA	37.16	2.29	14.97	1.05	2.63
Age	33.19	3.49	6.74	34.92	17.49

Figure captions:

Figure 1: Example of a regression tree.

Figure 2: Example of a multilayer perceptron.

Figure 3: Example of a SVM transformation.

Figure 4: Example of k-nearest neighbours.

Figure 5: Performance of the SVM model using the training data set.

Figure 6: Performance of the SVM model using the testing data set.

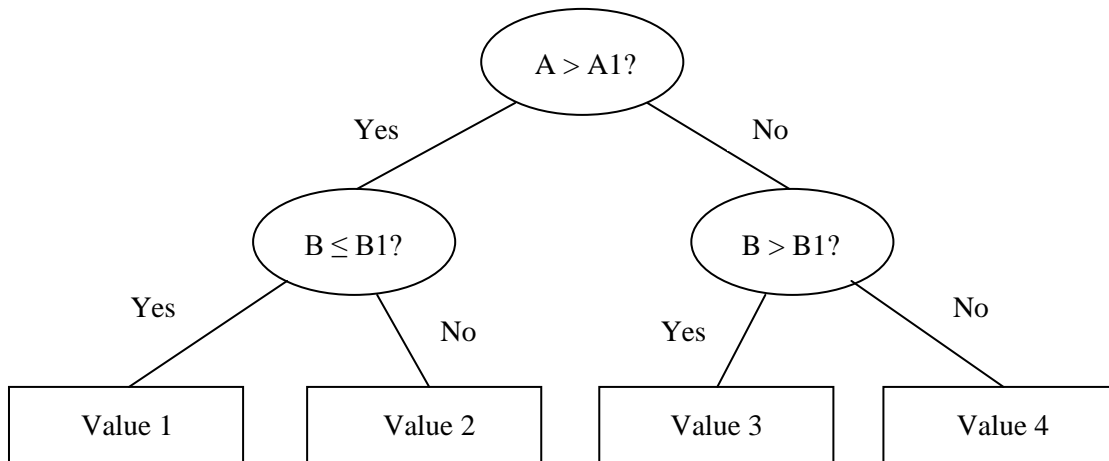


Figure 1: Example of a regression tree.

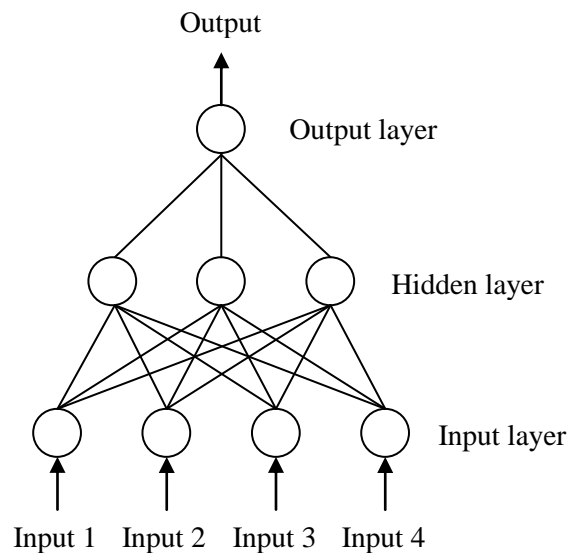


Figure 2: Example of a multilayer perceptron.

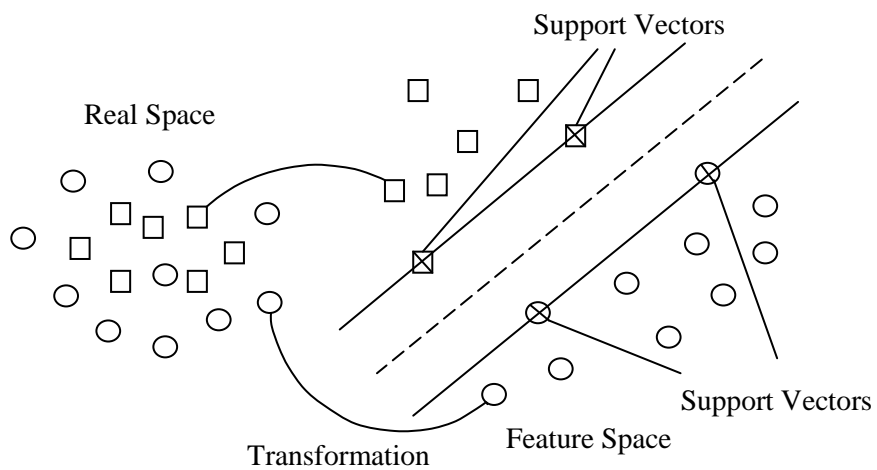


Figure 3: Example of a SVM transformation.

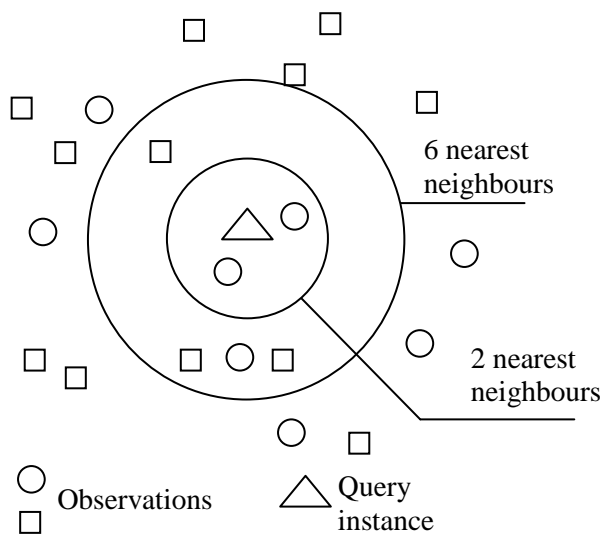


Figure 4: Example of k-nearest neighbours.

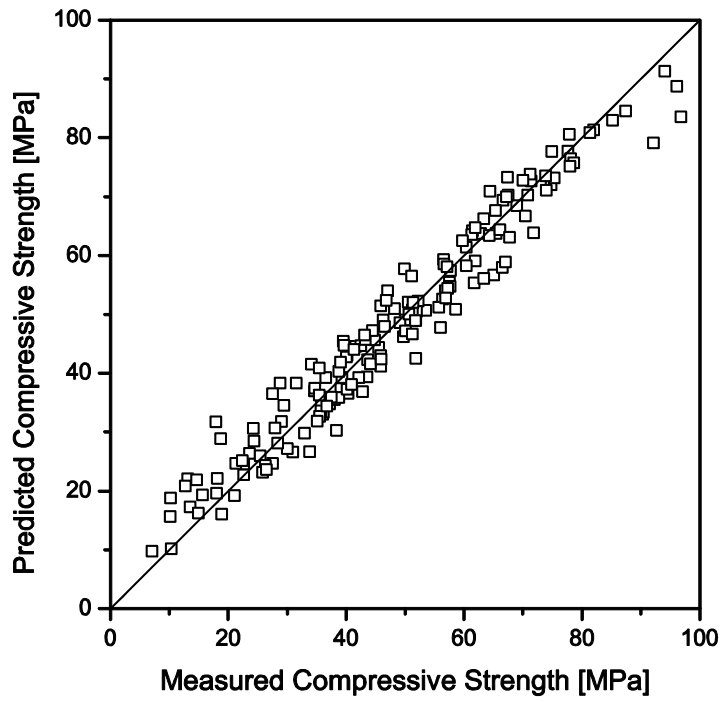


Figure 5: Performance of the SVM model using the training data set.

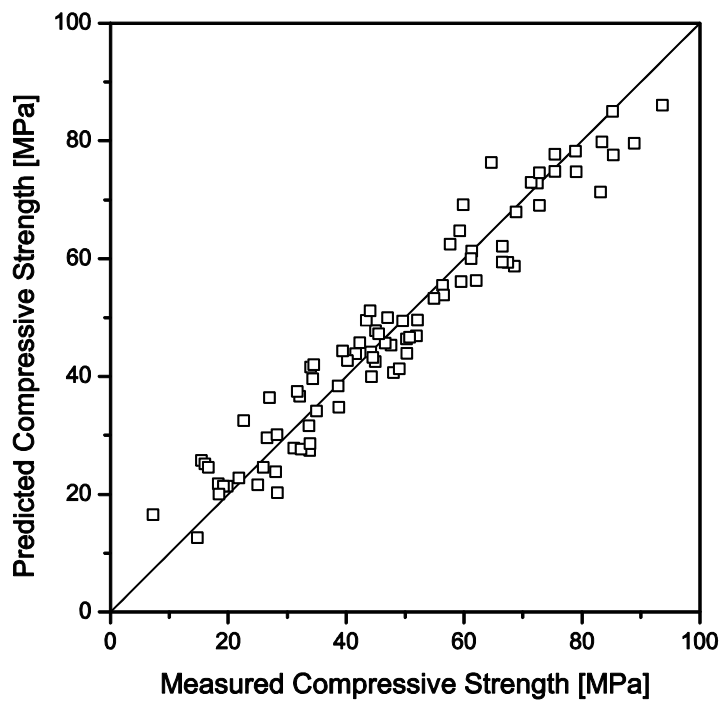


Figure 6: Performance of the SVM model using the testing data set.