24th European Conference on Artificial Intelligence - ECAI 2020
Santiago de Compostela, Spain

An experimental study on the behaviour of Fuzzy Quantification Models

Andrea Cascallar-Fuentes and **Alejandro Ramos-Soto** and **Alberto Bugarín-Diz**[1]

Abstract. In this paper we evaluate empirically whether there exist significant differences in the numerical results produced by six well-known fuzzy quantification models when applied to the evaluation of unary and binary fuzzy quantified statements on numerical data sets. The models we analyzed are: Zadeh's scalar and fuzzy cardinality, Yager's OWA, Delgado's GD, Sugeno integral and Vila's VQ. These models were tested by evaluating the degree of fulfillment they produced on fifteen numerical data sets from the UCI Machine Learning repository for all the possible fuzzy quantified statements generated by partitions of up to seven quantifiers and linguistic terms of the variables involved. We conducted tests of statistical significance for these evaluation results under a pair-wise comparison. Results indicate that no significant differences were found among the models for unary quantifiers involving a single imprecise property, with a single exception of very limited outreach. For binary quantified statements involving two imprecise properties, significant differences were observed in general among all the pairs of fuzzy quantification models under study. Therefore, in spite of unary models fulfill different theoretical properties, the models under study exhibit very similar empirical behaviour. For binary models, results point out that the selection of a particular model should be guided by other criteria (e.g. the properties they fulfill) different than their experimental behaviour, which is empirically proved to be different.

1 INTRODUCTION

The presence and use of linguistic quantifiers in human language is a very powerful tool for representing and describing knowledge about the quantity of elements that fulfill one or more properties [13].

Let us consider the following quantified statements as initial examples: ”Women's voting was about 60%” and ”Almost all workers are young.” In both cases, the statements express linguistically the number (”About 60%”, ”Almost all”) of elements in a given referential (women, workers) that respectively fulfill the corresponding properties, which in the examples are crisp (”having voted”) and imprecise (”being young”). From the linguistic point of view, determiners are the elements that usually develop the quantification role in language. Among the huge variety of determiners we use in language (lexical, proportional, absolute, exception, partitive, ...)[28, 25, 16], most of the attention in the related literature has been paid to absolute quantifiers, which express quantities over the total number of elements of the referential that fulfill the properties (e.g. “Two or more”, “A few”), and relative quantifiers, which make the counting depending on the total number of elements of the referential (e.g. “A half”, “Almost all”). Regarding the number of properties considered in the quantified statements (the n-arity of the quantifier) it usually ranges from on (unary quantifiers) to four (quaternary quantifiers), although literature has mostly focused on unary and binary quantifiers [2, 27, 29].

Unary quantified statements have the following structure: “Q X are S” where Q is a quantifier (e.g., “some”), X is a referential set (e.g., “students”), and S is a linguistic value (e.g., “tall”). Thus, an example of unary quantified statement is: “Some students are tall”. Binary quantified statements have the structure ”Q KX are S”, where an additional linguistic value K is included (e.g., ”blonde”). Thus, an example of binary quantified statement is: “Some *blonde* students are tall”.

In general, quantified sentences are a versatile tool for modelling natural language expressions which are used in a wide range of areas [13]. For instance, in multiple-criteria decision-making, fuzzy quantification models were proposed for aggregating the criteria according to their importance. Another fruitful application is fuzzy querying on databases, since natural language statements can be modelled by quantified sentences, being also suitable in the information retrieval area.

Quantified sentences are also used for building linguistic descriptions of data (LDD) [26, 35], which provide quantitative information about the fulfillment of some properties of interest in a numerical data set. Since the quantitative information, as well as the properties, is, in general, imprecise or fuzzy, many LDD models use the concept of quantified protoform [50] and follow the computing with words paradigm, where computations are performed on linguistic terms modeled as fuzzy sets [47, 52, 51], and its evolution, computing with perceptions [48, 49]. The information included in LDD may, in some cases, be directly consumed by users (as a way of conveying the information hidden in the data) but, in most cases, LDD are used in the content determination stage of the Natural Language Generation pipeline [32, 33]. Within the natural language generation field (NLG), many systems have been developed over the years with the aim of generating comprehensible texts from different data sources for a wide variety of application domains [24]. In NLG, LDD are actually pieces of information, usually described in an intermediate language, which are abstracted and combined with other information sources in order to produce (after performing the planning stages) the final natural language narrative which is conveyed to and consumed by the users.

Evaluating quantified sentences involves the use of a fuzzy quantification model, which calculates the fulfillment degree of the sentence (a value in the range [0,1]). The fulfillment degree in quantified sentences is a measure that combines the cardinality, i.e., how many elements in the referential match the property in the statement, and

---

[1] Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Spain, email: {andrea.cascallar.fuentes, alejandro.ramos, alberto.bugarin.diz}@usc.es

the compatibility between the cardinality and the quantifier. Several fuzzy quantification models have been proposed in the literature and were later studied from a theoretical perspective in terms of the properties they fulfill [25, 14, 13, 3, 9, 13, 16, 17, 18, 19, 37]. An extensive list of properties (including monotonicity, continuity, correct generalization, negation, antonymy and duality, among others) has been described, considering different aspects, that help to characterize the behaviour of the fuzzy quantification models. From this perspective, all the fuzzy quantification models exhibit different behaviour, since all of them fulfill different properties. Also some of them exhibit non plausible behaviour for some uses, since they fail to fulfill some relevant properties. But the behaviour of the models has not been studied yet from a practical or pragmatical perspective, by analyzing the real quantitative differences existing among them. This experimental approach, which has been adopted in other research fields (such as Machine Learning, for instance [21, 22]) is being done for the first time in this paper.

Our aim is to experimentally test whether there are significant differences between the most widely used unary and binary fuzzy quantification models used in quantified sentences. Therefore, this paper attempts to extend the previously mentioned theoretical studies to determine if significant differences exist among the methods and assess whether the presence or absence of differences justifies the application of one method over the others.

This paper is structured as follows: firstly, in Section 2 we describe the fuzzy quantification models included in our experimentation. In Section 3, we describe the selected data sets to perform our comparison and the definition of their associated protoform components (namely, the fuzzy variables, their partitions and the fuzzy quantifiers). Section 4 presents the experimental comparison between the fuzzy quantification models for unary and binary cases. Finally, Section 5 closes the paper with some final remarks.

2 FUZZY QUANTIFICATION MODELS

In Fuzzy quantification models, Zadeh [43, 44, 45, 46] proposed an extension of the classical existential and universal quantifiers (“exist” and “for all”), as well as other crisply defined quantifiers (e.g. ”more than 40%”) to imprecise (fuzzy) quantifiers with a higher degree of expressiveness, such as “a few” or “most of”. Later on, following a different perspective research line for the proposal of imprecise quantification models, the Theory of Generalized Quantifiers (TGQ) was developed [4, 6, 28] independently.

In [25], a generalization of the TGQ based on quantifier fuzzification mechanisms (QFM) was proposed. This generalization allows to define a fuzzy quantifier based on a transformation from semi-fuzzy quantifiers, which are easier to design. This mechanism can be applied to a wide range of quantifiers (not just absolute and relative ones), such as comparative, exception, ternary or quaternary quantifiers.

Evaluating a quantified sentence, as described above, involves computing its fulfillment degree. In this evaluation two elements must be considered: *i)* the cardinality, i.e., how many elements in the referential fulfill the (fuzzy) linguistic values stated for the variables in the statement; *ii)* the compatibility between the cardinality and the quantifier.

We describe in what follows these two elements for the six fuzzy quantification models in our study, which are the most frequently used in the literature.

2.1 Unary quantification models

In this section we present the unary fuzzy quantification models we have considered in our study. They are used in fuzzy quantified statements that are referred in the literature as unary quantified statements, which follow the “*Q X* are *S* prototype.

2.1.1 *Sum-based evaluation methods*

Zadeh’s method. This method [46] is based on the scalar cardinality “power” defined by Zadeh as $P(A) = \sum_{i=1}^{n} A(x_i)$.

The evaluation of unary quantified sentences for relative quantifiers is defined as:

$$Z_Q(A) = Q\left(\frac{P(A)}{|X|}\right) \tag{1}$$

Yager’s method based on OWA operators This method is a special case of the Choquet integral [8, 7]. For unary quantified statements, the degree of truth based on Choquet integral is defined as:

$$C_Q(A) = \sum i = 1nb_i \times (Q(i/n) - Q\left((i-1)/n\right)) \tag{2}$$

Yager’s method [40] can only be used with coherent[2] and relative quantifiers.

Being $w_i = Q\left(\frac{i}{|X|}\right) - Q\left(\frac{i-1}{|X|}\right), i \in \{1, ..., n\}$ and $Q(0) = 0$, the evaluation is:

$$Y_Q(A) = \sum_{i=1}^{n} w_i b_i \tag{3}$$

where b_i is the i-th higher value of the fulfillment degree to the fuzzy set A.

Delgado’s GD method The GD method [15, 13] is a quantification model of the so-called-G-family that belongs to a method family based on a fuzzy cardinality E defined as follows:

$$GD_Q(A) = \underset{i\in\{0,\ldots,n\}}{\oplus} \left(E(A,i) \otimes Q\left(\frac{i}{n}\right)\right) \tag{4}$$

Using the product as t-norm and the Lukasiewicz’s t-conorm, the evaluation of a unary quantified statement with relative quantifiers is as follows:

$$GD_Q(A) = \sum_{i=0}^{n} ED(A,i) \times Q\left(\frac{i}{n}\right) \tag{5}$$

where $ED(A,k) = b_k - b_{k+1}$ with $b_0 = 1$ and $b_{n+1} = 0$ is the ED fuzzy cardinality [14], a particular case of the E cardinality, using the minimum t-norm, Lukasiewicz’s t-norm, the maximum t-conorm and the standard negation.

2.1.2 *Max-min-based evaluation methods*

Sugeno integral based method The Sugeno integral [8] is another method to evaluate quantified sentences which also requires coherent quantifiers. In the relative quantifier case, the evaluation is:

$$S_Q(A) = \max_{1\leq i\leq n} \min\left(Q\left(\frac{P(A)}{|X|}\right)\right) \tag{6}$$

---

[2] A quantifier Q is coherent if $Q(x_i) \leq Q(x_{i+1}) \forall x_i < x_{i+1}$ and $Q(0) = 0, Q(1) = 1$.

ZS method This method [9] is based on Zadeh's fuzzy cardinality:

$$Z(A,k) = \begin{cases} 0 & \text{if } \nexists\alpha \mid |A_\alpha| = k \\ sup\{\alpha \mid |A_\alpha| = k\} & otherwise \end{cases} \tag{7}$$

The evaluation for unary quantified statements with relative quantifiers is:

$$ZS_Q(A) = \max_{k\in\{0,\dots,n\}} \min\left(Z(A,k), Q\left(\frac{k}{n}\right)\right) \tag{8}$$

It can be proved [13] that (8) is equivalent to:

$$ZS_Q(A) = \max_{\alpha\in M(A)} \min\left(\alpha, Q(|A_\alpha|)\right) \tag{9}$$

where $M(A) = \{\alpha \in (0,1] \,|\exists x_i \in X \; with \; A(x_i) = \alpha\} \cup \{1\}$ so the method evaluation can be performed without calculating the Z cardinality.

2.2 Binary quantification models

In this section we present the binary fuzzy quantification models we have considered in our study. They are used in fuzzy quantified statements that are referred in the literature as binary quantified statements, which follow the "*Q KX* are *S*" prototype.

2.2.1 *Sum-based evaluation methods*

Zadeh's method This method [46] is based on the relative cardinality of A and D:

$$P(A/D) = \frac{P(A\cap D)}{P(D)} \tag{10}$$

The evaluation of relative quantifiers is as follows:

$$Z_Q(A/D) = Q\left(P(A/D)\right) = Q\left(\frac{P(A\cap D)}{P(D)}\right) \tag{11}$$

Yager's method based on OWA operators Yager's model [40] can only be generalized to binary sentences for coherent and relative quantifiers. Its parameters are calculated as follows:

$$w_i = Q(S_i) - Q(S_{i-1}) \; i \in \{1,\dots,n\} \tag{12}$$

where

$$S_i = \frac{1}{d}\sum_{j=1}^{i} e_i, \; d = \sum_{k=1}^{n} e_k \tag{13}$$

being e_k the *k*-th low value of D set's fulfillment degree and $S_0 = 0$. The evaluation is:

$$Y_Q(A/D) = \sum_{i=1}^{n} w_i c_i \tag{14}$$

where c_i is the i-th highest value of the set of fulfillment degrees of $(\neg D \vee A)$.

Delgado's GD method The generalization of the GD method [15] uses the fuzzy cardinality ER, which utilizes the product as t-norm and the Lukasiewicz's t-conorm, as follows:

$$GD_Q(A/D) = \sum_{c\in CR(A/D)} ER(A/D,c) \times Q(c) \tag{15}$$

where

$$CR(A/D) = \{\frac{|(A\cap D)_\alpha|}{|D_\alpha|} \; with \; \alpha \in M(A/D)\} \tag{16}$$

and

$$\begin{aligned} M(A/D) = M(A\cap D) \cup M(D), \; and \\ M(A) = \{\alpha \in (0,1] \,|\exists x_i \in X \; with \; A(x_i) = \alpha\} \end{aligned} \tag{17}$$

It can be proved [13] that the evaluation 15 is equivalent to:

$$GD_Q(A/D) = \sum_{\alpha_i\in M(A/D)} (\alpha_i - \alpha_{i+1}) \times Q\left(C(A/D,\alpha_i)\right) \tag{18}$$

where if $M(A/D) = \{\alpha_1, \dots, \alpha_m\}$ an α-cut set defined in 16 with $1 = \alpha_1 > \dots > \alpha_m > \alpha_{m+1} = 0$, then:

$$C(A/D,\alpha_i) = \frac{|(A\cap D)_{\alpha_i}|}{|D_{\alpha_i}|} \tag{19}$$

Thus, the evaluation of a binary quantified statement can be performed without calculating the $ER(A/D)$ cardinality.

2.2.2 *Max-min-based evaluation methods*

Vila, Cubero, Medina and Pons' method This method [37] uses the "or" or "orness" degree defined for coherent quantifiers. *orness*(∃) = 1 and *orness*(∀) = 0. Every coherent quantifier Q between ∃ and ∀ has an orness degree in the [0, 1] interval:

$$o_Q = \sum_{i=1}^{n}\left(\frac{n-i}{n-1}\right) \times \left(Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right)\right) \tag{20}$$

Then, the evaluation for a binary quantified statement is:

$$\begin{aligned} V_Q(A/D) = o_Q \max_{x\in X}\left(D(x) \wedge A(x)\right) \\ +(1-o_Q)\min_{x\in X}\left(A(x) \vee (1-D(x))\right) \end{aligned} \tag{21}$$

ZS method This method [9], which uses the fuzzy cardinality ES, consists in a max-min composition between that cardinality and the quantifier, being the evaluation as follows:

$$ZS_Q(A/D) = \max_{c\in CR(A/D)} \min\left(ES(A/D),c), Q(c)\right) \tag{22}$$

It can be proved [13] that 22 is equivalent to:

$$ZS_Q(A/D) = \max_{\alpha\in M(A/D)} \min(\alpha, Q\left(\frac{|(A\cap D)_\alpha|}{|D_\alpha|}\right) \tag{23}$$

Thus, once again the evaluation of a binary quantified statement can be performed without calculating the $ES(A/D)$ cardinality independently.

3 MATERIAL AND METHODS

3.1 Data sets

We have used in the experiments fifteen data sets which have been used for different Artificial Intelligence-related tasks, such as classification, regression or others, and are available in the UCI machine learning repository [20]. The most relevant quantitative features of the data sets we considered are described in Table 3.1. The data sets meet the following conditions:

- They have at least two attributes.
- They have at least one numerical attribute.
- No large-scale data set were considered.

The number of instances in the collection of 15 data sets included in the experiment is ample, and ranges from 100 to 50,000. This allowed us to test the methods performance under different data set size scenarios. It is relevant to note that we are considering larger data sets than the ones reported in the related literature of linguistic descriptions of data using fuzzy quantification (for instance, 282 instances in [1], 513 in [10] or 1,268 in [38, 39]).

Data sets with missing values were used, but some pre-processing was performed to remove the rows which contained them. Furthermore, in the "Glass" data set, the attribute "Id" was removed since it does not describe a data feature.

3.2 Fuzzy Quantified Statements

3.2.1 Linguistic variables

The sets of linguistic terms that are used to summarize and/or qualify a referential in quantified sentences are known as linguistic variables [32], e.g., *speed* = $\{$low, medium, high$\}$. This concept was originally introduced by Zadeh in his early works as a more extensive idea that involves other elements such as operators and hedges [44]. However, the simpler definition provided above is widely used for the purposes of LDD and quantified sentences in general [32].

The data sets in our study contain both categorical and numerical variables. Consequently, we created linguistic variables for both variable types, so that fuzzy quantified statements containing both kinds could be computed.

In the case of categorical variables, the different values or classes are directly taken as the linguistic terms of the corresponding linguistic variable, which were modeled as crisp sets (or singletons). For instance, the attribute *"class"* of the *"Iris"* data set has the following values: $\{$"Iris Setosa", "Iris Versicolor", "Iris Virginica"$\}$. From this kind of variables, crisp categorical variables were created using their values as linguistic terms.

On the other hand, for numerical variables linguistic terms were modeled as trapezoidal fuzzy sets. For each numerical variable in a data set, we generated four different linguistic variables, each one containing a different number of linguistic terms. Three of these linguistic variables correspond to a fuzzy partition with equidistant fuzzy sets including, respectively 3, 5 and 7 linguistic terms. The remaining linguistic variable involved one partition with 5 randomly defined fuzzy sets. As in most approaches in LDD, the fuzzy partitions for every linguistic variable were defined as trapezoidal strong fuzzy partitions [34], i.e., for each point the sum of the fulfillment degree is 1.

Figure 1 shows an example of an equidistant partition, where a single parameter α models the distance between each pair of contiguous fuzzy sets. This parameter is calculated for each numeric variable in a data set as $\alpha = (\max - \min)/(2n-1)$ being $\min, \max$ respectively the minimum and maximum values of the corresponding numerical domain and n the number of terms.

3.2.2 Quantifiers

As mentioned before, in quantified sentences such as "Most *blonde* people have blue eyes", quantifiers are necessary to evaluate the (fuzzy) amount of individuals in the referential that fulfill a given condition (in the form of a summarizer or a qualifier).

We selected seven quantifiers $\{$"At least one", "A few", "Some", "About half", "Most", "Almost all", "All"$\}$. All of them were defined as coherent fuzzy quantifiers, as this was a necessary condition by some of the fuzzy quantification models (Yager's method, Sugeno integral based method, and Vila et al. method), which can only be applied to this type of quantifiers.

We designed two different fuzzy partitions for these quantifiers: equidistant and random. Figure 2 shows the equidistant coherent definition for these quantifiers, since they are defined as monotonically increasing while fulfilling that $Q(0) = 0$ and $Q(1) = 1$. Therefore, they can be evaluated under all the six fuzzy quantification models described in Section 2.

3.3 Experiments

As mentioned before, our objective in this study was to perform a comparison among the selected fuzzy quantification models described in Section 2 with the aim of detecting significant differences, if any. This study consisted in two stages (Figure 3): *i)* Generation of linguistic descriptions, which involves two sub-stages: *i-a)* "descriptions set," where we generate all possible unary and binary fuzzy quantified statements for every fuzzy quantification model and data set and *i-b)* "quantification stage," where we evaluate them with the six fuzzy quantification models; *ii)* based on the lists of resulting fuzzy quantified statements, we run tests of statistical significance to detect significant differences between pairs of fuzzy quantification models.

3.3.1 Generation of linguistic descriptions

At this stage, we generated all possible unary and binary fuzzy quantified statements for each data set described in Section 3.1. First, we

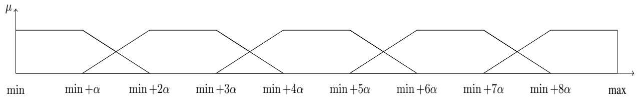


Figure 1. Linguistic variable definition from numerical variable with 5 equidistant linguistic terms.

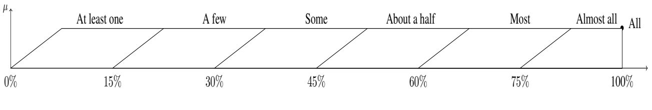
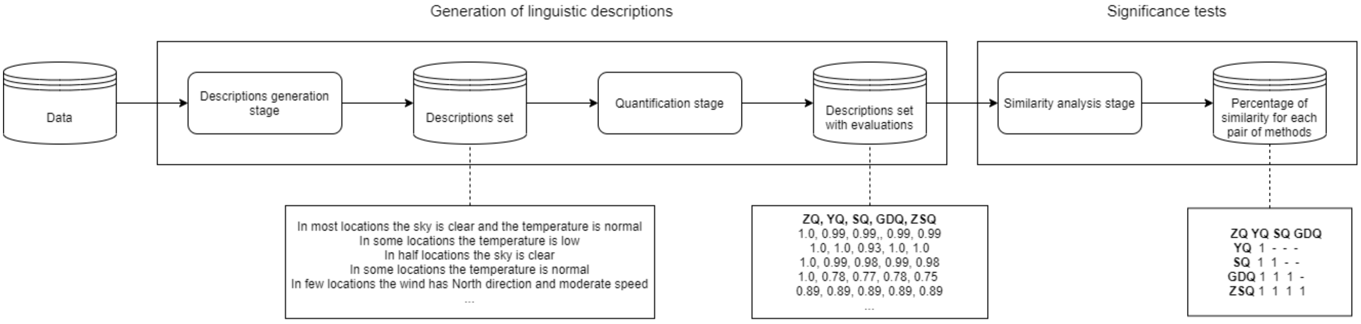


Figure 2. Equidistant definition of the quantifiers.

Table 1. Some relevant features of the data sets used in the experiments

Data set	#instances	#attributes			Associated task	Missing values
		#total	#numerical	#categorical		
Abalone [20]	4,177	8	7	1	Classification	No
Acute inflammations [11]	120	8	1	7	Classification	No
Liver disorders [20]	345	7	6	1	N/A	No
Balance [20]	625	5	4	1	Classification	No
Iris [20]	150	5	4	1	Classification	No
Blood transfusion service center [42]	748	5	4	1	Classification	No
Credit approval [20]	690	16	15	1	Classification	Yes
Wine [20]	178	14	13	1	Classification	No
Breast cancer Wisconsin [5]	699	11	10	1	Classification	Yes
Bank marketing [30]	45,211	17	7	10	Classification	No
Air quality [12]	9,358	14	14	0	Regression	Yes
Airfoil self-noise [20]	1,503	6	6	0	Regression	No
Glass identification [20]	214	11	10	1	Classification	No
Energy efficiency [36]	768	10	10	0	Classification, regression	No
Concrete compressive strength [41]	1,030	9	9	0	Regression	No



In most locations the sky is clear and the temperature is normal
In some locations the temperature is low
In half locations the sky is clear
In some locations the temperature is normal
In few locations the wind has North direction and moderate speed
...

ZQ, YQ, SQ, GDQ, ZSQ
1.0, 0.99, 0.99,, 0.99, 0.99
1.0, 1.0, 0.93, 1.0, 1.0
1.0, 0.99, 0.98, 0.99, 0.98
1.0, 0.78, 0.77, 0.78, 0.75
0.89, 0.89, 0.89, 0.89, 0.89
...

	ZQ	**YQ**	**SQ**	**GDQ**
YQ	1	-	-	-
SQ	1	1	-	-
GDQ	1	1	1	-
ZSQ	1	1	1	1

Figure 3. Description of the experimentation stages with their inputs and outputs.

generated linguistic variables from all attributes in a data set, creating both crisp categorical and fuzzy numerical linguistic variables based on the nature of the source variables, as described in Section 3.2.1.

Our unary quantified statements follow the standard form previously described (*"Q X are S"*), being Q is the set of equidistant or random quantifiers and S is the set of linguistic terms of each of the linguistic variables. For each data set, we generated all possible statements from the entire set of generated linguistic variables.

Likewise, we generated binary quantified statements following the described structure in Section 1 (*"Q KX are S"*). Once again, Q and S maintain their respective roles from unary quantified statements, and K is another term in the set of linguistic terms of the corresponding linguistic variables, where $K \neq S$. As in unary quantified statements, we generated all possible statements by obtaining the entire set of combinations of Q, K and S.

The next step of this stage is the evaluation of the statements. For each data set and fuzzy quantification model we performed two different studies: *i)* with a complete set composed of both categorical and fuzzy linguistic variables; and *ii)* considering only the subset of fuzzy linguistic variables generated from numerical attributes.

The reason for these two separate studies lies in the convergent behaviour of fuzzy methods when dealing with crisp sets. Thus, we chose to additionally study the same data sets considering only numerical variables (and consequently only fuzzy linguistic variables) to avoid any possible bias caused by the inclusion of categorical variables. Table 2 shows a summary of the number of experiments we undertook for each empirical study.

For binary quantified statements, the study based only on the fuzzy linguistic variables was performed with 14 data sets, since the *Acute inflammations* data set [11] had to be discarded as it only contains one numerical attribute. This means only one fuzzy linguistic variable can be generated and therefore can not be used to produce binary quantified statements, since these require at least two: one for K and another one for S.

The result of this stage for each data set is a pair of lists of all possible statements (one for unary quantified and another one for binary quantified sentences) with the results from the evaluation with the compared fuzzy quantification models. These lists are ordered by their associated degree of fulfillment (in descending order). Thus, for a given experiment, the test to determine the possible difference between a pair of fuzzy quantification models consists in a comparison of their corresponding sentence rankings ("all vs. all" comparison, as we will describe in what follows).

Table 2. Number of experiments performed for each scenario.

Quantifier	# Linguistic values	# Experiments
7 equidistant	3 equidistant	1
	5 equidistant	1
	7 equidistant	1
	3 random	5
	5 random	5
	7 random	5
7 random	3 equidistant	5
	5 equidistant	5
	7 equidistant	5
Total		**33**

3.3.2 Significance tests

We performed a non parametric test since our data do not meet the following conditions: normality (the distribution of the data follows a Gaussian distribution) and homocedasticity (the distributions of different groups are equal).

Considering the number of the compared fuzzy quantification models, 5 for each description type, and the data sets size, we performed the Friedman test [23] and the post-hoc Nemenyi test [31] since we made an "all vs all" comparison. According to the Nemenyi post-hoc test two methods differ significantly when $p < 0.05$ and differ highly significant when $p < 0.01$. The tests were performed separately for unary and binary quantified statements.

For unary quantified statements it was theoretically proved in the literature [15] that two pairs of the fuzzy quantification models in Section 2 are equivalent under certain conditions:

- Delgado's GD method [15] is a generalization of Yager's method [40] for relative and coherent quantifiers.
- ZS method [9] is a generalization of the Sugeno integral base method [8] also for coherent quantifiers.

Our aim is to empirically extend these results for the other eight pairs. Therefore, in order to keep consistence with the theoretically proved result, we state the null hypothesis H_0 as follows:

H_0: **There are not significant differences between the two compared fuzzy quantification models when used for evaluating the degree of fulfillment of fuzzy quantified statements.**

With regard to binary quantified statements, there are no studies that compare theoretically or experimentally the behaviour between a pair of fuzzy quantification models. In the absence of previous studies, we kept the same null hypothesis H_0 defined in the unary quantified statements scenario.

4 EXPERIMENTAL RESULTS

We have conducted a total number of 957 experiments (495 with the entire linguistic variables set and 462 with the fuzzy linguistic variables set). For detailed information about the performed tests, the complete results are available as supplementary material [3].

A summary of the results of the tests for unary models is presented in Table 3, with for linguistic variables, and in Table 4, only for fuzzy variables. These tables show the percentage of experiments where no significant differences were detected between the corresponding pair of fuzzy quantification models.

Analysing the results in Table 3, we can conclude they support the null hypothesis and the theoretical affirmations presented in [15] because in almost all experiments there were no differences between the behaviour of the fuzzy quantification models.

However, three pairs of methods, ZQ with SQ, YQ with SQ, SQ with GDQ, showed differences between their behaviour in two of five executions of the *"Wine"* data set with random partitions quantifiers. Thus, in general, only in 0.21% experiments significant differences were detected for each pair of methods, which in our opinion is not representative enough.

Analysing the properties of these pairs of methods in [13] where significant differences were detected, there exists at least one property that is fulfilled by one of the methods of the pair but not by the other: induced operators for the ZQ - SQ pair, monotonicity in the quantifiers for SQ - YQ and duality for SQ - GDQ. This could be a motivation for the significance of the differences we observed, although this hypothesis needs further study and testing. On the other hand, for the pairs of methods where no significant differences were detected, we found that at least one property exists which is fulfilled by one of the models of each pair whereas not by the other. Therefore, our study shows these pairs of methods behave similarly when evaluating unary quantified statements although they should behave differently from the point of view of the properties they fulfill.

Table 3. Percentages of cases where unary quantified models show similar behaviour for all the variables (crisp and fuzzy).

	ZQ	YQ	SQ	GDQ
YQ	100			
SQ	99.57	99.57		
GDQ	100	100	99.57	
ZSQ	100	100	100	100

Results based only on fuzzy linguistic variables (Table 4) show also the detected differences as in the previous case and also between GDQ and ZSQ only in one of the 462 fuzzy tests.

Table 4. Percentages of cases where unary quantified models show similar behaviour for the subset of fuzzy variables.

	ZQ	YQ	SQ	GDQ
YQ	100			
SQ	99.57	99.57		
GDQ	100	100	99.57	
ZSQ	100	100	100	99.78

Comparing the results of these two experiments, there is not a high dependency between the linguistic variable type and the results, since their results differ in 0.52% of the experiments.

Despite these differences, the null hypothesis H_0 is accepted in this case since detecting significant differences in 2 out of 195 tests between ZQ - SQ, YQ - SQ and SQ - GDQ and 1 difference between GDQ - ZSQ in only one data set is not representative enough of their behaviour.

Results from binary quantified statements tests are presented in Table 5. In this scenario the percentage of cases with non-significant differences is lower than 50% of the experiments in almost all pairwise comparisons, except between ZQ and GDQ (742.17% of cases with non-significant differences).

Besides, two pairs of methods, ZQ - VQ and GDQ - VQ, have the lowest percentage of similarity (5.80%), showing therefore the most different behaviour in the evaluation of binary quantified statements. This result supports the similarity of ZQ and GDQ with binary quantified statements because these two methods, which have a similar

---

[3] https://tinyurl.com/qs8au5b

behaviour evaluating binary quantified statements, have the same behaviour with respect to VQ.

In contrast with the previous case, here the null hypothesis H_0 is rejected because in almost all method pairs significant differences were detected in more than 50% of performed tests, except in the ZQ - GDQ comparison.

Table 5. Percentages of cases where unary quantified models show similar behaviour for all the variables (crisp and fuzzy).

	ZQ	YQ	VQ	GDQ
YQ	19.42			
VQ	5.80	12.75		
GDQ	72.17	21.16	5.80	
ZSQ	29.28	41.74	9.28	32.17

5 CONCLUSIONS AND FUTURE WORK

In this work, we presented an experimental comparison between the following six well-known and widely used fuzzy quantification models: Zadeh's scalar [46] and fuzzy cardinality [9], Yager's OWA [40], Delgado's GD [15], Sugeno integral [8] and Vila's VQ [37]. We tested their behaviour when evaluating unary and binary quantified sentences on fifteen data sets from the UCI machine learning repository [20].

We have analysed experimentally whether there exist significant differences between them when applied to the calculation of the degree of fulfillment in fuzzy quantified statements.

Tests results were evaluated with a pair-wise comparison performing statistical significance tests with a null hypothesis H_0 for unary and binary quantified statements which state there are not significant differences between a pair of fuzzy quantification models when calculating the fulfillment degree of the fuzzy quantified statements. H_0 is inspired on and extends the previous theoretical result [15] which show that two of the fuzzy quantification models we have studied ([15] and [9]) are respectively generalizations of [40] and [8] under certain conditions.

The experimentation for unary quantified statements only showed significant differences between three pairs of fuzzy quantification models (ZQ with SQ, YQ with SQ and GDQ with ZSQ) in 4 of 26 experiments with random partitions quantifiers of one specific data set. Thus, this study confirmed the null hypothesis for 7 pairs of fuzzy unary quantification models in the entire set of experiments. In only 4 of a total 377 experiments for three pairs of fuzzy quantification models significant differences were actually observed.

Therefore, these results point out that the selection of a fuzzy quantification model for an specific case should follow another criteria than their experimental behaviour (e.g., its theoretical properties or computational cost), since from a quantitative point of view, the differences in their empirical behaviour are, in general, not significant.

On the contrary, the experiments for binary quantified statements showed significant differences between two fuzzy quantification models at least in a 50% of experiments, except between ZQ and GDQ, where only in a 25.13% of experiments significant differences were detected. Therefore, since in 9 pair comparisons the methods behaviour is significant different in more than a 50% of the cases, the null hypothesis H_0 was rejected.

Therefore, these results point out that the selection of a fuzzy quantification model for an specific case should be very careful, since from a quantitative point of view their behaviour is significantly different. Other criteria, such as the theoretical properties these models fulfill (of fail to) become more relevant for selecting the most appropriate model for a given use or application.

As future work, we are extending our experimentation with binary quantified statements in the following ways: *i)* adding new data sets which allow us to test these methods in a wider range of cases; *ii)* considering other definitions of quantifiers and partitions of linguistic terms, consequently increasing the number of experiments; *iii)* performing a cluster-based analysis of the results of the statistical significance tests for all cases, in order to explore if it is possible to identify clear groups of fuzzy quantification models; and, finally, *iv)* extending the current analysis to other fuzzy quantification models, such as the F^A [19], among others.

ACKNOWLEDGEMENTS

This research was funded by the Spanish Ministry for Science, Innovation and Universities (grant TIN2017-84796-C2-1-R) and the Galician Ministry of Education, University and Professional Training (grants ED431C 2018/29, "accreditation 2016-2019, ED431G/08" and ED431G2019/04). All grants were co-funded by the European Regional Development Fund (ERDF/FEDER program).

REFERENCES

[1] Rui Jorge Almeida, Marie-Jeanne Lesot, Bernadette Bouchon-Meunier, Uzay Kaymak, and Gilles Moyse, 'Linguistic summaries of categorical time series for septic shock patient data', in *FUZZ-IEEE 2013, IEEE International Conference on Fuzzy Systems, Hyderabad, India, 7-10 July, 2013, Proceedings.*, pp. 1–8. IEEE, (2013).

[2] Alberto Alvarez-Alvarez and Gracián Triviño, 'Linguistic description of the human gait quality', *Engineering Applications of Artificial Intelligence*, **26**(1), 13–23, (2013).

[3] Senén Barro, Alberto Bugarín, Purificación Cariñena, and Félix Díaz-Hermida, 'A framework for fuzzy quantification models analysis', *IEEE Trans. Fuzzy Systems*, **11**(1), 89–99, (2003).

[4] Jon Barwise and Robin Cooper, 'Generalized quantifiers and natural language', in *Philosophy, language, and artificial intelligence*, 241–301, Springer, (1981).

[5] Kristin P. Bennett and Olvi L. Mangasarian, 'Robust linear programming discrimination of two linearly inseparable sets', *Optimization methods and software*, **1**(1), 23–34, (1992).

[6] Johan Benthem, 'Determiners and logic', *Linguistics and Philosophy*, **6**(4), 447–478, (1983).

[7] Patrick Bosc, 'On the comparison of the sugeno and the choquet integrals for the evaluation of quantified statements', in *Proceedings of the 3rd European Congress on Intelligent Techniques and Soft Computing (EUFIT'95), Aachen, Germany*, (1995).

[8] Patrick Bosc and Ludovic Lietard, 'Monotonic quantified statements and fuzzy integrals', in *NAFIPS/IFIS/NASA'94. Proceedings of the First International Joint Conference of The North American Fuzzy Information Processing Society Biannual Conference. The Industrial Fuzzy Control and Intellige*, pp. 8–12. IEEE, (1994).

[9] Miguel Delgado Calvo-Flores, Daniel Sánchez, and M. Amparo Vila, 'Un método para la evaluación de sentencias con cuantificadores lingüísticos', in *Actas del VIII Congreso Español sobre Tecnologías y Lógica Fuzzy: Pamplona, 8-10 de septiembre de 1998*, pp. 193–198. Departamento de Automática y Computación, (1998).

[10] Rita Castillo-Ortega, Nicolas Marín, and Daniel Sánchez, 'Linguistic local change comparison of time series', in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, pp. 2909–2915. IEEE, (2011).

[11] Jacek Czerniak and Hubert Zarzycki, 'Application of rough sets in the presumptive diagnosis of urinary system diseases', in *Artificial intelligence and security in computing systems*, 41–51, Springer, (2003).

[12] Saverio De Vito, Ettore Massera, Marco Piga, Luca Martinotto, and Girolamo Di Francia, 'On field calibration of an electronic nose for

benzene estimation in an urban pollution monitoring scenario', *Sensors and Actuators B: Chemical*, **129**(2), 750–757, (2008).

[13] Miguel Delgado, M. Dolores Ruiz, Daniel Sánchez, and M. Amparo Vila, 'Fuzzy quantification: a state of the art', *Fuzzy Sets and Systems*, **242**, 1–30, (2014).

[14] Miguel Delgado, Daniel Sánchez, Maria J. Martín-Bautista, and M. Amparo Vila, 'A probabilistic definition of a nonconvex fuzzy cardinality', *Fuzzy Sets and Systems*, **126**(2), 177–190, (2002).

[15] Miguel Delgado, Daniel Sánchez, and M. Amparo Vila, 'Fuzzy cardinality based evaluation of quantified sentences', *Int. J. Approx. Reasoning*, **23**(1), 23–66, (2000).

[16] Félix Díaz-Hermida, *Modelos de cuantificación borrosa basados en una interpretación probabilıstica y su aplicación en recuperación de información*, Ph.D. dissertation, PhD thesis, Universidade de Santiago de Compostela, 2006.

[17] Félix Díaz-Hermida, David E. Losada, Alberto Bugarín, and Senén Barro, 'A probabilistic quantifier fuzzification mechanism: The model and its evaluation for information retrieval', *IEEE Trans. Fuzzy Systems*, **13**(5), 688–700, (2005).

[18] Félix Díaz-Hermida, Marcos Matabuena, and Juan Carlos Vidal, 'The FA quantifier fuzzification mechanism: analysis of convergence and efficient implementations', *CoRR*, **abs/1902.02132**, (2019).

[19] Félix Díaz-Hermida, Martin Pereira-Fariña, Juan Carlos Vidal, and Alejandro Ramos-Soto, 'Characterizing quantifier fuzzification mechanisms: A behavioral guide for applications', *Fuzzy Sets and Systems*, **345**, 1–23, (2018).

[20] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. [Accessed November 2019].

[21] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, 'Do we need hundreds of classifiers to solve real world classification problems?', *Journal of Machine Learning Research*, (15), 3133–3181, (2014).

[22] M. Fernández-Delgado, M.S. Sirsat, E. Cernadas, S. Alawadi, S. Barro, and M. Febrero-Bande, 'An extensive experimental survey of regression methods', *Neural Networks*, **111**(Marzo), 11–34, (2019).

[23] Milton Friedman, 'The use of ranks to avoid the assumption of normality implicit in the analysis of variance', *Journal of the american statistical association*, **32**(200), 675–701, (1937).

[24] Albert Gatt and Emiel Krahmer, 'Survey of the state of the art in natural language generation: Core tasks, applications and evaluation', *Journal of Artificial Intelligence Research*, **61**, 65–170, (2018).

[25] Ingo Glöckner, 'DFS–an axiomatic approach to fuzzy quantification', *TR97-06, Techn. Fakultät, Univ. Bielefeld*, **2**(3), 10, (1997).

[26] Janusz Kacprzyk and Ronald R. Yager, 'Linguistic summaries of data using fuzzy logic', *International Journal of General System*, **30**(2), 133–154, (2001).

[27] Janusz Kacprzyk and Sławomir Zadrożny, 'Computing with words is an implementable paradigm: fuzzy queries, linguistic data summaries, and natural-language generation', *IEEE Transactions on Fuzzy Systems*, **18**(3), 461–472, (2010).

[28] Edward L Keenan and Dag Westerståhl, 'Generalized quantifiers in linguistics and logic', in *Handbook of logic and language*, 837–893, Elsevier, (1997).

[29] Nicolás Marín and Daniel Sánchez, 'On generating linguistic descriptions of time series', *Fuzzy Sets and Systems*, **285**, 6–30, (2016).

[30] Sérgio Moro, Paulo Cortez, and Paulo Rita, 'A data-driven approach to predict the success of bank telemarketing', *Decision Support Systems*, **62**, 22–31, (2014).

[31] Peter Nemenyi, *Distribution-free multiple comparisons*, Ph.D. dissertation, PhD thesis, Princeton University, 1963.

[32] Alejandro Ramos-Soto, Alberto Bugarín, and Senén Barro, 'On the role of linguistic descriptions of data in the building of natural language generation systems', *Fuzzy Sets and Systems*, **285**, 31–51, (2016).

[33] Ehud Reiter, 'An architecture for data-to-text systems', in *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pp. 97–104. Association for Computational Linguistics, (2007).

[34] Enrique H. Ruspini, 'A new approach to clustering', *Information and control*, **15**(1), 22–32, (1969).

[35] Gracián Triviño and Michio Sugeno, 'Towards linguistic descriptions of phenomena', *Int. J. Approx. Reasoning*, **54**(1), 22–34, (2013).

[36] Athanasios Tsanas and Angeliki Xifara, 'Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools', *Energy and Buildings*, **49**, 560–567, (2012).

[37] M. Amparo Vila, Juan-Carlos Cubero, Juan-Miguel Medina, and Olga Pons, 'Using owa operator in flexible query processing', in *The ordered weighted averaging operators*, 258–274, Springer, (1997).

[38] Anna Wilbik and Remco M. Dijkman, 'Linguistic summaries of process data', in *2015 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2015, Istanbul, Turkey, August 2-5, 2015*, pp. 1–7. IEEE, (2015).

[39] Anna Wilbik, Uzay Kaymak, and Remco M. Dijkman, 'A method for improving the generation of linguistic summaries', in *2017 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2017, Naples, Italy, July 9-12, 2017*, pp. 1–6. IEEE, (2017).

[40] Ronald R. Yager, 'On ordered weighted averaging aggregation operators in multicriteria decisionmaking', *IEEE Trans. Systems, Man, and Cybernetics*, **18**(1), 183–190, (1988).

[41] I-C Yeh, 'Modeling of strength of high-performance concrete using artificial neural networks', *Cement and Concrete research*, **28**(12), 1797–1808, (1998).

[42] I-Cheng Yeh, King-Jang Yang, and Tao-Ming Ting, 'Knowledge discovery on rfm model using bernoulli sequence', *Expert Systems with Applications*, **36**(3), 5866–5871, (2009).

[43] Lotfi A. Zadeh, 'Fuzzy sets', *Information and Control*, **8**(3), 338–353, (1965).

[44] Lotfi A. Zadeh, 'The concept of a linguistic variable and its application to approximate reasoning - I', *Inf. Sci.*, **8**(3), 199–249, (1975).

[45] Lotfi A. Zadeh, 'PRUF - A language for the representation of meaning in natural languages', in *Proceedings of the 5th International Joint Conference on Artificial Intelligence. Cambridge, MA, USA, August 22-25, 1977*, ed., Raj Reddy, p. 918. William Kaufmann, (1977).

[46] Lotfi A. Zadeh, 'A computational approach to fuzzy quantifiers in natural languages', in *Computational linguistics*, 149–184, Elsevier, (1983).

[47] Lotfi A. Zadeh, 'Fuzzy logic = computing with words', *IEEE Transactions on fuzzy systems*, **4**(2), 103–111, (1996).

[48] Lotfi A. Zadeh, 'From computing with numbers to computing with words: From manipulation of measurements to manipulation of perceptions', *Intelligent Systems and Soft Computing Prospects, Tools and Applications*, 3–40, (2000).

[49] Lotfi A. Zadeh, 'A new direction in AI: Toward a computational theory of perceptions', *AI magazine*, **22**(1), 73, (2001).

[50] Lotfi A. Zadeh, 'A prototype-centered approach to adding deduction capability to search engines-the concept of protoform', in *Intelligent Systems, 2002. Proceedings. 2002 First International IEEE Symposium*, volume 1, pp. 2–3. IEEE, (2002).

[51] Lotfi A. Zadeh and Janusz Kacprzyk, *Fuzzy logic for the management of uncertainty*, John Wiley & Sons, Inc., 1992.

[52] Lotfi A. Zadeh and Janusz Kacprzyk, 'CWW in Information/Intelligent Systems, vol. 1 foundations, vol. 2. aplications', *Physica-Verlag, Heidelberg*, (1999).