An Experimental Study on the Behaviour of Fuzzy Quantification Models

Andrea Cascallar-Fuentes and Alejandro Ramos-Soto and Alberto Bugarín-Diz

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 one (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 X and Y 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...
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, 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 behavior of the fuzzy quantification models. From this perspective, all the fuzzy quantification models exhibit different behavior, since all of them fulfill different properties. Also some of them exhibit non-plausible behavior for some uses, since they fail to fulfill some relevant properties. But the behavior 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.

\section{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.

\subsection{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.”

\subsection{Sum-based evaluation methods}

\textbf{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:

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

\textbf{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:

\[
C_Q(A) = \sum_{i=1}^{n} i \times (Q(i/n) - (i - 1)/n) \tag{2}
\]

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

Being \( w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right) \), \( i \in \{1, \ldots, n\} \) 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 \( A \).

\textbf{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:

\[
GD_Q(A) = \bigoplus_{i \in \{0, \ldots, n\}} \left( E(A, i) \otimes Q\left(\frac{i}{n}\right)\right) \tag{4}
\]

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:

\[
GD_Q(A) = \sum_{i=0}^{n} E(D(A, i)) \times Q\left(\frac{i}{n}\right) \tag{5}
\]

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.

\subsection{Max-min-based evaluation methods}

\textbf{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 \( Q(0) = 0, Q(1) = 1.\)
2.2.1 Sum-based evaluation methods

**Zadeh’s method** This method [9] is based on Zadeh’s fuzzy cardinality:

\[
Z(A, k) = \begin{cases} 
0 & \text{if } \frac{2\alpha}{|A_a|} = k \\
\sup \{\alpha \mid |A_a| = k\} & \text{otherwise}
\end{cases}
\]  

(7)

The evaluation for unary quantified statements with relative quantifiers is:

\[
ZS_Q(A) = \max_{\alpha \in M(A)} \min (\alpha, Q(|A_a|))
\]  

(8)

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

\[
ZS_Q(A) = \max_{\alpha \in M(A)} \min (\alpha, Q(|A_a|))
\]  

(9)

where \( M(A) = \{\alpha \in (0, 1) \mid \exists x_\alpha \in X \text{ with } A(x_\alpha) = \alpha\} \cup \{1\} \)

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:

\[
P(A/D) = \frac{P(A \cap D)}{P(D)}
\]  

(10)

The evaluation of relative quantifiers is as follows:

\[
Z_Q(A/D) = Q(P(A/D)) = Q\left(\frac{P(A \cap D)}{P(D)}\right)
\]  

(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:

\[
w_i = Q(S_i) - Q(S_{i-1}) \quad i \in \{1, ..., n\}
\]  

(12)

where

\[
S_i = \frac{1}{d} \sum_{j=1}^{i} c_i, \quad d = \sum_{k=1}^{n} e_k
\]  

(13)

being \( e_k \) the \( k \)-th low value of D set’s fulfillment degree and \( S_0 = 0 \). The evaluation is:

\[
Y_Q(A/D) = \sum_{i=1}^{n} w_i c_i
\]  

(14)

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

**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:

\[
GD_Q(A/D) = \sum_{c \in CR(A/D)} ER(A/D, c) \times Q(c)
\]  

(15)

where

\[
CR(A/D) = \left\{\frac{|(A \cap D)\alpha|}{|D\alpha|} \text{ with } \alpha \in M(A/D)\right\}
\]  

(16)

and

\[
M(A) = \{\alpha \in (0, 1) \mid \exists x_\alpha \in X \text{ with } A(x_\alpha) = \alpha\}
\]  

(17)

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

\[
GD_Q(A/D) = \sum_{\alpha \in M(A/D)} (\alpha_i - \alpha_{i+1}) \times Q(C(A/D, \alpha_i))
\]  

(18)

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

\[
C(A/D, \alpha_i) = \frac{|(A \cap D)\alpha_i|}{|D\alpha_i|}
\]  

(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. \( orness(\exists) = 1 \) and \( orness(\forall) = 0 \). Every coherent quantifier \( Q \) between \( \exists \) and \( \forall \) has an orness degree in the [0, 1] interval:

\[
o_Q = \sum_{i=1}^{n} \left(\frac{n-i}{n-1}\right) \times \left(Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right)\right)
\]  

(20)

Then, the evaluation for a binary quantified statement is:

\[
V_Q(A/D) = o_Q \max_{x \in X} (D(x) \land A(x)) \\
+ (1 - o_Q) \min_{x \in X} (A(x) \lor (1 - D(x)))
\]  

(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:

\[
ZS_Q(A/D) = \max_{\alpha \in M(A/D)} \min (ES(A/D), c), Q(c)
\]  

(22)

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

\[
ZS_Q(A/D) = \max_{\alpha \in M(A/D)} \min (\alpha, Q\left(\frac{|(A \cap D)\alpha|}{|D\alpha|}\right))
\]  

(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 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 set 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 $\alpha$ 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) “description 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
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 \≠ 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).
significant differences were detected between the corresponding pair variables. These tables show the percentage of experiments where no in Table 3, with for linguistic variables, and in Table 4, only for fuzzy

\[
H_0: \text{There are not significant differences between the two compared fuzzy quantification models when used for evaluating the degree of fulfillment of fuzzy quantified statements.}
\]

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].

<table>
<thead>
<tr>
<th>Quantifier</th>
<th># Linguistic values</th>
<th># Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 equidistant</td>
<td>3 equidistant</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5 equidistant</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>7 equidistant</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3 random</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5 random</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>7 random</td>
<td>5</td>
</tr>
<tr>
<td>7 random</td>
<td>3 equidistant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5 equidistant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>7 equidistant</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>33</td>
</tr>
</tbody>
</table>

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.

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).

<table>
<thead>
<tr>
<th></th>
<th>ZQ</th>
<th>YQ</th>
<th>VQ</th>
<th>GDQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>YQ</td>
<td>19.42</td>
<td>5.80</td>
<td>12.75</td>
<td></td>
</tr>
<tr>
<td>VQ</td>
<td>5.80</td>
<td>12.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDQ</td>
<td>72.17</td>
<td>21.16</td>
<td>5.80</td>
<td></td>
</tr>
<tr>
<td>ZSQ</td>
<td>29.28</td>
<td>41.74</td>
<td>9.28</td>
<td>32.17</td>
</tr>
</tbody>
</table>

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 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 \( F^S \) [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

[12] Saverio De Vita, Etore Massera, Marco Piga, Luca Martinotto, and Girolamo Di Francia, ‘On field calibration of an electronic nose for


