Uncertainty Modelling in Diagnostic Systems: An Adaptive Solution

Giacomo Piccinelli
Extended Enterprise Laboratory
HP Laboratories Bristol
HPL-98-37
February, 1998

E-mail: giapicc@hplb.hpl.hp.com

diagnosis, case-based reasoning, artificial intelligence, information retrieval, knowledge management

Uncertainty permeates the entire diagnostic process and its management is a fundamental issue in actual diagnostic systems. The type of information we can model about the context in which a problem occurred is crucial. The main components pictured in the definition of a context are observations (facts) but we argue that data on relevance and confidence may add precious information.

Focusing on case-based reasoning (CBR) paradigm, we present a model in which relevance and uncertainty become fundamental and dynamic components of both diagnostic knowledge and processes: fuzzy sets are the theoretic base of the model. A conversational CBR shell implementing nearest-neighbour (NN) retrieval mechanisms has been developed in order to test our proposal in terms of case-retrieval precision and we discuss the results obtained in some experiments. The "knowledge level" impact of our proposal is also discussed.

Uncertainty Modelling in Diagnostic Systems: An Adaptive Solution

Giacomo Piccinelli

Extended Enterprise Laboratory
Internet Business Management Department
Hewlett-Packard Laboratories
Bristol BS12 6QZ, U.K.

Tel.: +44 (0)117 9229610 – Fax: +44 (0)117 9229250
Email: giapicc@hplb.hpl.hp.com

Abstract

Uncertainty permeates the entire diagnostic process and its management is a fundamental issue in actual diagnostic systems. The type of information we can model about the context in which a problem occurred is crucial. The main components pictured in the definition of a context are observations (facts) but we argue that data on relevance and confidence may add precious information.
Focusing on case-based reasoning (CBR) paradigm, we present a model in which relevance and uncertainty become fundamental and dynamic components of both diagnostic knowledge and processes: fuzzy sets are the theoretic base of the model. A conversational CBR shell implementing nearest-neighbour (NN) retrieval mechanisms has been developed in order to test our proposal in terms of case-retrieval precision and we discuss the results obtained in some experiments. The "knowledge level" impact of our proposal is also discussed.

1 Introduction

The precise identification of the context in which a problem occurs is fundamental in order to diagnose its causes and, eventually, to fix it. The more accurate the information on the context, the more precise the diagnosis can be.
The goal of a diagnostic system is to maintain and extract from an information base facts, rules and any other type of indications that can help identifying problems. In this process meta-information (information on data) is fundamental [3,4,9]. Relevance is the most common (if not the only) parameter associated to the components of a context and it is quite useful as a discriminator. The problem with relevance as single meta-descriptor for context components is that it becomes a container for different aspects of information and the result is an average indication of the component "weight" in the context but the semantics of this weight becomes vague [15]. Case-based reasoning (CBR) paradigm [2,10] is particularly sensitive to the accuracy of knowledge

description and we oriented our research towards CBR systems enforcing “nearest neighbour (NN)” retrieval techniques [2,12].
We propose a model in which, capitalising on the expressiveness and flexibility of type 2 fuzzy sets, we explicitly manage both relevance and confidence as meta-descriptors of diagnostic knowledge. After an introduction to fuzzy set theory and a brief overview of CBR paradigm, we present a model for adaptive management [8] of both relevance and uncertainty [17] as information meta-descriptors. Experimental results show that precision substantially improves (up to a factor of 5) for the case-selection process but we also consider other the “knowledge-level” [5] implications of explicit uncertainty management.

2 Elements of fuzzy set theory

Standard *set theory* describes a set as the collection of all the elements for which a given (binary) predicate holds true. This definition actually splits a given world in two parts and the elements are distinguished only from the fact they belong to the set or not. We can consider a number of predicates at the same time and look at the intersection area but this solution becomes quickly unmanageable when the number of predicates grows. What we would like to do is to take our world *A* of elements and to associate an element coming from a potentially different world *B* to each of them depending on some sort of criteria. We can now partition the elements of A looking at their associated element in B.
Without losing in generality, we can think of B as the real numbers in the range [0,1]: the mapping of different cases is usually straightforward. We can imagine a set of bubbles including all the points of A with the same associated element. The step back to normal sets is simple: we only need to restrict ourselves to {0,1} as associated world.

Definition: *Given a pair of standard sets B and M, a* ***fuzzy set*** *F based on B is a pair (B, f) where f: B →M.*

In the usual terminology [11,18]: *B* is the “base set” or “support”, *M* is the “membership space” and *f* is the “membership function” mapping any element of the support in the correspondent membership value. When M is the interval [0,1] the fuzzy set is “normalised”. The membership function is the main component in the definition: intersection, union, complement, cardinality as well as other concepts of standard set theory are transferred to fuzzy sets working on f [18].
The basic definition of fuzzy set [16] can be generalised and the simplest extension is the recursive use of fuzzy sets in the definition of the membership space. The concept of type is introduced for a fuzzy set in order to express the "depth" of its membership space [16].

Definition*: Let us consider a normalised fuzzy set as having "type 1". A "****type m****" fuzzy set is a fuzzy set with base B whose membership values are type m-1 (m>1) fuzzy sets with base [0,1]*.

Thinking at the membership value associated to an element of the support as a description for that element [13,14], “type m” fuzzy sets introduce a hierarchical structure on the description. Each component at one level may be further specified in the lower levels and the deeper the hierarchy, the more precise the description.

We are mainly interested in type 2 fuzzy sets on top of which we define problem specific metrics and operations.

3 Uncertainty and Adaptivity in CBR diagnostic models

Case-based reasoning (CBR) paradigm [2] starts from the assumption that cognitive process is structured as a cycle. The first step is to gather some knowledge, then the knowledge is used to solve a problem and, depending on the result, we may decide to keep track of the new experience. Experience is accumulated either adding new information or adapting the existing knowledge. The idea is to solve a problem with the existing skills and, at the same time, improving these skills for future use.
A number of different solutions [1,8] have been developed for the actual implementation of this paradigm and the focus is on how to aggregate and store the atomic information (cases) and how to retrieve them. The solution of a problem depends on the ability of the system to retrieve similar cases for which a solution is already known [6]. If a perfect matching is not found the system has to choose cases in some way similar to the one describing the problem and to infer a potential solution. Inference process tends to be limited and the emphasis is on the retrieval of similar cases: the more common retrieval techniques are Inductive Retrieval (IR) and Nearest Neighbour (NN) [12]. In the case of IR a predefined access structure called "induction tree" hosts the cases and provides indications on their characteristics for a quicker access. NN techniques impose more flexible structure on the information at the cost of more expensive search procedures. Flexibility and access speed may be balanced depending on specific needs but in both solutions the characteristics of the information we found in the "cases" are crucial [7].
Thinking of a diagnostic "case", we try to define the context in which a failure occurs. The starting point is a set of facts (observations) but the same fact may have different relevance in different contexts. Collecting, in the case description, information on the relevance of facts allows being more precise in the retrieval (matching) process [7] and precision is fundamental when the dimension of the system knowledge base grows. The problem is that, while observations are hardly disputable, the relevance associated with them may depend on the experience of the observer. What usually happens is that confidence and relevance are empirically merged in a single value and this may corrupt the information.
To exemplify the potential problem we can think at a "warning LED" on a faulty device. In certain circumstances an engineer may be 100% sure that (in a scale from 1 to 10) the relevance of the fact "LED on" is 3 and the diagnosis is d_1. In a different situation the engineer may think that the fact is very important (relevance 10) but he or she is not sure about that (confidence 30%) and the diagnosis is d_2. In both cases it is likely that in the context describing the faulty device the fact "LED on" is assigned relevance value of 3. If something similar happens to the other facts we may end up with two diagnosis for what the system considers a single problem.
The better solution is to enrich the "case" with facts that model the "circumstances" in which observations are taken but it is not easy to identify all of them. A different (or complementary) solution is to explicitly model and manage the uncertainty associated with the observation [9]. The idea is to capture in this way the fact that there is something missing even if we don't know what it is. Certainty may be reinforced or reduced and adaptivity plays a fundamental role in this kind of process.

4 Model specification

We propose an integrated approach to both uncertainty and adaptivity problems based on type 2 fuzzy sets [18]. The result is a model in which both static and dynamic aspects coexist and support each other. The information selection technique is based on a “nearest neighbour” approach. We first introduce the static aspects of the model (knowledge component definition) and then the clustering and adaptivity features [8].

4.1 Information Description

Given a problem context C, our purpose is to obtain a compact but comprehensive description C_d of it. In our proposal, the key elements of the description are a set F of facts together with their relevance (in the context) and the confidence on that relevance (ex. {(Fact = No_Power, Relevance = 3, Confidence = 60%), ..,( )}).
The actual representation structure is a fuzzy set based construction that allows modelling, in a single point, different views on the same fact set. This means that facts present in a case description are the first aggregation element but we keep also track of the different relevance values and related confidence. This information is exploited in the definition of similarity concepts between two cases but it proves to be useful also for the dynamic aspects of the model.

Definition: *A **case description** C_d is a pair composed of a type 2 fuzzy set FS (W, f) and a value ε we call experience. W is a set of facts (represented by strings) and the function **f** maps every $w \in W$ in a fuzzy set RFS ([0,1], ρ) where $\rho : [0,1] \to [0,1]$.*

The RFS fuzzy sets is a relevance descriptor that represents what can be seen as a "confidence distribution" over the normalised set of relevance values: each fact has its own descriptor. We can assume that the absence of a fact from W is equivalent to its presence in association to an RFS where ρ is a constant function that returns the smallest real number greater than 0. The fact that ρ has value 0 in an interval [x, y] means that its behaviour in [x, y] is unspecified. We can also assume f extended over any super-set Ω of W where it returns a dummy RFS for every w in Ω - W. The experience parameter ε is fundamental for the adaptivity features of the model as it gives an indication on the “strength” of the present status of the description. When we collect some feedback on C_d suggesting to change (to adapt) part of it, we can refer to ε in order to establish the scope of the change.
From the definition of C_d, we notice that the emphasis is on the ρ functions. They collect the actual information on the confidence distribution and they represent the crucial point to work on for both retrieval and adaptivity processes. We discuss possible candidates in the next section. If we think of every C_d as a point in a complex “case space”, we now need to impose a metric on that space in order to manage concepts like similarity between descriptions that are fundamental in the perspective of clustering and retrieval activities. We propose a binary function D_σ (we call it "distance function") that compares on a component-by-component base two case descriptions summarising the result in a numeric value.

Definition: *Given a pair of functions ρ_1 and ρ_2 where ρ_i: [0,1] $\to$ [0,1] for i$\in$ {1,2}, σ an integer value and $\{[x^i, y^i]\}_{i=1..n}$ the set of intervals in [0,1] where the value of both ρ_1 and ρ_2 is not 0, we define the support function d_σ as*

$$d_\sigma(\rho_1, \rho_2) = \sum_{i=1..n} \int_{[x_i, y_i]} (\rho_1(x) - \rho_2(x))^{2\sigma} dx$$

Given a pair of case descriptions $C_d1=(W_1, f_1)$ and $C_d2=(W_2, f_2)$, we define the ***distance function*** *D_σ as*

$$D_\sigma (C_d1, C_d2) = \sum_{w \in W1 \cup W2} d_\sigma (f_1(w), f_2(w))$$

The σ parameter is a positive integer value that decides the sensitivity of the function: the bigger it is, the lower is the amplification of the differences between the components. The function exploits all the knowledge on relevance and associated confidence accumulated in the fuzzy set structure in order to take into consideration all the views on every component of the description. For the unspecified parts one of the ρ functions we assume a perfect matching with the other one.

4.2 Adaptivity

Knowledge evolution is a fundamental aspect of diagnostic cycle in terms of new information gathering as well as tuning of existing data. The solution we enforce takes advantage of the case description structure (C_d) and the interaction with the environment. If for the same case we have a C_d (S) from the system and a C_d (U) from the user, the idea is for the system to learn from the user. This doesn't mean that the system has to accept completely the user point of view replacing S with U but that we need to find an appropriate balance.
We need a special "unification" mechanism that merges two C_d in a meaningful way: the solution we propose is to link the weight of a C_d to the experience ε and to compute a weighted average value for all the components.

Definition: *Given two case descriptions $C_d1 <\varepsilon_1, (W_1, f_1)>$ and $C_d2 <\varepsilon_2, (W_2, f_2)>$ we define $M_{\alpha\beta}$ (C_d x $C_d \to C_d$) the* ***merging function*** *in α and β (real functions) as follows:*

$$M_{\alpha\beta}(C_d1, C_d2) = < \varepsilon, (W, f)>$$

where

$$\varepsilon = \alpha(\varepsilon_1) \bullet \beta(\varepsilon_2) \qquad W = W_1 \cup W_2$$

and, for all w in W:

$$f(w) = \text{RFS}([0,1], \rho)$$

where, given

$$f_1(w) = ([0,1], \rho_1) \quad \text{and} \quad f_2(w) = ([0,1], \rho_2)$$

for all the points x in which both ρ_1 and ρ_2 are specified, we have

$$\rho(x) = \frac{\alpha(\varepsilon_1)\, \rho_1(x) + \beta(\varepsilon_2) \cdot \rho_2(x)}{\alpha(\varepsilon_1) + \beta(\varepsilon_2)}$$

while if only one ρ_j is specified we have $\rho(x) = \rho_j(x)$ and if both ρ_1 and ρ_2 are not specified the same happens to ρ.

This process merges the knowledge coming from different sources in a unique C_d structure. A major problem is how to minimise the information loss and, at the same time, focusing on the information that is, in some sense, more valuable (more reliable). The experience value ε is a good reference for the maturity of the information coded into a C_d but a number of external elements may affect the evaluation process. Therefore, we introduced the adjustment parameters α and β. For the binary operator ♦ we have a range of choices depending on the policies we enforce: simple solutions are +, *max* or *min*. We can use these parameters in order to enforce ageing policies, security policies or source selection policies and, in this sense, we suggest the possibility to take advantage of user profiling, per user or per class of users, for a comprehensive plan on the α, β to use in different situations.
The position of the merging function within the model becomes clearer looking at its applications and the more important is in combination with the distance function for the management of clusters. The *adaptive association* procedure manages the evolution of the "cases base".

Definition: *Given two case descriptions C_dS and C_dU for the diagnosis* d, *the adjustment functions α (for C_dS) and β (for C_dU) and two real values* μ_1 *and* μ_2, *considering $\delta = D_\sigma(C_dS, C_dU)$ we define the* ***adaptive association*** *process as follows:*

- *if the δ is less than the threshold* μ_1, *we associate* d *to the case C_dS*
- *if the δ is greater than* μ_1 *but smaller than* μ_2, *we can merge C_dS and C_dU using the merging function M with parameters α and β and we associate* d *to the result of the merge*
- *if the δ is greater than* μ_2, *we associate the* d *to both O_dS and O_dU*

This means that, first possibility, if I already have a case description that matches the context to which the diagnosis *d* refers to than *d* is a possible diagnosis of the problem. If the two case descriptions are different, second possibility, but there are no substantial differences we can build a case representing an average point in between them and associate *d* with this new case. If we notice, last possibility, that the cases have substantial differences, we keep them distinct and we associate the diagnosis *d* to both of them.

5 Experimental results

In order to test the effectiveness of our proposal, we developed a conversational CBR shell based on the model defined in the previous sections and we investigated the structure of the knowledge base on different situations. The shell interface (Figures 1) supports the user through a suggestion mechanism that, looking at the symptoms specified at one point, presents a list of possible other symptoms inferred by the knowledge base. When the user finds a diagnosis that matches his/her problem he/she can select it obtaining more details and indications on possible solutions. We focus on the possibility for the user to give indications on both relevance and confidence and how this information impacts on the precision of the result.

New case **Retrieval**

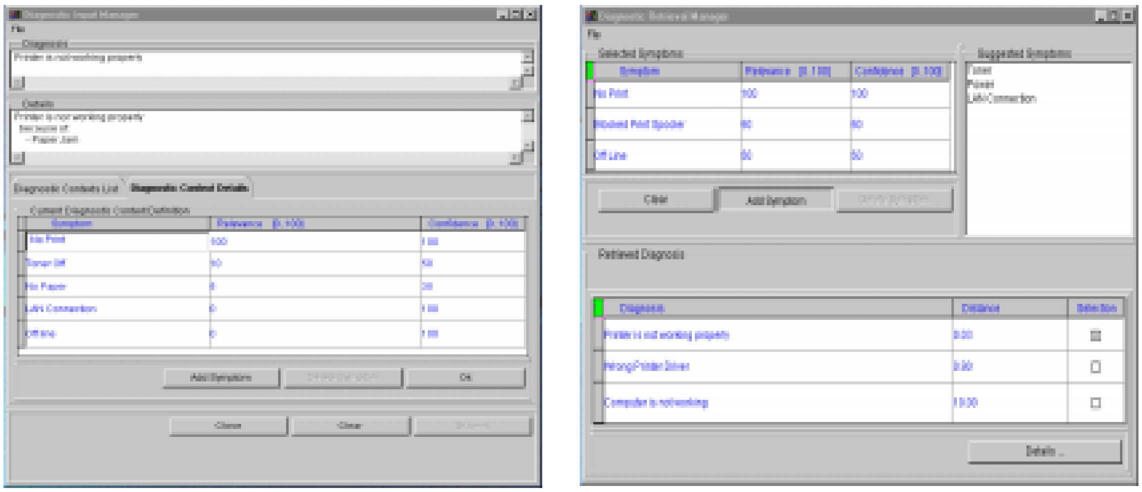

Figure 1: Conversational CBR shell

Concerning the characteristics of the knowledge base, adaptivity plays a major role and it may be interesting to look at the evolution and continuous refinement of a case descriptor. The graphs on (Figure 2) capture a series of snapshots of the confidence distribution over the relevance space for one fact in a case descriptor. The evolution line is from (a) to (e) and we can see how different points of view are managed. If, for example, we start (a) with a case in which “green LED” has relevance 3 with confidence 65%, we end up (e) with a more complete vision in which there is a strong confidence in the fact that the relevance of “green LED” is between 5 and 6 though also the range 2-3 has been reinforced. We work with a discrete version of the distribution function (relevance is on the x-axis and confidence on the y-axis) but we can tune the sensitivity of the function.

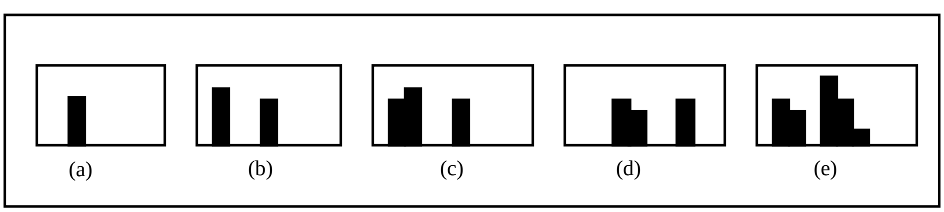

Figure 2: Evolution of component descriptor

In order to investigate the impact of explicit uncertainty management on the precision of the retrieval process, we tested the system under stress condition (up to 1000 diagnosis per case) looking at clustering problems. Having fixed the diagnostic data, we progressively reduced the sensitivity of the system to confidence information and we look at how the total number and the average dimension of clusters change. In (Figure

3) the x-axis represents the number of possible confidence levels while the y-axis represents (3.a) the average dimension of a cluster and (3.b) the total number of clusters. When we allow only one level of confidence we actually return to a standard system based only on relevance.
We notice that, on similar conditions, the number of diagnosis associated with what is modelled as a single case is up to 5 times greater in a standard system with respect to a reasonably sensitive system based the proposed model. This is reflected in the number of distinct clusters and in their average dimension.

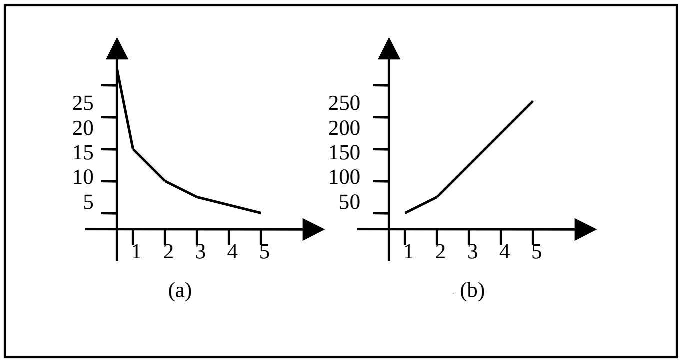

Figure 3: Clustering

In terms of retrieval it means that the NN algorithm can find more accurately the matching cases for the user description of a problem and supply him/her with a more restricted (but more precise) choice of diagnosis (solutions).

6 Discussion

In the previous section we focused on the benefits the model offers in term of selectivity (precision) in the case-retrieval process. This is fundamental for the ordinary activity of a diagnostic system but the explicit uncertainty modelling may have other applications in terms of knowledge-level information management.
Brachman and Levesque [5] define the knowledge-level information of a generic knowledge base (KB) as “the information KB offers about the world” that is, in our interpretation, information about the characteristic of the agents interacting with the KB. If a diagnostic system D is frequently asked, for example, to diagnose printer faults due to paper problems, the first thing we expect from D is it to have a lot of information concerning paper and printer. At the same time, if we notice that D is used mainly to diagnose printer problems related to paper we can infer that the users are not sufficiently informed on how to feed the paper into the printer or that a the printer actually have problem with a new type of paper. This kind of information may be useful for the diagnostic system in order to better understand the potential needs of its users and to give prompt replies but it is also a valuable feedback that may be used to improve the actual systems that D supports. In the example, either the paper manufacturer may be requested to modify the characteristics of the paper or the printer manufacturer may extend the user-manual section related to paper use.
Back to our model, the fact that confidence values for a symptom are low suggests that there isn’t a clear understanding of its meaning and it may need to be investigated more carefully. Looking at the confidence distribution of different symptoms of the same

case we can obtain indications on the reliability of the associated diagnosis proposals: if there is uncertainty on the causes of a problem (case) we may be more careful considering the proposed diagnosis.
Qualitative analysis of case descriptors may give indications on the system users, their needs and their problems. This extra layer of information provides a starting point for a more “user centred” diagnostic system where effectiveness derives not only from technological issues but also from a clearer understanding of the user (human or software agent).

8 Conclusions

The explicit modelling of uncertainty in diagnostic systems opens interesting possibilities in terms of both knowledge management and user interaction. We propose a model that, capitalising on the flexibility and expressiveness of fuzzy sets, captures both relevance and confidence aspects of the information related to diagnostic cases and that dynamically manages their evolution process.
We focus on an adaptive case-based (CBR) framework and tests prove that our model leads to an actual improvement in terms of case-selection precision with respect to standard relevance-based techniques. We also discuss the “knowledge-level” impact of our proposal and the benefits that may came from qualitative interpretations of the information gathered in the diagnostic system.

Bibliography

[1] D.W. Aha and L.A. Breslow. Refining Conversational Case libraries. In *Proc. 2nd International Conference on Case-Based Reasoning*. Providence, RI. Springer-Verlag, 1997.

[2] K.D. Althoff and S. Wess. Case-based reasoning and expert system development. In *Contemporary Knowledge Engineering and Cognition* (LNAI 622), Springer-Verlag 1991.

[3] A. Bookstein. Fuzzy requests: an approach to weighted Boolean searches. In *Journal of the American Society for Information Science*, Vol. 31, 1981.

[4] R. Bellman, R. Kalaba and L. Zadeh. Abstraction and pattern classification. In *Fuzzy Models for Pattern Recognition*, IEEE Press, 1992.

[5] R. Brachman and H. Levesque. The knowledge-level of a KMBS. In *On Knowledge Base Management Systems. Integrating Artificial Intelligence and Database Technology*. Springer-Verlag, 1986.

[6] C. Buckley, G. Salton and G.T. Yu. An evaluation of term dependence models in information retrieval. In *Research and Development in Information Retrieval (LNCS 146)*, Berlin, 1982.

[7] D.A. Buell and D.H. Kraft. Performance evaluation in a fuzzy retrieval system. In *Proc. of the 4th Int.Conf. on Information Retrieval*, Berkley, California, 1981.

[8] K. Hanney, M.T. Keane and P. Cunningham. What kind of adaptation do CBR systems need ? A review of current practices. In Adaptation of Knowledge for Reuse, *Proc. 1995 AAAI Fall Symposium*.

[9] Y.A. Cosmadopulos and R.W. Southwick. Usaing meta-level information for expert system control: a "blending" transformer approach. In *Research and Development in Expert Systems VI.* Proc. Expert System 89, Cambridge University Press, 1989.

[10] D. Janetzko and G. Strube . Case-based reasoning and model-based knowledge acquisition. In *Contemporary Knowledge Engineering and Cognition* (LNAI 622), Springer-Verlag 1991.

[11] A. Kaufmann. Introduction to the theory of fuzzy subsets. Academic Press, 1975.

[12] E.L. Rissland and J.J. Daniels. Using CBR to drive IR. In *Proc. 14th International Conference on Artificial Intelligence.* Vol. 1 (IJCAI-95). Montreal 1995.

[13] T. Radecki. Fuzzy set theoretical approach to document retrieval. In *Information Processing and Management*, Vol. 15, Pergamon Press, 1979.

[14] G. Salton and R.K. Waldstein. Term relevance weights in on-line information retrieval. In Information Processing and Management, Vol. 14, Pergamon Press, 1978.

[15] D. Wettschereck and D.W. Aha. Weighting Features. In *Proc. 1st International Conference on Case-Based Reasoning*. Lisbon, Portugal. Springer-Verlag 1995.

[16] L.A. Zadeh. Fuzzy sets. In *Information Control*, Vol. 8, Academic Press, 1965.

[17] L.A. Zadeh. The role of fuzzy logic in the management of uncertainty in expert systems. In *Fuzzy Set and Systems*, Vol. 11, North Holland, 1983.

[18] H.J. Zimmerman. Fuzzy set theory and its applications (2° Edition). Kluwer Academic Publishers, 1991.