Determination of Adequate Fuzzy Inequalities for their Usage in Fuzzy Query Languages

Marcel Shirvanian and Wolfram Lippe

Abstract—Although the usefulness of fuzzy databases has been pointed out in several works, they are not fully developed in numerous domains. A task that is mostly disregarded and which is the topic of this paper is the determination of suitable inequalities for fuzzy sets in fuzzy query languages. This paper examines which kinds of fuzzy inequalities exist at all. Afterwards, different procedures are presented that appear theoretically appropriate. By being applied to various examples, their strengths and weaknesses are revealed. Furthermore, an algorithm for an efficient computation of the selected fuzzy inequality is shown.

Keywords—Fuzzy Databases, Fuzzy Inequalities, Fuzzy Query Languages, Fuzzy Ranking.

I. INTRODUCTION

ALTHOUGH the usefulness of fuzzy databases has been pointed out in several works, they are not fully developed in numerous domains. A central aspect of databases is the query language with which the user can retrieve or modify data. Various specifications of fuzzy query languages exist but they are not technically mature. One of the drawbacks results from the extensive disregard of the question which inequalities are suitable. Since for many operations of a fuzzy query language, for example a selection or a join, not only equality constraints but also inequality constraints are utilized, appropriate inequalities are essential. Therefore, they are examined in this paper.

In contrast to real numbers, it is not clear whether a fuzzy set is greater or less than another one so that adequate procedures for this kind of determination must be found. To be more precise, measures are to be utilized that determine the degree to which a fuzzy set is greater than another one. In the following, these measures are denoted as fuzzy inequalities.

The remaining paper is organized as follows. Some of the present fuzzy inequalities in fuzzy query languages are described in subsection II.A. The next subsection deals with the theory of fuzzy ranking whose methods potentially can be used as fuzzy inequalities. The identification of theoretically adequate functions is handled in section III. In subsection IV.A these measures are applied to various examples in order to evaluate their quality on the basis of their results. An efficient computation of the best possible option is presented in subsection IV.B. Finally, a short conclusion is drawn in section V.

II. RELATED WORK

A. Present Fuzzy Inequalities in Fuzzy Query Languages

The majority of the analyses of fuzzy databases refer to the relational variant. Thus, it is not remarkable that also most fuzzy query languages are designed for fuzzy relational databases. Equivalent to the traditional relational query languages, fuzzy relational query languages are predominantly characterized by fuzzy relational algebras. Some of these algebras are specified in [6], [7], [10], [14], [34], [37], [40], [45], [46], [52]. Moreover, there are implementations of concrete languages, like the variants presented in [26], [51], [61].

The topic of fuzzy inequalities, however, is either ignored or only mentioned marginally in most of the works listed above so that almost no sufficient suggestions are made. An exception is represented by the fuzzy query language FSQL which is specified in [26]. The language provides four methods for the computation of the degree of the possibility and the necessity respectively to which a value of the fuzzy set is greater or less than another one so that adequate procedures are examined in this paper.

It has to be noted that these four measures match the methods introduced in [22] if a limitation on trapezoids is made. One of the referenced measures is defined by (5) and
produces the same results as (1) if trapezoids are used. The function $\mu$ returns the membership degree of the given fuzzy set and the given value; $\Omega$ is the universe of discourse.

$$PD(a, b) = \sup \min \{\mu_a(x), \mu_b(y)\}$$ (5)

It is obvious that (5) generates the extremal values if and only if the first or the last case of (1) holds. Evaluating the remaining situation, the output of (5) is the membership degree of the intersection of the two functions. This value must be used in any of these two functions in order to calculate the membership degree.

$$\frac{1}{\epsilon_a - \delta_a} x + \frac{\epsilon_a}{\epsilon_a - \delta_a} = \frac{1}{\gamma_b - \beta_b} x - \frac{\beta_b}{\gamma_b - \beta_b}$$

$$\Rightarrow \epsilon_a(\gamma_b - \beta_b) + \beta_b(\epsilon_a - \delta_a) = (\gamma_b - \beta_b)(\epsilon_a - \delta_a)$$

$$\Rightarrow x = \frac{\epsilon_a(\gamma_b - \beta_b) + \beta_b(\epsilon_a - \delta_a)}{\gamma_b - \beta_b + \epsilon_a - \delta_a}$$

$$= \frac{1}{\gamma_b - \beta_b} x - \frac{\beta_b}{\gamma_b - \beta_b}$$

The result matches the second case of (1) whereby the correlation between (1) and (5) is shown. The equality of the other measures can only inadequately be interpreted concerning some scenarios. On the other hand, the results of the four methods can only inadequately be interpreted concerning particular scenarios which will be demonstrated in subsection IV.A. Thus, it is important to study measures that do not belong to different comparison media.

B. Fuzzy Ranking

Interestingly, there are almost no papers which deal with the issue of fuzzy inequalities in details so that related analyses must be regarded. A suitable topic is the ranking of fuzzy sets because ranking two fuzzy sets, a fuzzy inequality is implicitly described.

Since, however, the goals of these two topics do not have to be identical, it is necessary to determine which kinds of fuzzy rankings can be utilized as fuzzy inequalities. Different attempts were undertaken to classify various rankings and to explore - partly by applying them to some examples - their strengths and weaknesses [5], [17], [23], [42], [53]–[55], [62]. The most detailed work is [17] whose classification, which is illustrated in the Fig. 1, is examined now. Some of the methods that are indicated in this figure were not part of the original paper because mostly they were published after this classification.

Fig. 1. A Taxonomy of Fuzzy Ranking Methods

Apart from the approaches above, there are some methods that cannot be assigned to a single technique. In [11] a universal method is shown which can be transformed into particular fuzzy rankings of the class $\alpha$-cut as well as the class area measurement. Otherwise, the variant presented in [20] contains elements of the centroid index and the area measurement. Moreover, the fuzzy ranking in [32] is based on the probability distribution and the area measurement which belong to different comparison media.

The concept of the linguistic expression can be ignored because thereby more than one value is produced. But fuzzy inequalities must calculate only a single membership degree, subject to the two given fuzzy sets.
The techniques probability distribution, left/right scores, centroid index and area measurement also have to be rejected because they include the distance between the fuzzy sets in their computations. Concerning the question whether a real number is greater than one, it is irrelevant if they are close-by or far away from each other. Thus, fuzzy inequalities should generate the maximum - 1 - or the minimum value - 0 - if and only if the supports of the two fuzzy sets are disjoint. The ranking methods of the remaining techniques must be analyzed individually in order to decide whether they are suitable as fuzzy inequalities.

Some characteristics for a fuzzy ranking method are proposed in [39]. The most important one for a fuzzy inequality $FI(a, b)$, that determines the degree to which the fuzzy set $a$ is greater than the fuzzy set $b$, is the fuzzy reciprocal whereby $FI(a, b) + FI(b, a) = 1$ must hold. Consequently, no independent calculation of the degree to which a fuzzy set is less than another one is necessary.

Before miscellaneous fuzzy inequalities are presented in the next section, it has to be mentioned that the fuzzy reciprocal holds for (2) and (3), but not (1) and (4). Therefore, the two last-mentioned variants should be neglected. But $FGEQ(a, b) + NFGT(b, a) = 1$ holds so that these measures are acceptable.

III. SPECIFICATION OF FUZZY INEQUALITIES

The approaches (1) to (4) belong to the technique comparison function. Alternative methods of this class have not to be considered because all of them can be seen as special cases of these four fuzzy inequalities. Hence, four techniques remain, namely degree of optimality, $\alpha$-cut, Hamming distance and proportion to optimal, whose fuzzy rankings must be examined in order to determine appropriate candidates.

A. Degree of Optimality

The procedures of the class degree of optimality rank a group of fuzzy sets by checking them against the greatest one of them. Most of these methods appear to be plain or inadequate. By contrast, the function declared in [12] evaluates two fuzzy sets by comparing all elements of $\Omega$ with each other. In doing so, an element of $a$ which is greater than an element of $b$ receives the degree 1. Conversely, the value 0 and in case of equality, the value 0.5 have to be assigned. The computation, which does not have to include the part with the factor 0, is defined as follows.

$$FIM(a, b) = \frac{\sum_{y \in \Omega} \min(\mu_a(x), \mu_b(y))}{\sum_{y \in \Omega} \sum_{x \in \Omega} \min(\mu_a(x), \mu_b(y))}$$

(6)

In theory, the sums have to be replaced by integrals if two continuous fuzzy sets are evaluated. But concerning the following analysis, (6) generates suitable results so that no integrals have to be computed. In addition, in [12] it is also recommended to perform an approximate calculation.

B. $\alpha$-Cut

The technique $\alpha$-cut, whose methods are regarded in this subsection, determines the ranking by means of some cuts of the fuzzy sets to the degree $\alpha$. Most of the procedures displayed before use only one particular $\alpha$-cut so that important information are ignored. The only fuzzy ranking that includes a sufficiently great amount of $\alpha$-cuts is specified in [35] and is presented now.

In order to determine the degree to which the fuzzy set $a$ is greater than $b$, the fuzzy set $c$ must be generated by means of (7). The not yet normalized value for an $\alpha$-cut arises by using (8).

$$c = \{ (\mu_a - b(z)/z) \in \Omega \}$$

with $\mu_a - b(z) = \sup_{z, y \in \Omega} \min(\mu_a(x), \mu_b(y))$

(7)

$$J_{a, b}(\alpha) = \begin{cases} 1 & \text{if } s = i \geq 0 \\ 0 & \text{if } s = i = 0 \\ -1 & \text{if } s \geq 0 \end{cases}$$

(8)

with $J_{int} = \frac{\max(s, 0) - \max(-i, 0)}{\max(s, 0) + \max(-i, 0)}$

and $s = \sup_{d \in [\epsilon, \eta]} d$ and $i = \inf_{d \in [\epsilon, \eta]} d$.

Now the values of (8) for the particular degrees of $\alpha$ must be merged. But the regular formula is too complex so that in [35] an approximation is proposed which is also utilized here. In the same paper an accurate computation for trapezoids as fuzzy sets is indicated which, however, can be ignored. The reasons for it are that the calculation is very complicated and sufficiently precise values result from the approximation defined by (9). The accuracy of the output is improved by increasing the value of the natural number $N$.

$$FIM(a, b) = \frac{2}{N^2} \sum_{n=0}^{N} n * J_{a, b}(\frac{n}{N}) - \frac{2}{N} J_{a, b}(1) + 1$$

(9)

C. Hamming Distance

The next fuzzy inequalities to be presented work with the Hamming distance of two fuzzy sets which is described by (10).

$$H(a, b) = \int_{-\infty}^{\infty} |\mu_a(x) - \mu_b(x)| \, dx$$

(10)

Before declaring the concrete approaches, some calculations are shown that are used by the subsequent fuzzy rankings. The fuzzy minimum is specified by (11), the fuzzy maximum by (12), the greatest upper set by (13) and the greatest lower set by (14).

$$\min(a, b) = \{ \mu_{\min}(z)/z \in \Omega \}$$

with $\mu_{\min}(z) = \sup_{z, y \in \Omega} \min(\mu_a(x), \mu_b(y))$

(11)
\[
\tilde{\mu}_{\text{max}}(z) = \sup_{x, y \in \Omega} \min(\mu_a(x), \mu_b(y))
\]

(12)

with

\[
\mu_{\text{max}}(z) = \sup_{x, y \in \Omega} \min(\mu_a(x), \mu_b(y))
\]

\[
\varrho = \left\{ \sup_{y \leq x} \mu_a(y) \mid x \in \Omega \right\}
\]

(13)

\[
\tau = \left\{ \sup_{y \leq x} \mu_a(y) \mid x \in \Omega \right\}
\]

(14)

Equivalent to the previous techniques, the class hamming distance also contains only a few fuzzy rankings which meet the requirements indicated before. One of them is illustrated in [39] which, however, can be disregarded because drastic deficits were revealed in [30]. Due to these disadvantages, three measures were developed in the last-mentioned work, two of which potentially can be used as fuzzy inequalities. They are defined by (15) and (16). The minimum is utilized as the t-norm for the computation of the intersection.

\[
FI_{K1}(a, b) = \frac{H(a \cap b, 0) + H(b, \tilde{\mu}_{\text{max}}(a, b))}{H(a, 0) + H(b, 0)}
\]

(15)

\[
FI_{K2}(a, b) = \frac{H(a \cap b, 0) + H(b, \tilde{\mu}_{\text{max}}(a, b)) + H(\varrho, \tilde{\mu}_{\text{max}}(\varrho, \tau))}{2H(a \cap b, 0) + H(\varrho, \varrho) + H(\tau, \tau)}
\]

(16)

Although these two measures appear relatively diverse at first sight, their results vary only minimally from each other which will be demonstrated later. Nevertheless, both procedures are analyzed because this difference will be pointed out to be significant. But these two formulas have a shortcoming, that is, a division by zero can take place which must be avoided. By applying (15), this happens if \(a\) and \(b\) are crisp sets whereas concerning (16), the two sets additionally must be identical. The values 0, 0.5 and 1 have to be produced for these particular scenarios.

D. Proportion to Optimal

The last class - proportion to optimal - shares properties with the technique degree of optimality. But the optimum now arises out of the given fuzzy sets. The method introduced in [36] is not presented here because it is a special case of the fuzzy ranking declared in [44]. The approach in [44] does not describe a concrete fuzzy ranking, but a group of procedures. In each case, the optima are specified by the fuzzy maximum and the fuzzy minimum because classical inequalities can be evaluated by means of the maximum and the minimum. The degree to which \(a\) is greater than \(b\) is computed by (17). The components \(\lambda\) and \(S\) are defined next.

\[
\mu_{SC}(a, b) = \lambda(S(a, \tilde{\mu}_{\text{max}}(a, b)), S(b, \tilde{\mu}_{\text{min}}(a, b)))
\]

(17)

\(S\) is a similarity measure and thus acquires the similarity of \(a\) to the fuzzy maximum and the similarity of \(b\) to the fuzzy minimum. Since the issue of similarity measures will not be addressed in this work, solely the measure is utilized that is named in [44] in the first place and which is defined by (18). Moreover, different kinds of measures, for example inclusion measures, can be applied which, however, are neglected here as well.

\[
S_1(a, b) = \int_{x \in \Omega} \min(\mu_a(x), \mu_b(x)) \, dx
\]

(18)

According to [44], \(\lambda\) is a function in which, amongst others, the arithmetic mean or any t-norm can be used. But if the similarity measure (18) is selected and a t-norm is applied on a comparison between a continuous fuzzy set and a crisp set, the output of (17) is always 0. Therefore, the arithmetic mean is utilized exclusively in the following which leads to (19).

\[
\mu_{SC}(a, b) = \frac{1}{2}(S_1(a, \tilde{\mu}_{\text{max}}(a, b)) + S_1(b, \tilde{\mu}_{\text{min}}(a, b)))
\]

(19)

Unfortunately, the fuzzy reciprocal does not hold for this formula. Hence, [44] recommends to include also the result with swapped arguments into the fuzzy ranking. In order to obtain a fuzzy inequality that fulfills all of the requirements outlined before, the ratio of both terms must be calculated. The approach, which together with the other presented methods is analyzed in the next subsection on the basis of examples, is specified by (20). In doing so, it is necessary to pay attention that the denominator does not become 0. This event only arises if two crisp sets are compared with each other.

\[
FI_{SC}(a, b) = \frac{\mu_{SC}(a, b)}{\mu_{SC}(a, b) + \mu_{SC}(b, a)}
\]

(20)

IV. DETERMINATION OF ADEQUATE FUZZY INEQUALITIES

A. Comparison of Fuzzy Inequalities

Concerning the determination of meaningful scenarios for the fuzzy inequalities, the analyses listed in subsection II.B are hardly helpful because thereby mostly comparisons with more than two fuzzy sets are examined. Only in [62] some useful examples are given. Consequently, a new collection of scenarios is used now which is observable in the Fig. 2. Naturally, fuzzy sets with disjoint supports do not need to be taken into account. The fuzzy set marked by the continuous line represents in each case the first argument \((a)\) whereas the other fuzzy set illustrates the second argument \((b)\). Since the fuzzy reciprocal holds for all measures - except (1) and (4) whose relationship has already been declared - , it is not necessary to apply the fuzzy inequalities with reversed arguments.

It was mentioned before that the methods described by (1) to (4) generate no informative values in some cases. Next, this is going to be demonstrated. For this purpose, the data that result from the application of these functions to the examples must be analyzed. The degrees which are displayed in the Table I are rounded after the fifth decimal place.

It is evident that these four methods together produce heterogeneous and meaningful values in many cases, for example in III, VI and IX. But particularly for the last scenario, in which a fuzzy set is compared with a crisp set, only minimum and maximum values are created. Therefore, the remaining fuzzy inequalities are now applied in order to discover whether thereby advantages are obtained.
presented here - $FI_{CT}$ - it is clear that appropriate data exist for many situations. But there are also scenarios in which the output of this method is quite questionable, in particular No. V. Although it appears to be that $a$ is less than $b$, the disproportionately small degree does not express the situation adequately because, for example, $a$ has the greater maximum value.

By contrast, $FI_M$ produces a very great value for the same example because the zones with high membership degrees are weighted too heavily. This especially has an impact on the scenarios VI and VII. In the first case, the fuzzy set $a$ is slightly greater than $b$ but its maximum value is somewhat less than the one of $b$. Accordingly, all other fuzzy inequalities generate a result greater than 0.5. Instead of that, $FI_M$ clearly declares $b$ as the greater fuzzy set. The result for the second case is close to the maximum value so that this scenario is also characterized inappropriate. Therefore, the utilization of $FI_M$ cannot be recommended.

The degrees of the remaining methods differ only minimally from each other or even are partly identical so that $FI_{K1}$ can be used as a reference. To put it simply, this approach computes the sum of the intersection area and of the zones in which $a$ dominates $b$. Then this intermediary result is divided by the sum of the areas of both fuzzy sets. These data seem to be consistent with respect to the examined examples. In [49] it is claimed that this measure represents the best choice if a preference relation is desired. The fuzzy ranking $FI_{K2}$, however, is not named in that paper so that it might be useful to determine the difference between these two methods.

The output of $FI_{K2}$ differs from the one of $FI_{K1}$ if and only if the kernels of both fuzzy sets do not intersect. The reason for it is that additionally the area between the two kernels and the intersection becomes a part of the computation of $FI_{K2}$. Thus, equivalent to $FI_M$, the zones with high mem-

\begin{table}[h]
\centering
\caption{Results of the Proposed Fuzzy Inequalities}
\begin{tabular}{|c|c|c|c|c|}
\hline
No. & $FI_{CT}$ & $FI_M$ & $FI_{K1}$ & $FI_{K2}$ & $FI SC$
\hline
I & 0.02315 & 0.0187 & 0.07143 & 0.0625 & 0.07143
\hline
II & 0.33881 & 0.35859 & 0.4 & 0.4 & 0.4
\hline
III & 0.41877 & 0.31114 & 0.41861 & 0.41861 & 0.41861
\hline
IV & 0.5492 & 0.70593 & 0.55 & 0.55495 & 0.55
\hline
V & 0.12412 & 0.447 & 0.2702 & 0.27387 & 0.26935
\hline
VI & 0.59805 & 0.40553 & 0.53519 & 0.53231 & 0.53522
\hline
VII & 0.66528 & 0.50659 & 0.69441 & 0.75506 & 0.7018
\hline
VIII & 0.72494 & 0.86832 & 0.70395 & 0.72222 & 0.70431
\hline
IX & 0.39759 & 0.2078 & 0.38026 & 0.37147 & 0.37966
\hline
X & 0.48794 & 0.51917 & 0.5 & 0.5 & 0.5
\hline
XI & 0.928 & 0.95989 & 0.872 & 0.87402 & 0.872
\hline
XII & 0.72 & 0.66267 & 0.64 & 0.64 & 0.64
\hline
\end{tabular}
\end{table}
bership degrees are weighted more heavily. But in contrast to $FI_M$, the method $FI_{K2}$ does not generate any implausible values. Since this kind of weighting is preferable, $FI_{K2}$ is more favorable than $FI_{K1}$.

Furthermore, $FI_{SC}$ can be ignored because the results of this method are very similar to the outputs of $FI_{K1}$. By searching for an alternative similarity measure that can be included in (17), varying fuzzy inequalities could be designed. Anyhow, the procedure $FI_{K2}$ can be seen as an appropriate solution so that this task appears to be needless.

Altogether, the five fuzzy inequalities (1) to (4) and (16) produce sufficient information to determine to what extent a fuzzy set is greater than another one. Hence, they should be included in a fuzzy query language. But in contrast to the first four methods, it is not clear whether $FI_{K2}$ can be calculated efficiently. This question is the topic of the next subsection.

B. Efficient Computation

It was already mentioned that $FI_{K2}$ evaluates two fuzzy sets by calculating the ratio of particular partial areas of them. Since only trapezoids are utilized here, it is not difficult to compute the relevant partial areas. Nevertheless, this process can be simplified further.

The method introduced in [60] calculates the difference of two fuzzy sets by using the extension principle, that is, it calculates the fuzzy set $c$ by means of (7). Afterwards, partial areas of the new fuzzy set are used in the determination of the final degree. It can be shown that this fuzzy inequality produces the same values as $FI_{K2}$ if at least exclusively normalized trapezoids are available. The proof of this equivalence is disregarded here.

Since the difference of two normalized trapezoids is such a fuzzy set by itself, the partial areas of only one fuzzy set have to be determined. Thus, it is even easier to compute the result. The pseudocode of the algorithm to determine the degree to which $a$ is greater than $b$ is illustrated in the Fig. 3. In each case, the algorithm calculates the relevant partial areas of the fuzzy difference.

V. CONCLUSION

A drawback of almost every fuzzy query language is that they do not possess adequate fuzzy inequalities which are important for many kinds of queries. The language FSQL which provides the four fuzzy inequalities (1) to (4) is an exception. But although the quality of their results is mostly reasonable, these methods are, on the one hand, inapplicable for some situations and are, on the other hand, only together meaningful. Hence, a single procedure is necessary that determines to what extent the fuzzy set $a$ is greater than the fuzzy set $b$. In this work different fuzzy inequalities were shown and evaluated on the basis of examples whereby some approaches were classified as problematic and others as acceptable. The method that due to its comprehensible results and its appropriate characteristics turned out to be the best option is specified by (16). Since the values of this fuzzy inequality can be calculated easily, it should also be available in a fuzzy query language.

\[ x_i = \frac{(e_i - \beta_a + \delta_i - \gamma_a + \delta_i - e_i)}{2} \]

if \( \delta_i < \gamma_a \)

\[ x_2 = \frac{(e_i - \beta_a + \delta_i - \gamma_a)}{2} \]

if \( \delta_i > \gamma_a \)

\[ x_3 = \left(1 - \frac{e_i - \beta_a + \delta_i}{e_i - \beta_a + \gamma_a - \delta_i}\right) \]

\[ x_4 = \left(1 - \frac{e_i - \beta_a + \delta_i}{e_i - \beta_a + \gamma_a - \delta_i}\right) \]

Fig. 3. Algorithm for the Computation of $FI_{K2}(a, b)$

ACKNOWLEDGMENT

The authors would like to thank Robert Graebert for numerous comments and suggestions which have been very helpful to improve the presentation of this paper.

REFERENCES


