Computer Assisted Mechanics and Engineering Sciences, 14: 197–207, 2007.

Holistic approach to diagnostics of engineering materials

Janusz Kasperkiewicz, Dariusz Alterman
Institute of Fundamental Technological Research, Polish Academy of Sciences
ul. Świętokrzyska 21, 00-049 Warsaw, Poland

(Received in the final form April 6, 2007)

A concept is presented of a system for automatic processing of the civil engineering data. It may concern designing, optimisation or diagnostics of constructional materials. The main point of interest was concrete and various concrete like composite materials. The applied methods are a combination of various soft computing techniques, like artificial neural networks, machine learning and certain techniques originating in statistics. The system is aimed at taking advantage of varied information available in publications, reports, monographs and direct experimental results, perhaps including even the grey information resources. After preparation of a database collected from laboratory or in-situ observations concerning the behaviour of various concrete materials, and gathered during the two last decades, a number of experiments were performed on the system dedicated mainly to prediction of compressive strength and frost resistance of concrete. The proposed approach allows more efficient control of information in problems of concrete technology.

1. Introduction

Typically, in civil engineering different soft computing methods for extracting knowledge from databases are applied separately. The observation concerns most probably also other fields of engineering. The soft computing methods are rather rarely used even as a so called *engine* inside a larger expert system. The approach proposed in what follows is a combined application of several such methods, aimed at estimation, prediction, design and/or optimisation of civil engineering composite materials, like concrete, fibrous concrete, etc.

Taken into account were different soft computing techniques: artificial neural networks, (ANNs), machine learning, (ML), and certain techniques generally related to statistics — not easy to classify precisely, but functionally close to artificial intelligence. Close to artificial intelligence are for example such techniques as rough sets, fuzzy sets, principal components analysis, correspondence analysis, cluster analysis, and — in certain problems — even evolutionary algorithms. Most of these concepts have been tried in the study, but only a few selected are mentioned in the paper.

There is a growing need of new techniques for procuring knowledge from vast domain of information resources of various origin, including also the so called grey information. There are many analysis tools available but they are never organised into a kind of a system. This concerns even quite large projects like recent WEKA [17], that offers to the user more than 70 different classifiers and various tools for data pre-processing.

All the procured information has to be initially evaluated, pre-processed and formatted. Its rapid analysis and evaluation is needed because of the recent progress in Civil Engineering materials and technologies [4], which results in a lack of full knowledge on the results of application of newest components, technologies and new methods of testing.

To a large extent soft computing or artificial intelligence methods are scarcely supported by formal mathematical proofs, which could theoretically justify natural expectations that a given procedure converges to the right solution. Typically a common method of evaluation of such methods, as well as a method for selecting and tuning the control parameters, is by experimenting. There is an argument for application of all the artificial intelligence methods that the soft computing approach has a higher tolerance for imprecision, uncertainty, partial truth or missing data, and in this way comes closer to the reality of experimental observations. This is in contrast to the possibilities of traditional computing, which can be described as being *hard*, with its algorithms strongly sensitive to various data deficiencies.

Present application of a rather simple implementation of the proposed approach resulted in a number of conclusions of practical importance, concerning prediction of strength and durability of hardened concrete.

2. Databases and data preparation

Even if the humans think usually in terms of qualitative information, (nice, good, evil, tasty), in Civil Engineering most typical data studied in materials technology are quantitative. The reason is that these data result mainly from various physical or chemo-physical measurements. But these information have a characteristic scatter and almost always are represented by numbers loaded with uncertainty, so that, naturally, the corresponding knowledge hidden in them is usually only approximate.

The total of the all human observations taken together represent a knowledge base, (KB). The experimental results selected and extracted from KB and pre-processed in appropriate way will be referred to as a database, (DB). Database is composed of records and the records are composed of different types of attributes, (components or variables).

Such databases in which the attributes are only numbers can relatively easy be processed with various mathematical algorithms, from popular statistical programs to more sophisticated soft computing tools, e.g. Artificial Neural Networks, (ANNs), also applied in the system discussed in the present paper.

In case of most ANNs types, (i.e. particular architectures of ANNs), acceptable as an input are only databases composed of complete records. It means that each record must have all the attribute values specified, and there must be no *voids* in the database. If certain records are defective or have blank attribute values they must be either reconstructed or excluded from further processing. The reconstruction of the data can sometimes be done by analysis of the remaining, complete records from the same database. The simplest solution is by taking as the missing values of an attribute the mean of the values of the same attribute in all the remaining complete records. Such approach, however, is correct only in exceptional situations. In general, when applying ANNs, it is usually assumed that the observations are correct and credible, and the network will operate only on well-defined scalars, vectors or matrices.

Apart from the problems with uncertain or defective data there is another problem connected with processing of the qualitative data, (i.e. the data that are immeasurable). This happens often in engineering databases, which cannot be presented as numerical matrices, being rather tables of numbers and of symbols, (alphanumeric strings). When trying to work on such a database applying for example a BP network, (BP for *Back Propagation*, but the same would be in case of any feed forward type ANN), the qualitative attributes must be coded into numbers. Such procedure is risky, because — generally speaking — there is no rational suggestion for any particular coding. Any choice of the numbering of nominal attributes may result in underestimation of certain attributes, (or in overestimation of other attributes), with the user most often completely ignorant of the consequences of a given choice. The only approach that seems free from such a drawback is to introduce additional variables (attributes) of Boolean type, one for each legal value of each nominal attribute. Of course the resulting database will have much larger volume, with a multitude of attributes in each record. With this solution the processing tool, for example one of available ANNs systems, has to be additionally equipped with a mechanism of checking the consistency of the prediction results. The check must guarantee that all new attributes may take the value of 1, (one, corresponding to the state "on"), only exclusively: two new attributes originating from a single old attribute must not

Table 1. An example of a definition of a structure of a database. The database concerns the properties of HPC materials (HPC — High Performance Concrete); irrelevant fields of the table are shaded in grey

No	Attribute	Attribute role	Units	Symbol	Attribute type	Quantitative data: Min	Quantitative data: Max	No of cases	Qualitative data – list of legal values
1	Label	info	[–]	No	label				
2	Cement content	input	[kg/m^3]	C	con	42	1009		
3	Water content	input	[kg/m^3]	W	con	61	630		
4	Silica fume content	input	[kg/m^3]	SF	con	0	297.8		
5	Air-entrainment content	input	[kg/m^3]	AE	con	0	5.4		
6	Superplasticiser content	input	[kg/m^3]	SP	con	0	48		
7	Fine aggregate content	input	[kg/m^3]	FA	con	0	1977		
8	Lightweight aggregate cont.	input	[kg/m^3]	FA_L	con	0	1035		
9	Coarse aggregate content	input	[kg/m^3]	CA	con	0	2190		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	Fibre type	input	[–]	fibtype	nom			6	{no, steelH, steelS, steelX, glass, carbon, polprop, other}
20	Fibre quantity	input	[kg/m^3]	fibquota	con	0	990		
21	Aggregate type	input	[–]	aggrtype	nom			7	{unknown, granite, basalt, gravel, mixed, pos_granit, pos_basalt}
22	Cement provider	input	[–]	cemprov	nom			4	{nowiny, gorazdze, malogoszcz, unknownn}
23	Additive type	input	[–]	addtype	nom			7	{no, si, pfa, metka, slag, pos_si, pos_pfa}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
39	Modulus of elasticity	output	[GPa]	E	lin	...	...	...	...
40	Slump	output	[mm]	slump	con	0	280		
41	Strength fc28	output	[MPa]	fc28	lin	10	120		
42	Frost resistance	output	[–]	fr_resist	nom			3	{class1,class2,class3}
43	Data source	source	[–]	ref	label				

Table 2. A fragment of the database of the structure as in Table 1. The row at the top specifies the symbols of the attributes

No	C	W	SF	AE	SP	FA	FA_L	CA	...	fib type	fib quota	aggr type	cem prov	add type	...	E	slump	fc28	fr_resist	ref
94	152	119	0	0.09	1.6	643	0	1197	...	steelH	10	gravel	gorazdze	no	...	?	145	38	class1	**C1**
1702	400	160	0	0.12	4	516	0	1393	...	#N/A	0	granite	unknown	no	...	?	25	53	class3	**F**
2	339	132	0	1.5	0	1111	0	758	...	?	?	granite	unknown	?	...	?	?	37	class1	**B12**
622	430	168	?	?	?	?	?	?	...	?	?	?	?	?	...	?	?	51	class2	**D3**
1700	400	160	0	0	2	516	0	1393	...	#N/A	0	basalt	nowiny	no	...	?	?	53	class3	**F**
107	361	141	40	0	20.2	725	0	1205	...	#N/A	0	unknown	unknown	si	...	44	75	71	class1	**G**
109	320	139	80	0	29.1	716	0	1191	...	#N/A	0	unknown	unknown	si	...	43	60	74	class1	**A**
1701	400	160	0	0	4	516	0	1393	...	#N/A	0	basalt	nowiny	no	...	?	20	57	class3	**F**
108	341	140	60	0	23.8	719	0	1196	...	#N/A	0	unknown	unknown	si	...	43	60	75	class1	**H**
623	430	168	?	?	?	?	?	?	...	?	?	?	?	?	...	?	?	51	class3	**E**
93	153	120	0	0.44	5.2	639	0	1192	...	steelH	10	gravel	gorazdze	no	...	?	185	34	class1	**B**
1695	360	165	0	5.4	0	659	0	1233	...	#N/A	0	basalt	nowiny	no	...	53	170	49	class3	**F**
105	151	117	0	0.17	3.1	638	0	1187	...	#N/A	0	unknown	unknown	no	...	?	174	43	class1	**E**
106	401	141	0	0	12.3	729	0	1211	...	#N/A	0	unknown	unknown	no	...	43	75	56	class3	**F**
1699	400	160	0	0	2	516	0	1393	...	#N/A	0	basalt	nowiny	no	...	?	130	50	class3	**F**
1696	382	194	0	1.5	0	769	0	993	...	#N/A	0	basalt	nowiny	no	...	46	120	53	class3	**F**
1707	400	111	40	0	8.8	734	66	875	...	#N/A	0	unknown	unknown	si	...	47	100	74	class1	**S7**
1705	400	140	40	0	8.8	967	0	0	...	#N/A	0	granite	unknown	si	...	53	50	92	class1	**D**
95	153	120	0	0.05	2.2	648	0	1206	...	steelH	10	gravel	gorazdze	no	...	?	113	38	class2	**E**
1706	400	97	40	0	8.8	617	99	875	...	#N/A	0	unknown	unknown	si	...	46	70	79	class1	**S7**
3	321	128	0	1.5	0	1188	0	702	...	?	?	granite	unknown	?	...	?	?	48	class1	**B13**

appear simultaneously as both being "on", (that is both being equal to 1). The whole procedure may then become very computationally expensive.

An example of the structure of a complex database is presented in Table 1. A fragment of the database with this structure can be seen in Table 2. In the table symbols #N/A mean: "not applicable", the question marks, (?), mean lack of information, and the additional **ref** column contains certain codes for respective references, (data sources).

The database as in the example above was collected from available reports and publications or from own laboratory experiments [3]. There is no standard concerning how such data should be formatted and described for civil engineering purposes, as documents aimed at unification of data are scarce, (an example is the recommendation from 1999 of the American Concrete Institute [12]). It is unavoidable that the database to be used for practical predictions purposes will have certain degree of ambiguity.

A database like the one presented in Table 2 can usually be treated as a "raw" or general database, which will be a source for its subset: a certain "working" database. To characterize such working database will only be possible after the aims and scope of the whole data processing is decided.

An important part of the database preparation may be generation and selection of so called *derived* variables. This however can be done only in the later stage of the procedure, after the quality of the "natural" data has been evaluated. In a way, this stage of the processing may be described as the database *post-preparation*.

3. Soft computing tools and procedures

Typically, artificial neural networks operate on numerical data. Such is the case of various feed forward, back propagation networks, build of external and hidden layers of neurons with appropriate, adaptable weights. All the weighted connections, like feedforward, feedback, lateral, or time-delayed connections are fed with numbers, and the data to be analysed by the network must be presented in proper form, i.e. in the form of numbers. In general case also images, sounds and other signals can be analysed by the network but these must first be coded into strings of digits.

Most of the ANNs solutions have the limitation as described above, but not all of them. For example the architectures of ANNs are from certain point of view different when they are based on the idea of SOM — Self-Organizing Maps, (unsupervised learning). In this case, even if the network also handles the numbers, an idea of so called a *concept* is employed. In these solutions, (and there is similarity for example in case of Hopfield networks), the system retains memory of selected states, and is able to store and retrieve patterns, defined as certain combinations of the attributes. For example in case of IAC network, (*Interactive Activation and Competition*) [6], it is not only possible to analyse data with nominal attributes, but also to work on incomplete data. Such system can be therefore applied even in classification of defective records.

The pre-processing of Civil Engineering data, like in case of any other data types, involves identification of the outliers. It should be followed by their evaluation, because in this case the apparent outliers not necessarily are a false information.

An important step is introduction of derived attributes, and — when needed — transformation of the qualitative into quantitative data: it may be necessary to replace all the nominal and structural attributes by linear or continuous attributes, (real numbers). Another very important element of the data pre-processing is clustering and reordering of the records. These actions seem especially promising

Finally, sometimes a normalization of the attributes is needed (e.g. by transforming domains into a unit multidimensional cube). In professional programs, however, appropriate routines are often already built-in, so they are applied in automatic way.

In principle, feed-forward ANNs can compute any computable function, i.e., they can do everything a normal digital computer can do. The resulting data resemble a continuous function representation, which in certain situations may however be a drawback. When the reality of the problem corresponds to a distinctly discontinuous function, the network should rather refrain from predictions by interpolation. This would be e.g. the case of a process represented by a combination of step functions and separate peaks. Regression analysis and many ANNs often produce interpolated faulty, incorrect answers. A possibility of the system to refuse such answers is rare and exists for example in case of the Fuzzy ARTMAP, applied in the system discussed in this paper. Which of various possible ANNs architectures will be most appropriate depends however on the actual data structure.

Most of the soft computing tools tried in the present investigations are available for experimenting via Internet [1, 2, 7, 13–16]. Algorithms finally used to process the data were those of BP, Fuzzy ARTMAP and aiNet artificial neural networks, AQ19 and See5 machine learning programs, GradeStat, and certain other statistical procedures. The ANNs solutions are rather well known and applied relatively often, but the popularity of particular algorithms is quite unequal. Machine Learning (ML) is less popular in problems of engineering materials analysis. ML programs work by propositional learning, discovering rules, formulated as sentences composed of so called selectors, (the simple conditions). At some ML solutions the search of the rules goes in form of analysis of decision trees.

Both ANNs and ML solutions can be supported by the instruments from other fields of computational sciences, like statistics or evolutionary computing. The last possibility has not been employed in the system. Also the possibilities of solutions like PCA, (*Principal Components Analysis*) [16], or rough sets approach [13], which are promising, have not been used in the present system.

4. The data processing system

A very general layout of a system for data processing is presented in Fig. 1. The main parts of the system are blocks of data preparation (a), of data processing and training (b), and of system exploitation (c). The proposed system is rather big and complex, and presentation of the details of the scheme is not possible in a short paper, so the details as proposed in [3], are not displayed in Fig. 1. Only as an example, a small sub-block of the system is shown in Fig. 2.

The system can be referred to as *holistic*, because the intention was both - to apply different soft computing tools and to enable processing a vast range of dataset formats. The whole system and also its components represent a hybrid approach to data recognition and analysis. It can be seen in Fig. 2 that when trying to repair a faulty dataset it is possible to select among several different solutions. The user can either generalise directly the available results, applying ANNs or various statistical programs, or to look for engineering formulas that can be found in recommendations or monographs, (empirical formulas). The user can also refer to certain knowledge, (non-disclosed), hidden in design and simulation computer programs produced by other investigators.

In the exploitation block of the system there are mainly two actions expected, (cf. Fig.1), which are *data prediction* and *data classification*. The same functions, however, may also be activated as a part of the data preparation procedures, used to evaluate and select certain groups of records and to eliminate and/or replenish defective records.

Which one of the two above actions is to be applied is decided as an optional decision of the user at the task definition stage, (point g in Fig.1), anyhow the data can be collected, pre-processed and processed completely independent of the task definition. In this way, which is easy with the present electronic information storage systems, a databank can be created *in advance*, as a *raw* database, even without precise definition of the task, for future exploitation referral.

If the data to be analysed is defective, and is lacking certain values of numerical attributes, the respective records could be repaired by predicting the attributes using an ANN trained with data from the remaining complete records. In such case the rules can also be searched for by ML, (Machine Learning), applying one of several possible approaches. Of course there are limits in similar repair procedures related to the proportion in the dataset of defective records, (or attributes), and their

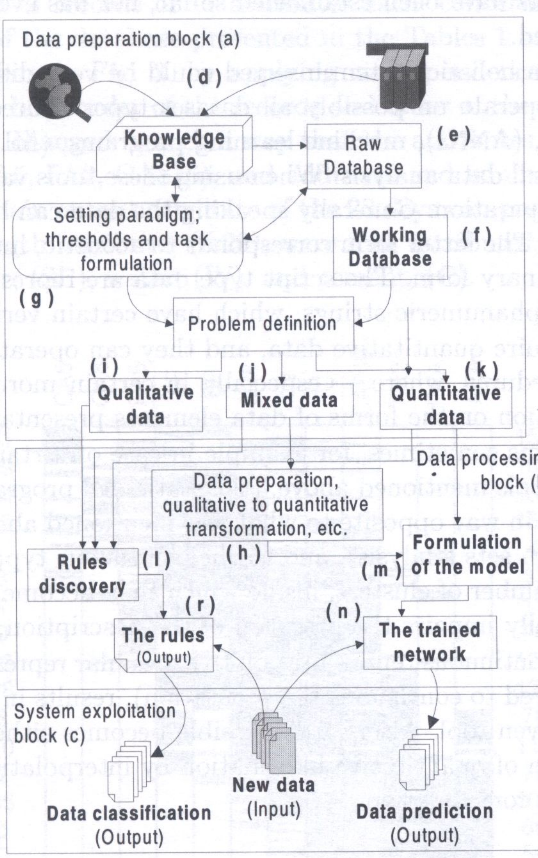

Fig. 1. General scheme of the system

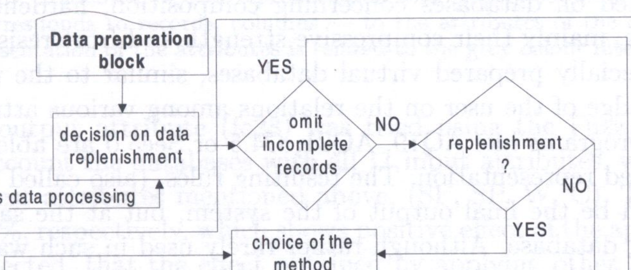

Fig. 2. An example of the sub-block from the system. Such segments are elements of sub-blocks 'l' and 'm' in Fig. 1

uniformity. Neither such limits have been established so far, nor has even an appropriate approach to the problem been suggested.

It was assumed that into a holistic system inserted could be very different soft computing tools and that it will be able to operate on possibly all dataset types. Three basic types of such tools are artificial neural networks, (ANNs), machine learning programs, (ML), and a number of special algorithms related to statistical data analysis. When using these tools various solutions set different requirements for the data preparation. Generally speaking the data can be prepared as a script type data or as a signal type data. The latter form corresponds to acoustic, image or video signals, which usually are submitted in a binary form. The script type data are represented by series of numbers and/or symbols in form of alphanumeric strings, which have certain verified practical meaning.

Many ANNs solutions require quantitative data, and they can operate only on numbers. This is different in case of ML procedures, where — especially in certain more advanced applications — there is practically no limitation on the forms of data elements presentation.

However, there are problems sometimes, for example in case of certain programs from the third group of the computational tools mentioned above, (the statistical programs). There are procedures that require data preparation in way opposite to what was mentioned above. They work on granular data, (e.g. in the case of rough sets analysis), and so the continuous type attributes must be transformed, (projected), into a number of clusters, inside which they become indistinguishable. Yet, the transformation, which naturally impairs the precision of the description, is a relatively easy task.

Transformation from the continuous representation to a granular representation, (or — depending on needs — vice versa: clustered to continuous representation), results in datasets more appropriate for further processing by a given tool. Afterwards possible becomes either the prediction of magnitudes, (like in, e.g., evaluation of quantitative information by interpolation), or generation of rules, concerning the dataset under consideration.

5. Examples of obtained results

A number of experiments aimed at different predictions and improvement of the quality of predictions were conducted on databases concerning composition, hardened material structure and properties of concretes, mainly their compressive strength and frost resistance. Some data processing concerned also specially prepared virtual databases, similar to the real ones described above, but with a full knowledge of the user on the relations among various attributes.

Machine learning programs like AQ19, AQ21, C4.5 or See5.0 are able to analyse data of a general, almost unrestricted representation. The resulting rules, (also called hypotheses), of a precisely defined credibility, can be the final output of the system, but at the same time can also be used to modify the working database. Although rather rarely used in such way, discrete but really fine-grained data may enable also certain quantitative prediction of the results.

An important element of the system is possibility of re-ordering and clustering of the data. Re-ordering may be useful in preparation of the datasets for ML processing. This is so because in certain situations different rules may be formulated depending on which group of records were submitted first to the system.

The clustering allows identification of data subsets for which certain relations can be searched for, different from those in other subsets [9]. For the above purposes applied can be algorithms like nearest neighbour clustering available in [15], or the correspondence analysis [7, 11].

The main experiments were done on the datasets similar to those described above, (cf. Tables 1 and 2).

A typical procedure was to start with identification of the defects in the dataset and their repair. Then the data have been reordered and clustered, taking into account possibilities of GradeStat and other statistical tools, and finally they were used for training and prediction, or for rules generation.

The results from the first action were positive, and similar to those described in previous papers, like e.g. in [8, 10]. An example of the rules found by the system is as follows.

The database composed of records of 11 attributes each, concerning HPC, (High Performance Concrete; it was a subset of the database presented in the Tables 1 and 2), was analysed using the GradeStat. As it can be seen in Fig. 3, by applying correspondence analysis the dataset can be reordered in what concerns both order of the records and order of the attributes. The original order of the attributes have been changed in such a way that the attributes concerning the description of aggregates, (CA8, CA6, FA5, FA2, CA816), and additives and admixtures, (PFA and AEA), have been located far from the main output attribute of the 28 days compressive strength of the concrete, (fc28). Closest to the strength attribute, (fc28), were four attributes, concerning contents of silica fume, (SF), superplasticiser, (SP), water, (W), and cement, (C).

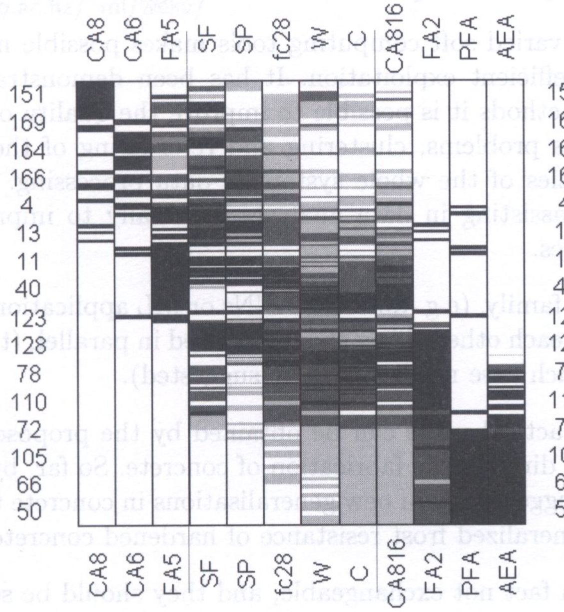

Fig. 3. A map, produced by GradeStat, of over-representation with respect to uniform distribution of HPC database records. Rows corresponds to records, columns — to the attributes of the dataset. The degree of over-representation of the attributes is related to the grey colour intensity

The prediction of the output attribute (fc28) was tried using the Fuzzy ARTMAP program, (*Beton.exe*), taking into account the databases with all 11 input attributes, with 8 attributes, or — finally — with only 4 closest attributes mentioned above, (SF, SP, W, C). The average prediction error was 19%, 14% and 11%, respectively, which shows positive effect of the applied data elimination of the attributes. It is expected, that the effect obtained by applying other ANNs solutions would be similar.

An example of what concerns another effect of the system exploitation — the rule generation, is related to evaluation of the frost resistance of hardened concrete. This property can be evaluated according to a number of national standards [5], differing in the object of testing. The test can be aimed either at the internal destruction of concrete or at the concrete surface scaling.

Experimenting with the databases as described before and concentrating on rule generation using mainly AQ19 ML program, a set of rules have been obtained, combining the two above mentioned, different types of approach, which is rather uncommon. The concrete can then be considered frost resistant in general terms, independent of the testing methods, if condition (1) is fulfilled. Produced was also an opposite indication, of concrete being generally sensitive to frost. Such concrete types would be described by condition (2):

$$[\mathrm{w\_c} < 0.43] \quad [\mathrm{fc28} > 55.00]\,, \tag{1}$$

$$[\mathrm{w\_c} > 0.32] \quad [\mathrm{A\_hr} < 6.0] \quad [\mathrm{fc28} < 57.00]\,. \tag{2}$$

There are two new symbols in the formulae above: w_c — a water cement-ratio, which is an example of a derived variable, and A_hr — the volume of air observed in the hardened concrete by image analysis.

The reliability of the results is closely related to the size and quality of the database analysed by the system. With a small test series, which were concerning the experiments performed at IFTR (IPPT PAN), these rules were giving 100% correct classification.

6. Conclusions

The final conclusions of the experiments performed on the suggested system are the following.

- Combined employment of varied soft computing tools makes possible much better data understanding and their more efficient exploitation. It has been demonstrated that by combining different soft computing methods it is possible to improve the quality of data, (by dealing with outliers and missing values problems, clustering and re-ordering of the data, etc.), coming to better prediction possibilities of the whole system of data processing. Data re-ordering seems to be an interesting tool assisting in data analysis, especially to improve the effectiveness of machine learning procedures.
- Different tools of the same family, (e.g. different ANNs or ML applications), work synergistically, and they may complement each other. They should be used in parallel, (the technique of properly combining the results in such case needs yet to be suggested).
- Valuable new results of practical value can be obtained by the proposed system; these will be the conclusions concerning directly the fabrication of concrete. So far, by applying the proposed system it was possible to suggest certain new generalisations in concrete technology, in particular — for evaluation of the generalized frost resistance of hardened concrete.
- Various AI solutions are in fact not exchangeable, and they should be selected cautiously.

It seems possible that a system like the one discussed in this paper in far future can be connected by a feedback with a package of structural engineering computing. This combination should enable a holistic optimization of constructions with simultaneous design of the component materials in the construction.

References

[1] aiNet — ANNs; a trial version, (2.29 MB); at: http://www.ainet-sp.si/NN/En/nn.htm
[2] AQ21 — Machine Learning programs, to be asked for at http://www.mli.gmu.edu/mdirections.html
[3] D. Alterman. *Evaluation of Concrete Materials by Automatic Reasoning (in Polish)*, doctoral dissertation, manuscript, 180 pp. IFTR PAS, Warszawa, 2005.
[4] A.M. Brandt, ed. *Optimization methods for material design of cement-based composites*, 314 pp. E&FN SPON, London, 1998.
[5] A.M. Brandt, J. Kasperkiewicz, eds. *Diagnosis of Concretes and High Performance Concrete by Structural Analysis* (in Polish, 218 pp). IFTR PAS–NATO Sci. Aff. Div., Warszawa, 2003.
[6] S. Dennis, D. McAuley. Connectionist models of cognition. *The BrainWave Connectionist Simulator*, at: http://www.itee.uq.edu.au/~cogs2010/cmc/
[7] GradeStat — grade data analysis package; a free application, (1.26 MB), at http://www.ipipan.waw.pl/~gradestat/
[8] J. Kasperkiewicz. Application of artificial neural networks in designing of concrete mixes, (in Polish). *Cement-Wapno-Beton*, **1/63**, 224–228, 1996.
[9] J. Kasperkiewicz, D. Alterman. On effectiveness of pre-processing by clustering in prediction of C.E. technological data with ANNs. *Proc. of the International Conference: New Trends in Intelligent Information Processing and Web Mining*, Zakopane, June 2–5, 2003. Springer-Verlag, pp. 261–266.
[10] J. Kasperkiewicz, J. Racz, A. Dubrawski. HPC strength prediction using artificial neural network. *Journal of Computing in Civil Engineering*, **9**: 279–284, 1995.

[11] T. Kowalczyk, E. Pleszczyńska, F. Ruland, eds. *Grade models and methods for data analysis. With applications for the analysis of data populations*, 477 pp. Springer-Verlag, 2004.

[12] C.B. Oland, C.F. Ferraris. *Guide to Recommended Format for Concrete in Materials Property Database*, 50 pp. Reported by ACI Committee 126, ACI 126.3R-99, September 1999,

[13] ROSETTA — a Rough Set Toolkit for Analysis of Data; available for non-commercial use, (1.78 MB); at http://rosetta.lcb.uu.se/general/download/

[14] See5/C5.0-demonstration — 10 days evaluation version for Windows XP (608 kB); at: http://www.rulequest.com/download.html

[15] SPSS for Windows — a 14 days evaluation version, (130 MB); at: http://www.spss.com/

[16] Statistics and Neural Network Toolboxes of MATLAB; possibility to ask for a 30 days trial software; http://www.mathworks.com/web_downloads/

[17] WEKA — the package of Data Mining Software, (about 10 MB); available on the web at http://www.cs.waikato.ac.nz/~ml/weka/