Strengthening fuzzy gradual rules through "all the more" clauses

Bernadette Bouchon-Meunier, Anne Laurent, Marie-Jeanne Lesot, Maria Rifqi

Strengthening fuzzy gradual rules through "all the more" clauses

Bernadette Bouchon-Meunier, *Senior Member, IEEE*, Anne Laurent, Marie-Jeanne Lesot and Maria Rifqi

***Abstract*— Fuzzy gradual rules of the form *the more X is A, the more Y is B* linguistically express information about the correlation between attributes and their co-variation. They thus provide valuable information summarizing the trends observed in a given data set. In this paper, we consider strengthened fuzzy gradual rules, *i.e.* gradual rules enriched with a clause introduced by the expression "all the more": such rules of the form *the more X is A, the more Y is B, all the more Z is C* offer additional precisions on the relation between the attributes. We study the definition of such strengthened rules, discussing their possible semantics, considering several interpretations of fuzzy gradual rules. We then propose quality criteria as well as a mining algorithm.**

I. Introduction

The extraction of information describing digital data, their inner trends and exceptional behaviors, can take many different forms, leading to various pieces of knowledge delivered to experts. In this paper, we focus on fuzzy gradual rules that convey information in the form of attribute co-variations, such as *the closer the wall, the harder the brakes are applied*, or more generally *the more* X_1 *is* A_1*, and ... and the more* X_n *is* A_n*, then the more* Y_1 *is* B_1*, and ... and the more* Y_p *is* B_p: such rules consider attributes (*e.g.* wall distance, braking, or more generally X_i and Y_i), associated to fuzzy modalities (*e.g.* close, hard, or more generally A_i or B_i), and data described by their membership degrees to these fuzzy modalities. They then establish links between the values of these membership degrees. Such rules were first proposed and interpreted as a special case of inference rules [1], [2], [3], [4], modeled by a fuzzy r-implication applied to each data point individually. A different approach of fuzzy gradual rules interprets them as global tendencies across the whole data set, imposing correlations on the variations of the membership degrees [5], [6], [7], [8], [9].

In this paper, we consider, for both interpretations, enriched gradual rules, that add strengthening information, linguistically expressed by clauses introduced by the expression "all the more". Such rules are of the form *the closer the wall, the harder the brakes are applied, all the more the higher the speed*: in this example, the information about the speed enriches by an additional precision the relation established between wall distance and braking.

B. Bouchon-Meunier, M.-J. Lesot and M. Rifqi are with the University Pierre et Marie Curie-Paris6, CNRS UMR 7606, LIP6, 104 avenue du Président Kennedy, Paris, F-75016 Paris. (phone: 33(0)144 278 886; email: {bernadette.bouchon-meunier,marie-jeanne.lesot,maria.rifqi@lip6.fr})

A. Laurent is with the University Montpellier 2, CNRS UMR 5506, 161 rue Ada, 34095 Montpellier (phone: 33(0)467 418 631 ; email: anne.laurent@lirmm.fr).

We discuss the question of the semantics of this strengthening clause, considering the frameworks corresponding to various interpretations of fuzzy gradual rules and the strengthening effect in each case. We then propose several quality criteria adapted to these different semantics. We also present a mining algorithm to automatically extract such rules from data sets.

The paper is organized as follows: Section 2 recalls the principles of fuzzy gradual rules, discussing their interpretations and recalling their extraction methods. In Section 3, we discuss the strengthening semantics and we detail in Section 4 the notions of support, confidence, and other quality criteria that can be considered in each case. Lastly in Section 5 we describe an algorithm to automatically extract such strengthened gradual rules.

II. Fuzzy gradual rules

Several definitions and approaches exist for the notion of fuzzy gradual rules, whose aim is to establish correlation between attributes to summarize relevant trends observed in the data. In this section, we first recall the classic notations and definitions of fuzzy gradual item, itemset and rules, as given *e.g.* by [6], [8]. We then compare the main two approaches, that respectively apply rules individually to each data point, in a fuzzy implication interpretation, or consider the rules as co-variation constraints applied across the whole data set, recalling the associated quality criteria.

A. Formalizations of fuzzy gradual item, itemsets and rules

A fuzzy data set $\mathcal{D}$ describes data through attributes corresponding to fuzzy linguistic variables, associated to fuzzy modalities: one can for instance have an attribute X = speed, with the modality set {*low, normal, high*}. The data are then described by membership degrees that indicate the extent to which their characteristics belong to the considered modalities. Table I illustrates such a data base, containing 8 data points described by 3 attributes. For instance the speed of the first object belongs with degree 0.2 to the modality *low*, 0.3 to *normal* and 0.5 to *high*.

Definition 1: Given an attribute X defined on a universe U, a fuzzy subset A defined on U describing one of X modalities and $* \in \{<, >\}$, a *fuzzy gradual item* is defined as the triplet $(X, A, *)$.

A triplet $(X, A, >)$ is then to be understood as *the more* X *is* A, or more precisely *the higher the membership degree of* X *to* A; likewise, $(X, A, <)$ is to be understood as *the less* X *is* A, or *the lower the membership degree of* X *to* A.

TABLE I
EXAMPLE OF A FUZZY DATA SET

Id.	Speed			Wall distance		Braking		
	low	normal	high	close	far	light	average	hard
o_1	0.2	0.3	0.5	0.4	0.6	0.6	0.4	0.2
o_2	0.2	0.2	0.6	0.5	0.5	0.2	0.7	0.1
o_3	0	0.1	0.9	0.7	0.3	0	0.6	0.4
o_4	0	0.2	0.8	0.8	0.2	0.1	0.3	0.6
o_5	0.1	0.7	0.3	0.3	0.7	0.3	0.3	0.4
o_6	0.2	0.3	0.5	0.9	0.1	0.5	0.3	0.2
o_7	0	0.6	0.4	0.8	0.2	0.1	0.1	0.8
o_8	0.1	0.2	0. 7	0.9	0.1	0	0.3	0.7

For instance (wallDistance, close, $>$) represents the fuzzy gradual item *the closer the wall*.

Definition 2: A *fuzzy gradual itemset* is defined as a combination of several gradual items, semantically interpreted as their conjunction

For instance $M = \{(X_1, A_1, >), (X_2, A_2, <)\}$ is interpreted as *the more X_1 is A_1 and the less X_2 is A_2*.

Definition 3: A *fuzzy gradual rule*, denoted by $M_1 \to M_2$ is defined as a pair of fuzzy gradual itemsets M_1 and M_2, on which a causality relationship is imposed; M_1 is called the *antecedent*, M_2 the *consequent*.

It can for instance take the form (wallDistance, close, $>$) $\to$ (braking, hard, $>$) meaning *the closer the wall, the harder the brakes are applied*.

Different semantics can be associated to these rules. We recall them below, distinguishing between the interpretations as fuzzy implication and as co-variation constraints, detailing the various approaches implementing the latter.

In the following, we do not distinguish between fuzzy sets and their membership functions: for any fuzzy subset A and for any data point x in the data set $\mathcal{D}$, we denote $A(x)$ the membership degree of x to A. For an itemset containing p items, $M = \{(X_j, A_j, *_j), j = 1..p\}$, we consider the membership degree to the itemset as $M(x) = \top_{j=1..p}(A_j(x))$ where $\top$ denotes a a t-norm, *e.g.* the minimum.

B. Fuzzy implication interpretation

Fuzzy gradual rules were first defined as logical inference rules, in a fuzzy logic framework [1], [2] : a fuzzy gradual rule $M_1 \to M_2$ is considered to hold if the membership degrees to the fuzzy modalities involved in the rule satisfy a fuzzy implication, for each data point. Such a rule can be interpreted as a fuzzy generalization of association rules [3], stating that the (fuzzy) presence of M_1 implies, in a logical implication sense, the (fuzzy) presence of M_2. For instance, the rule *the closer the wall, the harder the brakes are applied*, can be seen as the fuzzy extension of an association rule relating the presence of binary attributes *close wall* and *hard braking*.

In this interpretation, the support of the rule is then computed as the sum of the contribution of each data point to the implication:

Definition 4 (fuzzy implication interpretation): The *support* of a fuzzy gradual rule $R = M_1 \to M_2$ is computed as

$$supp(M_1 \to M_2) = \sum_{x \in \mathcal{D}} i(M_1(x), M_2(x)) \qquad (1)$$

denoting i a residuated implication operator, *e.g.* the Goguen implication $i(a, b) = \min(1, b/a)$ if $a \neq 0$, 1 otherwise.

It can be noticed that this support definition does not apply to an itemset as defined in the classic association rules framework: it applies to the rule, in an asymmetrical way that distinguishes between $M_1 \to M_2$ and $M_2 \to M_1$, and does not only consider the itemset $M = M_1 \cup M_2$.

As an example, consider the data given in Table I and R the rule *the closer the wall, the harder the brakes are applied* that we will use throughout the paper. R support is then 0.5+0.2+0.57+0.75+1+0.22+1+0.77 = 5.01.

It can be noticed that other implication operators lead to different types of rules than fuzzy gradual rules: in particular, s-implications lead to so-called certainty rules. They indeed model certainty variations in the conclusion, leading to rules such as *the later the waking, the more certain the lateness* [2], [3].

C. Co-variation constraint interpretation

A second approach on fuzzy gradual rules interpret them as co-variation constraints applied to the whole data set, instead of considering each point individually: the rule is understood as expressing a global tendency across the data, as correlation on the variations of the membership degrees, outside a logical implication framework. In the case of two items $(X, A, >)$ and $(Y, B, >)$, the rule *the more X is A, the more Y is B*, written $(X, A, >) \to (Y, B, >)$, is considered to hold if an increase in the membership degrees to A comes along with an increase in the membership degrees to B. Three categories of methods can be distinguished in this context, depending on whether they rely on regression analysis, induced ranking correlations or order compliant data subsets, as detailed below.

1) Regression analysis: In order to identify fuzzy gradual rules interpreted as co-variation relationships, it was first proposed to perform a regression analysis between membership degrees [5], so as to highlight gradual dependencies.

More precisely, when considering a rule $M_1 \rightarrow M_2$, a regression method is first applied to the membership degree couples $(M_1(x), M_2(x))$, for all $x \in \mathcal{D}$, *i.e.* considering all data points simultaneously. The validity of the rule $M_1 \rightarrow M_2$ is then evaluated from the quality of the regression: for linear regression [5], it is thus measured by the normalised mean squared error R^2, together with the slope of the regression line. Thus the attribute pairs for which the membership degrees are insufficiently correlated are rejected, as well as pairs for which the membership to M_1 remains almost constant while that to M_2 varies (or reciprocally).

2) Induced rankings correlation: Whereas the regression analysis method relies on the numerical values of the membership degrees, other works consider the co-variation constraint only in terms of the induced order: they state that the ordering induced by the membership degrees to M_1 must be identical to that derived from M_2. In the case of itemsets of size 1, the gradual rule *the more X is A, the more Y is B* is considered to hold if $\forall x, x' \in \mathcal{D}$, $A(x) < A(x')$ implies $B(x) < B(x')$. In the case of dependencies such as *the more X is A, the less Y is B*, the constraint imposes that the induced orders must be reversed.

In [6], the support of an itemset is thus measured as the proportion of so-called concordant data couples, *i.e.* couples that satisfy the constraints expressed by all gradual itemsets involved in the rule:

Definition 5 (induce rankings correlation interpretation): The *support* of an itemset $M = \{(X_j, A_j, *_j),\ j = 1..p\}$ is computed as

$$supp(M) = \frac{|\{(x, x')/\forall j \in [1, p]\ A_j(x) *_j A_j(x')\}|}{|\mathcal{D}|(|\mathcal{D}| - 1)} \qquad (2)$$

The support of a rule $M_1 \rightarrow M_2$ is then defined as the support of the itemset $M = M_1 \cup M_2$, as in the classic association rule framework.

Considering the data in Table I and the rule *the closer the wall, the harder the brakes are applied*, the support is then 0.57: all couples (to simplify the notation, we only give the object indexes) $(1, 2), (1, 5), (1, 6), (2, 5), (3, 5), (3, 6), (4, 6), (4, 7), (5, 6),$ $(6, 7), (6, 8), (7, 8)$, together with, for each of them the exchanged couples $((2, 1), (5, 1), ...)$, are discordant[1]. Thus, there are 24 discordant couples and 32 concordant ones for a total number of 56, leading to the support value 32/56 = 0.57.

In [6] it is then proposed to formulate the extraction of such rules as the discovery of association rules in a set of transactions derived from the data set: transactions t are built for all data couples (x, x'), items are defined as $A*$, $* \in \{<, >\}$, for all modalities in $\mathcal{D}$. A transaction t then possesses an item $A*$ if the couple (x, x') it corresponds to satisfies the constraint imposed by $A*$, *i.e.* $A(x) * A(x')$. Fuzzy gradual rules in $\mathcal{D}$ then correspond to association rules in this modified data set. In order to reduce the computational complexity, efficient approximation schemes are proposed [6].

In [9] the same support definition is considered, and interpreted in terms of ranking correlation: it is computed as a generalised Kendall coefficient using an efficient binary representation of the data pairs, as proposed in [8].

3) Ranking-compliant data subsets: In [7], [8] a different interpretation of the co-variation constraint semantics of fuzzy gradual rules is considered: it proposes to identify subsets $\mathcal{D}^*$ of the initial data set $\mathcal{D}$ that satisfy the ordering constraint expressed by a given itemset, *i.e.* subsets of data that can be ordered so that all couples from $\mathcal{D}^*$ satisfy the constraints expressed by the gradual itemsets.

Definition 6 (ranking-compliant data subset interpretation): The *support* of an itemset $M = \{(X_j, A_j, *_j),\ j = 1..p\}$ is computed as

$$supp(M) = \frac{1}{|\mathcal{D}|} \max_{\mathcal{D}^* \in \mathcal{L}} |\mathcal{D}^*| \qquad (3)$$

where $\mathcal{L}$ denotes the set of all maximal data subsets $\mathcal{D}^* = \{x_1, ... x_m\} \subset \mathcal{D}$ for which there exists a permutation π such that $\forall j \in [1, p], \forall k \in [1, m-1],\ A_j(x_{\pi_k}) *_j A_j(x_{\pi_{k+1}})$. The maximality constraint imposes that for any $\mathcal{D}^*$, no object can be added to $\mathcal{D}^*$ without loosing the ranking compliance property.

Considering the data in Table I and R the rule *the closer the wall, the harder the brakes are applied*, $\mathcal{L} = \{\{1, 3, 4, 7\}, \{2, 3, 4, 7\}, \{1, 3, 4, 8\}, \{2, 3, 4, 8\}\}$: all subsets contain 4 elements, and R support equals 4/8 = 0.5.

It can be underlined that this definition of support is independent of the amplitude of the constraint violation: if an object does not satisfy the itemset, *i.e.* if it does not fit with the considered order, it is simply removed from $\mathcal{D}^*$. On the contrary its penalisation is variable both for the regression and the ranking correlation approaches: in the regression approach, a point with a high deviation amplitude is likely to distort the regression coefficient to a large extent; in the ranking correlation approach, such a point is likely to give rise to a high number of discordant couples (x, x').

In [7] a heuristic is proposed to compute this support, in a level-wise process that considers itemsets of increasing sizes. It consists in discarding, at each level, the objects whose so-called conflict set is maximal, *i.e.* the data that prevent the maximal number of other objects to be sorted.

In [8], an exact and very efficient method is proposed, based on precedence graphs: the data are represented through a graph whose nodes are defined as the objects and the edges express the precedence relationships. The graph is represented by its adjacency matrix in a bitmap form: for an itemset $M = \{(X_j, A_j, *_j), j = 1..p\}$, the coefficient corresponding to (x, x') is 1 if $\forall j \in [1, p]\ A_j(x) *_j A_j(x')$, 0 otherwise. The support of the considered itemset can then be obtained as the length of the maximal path in the graph. The support defined in Equation (2) can be obtained as the sum of the elements in this matrix, as proposed in [9].

[1]When only two items are considered, it is sufficient to consider the object pairs, because both couples (i, j) and (j, i) have the same status, *i.e.* if both are concordant or both are discordant. For longer itemsets, they must be handled separately: it can *e.g.* be that both (i, j) and (j, i) are discordant, because of different gradual item constraints.

The relevance of this approach comes from its very high efficiency to generate gradual itemsets of size $p+1$ from itemsets of size p: indeed it holds that if M is an itemset generated using M' and M'', its adjacency matrix $Adj_M = Adj_{M'} \& Adj_{M''}$ where & is the bitwise AND operation.

III. Strengthening effect through "all the more" clauses

The semantics of fuzzy gradual rules have been widely studied, and several interpretations have been proposed, as recalled in the previous section. In this section, we consider strengthened fuzzy gradual rules, *i.e.* fuzzy gradual rules enriched by an additional reinforcement clause introduced by the linguistic expression "all the more". Such strengthened fuzzy gradual rules can be illustrated by the example *the closer the wall, the harder the brakes are applied, all the more the higher the speed* and formalized as

Definition 7: A *strengthened gradual rule* is a triplet of fuzzy gradual itemsets (M_1, M_2, M_3) and is denoted by $M_1 \to M_2; M_3$. M_1 is the antecedent, M_2 the conclusion, and M_3 the strengthening clause, linguistically introduced by the expression *all the more*.

In the following we discuss the semantics that can be attached to this strengthening effect, comparing various interpretations, in particular with respect to the different semantics of fuzzy gradual rules. We focus on the interpretation as co-variation constraints across the whole data, recalled in Section II-C, and in particular on the ranking compliant data subset approach (see Section II-C.3 and Definition 6) that identifies data subsets on which the ranking constraint holds. In Section III-D we discuss the strengthening effect for the fuzzy implication interpretation of fuzzy gradual rules.

A. Difference with conjunctive gradual rules

It is first important to underline the difference between strengthened rules and conjunctive rules of the form $(M_1 \wedge M_3) \to M_2$. With the considered illustrative example, this rule would be *the closer the wall and the higher the speed, the harder the brakes are applied*. Both rules can hold for a given data set, still, their semantics differ.

Indeed, for such conjunctive gradual rules, M_3 plays a causal role on M_2 and in the co-variation interpretation of fuzzy gradual rule imposes a strong constraint: it requires that the rankings induced by the three itemsets M_1, M_2 and M_3 are identical or highly consistent.

In the strengthening case, the ranking constraint only concerns the two itemsets M_1 and M_2, which restricts to a lesser extent the number of objects on which the rule applies: the reinforcement effect rather applies on the rule $M_1 \to M_2$ than on M_2, even if it does not exclude a causal effect of M_3 on M_2.

B. Constraint on variation strength

One interpretation of the strengthening effect considers it as a constraint on the variation intensity of M_2, when the rankings according to M_1 and M_2 agree: the relation "all the more" is understood as a intensification of the variations of the attributes involved in M_2 according to M_3 values.

Considering the illustrative example *the closer the wall, the harder the brakes are applied, all the more the higher the speed*, this interpretation means that first an increase in the wall closeness implies an increase in the braking strength, and second that the increase is correlated to high speed.

Formally, this interpretation of strengthening can be expressed as the conjunction of fuzzy gradual rules: a rule $M_1 \to M_2$, and a rule of the form $M_3 \to \Delta M_2$ where ΔM_2 denotes M_2 variations.

C. Reinforced presence

Another interpretation evaluates the strengthening influence of M_3 on the rule $M_1 \to M_2$ as the fact that the rule is better satisfied when the candidate objects are the ones possessing M_3 rather than when the whole data set is considered: it should be easier to apply the rule when one focuses on data possessing M_3. Equivalently, the strengthening influence of M_3 on the rule $M_1 \to M_2$ is interpreted through the extent to which M_3 is possessed by objects satisfying the rule.

The reinforcement is thus used to partition the data (considering a fuzzy partition) and to compare the rule validity on a data subset and the whole data set: this interpretation combines the fuzzy gradual rule with a (fuzzy) presence condition.

For this interpretation, the fuzzy gradual rule approach through ranking compliant data subsets is in particular relevant: it is indeed easy to measure a reinforced presence of M_3 when the rule $M_1 \to M_2$ applies, as the method explicitly extracts data subsets for which the rule holds. Therefore we will use this method in the following, in particular to define the quality criteria.

Still, the other ranking-based approaches of fuzzy gradual rules also make it possible to take into account a weighted influence of M_3 presence: in the regression case, it is possible to modify the R coefficient, *e.g.* to give less weights to regression errors that coincide with low presence of M_3. In the ranking correlation approach, the penalisation associated to a discordant pair can also depend on the membership degrees to M_3.

D. Case of fuzzy implication interpretation

In the framework of fuzzy gradual rules understood as fuzzy implications satisfied individually for each data point, a strengthened rule $M_1 \to M_2; M_3$. M_1 can be interpreted following different lines. Indeed, in a such framework, it makes sense to consider an additional attribute in the data base, whose values are the truth values of the implication $M_1 \to M_2$, *i.e.* $i(M_1(x), M_2(x))$ for all $x \in \mathcal{D}$. The strengthened rule can then be understood as gradual relation between the membership degree to M_3 and the new additional attribute, meaning that the reinforcement holds if the implication is all the truer as M_3 is true.

The strengthened rule is then identical to the cascaded fuzzy gradual rule $M_3 \to (M_1 \to M_2)$ in the fuzzy implication interpretation of gradual rules. The support of such a rule can then be computed by transposing Equation (1) as $\sum_{x\in\mathcal{D}} i(M_3(x), i(M_1(x), M_2(x))$.

It can be underlined that for s-implications, the cascaded rule is equivalent to $(M_1 \wedge M_3) \rightarrow M_2$ as the previous quantity equals $i(\top(M_3(x), M_1(x)), M_2(x))$. Now the Łukasiewicz implication, $i(a, b) = \min(1 - a + b, 1)$, is both an r- and an s-implication, combining the gradual rule interpretation with the previous property. It thus bridges the gap between reinforced and conjunctive rules.

Besides, this interpretation sheds new light on the interpretation presented in the previous subsection III-C, although the latter does not belong to the logical framework: it also considers a fuzzy presence semantics. Still, the presence of M_3 is not taken into account individually for each data point, but it is evaluated globally on the ranking compliant data subset.

IV. Strengthening quality criteria

In this section, we describe the quality measures, and in particular support and confidence, proposed to assess strengthened fuzzy gradual rules according to the semantics described in Section III-C. As mentioned previously, the ranking compliant data subset recalled in Section II-C.3 appears to be particularly relevant for this interpretation. Therefore we exploit its properties in the proposed quality criteria, especially the identification of subsets of data on which the ranking constraint holds.

A strengthened rule $M_1 \rightarrow M_2; M_3$ contains two components, namely the fuzzy gradual rule $M_1 \rightarrow M_2$, and its reinforcement, thus it must be assessed according to these two elements. Therefore, we propose to characterize a reinforcement rule both by the support and confidence of the rule $M_1 \rightarrow M_2$, and by the strengthened support and confidence in reinforcement, that measure the quality of the latter. In the following, we propose definitions for these criteria.

A. Strengthened support

As in the classic case, we propose to define the support of a strengthened rule by an evaluation of its frequency. To that aim, we use a sigma-count approach, to define the degree to which M_3 is present among the objects that satisfy the rule $M_1 \rightarrow M_2$: the higher it is, the better the reinforcement holds.

1) Definition: More formally, we propose the following definition

Definition 8: The *strengthened support* of a fuzzy gradual rule $R = M_1 \rightarrow M_2; M_3$ is defined as

$$s\_supp(R) = \max_{i=1..k} \sum_{x\in\mathcal{D}_i^*} M_3(x) \tag{4}$$

denoting $\mathcal{L} = \{\mathcal{D}_i^*, i = 1..p\}$ the set of all maximal data subsets that can be permuted to satisfy the ranking constraints on M_1 and M_2.

Introducing the notation $M_3(\mathcal{D}^*) = \sum_{x\in\mathcal{D}^*} M_3(x)$, *i.e.* the fuzzy cardinal of $\mathcal{D}^*$ according to M_3, the support equals $s\_supp(R) = \max_{i=1..p} M_3(\mathcal{D}_i^*)$.

2) Example: Considering the data given in Table I and the R rule *the closer the wall, the harder the brakes are applied, all the more the higher the speed*, one has $\mathcal{L} = \{\{1, 3, 4, 7\}, \{2, 3, 4, 7\}, \{1, 3, 4, 8\}, \{2, 3, 4, 8\}\}$, and $M_3(\mathcal{D}_1^*) = M_3(\{1, 3, 4, 7\}) = 0.5 + 0.9 + 0.8 + 0.4 = 2.6$, $M_3(\mathcal{D}_2^*) = M_3(\{2, 3, 4, 7\}) = 0.6 + 0.9 + 0.8 + 0.4 = 2.7$, $M_3(\mathcal{D}_3^*) = M_3(\{1, 3, 4, 8\}) = 0.5 + 0.9 + 0.8 + 0.7 = 2.9$, and $M_3(\mathcal{D}_4^*) = M_3(\{2, 3, 4, 8\}) = 0.6+0.9+0.8+0.7 = 3$. Thus the strengthened support is 3.

3) Properties: It must be underlined that the proposed definition is not symmetric, due to the specific role of M_3. Besides, as discussed in previous section, this definition does not take into account a co-variation of the membership degrees to the strengthening clause M_3 with those of the antecedent and consequent of the rule: it uses a sigma-count on M_3 presence. This approach is in particular relevant to handle reinforcement by binary attributes, as for instance *the greater the meat purchase, the averager the vegetable purchase, all the more if the city is Paris*. Such rules can easily be dealt with in this context: the membership degrees equal 1 for all objects possessing the attribute.

One can replace in the previous definition the sigma-count by a thresholded sigma-count, using a user-defined threshold: this makes it possible to avoid taking into account objects that are not representative enough of M_3.

Lastly, as the support defined for classic association rules, the following property holds:

Property 1: The strengthened support is anti-monotone with respect to the size of the M_3 itemset.

Proof: if one considers two rules $R_1 = M_1 \rightarrow M_2; M_3$ and $R_2 = M_1 \rightarrow M_2; M_4$ with $M_3 \subset M_4$, it holds that $s\_supp(R_1) \geq s\_supp(R_2)$. Indeed, the two rules R_1 and R_2 have the same basic form, and thus, the same set $\mathcal{L}$ of maximal ranking compliant data subsets. Moreover, for any $x \in \mathcal{D}$, $M_3(x) \geq M_4(x)$. Indeed itemsets are interpreted as conjunction of the items they contain. Thus, noting M the itemset such that $M = M_4 \backslash M_3$, $M_4 = M_3 \cup M$, and $\forall x$ $M_4(x) = \top(M_3(x), M(x)) \leq M_3(x)$. Thus it holds that $\forall \mathcal{D}_i^* \in \mathcal{D}$, $M_4(\mathcal{D}_i^*) \leq M_3(\mathcal{D}_i^*)$, which leads to the desired result.

This property will in particular be used for the mining algorithm described in Section V

B. Strengthened confidence

The confidence in the strengthened fuzzy gradual rule then measures the relevance of the strengthening effect as compared to the non strengthened rule $M_1 \rightarrow M_2$.

1) Definition: Following the classic scheme of confidence, defined as the quotient between the rule support by the antecedent support, a first definition could consist in comparing the strengthened support, as defined in Equation (4), to the support of the non strengthened rule: this would lead to define $s\_conf(R) = s\_supp(M_1 \rightarrow M_2; M_3)/supp(M_1 \rightarrow M_2) = \max_{i=1..p} M_3(\mathcal{D}_i^*)/\max_{i=1..p} |\mathcal{D}_i^*|$.

Now this definition is not satisfactory, because it compares values that may be obtained from different data subsets: the numerator is obtained from the ranking compliant subset with maximal M_3 fuzzy cardinality, whereas the denominator comes from the ranking compliant subset with maximal cardinality. Now these subsets need not be the same, and the quotient of two values corresponding to different objects does not have an intuitive meaning.

Therefore, we propose to compute the quotient between M_3 fuzzy cardinality and cardinality for each candidate data subset individually, and to then take the maximal value:

Definition 9: The *strengthened confidence* of a fuzzy gradual rule $R = M_1 \rightarrow M_2; M_3$ is defined as

$$s\_conf(R) = \max_{i=1..p} \frac{M_3(\mathcal{D}_i^*)}{|\mathcal{D}_i^*|} \quad (5)$$

denoting $\mathcal{L} = \{\mathcal{D}_i^*, i = 1..p\}$ the set of all maximal data subsets that can be permuted to satisfy the ranking constraints on M_1 and M_2.

This strengthened confidence implements the semantics proposed in Section III-C evaluating the average membership degree to M_3 among data satisfying the ranking constraint. Several data subsets satisfying the condition, we consider the maximal obtained value. This choice implies that we consider that the confidence can be established from any data subset on which the strengthened rule can rely and that it is sufficient that there exists one such subset.

In the previous example that considers the rule *the closer the wall, the harder the brakes are applied, all the more the higher the speed*, all maximal subsets $\mathcal{D}_i^*$ have the same cardinality, 4, the strengthened confidence equals 3/4 = 0.75

2) Properties: It can be noted that, as the strengthened support, and contrary to the classic confidence measure, the following property holds:

Property 2: The strengthened confidence is anti-monotone with respect to the size of the M_3 itemset.

Proof: denoting as previously $R_1 = M_1 \rightarrow M_2; M_3$ et $R_2 = M_1 \rightarrow M_2; M_4$ with $M_3 \subset M_4$, it holds that $M_4(\mathcal{D}_i^*) \leq M_3(\mathcal{D}_i^*)$ as previously. Dividing both sides by $|\mathcal{D}_i^*|$, and taking the maximum over i, this leads to the anti-monotony property.

This allows to directly extract strengthened with high confidence that satisfy a user defined threshold, instead of having to filter out the rules with low confidence after they have been extracted: it makes it possible to efficiently mine strengthened rules of interest.

C. Other quality criteria

Other classic quality criteria for association rule evaluation (see *e.g.* [10]) can also be extended, and in particular the lift measure [11]: the strengthened support and confidence measures are sensitive to the frequency of the M_3 itemset (in the same manner as classic support and confidence are sensitive to the frequency of the conclusion itemset), whereas the lift criterion is not. In particular, the lift criterion makes it possible to reject candidate association rules that conclude to modalities present in all data. Likewise, the aim of the strengthened lift is to reject reinforcement clauses based on modalities such that $\forall x$, $M_3(x) = 1$.

Definition 10: The *strengthened lift* of a fuzzy gradual rule $R = M_1 \rightarrow M_2; M_3$ is defined as

$$s\_lift = \max_{i=1..p} \frac{M_3(\mathcal{D}_i^*)}{|\mathcal{D}_i^*|} \times \frac{|\mathcal{D}|}{M_3(\mathcal{D})} \quad (6)$$

denoting $\mathcal{L} = \{\mathcal{D}_i^*, i = 1..p\}$ the set of all maximal data subsets that can be permuted to satisfy the ranking constraints on M_1 and M_2.

It compares the average membership degree to M_3 among data satisfying the ranking constraint to the global average membership degree to M_3 across the whole data set.

V. Algorithm to mine strengthened fuzzy rules

The problem is then to extract all strengthened fuzzy gradual rules whose support, confidence, strengthened support and strengthened confidence are larger than user-defined thresholds. The rules looked for are maximal rules, *i.e.* those that contain the highest possible number of fuzzy gradual items while satisfying the thresholds. This maximality constraint applies to all three itemsets constituting antecedent, consequent and strengthening clause.

A strengthened fuzzy gradual rule $M_1 \rightarrow M_2; M_3$ contains two components, the fuzzy gradual rule $M_1 \rightarrow M_2$ and the strengthened clause M_3 that must satisfy the support and confidence constraints independently. Thus the mining algorithm is made of two steps: the first one consists in identifying the fuzzy gradual rules that satisfy the support and confidence constraints, the second aims at establishing the strengthened itemsets that can reinforce these rules.

The first step can be performed using the GRITE algorithm [8] that extracts all frequent fuzzy gradual itemsets with their support. Moreover, it identifies data subsets that can be sorted so as to satisfy the ranking constraints expressed by these itemsets.

It can be noted that the definition of strengthened support and confidence (Equations (4) and (5)) actually depend on the frequent itemsets $M_1 \cup M_2$ and not on the rules $M_1 \rightarrow M_2$. Thus they do not require the identification of the causality relations between the items constituting the extracted itemsets.

The second step that aims at establishing the strengthening itemsets for the previous itemsets follows the principles of the APRIORI algorithm, using the strengthened support and confidence as quality measure: strengthened itemsets are extracted in a level-wise approach, progressively building the itemsets of increasing size, exploiting the anti-monotony properties of these criteria.

To that aim, given a fuzzy gradual rule R to strengthen (or a frequent fuzzy gradual itemset), the candidate strengthening itemsets of size $k+1$ are generated from the candidates of size k, using the classic APRIORIgen process. They are then assessed using strengthened support and confidence, defined in Equations (4) and (5). The candidates for which one of these values is below the user-defined thresholds are discarded. These two steps are iterated until no more candidate can be generated.

During the initialisation phase, the candidate strengthening itemsets containing single items are defined as modalities of attributes that do not appear in the itemsets to strengthen. Indeed, rules involving several modalities of a given attributes with different status (antecedent, conclusion, reinforcement) could only contain trivial information, regarding mutual exclusion of modalities for instance, and would not be of interest.

VI. Conclusion

In this paper we proposed a new approach to strengthen fuzzy gradual rules, reinforced using "all the more" clauses, to extract more information summarizing data sets. We studied the semantics of such rules, proposed criteria to measure the relevance of such rules and described a candidate level-wise extraction algorithm.

Beside the application of this algorithm to several data sets in order to empirically evaluate the identified knowledge, future works aim at extending the strengthening semantics to other types of fuzzy graduality, in particular expressed by certainty rules. More generally, extensions to other types of rules, such as association rules or non fuzzy gradual rules are to be examined. They would provide a way to extract contextual information, *i.e.* rules that do not aply to the whole data but only to data subsets.

References

[1] B. Bouchon-Meunier and S. Desprès, "Acquisition numérique / symbolique de connaissances graduelles," in *3èmes Journées Nationales du PRC Intelligence Artificielle*. Hermès, 1990, pp. 127–138.

[2] D. Dubois and H. Prade, "Gradual inference rules in approximate reasoning," *Information Sciences*, vol. 61, no. 1-2, pp. 103–122, 1992.

[3] E. Hüllermeier, "Implication-based fuzzy association rules," in *Proc. of PKDD'01*, 2001, pp. 241–252.

[4] S. Galichet, D. Dubois, and H. Prade, "Imprecise specification of ill-known functions using gradual rules," *Int. Journal of Approximate Reasoning*, vol. 35, no. 3, pp. 205–222, 2004.

[5] E. Hüllermeier, "Association rules for expressing gradual dependencies," in *Proc. of PKDD'02*, 2002, pp. 200–211.

[6] F. Berzal, J.-C. Cubero, D. Sanchez, M.-A. Vila, and J. M. Serrano, "An alternative approach to discover gradual dependencies," *Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15, no. 5, pp. 559–570, 2007.

[7] L. Di Jorio, A. Laurent, and M. Teisseire, "Fast extraction of gradual association rules: A heuristic based method," in *Proc. of CSTST'08*, 2008.

[8] ——, "Mining frequent gradual itemsets from large databases," in *Proc. of IDA'09*, 2009.

[9] A. Laurent, M.-J. Lesot, and M. Rifqi, "Graank: Exploiting rank correlations for extracting gradual dependencies," in *Proc. of FQAS'09*, 2009.

[10] P. Lenca, B. Vaillant, P. Meyer, and S. Lallich, "Association rule interestingness measures: Experimental and theoretical studies," in *Quality Measures in Data Mining*, F. Guillet and H. J. Hamilton, Eds. Springer, 2007, vol. 43, pp. 51–76.

[11] S. Brin, R. Motwani, and C. Silverstein, "Beyond market baskets: Generalizing association rules to correlations," in *Proc. of the ACM SIGMOD/PODS '97 Joint Conference*, 1997, pp. 265–276.