Fuzzy Orderings for Fuzzy Gradual Dependencies: Efficient Storage of Concordance Degrees

Perfecto Malaquias Quintero Flores, Federico del Razo Lopez, Anne Laurent, Pascal Poncelet, Nicolas Sicard

Fuzzy Orderings for Fuzzy Gradual Dependencies: Optimization I

Malaquias Quintero
LIRMM - University Montpellier 2
CNRS UMR 5506
Montpellier F-34392, France
School of Computer and Systems
Apizaco Institute of Technology
Apizaco, Tlaxcala, Mex.
Email: quinterofl@lirmm.fr

Anne Laurent and Pascal Poncelet
LIRMM - University Montpellier 2
CNRS UMR 5506
Montpellier F-34392, France
Email: laurent@lirmm.fr
Pascal.Poncelet@lirmm.fr

Nicolas Sicard
LRIE de l'EFREI
Villejuif, Paris 94 800, France
Email: nicolas.sicard@efrei.fr

***Abstract*—In mining *gradual patterns* the idea is to express co-variations of attributes, taking the *direction of change* of attribute values into account. These patterns are such as {*the more A, the more B*}, {*the more A, the more B, the less C*} or { *the higher the speed, the higher the danger* }. These patterns are denoted as {*A*≥*B*≥ }, { *A*≥*B*≥*C*≤} or { *speed*≥*danger*≥} respectively. Such patterns hold if the variation constraints simultaneously hold on the attributes. However, it is often hardly possible to compare attribute values, either because the values are taken from noisy data, or because it is difficult to consider that a small difference between two values is meaningful. In this context, we focus on the use of fuzzy orderings to take this into account.**

***Index Terms*—Mining *gradual patterns*, fuzzy orderings, fuzzy gradual patterns.**

I. Introduction

Given a database *D* an *association rule* is defined as a rule of the form *If A Then B* expressing the dependency between the so-called itemsets (binary attributes) *A, B* from the schema of *D*. The intended meaning of such a rule is that, if *A* is present in a transaction, then *B* is likely to be present too. An association rule is of the form:

$$R : I_{sa} \Rightarrow I_{sc}$$

where I_{sa} and I_{sc} are two itemsets. Two measures are usually defined to assess such rules: **The frequency/support** is the frequency of the union of the condition I_{sa} and consequence I_{sc} ie.

$$Freq(R) = Freq(I_{sa} \cup I_{sc})$$

The confidence measures the probability of knowing or occurrence of I_{sc} given I_{sa}, ie.

$$Conf(R) = \frac{Freq(I_{sa} \cup I_{sc})}{Freq(I_{sa})}$$

In the fuzzy case, the presence of an item in a transaction is a matter of degree. Another type of rule, called *gradual dependency*, conveys information in the form of attribute covariations, such as *the higher the age, the higher the salary*, meaning that the age of the persons increases together with its salary. Gradual dependencies consider tendencies across the whole data set, in terms of correlation of the attribute variations. This idea is closely connected to the so called *gradual rules* in fuzzy logic [9].

The automatic extraction of gradual dependencies or gradual association rules is one of the topics addressed in the field of data mining, for the modelling of frequent co-variations over a set of objects described by numerical attributes of data sets, such as biological databases, survey databases, data streams or sensor readings. In mining *gradual dependency* the idea is to express dependencies between the *direction of change* of attribute values.

As for the association rule extraction, the process consists of two steps: first frequent gradual patterns (also known as itemsets) are extracted. Then causality relations between the items are extracted. In mining frequent *gradual itemsets*, the goal is to discover frequent co-variations between attributes[10] [11].

When considering such gradual patterns and gradual rules, it is thus important to be able to count to which extent attributes co-variate. In this context, varied measures have been defined in the literature. However, few works have focused on how to exploit fuzzy orderings for handling noisy data.

For instance, when considering biological data from RNA/DNA chips, it would be semantically false to consider that two close values can be easily ordered. In this paper, we thus focus on an approach that evaluates frequent gradual patterns in terms of the robust rank correlation measure on the basis of fuzzy orderings.

The paper is organized as follows: in Section II, we introduce the preliminary definitions and related work. The Section III is devoted to a review of fuzzy ordering-based rank correlation coefficient. In Section IV, we present our approach. Finally we present in Section V our conclusions and future research.

II. Preliminary Definitions and Related Work

In this section, after recalling the definitions of gradual item, gradual itemset, gradual dependencies, rank correlation, fuzzy rank correlation as given in [9], [10], [11], we present the related works on gradual pattern mining, rank correlation for extracting gradual itemsets, mining gradual dependencies based on fuzzy rank correlation, fuzzy ordering-based rank

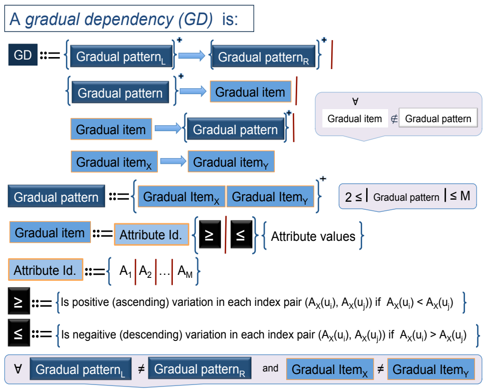


Fig. 1. Preliminary definitions: Gradual dependency, gradual pattern, and gradual item.

correlation coefficient, and on parallel frequent gradual pattern mining.

A. Preliminary Definitions

Gradual dependencies extraction applies to a data set D defined as a set of tuples T over a schema S of I attributes with m numerical values.

A *gradual item* is defined as a pair (I,θ) where I is an attribute in D and θ a comparison operator in $\{ \geq, \leq \}$. They represent the fact that the attribute values increase (in case of $\geq$) or decrease (in case of $\leq$).

A *gradual pattern/gradual itemset* is defined as a combination of several gradual items, semantically interpreted as their conjunction $g = \{ (I_1,\theta_1), (I_2,\theta_2), \ldots, (I_k,\theta_k) \}$ of cardinality greater than or equal to 2, see Fig. fig:GradualDep. For example, (Age,$\geq$) is a gradual item, while $\{(Age, \geq), (Salary, \leq), (Loans, \geq)\}$ is a gradual itemset, with a cardinality equal to 3.

The *support* of a gradual itemset in a data set D can be defined in varied manners [6], [11]. For instance, it can be defined as the number of tuples that can be ordered to support all item comparisons [11].

B. Related Works

Two kinds of dependencies can be distinguished: a first category considers linguistic variables represented by fuzzy sets and imposes covariation of the membership degrees across all data, for example, *the more the age is middle-aged, the less the number of cars is low*, where *middle-aged* and *low* refer to modalities of the linguistic variables *age* and *number of cars* respectively. A second, category directly considers the numerical values of the attributes and applies to attribute covariation on the whole attribute universe [11].

There are different interpretations of gradual dependency, as following: *(1) based in regression, (2) based in correlation, (3) approach based on conflict sets, and (4) approach based on the precedence grap*. Consult [11] for more information.

Laurent, Lesot, and Rifqi in [11] present an approach called GRAANK that combines the interpretation of gradual dependency of rank correlation measures and an algorithm on the precedence graph, named GRITE represented by its adjacency matrix, in a bitmap. The proposed algorithm thus follows the principle of the APRIORI algorithm, modifying the step of candidate evaluation, where for all candidate itemsets, compute their support as the sum of their binary matrices divided by $n(n$ - 1)/2 where n is the number of objects.

Koh and Hullermeier in [9] present a framework for mining gradual dependencies based on the use of fuzzy rank correlation for measuring the sthrenght of a dependency. The approach is a unification of previous approaches to evaluate gradual dependencies and captures both qualitative and quantitative measures of association as special cases. A gradual dependency $A \rightarrow B$ is evaluated in terms of two measures, namely the number of concordant pairs, *CT*, and the rank correlation *Fuzzy_*γ as defined in (2). Comparing this approach with the classical setting of association analysis, *CT* plays the role of the support of a rule, while *Fuzzy_*γ corresponds to the confidence. These measures can also be interpreted within the formal framework proposed by Dubois and Hullermeier in [7], in which every observation (in the case of a pair of points (*A(u), B(u)*) and (*A(v), B(v)*) is considered, to a certain degree, as an *example* of a pattern, as a *counterexample*, or as being *irrelevant* for the evaluation of the pattern. In the framework and the algorithm of Koh and Hullermeier, these degrees are given, respectively, by the degree of concordance, the degree of discordance, and the degree to which the pair is a tie. Formally they define the support and confidence of a gradual dependency $A \rightarrow B$ as follows:

$$supp(A \rightarrow B) = CT \tag{1}$$

$$conf(A \rightarrow B) = Fuzzy\_\gamma = \frac{CT - DT}{CT + DT} \tag{2}$$

where

$$CT = \sum_{u_i} \sum_{u_j} C(u_i, u_j) \tag{3}$$

$$CT = \sum_{u_i} \sum_{u_j} T(L(A(u_i), A(u_j)), L(B(u_i), B(u_j))) \tag{4}$$

$$DT = \sum_{u_i} \sum_{u_j} D(u_i, u_j) \tag{5}$$

$$DT = \sum_{u_i} \sum_{u_j} T(L(A(u_i), A(u_j)), L(B(u_j), B(u_i))) \tag{6}$$

Laurent et al. in [10] present an efficient parallel mining of gradual patterns and gradual rules on multicore procesor based on the algorithm named GRITE (Gradual Itemset Extraction) and a model of parallelization multithreading type master-workers, where only parallelized the evaluation phase of frequent itemsets. In that framework, Laurent et al. consider the support of a gradual itemset *P* in a database *DB* as the ratio of the cardinality of *P* in *DB* denoted by $\lambda(P, DB)$ over the cardinality of *DB* denoted by $|DB|$. That is, supp(P,DB)=$\frac{\lambda(P,DB)}{|DB|}$.

Do et al. in [6] present PGLCM (Efficient Parallel Mining of Closed Frequent Gradual Itemsets) based on the parallelization of the GLCM algoritm based on the LCM algoritm (*Linear* *time Closed itemset Miner*) using the Melinda library. In this framework, Do et al. consider a gradual itemset *P*= $\{(i_{k_1},v_{k_1})$, . . . , $(i_{k_j},v_{k_j})\}$ where $\{k_1,...,k_j\} \subseteq \{1,...,n\}$ and the $k_1,...,k_j$ are all distinct. Two tuples *t* and *t'* can be ordered with respect to *P* if all the values of the corresponding i_k items from the gradual itemset can be ordered with respect to variation $v \in \{\uparrow, \downarrow\}$ where ↑ stands for a positive (ascending) variation, ↓ for a negative (descending) variation and the formal definition of the suport of *P* is $support(P)$=$\frac{max_{L\in l}(|L|)}{|R|}$, i.e. it is the size of the longest list of tuples that respects a gradual itemset *P*, where *L* = $\{t_1,...,t_m\}$ be a list of tuples from a set of tuples *R* defined over the schema *S* = $\{I_1,...,I_n\}$ of a dataset.

III. An Overview of Robust Rank Correlation Coefficients on the Basis of Fuzzy Orderings

A. Rank Correlation Measures: An Overview

Correlation measures are among the most basic tools in statistical data analysis and machine learning. They are applied to pairs of observations (n≥2) of two variables *X* and *Y*

$$(x_i, y_i)_{i=1}^n \tag{7}$$

$$\mathbf{x} = (x_1, x_2, ..., x_n) \tag{8}$$

$$\mathbf{y} = (y_1, y_2, ..., y_n) \tag{9}$$

of two linearly ordered domains **X** and **Y** to measure to which extent the two observations comply with a certain model. The most prominent representative is surely *Pearson's product moment coefficient*, often called *correlation coefficient* for short. *Pearson's product moment coefficient* is applicable to numerical data and assumes a linear relationship as the underlying model; therefore, it can be used to detect linear relationships, but no non−linear ones [4].

Rank correlation measures are intended to measure to which extent a monotonic function is able to model the inherent relationship between the two observables. They neither assume a specific parametric model nor specific distributions of the observables. They can be applied to ordinal data and, if some ordering relation is given, to numerical data too [4]. Therefore, rank correlation measures are ideally suited for detecting monotonic relationships, in particular, if more specific information about the data is not available [5], [9]. The two most common approaches are *Spearman's rank correlation coefficient* (short *Spearman's rho*) and *Kendall's tau (rank correlation coefficient)*.

The goal of a rank correlation measure is to measure the dependence between the two variables in terms of their tendency to increase and decrease in the same or the opposite direction. If an increease in *X* tends to come along with an increase in *Y*, then the (rank) correlation is positive. The other way around, the correlation is negative if an increase in *X* tends to come along with a decrease in *Y*. If there is no dependency of either kind, the correlation is (close to) 0. Several rank correlation measures are defined in terms of the number *C* of *concordant*, the number *D* of *discordant*, and the number *N* of *tied* data points [9]. For a give index pair *(i,j)* ϵ { 1, . . . , *n* }$\hat{2}$, we say that *(i,j)* is concordant, discordant or tied depending on whether $(x_i,\ x_j)(y_i,\ y_j)$ is positive, negative or 0, respectively. A well-known example is Goodman and Kruskal's *gamma rank correlation*, which is defined as:

$$\gamma = \frac{C-D}{C+D} \tag{10}$$

B. Fuzzy Orderings

Fuzzy relation, fuzzy equivalence relation, and fuzzy ordering are concepts that have been introduced with the aim; to model human-like decisions by taking the graduality of human thinking and reasoning into account. Fuzzy orderings have broad utility. They can be applied, for example, when expressing our preferences with a set of alternatives. Compared to crisp orderings, they have greater expressive power. They allow us to express not only that we prefer an alternative to another one, but also the strength of this preference [8]. The study of *similarity*, *fuzzy relation*, *fuzzy ordering*, *similarity relation*, and the notion of *equivalence* was started by Zadeh [12] in 1971, in that paper he defined the notion of *similarity* as a generalization of the notion of equivalence, and a *fuzzy ordering* as a generalization of the concept of ordering.

A fuzzy relation $S : X^2 \rightarrow$[0,1] is called *similarity relation* on a domain *X* with respect to a t-norm *T*, for brevity *T-similarity*, if and only if the following three axioms hold for all *x, y, z* ∈ *X*:
(i) *S*-reflexivity: μ_S*(x,x)*=1,
(ii) *S*-symmetry: μ_S*(x,y)*=μ_S*(y,x)*, and
(iii) *T*-transitivity: $\mu_T(\mu_S(x,y),\ \mu_S(y,z)) \leq\ \mu_S(x,z)$.

Where $\mu_S(x,y)$, $\mu_S(y,z)$ and $\mu_S(x,z)$ are the grade of membership of the ordered pairs *(x,y)*, *(y,z)*, and *(x,z)* in *S*, with respect to a triangular norm (*t-norm*) *T*.

A fuzzy relation $E : X^2 \rightarrow$[0,1] is called *fuzzy equivalence relation* on a domain *X* with respect to a t-norm *T*, for brevity *T-equivalence*, if and only if the following three axioms are fulfilled for all *x, y, z* ∈ *X*:
(i) *E*-reflexivity: μ_E*(x,x)*=1,
(ii) *E*-symmetry: μ_E*(x,y)*=μ_E*(y,x)*, and
(iii) *T*-transitivity: $\mu_T(\mu_E(x,y),\ \mu_E(y,z)) \leq\ \mu_E(x,z)$.

Where $\mu_E(x,y)$, $\mu_E(y,z)$ and $\mu_E(x,z)$ are the grade of membership of the ordered pairs *(x,y)*, *(y,z)*, and *(x,z)* in *E*, with respect to a triangular norm (*t-norm*) *T*.

The concept of fuzzy order was introduced by generalizing the notion of (i) reflexivity $\mu_R(x,x)$ for any *x* ∈ *X*, (ii) antisymmetry ($\mu_R(x,y)$ *and* $\mu_R(y,x)$) imply *x* = *y*, and (iii) transitivity ($\mu_R(x,y)$ *and* $\mu_R(y,z)$) imply $\mu_R(x,z)$), where *R* is a fuzzy relation called an *order relation* in *X* if it satisfies (i), (ii), and (iii). A set *X* in which an order relation has been given is called an *ordered set* (*semi-ordered set* or *partially ordered set*), i.e. a *fuzzy ordering* is a fuzzy relation which is transitive. A *fuzzy partial ordering*, *P*, is a fuzzy ordering which is reflexive and antisymetric ($\mu_P(x,\ y) > 0$ *and* $x{\neq}y$) imply $\mu_P(y,\ x)$ =0. A *fuzzy linear ordering* is a fuzzy partial ordering in which $x{\neq}y$ imply $\mu_S(x,\ y)$>0 or $\mu_S(y,\ x)$>0. A *fuzzy preordering* is a fuzzy ordering which is reflexive. A *fuzzy weak ordering* is a fuzzy preordering in which $x \neq y$ imply $\mu_S(x, y) > 0$ or $\mu_S(y, x) > 0$.

In the last decade Ulrich Bodenhofer [1],[2] and [3] has presented a general framework for comparing fuzzy sets with respect to a general class of fuzzy orderings. This approach includes known techniques based on generalizing the crisp linear ordering of real numbers by means of the extension principle, applicable to any fuzzy subsets of any kind of universe for which a fuzzy ordering is known−no matter whether linear or partial. A approach for fuzzification of the ordering relation and ways to compare fuzzy sets with different heights, and ways of how to refine the ordering relation by lexicographic hybridization with a different ordering method. A formal study of fuzzy orderings with applications to statistical analysis of numerical data, has been made by Bodenhofer and Klawonn [4], [5].

A fuzzy relation $L : X^2 \rightarrow$[0,1] is called *fuzzy ordering* with respect to a t-norm *T* and a *T*-equivalence $E : X^2 \rightarrow$[0,1], for brevity *T-E-ordering*, if and only if the following three axioms are fulfilled for all $x, y, z \in X$:

(i) *E*-Reflexivity: $\mu_E(x,y) \leq \mu_L(x,y)$

(ii) *T-E*-Antisymmetry: $\mu_T(\mu_L(x,y),\mu_L(y,x)) \leq \mu_E(x,y)$

(ii) *T*-transitivity: $\mu_T(\mu_L(x,y),\mu_L(y,z)) \leq \mu_L(x,z)$.

Where $T-E-ordering$ L is *strongly complete* if $\mu_T(\mu_L(x,y),\ \mu_L(y,x))$ =1 for all $x,y \in X$, $\mu_{E_r}(x,y)$=max(0,1-$\frac{1}{r} * |x-y|$) is a μ_{T_L}-Equivalence on ***R*** (assume $r>0$), and $\mu_{T_L}(x,y)$ denoted the Lukasiewicz *t-norm*.

$$\mu_{T_L}(x, y) = max(0, x + y - 1) \tag{11}$$

For all $x,\ y \in X$, and based on the definition of strongly complete fuzzy orderings [4] and [5],

$$\mu_{L_r}(x, y) = min(1, max(0, 1 - \frac{1}{r} * (x - y))) \tag{12}$$

is a strongly complete $T_L - E_r-$ordering on ***R***. In order to generalize the notion of concordant and discordant pair, a binary fuzzy relation $R : X^2 \rightarrow$[0,1] is called a *strict fuzzy ordering* with respect to a t-norm *T* and a *T*-equivalence *E*, for brevity *strict T-E-ordering*, if it is irreflexive $\mu_R(x,x)=0$ for all $x \in X$, *T*-transitive, and *E*-extensional $\mu_T(\mu_E(x,x^{'}),\mu_E(y,y^{'}),\mu_R(x,y)) \leq \mu_R(x^{'},y^{'})$, for all $x,\ y,\ z \in X$. Given a $T_L - E-$ordering *L* strongly complete, it can be proven that the fuzzy relation R_x is defined as:

$$\mu_{R_x}(x_1, x_2) = 1 - \mu_{L_x}(x_2, x_1) \tag{13}$$

Analogously for all $y \in Y$ R_y is defined as:

$$\mu_{R_y}(y_1, y_2) = 1 - \mu_{L_y}(y_2, y_1) \tag{14}$$

C. A Fuzzy Ordering-Based Rank Correlation Coefficient

Bodenhofer and Klawonn in [4] and [5] demonstrate that established rank correlation measure are not ideally suited for measuring rank correlation for numerical data that are perturbed by noise, they propose to use robust rank correlation meassures based on fuzzy orderings named Fuzzy Rank Correlation and demonstrate that the new measures overcome the robustness problems of existing rank correlation coefficients. The formal description is: Assume that the data are given as in (7), (domainx), and (domainy), where $x_i \in X$ and $y_i \in Y$ for all $i=1,\ .\ .\ .\ ,\ n$, this means that we have two T_L-equivalences $E_x : X^2 \rightarrow$[0,1] and $E_y : Y^2 \rightarrow$[0,1], a strongly complete $T_L - E_x-$ordering $L_x : X^2 \rightarrow$[0,1] with a strict $T_L - E_x-$ordering on *X* define as in (13) and a strongly complete $T_L - E_y-$ordering $L_y : Y^2 \rightarrow$[0,1] with a strict $T_L - E_y-$ordering on *Y* define as in (14).

According to the gamma rank correlations measure and given an index pair (*i, j*) where $i = (x_i, y_i)$ and $j = (x_j, y_j)$, we can compute the degree to which (*i, j*) is a concordant pair as:

$$C(i, j) = \mu_{T_L}(\mu_{R_x}(x_i, x_j), \mu_{R_y}(y_i, y_j)) \tag{15}$$

And the degree to which (*i, j*) is a discordant pair as

$$D(i, j) = \mu_{T_L}(\mu_{R_x}(x_i, x_j), \mu_{R_y}(y_j, y_i)) \tag{16}$$

The numbers of concordant pairs ***CT*** and discordant pair ***DT***, respectively, as:

$$CT = \sum_{i=1}^{n} \sum_{j \neq i} C(i, j) \tag{17}$$

$$DT = \sum_{i=1}^{n} \sum_{j \neq i} D(i, j) \tag{18}$$

So the *fuzzy ordering-based rank correlation meassure* γ can be computed as:

$$Fuzzy\_\gamma = \frac{CT - DT}{CT + DT} \tag{19}$$

Where $\mu_{T_L}(x,y)$, $\mu_{R_x}(x_1,x_2)$, $\mu_{R_y}(y_1,y_2)$, $\mu_{L_x}(x_2,x_1)$ and $\mu_{L_y}(y_2,y_1)$ by fuzzy orderings we can compute as in (11), (13), (14), and (12) respectively.

IV. Fuzzy Ordering-Based Rank Correlation Coefficient for Mining of Gradual Itemsets

A. Notations

The automatic extraction of gradual dependencies consists of two steps: 1. extraction of frequent gradual itemsets, and 2. extraction of causality relations between the items. In this work, we focus on the first step, and we consider the following notations: A data set $\mathcal{D}_S$, constituted of $\mathcal{N}$ objects or transactions (data record) denote by $\mathcal{T}$= $\{t_1, ..., t_{\mathcal{N}}\}$ described by $\mathcal{M}$ numerical attributes denote by $\mathcal{A}$= $\{A_1, ..., A_{\mathcal{M}}\}$. Table of Fig. 2. shows an example data set where $\mathcal{T}$= $\{t_1, t_2, t_3, t_4, t_5\}$ transactions and $\mathcal{A}$= $\{A_1 : age, A_2 : salary, A_3 : loans, A_4 : cars\}$ attributs, its graphic illustration is shown in the diagram and graphics of Fig. 2.

In this framework, let us consider a *gradual pattern* $GP_{l,p}$::=$\{I_l\}^k$ where $\{I_l\}^k$ ::= I_1 ... I_k, such that $I_1 \neq I_2 \neq$ ... $\neq I_k$, for k ::= 2 | 3 | ... | $\mathcal{M}$, each *gradual item* I_l::=$\mathcal{A}v$, where $\mathcal{A}$::= A_1 | A_2 | ... | $A_{\mathcal{M}}$, each A_m::=$id\_attribut$ $[vector\ of\ numeric\ values\ u_i]$ for *i*=1, 2,..., $\mathcal{N}$, and *v* ::= $\geq$ | $\leq$, represent a positive (ascending) variation in the numeric values of the attribute A_m (in case v ::= $\geq$) or a negative (descending) variation (in case v ::= $\leq$), see Fig. 3 a). For instance $GP_{3,p}$::={ $A_1 \geq A_2 \geq A_4 \leq$ } is interpreted as a *gradual pattern* of size $k = 3$ and level $l = 3$ illustrated in Fig. 3 b), where for case of the data set of table in Fig 2 it imposes an ascending variation on the values of attributes *age*(u_i, u_j) and *salary*(u_i, u_j) and a descending variation on the values of attribute *cars*(u_i, u_j) and are concordant pairs simultaneously.

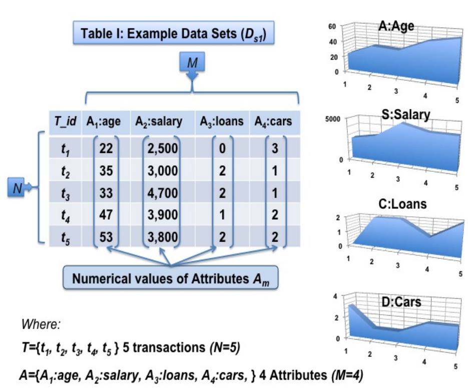
B. Algorithm of Extraction of Frequent Gradual Itemsets

In this context we propose an algorithm that evaluates gradual dependencies in terms of a fuzzy rank correlation coefficient, as described in the algorithms 1 and 2, where we apply the APRIORI algorithm to generate candidates from the $k-$itemsets to take advantage of the fact that any subset of a frequent itemset is also a frequent itemset and all infrequent itemsets can be pruned if it has an infrequent subset, see Figs 4 and 5. We implemented the Fuzzy Ordering-Based Rank Correlation Coefficient ($Fuzzy\_\gamma$) according to the formal description presented in the previous section, this in order to evaluate candidates itemsets and mining frequent gradual itemset.

C. Properties of the Proposed Method

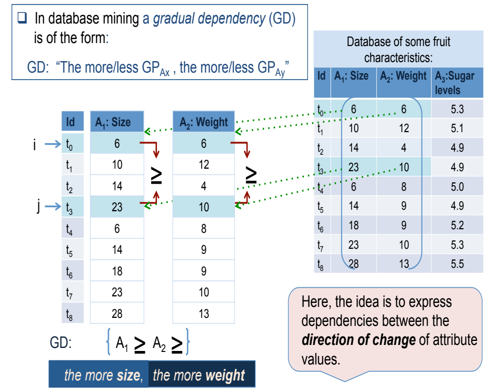
For us, in this work, the problem to address is the automatic extraction of frequent gradual itemsets, in which, relations between the directions of changing the values of the attributes involved are non-linear and/or affected by noise. Consequently, we propose a method of automatic extraction of frequent gradual itemsets on the basis of fuzzy orderings. To illustrate this, we consider the data set described in table and graphs of Fig. 2. Table I contains the list de concordant couplas, the numbers of concordant pairs *CT* and discordant pair *DT*, the support, and the fuzzy rank correlation coefficient ($Fuzzy\_\gamma$), for several gradual itemsets.

Properties of the proposed method and algorithms are: (i) In order to compute the degree to which each index pair $\mathcal{C}(i,j) \leftarrow min(\mathcal{R}_C[0], \mathcal{R}_C[1], ..., \mathcal{R}_C[k-1])$ are concordant pairs in itemsets $|I_s| > 2$, we exploit the properties of associativity and commutativity of t-norm of (15), (ii) In order to compute the degree to which each index pair



T_id	A_1:age	A_2:salary	A_3:loans	A_4:cars
t_1	22	2,500	0	3
t_2	35	3,000	2	1
t_3	33	4,700	2	1
t_4	47	3,900	1	2
t_5	53	3,800	2	2

Fig. 2. Notations of a Data Set.



Id	A_1: Size	A_2: Weight	A_3:Sugar levels
t_0	6	6	5.3
t_1	10	12	5.1
t_2	14	4	4.9
t_3	23	10	4.9
t_4	6	8	5.0
t_5	14	9	4.9
t_6	18	9	5.2
t_7	23	10	5.3
t_8	28	13	5.5

Fig. 3. Ilustration: a) Variations of a gradual item, b) A gradual itemset of size $k = 2$.

Algorithm 1: Fuzzy Gradual Dependencies Mining

Data: Transactions Database D_s, $\{A_M\}$, $minSupp$
Result: Fuzzy Frequent Gradual Dependencies $\mathcal{F}_{FGP}$

$\mathcal{F}_{FGP} \leftarrow \emptyset$;
$\mathcal{S}et\_g(I_l) \leftarrow Gen\_gItems(\{A_M\} \times v\{\leq, \geq\})$;
$k \leftarrow 2$;
$\mathcal{C}_{Gp_k} \leftarrow GenCand(\mathcal{S}et\_g(I_l), size(k), level(k))$;
$\mathcal{F}_k \leftarrow fuzzyOrderings(\mathcal{C}_{Gp_k}, aValues)$;
$\mathcal{L}\_F_k \leftarrow \mathcal{F}_k.\mathcal{L}istF_k$;
$\mathcal{F}_{FGP} \leftarrow \mathcal{F}_{FGP} \cup \{\mathcal{L}\_F_k\}$;
$k++$;
repeat
  $\mathcal{F}_k \leftarrow \emptyset$;
  $q \leftarrow 1$;
  $\mathcal{C}_{Gp_k} \leftarrow GenCand(\{\mathcal{L}\_F_{k-1}\}, size(k), level(k))$;
  $Cgp_{k,q} \leftarrow FirstCandidate \in \mathcal{C}_{Gp_k}$;
  foreach $Cgp_{k,q} \in \mathcal{C}_{GPk}$ **do**
    $Cgp_{k,q}.M \leftarrow T(Cgp_{k-1,a}.M, Cgp_{k-1,b}.M)$;
    $Support(Cgp_{k,q}) \leftarrow EvalSupport(Cgp_{k,q}.M)$;
    /* *minSupp stands for a user-specified minimum support value* */
    if $Support(Cgp_k) \geq minSupp$ **then**
      $\mathcal{F}_k.\mathcal{L}istF_k \leftarrow \mathcal{F}_k.\mathcal{L}istF_k \cup \{Cgp_{k,q}\}$;
    $q++$;
    $Cgp_{k,q} \leftarrow NextCandidate \in \mathcal{C}_{Gp_k}$;
  $Delet(\mathcal{F}_{k-1}.Matrices)$;
  $\mathcal{F}_{FGP} \leftarrow \mathcal{F}_{FGP} \cup \{\mathcal{F}_k\}$;
  $\mathcal{L}\_F_k \leftarrow \mathcal{F}_k.\mathcal{L}istF_k$;
  k++;
until $\mathcal{F}_{FGP}$ *does not grow any more*;

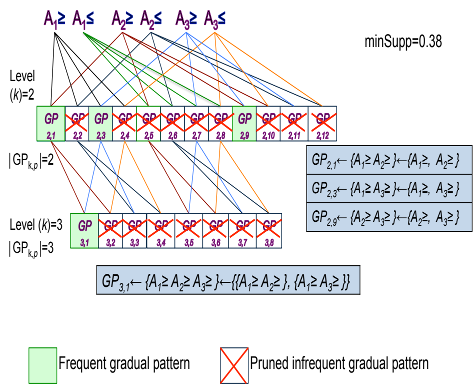


Fig. 4. Ilustration of APRIORI algorithm to generate candidates gradual patters.

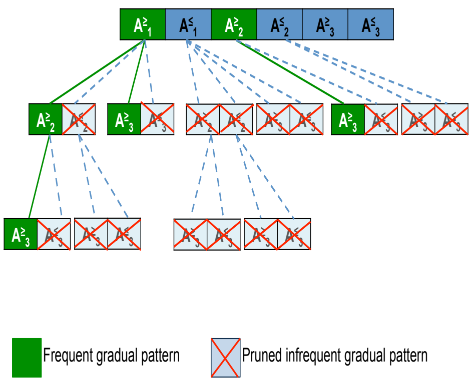


Fig. 5. Ilustration of APRIORI algorithm to generate candidates gradual patters.

TABLE I

EXAMPLES OF LISTS OF CONCORDANT COUPLES OF GRADUAL ITEMSETS

Itemset	List of concordant couples	*CT*	*DT*	*Support*	*Fuzzy_γ*
$A_1 \geq A_2 \geq$	{(0,1)(0,2)(0,3)(0,4)(1,3)(1,4)}	6	4	6/10	0.2
$A_1 \geq A_2 \geq A_3 \leq$	{(0,1)(0,2)(0,3)(0,4)}	4	5	4/10	-0.111
$A_1 \geq A_2 \geq A_3 \geq A_4 \leq$	{(0,1)(0,2)(0,3)(0,4)}	4	6	4/10	-0.2
$A_1 \geq A_3 \leq$	{(0,1)(0,2)(0,3)(0,4) }	4	5	4/10	-0.111
$A_2 \geq A_3 \leq$	{ (0,1)(0,2)(0,3)(0,4) }	4	4	4/10	0.00

$\mathcal{D}(i,j) \leftarrow min(\mathcal{R}_d[0], \mathcal{R}_d[1], ..., \mathcal{R}_d[k-1])$ are discordant pairs in itemsets $|I_s| > 2$, we exploit the properties of associativity and commutativity of t-norm of (16), (iii) Each $\mathcal{R}_C[ri]$, for the concordant case, is computed as $\mathcal{R}_C$[ri]← $\mathcal{R}_I(I.A_{ri}(u_i),\ I.A_{ri}(u_j))$, and for the discordant case as $\mathcal{R}_d$[ri]← $\mathcal{R}_I(I.A_{ri}(u_j),\ I.A_{ri}(u_i))$, (iv)The concordance degrees $\mathcal{C}(i,j)$ are stored in an $\mid N \mid \times \mid N \mid$ matrix, from which the total number of concordant pairs $\mathcal{CT}$ of an itemset I_s is computed by summing all entries, and (v) Finally, the support of itemset I_s is computed as: *Support*(I_s)= *CT*/$(n*(n-1))$, and the set of frequent gradual itemsets $\mathcal{I}_F$ is updated as $\mathcal{I}_F \leftarrow \mathcal{I}_F \cup \{I_s\}$ if *Support*(I_s) is $\geq$ *minSupp* (ε).

Algorithm 2: Fuzzy Ordering-Based Correlation

Data: Set of candidate gradual patters ($\mathcal{C}_{Gp_k}$), Size(k), $aValues$, and $minSupp$

Result: Frequent Gradual Patterns($\mathcal{F}_{k=2}$), $Support()$

$\mathcal{CT} \leftarrow \emptyset$; /* *Concordant and Support* */
$\mathcal{DT} \leftarrow \emptyset$; /* *Discordant* */
$q \leftarrow 1$;
$Cgp_{2,q} \leftarrow FirstCandidate \in \mathcal{C}_{Gp_{k=2}}$;
foreach $Cgp_{2,q} \in \mathcal{C}_{Gp_{k=2}}$ **do**
  /* *Compute of concordant $\mathcal{C}$ and discordant $\mathcal{D}$ pair $(u_i,u_j) \in aValues$* */
  for *i=0, 1, 2, ... ,* $\mathcal{N}-1$ **do**
    foreach $j \in$ { *0, ... ,* $\mathcal{N}-1$ } *and j* $\neq$*i* **do**
      /* *Compute: Relatinships R_I of each Item I* $\{A\geq \mid A\leq\} \in Cgp_{2,q}$ */
      for *ri=0, ... , 2-1* **do**
        if *variation is* " $\geq$ " **then**
          $\mathcal{R}_C$[ri]← $\mathcal{R}_{A_{ri}}(u_i,u_j)$;
          $\mathcal{R}_d$[ri]← $\mathcal{R}_{A_{ri}}(u_j,u_i)$;
        if *variation is* " $\leq$ " **then**
          $\mathcal{R}_C$[ri]← $\mathcal{R}_{A_{ri}}(u_j,u_i)$;
          $\mathcal{R}_d$[ri]← $\mathcal{R}_{A_{ri}}(u_i,u_j)$;
      /* *Compute: to each index pair $\in \mathcal{R}_I$ is concordant $\mathcal{C}$* */
      $\mathcal{C}(i,j) \leftarrow T\_norm(\mathcal{R}_C[0], \mathcal{R}_C[1])$;
      $\mathcal{CT} \leftarrow \mathcal{CT} + \mathcal{C}(i,j)$;
      /* *Compute: to each index pair $\in \mathcal{R}_I$ is discordant $\mathcal{D}$* */
      $\mathcal{D}(i,j) \leftarrow T\_norm(\mathcal{R}_d[0], \mathcal{R}_d[1])$;
      $\mathcal{DT} \leftarrow \mathcal{DT} + \mathcal{D}(i,j)$;
      $j++$;
    $i++$;
  $\mathcal{S}upport \leftarrow \mathcal{CT}/(n*(n-1))$;
  if $\mathcal{S}upport \geq minSupp\ (\varepsilon)$ **then**
    $\mathcal{F}_k.\mathcal{L}istF_k \leftarrow \mathcal{F}_k.\mathcal{L}istF_k \cup \{Cgp_{2,q}\}$;
    $\mathcal{F}_k.\mathcal{M}atrices \leftarrow \mathcal{F}_k.\mathcal{M}atrices \cup \mathcal{C}(i,j)$;
  q++;
$\mathcal{E}nd$;

V. Conclusions and Remarks

In this paper, we have presented a review of the basis and new models of fuzzy orderings, also we propose an original approach for extracting gradual itemsets. In our approach apply the APRIORI algorithm to generate candidates from the $k-$itemsets to take advantage of the fact that any subset of a frequent itemset is also a frequent itemset and all infrequent itemsets can be pruned if it has an infrequent subset, in order to evaluate candidates itemsets and mining frequent gradual itemset we implemented the Fuzzy Ordering-Based Rank Correlation Coefficient ($Fuzzy\_\gamma$) according to the formal description of Bodenhofer and Klawonn [4], [5] and Zadeh [12].

An important aspect to be addressed in future work includes the study of other optimizations in order to improve the efficiency of our approach (for example, the parallelization of our algorithm). Thus, in order to guarantee scalability, efficient pruning techniques are needed to avoid unnecessary comparisons. We will also study how causality can be defined based on this work, and efficiently extracted.

REFERENCES

[1] Bodenhofer, U.: Orderings of Fuzzy Sets Based on Fuzzy Orderings Part I: The Basic Approach. In Mathware & Soft Computing 15, 201-218 (2008)

[2] Bodenhofer, U.: Orderings of Fuzzy Sets Based on Fuzzy Orderings Part II: Generalizations. In Mathware & Soft Computing 15, 219-249 (2008)

[3] Bodenhofer, U.: Fuzzy Orderings of Fuzzy Sets. In Proc. 10th IFSA World Congress, Istanbul, pp. 500 - 507. http://www.bioinf.jku.at/people/bodenhofer/private/publications/Conferences.html (july, 2003)

[4] Bodenhofer, U., and Klawonn, F.: Towards Robust Rank Correlation Measures for Numerical Observations on the Basis of Fuzzy Orderings. In 5th Conference of the European Society for Fuzzy Logic and Technology, pp. 321 - 327, Ostrava, Czech Republic, University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, (septembre, 2007)

[5] Bodenhofer, U., and Klawonn, F.: Roboust Rank Correlation Coefficients on the Basis of Fuzzy Orderings: Initial Steps. In Mathware & Soft Computing 15, 5-20 (2008).

[6] Do T.D.T., Laurent, A., and Termier, A.: PGLMC: Efficient Parallel Mining of Closed Frequent Gradual Itemsets. In Proc. International Conference on Data Mining (ICDM), Sydney, Australia, 138-147, IEEE Computer Society Press (2010)

[7] Dubois, D., Hullermeier, E., and Prade, H.: A systematic approach to the assessment of fuzzy association rules. Data Mining and Knowledge Discovery, 13 (2):167-192, (2006)

[8] Ismat B., and Samina A.: Numerical Representation of Product Transitive Complete Fuzzy Orderings. In Journal Elsevier Mathematical and Computer Modelling. 53, pp. 617 - 623 (2011)

[9] Koh, H-W., and Hullermeier, E.: Mining Gradual Dependencies Based on Fuzzy Rank Correlation. In Combining Soft Computing and Statistical Methods in Data Analysis. Borgelt, C., Rodrguez, G.G., Trutschnig, W., Lubiano, M.A., Gil, M.., Grzegorzewski, P., Hryniewicz, O. (Eds.) Advances in Intelligent and Soft Computing, Vol. 77, pp. 379-386. Springer Heidelberg (2010)

[10] Laurent, A., Négrevergne, B., Sicard, N., and Termier, A.: Efficient Parallel Mining of Gradual Patterns on Multicore Processors. In AKDM2, Advances in Knowledge Discovery and Management. Springer (2010)

[11] Laurent, A., Lesot, M.-J., and Rifqi, M.: GRAANK: Exploiting Rank Correlations for Extracting Gradual Itemsets. In Proc. of the Eighth International Conference on Flexible Query Answering Systems (FQAS'09), T. Andreasen et al. (Eds), LNAI 5822, pp. 382-393, Springer-Verlag Berlin Heidelberg (2009)

[12] Zadeh, L., A.: Similarity relations and fuzzy orderings. In Information Sciences. Volume 3, Issue 2, pp. 177 - 200, April (1971)