Quantifier Selection for Linguistic Data Summarization

Ingo Glöckner

***Abstract*— Fuzzy quantifiers like "about sixty percent" are useful tools for expressing linguistic summaries. But, how can we determine the quantifier which best describes the given data? The quality indicators proposed for quantifier selection still make a rather heuristic impression. The paper therefore investigates a more principled way of controlling quantifier selection: a quantifier should be selected for summarization only when it is used in its prototypical sense. We capture this pragmatic issue of appropriate use by defining an associated *pragma quantifier* which expresses the paradigmatic cases best described by the considered quantifier. The quantifier selection will be based on an appropriateness score of the summary given by the degree of truth of the pragma quantifier. We further show that pragma quantifiers are typically neither absolute nor proportional, and thus demand generalized models of fuzzy quantification and new implementation techniques.**

I. Introduction

A. Basic concepts and earlier work

A framework for generating linguistic summaries from data and for evaluating their usefulness has been developed by Yager [2], [3], [4] and refined in the work of Kacprzyk, Strykowski and Zadrozny [5], [6], [7].

The general setting can be described as follows. We have a non-empty finite base set $E \neq \varnothing$ of individuals of interest, and a description of these individuals in terms of fuzzy sets $X \in \widetilde{\mathscr{P}}(E)$. In practice, the data will likely be described by a set A of attributes $a : E \longrightarrow V_a$ which assign an attribute value $a(e)$ to each individual. The above-mentioned fuzzy sets on E, in turn, will only indirectly be given in terms of fuzzy sets declared on the attribute values. Thus, a fuzzy set $X' \in \widetilde{\mathscr{P}}(V_a)$ declared on the attribute range of $a \in A$ gives rise to the corresponding fuzzy set $X \in \widetilde{\mathscr{P}}(E)$ defined by $\mu_X(e) = \mu_{X'}(a(e))$. The attribute-based description of the data in a database will not be of relevance in this paper, however, so that we will drop it for simplicity. From this simplified viewpoint, then, a linguistic summary has the form "Q objects are X's" or "Q X_1's are X_2's" (with an associated 'validity' or 'truthfulness' score $\tau \in [0,1]$).

Generally speaking, we may discern 'descriptive summarization' based on complete knowledge of the data collection, and 'inferential' or 'inductive summarization' which tries to generalize from a representative sample to hypotheses about the total collection. Moreover, we can discern explorative data summarization, which is not supposed to compute a comprehensive description of the data but rather to extract descriptions of interesting regularities, and summarization in a narrow sense, i.e. generation of a summary which must cover the contents of the total collection. Most systems for fuzzy data summarization belong to the first category [5], [6], [4], which has apparent junctures with data mining [8], [9]. An example of the second kind is TabVer [10], a system which generates natural language summaries of tables in a domain of environmental impact assessments.

Ingo Glöckner is with the Department of Computer Science (Prakt. Informatik VII), FernUniversität in Hagen, 58084 Hagen, Germany (phone: +49-(0)2331-987-4521; fax: +49-(0)2331-987-392; email: iglockner@web.de).

B. The assumed framework for fuzzy quantification

While existing data summarization systems are mostly based on Zadeh's Σ-count or FG-count approach to fuzzy quantification [11] or on Yager's OWA operators [12], we will assume a broader framework here which avails us with a uniform analysis of all kinds of linguistic quantifiers (see also Bosc and Lietard [13], Ralescu [14], Liu and Kerre [15] and Díaz-Hermida et al [16] for relevant work on fuzzy quantification). The general framework for fuzzy quantification assumed here (developed in [17], [18], [19]) is inspired by the linguistic Theory of Generalized Quantifiers (TGQ) [20]. It extends the notion of a (two-valued) generalized quantifier to the case of fuzzy arguments and gradual quantification results in the obvious way:

Definition 1: An $n$ary *fuzzy quantifier* on a base set $E \neq \varnothing$ is a mapping $\widetilde{Q} : \widetilde{\mathscr{P}}(E)^n \longrightarrow [0,1]$ (E needs not be finite).

For example, $\widetilde{Q}(X_1, X_2) = \sup\{\min(\mu_{X_1}(e), \mu_{X_2}(e)) : e \in E\}$ is a fuzzy quantifier suitable for modelling *at least one* X_1 *is* X_2. Describing the relationship between natural language (NL) quantifiers and matching fuzzy quantifiers is not an easy task, though, mainly because one cannot give a simple cardinality-based definition when the arguments are fuzzy. We therefore introduce semi-fuzzy quantifiers which serve as a simplified description of the target quantifier.

Definition 2: An $n$ary *semi-fuzzy quantifier* on a base set $E \neq \varnothing$ is a mapping $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$.

Semi-fuzzy quantifiers are easier to define compared to fuzzy quantifiers because one needs not describe their interpretation for fuzzy arguments. For example, a binary quantifier $[\geq 80\%]$ which models *at least 80%* can be defined by

$$[\geq 80\%](Y_1, Y_2) = \begin{cases} 1 & : \quad Y_1 \neq \varnothing \wedge \frac{|Y_1 \cap Y_2|}{|Y_1|} \geq 0.8 \\ 0 & : \quad Y_1 \neq \varnothing \wedge \frac{|Y_1 \cap Y_2|}{|Y_1|} < 0.8 \\ \frac{1}{2} & : \quad Y_1 = \varnothing\,. \end{cases} \tag{1}$$

The specification of a natural language quantifier in terms of a semi-fuzzy quantifier will be linked to the target fuzzy quantifier (which also accepts fuzzy arguments) by applying a quantifier fuzzification mechanism.

Definition 3: A *quantifier fuzzification mechanism* (QFM) $\mathscr{F}$ assigns a fuzzy quantifier $\mathscr{F}(Q) : \widetilde{\mathscr{P}}(E)^n \longrightarrow [0,1]$ to each semi-fuzzy quantifier $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$.

TABLE I

MAIN TYPES OF LINGUISTIC QUANTIFIERS

Type	example	definition
absolute unrestrictive	There are more than 3 Y's	$Q(Y) = q(\lvert Y\rvert)$
absolute	More than 3 Y_1's are Y_2's	$Q(Y_1,Y_2) = q(\lvert Y_1 \cap Y_2\rvert)$
exception	All except 3 Y_1's are Y_2's	$Q(Y_1,Y_2) = q(\lvert Y_1 \setminus Y_2\rvert)$
proportional	Two of three Y_1's are Y_2's	$Q(Y_1,Y_2) = \begin{cases} f(\frac{\lvert Y_1\cap Y_2\rvert}{\lvert Y_1\rvert}) & \lvert Y_1\rvert > 0 \\ v_0 & \text{else} \end{cases}$
cardinal comparative	More Y_1's than Y_2 are Y_3	$Q(Y_1,Y_2,Y_3) = q(\lvert Y_1 \cap Y_3\rvert, \lvert Y_2 \cap Y_3\rvert)$

C. Types of linguistic quantifiers to be considered

Table I lists the main types of linguistic quantifiers of interest for data summarization.[1] The coefficient $v_0 \in [0,1]$ that occurs in the definition of proportional quantifiers fixes the result of $Q(Y_1,Y_2)$ if there are no Y_1's at all. There are few intuitions regarding the proper interpretation in this case, but it is usually possible to let $v_0 = \frac{1}{2}$ (undecided).

The use of proportional quantifiers is typical of inductive data summarization where the observed regularities in the data are expressed as relative proportions like "most X_1's are X_2's". The preference for proportions is natural here and reflects the assumption that similar relative frequencies will be found in the total collection and in the considered sample. The remaining kinds of quantifiers which describe absolute numbers, numbers of exceptions or differences in absolute numbers are more useful for descriptive data summarization where absolute numbers are also of interest.

D. The problem of quantifier selection

From the perspective of natural language use (i.e. pragmatics), truthfulness of $Q(Y_1,Y_2)$ only indicates a possible use of Q – which might represent a very unusual case of applying Q, however. Therefore the semantically motivated definition of Q (the quantifier used in the summary), and the truth score $\tau = \mathscr{F}(Q)(X_1,X_2)$ which judges the validity of the summary "Q X_1's are X_2's", are not restrictive enough to guide quantifier selection to the most appropriate choice of Q. Consider *at least eighty percent*, for example. Clearly the corresponding quantifier should also be true if *all* X_1's are X_2's. However, only the quantifier *all* is appropriate for describing this situation, while *at least eighty percent* has a very low appropriateness grade in this case.

Existing approaches to linguistic data summarization have introduced various quality indicators for quantifier selection to solve this problem. While Yager [3] uses only the validity score and a metric for informativeness, Kacprzyk and Strykowski [5] use a multi-dimensional measure based on the degree of truth, the degree of imprecision, the degree of covering, the degree of 'appropriateness' and the length of the summary. The proposed quality indicators are rather heuristic in nature, though, while in this paper, we target at a more principled solution. The lack of specificity of the truthfulness score τ, which covers *all possible* uses of Q, shaped the idea of introducing a separate quantifier, called the *pragma quantifier* associated with Q, which is not concerned with the general semantical definition of Q but only with the most typical use of the quantifier for summarizing data. Consequently, the validity score obtained by evaluating the *pragma* quantifier, called the *appropriateness score* of the summary, will give a better indication as to the suitability of the quantifier because the summary will only be chosen if we have a paradigmatic case of applying the quantifier.

[1] See also the classification of semi-fuzzy quantifiers of Díaz-Hermida et al [21] which discerns further types of quantifiers.

II. MAIN RESULTS

A. The notion of a pragma quantifier

As explained above, our goal is that of replacing the validity score τ which expresses the truthfulness of the summary "Q X_1's are X_2's" by an improved score τ^* which judges the appropriateness of the quantifier to describe the data. To achieve this, we introduce the *pragma quantifier* $Q^* : \mathscr{P}(E)^n \longrightarrow [0,1]$ associated with a semi-fuzzy quantifier $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$, which is defined on the same base set E and accepts the same number of arguments. (For simplicity, we restrict ourselves to binary quantifiers in the following.) We take $Q^*(Y_1,Y_2) \in [0,1]$ as judging the degree to which the summary "Q Y_1's are Y_2's" is an appropriate description of the data. In other words, $Q^*(Y_1,Y_2) \in [0,1]$ expresses the degree to which the given arguments Y_1,Y_2 represent a prototypical case of Q. Recall the binary quantifier $[\geq 80\%]$ (*at least 80%*) defined by (1), for example. It would be rather unusual to summarize a proportion of 91% as "at least 80 percent" because there exists a more suitable description "at least 90 percent" on the same granularity level. The truth-conditional definition in (1) does not account for this suitability aspect, though. A pragma quantifier $[\geq 80\%]^*$ which achieves this will look roughly as follows:

$$[\geq 80\%]^*(Y_1,Y_2) = \begin{cases} \mathrm{tr}_{0.8,0.8,0.85,0.9}(\frac{\lvert Y_1\cap Y_2\rvert}{\lvert Y_1\rvert}) & : \; Y_1 \neq \varnothing \\ 0 & : \; Y_1 = \varnothing \end{cases} \quad (2)$$

where $\mathrm{tr}_{0.8,0.8,0.85,0.9}$ is a trapezoid membership function with full membership in the interval $[0.8, 0.85]$ and a subsequent decline of membership until zero membership is reached for a proportion of 0.9. By setting $v_0 = 0$, we make explicit that the quantifier $[\geq 80\%]$ should not be used if $Y_1 = \varnothing$.

In principle, the pragma quantifier Q^* can be freely chosen such that it best models the actual use of the quantifier. The only condition that we impose on the relationship between Q and Q^* is that paradigmatic use of the quantifier (as expressed by Q^*) should be stronger than possible use of the quantifier (as expressed by Q which only captures the truthfulness aspect). In formal terms, we thus require that

$$Q^*(Y_1,Y_2) \leq Q(Y_1,Y_2) \quad (3)$$

for all $Y_1,Y_2 \in \mathscr{P}(E)$. The pragma quantifier has been introduced as a semi-fuzzy quantifier in order to simplify specification. It seems reasonable to assume that for describing the prototypical uses of the quantifier, one only needs to specify the protoyipical uses for crisp arguments. We must use a QFM in order to make the pragma quantifier applicable to the fuzzy sets in the summary. Based on the assumed QFM $\mathscr{F}$, we define the appropriateness score of the summary "Q X_1's are X_2's" as $\tau^* = \mathscr{F}(Q^*)(X_1, X_2)$.

Generally speaking, quantifier selection should choose the quantifier which maximizes the appropriateness score τ^* rather than the original validity score τ.[2] Due to our lack of knowledge concerning the precise way in which a reader will interpret the summary, it is methodologically preferable to generate a summary which describes the data in a prototypical way: A selection based on the appropriateness score τ^* will favour those quantifiers for which the described situation is a paradigmatic case. This means that the impression of the reader concerning the cardinalities of the involved collections and relative proportions will be similar to the cardinalities and proportions actually found in the data.

B. Empirical evidence on the shape of pragma quantifiers

Newstead et al [22, p. 180] report a study on the effect of set size (i.e. cardinality of the first argument) on the interpretation of quantifiers which is relevant in our present context:

> "This study used a total of 11 quantifiers 'all', 'most', 'lots', 'many', 'half', 'several', 'some', 'some…not', 'few', 'a few' and 'none'. Four different set sizes were used, 12, 60, 108, 1000. Test items were of the kind: 'If lots of a group of 108 people are male, then ____ people are male.' The subjects task was to indicate what they thought was the single most appropriate whole number."

Thus the subjects specified the most typical choice of $|Y_1 \cap Y_2|$ for a quantifier, given the set size $|Y_1|$. Obviously the 'most appropriate whole numbers' so obtained do not cover the full meaning range of the quantifier but rather describe its prototypical usage. The experiment of Newstead et al revealed, among other things, that with low-magnitude quantifiers like *a few*, *few* and *some*, the proportion signified declined as set size increased, see [22, p. 180]. In other words, the prototypical usage of these quantifiers can not be expressed by a proportional quantifier (this requires a fixed proportion). The experiment also showed an incline in the absolute numbers as set size increased, which means that is also not possible to use an absolute quantifier.

C. Typical characteristics of pragma quantifiers

As shown in the last section, pragma quantifiers do not fit into the picture of the absolute/proportional distinction. The question arises which other criteria will characterize these quantifiers. We first notice that the pragma quantifiers of interest (like all quantifiers of relevance to data summarization) are generally *quantitative* in the Mostowskian sense:

Definition 4: A semi-fuzzy quantifier $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$ is *quantitative* if

$$Q(Y_1, \ldots, Y_n) = Q(\widehat{\beta}(Y_1), \ldots, \widehat{\beta}(Y_n))$$

for every bijection $\beta : E \longrightarrow E$ and $Y_1, \ldots, Y_n \in \mathscr{P}(E)$, where $\widehat{\beta}$ is the powerset mapping $\widehat{\beta}(Y) = \{\beta(e) : e \in Y\}$.

Thus, a renaming (permutation) of the arguments will not affect the quantification results, i.e. the quantifier cannot depend on any specific properties of individual elements of E.The relevance of quantitativity stems from the following alternative characterization of quantitative quantifiers:

Proposition 1: A semi-fuzzy quantifier on a finite base set is quantitative exactly if there are Boolean combinations $\Phi_1(Y_1, \ldots, Y_n)$,…, $\Phi_K(Y_1, \ldots, Y_n)$ and a mapping $q : \{0, \ldots, |E|\}^K \longrightarrow [0,1]$ such that

$$Q(Y_1, \ldots, Y_n) = q(|\Phi_1(Y_1, \ldots, Y_n)|, \ldots, |\Phi_K(Y_1, \ldots, Y_n)|)$$

for all $Y_1, \ldots, Y_n \in \mathscr{P}(E)$.

Thus, every quantitative quantifier only depends on the cardinalities of its arguments and their Boolean combinations. The description of pragma quantifiers is further simplified by a regularity of these quantifiers known as conservativity.

Definition 5: A quantifier $Q : \mathscr{P}(E)^2 \longrightarrow [0,1]$ is *conservative* if $Q(Y_1, Y_2) = Q(Y_1, Y_1 \cap Y_2)$ for all $Y_1, Y_2 \in \mathscr{P}(E)$.

It is well-known from TGQ that conservativity is shown by most NL quantifiers. In the quantitative case, we obtain a very simple characterization of conservative quantifiers:

Proposition 2: Let $Q : \mathscr{P}(E)^2 \longrightarrow [0,1]$ be quantitative. Q is conservative if and only if there exists a mapping $q : \{0, \ldots, |E|\}^2 \longrightarrow [0,1]$ such that $Q(Y_1, Y_2) = q(|Y_1|, |Y_1 \cap Y_2|)$ for all $Y_1, Y_2 \in \mathscr{P}(E)$.

This definition in terms of the pattern $Q(Y_1, Y_2) = q(|Y_1|, |Y_1 \cap Y_2|)$ appears very natural for the pragma quantifiers of interest and sufficiently expressive to catch all of them. In this case, the observed dependency of the prototypicality standard on the domain size becomes a dependency on the first argument $|Y_1|$ which represents this domain size. The prototypicality standard $q^*(|Y_1 \cap Y_2|) = q(|Y_1|, |Y_1 \cap Y_2|)$ for a given domain size $|Y_1|$ can then be applied to $|Y_1 \cap Y_2|$ to determine the outcome of quantification. This characterization of the quantificaton result in terms of the size of the domain $|Y_1|$ and the number of Y_1's which are Y_2's is exactly what one would expect of a pragma quantifier.

Let us consider a prototypical model of *a few* to illustrate these points, where $\mathrm{tr}_{a,b,c,d}$ is a trapezoid function as in (2):

$$r^* = r^*(|Y_1|) = 0.1 + \tfrac{0.6}{1+\log_{10}\max(1,|Y_1|)} \tag{4}$$

$$q^*(|Y_1 \cap Y_2|) = \begin{cases} 0 & : \ |Y_1| \le 4 \\ \mathrm{tr}_{r^*-0.2, r^*-0.1, r^*+0.1, r^*+0.2}\left(\frac{|Y_1 \cap Y_2|}{|Y_1|}\right) & : \ \text{else} \end{cases} \tag{5}$$

According to this model, *a few Y_1's are Y_2's* is not applicable at all if Y_1 has four or less elements, because we then have $Q(Y_1, Y_2) = q^*(|Y_1 \cap Y_2|) = 0$ according to the prototypicality standard $q^*(|Y_1 \cap Y_2| = q(|Y_1|, |Y_1 \cap Y_2|)$ determined by $|Y_1|$ (one should better enumerate or use other quantifiers in this case). For $|Y_1| \ge 5$, we obtain a decline of the prototypical proportion from 45% (i.e. about two elements for $|Y_1| = 5$) through 20% (i.e. about 20 elements for $|Y_1| = 100$) to a final proportion of 10% for large base sets. Thus, quantitative

[2] There are certain other effects to be considered. For example, if the data set used for summarization is only a sample of the total collection of interest, then the granularity of the considered quantifier must be balanced with the precision and representativeness of the sample.

conservative quantifiers can be used to describe pragma quantifiers whose definition depends on the size of Y_1.

From a semantical point of view, many linguistic quantifiers can be considered monotonic in their last argument, assuming the usual definition of monotonicity:

Definition 6: A quantifier is said to be *nondecreasing in the ith argument* if $Y_i \subseteq Y_i'$ always results in $Q(Y_1,\dots,Y_n) \leq Q(Y_1,\dots,Y_{i-1},Y_i',Y_{i+1},\dots,Y_n)$ independently of the other $Y_j \in \mathscr{P}(E)$; similarly for nonincreasing quantifiers where $\leq$ is replaced with $\geq$.[3]

For example, $[\geq 80\%]$ as defined by (1) is monotonically nondecreasing in the second argument, while a quantifier like "less than ten Y_1's are Y_2's" is monotonically nonincreasing in both arguments. When turning from a quantifier to its pragma form, the monotonicity type will become a bit more complex, since we expect a unimodal or 'sz-shaped' area of typical use for each quantifier. This notion is made precise by the definition of *convexity* of a quantifier in a given argument.

Definition 7: A quantifier $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$ is called *convex* in its *i*th argument if $Y_i \subseteq Y_i' \subseteq Y_i''$ always results in

$$Q(Y_1,\dots,Y_{i-1},Y_i',Y_{i+1},\dots,Y_n) \geq \min(Q(Y_1,\dots,Y_n), Q(Y_1,\dots,Y_{i-1},Y_i'',Y_{i+1},\dots,Y_n))$$

independently of the chosen $Y_j \in \mathscr{P}(E)$.

Typical examples comprise *between ten and twenty* (which is convex in both arguments) and *about 60 percent* (which is convex in the second argument). The pragma quantifier $[\geq 80\%]^*$ defined by (2) which describes prototypical usage of *at least 80%* is also convex in the second argument.

Apparently Q is convex in the second argument exactly if the mapping q is also convex in its second argument, i.e.

$$q(a,b) \geq \min(q(a,b'),q(a,b')) \tag{6}$$

for all $a,b,b',b'' \in \{0,\dots,|E|\}$ with $b' \leq b \leq b''$. Keeping the size of the domain $a = |Y_1|$ fixed, this means that $q^*(b) = q(a,b)$ will be a unimodal or sz-shaped function of b whose peak area represents the paradigmatic use of the quantifier for the given domain size.[4] This characterization of pragma quantifiers as convex quantifiers captures what we would expect of a quantifier which models the prototypical cases.

D. Axiomatization of the models

At this point, we know how the pragma quantifiers behave for two-valued arguments. However, we are ultimately interested in the appropriateness score $\tau^* = \mathscr{F}(Q^*)(X_1,X_2)$ for *fuzzy* arguments $X_1,X_2 \in \widetilde{\mathscr{P}}(E)$ which expresses the actual suitability of Q for summarizing the relationship between the X_1's and X_2's. Of course, we also need $\mathscr{F}$ to determine the regular validity score $\tau = \mathscr{F}(Q)(X_1,X_2)$ of the summary.

The general framework for fuzzy quantification makes no provisions that the results of $\mathscr{F}(Q)$ be meaningful. We must constrain the considered QFMs in order to identify the plausible choices of $\mathscr{F}$ from a linguistic perspective. We need a construction of induced fuzzy truth-functions to describe such a class of models. The construction assigns a suitable choice of fuzzy connectives to the given QFM.

Definition 8: Let $\mathscr{F}$ be a QFM and $f : \{0,1\}^n \longrightarrow [0,1]$ a (semi-fuzzy) truth function. The *induced fuzzy truth function* $\widetilde{\mathscr{F}}(f) : [0,1]^n \longrightarrow [0,1]$ is defined by $\widetilde{\mathscr{F}}(f) = \mathscr{F}(f \circ \eta^{-1}) \circ \widetilde{\eta}$, where $\eta : \{0,1\}^n \longrightarrow \mathscr{P}(\{1,\dots,n\})$ and $\widetilde{\eta} : [0,1]^n \longrightarrow \widetilde{\mathscr{P}}(\{1,\dots,n\})$ are defined by $\eta(y_1,\dots,y_n) = \{i : y_i = 1\}$ and $\mu_{\widetilde{\eta}(x_1,\dots,x_n)}(i) = x_i$, respectively.

The fuzzy set operations $\widetilde{\cup} : \widetilde{\mathscr{P}}(E)^2 \longrightarrow \widetilde{\mathscr{P}}(E)$ (fuzzy union) and $\widetilde{\neg} : \widetilde{\mathscr{P}}(E) \longrightarrow \widetilde{\mathscr{P}}(E)$ (fuzzy complement) will be defined element-wise in terms of the fuzzy disjunction $\widetilde{\vee} = \widetilde{\mathscr{F}}(\vee)$ and the fuzzy negation $\widetilde{\neg} = \widetilde{\mathscr{F}}(\neg)$. Based on these operations, we define the target class of well-behaved models.

Definition 9: A QFM $\mathscr{F}$ is called a *determiner fuzzification scheme* (DFS) if it satisfies the following conditions for all semi-fuzzy quantifiers $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$ and fuzzy arguments $X_1,\dots,X_n \in \widetilde{\mathscr{P}}(E)$:

(a) $\mathscr{F}(Q) = Q$ if $n = 0$;
(b) $\mathscr{F}(Q)(Y) = Q(Y)$ for crisp $Y \in \mathscr{P}(E)$, $n = 1$;
(c) $\mathscr{F}(\pi_e) = \widetilde{\pi}_e$ for all $E \neq \varnothing$, $e \in E$, where $\pi_e(Y) = 1$ iff $e \in Y$ and $\widetilde{\pi}_e(X) = \mu_X(e)$;
(d) $\mathscr{F}(Q')(X_1,\dots,X_n) = \widetilde{\neg}\mathscr{F}(Q)(X_1,\dots,X_{n-1},\widetilde{\neg}X_n)$ if $Q'(Y_1,\dots,Y_n) = \widetilde{\neg}Q(Y_1,\dots,Y_{n-1},\neg Y_n)$ for all crisp Y_i;
(e) $\mathscr{F}(Q')(X_1,\dots,X_{n+1}) = \mathscr{F}(Q)(X_1,\dots,X_{n-1},X_n \,\widetilde{\cup}\, X_{n+1})$ if $Q'(Y_1,\dots,Y_{n+1}) = Q(Y_1,\dots,Y_{n-1},Y_n \cup Y_{n+1})$ for all crisp Y_i;
(f) $\mathscr{F}(Q)(X_1,\dots,X_n) \geq \mathscr{F}(Q)(X_1,\dots,X_{n-1},X_n')$ if $X_n \subseteq X_n'$ given that $Q(Y_1,\dots,Y_n) \geq Q(Y_1,\dots,Y_{n-1},Y_n')$ for all crisp Y_i, $Y_n \subseteq Y_n'$;
(g) $\mathscr{F}(Q \circ \times_{i=1}^n \widehat{\mathscr{F}}(f_i)) = \mathscr{F}(Q) \circ \times_{i=1}^n \widehat{f_i}$ for all $f_i : E' \longrightarrow E$, where $\widehat{f}(Y) = \{f(e) : e \in Y\}$ for all crisp Y and $\mu_{\widehat{\mathscr{F}}(f)(X)}(e) = \mathscr{F}(\pi_e \circ \widehat{f})(X)$.

A DFS is called a *standard DFS* if it induces the standard set of connectives min, max and $1-x$. These models are the most regular but non-standard choices are also conceivable.

Let us briefly consider the rationale for introducing the postulates (a) through (g): Condition (a) ensures that $\mathscr{F}$ preserve constants (i.e. nullary quantifiers) while (b) ensures that $\mathscr{F}$ properly generalize unary quantifiers. The axiom system then guarantees that every quantifier will be properly generalized to the fuzzy case, i.e. $\mathscr{F}(Q)$ will always coindice with Q for crisp arguments. (c) expresses the compatibility of $\mathscr{F}$ with membership assessment, which can be viewed as a special case of quantification. (d) expresses the compatibility of $\mathscr{F}$ with dualization; combined with the other conditions, $\mathscr{F}$ will also be compatible negation and the formation of antonyms of quantifiers. (e) demands that $\mathscr{F}$ be compatible with unions of arguments. The condition helps to ensure the coherence of interpretations for quantifiers of different arities. (f) requires $\mathscr{F}$ to preserve (non-increasing) monotonicity of a quantifier in the nth argument; the axiom system then guarantees that every DFS will preserve arbitrary monotonicity properties. (g) expresses the compositionality of $\mathscr{F}$ with fuzzy images. The condition is necessary to ensure

[3]The analogous definition of monotonic fuzzy quantifiers is based on arguments in $\widetilde{\mathscr{P}}(E)$, and a comparison by the fuzzy subsethood relation.

[4]For $[\geq 80\%]^*$ the peak of membership is reached when $b/a = 0.8$.

TABLE II

KNOWN CLASSES OF STANDARD MODELS: AN OVERVIEW

Type	Construction
$\mathscr{F}_\Omega$-DFS	From supervaluation results of three-valued cuts: $X_\gamma^{\min} = \begin{cases} X_{\geq \frac{1}{2}+\frac{1}{2}\gamma} & \gamma \in (0,1] \\ X_{>\frac{1}{2}} & \gamma = 0 \end{cases}$ $X_\gamma^{\max} = \begin{cases} X_{> \frac{1}{2}-\frac{1}{2}\gamma} & \gamma \in (0,1] \\ X_{\geq\frac{1}{2}} & \gamma = 0 \end{cases}$ $\mathscr{T}_\gamma(X_i) = \{Y : X_\gamma^{\min} \subseteq Y \subseteq X_\gamma^{\max}\}$ $S_{Q,X_1,\dots,X_n}(\gamma) = \{Q(Y_1,\dots,Y_n) : Y_i \in \mathscr{T}_\gamma(X_i)\}$ $\mathscr{F}_\Omega(Q)(X_1,\dots,X_n) = \Omega(S_{Q,X_1,\dots,X_n})$
$\mathscr{F}_\xi$-DFS	From suprema and infima of supervaluations: $\top_{Q,X_1,\dots,X_n}(\gamma) = \sup S_{Q,X_1,\dots,X_n}(\gamma)$ $\bot_{Q,X_1,\dots,X_n}(\gamma) = \inf S_{Q,X_1,\dots,X_n}(\gamma)$ $\mathscr{F}_\xi(Q)(X_1,\dots,X_n) = \xi(\top_{Q,X_1,\dots,X_n}, \bot_{Q,X_1,\dots,X_n})$
$\mathscr{M}_\mathscr{B}$-DFS	From fuzzy median of supervaluation results: $Q_\gamma(X_1,\dots,X_n)$ $= \mathrm{med}_{\frac{1}{2}}(\top_{Q,X_1,\dots,X_n}(\gamma), \bot_{Q,X_1,\dots,X_n}(\gamma))$ $\mathscr{M}_\mathscr{B}(Q)(X_1,\dots,X_n) = \mathscr{B}(Q_\gamma(X_1,\dots,X_n)_{\gamma\in[0,1]})$

the coherence of interpretations for quantifiers defined on different base sets.

The choice of postulates (a) through (g) was based on a large catalogue of semantic desiderata from which a minimal (independent) system of core requirements was then distilled. The total list of desiderata validated by these models is discussed in [19]. Among other things, inequalities between quantifiers are preserved by a DFS. This means that for a pragma quantifier Q^*, we generally have

$$\mathscr{F}(Q^*)(X_1,X_2) \leq \mathscr{F}(Q)(X_1,X_2) \quad \text{for all } X_1,X_2 \in \widetilde{\mathscr{P}}(E),$$

and thus $\tau^* \leq \tau$, i.e. paradigmatic use is stronger than possible use of a fuzzy quantifier.

E. Concrete examples of models

Table II lists three general constructions of models which result in the classes of $\mathscr{F}_\Omega$, $\mathscr{F}_\xi$ and $\mathscr{M}_\mathscr{B}$ models.[5] The $\mathscr{F}_\Omega$-DFSes form the broadest class of standard DFSes currently known. They can also be constructed from argument similarities using the extension principle [19]. All practical $\mathscr{F}_\Omega$ DFSes belong to the more regular $\mathscr{F}_\xi$ class, though, which comprises all Q-continuous $\mathscr{F}_\Omega$ models. The most prominent example of an $\mathscr{F}_\xi$-DFS is the following model $\mathscr{F}_{\mathrm{owa}}$, which generalizes Yager's basic OWA approach [12]:

$$\mathscr{F}_{\mathrm{owa}}(Q)(X_1,\dots,X_n) = \tfrac{1}{2}\int_0^1 [\top_{Q,X_1,\dots,X_n}(\gamma) + \bot_{Q,X_1,\dots,X_n}(\gamma)]\, d\gamma.$$

The $\mathscr{M}_\mathscr{B}$-DFSes comprise the most regular models. They can be characterized as the subclass of those $\mathscr{F}_\xi$ models which propagate fuzziness (in the sense of being compatible with a natural fuzziness order). An example is

$$\mathscr{M}(Q)(X_1,\dots,X_n) = \int_0^1 Q_\gamma(X_1,\dots,X_n)\, d\gamma.$$

[5]Here, $X_{\geq\alpha}$ denotes the α-cut and $X_{>\alpha}$ the strict α-cut, respectively. Moreover, $\mathrm{med}_{1/2}(x,y)$ is the fuzzy median, i.e. the second-largest of the three values x, y, $\frac{1}{2}$.

The most prominent example is the following DFS $\mathscr{M}_{\mathrm{CX}}$, however, which generalizes Zadeh's FG-count approach [11]. The model can be succinctly described as follows:

$$\mathscr{M}_{\mathrm{CX}}(Q)(X_1,\dots,X_n) = \sup\{Q^L_{V,W} : V,W \in \mathscr{P}(E)^n, V_i \subseteq W_i\}$$
$$Q^L_{V,W} = \min(\Xi_{V,W}, \inf\{Q(Y_1,\dots,Y_n) : V_i \subseteq Y_i \subseteq W_i, \text{ all } i\})$$
$$\Xi_{V,W} = \min_i \min(\inf\{\mu_{X_i}(e) : e \in V_i\}, \inf\{1-\mu_{X_i}(e) : e \notin W_i\})\,.$$

All of these examples are standard models, i.e. compatible with the standard choice fuzzy connectives. An interesting non-standard model, which is straightforward from a random-sets view of fuzzy quantification, has been described by Díaz-Hermida et al [16]. However, their model is only defined in the finite case, and it is far from obvious how it could be extended to arbitrary quantifiers. In fact, the requirement that a QFM handle non-finite base sets might be too restrictive in certain cases, and we will therefore admit *discrete QFMs (DFSes)* as well which are limited to finite base sets but otherwise defined like the general concepts.

F. Special requirements on the models

Given the context-dependence of linguistic terms and quantifying expressions, committing to a particular choice of numerical membership grades is an intricate problem which can probably only be solved if there is a limited application domain which fixes the context. In any case, the chosen membership grades will not be perfectly dependable and we would like the process of summary generation to be robust against this factor. For that purpose, we consider four criteria which will ensure a certain stability of the quality scores τ and τ^* against the variability observed in the assignment of numerical membership degrees. These criteria will be used to assess the robustness of the example models in order to identify the most suitable choice which we will use for computing quantification scores.

Robustness against slight unsystematic changes in the membership grades of arguments and of the quantifier is ensured by the following continuity criteria.

Definition 10: A QFM $\mathscr{F}$ is called *arg-continuous* if $\mathscr{F}$ maps all $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$ to continuous fuzzy quantifiers $\mathscr{F}(Q)$, i.e. for all $X_1,\dots,X_n \in \widetilde{\mathscr{P}}(E)$ and $\varepsilon > 0$ there exists $\delta > 0$ with $d(\mathscr{F}(Q)(X_1,\dots,X_n), \mathscr{F}(Q)(X'_1,\dots,X'_n)) < \varepsilon$ for all $X'_1,\dots,X'_n \in \widetilde{\mathscr{P}}(E)$ with $d((X_1,\dots,X_n),(X'_1,\dots,X'_n)) < \delta$; where

$$d((X_1,\dots,X_n),(X'_1,\dots,X'_n)) = \max_i \sup_e [\mu_{X_i}(e) - \mu_{X'_i}(e)]\,.$$

Definition 11: A QFM $\mathscr{F}$ is called *Q-continuous* if for each semi-fuzzy quantifier $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$ and all $\varepsilon > 0$, there exists $\delta > 0$ such that $d(\mathscr{F}(Q), \mathscr{F}(Q')) < \varepsilon$ whenever $Q' : \mathscr{P}(E)^n \longrightarrow [0,1]$ satisfies $d(Q,Q') < \delta$; where

$$d(Q,Q') = \sup\{|Q(Y_1,\dots,Y_n) - Q'(Y_1,\dots,Y_n)| : Y_i \in \mathscr{P}(E)\}$$
$$d(\widetilde{Q},\widetilde{Q}') = \sup\{|\widetilde{Q}(X_1,\dots,X_n) - \widetilde{Q}'(X_1,\dots,X_n)| : X_i \in \widetilde{\mathscr{P}}(E)\}$$

for $\widetilde{Q} = \mathscr{F}(Q)$, $\widetilde{Q}' = \mathscr{F}(Q')$.

We can also demand that the models be robust against larger *systematic* changes. Perhaps those models of fuzzy quantification are more successful in this respect which do not ascribe significance to the precise choice of membership grades, but only to their relative order. Roughly speaking, this means that the scale of membership assignments should be regarded an ordinal scale. Though the standard connectives min and max are compatible with this view, we must impose a bit more structure on the scale because the membership grades are also characterized by the relationship to their associated negations. Let us now consider an operation which changes the scaling of membership degrees.

Definition 12 (Model transformation scheme): Let $\mathscr{F}$ be a QFM and $\sigma : [0,1] \longrightarrow [0,1]$ a bijection. For every $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$ and all $X_1, \dots, X_n \in \widetilde{\mathscr{P}}(E)$, we define

$$\mathscr{F}^\sigma(Q)(X_1,\dots,X_n) = \sigma^{-1}(\mathscr{F}(\sigma \circ Q)(\sigma X_1, \dots, \sigma X_n))$$

where $\sigma X_i \in \widetilde{\mathscr{P}}(E)$ is given by $\mu_{\sigma X_i}(e) = \sigma(\mu_{X_i}(e))$.
Based on these transformations, we can define the stability of a model against systematic changes of membership values.

Definition 13: A QFM $\mathscr{F}$ is called *stable under symmetric rescaling* if $\mathscr{F}^\sigma = \mathscr{F}$ for every nondecreasing bijection $\sigma : [0,1] \longrightarrow [0,1]$ with $\sigma(1-x) = 1-\sigma(x)$ for all $x \in [0,1]$.
A conforming model will only depend on the ordering of membership grades and their symmetries under negation.

The second aspect of our quasi-ordinal setting would be the possibility to work with a finite set of membership degrees, e.g. for mapping a Likert scale {very true, quite true, undecided, quite false, very false} to a numeric scale like $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. Let us assume that such scales will be embedded into $[0,1]$ in such a way that they can be negated by the standard negation $1-x$.

Definition 14: A QFM $\mathscr{F}$ *admits discrete scaling* if for every finite set $\Upsilon = \{\upsilon_1, \dots, \upsilon_m\} \subset [0,1]$ with $\{0,1\} \subseteq \Upsilon$ which is closed under negation (i.e. $x \in \Upsilon$ entails $1-x \in \Upsilon$), every Υ-valued semi-fuzzy quantifier $Q : \mathscr{P}(E)^n \longrightarrow \Upsilon$ and all Υ-valued arguments $X_1, \dots, X_n \in \widetilde{\mathscr{P}}(E)$ with $\mu_{X_i} : E \longrightarrow \Upsilon$, it also holds that $\mathscr{F}(Q)(X_1, \dots, X_n) \in \Upsilon$.
For example, if the quantifier and arguments only use the above five element scale, then the quantification result of a conforming model will also be an element of $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ and no intermediate truth values will be introduced.

Concerning the stability criteria defined above, we can assert that $\mathscr{M}$, $\mathscr{M}_{\mathrm{CX}}$ and $\mathscr{F}_{\mathrm{owa}}$ as well as the random-sets based 'discrete' model of Díaz-Hermida et al [16] satisfy the continuity requirements. However, only $\mathscr{M}_{\mathrm{CX}}$ admits discrete scaling. In addition, only $\mathscr{M}_{\mathrm{CX}}$ shows stability under symmetric rescaling. In other words, $\mathscr{M}_{\mathrm{CX}}$ only considers relative order and negation symmetries, while the other models depend on the exact choice of membership grades.

G. Implementation of pragma quantifiers

In [18], we have presented a methodology for implementing absolute quantifiers, exception quantifiers, proportional quantifiers and cardinal comparatives in the models $\mathscr{M}$, $\mathscr{M}_{\mathrm{CX}}$ and $\mathscr{F}_{\mathrm{owa}}$. This section, which contains the main technical contribution of the paper, explains how these methods can be used for implementing quantitative conservatitive quantifiers (and thus pragma quantifiers) in these models as well. In the following, we assume a finite base set $E \neq \varnothing$ of cardinality $|E| = N$. For given fuzzy arguments $X_1, \dots, X_n \in \widetilde{\mathscr{P}}(E)$, the set of relevant cutting levels is given by $\Gamma(X_1,\dots,X_n) = \left\{2\mu_{X_i}(e) - 1 : \mu_{X_i}(e) \geq \frac{1}{2}\right\} \cup \left\{1 - 2\mu_{X_i}(e) : \mu_{X_i}(e) < \frac{1}{2}\right\} \cup \{0,1\}$. The computation of quantifiers will be based on an ascending sequence of cutting levels $0 = \gamma_0 < \gamma_1 \cdots < \gamma_{m-1} < \gamma_m = 1$ with $\{\gamma_1, \dots, \gamma_m\} \supseteq \Gamma(X_1, \dots, X_n)$ (usually we will have an equality here). The proposed algorithms rest on a pre-computation of membership histograms which has complexity $\mathscr{O}(N \log m)$, see [18], [19] As shown in table III, the complexity of the subsequent evaluation stage then depends on the quantifier type and its monotonicity pattern, but it is independent of the chosen model $\mathscr{F}_{\mathrm{owa}}$, $\mathscr{M}$ or $\mathscr{M}_{\mathrm{CX}}$. The results listed for absolute comparatives can be improved to $\mathscr{O}(mN)$ (in the general case) and $\mathscr{O}(m)$ (in the convex case) provided that Q depends only on $|Y_1 \cap Y_3| - |Y_2 \cap Y_3|$.

TABLE III
IMPLEMENTATION: COMPLEXITY OF ALGORITHMS FOR EVALUATING $\mathscr{F}_\xi(Q)(X_1,\dots,X_n)$ IN THE PROTOTYPICAL MODELS

Type of Q	example	complexity
abs. unrestrictive	There is an odd number of X's	$\mathscr{O}(mN)$
— convex	There are about 50 X's	$\mathscr{O}(m)$
— monotonic	There are at least 20 $X's$	$\mathscr{O}(m)$
absolute	About 50 X_1's are X_2's	see *abs. unrst.*
exception	All except 10 X_1's are X_2's	see *abs. unrst.*
proportional	10% or 20% of X_1's are X_2's	$\mathscr{O}(mN^2)$
— convex	About 50% of X_1's are X_2's	$\mathscr{O}(mN)$
— monotonic	Most X_1's are X_2's	$\mathscr{O}(m)$
abs. comparative	$\lvert X_1 \cap X_3\rvert - \lvert X_2 \cap X_3\rvert$ is prime	$\mathscr{O}(mN^2)$
— convex	Two times more X_1's than X_2's are X_3's	$\mathscr{O}(mN)$
— monotonic	More X_1's than X_2's are X_3's	$\mathscr{O}(m)$

Unfortunately, pragma quantifiers are not covered by this table because pragma quantifiers are typically neither absolute nor proportional, but rather more general conservative hybrids (see sections II-B and II-C). Recalling that pragma quantifiers are typically quantitative and conservative, we will therefore develop a computational analysis of such quantifiers which shows how to implement pragma quantifiers in the above models $\mathscr{F}_{\mathrm{owa}}$, $\mathscr{M}$ and $\mathscr{M}_{\mathrm{CX}}$.

Hence let $0 = \gamma_0 < \gamma_1 < \dots < \gamma_{m-1} < \gamma_m = 1$ be an ascending sequence of $\gamma_j \in [0,1]$ (as above). For $\gamma = 0, \dots, m-1$, we abbreviate $\overline{\gamma}_j = \frac{\gamma_j + \gamma_{j+1}}{2}$. We further let $\top_j = \top_{Q,X_1,\dots,X_n}(\overline{\gamma}_j)$, $\perp_j = \perp_{Q,X_1,\dots,X_n}(\overline{\gamma}_j)$ and $C_j = \mathrm{med}_{1/2}(\top_j, \perp_j)$. As a prerequisite for implementing pragma quantifiers, let us now express the preferred model $\mathscr{M}_{\mathrm{CX}}$ as a function of the finite sample $\Gamma(X_1, \dots, X_n)$ of (three-valued) cut levels.

Proposition 3: Let $Q : \mathscr{P}(E)^n \longrightarrow [0,1]$, $X_1, \dots, X_n \in \widetilde{\mathscr{P}}(E)$ and $0 = \gamma_0 < \gamma_1 < \dots < \gamma_{m-1} < \gamma_m = 1$ be given, $\Gamma(X_1, \dots, X_n) \subseteq \{\gamma_0, \dots, \gamma_m\}$. For $j \in \{0, \dots, m-1\}$ let $B_j = 2\perp_j - 1$ if $C_0 \geq \frac{1}{2}$ and $B_j = 1 - 2\top_j$ otherwise. Further let

$$\widehat{J} = \{j \in \{0, \dots, m-1\} : B_j \leq \gamma_{j+1}\}, \quad \widehat{j} = \min \widehat{J}.$$

Then

$$\mathscr{M}_{\mathrm{CX}}(Q)(X_1,\dots,X_n) = \begin{cases} \frac{1}{2}+\frac{1}{2}\max(\gamma_{\widehat{j}},B_{\widehat{j}}) & : \quad \bot_0 > \frac{1}{2} \\ \frac{1}{2}-\frac{1}{2}\max(\gamma_{\widehat{j}},B_{\widehat{j}}) & : \quad \top_0 < \frac{1}{2} \\ \frac{1}{2} & : \quad \text{else.} \end{cases}$$

A similar reformulation is possible for $\mathscr{F}_{\mathrm{owa}}$ and $\mathscr{M}$, see [18] and [19]. These formulas enable us to evaluate fuzzy quantifications in $\mathscr{F}_{\mathrm{owa}}$, $\mathscr{M}$ and $\mathscr{M}_{\mathrm{CX}}$. However, we must still optimize the computation of $\top_j$ and $\bot_j$ because a naive implementation which considers each $Y_i \in \mathscr{T}_{\overline{\gamma}_j}(X_i)$ will not give us acceptable performance. In the following, we utilize the fact that typical pragma quantifiers are quantitative and conservative in order to effect this optimization.

First of all, it is apparent from the analysis $Q(Y_1,Y_2) = q(|Y_1|,|Y_1 \cap Y_2|)$ of these quantifiers given in proposition 2 that $\top_j$ and $\bot_j$ can be rewritten as

$$\top_j = \max\{q(c_1,c_2) : (c_1,c_2) \in R_j\}$$
$$\bot_j = \min\{q(c_1,c_2) : (c_1,c_2) \in R_j\}$$

where $\qquad R_j = \{(|Y_1|,|Y_1 \cap Y_2|) : Y_i \in \mathscr{T}_\gamma(Y_i)\}$.

In numeric terms, R_j can be described as follows.

Proposition 4: For conservative quantifiers,

$$R_j = \{(c_1,c_2) : \ell_1 \le c_1 \le u_1, \\ \max(\ell_2, c_1 - u_3) \le c_2 \le \min(u_2, c_1 - \ell_3)\},$$

where $\ell_r = |Z_r|_\gamma^{\min} = |(Z_r)_\gamma^{\min}|$ and $u_r = |Z_r|_\gamma^{\max} = |(Z_r)_\gamma^{\max}|$, $\gamma = \overline{\gamma}_j$, depend on $Z_1 = X_1$, $Z_2 = X_1 \cap X_2$ and $Z_3 = X_1 \cap \neg X_2$, assuming the standard fuzzy intersection and complement.[6] We conclude that

$$\top_j = \max\{q(c_1,c_2) : \ell_1 \le c_1 \le u_1, \\ \max(\ell_2, c_1 - u_3) \le c_2 \le \min(u_2, c_1 - \ell_3)\}$$
$$\bot_j = \min\{q(c_1,c_2) : \ell_1 \le c_1 \le u_1, \\ \max(\ell_2, c_1 - u_3) \le c_2 \le \min(u_2, c_1 - \ell_3)\}.$$

In order to compute a quantification result based on this formula, one must consider every choice of j (i.e. m cutting levels) and (at worst) $N = |E|$ choices of c_1 and c_2 each. Thus, the complexity of evaluating a quantitative conservative quantifier is $\mathscr{O}(mN^2)$ in the general case. However, we know that typical pragma quantifiers are convex in their second argument, which means that q will be convex in its second argument as well, see (6). Keeping $a \in \{0,\dots,|E|\}$ fixed, this means that there exists a 'peak' element $b_{pk}(a) \in \{0,\dots,|E|\}$ such that $q(a,b)$ is nondecreasing for all $b \le b_{pk}(a)$ and nonincreasing for all $b \ge b_{pk}(a)$. We therefore obtain

$$\top_j = \max\{q'(c_1) : \ell_1 \le c_1 \le u_1\}$$
$$\bot_j = \min\{\min(q(c_1, \max(\ell_2, c_1 - u_3)), \\ q(c_1, \min(u_2, c_1 - \ell_3))) : \ell_1 \le c_1 \le u_1\}$$

where

$$q'(c_1) = \begin{cases} q(c_1, \min(u_2, c_1 - \ell_3)) & : \quad \min(u_2, c_1 - \ell_3) < b_{pk}(c_1) \\ q(c_1, \max(\ell_2, c_1 - u_3)) & : \quad \max(\ell_2, c_1 - u_3) > b_{pk}(c_1) \\ q(c_1, b_{pk}(c_1)) & : \quad \text{else.} \end{cases}$$

[6] A method for efficiently computing ℓ_r and u_r from the histogram of Z_r is explained in [18], [19].

TABLE IV

QUANTIFIER SELECTION TABLE OF THE TABVER SYSTEM

$\lvert Y_1 \rvert$	$\lvert Y_1 \cap Y_2 \rvert$							
	2	3	4	5	6	7	8	9
2	both							
3	—	all						
4	—	—	all					
5	—	—	nall	all				
6	—	—	most	nall	all			
7	—	—	mhalf	most	nall	all		
8	—	—	ahalf	mhalf	most	nall	all	
9	—	—	ahalf	mhalf	mhalf	most	nall	all

Assuming that b_{pk} can be determined in constant time, this means that in the convex case, we must consider m choices of j and at most N choices of c_1. The complexity of evaluating a convex quantitative conservative quantifier (i.e. our typical pragma quantifier) therefore reduces to $\mathscr{O}(mN)$, which permits efficient computation in practice.

H. Summarization example

In the TabVer system [10] mentioned in the introduction, quantifier selection was based on a simple lookup table for small cardinalities (shown in Table IV) and a fixed choice of prototypical intervals to control quantifier selection for cardinalities $|Y_1| \ge 10$, i.e. *ahalf* $[0.45, 0.55)$, *mhalf* $[0.55, 0.7)$, *most* $[0.7, 0.9)$, *nall* $[0.9, 1)$, and *all* $\{1\}$. (Here, *ahalf* means 'around half', *mhalf* 'more than half', and *nall* 'nearly all'). There was no need to introduce quantifiers for the low percentage range (marked '—' in the table) because TabVer would typically enumerate such cases or summarize their complement. The TabVer system assumed crisp arguments throughout, i.e. $|Y_1|$ and $|Y_1 \cap Y_2|$ were directly used for quantifier lookup. However, the lookup table and selection intervals can also be viewed as describing typical usage of the considered quantifiers. The corresponding six pragma quantifiers are then defined by $Q^*(Y_1,Y_2) = q(|Y_1|, |Y_1 \cap Y_2|$, where for simplicity we let $q(|Y_1|, |Y_1 \cap Y_2|) = 1$ if $(|Y_1|, |Y_1 \cap Y_2|)$ is marked by Q in the quantifier table, or $|Y_1| \ge 10$ and $|Y_1 \cap Y_2|/|Y_1|$ is contained in the selection interval of the quantifier; otherwise we let $q(|Y_1|, |Y_1 \cap Y_2|) = 0$. In this way, we obtain a system of (two-valued) pragma quantifiers which can be used to control quantifier selection.[7] Because of its special robustness (see section II-F), the model $\mathscr{M}_{\mathrm{CX}}$ was used for computing appropriateness grades $\tau^* = \mathscr{M}_{\mathrm{CX}}(Q)(X_1,X_2)$. The appropriateness scores τ^* obtained for three summarization problems are shown in Table V. In the first case, appropriateness scores where computed for $X_1 = a/1.0 + b/0.9 + c/0.8 + d/0.7$ and $X_2 = a/0.7 + b/0.8 + c/0.7 + d/0.9$. The best result of $\tau^* = 0.7$ was observed for the quantifier *all*. In the second case, τ^* was computed for a larger summarization problem $X_1' = a/1.0 + b/0.9 + c/0.8 + d/0.8 + e/0.7 + f/0.7 + g/0.1 + h/1.0 + i/0.9$ and $X_2' = a/0.05 + b/0.9 + c/0.7 + d/1.0 + e/1.0 + f/0.8 + g/0.1 + h/0.6 + i/0.9$. Here, the best appropriateness score $\tau^* = 0.6$

[7] Of course, a more complicated gradual modelling, e.g. by trapezoid functions as in (5), would also be possible and probably more appropriate.

TABLE V

APPROPRIATENESS SCORES τ^* FOR THREE SUMMARIZATION PROBLEMS

Quantifier	ahalf	mhalf	most	nall	all	both
τ^* for X_1, X_2	0.0	0.0	0.0	0.0	0.7	0.2
τ^* for X_1', X_2'	0.2	0.3	0.4	0.6	0.05	0.05
τ^* for X_1', X_2''	0.3	0.4	0.5	0.5	0.5	0.1

was obtained for the quantifier *nall* (nearly all). In the third example, X_2' was exchanged by $X_2'' = a/0.5+b/0.5+c/0.7+d/1.0+e/1.0+f/0.8+g/0.5+h/0.6+i/0.9$. In this case, X_2'' is so fuzzy that a unique decision between *most*, *nearly all* and *all* only based on τ^* is no longer possible. In fact, *most* is the best choice here because from the truthfulness point of view, it includes *nearly all* and *all* as special cases.

III. CONCLUSIONS

This paper was concerned with the problem of quantifier selection in fuzzy data summarization. We started from an example demonstrating that one cannot simply choose any quantifier which validates the summary. Although the generated summary should of course be verified by the data, the truthfulness score can be misleading because it covers all possible uses of the quantifier regardless of their typicality. By contrast, we proposed to select only those quantifiers for which the situation described is a paradigmatic case. This notion of prototypical usage is formalized by a *pragma quantifier* which gives rise to the appropriateness score τ^* on which the quantifier selection will be based.

The psychological literature on quantifiers used in rating scales gives some impression of the general shape of such pragma quantifiers. Specifically, they are neither fixed proportional nor absolute quantifiers, and their interpretation typically shows some effects of the size of the considered domain. However, these expressive quantifiers can be viewed as special cases of conservative quantifiers. We investigated the computational implications of pragma quantifiers by developing the relevant formulas for implementing these quantifiers in important models. While regular quantifiers are often monotonic, pragma quantifiers appear to be typically convex (i.e. unimodal or sz-shaped). This means that the computational effort to compute quantification results will be higher than for monotonic quantifiers. However, the analysis of computational complexity revealed that the new class of pragma quantifiers can still be handled efficiently.

The proposed method for quantifier selection is mainly intended for generating summaries involving vague quantifiers like *a few*, *many* etc. Selection should be based on a predefined system of such quantifiers (as in the above example). This makes it possible to define a balanced (non-conflicting) system of pragma quantifiers where the prototypical cases of using a quantifier also depend on all other quantifiers available for summarization. As shown by the example in section II-H, the method works well if there is a balance between the width of the selection intervals and the amount of fuzziness observed in the arguments. If the selection intervals are too narrow given the fuzziness of the data (i.e. if one gets several 0.5 values for τ^*), one should try and utilize inclusions between the top-ranked quantifiers in order to establish the most general choice. Alternatively, one could fall back on the truthfulness score τ to break symmetries or use a combined factor $\tau \cdot \tau^*$ for a finer ranking. It would also be instructive to try the random-sets based model of Díaz-Hermida et al [16] which is computationally demanding but potentially more discriminating than a standard DFS.

REFERENCES

[1] D. Dubois, H. Prade, and R. Yager, Eds., *Fuzzy Information Engineering*. New York: Wiley, 1997.

[2] R. Yager, "A new approach to the summarization of data," *Information Sciences*, vol. 28, pp. 69–86, 1982.

[3] ——, "On linguistic summaries of data," in *Knowledge Discovery in Databases*, W. Frawley and G. Piatetsky-Shapiro, Eds. AAAI/MIT Press, 1991, pp. 347–363.

[4] D. Rasmussen and R. Yager, "A fuzzy SQL summary language for data discovery," in *Fuzzy Information Engineering*, D. Dubois, H. Prade, and R. Yager, Eds. New York: Wiley, 1997, pp. 253–264.

[5] J. Kacprzyk and P. Strykowski, "Linguistic summaries of sales data at a computer retailer: A case study," in *Proc. of 8th International Fuzzy Systems Association World Congress (IFSA '99)*, 1999, pp. 29–33.

[6] S. Zadrożny and J. Kacprzyk, "On database summarization using a fuzzy querying interface," in *Proc. of 8th International Fuzzy Systems Association World Congress (IFSA '99)*, 1999, pp. 39–43.

[7] J. Kacprzyk, "Intelligent data analysis via linguistic data summaries: A fuzzy logic approach," in *Classification and Information Processing at the Turn of the Millenium*, R. Decker and W. Gaul, Eds. Springer, 2000, pp. 153–161.

[8] K. Cios, W. Pedrycz, and R. Swiniarski, *Data Mining Methods for Knowledge Discovery*. Dordrecht: Kluwer, 1998.

[9] T. Lin, Y. Yao, and L. Zadeh, Eds., *Data Mining, Rough Sets and Granular Computing*. Heidelberg: Physica-Verlag, 2002.

[10] I. Glöckner, A. Grieszl, M. Müller, and M. Ronthaler, "TabVer: A case study in table verbalization," in *KI-94: Advances in Artificial Intelligence*, ser. LNAI, B. Nebel and L. Dreschler-Fischer, Eds., vol. 861. Berlin: Springer, 1994, pp. 94–105.

[11] L. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," *Computers and Mathematics with Applications*, vol. 9, pp. 149–184, 1983.

[12] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[13] P. Bosc and L. Lietard, "On the comparison of the Sugeno and Choquet fuzzy integrals for the evaluation of quantified statements," in *Proceedings of EUFIT '95*, 1995, pp. 709–716.

[14] A. Ralescu, "A note on rule representation in expert systems," *Information Sciences*, vol. 38, pp. 193–203, 1986.

[15] Y. Liu and E. Kerre, "An overview of fuzzy quantifiers. (I). Interpretations," *Fuzzy Sets and Systems*, vol. 95, pp. 1–21, 1998.

[16] F. Díaz-Hermida, D. Losada, A. Bugarín, and S. Barro, "A probabilistic quantifier fuzzification mechanism: the model and its evaluation for information retrieval," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 5, pp. 688–700, 2005.

[17] I. Glöckner, "DFS – an axiomatic approach to fuzzy quantification," Technical Faculty, University Bielefeld, 33501 Bielefeld, Germany, TR97-06, 1997.

[18] I. Glöckner, "Evaluation of quantified propositions in generalized models of fuzzy quantification," *International Journal of Approximate Reasoning*, vol. 37, no. 2, pp. 93–126, 2004.

[19] I. Glöckner, *Fuzzy Quantifiers: A Computational Theory*, ser. Studies in Fuzziness and Soft Computing. Berlin: Springer, 2006, vol. 193.

[20] J. Barwise and R. Cooper, "Generalized quantifiers and natural language," *Linguistics and Philosophy*, vol. 4, pp. 159–219, 1981.

[21] F. Díaz-Hermida, P. Cariñena, A. Bugarín, and S. Barro, "Definition and classification of semi-fuzzy quantifiers for the evaluation of fuzzy quantified sentences," *International Journal of Approximate Reasoning*, vol. 34, no. 1, pp. 49–88, 2003.

[22] S. Newstead, P. Pollard, and D. Riezebos, "The effect of set size on the interpretation of quantifiers used in rating scales," *Applied Ergonomics*, vol. 18, no. 3, pp. 178–182, 1987.