FuzzySentClass: Interval-valued fuzzy approach to the Sentiment Analysis Problem via SentiWordNet

Lidiane Costa da Silva*, Rafael Rodrigues Bastos*‡, Francisco Dias Franco*
Bruno Moura Paz de Moura†, Adenauer Correa Yamin* and Renata Hax Sander Reiser*

*Federal University of Pelotas (UFPEL) - Pelotas, Brazil
Laboratory of Ubiquitous and Parallel Systems (LUPS/CDTEC)
{lcdsilva, rrbastos, fdfranco, bmpdmoura, adenauer, reiser}@inf.ufpel.edu.br

†Federal Universitry of Pampa (UNIPAMPA) - Bagé/RS, Brazil
Diretoria de Tecnologia da Informação e Comunicação
brunomoura@unipampa.edu.br

‡Federal Institute of Education, Science and Technology
Sul-rio-grandense (IFSUL) - Bagé/RS, Brazil
Coordenadoria de Tecnologia da Informação e Comunicação
rafaelbastos@ifsul.edu.br

***Abstract*—Sentiment analysis, especially social network analysis (SNA), is a relevant research area. In recent years, this domain has become an active research question in data mining, natural language processing, and sentiment analysis (opinion mining). It consists of analyzing and extracting emotions, opinions, or attitudes from reviews of products, services, music, and movies, classifying them into positive, neutral, and negative, or even extracting the degree of importance (polarity). In this article, we propose a new approach using the Interval-valued Fuzzy Logic called FuzzySentClass to classify tweets based on lexicon using SentiWordnet. Our approach consists of classifying tweets according to three classes: positive, neutral, and negative, applying the FuzzySentClass that considers the steps of Fuzzification, Inference, and Defuzzification of an interval-valued fuzzy system. For the stage of obtaining results, the Juzzy platform was considered. The obtained results are evaluated based on the accuracy of the classifications obtained in the executions varying the type reducer in the FuzzySentClass Defuzzification step. In addition, the interval entropy approach is used to measure the imprecision information of achieved results. Our approach reached an accuracy of 83.22 with the centroid type reducer and 82.63 with the center of sets type reducer. And, resulting on the values of 0.117469 and 0.149853 as the maximum diameter of interval entropy for IvFS related to input and output variables, respectively.**

***Index Terms*—Interval-valued Fuzzy Logic, Interval-valued Fuzzy Sets, Sentiment Analysis, Interval Entropy**

I. Introduction

The evolution of the web, mainly related to the volume, speed, and variety of information about the opinions of network users, is increasing, making research in the area of sentiment analysis (SA) a strong trend and providing support for applications which help in decision-making contexts modeling uncertainty and imprecision. This area of research can provide essential and targeted information for business domain analysts (involving digital services, hosting, security) [1].

Social network analysis (SNA) is a research area collaborating in data mining and natural language processing (NLP). Sentiment analysis (opinion mining) consists of analyzing and extracting emotions, opinions, or attitudes from product reviews or movie reviews to classify them as positive, negative or neutral, providing information about opinions' degree of importance as polarity [2], [3].

Sentiment Analysis frequently assists decision-making as a closely related research area to Computer Science. Opinions (such as reviews, ratings, or comments) can be published through written or expressed text in digital media files (such as audio and video). Sentiment analysis can be used to verify the sentiment expressed by opinions posted on the Web [4].

Based on social networking popularization increases, the access to information provided by websites and their applications (such as Facebook and Twitter), instant messaging (such as WhatsApp, Discord, and Telegram), forums (such as Disqus and Reddit), and videos (such as Youtube), the users of these sites and apps were able to get the acknowledgment of other people's views on the subject they were looking for, as well as discuss and express their own opinions on the subject. Thus, as they interacted with more people, they consolidated their views about the subject.

People's opinions about products and services are significant for e-commerce or service providers. A positive opinion can boost your sales, as it can positively influence an internet user's perspective, making him buy that product or hire that service.

Identify applicable funding agency here. If none, delete this.

On the other hand, a negative opinion can negatively impact the user, making him look for products or services with better reviews.

In order to contribute to the specification of the sentiment values, this work considers an Interval Type-2 Fuzzy Logic (IT2FL), which is also named Interval-valued Fuzzy Logic, modeling not only the specialist uncertainties but also the imprecision of technological devices mapping products and services of web. Therefore, we can deal with approximate reasoning and give a closer view of the exact sentiment values that will help the producers, consumers, or any interested person to make an effective decision according to their product or service interest.

The paper is structured as follows. The first section deals with the contextual foundations of the work. Related works are presented in section II. Section III introduces basic concepts of Interval-valued Fuzzy Logic (IvFL). In section IV, details of the *FuzzySentClass* component and its conceptions are discussed, including database, fuzzification, rule base, inference, and defuzzification. Section V describes the experimental evaluation. Finally, section VI presents conclusions and future work.

II. Related Works

A new approach based on fuzzy logic for text classification, especially Twitter message classification is introduced in [5]. The inputs used in the proposed fuzzy logic-based model are H (word importance); (ii) F (tweet score); (iii) M (tweet duration); (iv) I (is a feature that indicates how many words in a tweet are equal to those words in the key list); (v) G a weight defined for the tweet; (vi) a function that represents the proportion of words used; (vii) patterns identified in tweets when confronted with a key list, the output is the degree of relevance for each analyzed tweet. They compared five commonly used defuzzification methods as experimental results and concluded that the centroid method is more effective and efficient than the other methods. In addition, they performed a comparison with the well-known keyword search method and the results revealed that the proposed fuzzy logic-based approach is more suitable for classifying relevant and irrelevant Twitter messages.

In [6], the authors presented an approach using neural networks and fuzzy sets to improve the quality of sentiment classification. This classification method uses fuzzy logic and neural network to design a classifier. In the fuzzification step, Gaussian membership functions were used, and the defuzzification method used is through the operator of a Multilayer Perceptron Backpropagation Network (MLPBPN).

In [7], the fuzzy system based on rules to obtain degrees of sentiment, trapezoidal membership functions were used in the fuzzification step and considering the maximum for defuzzification. As experimental results of this work, the authors compared the accuracy of the proposed approach with the precision of two other machine learning algorithms Naive Bayes and Decision Trees. The results show that the proposed fuzzy method reached the same level of performance as the two different algorithms.

In Liu et al. [8], the project addressing a fuzzy system based on rules as a computational model for accurate and interpretable sentiment analysis is reported. It consider Tsukamoto's fuzzy rule-based system supporting a fuzzy controler to classification problems. In the fuzzification step, trapezoidal membership functions were used, the min/max method for rule application and aggregation, and the MAX method for the defuzzification step. As experimental results, the fuzzy rule approach was compared with computational models learned using popular machine learning algorithms (Naive Bayes and C4.5) in sentiment classification. Four sets of movie review data were employed, improving the fuzzy rule learning approach when compared to the well-known Naive Bayes and C4.5 algorithms. Thus, indicating the suitability of the fuzzy rule approaches for learning task sentiment analysis.

In [9], the authors propose a new hybrid approach to classify tweets based on fuzzy logic and a lexicon-based method using SentiWordnet. This approach consists on classifying tweets according to three classes: positive, negative or neutral evaluations, using SentiWordNet integrated to fuzzy logic with its three essential steps: Fuzzification, Inference, and Defuzzification. The dataset of tweets to be classified and the classification result were stored in the Hadoop Distributed File System (HDFS), and the Hadoop MapReduce programming model considered for the application design.

This paper considers main attributes described in related works, providing a FuzzySentClass logical approach based on Interval-valued Fuzzy Logic, considering both, uncertainties and imprecision experts opinions, to model attitudes and reviews of products, services, music, and movies related to the degree of classification of analyzed tweets. The proposal was validated through executions considering an Interval-valued Fuzzy System designed with the Juzzy platform using SentiWordNet.

From Table I summarizing data, main characteristics of tools, like Knime [1] and HDFS [2] were used. The number of input variables (linguist terms) in most cases is 2, and the number of output variables is 1. In addition, the most used fuzzification method is trapezoidal membership functions, the inference considered Mandani and Tsukamoto. See, defuzzification is presented by several forms, and the most used connector is AND.

TABLE I: Comparison of related works.

Work	Tools	In/Out	Fuz	Inf	Def	Con
[5]	INA	7/1	Tra MF	Ma	♠ ◇ ♣ △ ♡	AND
[6]	NN	2/1	Gau MF	MLPBPN	MAX	INA
[7]	Knime	2/1	Tra MF	Tsukamoto	MAX	AND
[8]	INA	2/1	Tra MF	Tsukamoto	MAX	AND
[9]	HDFS	2/1	Tra, Tri MF	INA	INA	AND
★	Juzzy	2/1	Tra MF	Ma	♠ ■	AND

♠ Centroid ■ Center of Sets ◇ Bisector ♣ MOM △ SOM ♡ LOM Gaussian MF (Gau MF) Triangular MF (Tri MF) Trapezoidal MF (Tra MF) Mandani (Ma) Neural Network (NN) Information not available (INA) ★ FuzzySentClass

[1] https://www.knime.com/

[2] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

III. Fundations of Fuzzy Logic

Lotf Zadeh introduced T2FL in 1975 as an extension of the traditional FL [10] modeling the inherent uncertainties related to the antecedent and consequent membership functions, enabling the manipulation of imprecise terms throughout its fuzzy inference system [11].

Type-2 fuzzy sets (T2FS) emerged when no procedure was available to select the crisp membership degree $\mu_A(x)$ of an element $x \in \chi$ in a fuzzy set A, meaning that it is not a single real value [12]. These sets are handy in situations where there exists uncertainty about the degrees, forms or parameters of the membership functions [13], providing potential strategy on the treatment of uncertainties in information models based on multiple-criteria obtained from distinct specialists and/or extracted from simulators [14].

This logical proposal considers Interval Type-2 Fuzzy Logic (IT2FL) also named as Interval-valued Fuzzy Logic (IvFL), based on T2FS theory, modelling the uncertainty and imprecision of sentimental analysis attributes as interval-valued membership degrees related to an interval-valued fuzzy set (IvFS) A [15]. Thus, extending the Fuzzy Set (FS) theory, IvFS theory can model vagueness with an additional ability, considering imprecision (non-specificity) as another important aspect of uncertainty, reflecting this uncertainty by the length of the interval-valued membership degrees.

Definition 1: [13] A T2FS A is characterized by a type-2 membership function $\mu_A(\mathrm{x}, u)$ and given as follows:

$$A = \{((\mathrm{x}, u), \mu_A(x, u)) : \mathrm{x} \in \chi, u \in J_{\mathrm{x}} \subseteq [0,1]\}. \quad (1)$$

in which $0 \leq \mu_A(x, u) \leq 1$. And, u and $\mu_A(x, u)$ are called the primary and secondary membership function of $\mathrm{x} \in \chi$. A T2FS assigns to an element x in the universe χ a mapping $A(x, .)\colon [0,1] \to [0,1]$ is given as $A(\mathrm{x}, u) = A(\mathrm{x})(u)$, for every $\mathrm{x} \in \chi$, $u \in [0,1]$. In particular, in Type-1 fuzzy sets (T1FS) $A(\mathrm{x})$ is a real number in $[0,1]$, for every $\mathrm{x} \in \chi$.

Definition 2: [15] When all $\mu_A(\mathrm{x}, u) = 1$, then A is an interval type-2 fuzzy set (IT2FS), corresponding to

$$A(\mathrm{x}) = \{(u, 1) : u \in J_{\mathrm{x}} \subseteq [0,1]\}, \forall \mathrm{x} \in \chi.$$

Observe that Interval-valued Fuzzy Sets (IvFS) [16] are a particular case of T2FS. Let $L([0,1])$ be the set of all closed subintervals in $[0,1]$.

Definition 3: An Interval-valued Fuzzy Set (IvFS) is defined by the function $A : \chi \to L([0,1])$, such that the membership degree of an element $\mathrm{x} \in \chi$ is given as $A(\mathrm{x}) = [\underline{A}(\mathrm{x}), \overline{A}(\mathrm{x})] \in L([0,1])$. And, for each $\mathrm{x} \in \chi$, $\underline{A} : \chi \to [0,1]$ and $\overline{A} : \chi \to [0,1]$ are functions defining the lower and the upper bound of the membership $A(\mathrm{x})$, respectively.

In addition, let A, B be IvFS, the corresponding complement, union and intersection are also IT2FS given as:

$$A_C(\mathrm{x}) = [1 - \overline{A}(\mathrm{x}), 1 - \underline{A}(\mathrm{x})];$$
$$A(\mathrm{x}) \cup B(\mathrm{x}) = [\max(\underline{A}(\mathrm{x}), \underline{B}(\mathrm{x})), \max(\overline{A}(\mathrm{x}), \overline{B}(\mathrm{x}))];$$
$$\mu_{A \cap B}(\mathrm{x}) = [\min(\underline{A}(\mathrm{x}), \underline{B}(\mathrm{x})), \min(\overline{A}(\mathrm{x}), \overline{B}(\mathrm{x}))], \forall \mathrm{x} \in \chi.$$

For each $X = [\underline{X}, \overline{X}] \in L([0,1])$, $M(X) = \frac{\underline{X} + \overline{X}}{2}$ and $W(X) = \frac{\overline{X} - \underline{X}}{2}$ are the medium point and diameter of the extremes of interval X, respectively.

In this paper, we denote $A(\mathrm{x}) = X, B(\mathrm{x}) = Y, \forall \mathrm{x} \in \chi$, U as the set of all real intervals in the unit interval $[0,1]$ and $L([0,1])$ as the set of interval fuzzy values. The partial order on $L([0,1])$ is the product order [17] given as:

$$X \leq Y \text{ iff } \underline{X} \leq \underline{Y} \text{ and } \overline{X} \leq \overline{Y}, \text{for } X, Y \in L([0,1]).$$

A system based on IvFL can estimate input and output functions by using heuristic and interval techniques. Figure 1, graphically illustrates the inference system architecture based on IvFL. Its main blocks are briefly described as follows:

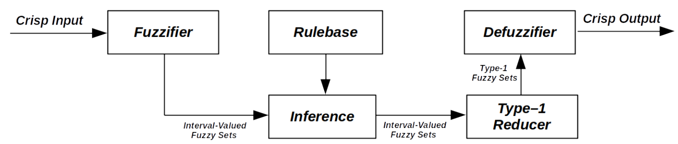


Fig. 1: Interval-Valued Fuzzy Controller Architecture

1 **Fuzzification Interface (Fuzzifier)**: The fuzzification process based on IVFL is performed according to the nature and definition of a type-2 fuzzy set. It associates an input value with an interval function and not simply with a single value in U. In other words, the uncertainty regarding the input membership function is inserted into the inference mechanism. Thus, for each IVFS A, an input vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \chi^n$, for $n \in \mathbb{N}^*$ is related to a pair of vectors in $\mathbb{U}^n$ given as follows:

$$\left(\overline{A}(x_1), \overline{A}(x_2), \ldots, \overline{A}(x_n)\right), \left(\underline{A}(x_1), \underline{A}(x_2), \ldots, \underline{A}(x_n)\right).$$

2 **Rule Base (RB)**: Composed of rules classifying linguistic variables (LVs) according to the IVFS;

3 **Logic Decision Unity (Inference)**: It executes inference operations between the input data and the rules defined in the RB to obtain performance by the system action;

4 **Defuzzification**: Considering two main stages of IVFS:

(i) **Type-1 Reducer** transforms an IVFS into a fuzzy set, that is, it provides the best fuzzy set that represents the IVFS, satisfying the following premise: when all uncertainties disappear, the result of the Interval-valued Fuzzy Ruled Based System (IV-FRBS) is reduced to a Fuzzy Ruled Based System (FRBS) [18];

(ii) **Defuzzifier** provides an output given as the average of the extremes $\underline{Y}$ and $\overline{Y}$, expressed as:

$$y = \frac{\underline{Y} + \overline{Y}}{2} = \frac{\underline{A}(x) + \overline{A}(x)}{2}, \forall x \in \chi, \quad (2)$$

and corresponding to the lower and upper bounds, related to the image by the membership function A applied to an element x in the universe χ. They are calculated using the iterative method of Karnik and Mendel (KM algorithm) [19].

The defuzzification step can still be obtained using a conventional method such as the centroid, as the final value of an inference system performance.

A. Fuzzy Connectives and Total Orders on $L([0,1])$

In the conception of the FuzzySentClass approach, methodologies and metrics for comparison and analysis of results based on admissible orders [20] is used, considering the need to compare the intervals produced as output for the fuzzy inference system of the FuzzySentClass approach, aiming at video traffic classification, with modeling based on T2FS.

The expectation with the use of admissible orders is to circumvent situations where two intervals can be understood as incomparable by usual methods of ordering real intervals. An example in this sense would be the product order "$\leq_{L([0,1])}$", which is a partial order relation, not a total one, and therefore allows two interval results to be incomparable ($X \nleq_{L([0,1])} Y$ and $Y \nleq_{L([0,1])} X$).

Linear orders in $L([0,1])$ are reflexive, antisymmetric, transitive and total binary relations. That is, a linear order on $L([0,1])$ is an order where any two pairs of subintervals of the unit interval $[0,1]$ are comparable.

A partial order $\leq_{L([0,1])}$ in $L([0,1])$ can be extended by an admissible order $\preceq_{L([0,1])}$, whenever the order $\preceq_{L([0,1])}$ is linear (total) and preserves the relationships already established by the partial ordering $\leq_{L([0,1])}$.

Admissible order classes are currently studied in many approaches that make use of fuzzy logic valued at intervals [21], and recent results already guarantee the total ordering and preservation of the diameters of interval data [22]. To substantiate this Thesis, the definition of the connectives considering these study proposals will be briefly characterized and exemplified below.

According to [23], admissible orders on $L([0,1])$ w.r.t. the product order can be obtained by aggregations. Additional information is available at [24] [22], [25]. See, the main used examples in the following:

1. Sejam $M_1, M_2 : [0,1]^2 \rightarrow [0,1]$ fuzzy aggregations such that, $\forall X, Y \in [0,1]$,

$$M_1(\underline{X},\overline{X}){=}M_1(\underline{Y},\overline{Y}) \vee M_2(\underline{X},\overline{X}){=}M_2(\underline{Y},\overline{Y}) \Leftrightarrow X{=}Y$$

The $\preceq_{M_1,M_2}$ admissible order relation is given as

$$X \preceq_{M_1,M_2} Y \Leftrightarrow \begin{cases} M_1(\underline{X},\overline{X}) < M_1(\underline{Y},\overline{Y}); \text{or} \\ M_1(\underline{X},\overline{X}) = M_1(\underline{Y},\overline{Y}) \text{ and } M_2(\underline{X},\overline{X}) \leq M_2(\underline{Y},\overline{Y}). \end{cases} \quad (3)$$

2. In particular, by Eq. (3), if $M_1(x,y) = (x+y)/2$ and $M_2(x,y) = y$, there is the Xu-Yager order [26] in $L([0,1])$, given by the expression:

$$[\underline{X},\overline{X}] \preceq_{XY} [\underline{Y},\overline{Y}] \Leftrightarrow \begin{cases} \underline{X}+\overline{X} < \underline{Y}+\overline{Y}; \text{or} \\ \underline{X}+\overline{X} = \underline{Y}+\overline{Y} \text{ and } \overline{X}-\underline{X} \leq \overline{Y}-\underline{Y}, \forall\, X,Y \in \mathbb{U}. \end{cases} \quad (4)$$

3. The Lexicographic Orders $\preceq_{Lex1}$related to the first variable, and $\preceq_{Lex2}$ the second variable, are respectively defined by the expressions:

$$[\underline{X},\overline{X}] \preceq_{Lex1} [\underline{Y},\overline{Y}] \Leftrightarrow \begin{cases} \underline{X} < \underline{Y}; \text{or} \\ \underline{X} = \underline{Y} \text{ and } \overline{X} \leq \underline{Y}; \end{cases} \quad (5)$$

$$[\underline{X},\overline{X}] \preceq_{Lex2} [\underline{Y},\overline{Y}] \Leftrightarrow \begin{cases} \overline{X} < \overline{Y}; \text{or} \\ \overline{X} = \overline{Y} \text{ and } \underline{X} \leq \underline{Y}. \end{cases} \quad (6)$$

In this case, there is also a particular case of Eq.(3), considering $M_1(x,y) = x$ and $M_2(x,y) = y$ for $\preceq_{Lex1}$, and the corresponding reverse projection for $\preceq_{Lex2}$.

The logical connectives in the lattice $(L([0,1], \preceq_{L([0,1)])})$, are defined analogously to $(L([0,1], \leq_{L([0,1)])})$, preserving related properties of total ordering provided by the admissible order $\preceq_{L([0,1)])}$, refining the partial order $\leq_{L([0,1)])}$.

Next, the negations and aggregations in $(L([0,1]$ applied in the FuzzySentClass model are presented, which were defined considering the admissible order of Xu-Yager, $\preceq_{XY}$).

Definição 1: According to [20], the function $\mathbb{N}_{XY} : L([0,1])^2 \rightarrow L([0,1])$ given as follows:

$$\mathbb{N}_{XY}(X) = \begin{cases} \left[1 - \frac{\overline{X}+3\underline{X}}{2}, 1 - \frac{\overline{X}-\underline{X}}{2}\right], & \text{if } \overline{X}+\underline{X} \leq 1 \text{ ;} \\ \left[\frac{\overline{X}-\underline{X}}{2}, 2 - \frac{3\overline{X}+\underline{X}}{2}\right], & \text{otherwise,} \end{cases} \quad (7)$$

is a strong IVFN referring to the Xu-Yager order, with equilibrium point $E_{XY} = [\frac{1}{4}, \frac{3}{4}]$.

Definição 2: The function $\mathbb{M}_n : L([0,1])^n \rightarrow L([0,1])$ given by the expression

$$\mathbb{M}_n(\overrightarrow{X},\overrightarrow{Y}) = \begin{cases} \mathbf{0}, \text{ if there exist } X_i = \mathbf{0}, 0 \leq i \leq n, \\ \left[\frac{1}{n}\Sigma_{i=1}^n \underline{X_i}, \frac{1}{n}\Sigma_{i=1}^n \overline{X_i}\right], \text{ otherwise;} \end{cases} \quad (8)$$

is an aggregation related to Xu-Yager order.

In the following definition, the function EN is defined by entropy-like properties valued at intervals, but restricted to the interval values on the lattice $(L([0,1], \preceq_{L([0,1)])})$.

B. Entropy and Total Orders on $L([0,1])$

In the case of the interval fuzzy approach, the measure of information imprecision has been added in the calculation of the fuzzy entropy, modeling the lack of precise knowledge of specialists about the degree of membership of an element, that is when its degree of membership is interpreted by an interval, defined by an interval membership function.

In this work, the output of an interval function may exhibit less or more uncertainty than its inputs. In this context, it is interesting to analyze the width of the membership interval entropy of input and output variables, which must be related by preserving the intervals referring to all process information, variables, and operators applied in the modeling.

The concept of interval entropy is considered, modeling the inaccuracy of the input data and preserving such inaccurate information, via the interval diameter, during the computations until obtaining the outputs. This entropy concept is constructed by aggregating functions and normal-functions EN. Furthermore, they allow the comparison and/or ordering of interval results by applying the concept of linear admissible orders.

This interval approach to entropy, presented below, will be applied as a metric in the validation of the fuzzy sets generated for the input and output attributes, defining support for the approximate reasoning modeled for the FuzzySentClass approach. For additional IvFS concepts, see [23], [27]–[29] and even more recently the results presented in [22], [25].

Definition 4: By [23, Def.5], let $\mathbb{N} : L([0,1]) \to L([0,1])$ a strong IVFN referring to the total order $\preceq_T$ with ε-equilibrium point. The function $EN_{IV} : L([0,1]) \to L([0,1])$ which satisfies the following properties:

1) $EN_{IV}(\varepsilon) = [1 - \omega(\varepsilon), 1]$;
2) $EN_{IV}(X) = 0_L$ if, and only if, $X = 0_L$ or $X = 1_L$;
3) Se $Y \preceq_T X \preceq_T \varepsilon$ or $\varepsilon \preceq_T X \preceq_T Y \preceq_T$, whenever $w(X){=}w(Y)$, then $EN_{IV}(X) \geq_T EN_{IV}(Y)$;

EN_{IV} is an interval-valued normal-function (IvNF) EN referring to an IvFN $\mathbb{N}$.

Example 1: Let the Xu-Yager $\preceq_{XY}$-order given by Eq. 4, $\mathbb{N}_{XY} : L([0,1]) \to L([0,1])$ a strong IvFN referring to the order $\preceq_{XY}$, given by Eq. 7. A função $EN_{IV} : L^2([0,1]) \to L([0,1])$, when $p \in [1, \infty)$, defined by:

$$EN_I(X){=}[1{-}|2M(X){-}1|^p{-}W(X), 1{-}|2M(X){-}1|^p], \quad (9)$$

defines an interval-valued normal-function EN, referring to $\mathbb{N}_{XY}$.

Now, in the following, let $\mathcal{F}_{L(0,1)}$ denoting the set of all IvFS defined on $(L([0,1]), \preceq_{L([0,1])})$.

Definição 3: Consider a $\preceq_T$-order, $\mathbb{N} : L([0,1]) \to L([0,1])$ as a strong IvFN referring to the $\preceq_T$-order. The function $E : \mathcal{F}_{L(0,1)} \to L([0,1]))$ is an interval-valued entropy function (IvFE) with respect to the strong IvFN $\mathbb{N}$ whenever, for every $A, B \in \mathcal{F}_{L(0,1)}$ the following properties are satisfied

($\mathbb{E}1$) $E(A) = \mathbf{0}$ if, and only if, A is a crisp set;
($\mathbb{E}2$) $E(\tilde{\epsilon}) = 1 - W(\epsilon)$, when $\tilde{\epsilon}$ denote the IvFS $A \in \mathcal{F}_{L(0,1)}$ such that $A(\mathrm{x} = \epsilon)$, for $\mathrm{x} \in \chi$;
($\mathbb{E}3$) $W(A(\mathrm{x})){=}W(B(\mathrm{x})) \Rightarrow E(A) \preceq_T E(B)$ and $W(A(\mathrm{x})){\preceq_T}W(B(\mathrm{x})){\preceq_T}\epsilon$ or $\epsilon{\preceq_T}W(B(\mathrm{x})){\preceq_T}W(A(\mathrm{x}))$.

Proposition 1: Be a universe set $\chi_n = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n\}$. Let be the total order $\leq_T$, on which a strong IVFN has been defined $\mathbb{N} : L([0,1]) \to L([0,1])$ and an interval-valued fuzzy aggregation $\mathbb{M} : L([0,1])^n \to L([0,1])$ that satisfies:

$\mathbb{A}7$ $\mathbb{M}(X, X, \ldots, X) = XS$, $\forall X \in L([0,1])$;
$\mathbb{A}8$ $\mathbb{M}(X_1, X_2, \ldots, X_n){=}\mathbf{0} \Leftrightarrow X_1{=}X_2{=}\ldots{=}X_n{=}\mathbf{0}$.

Example 2: According to [20], the function $\mathbb{M}_n : L([0,1])^n \to L([0,1])$ given by

$$\mathbb{M}_{XY}(\overrightarrow{X}) = \begin{cases} \mathbf{0}, \text{ if } X_i = \mathbf{0},\ 0 \leq i \leq n, \\ \left[\frac{1}{n}\Sigma_{i=1}^n \underline{X_i}, \frac{1}{n}\Sigma_{i=1}^n \overline{X_i}\right], \text{ otherwise}; \end{cases} \quad (10)$$

is an aggregation function w.r.t. $\preceq_{XY}$-order.

Example 3: Let $\tilde{A} \in \mathcal{F}_{L([0,1])}$ in $\chi_n = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n\}$. Consider the $\leq_{XY}$- order given by Eq. 4, $\mathbb{N}_{XY} : L([0,1]) \to L([0,1])$ as a strong IvFN referring to the $\leq_{XY}$-order , given by Eq. 7. For $p \in [1, \infty)$, the function $EN_{IV} : \mathcal{F}_{L([0,1])} \to L([0,1])$ defined by:

$$EN_{IV}(\tilde{A}){=}\mathbb{M}_{i=1}^n \left([1{-}|2(X_M){-}1|^p{-}W(X), 1{-}|2(X_M)){-}1|^p]\right), \quad (11)$$

is an interval-valued entropy function, referring to $\mathbb{N}_{XY}$, whenever $M(X) = \frac{\underline{X}+\overline{X}}{2}$ e $W(X) = \frac{\overline{X}-\underline{X}}{2}$ where $\mu_{\tilde{A}}(\mathrm{x}) = [\underline{X}, \overline{X}] \in L([0,1])$, $\forall \mathrm{x} \in \chi_n$.

IV. Modeling Fuzzy System

The FuzzySentClass is responsible for verifying the analyzed tweet's classification level (polarity). The FuzzySentClass system considers a rule base acting in three stages: Fuzzification, Inference, and Defuzzification, returning as output the classification (positive, neutral, negative) of each tweet analyzed in a data set.

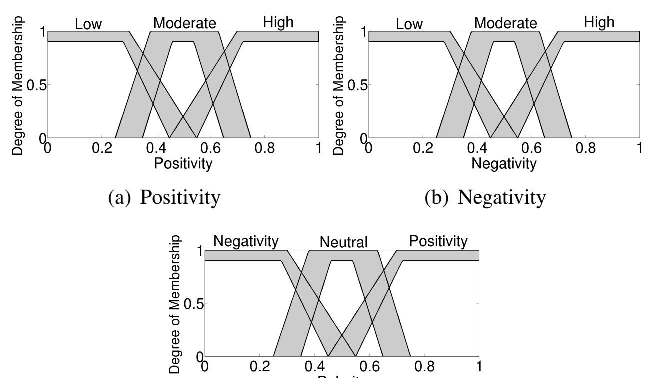
A. Definitions of Membership Functions

By studying the variables considering the related works found in the literature, the Linguistic Variables about each of the uncertainty variables were transformed into interval-valued fuzzy sets, using the trapezoidal form in the graphic representation of their membership functions.

From reading the tweets of the analyzed dataset, the values for the input variables are obtained: (i) Positivity; (ii) Negativity.

The graphical modeling presented in Figures 2(a) and 2(b) are related to the membership functions for all linguistic variable Positivity and Negativity considered for the input data, respectively. And, the modeling for linguistic output variables can be visualized in Figure 2(c). Such representation considers:

- The linguistic terms defining fuzzy sets for the Positivity variables are: Low, Moderate, and High;
- Related interval-valued fuzzy sets for the Negativity variable are: Low, Moderate, and High;
- The corresponding interval-valued fuzzy sets of the Polarity variable are named as Negativity, Neutral, and Positivity, representing the classification level regarding the tweet analyzed by the FuzzySentClass approach.



(a) Positivity

(b) Negativity

(c) Polarity

Fig. 2: Positivity, Negativity, and Polarity

B. Fuzzification

In this phase, the crisp input values are mapped to the fuzzy domain using trapezoidal membership functions to obtain each fuzzy set defined in the variables considered in the FuzzySentClass.

C. Rule Base

The Rule Base (RB) of FuzzySentClass is developed to be easily understandable and editable since there is no difficulty in adding new rules whether other input variables are desired to be manipulated. RB considers three factors for its construction:

- LV's name the FS's, turning the modeling closer to the real world system;
- The type "AND" connections are taken into account to create the relationship among the input variables;
- The type of implications *generalized modus ponens* (affirmative): "if X is A, then Y is B".

D. Inference

In the Inference process, the composition operators performed over FS relating the antecedents of rules to implications using the generalized modus ponens operator.

(i) performing the application of fuzzy operators and the input consists of three values resulting from fuzzification. The "AND" fuzzy operator aggregates the main rules and the method MIN (minimum) on the three returned values of fuzzification;

(ii) Implication Fuzzy Method Application: this step performs a combination of the value obtained in the fuzzy operator applied and the values of FS output rule, using the method MIN (minimum) on these combinations;

(iii) Aggregation Fuzzy Method Application: resulting composition of the fuzzy output of each rule by using the method MAX (maximum), thus creating a single fuzzy region to be analyzed by the next Fuzzy process module.

E. Defuzzification

With the research progress, the region transformation happens to be the result of the inference in a discrete value (which is the utilization). The utilized technique for modeling the system was FuzzySentClass center of the area.

This method calculates the centroid (u) of the area consisting of the output of the fuzzy inference system (connection of all contributions rules stated in sections IV-C and IV-D).

V. Experimental Results

In this section, we present some experimental results of our work. For that, we use a Twitter dataset [30]. The files referring to the source codes used in this work and the datasets are available in the public repository on the GitHub[3] platform.

A. Data Pre-Processing Stage

Before running the experiment with JuzzySentimentAnalysis, it was necessary to perform some dataset pre-treatment steps in order to use the SentiWordNet.

The SentiWordNet [31] is a lexical resource explicitly devised for supporting sentiment classification and opinion mining applications. The terms are grouped based in cognitive synonyms named synset. Each synset has a degree of positivity and negativity according to sentiment expressed. As the SentiWordNet has only english words, all tweets in another language were translated using Apache OpenNLP library[4]. In sequence, tweet tokenization is organized, splitting all tweet terms, removing spaces, commas, numbers, punctuation's, mentions with "@", URL and other symbols. So, each tweet becomes a tuple. Thus, stopwords are removed for each tuple, which are useless words for sentiment analysis, such as "a", "the", "to", "at". We did this based on dictionary available at Kaggle [5] platform.

Therefore, the next step we focused on to stemming each term of tuples using Porter Stemming Algorithm [6]. In this step the dataset becomes a set of tuples, which contains opinion words that need to be identified according of the part of speech. For this, we use the Apache OpenNLP POS Tagger tool. In this work, we consider only adjectives, adverbs, nouns and verbs.

After these data pre-processing stages, the dataset was ready to use the SentiWordNet dictionary. So, we get the corresponding value for positive and negative for each term, then we get the average of each one to the tweet. And, at this moment the system is ready to start the Fuzzy classification.

B. Classification Stage

The Table II presents the our experimental results. Based on simulations performed on the FuzzySentClass approach, the classification rate achieve 83%, and the corresponding significant error rate is 16.88%. This marks are approximate to the results reported, e.g., in [9], which are based on trapezoidal representations for membership functions, achieving 84.00% for classification rate and 16.00% for corresponding error rate. However, our approach provide an interval membership degree for each element in a fuzzy set of output data, over that we can perform an interval entropy measure.

TABLE II: Results on Classification Rate and Error Rate

Approach	Defuzzification	CR	ER
FuzzySentClass	Centroid	83.22%	16.88%
FuzzySentClass	Center of sets	82.63%	17,37%

Table III presents the results using the interval entropy as a metric for evaluate the imprecise information related to the input and output attributes.

The interval entropy is defined by Eq. (11) when $p = 1$ and aggregation $\mathbb{M}$ is given by Eq.(10). In this interpretation, the results present small values even for the greatest diameter of the interval entropies related to the fuzzy sets of linguistic terms (Negative, Neutral and Positive) in the output variable Polarity. The reduce diameter of intervals shows that the organized information has been preserved from the input variables (taking 0.117469 in the Low Positivity IvFS) to the output variable (taking 0.149853 in the Neutral Polariry IvFS).

[3] https://github.com/rafaelrodriguesbastos/FuzzySentClass

[4] https://opennlp.apache.org

[5] https://www.kaggle.com/datasets/rowhitswami/stopwords

[6] https://tartarus.org/martin/PorterStemmer/

TABLE III: Entropy Metric Results

Variables	**Input Variable**								
	Low			**Moderate**			**High**		
	Inf	**Sup**	**W**	**Inf**	**Sup**	**W**	**Inf**	**Sup**	**W**
Positivity	0.049181	0.16665	0.117469	0.0057449	0.11128	0.1055351	0.02517	0.061102	0.035932
Negativity	0.032219	0.14044	0.108221	0.0037964	0.083434	0.0796376	0.025972	0.053601	0.027629
	Output Variable								
	Negative			**Neutral**			**Positive**		
Polarity	**Inf**	**Sup**	**W**	**Inf**	**Sup**	**W**	**Inf**	**Sup**	**W**
CoC	0.018335	0.16369	0.145355	0.014677	0.16453	0.149853	0.029564	0.18111	0.151546
C	0.013443	0.15763	0.144187	0.0030352	0.14154	0.1385048	0.020822	0.1703	0.149478

CoC - Center of Sets Reducer **C** - Centroid Reducer **W** - Diameter

The uncertainty was preserved by the interval diameters, reflecting the lack of precise knowledge and the imprecision about the membership degrees of IvFS related to input attributes, which are propagate by the computations of the fuzzy controller processes, to corresponding IvFS related to the output attribute (polariry).

This analysis also considered the SentWordNet data set and two type reduction operators in the FuzzySentClass executions to obtain results, which are: (i) Centroid (C) and (ii) Center of Sets (CoS).

In the FuzzySentClass approach, the interval entropy analysis performed over data extracted in the IvFS interpreting the Polarity attribute presents better results for the Centroid reducer then the Center of Sets. This values referring to the calculation of the diameter of the linguistic terms of the output variable Polarity.

Moreover, for all analyzed cases, there are intervals with reduced diameter. It also means that smaller interval entropy, then better organized information in the IvFS modelling the FuzzySentClass approach.

VI. Conclusion

The *FuzzySentClass*, is presented as new approach for classifying tweets related to sentiment polarity using IvFL. The results were compared with two of the type-1 reducing methods (Center of Sets, Centroid). Our evaluation considered SentWordNet for classifying tweets, showing which the proposed method managed to reach 83.22% and 82.63% of accuracy. To analyze the disorganization of information, we offer an approach using interval entropy. The main advantage of the FuzzySentClass is related to the use of the IvFL, which adds to the model the treatment for inaccuracies and uncertainties in the full extension of the fuzzy controller process. In most cases, the proposed approach using IvFL achieved better results with the centroid type-1 reducer.

Further work addresses new aggregation functions in the inference process and prospects other datasets in the classification process, improving the result comparisons among other fuzzy controller in the literature.

References

[1] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on twitter using soft computing techniques," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 1, 2020. DOI: 10.1002/cpe.5107. [Online]. Available: https://doi.org/10.1002/cpe.5107.

[2] V. Chang, "A proposed social network analysis platform for big data analytics," *Technological Forecasting and Social Change*, vol. 130, pp. 57–68, 2018, ISSN: 0040-1625. DOI: https://doi.org/10.1016/j.techfore.2017.11.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004016251731524X.

[3] N. A. Ghani, S. Hamid, I. A. Targio Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Computers in Human Behavior*, vol. 101, pp. 417–428, 2019, ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2018.08.039. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S074756321830414X.

[4] U. B. Corrêa, "Análise de sentimento baseada em aspectos usando aprendizado profundo: Uma proposta aplicada à língua portuguesa," Doutorado em Computação, Universidade Federal de Pelotas, Pelotas, 2021, p. 123.

[5] K. Wu, M. Zhou, X. S. Lu, and L. Huang, "A fuzzy logic-based text classification method for social media data," in *2017 IEEE Intl.Conference on Systems, Man, and Cybernetics, SMC 2017, Banff, AB, Canada, October 5-8, 2017*, IEEE, 2017, pp. 1942–1947. DOI: 10.1109/SMC.2017.8122902. [Online]. Available: https://doi.org/10.1109/SMC.2017.8122902.

[6] J. B. Sathe and M. P. Mali, "A hybrid sentiment classification method using neural network and fuzzy logic," in *2017 11th Intl.Conference on Intelligent Systems and Control (ISCO)*, 2017, pp. 93–96. DOI: 10.1109/ISCO.2017.7855960.

[7] C. Jefferson, H. Liu, and M. Cocea, "Fuzzy approach for sentiment analysis," in *2017 IEEE Intl.Conference on Fuzzy Systems (FUZZ-IEEE)*, 2017, pp. 1–6. DOI: 10.1109/FUZZ-IEEE.2017.8015577.

[8] H. Liu and M. Cocea, "Fuzzy rule based systems for interpretable sentiment analysis," in *2017 Ninth Intl.Conference on Advanced Computational Intelligence (ICACI)*, 2017, pp. 129–136. DOI: 10.1109/ICACI.2017.7974497.

[9] Y. Madani, M. Erritali, J. Bengourram, and F. Sailhan, "A hybrid multilingual fuzzy-based approach to the sentiment analysis problem using sentiwordnet," *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, vol. 28, no. 3,

pp. 361–390, 2020. DOI: 10.1142/S0218488520500154. [Online]. Available: https://doi.org/10.1142/S0218488520500154.

[10] ". Zadeh", "The concept of a linguistic variable and its application to approximate reasoning—i," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975, ISSN: 0020-0255. DOI: http://dx.doi.org/10.1016/0020-0255(75)90036-5.

[11] J. M. Mendel, "Fuzzy sets for words: A new beginning," in *Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE Intl.Conference on*, vol. 1, 2003, pp. 37–42. DOI: 10.1109/FUZZ.2003.1209334.

[12] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 6, pp. 643–658, 1999. DOI: 10.1109/91.811231.

[13] N. N. Karnik and J. M. Mendel, "Introduction to type-2 fuzzy logic systems," in *1998 IEEE Intl.Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence*, vol. 2, 1998, 915–920 vol.2. DOI: 10.1109/FUZZY.1998.686240.

[14] K. Mittal, A. Jain, K. S. Vaisla, O. Castillo, and J. Kacprzyk, "A comprehensive review on type 2 fuzzy logic applications: Past, present and future," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103 916, 2020.

[15] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. Fuzzy Systems*, vol. 14, no. 6, pp. 808–821, 2006. DOI: 10.1109/TFUZZ.2006.879986.

[16] M. Gehrke, C. Walker, and E. Walker, "Some comments on interval valued fuzzy sets," *Intl.Journal of Intelligent Systems*, vol. 11, no. 10, pp. 751–759, 1996.

[17] E. Klement, R. Mesiar, and E. Pap, "Triangular norms. position paper I: Basic analytical and algebraic properties," *Fuzzy Sets and Systems*, vol. 143, no. 1, pp. 5–26, 2004.

[18] D. Wu and M. Nie, "Comparison and practical implementation of type-reduction algorithms for type-2 fuzzy sets and systems," in *FUZZ-IEEE*, IEEE, 2011, pp. 2131–2138, ISBN: 978-1-4244-7315-1. [Online]. Available: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5976945.

[19] J. M. Mendel, "On km algorithms for solving type-2 fuzzy set problems," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 426–446, 2013. DOI: 10.1109/TFUZZ.2012.2227488.

[20] H. Zapata, H. Bustince, S. Montes, *et al.*, "Interval-valued implications and interval-valued strong equality index with admissible orders," *Int. J. Approx. Reason.*, vol. 88, pp. 91–109, 2017. DOI: 10.1016/j.ijar.2017.05.009. [Online]. Available: https://doi.org/10.1016/j.ijar.2017.05.009.

[21] L. Costa, M. Matzenauer, A. C. Yamin, R. Reiser, and B. R. C. Bedregal, "Interval extension of the generalized atanassov's intuitionistic fuzzy index using admissible orders," in *2019 IEEE Intl. Conference on Fuzzy Systems, FUZZ-IEEE 2019, New Orleans, LA, USA, June 23-26, 2019*, IEEE, 2019, pp. 1–6. DOI: 10.1109/FUZZ-IEEE.2019.8858816.

[22] H. Bustince, C. Marco-Detchart, J. Fernandez, C. Wagner, J. Garibaldi, and Z. Takáč, "Similarity between interval-valued fuzzy sets taking into account the width of the intervals and admissible orders," *Fuzzy Sets and Systems*, vol. 390, pp. 23–47, 2020, Similarity, Orders, Metrics, ISSN: 0165-0114. DOI: https://doi.org/10.1016/j.fss.2019.04.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165011418304494.

[23] Z. Takác, H. Bustince, J. M. Pintor, C. Marco-Detchart, and I. Couso, "Width-based interval-valued distances and fuzzy entropies," *IEEE Access*, vol. 7, pp. 14 044–14 057, 2019. DOI: 10.1109/ACCESS.2019.2893800. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2893800.

[24] H. Bustince, J. Fernández, A. Kolesárová, and R. Mesiar, "Generation of linear orders for intervals by means of aggregation functions," *Fuzzy Sets and Systems*, vol. 220, pp. 69–77, 2013.

[25] F. Santana, B. Bedregal, P. Viana, and H. Bustince, "On admissible orders over closed subintervals of [0,1]," *Fuzzy Sets and Systems*, vol. 399, pp. 44–54, 2020, Fuzzy Intervals, ISSN: 0165-0114. DOI: https://doi.org/10.1016/j.fss.2020.02.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011420300555.

[26] Z. Xu and R. R. Yager, "Some geometric aggregation operators based on intuitionistic fuzzy sets," *Intl.journal of general systems*, vol. 35, no. 4, pp. 417–433, 2006.

[27] M. J. Asiain, H. Bustince, R. Mesiar, A. Kolesárová, and Z. Takác, "Negations with respect to admissible orders in the interval-valued fuzzy set theory," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 556–568, 2018. DOI: 10.1109/TFUZZ.2017.2686372. [Online]. Available: https://doi.org/10.1109/TFUZZ.2017.2686372.

[28] M. J. Asiain, H. Bustince, R. Mesiar, A. Kolesárová, and Z. Takáč, "Negations with respect to admissible orders in the interval-valued fuzzy set theory," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 556–568, 2017.

[29] U. Bentkowska, H. Bustince, A. Jurio, M. Pagola, and B. Pekala, "Decision making with an interval-valued fuzzy preference relation and admissible orders," *Applied Soft Computing*, vol. 35, pp. 792–801, 2015.

[30] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N project report, Stanford*, vol. 1, no. 12, p. 2009, 2009.

[31] S. Baccianella, A. Esuli, F. Sebastiani, *et al.*, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining.," in *Lrec*, vol. 10, 2010, pp. 2200–2204.