Evaluation of pressure soundings exemplified with the Data Mining methods

Problem definition

Soil characteristics that can be used as a starting base for the determination of characteristic values according to DIN V ENV 1997-1:1996-04, 2.4.3, are to be deduced from test results. Pressure sounding as an indirect soil exploration method does not produce any explicit values of the geotechnical parameters. Because of this, quantitative coherence between sounding results and a geotechnical characteristic for the type of soil in question must be either statistically established or deterministically fixed in a verifiable way within predefined limits of validity according to DIN 4094-1:2002. But often classic statistical methods are complicated to use and produce results that are hard to see through.

Research method

As a new computer based method of data analysis, Data Mining is to be adjusted to the problem of pressure sounding evaluation.

Data Mining is defined as a process to identify patterns and structures within data and to constitute rules from great amounts of data.

Data Mining uses different methods:

- Rules of associativity

The rules of associativity say that if a certain event X occurs another event Y often occurs, too. The *confidence* of this rule states the percentage in which the implications actually apply. The *support* value says in how many cases the rules apply at all.

X=> Y (*confidence, support*)

- Methods of classification

The rule of classification tries to predict the value of a dependant variable of aim (the class) from the values of known variables. Rules of classification are needed if missing information is to be deduced from an objects known characteristics. One method of classification is the classification using the decision diagrams.

- Bayes' Classification

This method is based on Bayes' rule of conditional likelihood. It says that if you have a hypothesis H and an event E fitting this hypothesis, the likelihood of H given by the event E is:

P(H½ E) = P(E½ H)P(H) / P(E)

- Clustering

Classes are being established by ably grouping the objects in clusters. In order to obtain as great a separation as possible, the objects within a class must be as similar as possible.

- Genetic algorithms

These algorithms serve the search for the global extreme value of a function. The following steps are gone trough during this process:

- Selection;
- Genetic operation (crossing and mutation)

A tree whose knots represent user prefixed basis functions (summation, division etc.) is used as the chromosome.

State of the art

Several calculations have been run with the help of Microsoft's Data Mining- Package as part of SQL Server2000 in order to analyse the correlation between soil characteristics and nation-wide data from pressure soundings taken from literature. Amongst other things, the influence of the end pressure and the occurring hydrostatic excessive pressure on cohesive soils, expressed by the values of

$(q_t-u_1)/\sigma_{v0}$' and $B_q=\Delta u_2/(q_t-\sigma_{v0})$,

have been examined for the degree of overconsolidation OCR, whereby q_t is the corrected end resistance, σ_{v0} is the total vertical tension and B_q is the pore pressure ratio. The value u_1 represents the pore pressure on the cone face. The characteristic Δu_2 is equivalent to the difference between the entry excess pore pressure behind the cone tip and the equilibrium pore pressure u_0.

It has been established that the influence of the parameter $(q_t-u_1)/\sigma_{v0}$ on the OCR values is significantly greater by overconsolidated Soil than that of B_q, while the B_q plays a decisive role in the determination of the OCR values by normal consolidated soil.

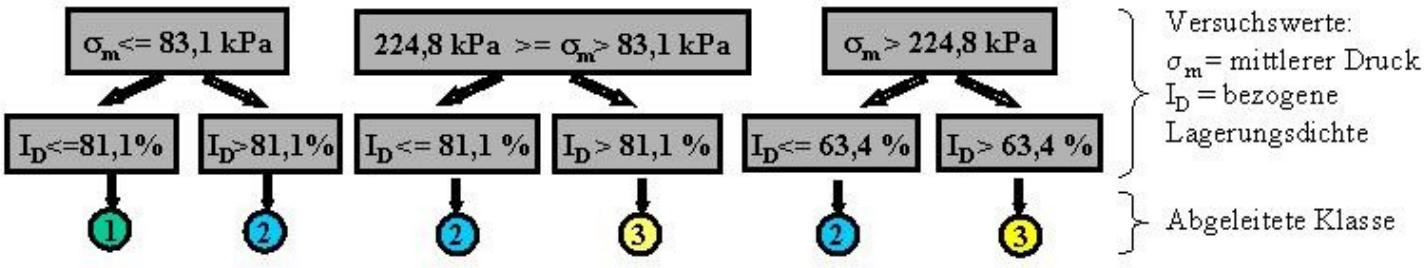
For non- cohesive soils, the coherence between the soil compactness, the middle tension on the sounding area and the probe's end pressure has been analysed with the help of the Data Mining program See5 (www.rulequest.com). The application of the classification methods implemented in this program produce results that are easy to interpret and that are according to statutes known from literature.

Example

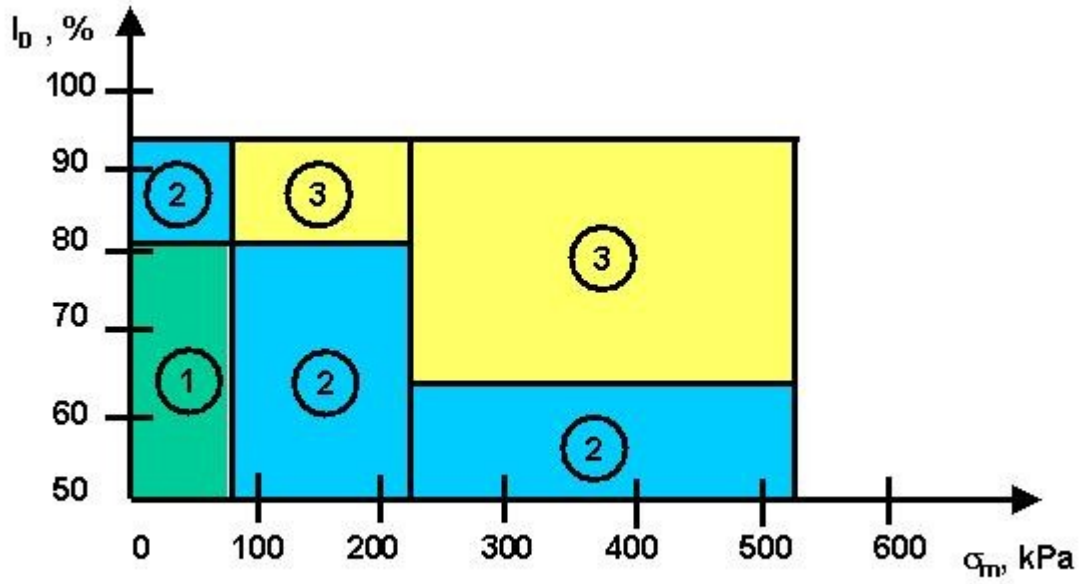
Classification of a cohesiveless soil according to its cone resistance (q_c - value):

Definition der Klassen:

Klasse	q_c [MPa]
1	<15
2	15 - 30
3	30 - 47



Picture 1: Classification using a decision tree



Picture 2: Results of the sub-division in classes in accordance to the decision tree

Future prospects

Data Mining methods are to be adjusted to the needs of CPT results analysis and other indirect methods of investigation to allow for an optimal search for the regularities and coherence between the measurement result and the soil parameters and in addition to that take the scattering of characteristics into consideration. The different software packages available at the institute and the different methods of Data Mining are being presented and tested for their qualification for the analysis of CPT results .

With the help of Data Mining, the rules for soil classification into defined classes as well as for the determination of their mechanical characteristics will be set up based on the results of indirect soil investigation methods. Also, the validity of the coherence and rules found will be analysed.

Literature

- Ziegler, M.; Prokhorova, A.: Verbesserung der Interpretation von Sondierungsergebnissen mit Hilfe von Data Mining-Methoden. In: Vorträge der Baugrundtagung 2004 in Leipzig / Deutsche Gesellschaft für Geotechnik e.V. Essen, Verlag Glückauf, 2004. S. 213–220.
- Prokhorova, A.: Anwendung der Data Mining - Methoden für Prognosen in der Geotechnik. In: Ziegler, M. (Hrsg.): Stochastische Prozesse in der Geotechnik. Aachen, Mainz, 2003. S. 65-79. (Schriftenreihe Geotechnik im Bauwesen ; 1)
- Prokhorova, A.; Ziegler, M.: Better understanding of field measurements by means of data mining. In: Natau, O.; Fecker, E.; Pimentel, E. (Eds.): Geotechnical measurements and modelling. Lisse (u.a.): Balkema, 2003. S. 291-296.
- Prokhorova, A.; Ziegler, M.: Anwendung von Data mining in der Geotechnik. In: 2. Siegener Kolloquium Erkundung und Felduntersuchungen in der Geotechnik. Siegen: Universitätsverl., 2003. S. 1-21.