



HAL
open science

Symbolic Approximate Reasoning Within Unbalanced Multi-sets: Application to Autism Diagnosis

Nouha Chaoued, Amel Borgi, Anne Laurent

► **To cite this version:**

Nouha Chaoued, Amel Borgi, Anne Laurent. Symbolic Approximate Reasoning Within Unbalanced Multi-sets: Application to Autism Diagnosis. AICCSA: ACS/IEEE Conference on Computer Systems and Applications, Oct 2017, Hammamet, Tunisia. pp.1494-1501, 10.1109/AICCSA.2017.74. hal-01847703

HAL Id: hal-01847703

<https://hal.science/hal-01847703v1>

Submitted on 25 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Symbolic Approximate Reasoning within Unbalanced Multi-Sets: Autism Diagnosis

Nouha Chaoued^{1,3,*}, Amel Borgi², and Anne Laurent³

¹ Université de Tunis El Manar, Faculté des Sciences de Tunis, LIPAH - LR 11 ES 14, 1002, Tunis, Tunisie

² Université de Tunis El Manar, Institut Supérieur d'Informatique, Faculté des Sciences de Tunis, LIPAH - LR 11 ES 14, 1002, Tunis, Tunisie
Amel.Borgi@insat.rnu.tn

³ Université de Montpellier, LIRMM - CNRS UMR 5506, F-34392 Montpellier, France
firstname.lastname@lirmm.fr

Abstract. Knowledge-Based Systems aim to reason and solve complex problems with a knowledge base, *i.e.* facts and rules. The reasoning process can be based on approximate reasoning. It allows deducing new facts from imprecise information. In this work, we focus on dealing with unbalanced linguistic terms in the multi-valued logic context. In this paper, we introduce a new method for approximate reasoning within unbalanced terms. This approach is based on a Generalized Modus Ponens model using Generalized Symbolic Modifiers. Moreover, we implement a tool for autism diagnosis based on the KBS **RAMOLI** with our inference engine.

Keywords: Unbalanced Multi-Sets, Approximate reasoning, Generalized Modus Ponens, Generalized Symbolic Modifiers, Autisme diagnosis

1 Introduction

In Artificial Intelligence, an important field is to develop machines with the human-like ability to reason with linguistic terms. Knowledge representation by computers must consider the fact that the information used in human thinking is often imprecise.

Zadeh proposed the fuzzy sets [1] to deal with the complexity of this representation based on fuzzy logic. Thereafter, De Glas introduced, in the multi-valued logic context, the multi-sets representation model [2, 3].

These two logics propose modeling membership degrees with words in the reasoning process. These words correspond to linguistic variables that take their values in a set of linguistic terms [4]. Each linguistic variable is associated to adverbs qualifying the belief degree of the veracity of a proposition.

In most Knowledge-Based Systems (KBS), the manipulated multi-sets include terms that are uniformly and symmetrically distributed (Fig. 1).

* Corresponding Author.

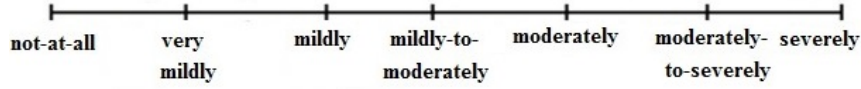


Fig. 1. Uniform Multi-set of 7 Linguistic Terms

Nevertheless, in other cases, linguistic information is represented with terms neither uniformly nor symmetrically distributed (Fig.2).

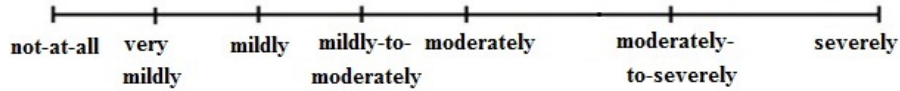


Fig. 2. Unbalanced Multi-set of 7 Linguistic Terms

KBSs include an inference process that can be based on approximate reasoning. It deduces meaningful outputs from imprecise inputs by means of the Generalized Modus Ponens (GMP) model.

In this work, we introduce an adaptation of the GMP model proposed in [5–7] to infer within unbalanced multi-sets. It is based on Generalized Symbolic Modifiers (GSM).

In the context of multi-valued logic, Bel Haj Kacem et al. [8] proposed a KBS designed for uniform multi-sets, named **RAMOLI**. We implement a new version that takes into account unbalanced multi-sets, named **ARUMS** (**A**pproximate **R**easoning within **U**nbalanced **M**ulti-**S**ets). This tool integrates our inference model considering unbalanced multi-sets. We test our proposal to autism diagnosis based on the **D**iagnostic and **S**tatistical Manual of **M**ental Disorders (DSM) 5 [9].

This paper is organized as follows. Section 2 introduces the basic concepts of multi-valued logic, Generalized Symbolic Modifiers, and the unbalanced multi-sets. Then, we present existing works treating the approximate reasoning in the multi-valued logic context in section 3. In section 4, we propose our GMP model dealing with unbalanced multi-sets. The section 5 presents the tests results for autism diagnosis.

2 Preliminaries

Multi-valued logic is based on De Glas’s theory [2]. In this approach, each linguistic variable is represented by an ordered and finite list of M linguistic symbols. They constitute a multi-set denoted by [2, 10]:

$$L_M = \{\tau_0, \tau_1, \dots, \tau_{M-1}\}; M \in \mathbb{N}^* \setminus \{1\} \quad (1)$$

τ_i is the membership degree to the multi-set L_M with $i \in [0, M - 1]$. The membership relation in multi-valued logic is partial:

$$x \in_\alpha A \Leftrightarrow x \text{ belongs to } A \text{ at a degree } \alpha \quad (2)$$

It should be noted that symbolic degrees are connected only by the total order relation \leq defined by [3]:

$$\tau_\alpha \leq \tau_\beta \Leftrightarrow \alpha \leq \beta; \forall \alpha \text{ and } \beta \in [0, M - 1] \quad (3)$$

Each degree τ_i of L_M is assimilated to a linguistic term corresponding to the precision degree of a proposition. It is a qualifier ϑ_α that express the imprecision of a predicate.

$$\begin{aligned} x \text{ is } \vartheta_\alpha A &\Leftrightarrow (x \text{ is } \vartheta_\alpha A) \text{ is true} \\ &\Leftrightarrow (x \text{ is } A) \text{ is } \tau_\alpha \text{ true} \end{aligned} \quad (4)$$

For example saying that *the communication is little impaired* means that *the communication* satisfies the predicate *impaired* with the degree τ_1 (Fig.3).

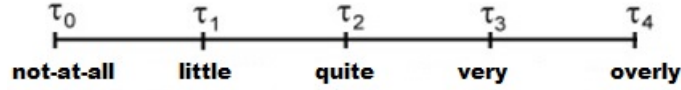


Fig. 3. Uniform Multi-set of 5 truth-degrees

The authors in [11] associate a numerical intensity level $prop(\tau_i)$ to each linguistic degree τ_i . It is understood as a proportion associated with τ_i : $prop(\tau_i) = \frac{i}{M-1}$; with M the size of the multi-set.

For the previous example *the communication is little impaired* means that *the communication* is considered by 25% ($\frac{1}{4} = prop(\tau_1)$) of asked persons as *impaired*.

In multi-valued context, the authors in [11–13] believe that any multi-valued symbol can be considered as a modification of another multi-valued symbol. A modifier allows to build a new term from an initial one, or to compare two values by finding out the modification that transforms from one to the other.

The membership relation corresponds to a symbolic degree τ_i from a multi-set L_M represented with [0-1] scale. Thus, the data modification is the transformation of the degree, the scale of the multi-set or the both. Indeed, some modifiers preserve the same multi-set but change the membership degree. Others erode or expand the multi-set. Akdag et al. [11, 12] proposed symbolic linguistic modifiers. They were generalized and formalized in [13, 14] as Generalized Symbolic Modifiers (GSM). Considering that the modifier is denoted by m , the result of applying m to τ_i is $\tau_{i'} : m(\tau_i) = \tau_{i'}$.

These modifiers are classified as:

- The weakening modifiers lead to the proportion decrease, i.e. $prop(\tau_{i'}) < prop(\tau_i)$. The four weakening modifiers defined in [13] are EW_ρ , DW_ρ , DW'_ρ and CW_ρ .
- The reinforcing modifiers involve the proportion increase, i.e. $prop(\tau_{i'}) > prop(\tau_i)$. The four reinforcing modifiers proposed in [13] are ER_ρ , DR_ρ , DR'_ρ and CR_ρ .
- The central modifiers preserve the proportion $prop(\tau_i)$ unchanged. Such modifiers act as a zoom on the base. The four central modifiers presented in [13] are EC_ρ , EC'_ρ , DC_ρ and DC'_ρ . The authors in [5] proposed another central modifier denoted by CC that conserves both the degree and the multi-set unchanged.

Kacem et al. [7] proposed a tool to aggregate modifiers. It takes into account the modifier mode, i.e. reinforcing or weakening. The authors introduce two operators:

- The M -norm A_T for modifiers conjunction based on T -norm.
- The M -conorm A_S based on a T -conorm for modifiers disjunction.

The authors consider three cases:

- Aggregation with CC ,
- Aggregation of two modifiers of the same type,
- Aggregation of two dual modifiers.

In their proposition only the modifiers CC , CW and CR are used. Tables 1 and 2 summarize the aggregation results for modifiers conjunction and disjunction respectively.

Table 1. Aggregation results by A_T [7]

A_T	CC	CR_α	CW_α
CC	CC	CC	CW_α
CR_β	-	$CR_\gamma, \tau_\gamma = T(\tau_\alpha, \tau_\beta)$	CW_α
CW_β	-	-	$CW_\gamma, \tau_\gamma = \neg T(\neg\tau_\alpha, \neg\tau_\beta)$

Table 2. Aggregation results by A_S [7]

A_S	CC	CR_α	CW_α
CC	CC	CR_α	CC
CR_β	-	$CR_\gamma, \tau_\gamma = S(\tau_\alpha, \tau_\beta)$	CR_β
CW_β	-	-	$CW_\gamma, \tau_\gamma = \neg S(\neg\tau_\alpha, \neg\tau_\beta)$

We recall that computing with words is an important issue in many domains dealing with imperfect data. In fact, researchers may describe their knowledge

with linguistic terms. Linguistic term sets are usually assumed to be uniformly distributed. The position of a term τ_i is $\frac{i}{M-1}$, which is its proportion. The distance between each pair of successive terms is the same, i.e. $\frac{1}{M-1}$ with M the number of terms in the set.

Nevertheless, in some cases, a particular sub-domain may be more informative than the remaining reference domain. Thus, the density of terms in this sub-domain is greater than in the others.

These linguistic term sets are not uniformly and symmetrically distributed. They are named unbalanced sets. Some authors [15–20] proposed examples from real life using these sets. Truck [21] indicates that such sets *"are commonly used to describe erratic phenomenons and irregular progression in the patterns such as weather events or behavior under the influence of alcohol"*. Xia and Xu [22] give the example of diminishing marginal utility in economics. They indicate that the investment of a same amount is different if the company has a bad performance or a good one. In this case, we need more precise bad information than good one. Hence, the gap between the grades expressing bad information are smaller than the one between the grades expressing good information.

In our work, we deal with unbalanced terms set in the context of multi-valued logic. This multi-set is defined as: $L_M = \{\tau_0, \tau_1, \dots, \tau_{(M-1)}\}; M \in \mathbb{N}^* \setminus \{1\}$; with M its cardinality.

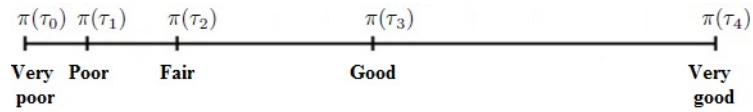
Each term τ_i represents a possible value for a linguistic variable. It is defined by its position $\pi(\tau_i)$ (Figure 4-a):

$$\pi : L_M \rightarrow [0, 1]$$

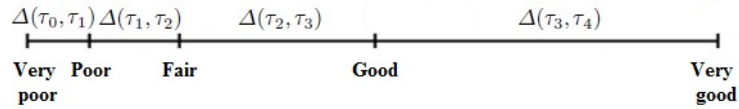
This function represents the distribution of the terms τ_i using real values between 0 and 1.

These terms are not necessarily symmetrically distributed or equidistant from each others, *i.e.* $\Delta(\tau_i, \tau_{i+1})$ are not equal for all $i \in [0, M - 1]$ (Figure 4-b):

$$\begin{aligned} \Delta : L_M \times L_M &\rightarrow [0, 1] \\ \Delta(\tau_i, \tau_j) &= \pi(\tau_j) - \pi(\tau_i) \end{aligned}$$



(a) Terms position



(b) Distances between terms

Fig. 4. Terms Definitions

In this paper, we propose an inference model for approximate reasoning within unbalanced multi-sets. This model is implemented in a tool for autism diagnosis.

3 Approximate Reasoning with Multi-Valued Logic

KBSs are able to perform reasoning without any exact measurement. Zadeh [4, 23–25] has introduced, in the fuzzy logic context, the concept of approximate reasoning. It allows inference with imprecise inputs to obtain meaningful outputs based on the Generalized Modus Ponens (GMP). GMP is an extension of the classic Modus Ponens for exact reasoning where new facts are obtained only when the observation and the rule premise are equal.

The schema of GMP in the multi-valued logic context is the following:

If $(X \text{ is } A)$ then $(Y \text{ is } B)$
 $(X \text{ is } A')$

$(Y \text{ is } B')$

With:

- A is the rule premise predicate;
- A' is the observation predicate;
- B is the rule conclusion predicate;
- B' is the inferred conclusion predicate;
- A' is approximately equal to A ;
- B' is approximately equal to B ;
- A, A', B and B' are represented by multi-sets $L_{M_A}, L_{M_{A'}}, L_{M_B}$ and $L_{M_{B'}}$, respectively.

A KBS leads to deduce new knowledge from the facts and rules existing in its knowledge base. Thus, new facts are obtained from ones that are approximately similar to the rule premise. The similarity between the premise and the observation allows determining the conclusion belief degree.

The approximate reasoning is composed of two steps. The first one aims to find out which modification the premise undergoes to obtain the observation. While, the second concerns the deduction of the inferred conclusion. Its membership degree depends on the rule conclusion degree and the relation between the premise and the observation.

The GMP model previously presented includes a free rule as its premise and its conclusion are completely true. While in the general case truth degrees are associated with the premise and the conclusion. The observation is represented by the same multi-set as the premise but with a different membership degree. These rules are named strong rules. Khoukhi [26] proposed a GMP with these rules as follows:

If $(X \text{ is } \vartheta_\alpha A)$ then $(Y \text{ is } \vartheta_\beta B)$
 $(X \text{ is } \vartheta_\gamma A)$

$(Y \text{ is } \vartheta_\lambda B)$

ϑ_α , ϑ_β , ϑ_γ , and ϑ_λ are linguistic qualifiers associated with the degrees τ_α , τ_β , τ_γ and τ_λ respectively (Eq.4).

The weakness of this type of reasoning is that it is concerned only with the degree of similarity between A and A' , i.e. the modification undergone by the premise to obtain the observation. However, the type of the modification is not considered.

In fuzzy logic context, Bouchon-Meunier [27] introduced a model of approximate reasoning based on linguistic modifiers. These later evaluate the similarity between the premise and the observation. The corresponding inference diagram adapted to the multi-valued logic context is the following:

$$\begin{array}{l} \text{If } (X \text{ is } \tau_\alpha A) \text{ then } (Y \text{ is } \tau_\beta B) \\ (X \text{ is } m(\tau_\alpha A)) \\ \hline (Y \text{ is } m'(\tau_\beta B)) \end{array}$$

With m and m' are linguistic modifiers.

Hence, deducing the inferred conclusion implies determining the modifier m' . It is obtained from the modifier m and the causality between the observation and the premise.

In the context of multi-valued logic, Bel Haj Kacem et al. [5–7] proposed an extended GMP model with GSM. The authors consider that the modifiers m and m' are equal. They suggested determining the modifier m by the means of the **det_mod** algorithm [6]. The modifier can be: *CR*, *CW*, *ER*, *DW*, *CC* [5,13] or the composition of two of them. Its identification depends on the granularities of the premise and the observation multi-sets (M_A and $M_{A'}$) and the truth degrees τ_α and τ_γ such that $m(\tau_\alpha A) = \tau_\gamma A'$.

The rule used in the previous GMP is called simple rule as the premise includes only one proposition. In the literature, the general case is the complex rule whose premise is a conjunction or a disjunction of propositions. The GMP with complex rules is as follows:

$$\begin{array}{l} \text{If } (X_1 \text{ is } \tau_{\alpha_1} A_1) \text{ op } (X_2 \text{ is } \tau_{\alpha_2} A_2) \dots \text{ op } (X_n \text{ is } \tau_{\alpha_n} A_n) \text{ Then } (Y \text{ is } \tau_\beta B) \\ (X_1 \text{ is } \tau_{\gamma_1} A_1) \text{ op } (X_2 \text{ is } \tau_{\gamma_2} A_2) \dots \text{ op } (X_n \text{ is } \tau_{\gamma_n} A_n) \\ \hline (Y \text{ is } \tau_\lambda B) \end{array}$$

With:

- **op** is one of the logical operators **and**, or **or**;
- A_1, A_2, \dots, A_n are the premise and observation predicates;
- n is the number of elements in the rule premise.

The inference engine aims to deduce the inferred conclusion $\tau_\lambda B$. It corresponds to $m(\tau_\beta B)$. The modifier m is the aggregation of the modifiers m_i [7]. Each modifier m_i transforms an element of the premise ($X_i \text{ is } \tau_{\alpha_i} A_i$) to an element of the observation ($X_i \text{ is } \tau_{\gamma_i} A_i$).

4 Approximate Reasoning within Unbalanced Multi-sets

Our aim in this paper is to extend the model proposed by Bel Haj Kacem et al. [5–7] to consider unbalanced multi-sets. To achieve this purpose, we use our algorithms for the representation of unbalanced multi-sets [28].

The first algorithm, named **UnbalancedToUniform**, allows to express an unbalanced term τ_{pos} , of a multi-set L_M , within a new uniform multi-set $L_{M'}$. This algorithm gives us the ability to use existing tools initially designed for uniform terms as aggregation tools and modifiers.

Each term τ_k of L_M can be represented within a specific uniform multi-set L_{M_K} . Hence, the granularity M' of the uniform multi-set, used to represent all unbalanced terms, equals the LCM of these sets granularity, i.e M_K .

The inputs of this algorithm are the M terms of the unbalanced set L_M , the normalized distances between each pair of successive terms $\Delta(\tau_i, \tau_{i+1})$, and the pair (τ_{pos}, α) with α a proportion error. α is a numerical value representing a symbolic translation associated with the term τ_{pos} . If α is positive, it is a reinforcement of τ_{pos} , otherwise it is a weakening. A pair $(\tau'_{\gamma_{pos}}, L_{M'})$ is given as the algorithm output. It corresponds to (τ_{pos}, α) in $L_{M'}$.

The second algorithm, named **UniformToUnbalanced**, aims to represent a uniform term τ'_i from $L_{M'}$ with a term τ_{pos} from an unbalanced set L_M . It tends to identify the term τ_{pos} whose position is the closest to $\pi(\tau'_i)$, with $\tau'_i \in L_{M'}$. τ_{pos} is specified by comparing the position of τ'_i with the sum of the distances $\Delta(\tau_i, \tau_{i+1})$ for all $\tau_i \in L_M$.

This algorithm inputs are the position of τ'_i , i.e. $\pi(\tau'_i)$, the M terms of the unbalanced multi-set L_M and the normalized distances $\Delta(\tau_i, \tau_{i+1})$ between each pairs of successive terms. As output, we obtain the pair (τ_{pos}, α) , $\tau_{pos} \in L_M$ and α the proportion error. α exists when no term τ_i corresponds to the exact position of τ'_i .

In this paper we treat complex strong rule using the following GMP model:

If $(X_1 \text{ is } \tau_{\alpha_1} A_1) \text{ op } (X_2 \text{ is } \tau_{\alpha_2} A_2) \dots \text{ op } (X_n \text{ is } \tau_{\alpha_n} A_n)$ then $(Y \text{ is } \tau_{\beta} B)$
 $(X_1 \text{ is } \tau_{\gamma_1} A'_1) \text{ op } (X_2 \text{ is } \tau_{\gamma_2} A'_2) \dots \text{ op } (X_n \text{ is } \tau_{\gamma_n} A'_n)$

$(Y \text{ is } \tau_{\lambda} B')$

With:

- **op** is one of the logical operators **and**, or **or**.
- A_1, A_2, \dots, A_n are predicates representing the premise
- A'_1, A'_2, \dots, A'_n are predicates representing the observation
- A_i and A'_i are semantically equivalent ($i \in \{1, \dots, n\}$)
- B and B' are the predicates representing the conclusion. They are semantically equivalent
- The unbalanced multi-sets $L_{M_{A_i}}$ ($i \in \{1, \dots, n\}$) (resp. L_{M_B}) and $L_{M_{A'_i}}$ ($i \in \{1, \dots, n\}$) (resp. $L_{M_{B'}}$) do not have necessarily the same granularity and terms distribution
- $\tau_{\gamma_i} = m_i(\tau_{\alpha_i})$ ($i \in \{1, \dots, n\}$); $\tau_{\lambda} = m(\tau_{\beta})$ with m_i and m are *GSM*

We tend to deduce the inferred conclusion τ_λ . Our inference process is as follows:

- Transform unbalanced terms τ_{α_i} , τ_β and τ_{γ_i} from the multi-sets $L_{M_{A_i}}$, L_{M_B} and $L_{M_{A'_i}}$ respectively into the uniform pairs $(\tau_{\alpha'_i}, L_{M'_{A_i}})$, $(\tau_{\beta'}, L_{M'_B})$ and $(\tau_{\gamma'_i}, L_{M'_{A'_i}})$ with the **UnbalancedToUniform** algorithm [28].
- Express all uniform terms within $L_{M''}$ such as:

$$M'' = 1 + LCM(M'_{A_1} - 1, \dots, M'_{A_n} - 1, M'_{A'_1} - 1, \dots, M'_{A'_n} - 1, M'_B - 1, M'_{B'} - 1)$$
 - $\tau_{\alpha''_i} = DC \frac{M''-1}{M'_{A_i}-1}(\tau_{\alpha'_i})$
 - $\tau_{\gamma''_i} = DC \frac{M''-1}{M'_{A'_i}-1}(\tau_{\gamma'_i})$
 - $\tau_{\beta''} = DC \frac{M''-1}{M'_B-1}(\tau_{\beta'})$
- Deduce the modifiers m_i : $m_i = \det_mod((\tau_{\alpha''_i}, L_{M''}), (\tau_{\gamma''_i}, L_{M''}))$
- Aggregate the modifiers m_i to determine the modifier m [7].
- Apply the modifier m to the rule uniform normalized conclusion $\tau_{\beta''}$:

$$\tau_{\lambda''} = m(\tau_{\beta''})$$
- Express the inferred conclusion by means of the uniform multi-set $L_{M'_{B'}}$:

$$\tau_{\lambda'} = EC \frac{M''-1}{M'_{B'}-1}(\tau_{\lambda''})$$
- Represent the uniform inferred conclusion $\tau_{\lambda'}$ on the unbalanced multi-set $L_{M_{B'}}$ by means of the **UniformToUnbalanced** algorithm [28].

To illustrate our method, we consider the context of autism diagnosis. The GMP schema in the multi-valued logic context is as follows:

If (*social interaction is mildly altered*) **and** (*stereotyped character is mildly present*)

Then (*child is mildly autistic*)

(*social interaction is severely altered*) **and** (*stereotyped character is moderately present*)

(*child is ϑ_λ autistic*)

We aim to determine the inferred conclusion τ_λ *autistic* corresponding to ϑ_λ .

The linguistic variable *present* indicates the degree of presence of an autism symptom. Some characters are altered for autistic child, they are represented within the linguistic variable *impaired*. The multi-set used to express these linguistic variables are: $L_7 = \{\text{not at all, very mildly, mildly, mildly to moderately, moderately, moderately to severely, severely}\}$ (Fig 2).

The linguistic variable *autistic* indicates the severity of the child affection by autism. The terms are the same as proposed in **CARS**: $L_4 = \{\text{not at all, mildly, moderately, severely}\}$ (Fig 5). In **CARS** [29], it is indicated that:

- A score below 30 does not allow a diagnosis of autism;

- Results between 30 and 37 represent a diagnosis of autism from mild to moderate;
- A score above 37 means a diagnosis of severe autism.

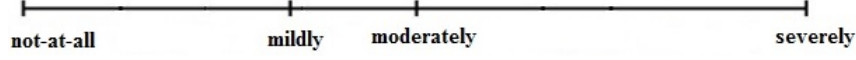


Fig. 5. Representation of Autistic Multi-set

We apply the **UnbalancedToUniform** algorithm [28] to the unbalanced terms sets L_7 and L_4 . The granularities of the uniform multi-sets are:

$$M' = 1 + LCM(9 - 1, 9 - 1, 9 - 1, 5 - 1, 5 - 1) = 9.$$

$$M'_{au} = 1 + LCM(4 - 1, 7 - 1, 3 - 1) = 7.$$

Hence, the GMP model with uniform multi-sets is as follows:

$$\begin{array}{l} \text{If } (\text{social interaction is } (\tau'_2, L_9)) \text{ and } (\text{stereotyped character is } (\tau'_2, L_9)) \\ \quad \text{Then } (\text{child is } (\tau'_2, L_7)) \\ \quad (\text{social interaction is } (\tau'_8, L_9)) \text{ and } (\text{stereotyped character is } (\tau'_4, L_9)) \\ \hline (\text{child is } (\tau'_\lambda, L_7)) \end{array}$$

The granularity of the normalized multi-set is:

$$M'' = 1 + LCM(9 - 1, 9 - 1, 7 - 1) = 25$$

To represent the terms within the normalized multi-set. Normally, we apply the modifier $DC_{\frac{M''-1}{M'_i-1}}$ to every term of the GMP:

$$\begin{array}{l} - \tau_{\alpha_1}'' = DC_{\frac{25-1}{9-1}}(\tau'_2) = DC_3(\tau'_2) = \tau_6'' \\ - \tau_{\alpha_2}'' = DC_{\frac{25-1}{9-1}}(\tau'_2) = DC_3(\tau'_2) = \tau_6'' \\ - \tau_{\beta}'' = DC_{\frac{25-1}{7-1}}(\tau'_2) = DC_4(\tau'_2) = \tau_8'' \\ - \tau_{\gamma_1}'' = DC_{\frac{25-1}{9-1}}(\tau'_8) = DC_3(\tau'_8) = \tau_{24}'' \\ - \tau_{\gamma_2}'' = DC_{\frac{25-1}{9-1}}(\tau'_4) = DC_3(\tau'_4) = \tau_{12}'' \end{array}$$

Afterwards, we determine the modifiers m_i that allows transforming each element of the premise:

$$m_1 = \text{det_mod}((\tau_6'', L_{25}), (\tau_{24}'', L_{25})) = CR_{18}$$

$$m_2 = \text{det_mod}((\tau_6'', L_{25}), (\tau_{12}'', L_{25})) = CR_6$$

These modifiers are aggregated as follows: $A_T(CR_{18}, CR_6) = CR_0$

$$\text{We use this modifier with the rule conclusion: } \tau_{\lambda}'' = CR_0(\tau_8'') = \tau_2''$$

This term is represented in the normalized uniform set L_{25} . To transform it into the initial unbalanced multi-set L_4 , we should first express it into the uniform multi-set, in this case L_7 : $\tau_{\lambda}' = EC_{\frac{25-1}{7-1}}(\tau_8'') = EC_4(\tau_8'') = \tau_2'$

The last step consists in finding the closest matching term back in L_4 . We use for that the **UniformToUnbalanced** algorithm proposed in [28]. The position of this term is the same of τ_1 in L_4 . Thus, the inferred conclusion is *the child is mildly autistic*.

5 Autism Diagnosis

We implement the tool **ARUMS** (**A**pproximate **R**easoning within **U**nbalanced **M**ulti-**S**ets) to take into account unbalanced multi-sets. It is based on the KBS designed for uniform multi-sets, **RAMOLI** [5].

Our tool integrates the inference model considering unbalanced multi-sets presented in the previous section. We test our proposal to diagnostic autism.

Autism is a pervasive developmental disorder characterized by the alterations of three principal areas [9]: verbal and nonverbal communication; social interaction; and behavior, interests and activities. Specialists notice that child behaviors are restricted and stereotyped.

The diagnosis is made by child psychiatrists. They observe the behavior of a child and ask its parents according to some standard protocols as **DSM5** [9]. An evaluation questionnaires exist based on **CARS** [29]. These questionnaires help psychiatrists to detect the autism symptoms.

In [30], the authors proposed to deal with autism diagnosis within uniform multi-sets. They extracted autism symptoms from the decision-making algorithm of **DSM5** [9]. With the help of child psychiatrists of Razi Hospital, they built a rules base by means of these symptoms.

The predicates used in our tool for autism diagnosis are:

- Nonverbal behaviors
- Ability to develop peer relationships
- Willingness to share
- Social reciprocity
- Emotional reciprocity
- Development of spoken language
- Conversation
- Stereotyped or idiosyncratic language
- Imitative play
- Stereotyped patterns of interest
- Patterns of interest abnormal in intensity
- Patterns of interest abnormal either in focus
- Inflexible or ritual,nonfunctional adherence
- Stereotyped and repetitive motor mannerisms
- Persistent preoccupation with parts of objects
- Social interactions
- Communication
- Stereotyped, restricted or repetitive characters
- Patient
- Reciprocity

- Patterns of interest

The rules base, used in the work of Bel Haj Kacem et al. [30] and the same as in our tool, includes:

- If Social interactions are mildly impaired and Communication is mildly impaired and Stereotyped, restricted or repetitive characters are mildly present then Patient is mildly autistic
- If Nonverbal behaviors are mildly impaired then Social interactions are mildly impaired
- If Ability to develop peer relationships is mildly impaired then Social interactions are mildly impaired
- If Willingness to share is mildly impaired then Social interactions are mildly impaired
- If Reciprocity is mildly impaired then Social interactions are mildly impaired
- If Social reciprocity is mildly impaired then Reciprocity is mildly impaired
- If Emotional reciprocity is mildly impaired then Reciprocity is mildly impaired
- If Development of spoken language is mildly impaired then Communication is mildly impaired
- If Conversation is mildly impaired then Communication is mildly impaired
- If Stereotyped or idiosyncratic language is mildly present then Communication is mildly impaired
- If Imitative play is mildly impaired then Communication is mildly impaired
- If Patterns of interest are mildly impaired then Stereotyped, restricted or repetitive characters are mildly present
- If Stereotyped patterns of interest are present then Patterns of interest are mildly impaired
- If Patterns of interest abnormal in intensity are present then Patterns of interest are mildly impaired
- If Patterns of interest abnormal either in focus are present then Patterns of interest are mildly impaired
- If Inflexible or ritual,nonfunctional adherence are present then Stereotyped, restricted or repetitive characters are mildly present
- If Stereotyped and repetitive motor mannerisms are present then Stereotyped, restricted or repetitive characters are mildly present
- If Persistent preoccupation with parts of objects are present then Stereotyped, restricted or repetitive characters are mildly present

We present in this section an implementation of our inference engine to diagnosis autism within unbalanced set. The multi-set representing the linguistic variable *autistic* is unbalanced in the **CARS** definition but not for the description of child behaviors, *i.e.* *impaired* and *present*. In fact, a note between 1 and 4 with a step of 0.5 is given to each symptom. For this reason, we use the uniform multi-set L_7 (Fig.1) for the linguistic variables *impaired* and *present* and the unbalanced multi-set L_4 (Fig.5) for *autistic* variable.

Table 3. Precision Recall F-measure for Autism Diagnosis tests

	Not autistic	Autistic		
		mildly autistic	moderately autistic	severely autistic
Accuracy	0.97	0.97		
Precision	0.92	0.67	1	0.35
Recall	1	0.8	0.08	1
F-measure	0.96	0.73	0.15	0.52

36 child with different autism severity degrees are treated by psychiatrist of Razi Hospital. The given questionnaire responses are integrated as observed facts in our knowledge base. Each patient will be described with 15 facts, *i.e.* the 15th first predicates indicated bellow, corresponding to the autism symptoms. For our tests, the treated child are:

- 11 not autistic;
- 6 mildly autistic;
- 13 moderately autistic;
- 6 severely autistic.

The inference engine applies rules to the knowledge base to deduce new knowledge. This process is iterated when a new fact in the knowledge base could trigger additional rules in the inference engine. This process allows to diagnosis if the child is autistic or not. It it is the case which is its severity degree.

The results of our tests are presented in the table 3.

We notice that our diagnosis tool gives good result to distinguish between autistic and not autistic child. In our test, only one child *mildly autistic* is diagnosed as *not autistic*. However, we obtain wrong diagnosis of autism severity. In particular *moderately autistic* child are considered as *severely autistic*.

6 Conclusion

Our contribution concerns approximate reasoning within unbalanced multi-sets. We propose to refine the model proposed by Kacem et al. [5–7]. The authors introduced an approximate reasoning based on GSM in the multi-valued context. We aim to extend their GMP model to consider unbalanced multi-sets. It allows inferring with complex strong rules.

We implement a KBS **ARUMS** that takes into account unbalanced multi-sets. To illustrate our method for approximate reasoning within unbalanced multi-sets, we treat the autism diagnosis based on **DSM5** and **CARS**.

As future work, we propose to use our tool with more patient to check the obtained results. We also aim to ameliorate the identification of autism severity degree.

Moreover, researchers propose, in the literature, aggregation operators in multi-valued logic context. Therefore, we aim to elaborate a tool to aggregate unbalanced terms without transforming them to a uniform multi-set.

References

1. Zadeh, L.A.: Fuzzy sets. *Information and Control* **8** (1965) 338 – 353
2. Glas, M.D.: Knowledge representation in a fuzzy setting. *Rapport interne* **89** (1989) 48
3. Akdag, H., Glas, M.D., Pacholczyk, D.: A qualitative theory of uncertainty. *Fundamenta Informaticae* **17** (1992) 333–362
4. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning - I. *Inf. Sci.* **8** (1975) 199–249
5. Kacem, S.B.H., Borgi, A., Ghédira, K.: Generalized modus ponens based on linguistic modifiers in a symbolic multi-valued framework. In: 38th IEEE International Symposium on Multiple-Valued Logic, May 22-23, 2008, Dallas, Texas, USA. (2008) 150–155
6. Kacem, S.B.H., Borgi, A., Tagina, M.: On some properties of generalized symbolic modifiers and their role in symbolic approximate reasoning. In: *Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence*, 5th International Conference on Intelligent Computing, Ulsan, South Korea, September 16-19, 2009. (2009) 190–208
7. Kacem, S.B.H., Borgi, A., Tagina, M.: Extended symbolic approximate reasoning based on linguistic modifiers. *Knowl. Inf. Syst.* **42** (2015) 633–661
8. Kacem, S.B.H., Borgi, A., Tagina, M.: Ramoli: A generic knowledge-based systems shell for symbolic data. In: 2013 World Congress on Computer and Information Technology (WCCIT). (2013) 1–6
9. Association., A.P., Association., A.P.: *Diagnostic and statistical manual of mental disorders : DSM-5*. 5th ed. edn. American Psychiatric Association Arlington, VA (2013)
10. Akdag, H., Pacholczyk, D.: Incertitude et logique multivalente, première partie : Etude théorique. *BUSEFAL* **38** (1989) 122–139
11. Akdag, H., Mellouli, N., Borgi, A.: A symbolic approach of linguistic modifiers. *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Madrid (2000) 1713–1719
12. Akdag, H., Truck, I., Borgi, A., Mellouli, N.: Linguistic modifiers in a symbolic framework. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **9** (2001) 49–61
13. Truck, I.: *Approches symbolique et floue des modificateurs linguistiques et leur lien avec l'agrégation: Application: le logiciel flous*. Phd thesis, University of Reims Champagne-Ardenne (2002)
14. Truck, I., Borgi, A., Akdag, H.: Generalized modifiers as an interval scale: Towards adaptive colorimetric alterations. In: *Advances in Artificial Intelligence*, 8th Ibero-American Conference on AI, Seville, Spain, November 12-15, 2002. (2002) 111–120
15. Wang, J., Hao, J.: A new version of 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Systems* **14** (2006) 435–445
16. Herrera-Viedma, E., López-Herrera, A.G.: A model of an information retrieval system with unbalanced fuzzy linguistic information. *Int. J. Intell. Syst.* **22** (2007) 1197–1214
17. Herrera, F., Herrera-Viedma, E., Martínez, L.: A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE T. Fuzzy Systems* **16** (2008) 354–370
18. Dong, Y., Xu, Y., Yu, S.: Computing the numerical scale of the linguistic term set for the 2-tuple fuzzy linguistic representation model. *IEEE Trans. Fuzzy Systems* **17** (2009) 1366–1378

19. Dong, Y., Zhang, G., Hong, W., Yu, S.: Linguistic computational model based on 2-tuples and intervals. *IEEE Trans. Fuzzy Systems* **21** (2013) 1006–1018
20. Dong, Y., Herrera-Viedma, E.: Consistency-driven automatic methodology to set interval numerical scales of 2-tuple linguistic term sets and its use in the linguistic GDM with preference relation. *IEEE Trans. Cybernetics* **45** (2015) 780–792
21. Truck, I.: Comparison and links between two 2-tuple linguistic models for decision making. *Knowledge-Based Systems* **87** (2015) 61–68
22. Xia, M., Xu, Z.: An approach to multiplicative linguistic group decision making based on possibility degrees. *International Transactions in Operational Research* (2015)
23. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning - II. *Inf. Sci.* **8** (1975) 301–357
24. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning - III. *Inf. Sci.* **9** (1975) 43–80
25. Zadeh, L.A.: A theory of approximate reasoning. *Machine intelligence* **9** (1979) 149–194
26. Khoukhi, F.: Approche logico-symbolique du traitement des connaissances incertaines et imprécises dans les systèmes à base de connaissances. Phd thesis, University of Reims Champagne-Ardenne (1996)
27. Bouchon-Meunier, B., Kreinovich, V.: Fuzzy modus ponens as a calculus of logical modifiers: Towards zadeh’s vision of implication calculus. *Inf. Sci.* **116** (1999) 219–227
28. Chaoued, N., Borgi, A., Laurent, A.: Linguistic modifiers with unbalanced term sets in multi-valued logic. In: *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Lisbon, Portugal, November 12-14, 2015.* (2015) 50–60
29. Schopler, E., Reichler, R.J., DeVellis, R.F., Daly, K.: Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of autism and developmental disorders* **10** (1980) 91–103
30. Kacem, S.B.H., Borgi, A., Othman, S.: A diagnosis aid system of autism in a multi-valued framework. (In: *Uncertainty Modelling in Knowledge Engineering and Decision Making - Proceedings of the 12th International FLINS Conference (FLINS 2016)*) 405–410