Ef□ficient Multi-Stage Reasoning with Fuzzy Words

Martin Spott
Intelligent Systems Lab, BT
Adastral Park, Orion pp 1/12, Martlesham Heath
Ipswich, IP5 3RE, UK
martin.spott@bt.com

Abstract— **Fuzzy systems are a natural choice for processing coarse granular information, but unfortunately, most fuzzy systems suffer from two drawbacks. Although knowledge is formulated at a coarse granular level using fuzzy sets, most information processing algorithms operate on the details at a fine granular level and are, therefore, computationally costly. Furthermore, the fuzzy results are not expressed with the predefined fuzzy sets that were used to describe fuzzy knowledge in a comprehensive way, and are therefore often difficult to understand. As a solution to these problems, we proposed a methodology in [1], [2] that represents and processes fuzzy information at the coarse granular level. Here, we show that even a chain of inferences, i.e., multistage reasoning can be processed entirely at the coarse granular level.**

I. INTRODUCTION

In many applications, we need to abstract from detailed information in order to make processing of the information feasible, since the number of details is too large. In other cases, information is a priori an imprecise abstraction from details, because more precise information is not available. Furthermore, it is often easier to understand information on a higher level of abstraction. For instance, precise voltages of batteries are of no use for many people. An information like "the battery is *almost flat*" might be much more helpful.

We can classify the precision of information by its level of *granularity*. We call the most precise units of information *fine granular*. If we, for example, deal with heights of persons, then the most precise information units might be *cm* or *mm*, depending on the accuracy of the measuring device. However, a loss in precision leads to *coarse granular* information units, which might be collections of fine granular units, for example. In this paper, we distinguish only two levels of granularity, fine and coarse. Generally, whole hierarchies of granularity might be useful in applications. The user can dig into details, if he likes to.

Fuzzy systems have been invented for processing coarse granular information. Unfortunately, most of them suffer from a serious drawback: although information like a fuzzy rule "*rich* customers buy *many* bottles of champagne" is formulated at the coarse granular level, the information is still processed at the fine granular level. An example is the *max-min-composition* (or more generally, *compositional rule of inference*) as inference mechanism for rule-based fuzzy systems [3]. In the worst case, you have to calculate a maximum on the set of details (fine granular pieces of information). Since the number of details can be very large, think of the number of customers of a supermarket chain, these methods are useless, if your reason for abstraction is saving computational costs. We would have expected that the computational expense depends only on the number of coarse granular information units but not on the number of fine granular ones. Nevertheless, some special cases of the max-min-composition like the MAMDANI controller [4] are quite efficient processing techniques [5].

A second problem is related to the interpretability of processing results. Rules in rule-based systems are often defined by experts who use *linguistic values* like *tall* for the size of a person, i.e., a meaning is attached to a fuzzy set. We will call a linguistic value a *fuzzy word* in this paper. Methods like the max-min-composition then aggregate a fuzzy input and the fuzzy rules and calculate fuzzy sets as output that represent the imprecise result of the information processing. These fuzzy sets can often not be interpreted in terms of the fuzzy words the experts used to define the rules. Therefore, the interpretability of information at the coarse granular level is lost. This holds, for example, for the popular MAMDANI controller. There are methods, indeed, that interpret fuzzy sets in terms of predefined words (*linguistic approximation*), but they are usually separate from the processing procedure. This might result in additional computational expense or inconsistencies in semantics. In summary, it would be nice, if processing results are inherently expressed with the predefined fuzzy words, which would guarantee that the results are easy to understand.

Aiming at efficient information processing at the coarse granular level and at understandable fuzzy results, we proposed a new method to combine fuzzy words in [1]. The main idea is to represent fuzzy information at the coarse granular level as a combination of predefined fuzzy words. We will briefly explain the concept in Sections II, III. Section IV revisits the results presented in [2] on efficient reasoning with fuzzy words. Instead of processing arbitrary fuzzy sets, we always base processing on the fuzzy words and obtain results in the form of combined fuzzy words. Therefore, computational costs depend only on the number of fuzzy words and the results are easy to understand. We describe the concept of linear inference functions and show that rule-based reasoning can be implemented as such. Finally, Sect. V extends the approach to multi-stage reasoning. We prove that a sequence of inferences can be executed entirely at the coarse granular level. Furthermore, a sequence can be reduced to a single stage inference.

II. Fuzzy Partition—Fuzzy Words

As pointed out in the introduction, we aim at processing *coarse granular pieces of information*. Before we can discuss ways to process knowledge, we have to define, what a coarse granular piece of information is. The starting point in our approach is a universe $\mathcal{U}$: a set of basic elements as fine granular pieces of information. Information can only be expressed with the basic elements: "Peter is *190cm* tall" with *190cm* as basic element of the set of possible heights $[1cm, 300cm]$.

At a coarse granular level, we abstract from the basic elements and use a more abstract language to express information: "Peter is *tall*." We call a term like *tall* a *word*. Since words like *tall* might be inherently fuzzy, we will represent words by fuzzy sets on a universe $\mathcal{U}$ and call them *fuzzy words*. By abstraction, the set $\mathcal{U}$ of details will then be replaced by a set of fuzzy words as basic pieces of information that partition the entire universe in a fuzzy way.

For the definition of fuzzy words, we express fuzziness as a combination of different crisp definitions. A fuzzy partition is then formed by a collection of crisp partitions. For example, we may ask people to define crisp partitions {*very small, small, medium, tall, very tall*} of $\mathcal{U} = [1cm, 300cm]$ that describe possible heights of men as crisp sets. We then have several crisp definitions for each of the words *very small, small, medium, tall, very tall*.

For the construction of the respective fuzzy words, we use the *context model* [6], [7]. In this model, a source c of information is called a *context*. In our example, each person is a context. Each context c defines a word W as crisp set $W = \Gamma(c)$ via a function $\Gamma : \mathcal{C} \rightarrow 2^{\mathcal{U}}$. $\mathcal{C}$ is the set of contexts, $2^{\mathcal{U}}$ the set of all subsets of $\mathcal{U}$. A person c, e.g., might define *tall* $= \Gamma(c) = [185cm, 200cm]$. To each context c, a weight is attached. This weight is modelled by a probability $P(c)$; it measures the correctness, importance or reliability of an information $\Gamma(c)$ given by c [6]. Finally, a fuzzy word $\tilde{W}$ aggregates all definitions $W(c) = \Gamma(c)$. $\tilde{W}$ is defined by a fuzzy set $\mu_{\tilde{W}}$, i.e., for all $u \in \mathcal{U}$

$$\mu_{\tilde{W}}(u) = P(\{c \in \mathcal{C} \mid u \in \Gamma(c)\}) = \sum_{\substack{c \in \mathcal{C}: \\ u \in \Gamma(c)}} P(c) \qquad (1)$$

As an example, we asked 22 persons to define the heights *very small, small, medium, tall* and *very tall* of men. Since we considered all persons equally reliable, the probability of each piece of information $\Gamma(c)$ is $1/22$. Therefore, a membership value like $\mu_{\text{tall}}(u)$ is equivalent to the relative number of persons that declare a man of $u\,cm$ height as *tall*.

If we use the context model to form a set of fuzzy words $\mathcal{P} = \{\tilde{W}_1, \tilde{W}_2 \ldots \tilde{W}_n\}$ from a collection of crisp partitions, it is easy to show that $\mathcal{P}$ forms a *partition of unity* [8]: $\forall u \in \mathcal{U} : \sum_{W \in \mathcal{P}} \mu_{\tilde{W}}(u) = 1$. From here on, we call a partition of unity simply *fuzzy partition*.

III. Combining Fuzzy Words

Having abstracted from fine granular information, we have a set of words we can use to describe phenomena. For example, we can describe the height of a man with one of the words *very small, small, medium, tall* or *very tall*. A restriction to such a small number of possible statements, in this case five different heights, is not very expressive and therefore not very useful for applications.

In [1], we proposed a method called *combining fuzzy words* in order to dramatically increase the expressiveness of fuzzy words. It works in a similar way, as we combine words of our vocabulary in spoken languages to form new statements. We, for example, say that a man is "*tall* or of *medium height*" (simple combination), or it is "more likely that his height is *medium* than that he is *tall*" (weighted combination). We introduced a probabilistic model for the combination of fuzzy words, which is consistent with the context model that we use for the definition of fuzzy sets. $\mathcal{P} = \{\tilde{W}_1, \tilde{W}_2 \ldots \tilde{W}_n\}$ be a set of words that form a fuzzy partition of the universe $\mathcal{U}$. We define a probability measure $P_{\mathcal{P}}$ on $\mathcal{P}$, or more precisely, on $2^{\mathcal{P}}$, that models the weights of the words. The result are statements like "The height of the man is *medium* with $P_{\mathcal{P}}(\textit{medium})$ and he is *tall* with $P_{\mathcal{P}}(\textit{tall})$". More formally: A statement "$\forall \tilde{W} \in \mathcal{P} : \quad \tilde{x} = \tilde{W}$ with probability $P_{\mathcal{P}}(\tilde{W})$" is called a *weighted combination of the fuzzy words* $\tilde{W} \in \mathcal{P}$; $\tilde{x}$ is a *linguistic variable* like *man's height* that can hold fuzzy values. In the following, we often leave out the word *weighted* for reasons of simplicity and just say *combination of fuzzy words*.

A combination of fuzzy words is a fuzzy set and is represented by a membership function, as a simple fuzzy word is. We therefore have two different levels of presentation for fuzzy information: the *coarse granular* or *symbolic level* at which statements are expressed using fuzzy words as symbols and the *fine granular level* where we deal only with membership functions. In order to describe the relation of the fine and the coarse level we introduced two operators in [1]. The *synthesis* of a fuzzy set translates a symbolic combination of fuzzy words into the equivalent membership function, whereas the *analysis* of a fuzzy set transforms a membership function into a symbolic combination of predefined fuzzy words. In this way, the synthesis calculates the exact definition of a combination of fuzzy words, and the analysis explains a fuzzy set in terms of comprehensible fuzzy words. Figure 1 illustrates this relation. We showed in [1] that our approach allows to transform a combination of fuzzy words into a fuzzy set on the fine granular level and back into a combination of fuzzy words without losing information. This is generally not possible with most of the techniques based on the max-min-composition or *fuzzy production rules* like [9].

A. Synthesis Of A Fuzzy Set

As introduced above, a synthesis transforms a combination of fuzzy words, given by a probability measure $P_{\mathcal{P}}$ attached to a fuzzy partition $\mathcal{P}$, into a fuzzy set. In order to achieve this, we enrich the context model that is used to define the fuzzy words by $P_{\mathcal{P}}$. The basic idea is to adapt the context weights $P(c)$ for each fuzzy word $\tilde{W} \in \mathcal{P}$ with $P_{\mathcal{P}}(\tilde{W})$. This leads to

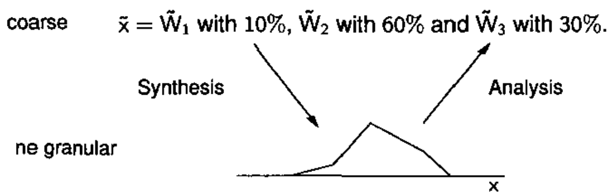

Fig. 1. Synthesis and Analysis of a fuzzy set

Theorem 1: Let $\mathcal{P}$ be a fuzzy partition of the universe $\mathcal{U}$ based on a set $\mathcal{C}$ of contexts with probability measure P on $2^{\mathcal{C}}$. For the fuzzy set $\tilde{A}$ that represents the weighted combination "$\forall \tilde{W} \in \mathcal{P}: \tilde{x} = \tilde{W}$ with probability $P_{\mathcal{P}}(\tilde{W})$" holds

$$\mu_{\tilde{A}}(u) = \sum_{\tilde{W} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{W})\, \mu_{\tilde{W}}(u) \tag{2}$$

The proof can be found in [1]. The theorem shows that the membership function $\mu_{\tilde{A}}$ of a weighted combination of fuzzy words $\tilde{W} \in \mathcal{P}$ can be computed directly from the membership functions $\mu_{\tilde{W}}$ of the words. Especially, it is not necessary to consider the contexts and their probabilities. Due to this result, we also write $\tilde{A} = \sum_{\tilde{W} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{W})\, \tilde{W}$ alternatively to (2).

B. Analysis Of A Fuzzy Set

The analysis of a fuzzy set $\tilde{A}$ calculates a combination of fuzzy words that describe $\tilde{A}$ in an understandable way. Generally, we can only expect that the combination is an approximation of $\tilde{A}$. This is the price we have to pay for the abstraction of knowledge. But the analysis should at least fulfil the following requirement:

$$\tilde{A} = \sum_{\tilde{W} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{W})\, \tilde{W} \Rightarrow \text{synthesis}(\text{analysis}(\tilde{A})) = \tilde{A} \tag{3}$$

In other words: If we analyse a fuzzy set that can be accurately represented by a combination of fuzzy words then the analysis must find this combination.

The starting point for an analysis is a fuzzy partition $\mathcal{P}$, i.e., a set of fuzzy words, and a fuzzy set $\tilde{A}$ that is to be described with the fuzzy words. Following (2), we are looking for a probability measure $P_{\mathcal{P}}$ such that the weighted combination of fuzzy words approximates $\tilde{A}$ as accurately as possible:

$$\mu_{\tilde{A}}(u) \approx \sum_{\tilde{W} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{W})\, \mu_{\tilde{W}}(u) \tag{4}$$

for all $u \in \mathcal{U}$. For now, we write $\alpha_{\tilde{W}}$ instead of $P_{\mathcal{P}}(\tilde{W})$ for the parameters we are looking for and make sure later on that they actually form a probability measure.

As measure for the quality of approximation, we use $Q = ||\mu_{\tilde{A}} - \sum_{\tilde{W} \in \mathcal{P}} \alpha_{\tilde{W}}\, \mu_{\tilde{W}}||_2$ with the norm $||f||_2 = \left(\int_a^b f(x)^2 dx\right)^{1/2}$ on the universe $[a, b]$. Alternatively, we can use a discrete version of the norm $|| \;\; ||_2$ for finite universes. The approximation method must determine a set of parameters $\alpha_{\tilde{W}}$ that minimise the quality function Q. We know from approximation theory that the solution to this optimisation problem is unique, if the membership functions $\mu_{\tilde{W}}$ are linearly independent. For the following reason, it is not a serious restriction to assume linear independency: If the membership functions of our fuzzy words are linearly dependent then at least one of the respective fuzzy words can be represented accurately by a weighted combination of the other fuzzy words. Therefore, we lose nothing, if we choose a linearly independent subset of membership functions, since we can always reconstruct the excluded functions. The optimal parameters $\alpha_{\tilde{W}}$ are calculated by solving a system of linear equations, cf standard books on numerical mathematics. The parameters $\alpha_{\tilde{W}}$ generally do not define a probability measure $P_{\mathcal{P}}$ on $2^{\mathcal{P}}$, since $\sum_{\tilde{W} \in \mathcal{P}} \alpha_{\tilde{W}} \neq 1$, and some of the $\alpha_{\tilde{W}}$ may be negative. A negative $\alpha_{\tilde{W}}$ can obviously not be interpreted as probability $P_{\mathcal{P}}(\tilde{W})$. As a solution, we propose the following procedure as analysis of $\tilde{A}$:

1) Determine the set M of fuzzy words, that overlap with $\tilde{A}$: $M := \{\tilde{W} \in \mathcal{P} \mid \text{supp}(\tilde{W}) \cap \text{supp}(\tilde{A}) \neq \emptyset\}$.
2) Optimise the weights $\alpha_{\tilde{W}}$ for all $\tilde{W} \in \mathcal{P}$.
3) If some of the $\alpha_{\tilde{W}}$ are negative, exclude the respective fuzzy words from M.
4) Define $\alpha_{\tilde{W}} = 0$ for all $\tilde{W} \notin M$ und normalise the weights $\alpha_{\tilde{W}}$ so that $\sum_{\tilde{W} \in \mathcal{P}} \alpha_{\tilde{W}} = 1$.
5) Define the probabilities $P_{\mathcal{P}}(\tilde{W})$

$$\forall \tilde{W} \in \mathcal{P} : P_{\mathcal{P}}(\tilde{W}) := \alpha_{\tilde{W}} \tag{5}$$

Hereby, the support of a fuzzy set $\tilde{A}$ is defined by $\text{supp}(\tilde{A}) = \{u \in \mathcal{U} \mid \mu_{\tilde{A}}(u) > 0\}$.

The proposed method indeed fulfils requirement (3):

Theorem 2: Let the fuzzy set $\tilde{A}$ be a combination of fuzzy words $\tilde{W} \in \mathcal{P}$ with linearly independent membership functions and weights $P'_{\mathcal{P}}(\tilde{W})$:

$$\forall u \in \mathcal{U} : \mu_{\tilde{A}}(u) = \sum_{\tilde{W} \in \mathcal{P}} P'_{\mathcal{P}}(\tilde{W})\, \mu_{\tilde{W}}(u) \tag{6}$$

Under these assumptions, for the proposed procedure for the analysis of $\tilde{A}$ holds

$$\forall \tilde{W} \in \mathcal{P} : P_{\mathcal{P}}(\tilde{W}) = P'_{\mathcal{P}}(\tilde{W}) \tag{7}$$

and, therefore, the analysis reconstructs the synthesised fuzzy set $\tilde{A}$.

The theorem can be proven by using the fact that the solution of the function approximation problem is unique [8]. It shows that the proposed analysing procedure of a fuzzy set $\tilde{A}$ finds a combination of fuzzy words that accurately describes $\tilde{A}$, if such a combination exists. If such a combination of fuzzy words does not exist, we generally end up with an approximation of the fuzzy set $\tilde{A}$. In both cases, the combination of fuzzy words is unique.

IV. Reasoning With Fuzzy Words

In the last sections, we showed how fuzzy information can be represented at the symbolic level. Information represented in this way is easy to understand and the representation is very compact. As already mentioned in the introduction, we also aim at information processing at the symbolic level in order to achieve comprehensible results with low computational costs.

Let us begin with a couple of terms frequently used in information processing: *knowledge base*, *observations*, *inference* and *conclusion*. The *knowledge base* comprises knowledge that is important and generally valid for the solution of the given problem. *Inference* is a procedure that draws *conclusions* or deduces information from the knowledge base. In the example of medical diagnosis, known relations between symptoms and causes form the knowledge base. Observed symptoms of patients are then used to infer possible causes. In this case, a conclusion (possible causes) is not drawn from the knowledge base alone, but from a combination of the knowledge base and an *observation*. The difference between an observation and an element of the knowledge base is that an observation is not generally valid. In the example, the observation depends on the patient.

The idea of information processing at the symbolic level is based on the assumption that important parts of the knowledge base are defined at the symbolic level, anyway. An example are fuzzy if-then rules like "if the person is *tall*, then it is *heavy*". Since *tall* and *heavy* are fuzzy words at the symbolic level, the rule is a relation at the symbolic level, as well. In most approaches for fuzzy rule-based reasoning like ZADEH'S *compositional rule of inference*, such a rule is transformed into a fuzzy relation at the fine granular level. A fuzzy observation of a person's height, given as a fuzzy set or fuzzy word, is then combined with the transformed rule and a fine granular conclusion will be drawn. By following such an approach, we obtain a computationally costly algorithm, since we operate at the fine granular level. The result is also at the fine granular level, i.e. not expressed with fuzzy words.

Instead, we propose to process information at the symbolic level. A fuzzy observation is either given at symbolic level, e.g., a person is *tall*, or at fine granular level as arbitrary fuzzy set. In the latter case, the fuzzy set will be analysed as described in Section III-B and in this way transferred onto the symbolic level. Then the symbolic observation and the symbolic knowledge base will be combined and a symbolic conclusion can be drawn.

A. Linear Inference Functions

Let us formalise the inference procedure and assume that an inference maps a combination of fuzzy words onto another combination of fuzzy words, thus it is a function $f_{\mathcal{K}} : \mathrm{Comb}(\mathcal{P}_{\mathcal{U}}) \rightarrow \mathrm{Comb}(\mathcal{P}_{\mathcal{V}})$. $\mathrm{Comb}(\mathcal{P}_{\mathcal{U}})$ and $\mathrm{Comb}(\mathcal{P}_{\mathcal{V}})$ denote the sets of all combinations of fuzzy words on universes $\mathcal{U}$ and $\mathcal{V}$. $f_{\mathcal{K}}$ also depends on the knowledge base $\mathcal{K}$. For reasons of simplicity we assume that the knowledge base does not change and just write f instead of $f_{\mathcal{K}}$.

Following the idea of the previous sections to represent fuzzy information as a combination of fuzzy words that form a partition of the underlying universe, we introduce an important subset of all inference functions that is *linear inference functions*. Being linear means that for all $\tilde{A}_i \in \mathcal{P}_{\mathcal{U}}$, $\alpha_i \in \mathbb{R}$ holds $f(\sum_i \alpha_i \tilde{A}_i) = \sum_i \alpha_i f(\tilde{A}_i)$. We will indeed introduce rule-based reasoning later on that can be expressed as a linear inference function.

The great thing about linear inference functions in the context of our theory is that the computational complexity of an inference is very low. Instead of calculating an inference $f(\tilde{A})$ for a combination $\tilde{A} \in \mathrm{Comb}(\mathcal{P}_{\mathcal{U}})$ from scratch, we can precalculate $f(\tilde{A}_i)$ for all fuzzy words $\tilde{A}_i \in \mathcal{P}_{\mathcal{U}}$. Due to the linearity of f, $f(\tilde{A})$ is just a linear combination of the precalculated $f(\tilde{A}_i)$. This is the simple but important

Theorem 3: Let $\mathcal{P}_{\mathcal{U}} = \{\tilde{A}_i\}_{i=1\ldots n}$ and $\mathcal{P}_{\mathcal{V}} = \{\tilde{B}_i\}_{i=1\ldots m}$ be sets of fuzzy words and $\tilde{A}$ a combination of fuzzy words $\sum_{i=1}^{n} \alpha_i \tilde{A}_i$. $f : \mathrm{Comb}(\mathcal{P}_{\mathcal{U}}) \rightarrow \mathrm{Comb}(\mathcal{P}_{\mathcal{V}})$ be a linear inference function with $f(\tilde{A}_i) := \sum_{j=1}^{m} \gamma_{ij} \tilde{B}_j$. Then $f(\tilde{A}) = \sum_{j=1}^{m} \beta_j \tilde{B}_j$ with $\beta_j = \sum_{i=1}^{n} \alpha_i \gamma_{ij}$.

For the proof, just use the linearity property of f. The theorem shows that the coefficients β_j we are looking for can be obtained by a simple vector-matrix-multiplication $\vec{b} = G^t \vec{a}$ with vectors $\vec{a} = (\alpha_i)_i$, $\vec{b} = (\beta_j)_j$ and matrix $G = (\gamma_{ij})_{ij}$.

Obviously, the computational complexity depends on the number of fuzzy words only and not on the number of details.

B. Rule-Based Reasoning

As an example of linear inference we introduced fuzzy rule-based reasoning in [2] that is based on Bayesian networks [10]. Each connection in a Bayesian network between nodes A and B represents a conditional probability $P(B|A)$. We can interpret $P(B|A)$ as a weighted if-then rule "if A then B with probability $P(B|A)$". Bringing this together with common fuzzy rules and our theory, we interpret a weighted fuzzy rule "if $\tilde{x} = \tilde{A}$, then $\tilde{y} = \tilde{B}$ with probability p" as conditional probability $P_{\mathcal{P}}(\tilde{B}|\tilde{A}) := P_{\mathcal{P}}(\tilde{y} = \tilde{B}|\tilde{x} = \tilde{A}) = p$ with $\mathcal{P} = \mathcal{P}_{\mathcal{U}} \times \mathcal{P}_{\mathcal{U}}$. Instead of the if-then phrase we write $\tilde{A} \stackrel{P_{\mathcal{P}}(\tilde{B}|\tilde{A})}{\longrightarrow} \tilde{B}$.

For the inference mechanism, we refer to probability theory:

$$P_{\mathcal{P}}(\tilde{B}_j) = \sum_i P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)\, P_{\mathcal{P}}(\tilde{A}_i) \tag{8}$$

which holds, since the fuzzy words $\tilde{A}_i$ are discrete elements of $\mathcal{P}_{\mathcal{U}}$ and $\bigcup_i \tilde{A}_i = \mathcal{P}_{\mathcal{U}}$. Using (8), an inference process that maps a fuzzy observation $\tilde{x} = \tilde{A}$ onto a fuzzy conclusion $\tilde{y} = \tilde{B}$, consists of three steps. First, we have to determine the probabilities $P_{\mathcal{P}}(\tilde{A}_i) = P_{\mathcal{P}}(\tilde{A}_i|\tilde{A})$. If $\tilde{A}$ is represented as a combination of fuzzy words $\tilde{A}_i$ like "the person is *tall*", we already know the probabilities. Otherwise, we analyse $\tilde{A}$ as described in Section III-B. In the second step, the probabilities $P_{\mathcal{P}}(\tilde{B}_j)$ are calculated using (8) and finally combined in conclusion $\tilde{B}$ using (2) (substitute $\tilde{W}$ with $\tilde{B}$).

Obviously, the inference formula (8) is a linear inference function as described in Section IV-A, i.e., the inference can be described as vector-matrix-multiplication $\vec{b} = G^t \vec{a}$ with vectors $\vec{a} = (P_{\mathcal{P}}(\tilde{A}_i))_i$, $\vec{b} = (P_{\mathcal{P}}(\tilde{B}_j))_j$ and matrix $G = (P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i))_{ij}$.

The conclusion of the inference is a combination of fuzzy words, i.e., it is easy to understand, and at the same time, it is well defined as a fuzzy set using (2). Furthermore, the *modus ponens* holds. That means that in case of a rule $\tilde{A}_i \stackrel{1}{\rightarrow} \tilde{B}$ in the rule set, an observation $\tilde{A}_i$ will be mapped onto

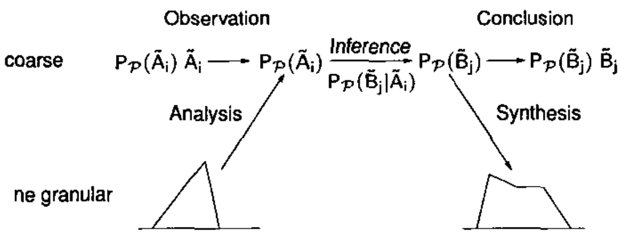

Fig. 2. Rule-based inference with fuzzy words

the conclusion $\tilde{B}$. This is not true for most fuzzy inference mechanisms, for example, MAMDANI'S popular approach. The problem is that if the modus ponens does not hold, the fuzziness of conclusions increases with every inference step. Therefore, after a couple of inferences, the level of fuzziness might be so large that the conclusion becomes meaningless. This will not happen with our approach.

V. MULTI-STAGE REASONING

Let us now consider a chain of n inferences. Assume linear inference functions $f_k : \mathrm{Comb}(\mathcal{P}_{\mathcal{U}_k}) \rightarrow \mathrm{Comb}(\mathcal{P}_{\mathcal{U}_{k+1}})$ and the chain of inferences $f_n(\ldots(f_2(f_1(\tilde{A}))))$. It is easy to see from Sect. IV-A that this can be represented by $\vec{b} = G^t_n \quad G^t_2 G^t_1\, \vec{a}$. Furthermore, we can transform this chain of inferences into a single-step inference by combining the matrices: $\vec{b} = G^t \vec{a}$ with $G^t = G^t_n \quad G^t_2 G^t_1$. In this way, the computational complexity can be further reduced, since the product of matrices can be calculated in advance.

In this simple case we assume that for each universe of discourse $\mathcal{U}_k$ there is only one partition of unity $\mathcal{P}_{\mathcal{U}_k}$, i.e., there is only one set of fuzzy words that can be used to represent information in each universe. In case of rule-based reasoning this means that the antecedents of rules in one rule-layer are based on the same fuzzy words as the consequents of the preceding rule-layer. Generally, this is not the case.

For the general case, we first only consider $n = 2$ for simplicity. Be the first inference step a mapping $f : \mathrm{Comb}(\mathcal{P}_{\mathcal{U}}) \rightarrow \mathrm{Comb}(\mathcal{P}_{\mathcal{V}})$, the second one a mapping $g : \mathrm{Comb}(\mathcal{P}'_{\mathcal{V}}) \rightarrow \mathrm{Comb}(\mathcal{P}_{\mathcal{W}})$. Thereby, we assume two different partitions $\mathcal{P}_{\mathcal{V}} = \{\tilde{B}_j\}_j$ and $\mathcal{P}'_{\mathcal{V}} = \{\tilde{B}'_k\}_k$ with fuzzy words $\tilde{B}_j$ and $\tilde{B}'_k$ that are used to describe information on the universe $\mathcal{V}$ as combinations in $\mathrm{Comb}(\mathcal{P}_{\mathcal{V}})$ and $\mathrm{Comb}(\mathcal{P}'_{\mathcal{V}})$. If we think of rule-based reasoning, this means that the rules in the first inference step use different fuzzy words in the consequences than the words we find in the antecedents of the rules in the second inference step.

This general setting is shown in Fig. 3. An observation is given as fuzzy set, which will be transformed into a combination of fuzzy words by analysis. Alternatively, the observation is a combination by nature. The first inference step maps the observation onto a combination

$$\tilde{B} = \sum_j \beta_j \tilde{B}_j \tag{9}$$

of $\tilde{B}_j$ as intermediate conclusion $\tilde{B}$. In order to apply the inference function g as second inference step, the intermediate conclusion needs to be expressed as combination $\sum_k \beta'_k \tilde{B}'_k$ with fuzzy words $\tilde{B}'_k$. The weights are obtained from an analysis, which is costly since it operates at the fine granular level. Then, the second inference step can be executed.

In order to avoid the costly analysis step before applying g we found a transformation as shown in Fig. 3 that directly transfers the representation by a combination of the $\tilde{B}_j$ into one of the $\tilde{B}'_k$. We show that the transformation mapping t is a linear inference function and how to define it by determining $t(\tilde{B}_j)$ for all j. We prove that the transformation leads to the same result as the sequence of synthesis and analysis.

In order to achieve this, we approximate each $\tilde{B}_j$ by a combination of $\tilde{B}'_k$:

$$\mu_{\tilde{B}_j}(v) \approx \sum_k w_{jk}\, \mu_{\tilde{B}'_k}(v) \tag{10}$$

and define $t(\tilde{B}_j) := \sum_k w_{jk}\, \tilde{B}'_k$. t is obviously linear. The weights w_{jk} can be determined using the function approximation approach described in Sect. III-B. We only apply step 2 of the 5-step procedure, i.e., pure function approximation. The following theorem proves that the transformation is identical to the sequence of synthesis and analysis as shown in Fig. 3.

Theorem 5.1: w_{jk} be the optimal weights in (10) and β'_k the weights we obtain through the analysis of the intermediate conclusion $\tilde{B}$ in Fig. 3 using the 5-step procedure. Let $I := \{k \mid \mathrm{supp}(\tilde{B}'_k) \cap \mathrm{supp}(\tilde{B}) \neq \emptyset\}$. If the membership functions of $\tilde{B}_j \in \mathcal{P}_{\mathcal{V}}$ and $\tilde{B}'_k \in \mathcal{P}'_{\mathcal{V}}$ are continuous, then for all k holds

$$\beta'_k = \begin{cases} \dfrac{\max\{0, \sum_j w_{jk}\, \beta_j\}}{\sum_{i \in I} \max\{0, \sum_j w_{ji}\, \beta_j\}}, & \text{if } k \in I \\ 0 & \text{else} \end{cases} \tag{11}$$

Proof: We assume a Pre-Hilbert Space $(C(\mathcal{V}), \|\ \|_2)$ with the space $C(\mathcal{V})$ of continuous functions on $\mathcal{V}$ and inner product $\langle f, f\rangle^{1/2} = \|f\|_2$. In (10) we computed optimal approximations of the fuzzy words $\tilde{B}_j$. According to approximation theory therefore holds for all j, l

$$\sum_k w_{jk}\, \langle \mu_{\tilde{B}'_k}, \mu_{\tilde{B}'_l}\rangle = \langle \mu_{\tilde{B}_j}, \mu_{\tilde{B}'_l}\rangle \tag{12}$$

For all l, $\alpha_k := \sum_j w_{jk}\, \beta_j$ is solution to the so-called *normal equation*:

$$\sum_k \sum_j w_{jk}\, \beta_j\, \langle \mu_{\tilde{B}'_k}, \mu_{\tilde{B}'_l}\rangle \tag{13}$$

$$= \sum_j \beta_j \sum_k w_{jk}\, \langle \mu_{\tilde{B}'_k}, \mu_{\tilde{B}'_l}\rangle \tag{14}$$

$$\overset{(12)}{=} \sum_j \beta_j\, \langle \mu_{\tilde{B}_j}, \mu_{\tilde{B}'_l}\rangle \tag{15}$$

$$\overset{(9)}{=} \langle \mu_{\tilde{B}}, \mu_{\tilde{B}'_l}\rangle \tag{16}$$

We know from approximation theory that, therefore, the α_k minimise the quality measure $Q = \|\mu_{\tilde{B}} \quad \sum_k \alpha_k\, \mu_{\tilde{B}'_k}\|_2$. The solution is unique, since we assumed linear independence of membership functions. This proves that we computed the best approximation of $\tilde{B}$ with fuzzy words $\tilde{B}'_k$, which is step 2 of the 5-step procedure for obtaining the β'_k. Steps 3–6 are incorporated in (11). ∎

The theorem shows that we do not need to analyse the intermediate conclusion $\tilde{B}$, but instead apply the linear transformation

$$t(\tilde{B}) = t(\sum_j \beta_j \tilde{B}_j) = \sum_j \beta_j t(\tilde{B}_j) = \sum_k \beta'_k \tilde{B}'_k \qquad (17)$$

with β'_k being defined in (11). Being a linear transformation, t can be represented as matrix $T = (w_{jk})_{jk}$ and the overall inference $\tilde{C} = g(t(f(\tilde{A})))$ can be written as $\vec{c} = G^t\, T^t\, F^t\, \vec{a} = H^t\, \vec{a}$ with $H = F\,T\,G$.

To conclude, we see that multi-stage inference with fuzzy words can be achieved entirely at the coarse granular level even in the general case of different partitions of each universe.

Regarding computational complexity, our approach is a least as good as others and outperforms most of them. Let us assume an n-stage inference, discrete universes with m elements each, and partitions with k fuzzy words each. Considering that we can pre-calculate the transformation matrix T, the complexity of our approach is $O(k^2 n)$ as for MAMDANI'S method. If we accumulate the matrices of the inference functions, our complexity can be reduced to $O(k^2)$. The general compositional rule of inference, on the other hand, requires $O(m^2 n)$, whereby m is usually much greater than k. In case, an initial analysis step of a fuzzy observation is necessary, we realistically require $O(m)$ for that. Thereby, we take into account that usually only two or three neighbouring fuzzy words overlap in partitions and the function approximation therefore only operates on sparse matrices. However, in most applications, observations are either crisp or directly given as combinations of fuzzy words, so an analysis is not necessary and its relatively high costs therefore do not compromise the usefulness of the approach.

VI. CONCLUSIONS

The implications of the results in previous sections are extremely interesting with respect to our goals of processing fuzzy information, which are computational efficiency and interpretability of information. Existing approaches process fuzzy information at the fine granular level; we, instead, transform information onto the coarse granular level, process it there, and transform it back, if necessary. Thereby, the representations at fine and coarse granular level are isomorphic.

We showed that, in this way, a very efficient multi-stage fuzzy rule-based reasoning mechanism can be realised. Apart from an initial transformation step onto the coarse granular level, all operations are done at the coarse granular level and therefore, the computational complexity depends on the number of fuzzy words and not on the number of details. Fuzzy conclusions drawn by the system are represented as combinations of fuzzy words and are, therefore, easy to understand.

An interesting topic for future research is the question if there are linear inference functions other than the rule-based approach that could benefit from these results with respect to comprehensibility of conclusions and computational complexity.

REFERENCES

[1] M. Spott, "Combining fuzzy words," in *Proc. of FUZZ-IEEE 2001*, Melbourne, Australia, 2001.

[2] ——, "Efficient reasoning with fuzzy words," in *Proc. of FSKD 2002*, Singapore, 2002.

[3] L. A. Zadeh, "A theory of approximate reasoning," in *Machine Intelligence*, J. Hayes, D. Mitchie, and L. Mikulich, Eds. New York: Wiley, 1979, pp. 149–194.

[4] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Intern. Journal of Man–Machine Studies*, vol. 7, 1975.

[5] H.-P. Chen and T.-M. Parng, "A new approach of multi-stage fuzzy logic inference," *Fuzzy Sets and Systems*, vol. 78, pp. 51–72, 1996.

[6] J. Gebhardt and R. Kruse, "The context model—an integrating view of vagueness and uncertainty," *Intern. Journal of Approximate Reasoning*, vol. 9, pp. 283–314, 1993.

[7] C. Borgelt, "Data mining with graphical models," Ph.D. dissertation, Otto-von-Guericke-Universität Magdeburg, 2000.

[8] M. Spott, "Schließen mit unscharfen begriffen (reasoning with fuzzy terms)," Ph.D. dissertation, Universität Karlsruhe, Germany, 2000.

[9] D. S. Yeung and E. C. C. Tsang, "Weighted fuzzy production rules," *Fuzzy Sets and Systems*, vol. 88, pp. 299–313, 1997.

[10] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of pausible inference*. Morgan Kaufmann, 1988.

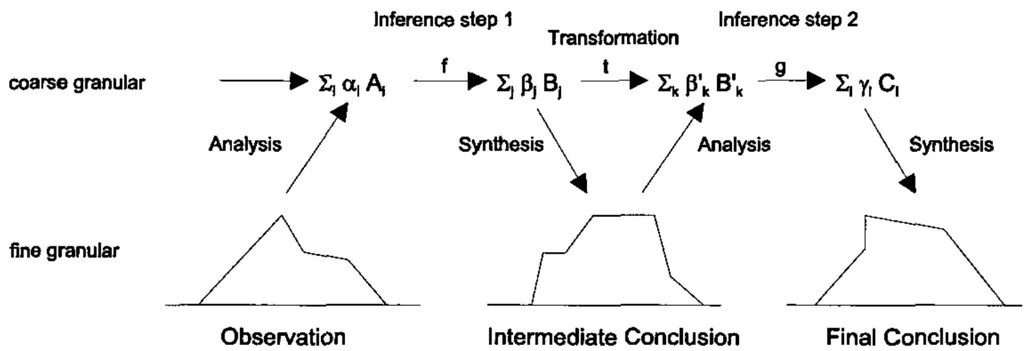

Fig. 3. Multi-stage reasoning with fuzzy words