Proc. Int. Symp. "Brittle Matrix Composites 9"
A.M. Brandt, J. Olek and I.H. Marshall, eds.
Warsaw, October 25-28, 2009
IFTR and Woodhead Publ., Warsaw 2009

APPLICATION OF MACHINE LEARNING FOR PREDICTION OF CONCRETE RESISTANCE TO MIGRATION OF CHLORIDES

Maria MARKS[1], Daria JÓŹWIAK-NIEDŹWIEDZKA[2], Michał A.GLINICKI[3]

Institute of Fundamental Technological Research, Polish Academy of Sciences
Pawinskiego 5 B, 02-106 Warszawa, Poland
e-mail: [1] mmarks@ippt.gov.pl, [2] djozwiak@ippt.gov.pl, [3] mglinic@ippt.gov.pl

ABSTRACT

The objective of this research was to develop rules for automatic categorization of concrete quality using selected artificial intelligence methods based on machine learning. The range of tested materials included concrete containing non-conventional additive of solid residue from coal combustion in fluidized bed boilers (CFBC fly ash). Performed experimental tests on chloride migration provided data for learning and testing of rules discovered by machine learning techniques. The rules generated by computer programs AQ21 and WEKA using J48 algorithm provided means for adequate categorization of plain concrete and concrete modified with CFBC fly ash as materials of good and acceptable resistance to chloride penetration.

Keywords
Concrete durability, chloride ion migration, circulated fluidized bed combustion fly ash (CFBC fly ash), machine learning

INTRODUCTION

The knowledge of relationships between the composition of concrete, its microstructure and technical properties, including durability in aggressive environments, is a primary objective of research within materials science of concrete. Due to a rapidly increasing number of concrete mix components and its properties, an increased use of complex technologies, a wide range of phase compositions and microstructural features, the simple engineering approach to these relationships might be insufficient.

Modern computation methods that belong to the group of artificial intelligence methods could provide practical support to concrete technology. Kasperkiewicz [1, 2] demonstrated a variety of possibilities of using artificial intelligence methods in civil engineering problems. Three basic concepts are artificial neural networks, machine learning and genetic algorithms. In case of all these approaches the user is not obliged to bother about the model of the process or phenomenon, because the system itself gains the knowledge from examples. It can generate thereupon answers in the form of unknown values of the attributes, classification of new examples of the same format or formulation of rules (hypotheses, generalisations) concerning the process under consideration. More details were given in relation to the applied solutions of Fuzzy ARTMAP and ML program AQ19. Further attempts of using machine learning methods to support phase identification in concrete during indentation testing were reported in [3]. The objective of current research was to develop rules for automatic categorization of concrete quality using machine learning techniques.

In the last decade, due to increased use of clean coal technologies in power generation, new types of coal combustion by-products became available. The composition and physical properties of new types of coal combustion by-products are significantly different than properties of well known fly ashes, widely used in concrete technology. The disposal problems of such non-conventional solid residues from coal combustion are growing, therefore some attempts were undertaken to apply such by-products for cement production or concrete mix production. Solid residues from coal combustion in circulated fluidized bed boilers, called circulated fluidized bed combustion (CFBC) fly ash, are characterized by different mineral and phase composition than conventional fly ash, by angular shape of grains and by the lack of glassy phase. In spite of such differences the research on the concrete strength development in time [4, 5] revealed promising perspectives of using fluidized bed fly ash in structural concrete. However, the durability of structural concrete modified with such an additive is still not well known. Therefore the undertaken research was focused on the resistance of concrete with fluidized bed fly ash to chloride ion aggression. Performed experimental tests on chloride migration provided data for learning and testing of rules discovered by machine learning techniques.

COMPOSITION OF CONCRETE MIXES AND TEST RESULTS OF CHLORIDE MIGRATION COEFFICIENT

The chloride migration coefficient in concrete specimens with different content of fluidized bed fly ash was measured [6]. Ordinary Portland cement CEM I 32.5 R from Małogoszcz cement plant, gravel fractions 2÷8 mm and 8÷16 mm, and sand fraction 0÷2 mm, were used. Two kinds of fluidized fly ash were tested: from hard coal combustion in the thermal-electric power station *Katowice* 'K' and from brown coal - lignite in power plant *Turów* 'T'. Chemical and physical properties of Portland cement type I and CFBC fly ash are shown in Table 1.

Table 1. Chemical composition and physical properties of Portland cement CEM I and fluidized bed fly ash from combustion of hard and brown coal [7]

Chemical compounds	PC type I	fluidized bed fly ash	
		from hard coal K	from brown coal T
SiO_2	21.4	47.18	36.47
Fe_2O_3	3.5	6.8	4.4
Al_2O_3	5.7	25.62	28.4
TiO_2	NA	1.08	3.84
CaO	64.1	5.84	15.95
MgO	2.1	0.15	1.65
SO_3	2.1	3.62	3.8
Na_2O	0.5	1.18	1.64
K_2O	0.92	2.36	0.62
Cl^-	0.029	0.1	0.03
CaO_{free}	0.9	3.4	4.75
Specific gravity [g/cm^3]	3.15	2.68	2.75
Loss on ignition, 1000^{o}C/1h	1.1	3.4	2.73

Three chemical admixtures were used: a plasticizer (magnesium lignosulfonates), a high range water reducer (polycarboxylane ether) and an air-entraining admixture (synthetic surfactants) were used to achieve approximately the same workability and porosity of fresh mix. Three concrete mixes were designed: series B with water to binder ratio w/b = 0.45, and air-entrained series C with w/b = 0.45 and series D with w/b = 0.42. In Table 2 the mixture proportions of tested concretes and the compressive strength of hardened concrete are shown.

Table 2. Composition of concrete mixes and compressive strength tested after 28 days

Concrete mix		Cement	Addition		Aggregate	Water	Plasticizer	HRWR	AEA	f_{c28}
			T	K						
		Content [kg/m^3]								[MPa]
Series B	B0	360	-	-	1859	162	3.2	4.3	-	55.0
	B15K	306	-	54	1854	162	3.2	3.2	-	56.2
	B30K	252	-	108	1847	162	3.2	3.2	-	51.6
	B15T	306	54	-	1850	162	3.2	4.7	-	60.3
	B30T	252	108	-	1841	162	3.2	5.6	-	58.7
Series C	C0	380	-	-	1822	171	3.4	2.7	0.4	46.3
	C15K	323	-	57	1813	171	3.4	2.5	0.6	47.2
	C30K	266	-	114	1803	171	3.4	3.4	0.6	46.8
	C15T	323	57	-	1810	171	3.4	3.8	0.6	45.3
	C30T	266	114	-	1800	171	3.4	4.8	0.6	46.3
Series D	D0	406	-	-	1586	175	-	0.0	3.2	22.7
	D20T	290	73	-	1431	151	-	2.0	2.9	21.0
	D40T	217	145	-	1423	150	-	4.0	5.8	26.1
	D20K	323	-	81	1593	167	-	2.2	3.2	38.3
	D40K	244	-	162	1606	157	-	4.5	6.5	43.0

HRWR- high range water reducer, AEA- air-entraining admixture,
0-no addition, T - fluidized fly ash from brown coal, K - fluidized fly ash from hard coal

The design of concrete mixes was performed according to the experimental method with replacement of cement mass by fluidized fly ash: 15% and 30% in series B and C, 20% and 40% in series D. The specimens were cast in cubical moulds 150 mm and in cylinder moulds ø100 mm x 200 mm. Fresh mixes were consolidated by vibration. After 48 hours the specimens were demoulded and cured in high humidity conditions RH > 90% at temperature 18÷20 ^{0}C until the age of 28 days.

Table 3. Estimation of the chloride resistance to chloride ions penetration, [9]

Diffusion coefficient	Resistance to chloride penetration
< 2 x 10^{-12} m^2/s	Very good
2 – 8 x 10^{-12} m^2/s	Good
8 – 16 x 10^{-12} m^2/s	Acceptable
> 16 x 10^{-12} m^2/s	Unacceptable

The Nordtest Method Build 492 [8] was used to determine the chloride migration coefficient. The principle of the test is that concrete specimen is subjected to external electrical potential applied across it and chloride ions are forced to migrate into concrete. The specimens are then split open and sprayed with silver nitrate solution, which reacts to give white insoluble silver chloride on contact with chloride ions. This provides a possibility to measure the depth to which a sample has been penetrated. The conformity criteria proposed by L. Tang [9] are based on the voltage magnitude, temperature of anolite measured on the beginning and the end of test and the depth of chloride ions penetration are shown in Table 3.

Table 4 presents the results of chloride migration coefficient determined after 28 days of maturity period for concretes series B, C and D. The highest values of D_{nssm} was obtained for concrete without fluidized fly ash replacement, both for non air-entrained and air-entrained. These values are only acceptable when criteria from Table 3 are used.

Table 4. Results of tests of chloride ions penetration, series B, C and D (mean values from 3 specimens)

Series	**Depth of chloride penetration** [mm]	$\boldsymbol{D_{nssm}}$ [x 10^{-12} m^2/s]	**Resistance to chloride penetration**
B0	27.2	15.25	Acceptable
B15K	20.3	8.68	Acceptable
B30K	15.2	4.98	Good
B15T	17.9	6.40	Good
B30T	12.2	3.02	Good
C0	26.3	13.83	Acceptable
C15K	19.0	7.53	Good
C30K	18.7	6.57	Good
C15T	23.1	9.35	Acceptable
C30T	28.2	10.08	Acceptable
D0	23.3	10.60	Acceptable
D20T	22.5	7.83	Good
D40T	21.7	5.69	Good
D20K	19.4	6.19	Good
D40K	14.1	1.58	Very good

MACHINE LEARNING METHODS

Data mining can be defined as the process of discovering patterns in a dataset. By a dataset we mean a *database,* i.e. collection of logically related records. Each record can be called an *example* or *instance* and each one is characterized by the values of a set of predetermined *attributes*. A few different styles of learning appear in data mining applications but the most common is a *classification* [10]. The aim of the classification process is to learn a way of classifying unseen examples based on the knowledge extracted from the provided set of classified examples. In order to extract the knowledge from the provided dataset the attribute set characterizing the examples have to be divided into two groups: the *class* attribute or attributes and the *non-class* attributes. It is obvious that for an unseen examples only non-class attributes are known, therefore the aim of data mining algorithms is to build such a knowledge model that allows predicting the example class membership based only on non-class attributes. The knowledge model is dependent on the way how the classifier is constructed and it can be represented by decision trees (e.g. algorithm C4.5) or classification rules (the AQ algorithms family). Regardless of the representation both types of algorithms create hypotheses.

In order to evaluate the classifier, i.e. to judge the hypotheses generated from the provided *training set* we have to verify the classifier performance on the independent dataset which is called *testing set*. Of course, both the training data and the test data should be representative samples of the underlying problem. The classifier predicts the class of each instance from the test set; if it is correct, that is counted as a success; if not - it is an error. In order to measure the overall performance of the classifier some quantitative analysis should be done.

The example of such a quantitative measure can be *classification accuracy.* This is the number of correct classifications of the instances from the test set divided by the total number of these instances. The measure is expressed as a percentage.

In a multiclass prediction, the result on a test set is often displayed as a two-dimensional *confusion matrix* with a row and column for each class. Each matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. Good results correspond to large numbers down the main diagonal and small, ideally zero, off-diagonal elements. The classification accuracy is the sum of the numbers down the main diagonal divided by the sum of the all numbers in the matrix.

Lets consider what can be done when the amount of data for training and testing is limited. The simplest way is to reserve a certain amount for testing and use the remainder for training. Of course the selection should be done randomly. In practical terms, it is common to hold out one-third of the data for testing and use the remaining two-thirds for training. The main disadvantage of this simple method is a problem that this random selection may be not representative. A more general way to mitigate any bias caused by the particular sample chosen for holdout is to repeat the whole process, training and testing, several times with different random samples. This process is called the *k-fold cross-validation.* In this technique you decide on a fixed number of folds – k. Then the data set U is split into k approximately equal portions ($U = E_1 \cup ... \cup E_k$) [11]. In each iteration i the set E_i is used for testing and the remainder $U \setminus E_i$ is used for training.

Overall classification accuracy is calculated as an average from the classification accuracy for each iteration $\eta(E_i)$, i.e. is defined as:

$$\bar{\eta} = \frac{1}{k}\sum_{i=1}^{k}\eta(E_i). \tag{1}$$

In order to generate rules describing the concrete resistance to chloride penetration many numerical experiments were performed using program AQ21 and algorithm J48 from the WEKA workbench. Algorithm AQ21, invented in the Machine Learning and Inference Laboratory of George Mason University [12], is based on covering approach alike most of the rule-based data mining algorithms. Therefore, the AQ21 algorithm generates subsequent rules until all the examples (sometimes not all) are covered. The idea of adding new rule or new term to existing rule is to include as many instances of the desired class (*positive examples*) as possible and to exclude as many instances of other classes (*negative examples*) as possible.

The second considered algorithm, J48, is available as a part of WEKA workbench, which was developed at the University of Waikato in New Zealand [13]. Algorithm J48 is an implementation of the last publicly available version of an algorithm C4.5 devised by J. Ross Quinlan. Construction of decision trees is based on a simple divide and conquer approach, which is well known in computer science. The main problem is connected with a selection of tests (splits of attributes) which should be placed in the nodes. The test is good if it allows to shorten the way from the root to the leaves representing classes. Decision trees can be converted to classification rules with ease.

SEEKING FOR THE RULES DESCRIBING CHLORIDE PENETRATION AFTER 28 DAYS

Results obtained from AQ21

As the results of the experiments done on specimens of concrete with different content of fluidized fly ash, as shown in tables 2 and 4, the following database consisted of 15 records was introduced. This database can be used to determine the rules describing the concrete resistance to chloride penetration after 28 days. The database with one nominal and six numerical attributes is presented in Table 5.

Table 5. The database

C1	pfT	pfK	W	A_fr	fc28	Resistance
360	0	0	162	2.1	55.0	Acceptable
306	0	54	162	1.8	56.2	Acceptable
252	0	108	162	1.3	51.6	Good
306	54	0	162	1.6	60.3	Good
252	108	0	162	1.6	58.7	Good
380	0	0	171	6.2	46.3	Acceptable
323	0	57	171	6.8	47.2	Good
266	0	114	171	5.8	46.8	Good
323	57	0	171	6.6	45.3	Acceptable
266	114	0	171	6.2	46.3	Acceptable
406	0	0	175	4.9	22.7	Acceptable
290	73	0	151	6.9	21.0	Good
217	145	0	150	7.8	26.1	Good
323	0	81	167	4.6	38.3	Good
244	0	162	157	4.6	43.0	Good

where:

C1 – content of cement, [kg/m^3],
pfT – content of fluidized fly ash from brown coal (power plant *Turów*), [kg/m^3],
pfK – content of fluidized fly ash from hard coal (power station *Katowice*), [kg/m^3],
W – content of water, [kg/m^3],
A_fr – air content in fresh mix, [%],
fc28 – compressive strength of concrete tested after 28 days, [MPa],
resistance – the resistance of concrete to chloride penetration (Acceptable, Good).

The last attribute (resistance) is a nominal one which takes on two possible values (Acceptable, Good). In the considered database to the class [Resistance=Acceptable] belongs 6 examples and to the class [Resistance=Good] belongs 9 examples.

The aim of an experiment is to generate the rules, which allow us to determine concrete resistance to chloride ions penetration. As an training set all the instances from the database were considered. The rules generated by an AQ21 algorithm are presented below:

```
[Resistance=Good]
    # Rule 1
 <-- [pfK>=55] : p=5,n=0,q=0.556
    # Rule 2
 <-- [C1<=258] : p=4,n=0,q=0.444

    # Rule 3
 <-- [pfT>=27 ] [W<=166] : p=4,n=0,q=0.444                    (2)
```

[Resistance=Acceptable]
# Rule 1
<-- [pfK<=55] [A_fr=1.7..6.75] : p=6,n=0,q=1

Rule 2
<-- [pfK<=55] [fc28=44.15..57.45] : p=5,n=0,q=0.833

where *p* denotes the number of positive examples covered by the rule, *n* denotes the number of negative examples covered by the rule (i.e. the number of records from the other classes satisfying the rule) and *q* denotes the quality of the rule.

The rules showed in (2) can be interpreted as follows (it should be underlined that the presented rules concern concretes with the overall mass of cement and additions equal 360, 380 or 406 [kg/m^3] (Table 2)).

[Resistance is **Good**]
IF
[pfK >= 55]
OR
[C1 <= 258]
OR
[pfT >= 27] i [W <=166]

[Resistance is **Acceptable**]
IF
[pfK <= 55] i [A_fr = 1.7..6.75]
OR
[pfK <= 55] i [fc28 = 44.15..57.45]

The rules verified on a training set show the 100% classification accuracy, what is illustrated by an confusion matrix:

	Acceptable	Good
Acceptable	6	0
Good	0	9

To predict the performance of a classifier on new data, we need to assess its classification accuracy on a dataset that played no part in the formation of a classifier – the test set. In order to estimate the performance many numerical experiments where performed both with static set holdout and cross validation.

In the first experiment the dataset was divided into training set consisting of two-thirds randomly selected instances and the testing set consisting of remainder instances. The following rules were generated by an AQ21 algorithm:

[Resistance=Good]
# Rule 1
<-- [pfK>=55] : p=4,n=0,q=0.667

Rule 2
<-- [C1<=306] [pfT<=93] [fc28<=53.9] : p=4,n=0,q=0.667

Rule 3

```
  <-- [C1<=258] : p=3,n=0,q=0.5
[Resistance=Acceptable]
     # Rule 1
  <-- [C1>=298] [pfK<=55] : p=3,n=0,q=0.75

     # Rule 2
  <-- [fc28=44.15..46.55] : p=2,n=0,q=0.5
```

The rules generated on a dataset with ten examples taken from all the series (3 from series B, 4 from series C i 3 from series D) and verified on a test set can be described by an confusion matrix:

	Acceptable	Good
Acceptable	2	0
Good	1	2

so the classification accuracy is 80%.

In order to estimate the classification accuracy the *k*-fold cross validation was also used. Assuming k=3 the classification accuracy obtained for each iteration is equal respectively 60%, 60% i 80%, so the overall classification accuracy is equal 66.7% (1). When the database consists of a very small number of records (less than 100) [11] the suggested value of parameter k is just the number of examples. Assuming k=15 we obtained the classification accuracy equal 53.3%.

Results obtained from J48

In order to generate the rules, which allow us to determine concrete resistance to chloride ions penetration the J48 algorithm was also used. As an training set all the instances from the database (Table 5) were considered (the same training set was used in experiment described in section 4.1). The decision tree generated by an J48 algorithm is presented below:

```
C1 <= 323
  pfK <= 54
    W <= 162: Good (5.0/1.0)
    W > 162: Acceptable (2.0)
  pfK > 54: Good (5.0)
C1 > 323: Acceptable (3.0),
```

where the first number in brackets denotes the number of examples from the training set covered by a selected leaf, and the second number – just after the sign "/" – indicates the number of incorrectly classified instances (negative examples). When there is only one number in brackets then it indicates the number of examples correctly classified (positive examples).

The obtained decision tree can be easily transformed into the following rules:

```
[Resistance=Good]
Rule1  [C1 <= 323] i [pfK <= 54] i [W <= 162]
Rule2  [C1 <= 323] i [pfK > 54]
                                                        (3)
[Resistance=Acceptable]
Rule1  [C1 <= 323] i [pfK <= 54] i [W > 162]
Rule2  [C1 > 323]
```

The classification accuracy for the rules verified on a training set is illustrated by an confusion matrix:

	Acceptable	Good
Acceptable	5	1
Good	0	9

As one can see one example from Acceptable class is classified incorrectly to Good class, the remaining 14 examples are classified correctly, so the classification accuracy is equal 93.3%.

When we held out one-third of the data for testing and used the remaining two-thirds for training (the same division as in section 4.1) we obtained the following results:

pfK <= 54
  W <= 162: Good (3.0/1.0)
  W > 162: Acceptable (3.0)
pfK > 54: Good (4.0)

i.e.
[Resistance=Good]
Rule1 [pfK <= 54] i [W <= 162]
Rule2 [pfK > 54]

[Resistance=Acceptable]
Rule1 [pfK <= 54] i [W > 162]

The rules generated on a training set and verified on a test set can be described by an confusion matrix:

	Acceptable	Good
Acceptable	1	1
Good	0	3

here also one example from Acceptable class is classified incorrectly to Good class. The classification accuracy is exactly the same as in experiment with static holdout from section 4.1 and is equal 80%.

When we used a *k*-fold cross validation for J48 algorithm the following results were obtained respectively for *k* equal 3 and 15:
for *k*=3

	Acceptable	Good
Acceptable	2	4
Good	1	8

so the classification accuracy is 66.7%,
for *k*=15

	Acceptable	Good
Acceptable	2	4
Good	2	7

so the classification accuracy is 60%,

CONCLUSIONS

The rules generated by computer programs AQ21 and WEKA using J48 algorithm provided means for automatic categorization of plain concrete and concrete modified with CFBC fly ash as materials of good and acceptable resistance to chloride penetration. Due to a small number of tested specimens the rules are applicable only to concrete mix composition of similar binder content and similar values of water to cement ratio.

Application of AQ21 and WEKA programs resulted in similar estimation of concrete resistance to chloride ion penetration. Further tests are needed for enlargement of experimental data basis.

REFERENCES

[1] Kasperkiewicz J., On possibilities of application of artificial inteligence methods in civil engineering (in Polish), Drogi i Mosty, nr 3, 2004, 15-36

[2] Kasperkiewicz J., The applications of ANNs in certain materials-analysis problems. Journal of Materials Processing Technology, 106 (2000), 74-79

[3] Kasperkiewicz J., Marks M., Wołowicz J., Possibilities of acoustic identification of phases in concrete (in Polish), LIV Konferencja Naukowa KILiW PAN i KN PZITB, Krynica 2008, Tom V, 325-332

[4] Roszczynialski W., Nocuń-Wczelik W., Gawlicki M., Małolepszy J., Fly ashes from fluidized bed coal combustion as a cement additive, (in Polish) Ceramics 66, 2001, Polish Ceramic Bulletin, 370-377

[5] Glinicki M.A., Ładyżyński K., The influence of addition of activated fly ash from fluidized bed combustion on properties of structural concrete (in Polish), VIII International Conference „Ashes from power generation", UPS, Miedzyzdroje 2001, 119-133

[6] Jóźwiak-Niedźwiedzka D., Effect of fluidized bed combustion fly ash on the chloride resistance and scaling resistance of concrete, Concrete in Aggressive Aqueous Environments, Performance, Testing and Modeling', 02-05 June 2009, Toulouse, France, RILEM proceedings PRO 63, vol. 2, 556-563

[7] Łagosz A., Małolepszy J, Śliwiński J., Tracz T., Utilization of fly ash from fluidized bed boilers as a mineral additive for concretes, (in Polish), V Konferencja „Dni Betonu – Tradycja i Nowoczesność", Wisła, 2008, 719-727

[8] Nordtest Method NT Build 492. Concrete. mortar and cement-based repair materials: Chloride migration coefficient from non-steady-state migration experiments. 1999

[9] Tang L., Chloride transport in concrete – Measurement and prediction, Publication P-96:6, Chalmers University of Technology, Department of Building Materials, Göteborg 1996

[10] Cichosz P., Learning systems (in Polish), WNT, Warszawa 2000

[11] Krawiec K., Stefanowski J., Machine learning and neural networks (in Polish), Wydawnictwo Politechniki Poznańskiej, 2003

[12] Wojtusiak J., AQ21 user's guide, George Mason University, MLI 04-3, September 2004

Witten I.H., Frank E., Data mining. Practical machine learning tools and techniques, Elsevier, 2005