An Approximation to Context-Aware Size Modeling for Referring Expression Generation

DOI: 10.1109/FUZZ-IEEE.2018.8491506

Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez
Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada, 18071, Granada, España

Abstract

In this paper we describe a methodology for modeling context-dependent fuzzy size categories like *small* and *large*. We consider in this work that the context is fixed by a collection of crisp size values, so that the relativity in the definition of the categories is related to the distances between sizes in the context. Modeling visual concepts like those related to size is a key point, for instance, in the generation of referring expressions (conjunctions of properties) identifying objects in a certain visual scene. Taking context into account in the fuzzy modeling process is crucial in order to get human-like results. We illustrate our approach with several exam-

ples, comparing the results with other usual approaches to size category modeling.

1 Introduction

Friendly human-machine interaction necessarily involves the use of natural language as the backbone of the communication of computer applications with users. For this reason, research in the field of Natural Language Generation draws more and more the research community's attention.

A particular kind of NLG systems are the so-called data-to-text systems [1], a type of systems that bases its operation on the production of *textual information* obtained from data analysis. Among others, there are two important issues that have to be solved when developing this type of systems:

- The first one is to establish a link between the properties observed in the data and the different language labels used to express them.
- The second one is to develop algorithms that, using the above link and based on a given quality criteria, produce the most appropriate text to transfer the desired information to the user.

Zadeh's Theory of Fuzzy Sets is a tool of special utility for both issues. With respect to the first one, given the graduality of most of the concepts expressed through natural language, the modeling of the semantics of these concepts fits better within a fuzzy framework than within a precise and rigid mold; in relation to the second one, since it is basically a KDD process and, as previously said, it works on gradual concepts, there are many Soft Computing tools useful to appropriately carry it out [2].

There are numerous data-to-text systems and they are applied in very diverse areas [2, 3]. Among them, *referential games* can be found [4]. A referential game is a special type of game based on the location, within a scene, of a given object from small texts that describe and distinguish, what is more important, it from others in the scene. These texts are called referring expressions [5, 6] and, usually, are in the form of simple noun phrases (e.g., "The small circle" or "The big red circle"). These sentences can be obtained through the copulative joint of a subset of properties that, while fulfilled by the object to be referred, are not fulfilled by the rest of objects.

In our research group, we are working on the development of a reference game for teaching the children basic visual concepts such as color, size or position of simple geometric figures [7]. The fuzzy modeling of these concepts responds to different levels of complexity, among which there is a component of context awareness. In this work, we focus on the management of one of those attributes that is especially sensitive to context: the *size*. To do this, we analyze the problem and different ways of dealing with its fuzzy modeling in the particular framework of referring expression generation. As we will see, our working hypothesis is that the *fuzzy* application of a property to the objects of a set, can be inferred from the similarity relationships that are observed among the objects in the reference set on which the property is defined.

The paper is organized as follows: after this introductory section, in Section 2, the problem is described in detail. Section 3 presents a proposal for the management of size based on the use of fuzzy clustering techniques. The fourth section analyzes the use of the performance of the approach on a set of examples. We discuss about combining context-dependent properties for referring expression generation in Section 5. Finally, Section 6 highlights some conclusions and outlines interesting guidelines for future work.

2 Context and size

The problem of how to give semantics to the different labels of a linguistic variable is a recurrent problem in the field of Fuzzy Sets Theory. In some cases, it is sufficient to get a set of membership functions designed by an expert. In others, these functions are inferred from the data. Even the goal may be different: in some cases, what matters is to obtain the modeling of the concept itself; in others, what matters is to optimize the performance of a system, not being so important the modeling itself of the different concepts. Whatever the objective, what is observed on numerous occasions is a high dependence on the context that makes the determination of such semantics a non-easy to solve task [8].

Only to provide some examples: in [9], fuzzy size concepts are defined by means of a sigmoidal function whose parameters are given by data from a catalogue of objects which contains information about the target fuzzy concept; in [10], linguistic labels are assigned to a continuous value space through the use of the Fuzzy C-Means clustering technique (as we will see, our proposal shares with this work the idea of using fuzzy clustering as the tool that generates semantics); in [11], the initial membership functions of a fuzzy system are adjusted by means of the gradient descent algorithm, resulting in a hybrid learning process; finally, in [12, 13], interested readers can find how to tune the membership functions of a fuzzy control system by making use of genetic algorithms.

In our case, the modeling of the concepts is used to assign properties to the objects, so that these properties can then be used to build valid referring expressions. Thus, we are not as interested in the modeling itself as in the allocation of properties that it produces on the objects. Concretely, as we have commented in the introduction, in this work we focus in the management of the *size* property.

Size is a property particularly relevant in many applications of data-to-text systems because, among other things, humans tend to use size to identify objects when there are other objects of the same type in the scene [14], even when crisp measurements are not available [15]. When dealing with size, approaches vary: some approaches are based on a set of if-then rules to choose an appropriate size label for an object depending on object properties [16]. We can distinguish between *overall size modifiers* and *individuating size modifiers*, according to whether the label refers to the overall size of the object or to a particular dimension of the object [17]. These rules and properties can be derived either from expert knowledge or by means of a training process for binary trees that try to predict a suitable size modifier according to object properties [18].

In any case, the management of size is context-dependent and is an open problem in this type of systems. Following, some examples to illustrate this idea are presented.

Consider the scene depicted in Figure 1. In this scene, to refer to object B, it is enough to say *the large object*. However, the same object B cannot be referred to in such a way in the scene depicted in Figure 2. In this case, a more appropriate referring expression would be *the medium-sized object*. The appearance of a third object C *changes* the way we can refer to object B. Thus, it is clear that working with a fixed restriction on, e.g., the diameter's length, does not allow to correclty interact with the user. The sensitivity to the context is clear. The problem becomes more complicated as the variety of sizes present in the scene increases.

Considers now the scenes depicted in Figures 3 and 4. In these scenes, the appearance of a new object changes the size labeling again. While in the scene of Figure 4, object B can be referred to by *the medium-sized object*, in the scene

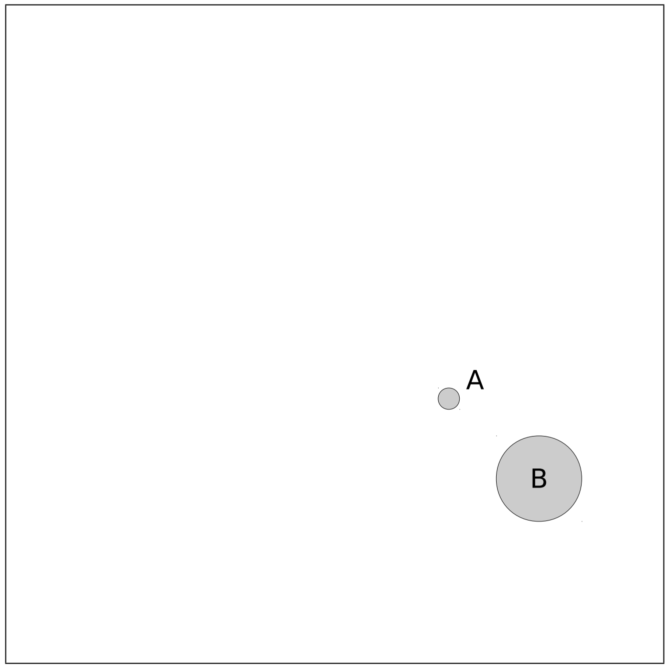

Figure 1: Scene with objects A and B

of Figure 3, however, the closeness in size of the larger object complicates the labeling of the object. Object B can be considered medium-sized, but it can also be considered to be large. C also has this problem, although compatibility with medium-sized is lower and compatibility with large grows. In both cases, if the size is used to refer to these objects, the success of the referring expression is questioned.

3 Our approach

We have seen that assigning a size label to an object in a scene is dependent on the size of the rest of the objects that appear in the scene. In this section, we introduce a method that allows us to derive the degree of compatibility of an object with each of the size labels based on the context determined by the

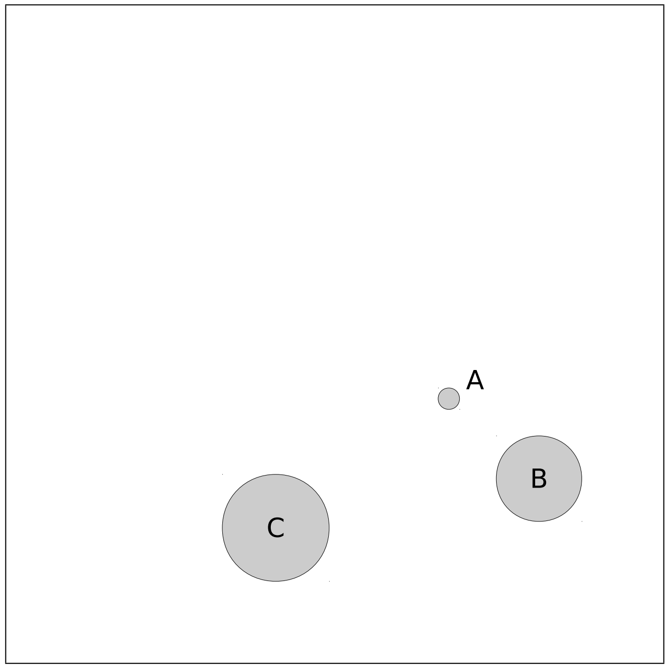

Figure 2: Scene with objects A, B, and C

given scene. Concretely, we work with a linguistic variable of size with three labels (namely, large, medium-sized, and small) and on the hypothesis that the similarity between the sizes of the objects in a precise referential set allows to infer the fuzzy semantics of the linguistic labels.

3.1 Formal framework

Let us consider a set of sizes $S = \{s_1, \ldots, s_n\}$ with $n \geq 2$, $s_i \in \mathbb{R}$, and $0 < s_1 < s_2 < \cdots < s_n$. For instance, S may have been obtained from an image I containing $m \geq n$ objects by measuring the size of each object (we consider the possibility that more than one object share the same size). Our approach relies on the following ideas:

- Our objective is to define the fuzzy categories *small*, *medium*, and *large*

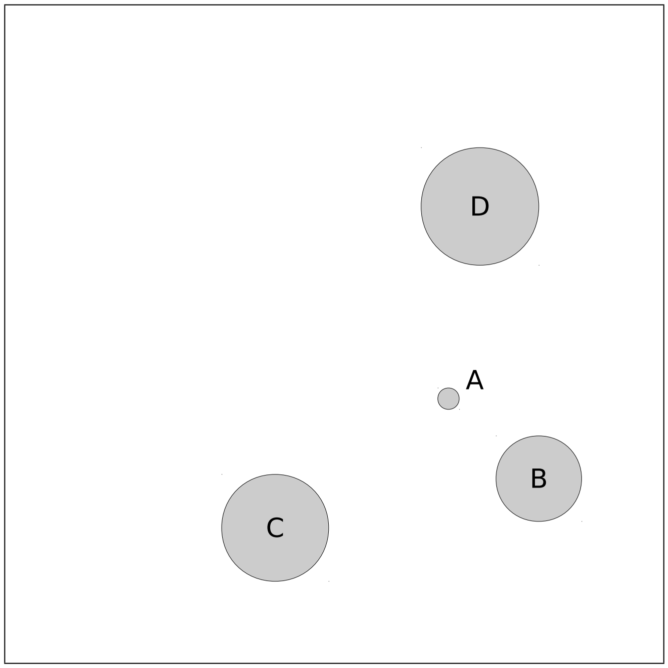

Figure 3: Scene with objects A, B, C, and D

as fuzzy subsets on S.

- We impose $small(s_1) = 1$, $small(s_n) = 0$, $large(s_1) = 0$, $large(s_n) = 1$. That is, s_1 and s_n are fully representative of *small* and *large*, respectively.
- We also designate sizes as representatives of *medium* when $n > 2$. When n is odd we designate a single representative $s_{(n+1)/2}$. When n is even we designate two representatives: $s_{n/2}$ and $s_{(n/2)+1}$.
- Membership of s_i to the fuzzy category *small* (resp. *large*) is expected to be proportional to the Euclidean distance between s_i and s_1 (resp. s_n). Similarly, when $n > 2$, membership to *medium* is proportional to the closer representative of this category.
- Models for the three fuzzy categories are obtained by fuzzy clustering. The

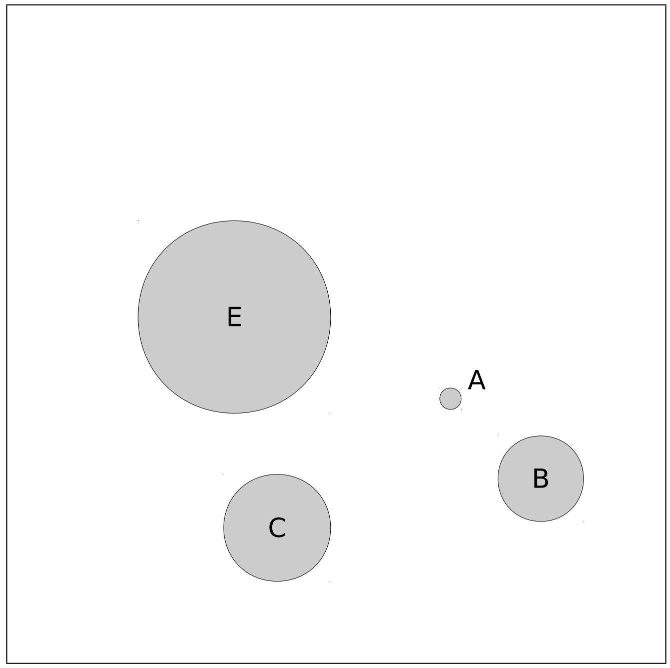

Figure 4: Scene with objects A, B, C, and E

fuzzy category *small* is modeled by the fuzzy cluster containing s_1 with degree 1. Similarly, the fuzzy category *large* is modeled by the fuzzy cluster containing s_n with degree 1. The fuzzy category *medium* is modeled by the union of fuzzy clusters containing representatives of *medium*. When $n = 2$, *medium* is modeled by the empty set.

3.2 Clustering

We shall employ an approach to clustering based on representations by levels [19, 20]. The starting point is a reflexive and symmetric binary fuzzy relation defined on S. This relation, which measures the similarity between object sizes, will be calculated from distances between sizes in S by considering a parameter d_m defining the distance beyond which the resemblance between sizes is 0, in other words, the distance beyond which we consider two object sizes completely different. Let $s_i, s_j \in S$ and let $d(s_i, s_j) = |s_i - s_j|$ be the Euclidean distance between them. Then, the resemblance $R^{d_m}(s_i, s_j)$ is calculated as:

$$R^{d_m}(s_i, s_j) = \max\left\{1 - \frac{d(s_i, s_j)}{d_m}, 0\right\} \tag{1}$$

It is easy to show that R^{d_m} so defined is reflexive and symmetric. When d_m is known, we shall note R^{d_m} as simply R for the sake of brevity. In the case $n = 2$, $d_m \leq s_n - s_1$ is required. When $n > 2$, let s', s'' with $s' \leq s''$ be the designated representatives for *medium* (if n is odd then $s' = s''$); we shall always consider parameter d_m so that $0 < d_m \leq min(d(s_1, s'), d(s'', s_n))$. With these conditions, we guarantee that for any $s \in \{s', s''\}$ it is

$$R(s_1, s) = R(s, s_n) = R(s_1, s_n) = 0$$

as we want s_1, s_n, and the pair s', s'' to be fully representative of the concepts *small*, *large* and *medium*, respectively. That means that an object being in one of the categories with degree 1 is expected to be in the rest of categories with degree 0.

Clustering based on representation by levels performs crisp clustering in the different levels $\Lambda(R) = \{R(s_i, s_j) \mid (s_i, s_j) \in \mathit{support}(R)\}$. Note that this is a finite set $\Lambda(R) = \{\alpha_1, \ldots, \alpha_k\}$ with $k \geq 1$ and $1 = \alpha_1 > \cdots > \alpha_{k+1} = 0$. For each $\alpha_k \in \Lambda(R)$, a crisp clustering based on coverings is obtained as follows:

1. Compute the α_k-cut of R, R_{α_k}, which is always a reflexive and symmetric crisp binary relation.

2. Compute the clustering of S at level α_k as the (unique) set of maximal cliques in R_{α_k}. Let the set of clusters obtained be $\mathcal{C}_{\alpha_k} \subset \{0,1\}^S$. Note that clusters so obtained may overlap, yielding a collection of crisp clusters

that form a covering of S. However, it is $\emptyset \notin \mathcal{C}_{\alpha_k}$ and $C^i_{\alpha_k} \not\subset C^j_{\alpha_k}$ for every $C^i_{\alpha_k}, C^j_{\alpha_k} \in \mathcal{C}_{\alpha_k}$, among other properties, since all clusters are maximal cliques.

Note that, by the restriction imposed on d_m, it is $R(s_1, s_n) = 0$ and hence $|\mathcal{C}_{\alpha_k}| \geq 2$ since s_1 and s_n are in different clusters. Also, it is easy to show that there is a single cluster $C^{small}_{\alpha_k} \in \mathcal{C}_{\alpha_k}$ such that $s_1 \in C^{small}_{\alpha_k}$ and a single cluster $C^{large}_{\alpha_k} \in \mathcal{C}_{\alpha_k}$ such that $s_n \in C^{large}_{\alpha_k}$ with $C^{small}_{\alpha_k} \neq C^{large}_{\alpha_k}$. Finally, following the ideas we explained above, in the case $n > 2$ we define

$$C^{medium}_{\alpha_k} = \bigcup \{ C^i_{\alpha_k} \mid C^i_{\alpha_k} \cap \{s', s''\} \neq \emptyset \} \tag{2}$$

That is, $C^{medium}_{\alpha_k}$ is the union of all clusters containing at least one of the representatives of the category *medium*. Note that $C^{medium}_{\alpha_k}$ may be empty.

From the clustering in every level, the fuzzy clusters modeling the three properties *small*, *medium*, and *large* are obtained following the ideas in [19, 20] as follows: $\forall s_i \in S$,

$$small(s_i) = \sum_{\alpha_k \mid s_i \in C^{small}_{\alpha_k}} (\alpha_k - \alpha_{k+1}) \tag{3}$$

$$medium(s_i) = \sum_{\alpha_k \mid s_i \in C^{medium}_{\alpha_k}} (\alpha_k - \alpha_{k+1}) \tag{4}$$

$$large(s_i) = \sum_{\alpha_k \mid s_i \in C^{large}_{\alpha_k}} (\alpha_k - \alpha_{k+1}) \tag{5}$$

3.3 Example

Let us consider $S = \{10, 20, 25, 30, 40\}$ ($n = 5$). Then, $s_1 = 10$ is the representative for *small*, $s_5 = 40$ is the representative for *large*, and $s_3 = 25$ is the single representative for *medium* since n is odd. Let $d_m = 15$, which is the

α_k	$C^{small}_{\alpha_k}$	$C^{medium}_{\alpha_k}$	$C^{large}_{\alpha_k}$
1	$\{10\}$	$\{25\}$	$\{40\}$
2/3	$\{10\}$	$\{20, 25, 30\}$	$\{40\}$
1/3	$\{10, 20\}$	$\{20, 25, 30\}$	$\{30, 40\}$

Table 1: Crisp clusters by levels with $d_m = 15$.

maximum possible value in this example. Since the set of distances between sizes is $\{0, 5, 10, 15, 20, 30\}$ then it is $\Lambda^{15}(R) = \{1, 2/3, 1/3\}$. Table 1 shows the crisp clusters in each level. On the basis of this information, and using Eqs. (3)-(5), we have the following clusters:

$$\begin{aligned} small &= 1/10 + (1/3)/20 \\ medium &= (2/3)/20 + 1/25 + (2/3)/30 \\ large &= (1/3)/30 + 1/40 \end{aligned}$$

In order to illustrate the effect of changing d_m, let us consider the same example with $d_m = 10$. We have $\Lambda^{10}(R) = \{1, 1/2\}$. Table 2 shows the crisp clusters in each level, from which the following clusters are obtained using Eqs. (3)-(5):

$$\begin{aligned} small &= 1/10 \\ medium &= 0.5/20 + 1/25 + 0.5/30 \\ large &= 1/40 \end{aligned}$$

We can see that values are equal or lower. It is easy to show in general that lower values of d_m provide lower membership values, yielding more separate clusters.

α_k	$C_{\alpha_k}^{small}$	$C_{\alpha_k}^{medium}$	$C_{\alpha_k}^{large}$
1	$\{10\}$	$\{25\}$	$\{40\}$
1/2	$\{10\}$	$\{20, 25, 30\}$	$\{40\}$

Table 2: Crisp clusters by levels with $d_m = 10$.

With these examples, we have illustrated how considering the largest and smallest size as representatives of the corresponding labels, and d_m as the distance defining their support, we can easily determine the memberships of sizes to the three labels.

As a final comment, let us note that sizes in this example are set symmetrically with respect to 25, and hence memberships to *small* and *large*, and also to *medium*, are assigned symmetrically with respect to distance to 25. We shall see non-symmetric cases in the following section.

4 Experiments

Once we have presented our approach, in this section we show some experiments over simple illustrative scenes for the sake of clarity, intended to illustrate the behaviour of our methodology, as well as to compare our results with other approaches. Specifically, we shall compare our results with those provided by a predefined fuzzy partition of the size domain (called universal partition), and also with a scaling of the universal fuzzy partition to the range of sizes defining the context (called scaled partition). For this comparison, we shall employ the fuzzy partition shown in Fig. 5 as universal partition. In this case, the scaled partition is a scaling of the universal partition, from the range [10,90] to the range [mins, maxs]; where mins is the minimum size of the objects in the context, and maxs is the maximum size of the objects in the context.

We have considered five scenes, shown in Figs. 6 to 10 (we include the object sizes on each Figure caption). Objects are labelled with capital letters

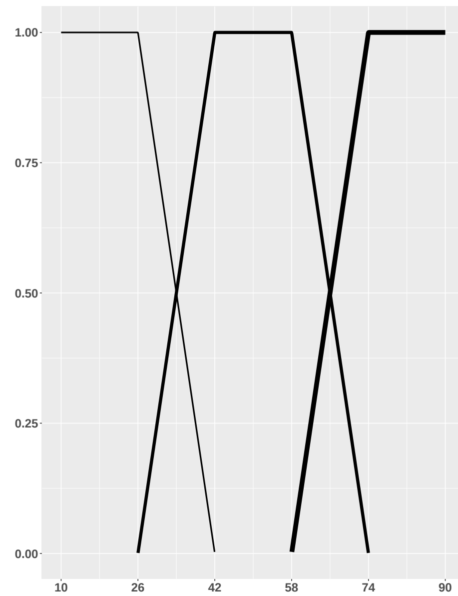

Figure 5: Universal partition

in increasing size order; their respective sizes are chosen to be distinguishable. Results are graphically represented in Table 3, including the model obtained for each of the size categories *large*, *medium*, and *small* using the two approaches based on the universal partition, and our method with d_m being the largest possible one according to the restrictions introduced in the previous section.

4.1 Example scene 1

In this example it is $S = \{10, 50, 90\}$. We can see that the obtained result is the same for the three approaches, particularly because the universal and scaled partitions are the same since the range of sizes in S is that of the universal partition. It also coincides with our approach since the sizes in this example are symmetrically defined and fall within the respective cores of the three categories.

4.2 Example scene 2

In this example it is $S = \{50, 55, 75\}$. In this case, the central value 55 is not situated in the middle of the extreme values. Whilst our method takes 50, 55 and 75 as fully representatives of *small*, *medium* and *large*, the universal partition considers that none of these values fit *small*, as it defines a non-contextual model for the categories. According to the universal partition, both 50 and 55 are medium, and 75 is large.

On its turn, the scaled partition consider that no size in S fits the category *medium* in this example, yielding *large* for 75 and *small* for both 50 and 55. This is acceptable in absolute terms, when we want to describe the object, but not in relative ones, since the size for object B is distinguishable from the size of A, and hence the referring expression "the middle sized object" can be used to successfully refer to B so that a human being can identify it. Therefore while the scaled partition tells us that objects A and B are indistinguishable by its size, our method states that we can use size to point each object to a human user successfully, which is our aim.

4.3 Example scene 3

In this example it is $S = \{50, 55, 75, 90\}$. The scene has been obtained by adding a new object D to the scene 2. This example illustrates how perception of size category, and hence membership of objects into those categories, can change because of the relative relation between sizes. The universal partition is again not working, particularly determining that there is no small object in the scene, something that is counterintuitive for human beings (object A). The scaled partition yields a good result in absolute terms again. On the other hand, our method performs better in order to interact with a human user. While the scaled partition states that we have to refer to the set of "small objects" (which can be the set {A,B,C} for a human user), our method tells us that there is only one small object which can be easily determined and that there is a set of "medium-sized objects", which is more identifiable as the user can compare with the objects A and D, and determine which objects are the "medium-sized" ones.

4.4 Example scene 4

In this example it is $S = \{10, 20, 50\}$. This example is similar to that of scene 2 in that we have three values, with the central one not being halfway between the extremes. In this case, it is *large* which is empty following the universal partition, whilst the scaled partition yields an almost empty set as representation for *medium* which, again, is counterintuitive for humans, as we have already shown.

4.5 Example scene 5

In this example it is $S = \{10, 20, 50, 60, 90\}$. It corresponds again to adding new sizes to a previous scene, in this case to scene 4. We have again that the universal and scaled partition are the same. We can see that both in the scaled partition and our method there is an important influence of the context; for instance, size 50 is absolutely large in scene 4, whilst it is absolutely medium in scene 5 due to the introduction of larger objects. In this case, the scaled partition tell us that exist a set of "large objects" which is not a singleton, again we can see that a human user will identify the "largest object" of the scene more easily than the set of "large objects", so the model obtained by our method is the most intuitive model.

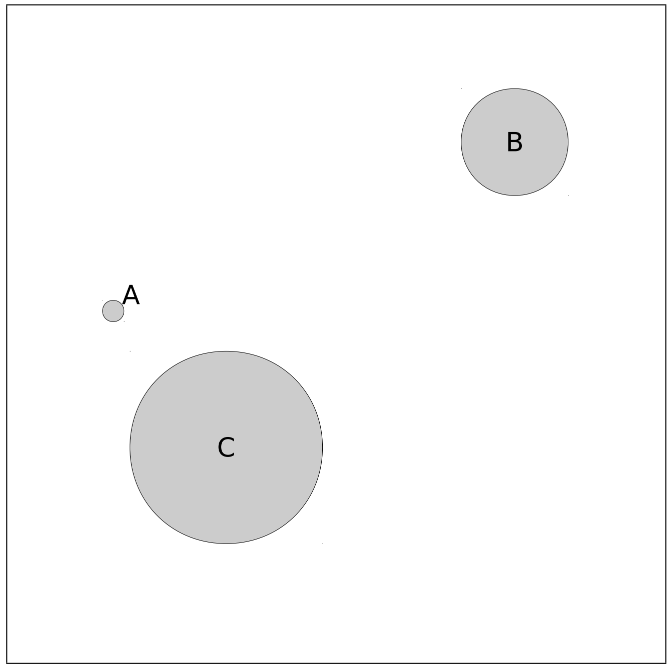

Figure 6: Example scene 1. Object sizes: {10,50,90}.

5 Discussion

As we discussed in the introduction, our proposal is developed within the context of referring expression generation. From a knowledge representation point of view, a referring expression can be seen as a conjunction of properties.

In the previous sections we have introduced a methodology to model size categories, for the sake of referring, taking into account the context. In the examples we have employed in this paper we have considered a collection of sizes corresponding to the objects in an image. However, this context may vary during the process of generating a referring expression due to the influence of other properties.

Consider for instance the image in Fig. 11. If we want to model the category *large* for generating the referring expression "The large object", we must take

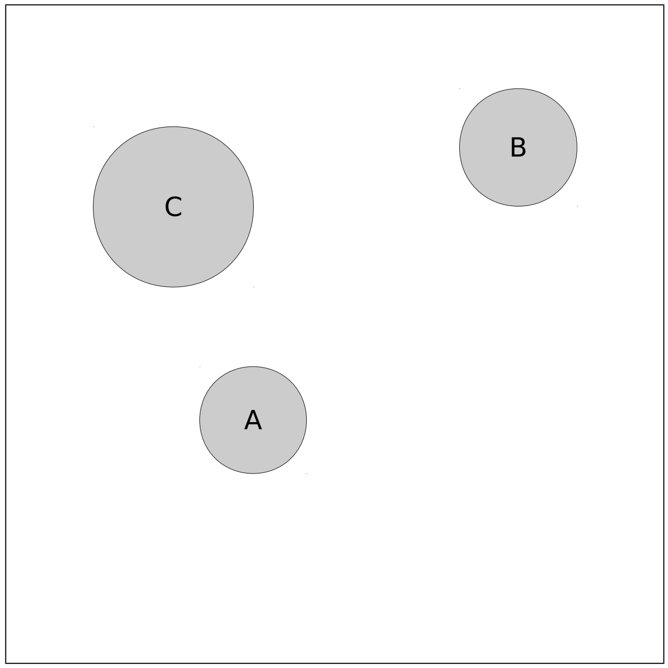

Figure 7: Example scene 2. Object sizes: {50,55,75}.

into account all the objects in the image. Hence, the collection of sizes will be $S = \{10, 20, 50, 60, 90\}$. However, if the referring expression we are evaluating is "The large black object", the context in which we are defining the category *large* is that of the black objects only, that is, $S = \{10, 20, 50\}$. That is, one of the properties appearing in the referring expression (black) restrict the collection of objects on which we want to identify the referent object by means of size, hence determining the context for defining such property. As we saw in our experiments (see scenes 4 and 5), the models for the different fuzzy size categories are very different in these two cases, something that will influence the ability of the different size categories to be part of a valid referring expression.

The situation is even more complicated when we are dealing with two fuzzy properties F_1 and F_2 during a referring expression generation process, because:

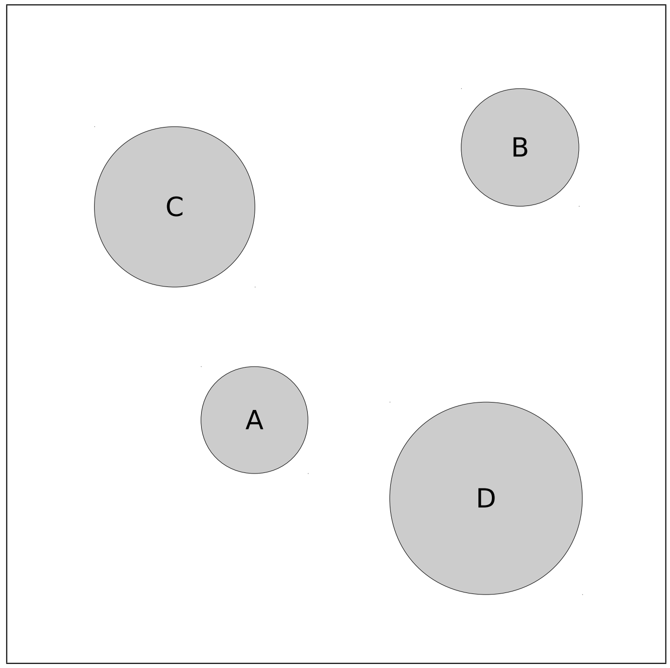

Figure 8: Example scene 3. Object sizes: {50,55,75,90}.

- Once a fuzzy property F_1 is employed, we have a fuzzy restriction on the set of objects. Hence, the context itself on which to determine the models for F_2 is a fuzzy subset of objects. For solving this problem, our methodology (which is designed for a context defined by a crisp set of objects) must be adapted. This will be an object of future research.

- The order in which the properties are applied matters. Different results are expected in general when defining F_1 on the context given by the objects satisfying F_2, and when defining F_2 on the context given by the objects satisfying F_1. This fact introduces an additional complexity in the referring expression generation problem since, instead of having a single fuzzy model for the conjunction $F_1 \wedge F_2$ as a potential referring expression, we have two possible fuzzy models depending on the ranking we consider:

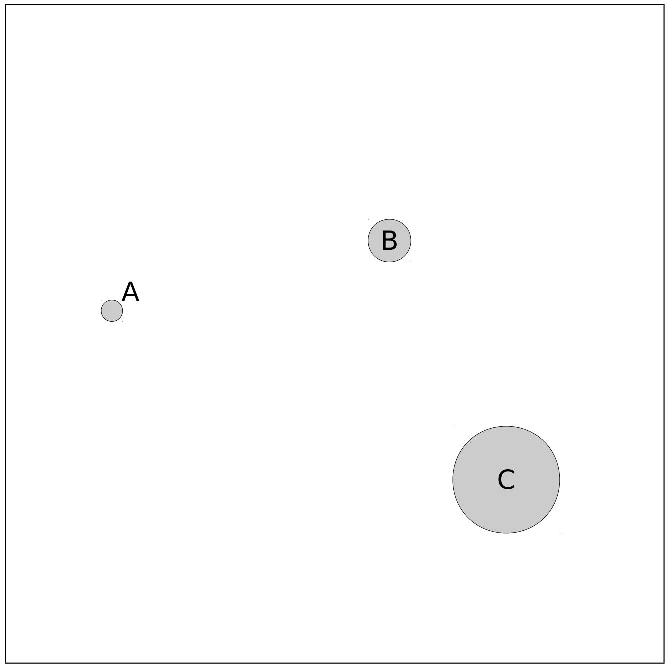

Figure 9: Example scene 4. Object sizes: {10,20,50}.

one for "the objects satisfying F_1 that satisfy F_2" and another one for "the objects satisfying F_2 that satisfy F_1". Note that when the semantics of both F_1 and F_2 are context-independent, both models are the same and can simply be described as "the objects satisfying F_1 and F_2", which is the way referring expressions are usually understood.

6 Conclusions and future work

Our approach to context-dependent fuzzy size category modeling is based on defining the context as a crisp set of sizes, determining certain sizes in the context as representatives of the different categories to model (in our work, only *small*, *medium*, and *large*), and computing membership to categories on the basis of a clustering procedure by levels based on distance between sizes.

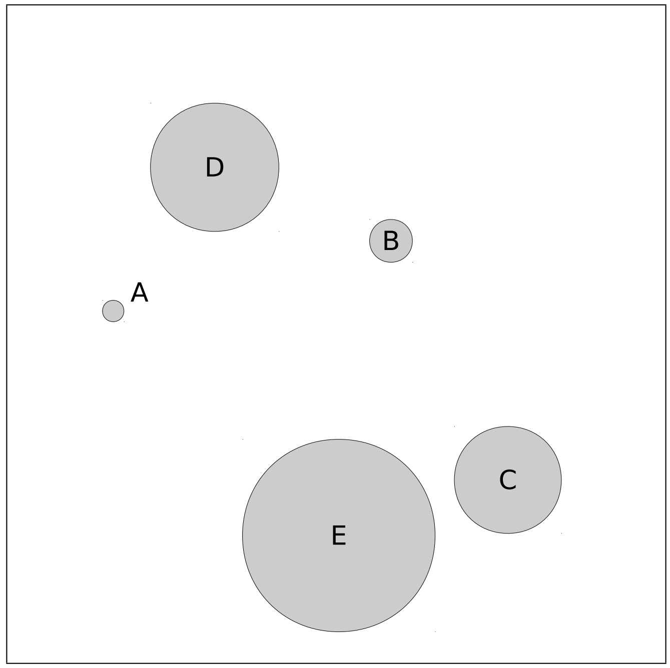

Figure 10: Example scene 5. Object sizes: $\{10,20,50,60,90\}$.

This way, the fuzzy sets modeling the fuzzy categories vary with the set of sizes involved, so that the categories are relative to the representative prototypes. In other words, removing a size or adding a new one to the context may change memberships to the fuzzy categories.

One important consequence of using context-dependent properties is that a referring expression cannot be seen as a symmetric conjunction of properties anymore, since the ranking of properties in the expression affects the context on which every property is defined. In the particular case of fuzzy properties, once a fuzzy property is considered, the context (set of objects satisfying the set property) on which to define the other one is a fuzzy set. Extending our methodology to the problem of fuzzy modeling on a fuzzy context will be dealt with in future works.

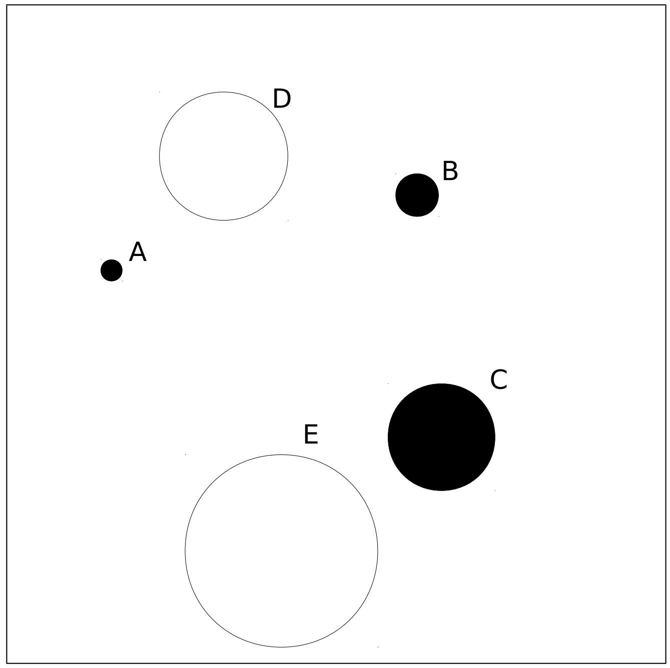

Figure 11: Objects with different color and size.

Acknowledgment

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund - ERDF (Fondo Europeo de Desarrollo Regional - FEDER) under project TIN2014-58227-P *Descripción lingüística de información visual mediante técnicas de minería de datos y computación flexible*. This work has also been partially supported by the Spanish Ministry of Education, Culture and Sport grant FPU16/05199 and by the University of Granada grant *Becas de Iniciación para estudiantes de Máster del Plan Propio de Investigación (2017)*.

References

[1] E. Reiter and R. Dale, *Building Natural Language Generation Systems*. Cambridge, UK: Cambridge University Press, 2000.

[2] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets and Systems*, vol. 285, pp. 6–30, 2016.

[3] A. Ramos-Soto, A. Bugarín, and S. Barro, "On the role of linguistic descriptions of data in the building of natural language generation systems," *Fuzzy Sets and Systems*, vol. 285, pp. 31–51, 2016.

[4] M. Franke, *On Scales, Salience and Referential Language Use*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 311–320.

[5] E. Krahmer and K. van Deemter, "Computational generation of referring expressions: A survey," *Computational Linguistics*, vol. 38, no. 1, pp. 173–218, 2012.

[6] K. van Deemter, *Computational Models of Referring: A Study in Cognitive Science*. MIT Press, 2016.

[7] N. Marín, G. Rivas-Gervilla, and D. Sánchez, "Scene selection for teaching basic visual concepts in the Refer4Learning app," in *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE Naples, Italy*, 2017.

[8] J. T. Cadenas, N. Marín, and M. A. V. Miranda, "Context-aware fuzzy databases," *Appl. Soft Comput.*, vol. 25, pp. 215–233, 2014.

[9] I. Zukerman, S. N. Kim, T. Kleinbauer, and M. Moshtaghi, "Employing distance-based semantics to interpret spoken referring expressions," *Comput. Speech Lang.*, vol. 34, no. 1, pp. 154–185, Nov. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.csl.2015.01.002

[10] D. Özdemir and L. Akarun, “A fuzzy algorithm for color quantization of images,” *Pattern Recognition*, vol. 35, no. 8, pp. 1785–1791, 2002.

[11] M.-S. Chen and S.-W. Wang, “Fuzzy clustering analysis for optimizing fuzzy membership functions,” *Fuzzy sets and systems*, vol. 103, no. 2, pp. 239–254, 1999.

[12] A. Homaifar and E. McCormick, “Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms,” *IEEE transactions on fuzzy systems*, vol. 3, no. 2, pp. 129–139, 1995.

[13] K. Shimojima, T. Fukuda, and Y. Hasegawa, “Self-tuning fuzzy modeling with adaptive membership function, rules, and hierarchical structure based on genetic algorithm,” *Fuzzy sets and systems*, vol. 71, no. 3, pp. 295–309, 1995.

[14] S. Brown-Schmidt and M. K. Tanenhaus, “Watching the eyes when talking about size: An investigation of message formulation and utterance planning,” *Journal of Memory and Language*, vol. 54, no. 4, pp. 592–609, 2006.

[15] M. Mitchell, K. van Deemter, and E. Reiter, “Natural reference to objects in a visual domain,” in *Proceedings of the 6th international natural language generation conference*. Association for Computational Linguistics, 2010, pp. 95–104.

[16] M. Mitchell, K. Van Deemter, and E. Reiter, “Two approaches for generating size modifiers,” in *Proceedings of the 13th European Workshop on Natural Language Generation*. Association for Computational Linguistics, 2011, pp. 63–70.

[17] ——, “On the use of size modifiers when referring to visible objects,” in *Proceedings of the Cognitive Science Society*, vol. 33, no. 33, 2011.

[18] M. Mitchell, K. van Deemter, and E. Reiter, "Applying machine learning to the choice of size modifiers," in *Proceedings of the 2nd PRE-CogSci Workshop*, 2011.

[19] D. Sánchez, M. Delgado, M. Vila, and J. Chamorro-Martínez, "On a non-nested level-based representation of fuzziness," *Fuzzy Sets and Systems*, vol. 192, pp. 159–175, 2012.

[20] D. Dubois and D. Sanchez, "Fuzzy Clustering based on Coverings," in *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, ser. Studies in Fuzziness and Soft Computing, C. Borgelt, M. A. Gil, J. M. C. Sousa, and M. Verleysen, Eds. Springer, 2013, vol. 285, pp. 319–330.

[21] E. Makalic, I. Zukerman, M. Niemann, and D. Schmidt, "A probabilistic model for understanding composite spoken descriptions," *PRICAI 2008: Trends in Artificial Intelligence*, pp. 750–759, 2008.

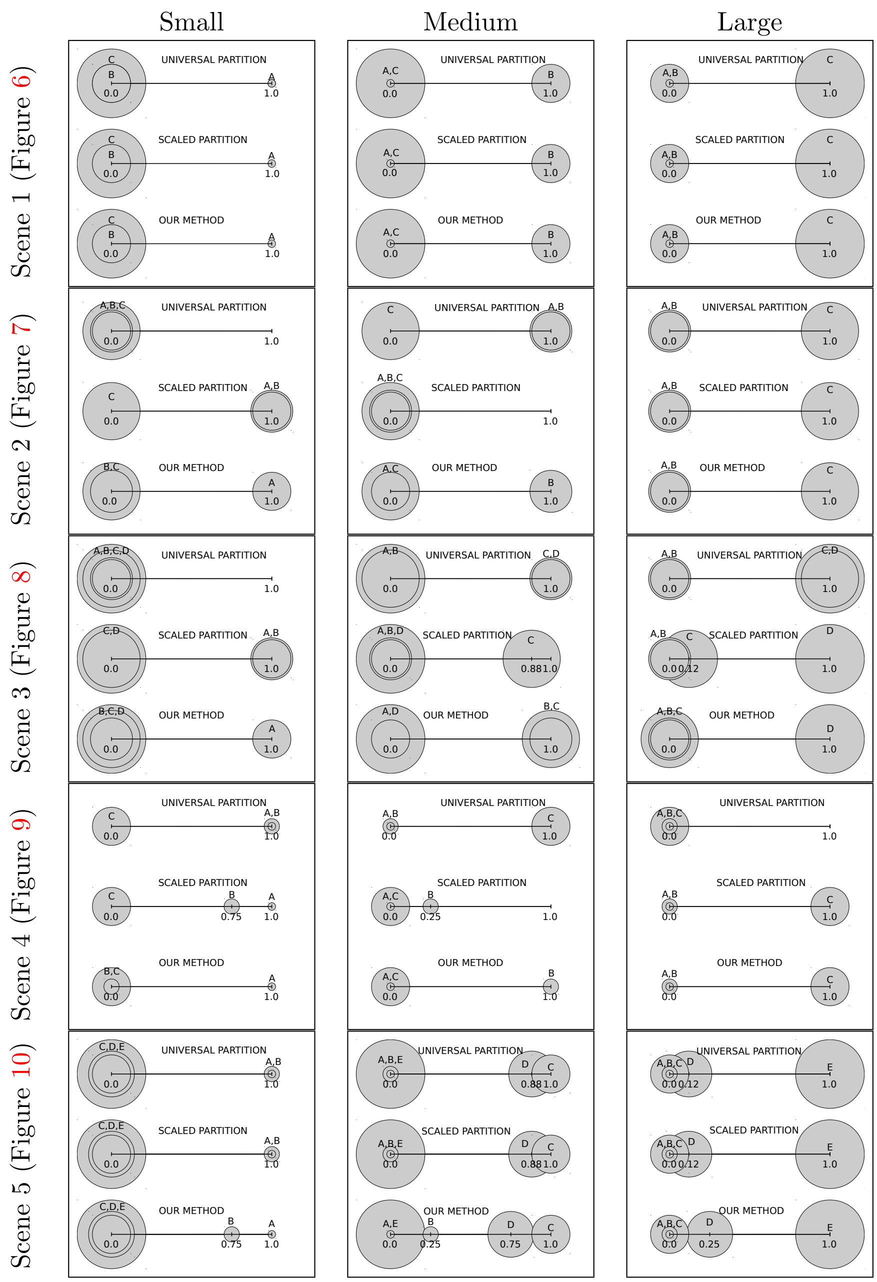

Table 3: Model representations. For each scene and each size category, we arrange the objects over the interval [0,1] according to the degree that each model assign to them for each size category.