Mathematical Research

Modelling Uncertain Data

edited by

Hans Bandemer

Volume 68

Akademie Verlag

[19] TANAKA, H. / WATADA, J. (1988).
Possibilistic linear systems and their applications to the linear regression model
Fuzzy Sets and Systems 27, 275-289

[20] WEBER, S. (1984). Measures of fuzzy sets and measures of fuzziness.
Fuzzy Sets and Systems 13, 247 - 271.

[21] YAGER, R. R. (1980). On a general class of fuzzy connectives,
Fuzzy Sets and Systems 4, 235 - 242.

[22] ZADEH, L. A. (1968). Probability measures of fuzzy events.
J. Math. Anal. Appl. 23, 421 - 427.

On the Specification of Fuzzy Data in Management

Andreas Geyer-Schulz[1]

Abstract. In decision-making we perceive uncertainty almost always as a major stumbling block for successfully applying quantitative models. Ususally we distinguish three different types of uncertainty, namely risk, imprecision and vagueness, all of which may be mixed up in the same situation. Consider, for example, buyer intention surveys for forecasting future product demand [11, p. 228]: We select a random sample of buyers (risk), assess their income (imprecision) and their intention to buy our product on a verbal scale ranging from "no chance" to "certain" (vagueness). In the rest of this article we concentrate on the modeling of vagueness with fuzzy rule languages. First, we establish the meaning of expressions like "very high" by assigning fuzzy subsets to them and we thus design fuzzy rule languages for qualitative modeling [3]. As example we introduce the Boston Consulting Group's approach for establishing a company's business portfolio plan [11, p. 76]. Our motivation comes from two sources, namely from the wide-spread use of qualitative techniques like portfolio analysis, scenario techniques and gap analysis in strategic management [11] and from the recent success of fuzzy rule languages in robotics and control [1, 10, 13]. We show that fuzzy rule languages are very high languages which, because of abstraction, considerably reduce problem complexity. In the setting of genetic based machine learning we can give a precise interpretation of this statement. The abstraction level of a language is inversely related with computational complexity: We prove, that in this context fuzzy classifier systems have a better performance than crisp classifier systems. We demonstrate these techniques for the Boston Consulting Group example.

The Boston Consulting Group Approach. The Boston Consulting Group Approach is one of the best-known methods for deriving a company portfolio plan with the objective of refreshing the company's portfolio of businesses by flushing out poor ones and adding promising new ones [11]. For this purpose management first identifies the strategic business units (SBUs) making up the company. In the next step management classifies SBUs according to market growth rates and relative market shares as pioneered by the Boston Consulting Group. In Figure 1 we show a result of this classification.
Assuming that each quadrant in Figure 1 represents a distinct type of cash-flow situation, we obtain the following classification of SBUs:

- "Stars": high-growth, high-share SBUs.
- "Cash Cows": low-growth, high-share SBUs.
- "Question Marks": high-growth, low-share SBUs.
- "Dogs": low-growth, low-share SBUs.

With the help of a fuzzy rule language whose syntax we define below and whose implementation is presented in [3] we may represent this classification scheme with the the following fuzzy rules:

[1] Currently: Department of Business Administration / Management Information Systems, University of Augsburg, Memmingerstraße 18, D-8900 Augsburg, Germany. On leave: Department of Applied Computer Science, Institute of Information Processing and Information Economics, Vienna University of Economics and Business Administration, Augasse 2-6, A-1090 Vienna, Austria.

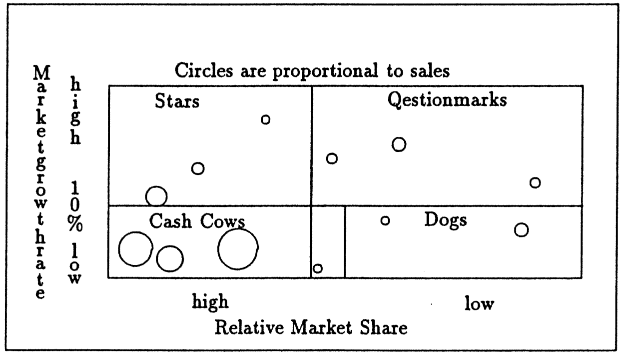

Figure 1: Business Portfolio Matrix

[1] STAR USED IF(MARKETGROWTH IS HIGH)AND(RELATIVE_SHARE IS HIGH)
[2] QUESTION_MARK USED IF(MARKETGROWTH IS HIGH)AND(RELATIVE_SHARE IS LOW)
[3] CASH_COW USED IF(MARKETGROWTH IS LOW)AND(RELATIVE_SHARE IS HIGH)
[4] DOG USED IF(MARKETGROWTH IS LOW)AND(RELATIVE_SHARE IS LOW)

The distribution of SBUs in the four quadrants of Figure 1 shows the company's current state of health. Top management should project a future matrix showing where each SBU is likely to be with no change in strategy [11]. Clearly it is a top management job to suggest a strategy to obtain an "optimal distribution" of SBUs. According to the Boston Consulting Group, four basic strategies are available for each SBU:

- Build: particularly appropriate for "Question Marks", a strategy for improving the market position by investing in the SBU.
- Hold: suitable for strong "Cash Cows", a strategy aiming at preserving the market position of the SBU.
- Harvest: useful for weak "Cash Cows" or "Question Marks" or "Dogs", a strategy for obtaining a short-term increase in cash flow regardless of long-term effects.
- Divest: particularly appropriate for "Dogs" and "Question Marks" that cannot be financed for growth, a strategy for selling or liquidating the SBU.

Of course, we may program this heuristic with the following fuzzy rules:

[1] BUILD USED IF(MARKETGROWTH IS HIGH)AND(RELATIVE_SHARE IS LOW)AND(SALES IS ABOVE LOWER MEDIUM)
[2] HOLD USED IF(MARKETGROWTH IS LOW)AND(RELATIVE_SHARE IS HIGH)AND(SALES IS ABOVE UPPER MEDIUM)
[3] HARVEST USED IF(MARKETGROWTH IS LOW)AND(RELATIVE_SHARE IS HIGH)AND(SALES IS BELOW MEDIUM)
[4] HARVEST USED IF(MARKETGROWTH IS LOW)AND(RELATIVE_SHARE IS LOW)AND(SALES IS ABOVE MEDIUM)
[5] HARVEST USED IF(MARKETGROWTH IS HIGH)AND(RELATIVE_SHARE IS LOW)AND(SALES IS BELOW MEDIUM)

[6] DIVEST USED IF(MARKETGROWTH IS LOW)AND(RELATIVE_SHARE IS LOW)AND(SALES IS BELOW MEDIUM)
[7] DIVEST USED IF(MARKETGROWTH IS HIGH)AND(RELATIVE_SHARE IS LOW)AND(SALES IS BELOW MEDIUM)

And by executing these fuzzy rules on a computer we are now in the position to pursue a hands on approach and to see what happens:

```
      SBU←ENTER_SBU
NAME OF SBU:UNIFOOD
MARKETGROWTH:VERY LOW
RELATIVE MARKET SHARE:VERY HIGH
SALES:RATHER LOWER MEDIUM
      CLASSIFY_SBU SBU
  UNIFOOD  IS A   CASH COW
      RECOMMEND_STRATEGY SBU
  HARVEST    UNIFOOD  WITH CERTAINTY    0.99994892
```

Model Prototyping. Even without specifying the quantitative meaning of our fuzzy rules we are already able to simulate a verbally formulated theory in order to test its consistency, its completeness and its face-validity [2]. The verbal results of such a model prototyping process are easy to understand and correspond quite well with conclusions drawn by managers and prescriptions given in management text books. We start to analyze qualitative models in a formal setting. The main benefit to the management is, that the specification of the quantitative meaning of the rule-language can be delegated to a second modeling-phase, thus allowing for a concentration on the essential by skipping all distracting details. Obviously, the fuzzy rule-language is used as a very high and abstract language in qualitative management.

What are the effects of using a very high and abstract language except intuitive appeal? On an informal level, we would not hesitate to claim that new rules are easier to learn and that it is easier to test, whether they make sense. It is just easy learning and easy testing. Next, let us move these informal ideas into a formal setting: consider, for instance, genetic based machine learning.

Fuzzy Classifier Systems. A classifier system is a machine learning system which learns rules to guide its performance in an arbitrary environment [6]. Its main components are a production system, an apportionment of credit function and a genetic algorithm. A *fuzzy* classifier system learns rules written in a *fuzzy* rule language and uses a *fuzzy* production system. "Learning a heuristic" means for a genetic algorithm to search a rule-base which performs well with regard to the objective function of the system.

The search space S of a classifier system whose production system is defined by the formal language L is the set W of all words of L which can be generated with at most d derivations. Each rule-base w is a word of W, and in real applications d is always fixed to a finite natural number. Because d is finite, the cardinality of S, denoted by $card(S)$, is finite and with regard to the objective function $z : S \rightarrow R$, a global optimum exists and can be found with a genetic algorithm [4]. Without additional information about S and $z(w)$, we can find the optimum by complete enumeration of S. In this case (which is the worst case) the computational complexity $t_{w_{opt}}$ is given by $t_{w_{opt}} = \sum_{w \in S} \cdot t(z(w))$, where $t(z(w))$ is the evaluation time of $z(w)$ with respect to a rule-base w. If we hold $t(z(w))$ constant, we get $t_{w_{opt}} = card(S) \cdot t(z(w))$. Even with additional information about S and z and more sophisticated search algorithms than complete enumeration, the computational complexity of learning is a function of search space size. Therefore, the search space size $card(S)$ is a useful quantitative measure for ease of learning and testing.

Let us now return to our original objective: to prove, that a fuzzy classifier system with the fuzzy rule language FRL used in the Boston Consulting Group example learns a rule-base faster than a classifier system with the crisp rule language RL. Since, in this setting, learning depends on search space size, we reformulate the previous proposition in terms of search space size. By $S_{FRL,d}$ we denote the search space S generated from FRL with at most d derivations. $\Phi_I : FRL \hookrightarrow RL$ is an embedding of FRL into RL which translates rule–bases in FRL to RL. $\phi_I(d) : N \rightarrow N_0$ maps the number of derivations d in FRL to the minimal number of derivations needed in RL, so that all rule–bases which may result from an application of Φ_I to rule–bases in $S_{FRL,d}$ are in $S_{RL,\phi_I(d)}$. Furthermore, k is the minimal number of derivations needed, so that at least one rule-base is contained in $S_{FRL,k}$.

Theorem 1 *$card(S_{FRL,d}) < card(S_{RL,\phi_I(d)})$, for all $d \geq k$.*

Proof In the proof we exploit a natural interconnection between formal languages and formal power series [7, 8]. The proof proceeds as follows:

1. We define the formal language FRL:

⟨rule base⟩ := ⟨rule⟩ | ⟨rule⟩ ⟨rule base⟩
⟨rule⟩ := ⟨frame⟩ "USED" "IF" ⟨truth value⟩
⟨truth value⟩ := ⟨truth value⟩ ⟨connective⟩ ⟨truth value⟩ | "(" ⟨noun⟩ "IS" ⟨verbal expression⟩ ")"
⟨verbal expression⟩ := ⟨adjective⟩ | ⟨adverb⟩ ⟨verbal expression⟩ | "(" ⟨verbal expression⟩ ")"
⟨connective⟩ ⟨verbal expression⟩
⟨adjective⟩ := "HIGH" | "LOW" | "MEDIUM" | "UNDEFINED" | "UNKNOWN"
⟨adverb⟩ := "ABOVE" | "BELOW" | "AROUND" | "UPPER" | "LOWER" | "MORE_OR_LESS" | "RATHER" | "VERY" | "NOT" | "NEITHER" | "POSSIBLY" | "TRULY" | "FUZZILY"
⟨connective⟩ := "AND" | "OR" | "BUT" | "NOR" | "TO" | "EXCEPT"
⟨noun⟩ := "MARKETGROWTH" | "RELATIVE_SHARE" | "SALES"
⟨frame⟩ := "STAR" | "QESTION_MARK" | "CASH_COW" | "DOG" | "BUILD" | "HARVEST" | "HOLD" | "DIVEST"

2. We define the formal language RL:

⟨rule base⟩ := ⟨rule⟩ | ⟨rule⟩ ⟨rule base⟩
⟨rule⟩ := ⟨frame⟩ "USED" "IF" ⟨truth value⟩ "WITH" ⟨α⟩
⟨truth value⟩ := ⟨truth value⟩ ⟨connective⟩ ⟨truth value⟩ | "(" ⟨noun⟩ "IN" ⟨interval list⟩ ")"
⟨connective⟩ := "AND" | "OR" | "BUT" | "NOR" | "TO" | "EXCEPT"
⟨noun⟩ := "MARKETGROWTH" | "RELATIVE_SHARE" | "SALES"
⟨frame⟩ := "STAR" | "QESTION_MARK" | "CASH_COW" | "DOG" | "BUILD" | "HARVEST" | "HOLD" | "DIVEST"
⟨interval list⟩ := ⟨interval⟩ | ⟨interval⟩ ⟨interval list⟩
⟨interval⟩ := "[" ⟨real⟩ "," ⟨real⟩ "]"
⟨real⟩ denotes the m-bit representation of a floating point number.
⟨α⟩ is a ⟨real⟩ in [0, 1].

3. With the help of the representation theorem of [12] we can construct a mapping Φ : $FRL \hookrightarrow RL$, which translates each fuzzy rule to an (infinite) indexed family over the real unit interval [0, 1]. By choosing a finite chain I with elements in [0, 1] we approximate Φ by $\Phi_I : FRL \hookrightarrow RL$ defined by the following set of meta production rules [5, 9]:

$\forall \alpha \in I$: ⟨frame⟩ "USED" "IF" ⟨truth value⟩
$\Longrightarrow$ ⟨frame⟩ "USED" "IF" ⟨truth value⟩ "WITH" ⟨α⟩

"(" ⟨noun⟩ "IS" ⟨verbal expression⟩ ")" $\Longrightarrow$ "(" ⟨noun⟩ "IN" ⟨interval list⟩ ")"

4. To count the number of words derivable from a nonterminal symbol in d derivations we define the counting function $\Pi_{FRL} : V_N \times N \rightarrow N_0$, where V_N denotes the nonterminal symbol set of FRL and the second argument the number of available derivations. Therefore, $i, j, k \in N$ and the value of the counting function is 0, if the derivations are exhausted.

$$\Pi_{FRL}(\langle\text{adjective}\rangle, 1) = 5,\ \Pi_{FRL}(\langle\text{adverb}\rangle, 1) = 13,\ \Pi_{FRL}(\langle\text{connective}\rangle, 1) = 6,$$
$$\Pi_{FRL}(\langle\text{noun}\rangle, 1) = 3,\ \Pi_{FRL}(\langle\text{frame}\rangle, 1) = 8.$$

In the following recursive scheme the "constant" nonterminal functions given above have been replaced by their value:

$$\Pi_{FRL}(\langle\text{rule base}\rangle, i) = \Pi_{FRL}(\langle\text{rule}\rangle, i-1) + \sum_{j+k=i-1} \Pi_{FRL}(\langle\text{rule}\rangle, j) \cdot \Pi_{FRL}(\langle\text{rule base}\rangle, k)$$

$$\Pi_{FRL}(\langle\text{rule}\rangle, i) = 8 \cdot \Pi_{FRL}(\langle\text{truth value}\rangle, i-2)$$

$$\Pi_{FRL}(\langle\text{truth value}\rangle, i) = 3 \cdot \Pi_{FRL}(\langle\text{verbal expression}\rangle, i-2) + \sum_{j+k=i-2} 6 \cdot \Pi_{FRL}(\langle\text{truth value}\rangle, j) \cdot \Pi_{FRL}(\langle\text{truth value}\rangle, k)$$

$$\Pi_{FRL}(\langle\text{verbal expression}\rangle, i) = 5 + 13 \cdot \Pi_{FRL}(\langle\text{verbal expression}\rangle, i-2) + \sum_{j+k=i-2} 6 \cdot \Pi_{FRL}(\langle\text{verbal expression}\rangle, j) \cdot \Pi_{FRL}(\langle\text{verbal expression}\rangle, k)$$

The search space size is then:

$$S_{FRL,d} = \sum_{i=1}^{d} \Pi_{FRL}(\langle\text{rule base}\rangle, i)$$

5. We define the counting function Π_{RL} in the same way:

$$\Pi_{RL}(\langle\text{connective}\rangle, 1) = 6,\ \Pi_{RL}(\langle\text{noun}\rangle, 1) = 3,\ \Pi_{RL}(\langle\text{frame}\rangle, 1) = 8,$$
$$\Pi_{RL}(\langle\text{real}\rangle, 1) = 2^{64},\ \Pi_{RL}(\langle\text{interval}\rangle, 2) = 2^{128}.$$

$$\Pi_{RL}(\langle\text{rule base}\rangle, i) = \Pi_{RL}(\langle\text{rule}\rangle, i-1) + \sum_{j+k=i-1} \Pi_{RL}(\langle\text{rule}\rangle, j) \cdot \Pi_{RL}(\langle\text{rule base}\rangle, k)$$

$$\Pi_{RL}(\langle\text{rule}\rangle, i) = 2^{67} \cdot \Pi_{RL}(\langle\text{truth value}\rangle, i-3)$$

$$\Pi_{RL}(\langle\text{truth value}\rangle, i) = 3 \cdot \Pi_{RL}(\langle\text{interval list}\rangle, i-2) + \sum_{j+k=i-2} 6 \cdot \Pi_{RL}(\langle\text{truth value}\rangle, j) \cdot \Pi_{RL}(\langle\text{truth value}\rangle, k)$$

$$\Pi_{RL}(\langle\text{interval list}\rangle, i) = 2^{128} + 2^{128} \cdot \Pi_{RL}(\langle\text{interval list}\rangle, i-3)$$

The search space size is then:

$$S_{RL,d} = \sum_{i=1}^{\phi_I(d)} \Pi_{RL}(\langle\text{rule base}\rangle, i)$$

6. We define $\phi_I(d)$ in such a way, that each word w of FRL, when translated to RL by Φ_I is derivable in at most $\phi_I(d)$ derivations and that $\phi_I(d)$ is the smallest number, for which this holds. For this purpose we develop the recursive scheme of the function $\varphi : V_N \times N \rightarrow N_0$ which counts the number of derivations in FRL and we adjust this scheme in order to take Φ_I into account.

$$\varphi(\langle\text{rule base}\rangle, i) = \max((1 + \varphi(\langle\text{rule}\rangle, i-1)), (1 + \max_{j+k=i-1}(\varphi(\langle\text{rule}\rangle, j) + \varphi(\langle\text{rule base}\rangle, k))))$$

For $\varphi(\langle\text{rule}\rangle, i)$ the consideration of the first meta production rule requires two changes. In RL we need one additional derivation for the α-level and each rule of FRL is translated in $card(I)$ rules in RL.

$$\varphi(\langle\text{rule}\rangle, i) = card(I) \cdot (3 + \varphi(\langle\text{truth value}\rangle, i-2)) + card(I) - 1$$

$$\varphi(\langle\text{truth value}\rangle, i) = \max(2 + \max_{j+k=i-2}(\varphi(\langle\text{truth value}\rangle, j) + \varphi(\langle\text{truth value}\rangle, k))), (2 + \varphi(\langle\text{verbal expression}\rangle, i-2)))$$

For deriving the scheme for $\varphi(\langle\text{verbal expression}\rangle, i)$ we have to analyze three cases corresponding to the three production rules. In the first case we translate an adjective into an interval, because the meaning of an adjective is defined by a convex fuzzy set. Therefore, we need one additional derivation in RL. In the second case we consider adverb adjective combinations. Since adverbs usually only influence the interval range but not the number of intervals, we need one derivation less in RL than in FRL. The last case covers the case of logical combinations of verbal expressions. In the case of combining n adjectives with "OR" we get an interval list with n intervals.

$$\varphi(\langle\text{verbal expression}\rangle, i) = \max(3, (1 + \varphi(\langle\text{verbal expression}\rangle, i-2)), (2 + \max_{j+k=i-2}(\varphi(\langle\text{verbal expression}\rangle, j) + \varphi(\langle\text{verbal expression}\rangle, k))))$$

Finally, $\phi_I(d) = \max_{i \leq d} \varphi(\langle\text{rule base}\rangle, i)$.

7. We establish the theorem by comparing the two series $\Pi_{FRL}(d)$ and $\Pi_{RL}(\phi_I(d))$ for all $d \geq k$ ■

References

[1] T. Akiyama, T. Sasaki: Extended Fuzzy Traffic Control Model on Urban Expressway. Preprints of Second IFSA Congress 1 (1987), 1, 278 – 279.

[2] A. Geyer-Schulz, A. Taudes: Fuzzy Rule-Based Simulation with Economic Applications. Proceedings of the European Simulation Multiconference (1987), 204-209.

[3] A. Geyer-Schulz: Fuzzy Rule-Based Expert Systems. In: J.R. Kraemer and P.C. Berry: APL Techniques in Expert Systems. ACM SIGAPL, Syracuse, NY 1988.

[4] A. Geyer-Schulz: On Learning in a Fuzzy Rule-Based Expert System. Kybernetika 28 (1992), 34–36.

[5] A. Geyer-Schulz: Fuzzy Classifier Systems. In: Fuzzy Logic: State of the Art, Kluwer Academic Publishers. To appear.

[6] D. E. Goldberg: Genetic Algorithms in Search, Optimization & Machine Learning. Addison–Wesley, Reading, MA 1989.

[7] I. P. Goulden, D. M. Jackson: Combinatorial Enumeration. John Wiley & Sons, New York 1983.

[8] W. Kuich, A. Salomaa: Semirings, Automata, Languages. EATCS Monographs on Theoretical Computer Science, Vol. 5, Springer Verlag, Berlin, 1986.

[9] R. Kruse, K. D. Meyer: Statistics with Vague Data. D. Reidel Publishing Company, Dordrecht, 1987.

[10] K. Hirota, Y. Arai, S. Hachisu: Real Time Pattern Recognition and Fuzzy Controlled Robot-Arm. Preprints of Second IFSA Congress 1 (1987), 1, 274 – 277.

[11] P. Kotler: Marketing Management – Analysis, Planning and Control. Prentice-Hall, Englewood Cliffs, New Jersey, 1980.

[12] C. V. Negoita, D. A. Ralescu: Representation Theorems for Fuzzy Concepts. Kybernetes 4 (1975), 4, 169 – 174.

[13] D. Yasuhiko, M. Honma: Improved Fuzzy-Set Temperature Distribution Control for Electric Furnace. Preprints of Second IFSA Congress 1 (1987), 1, 270 – 273.

Specification of Medical Data With Fuzzy Methods

Wolfgang Schüler [1]

Before you can start a therapy you must have some idea of a diagnosis. The present paper deals with a computer assisted diagnosis system based on fuzzy classification.

During the last years for supporting the medical decision process a lot of investigations were done to use computer assisted systems for diagnosing or differential diagnosing in almost all medical branches. The mathematical methods most frequently used for this purpose are taken from mathematical logic and from probability theory. Such techniques, as predicate logic, BAYESian reasoning, and discrimination analysis, were also suggested for interpreting X-ray pictures, however, up to now, found no common application. Possible reasons for the reserve of researchers and practitioners with respect to these methods are

- nearly all features are of qualitative character,
- independence of the features is questionable,
- the number of features is large,
- most of the features are subjective in nature due to visual interpretation of X-ray pictures by the observing physician,
- usual conditions as tacitly assumed, when applying methods from mathematical statistics, are obviously not met by the data,
- the number of diagnosis classes is large,
- feature presentation can be weak, hence the diagnostic decision can be highly uncertain,
- medical information is usually "soft" in its nature (e.g. when using natural or professional language: low – high, little – big, cast suspicion on ..., benign – malign, possible diagnosis ...).

For improving the quality of diagnosis with respect to tumors an expert system was developed using the theory of fuzzy sets for describing individual medical information with respect to patients as well as for specifying global information about diagnosis classes. The special field of application of this system is diagnosis with respect to early stages of carcinoma of the mammary gland.

Making a diagnosis of early stages of that carcinoma is a difficult problem. Detection of mamma tumor at the earliest is decisive for the patient's prognosis, since his survival rate depends strongly on the stage of the tumor in that moment, when treatment is started.

Basic strategies for detecting carcinoma of the mammary gland are clinical and radiological methods, especially mammography. In the resulting mammograms early stages of cancer are reflected by only very small variations of selected features. Hence assigning of the patient to each of a broad variety of diagnosis classes seems possible. The physician can detect a cancer only by combining and selection of several features. If the mammogram contains essential information with respect to more than four or five features, then it is beyond the explanation ability of the observer. On the other hand a high certainty of the diagnostic decision is required for tumor diseases already in an early stage, since the therapeutic strategy and hence the prognosis of the patient depends on that decision in a high degree. The security of diagnostic decisions at

[1] Hospital Küchwald, Clinic for Radiology, Department of Medical Physics, Bürgerstr. 2, O-9003 Chemnitz, Germany