Explainable AI and Interval Type 2 Fuzzy sets

Indian Institute of Information Technology , Allahabad, India

presented by

Prof. U. S. Tiwary
Indian Institute of Information Technology Allahabad
ust@iiita.ac.in

Explainable AI and Interval Type 2 Fuzzy sets

Indian Institute of Information Technology , Allahabad, India

Content

Introduction : From Expert Systems to Machine Learning (Automatically Learned Systems)

- Limitations of Expert Systems
- Rise of Black Box Machine Learning Models
- Challenges of Data driven Learnable Systems
- Explainable and Interpretable AI

Indian Institute of Information Technology , Allahabad, India

Indian Institute of Information Technology , Allahabad, India

Why Explainable Artificial Intelligence ?
(Understandability, Intelligibility & Comprehensibility)

1. Designers and developers must have the chance of analyzing the generated model and discerning its meaning. i.e. to understand the structure of the system.
2. End-users must use the system as a decision support, so that the model must explain the phenomena under study.

- Understandability and intelligibility can be viewed as synonyms that are associated to a functional understanding of the model in ML. In other words, it refers to grasp how the model works, without trying to elucidate its inner procedure or to shed light on its internal representation. For FRBSs, intelligibility is primarily associated with inference.
- Comprehensibility was defined as the learning algorithm's ability for encoding its model in such a way that it may be inspected and understood by humans. This definition narrows the focus to the model itself. It is based on the comprehensibility postulate argued by Michalski.

A. Fernandez, F. Herrera, O. Cordon, M. Jose del Jesus and F. Marcelloni, "Evolutionary Fuzzy Systems for Explainable Artificial Intelligence: Why, When, What for, and Where to?," in *IEEE Computational Intelligence Magazine*, vol. 14, no. 1, pp. 69-81, Feb. 2019.

doi: 10.1109/MCI.2018.2881645 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8610271&isnumber=8610087

Explainable Artificial Intelligence (EAI)

EAI can be obtained by means of the synergy between Fuzzy Systems (Top down approach) and Learnable Systems (Bottom up approach).

The key ability of Fuzzy Systems is to explore an AI system by extending it in unknown regions , so that the systems behaviour can be studied and the system is able to explain the process it followed to make the output decision.

Indian Institute of Information Technology , Allahabad, India

Explainable Artificial Intelligence

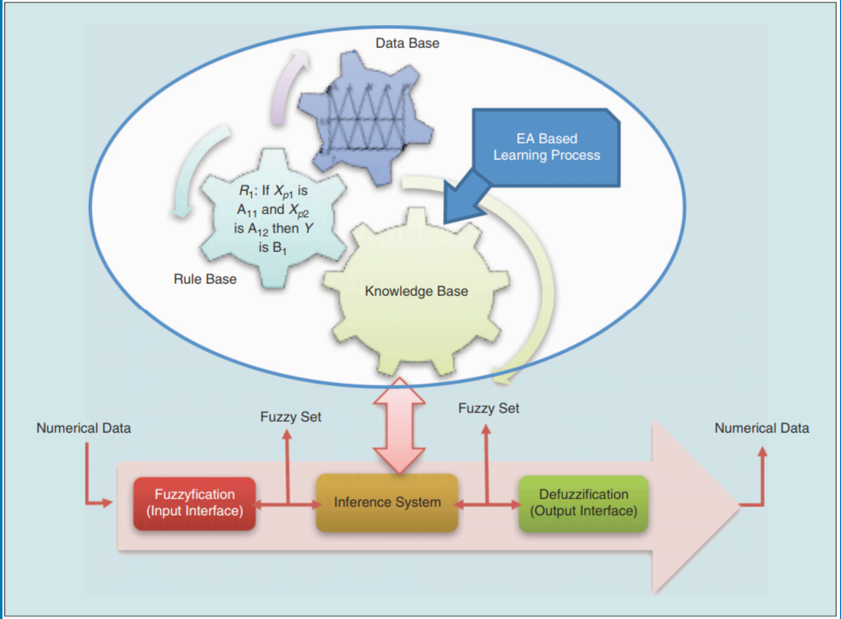


FIGURE 1 How EFSs are built on top of an FRBS. Inspired from [29].

A. Fernandez, F. Herrera, O. Cordon, M. Jose del Jesus and F. Marcelloni, "Evolutionary Fuzzy Systems for Explainable Artificial Intelligence: Why, When, What for, and Where to?," in *IEEE Computational Intelligence Magazine*, vol. 14, no. 1, pp. 69-81, Feb. 2019.

doi: 10.1109/MCI.2018.2881645 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8610271&isnumber=8610087

Explanaible Artificial Intelligence

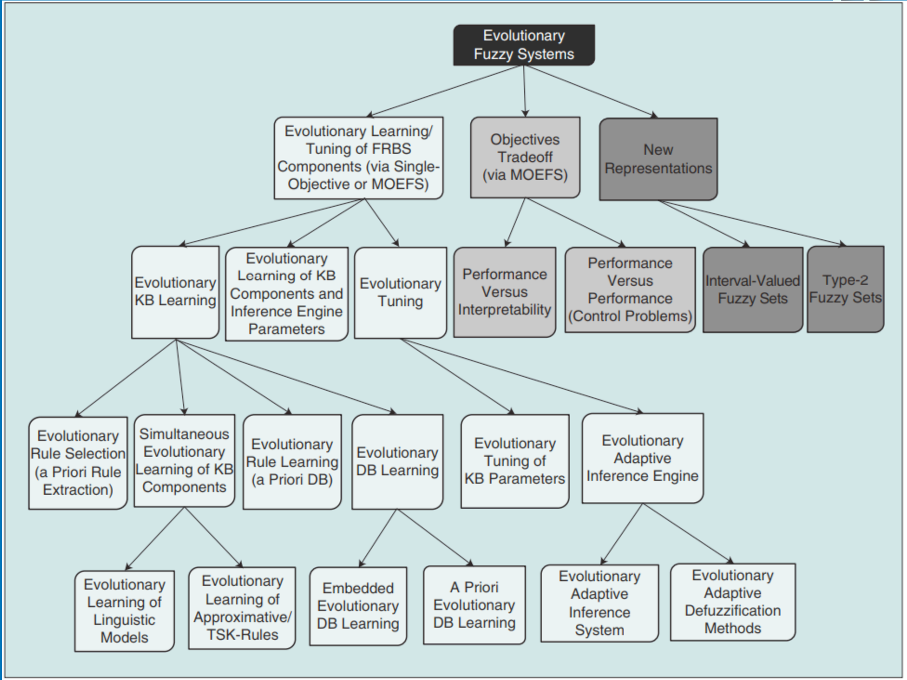


FIGURE 2 Evolutionary Fuzzy Systems Taxonomy. First presented in Fernández et al. [29].

A. Fernandez, F. Herrera, O. Cordon, M. Jose del Jesus and F. Marcelloni, "Evolutionary Fuzzy Systems for Explainable Artificial Intelligence: Why, When, What for, and Where to?," in *IEEE Computational Intelligence Magazine*, vol. 14, no. 1, pp. 69-81, Feb. 2019.

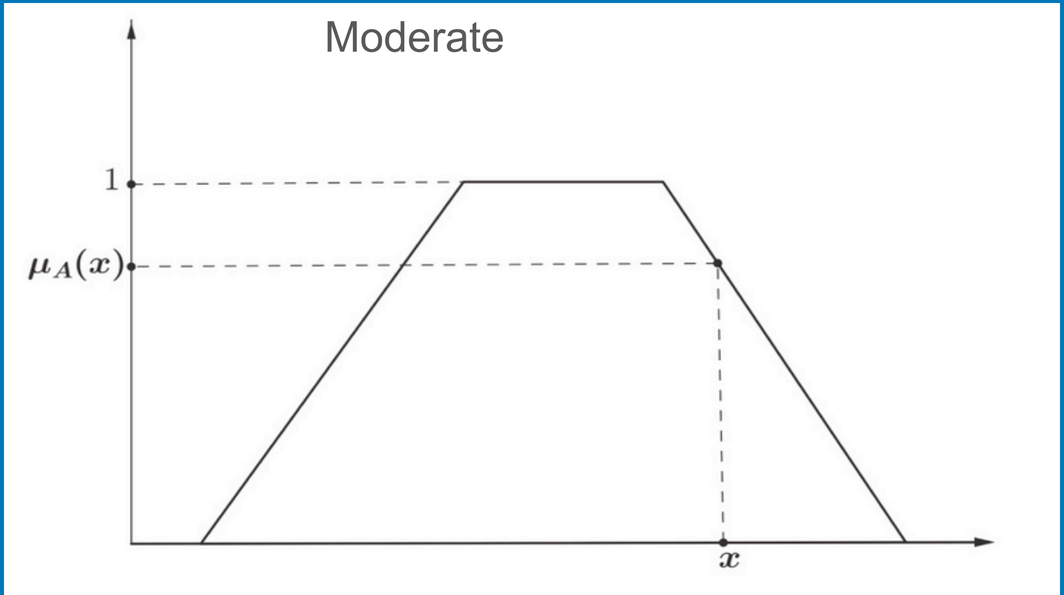
doi: 10.1109/MCI.2018.2881645 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8610271&isnumber=8610087

Indian Institute of Information Technology , Allahabad, India

Type 1 Fuzzy Set

- Fuzzy set are used to model the variable that donot have a crisp values.
- In language these variables are Linguistic variables and Fuzzy sets associated with them are Linguistic values.
- Example:
- **Linguistic Variable:** Temperature
- **Linguistic Values (**for Temperature**):** Cold, Moderate, Warm,etc.

Type 1 FS



Indian Institute of Information Technology , Allahabad, India

Extension Principle : Generalization of a mapping function f

- The extension principle is a basic concept of fuzzy set theory that provides a general procedure for extending crisp domains of mathematical expressions to fuzzy domains.
- This procedure generalizes an ordinary mapping of a function f to a mapping between fuzzy sets.

Fuzzy Relation and Composition

- Suppose that g is a function from X to Y, and A is a fuzzy set on X defined as

  $$A = \{(x_1, \mu_A(x_1)), (x_2, \mu_A(x_2)), \ldots, (x_n, \mu_A(x_n))\}$$

- *Then the extension principle states that the image of fuzzy set A under the mapping f can be* expressed as a fuzzy set $B \subseteq Y$.

- $B = f(A) = \{(y, \mu_B(y))\}$,

  where $\mu_B(y) = \max\limits_{x = f^{-1}(y)} \mu_A(x)$

Composition of Fuzzy Relations

- We can summarize various kinds of compositions as :

1) Composition of crisp sets A and B. It can represent a relation R between the sets A and B.

$$R = \{(x, y) \mid x \in A, y \in B\}, R \subseteq A \times B$$

2) Composition of fuzzy sets A and B. It is a relation R between fuzzy sets A and B.

$$R = \{((x, y), \mu_R(x, y)) \mid \mu_R(x, y) = \min[\mu_A(x), \mu_B(y)] \text{ or } \mu_R(x, y) = \mu_A(x) \bullet \mu_B(y)\}$$

3) Composition of crisp relations R and S

$S\ o\ R = \{(x, z) \mid (x, y) \in R, (y, z) \in S\}$ *where* $R \subseteq A \times B$, $S \subseteq B \times C$, *and* $S\ o\ R \subseteq A \times C$

4) Composition of fuzzy relations R and S

$$SR = S\ o\ R = \{((x, z), \mu_{SR}(x, z))\}$$

where $\mu_{SR}(x, z) = \max_y \min[\mu_R(x, y), \mu_S(y, z)]$

Fuzzy Rules and Fuzzy Inference

- R: If *x is A then y is B.*,
- which is sometimes abbreviated as
- R: A $\rightarrow$ B
- The expression describes a relation between two variables *x and y. This suggests that a fuzzy rule can be defined as a binary relation* R *on the product space X × Y.*

Fuzzy Implication

A fuzzy rule can be represented by a fuzzy relation
$R = A \rightarrow B$

- R can be viewed as a fuzzy set with a two-dimensional membership function
- $\mu_R(x, y) = f(\mu_A(x), \mu_B(y))$

  where the function f, called the fuzzy implication function, performs the task of transforming the membership degrees of x in A and y in B into those of (x, y) in $A \times B$.
- f is a min operator [Mamdani] and product operator [Larsen]

Expansion of Fuzzy Set

- *Type-n Fuzzy Set*
  - The value of membership degree might include uncertainty. If the value of membership function is given by a fuzzy set, it is a **type-2** fuzzy set.
  - This concept can be extended up to Type-*n* fuzzy set.

Type-n Fuzzy Set

- Fuzzy sets of type 2: $A : X \to \mathcal{F}([0,1]),$
- $\mathcal{F}([0,1])$: the set of all ordinary fuzzy sets that can be defined with the universal set [0,1].
- $\mathcal{F}([0,1])$ is also called a fuzzy power set of [0,1].

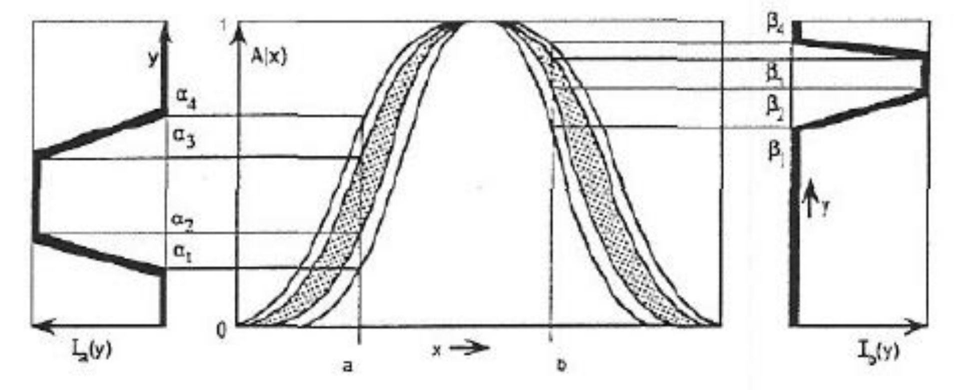
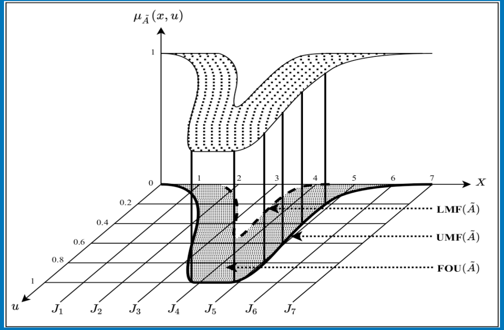


Figure: Fuzzy Set of Type 2

Interval Type 2 Fuzzy Sets (IT2 FS)

- IT2FS can model uncertainty and imprecision in a better way.
- Two type of uncertainty in language:
  - **Intralevel Uncertainty**
  - **Interlevel Uncertainty**
- Type1 FS can only model Intra level uncertainty.

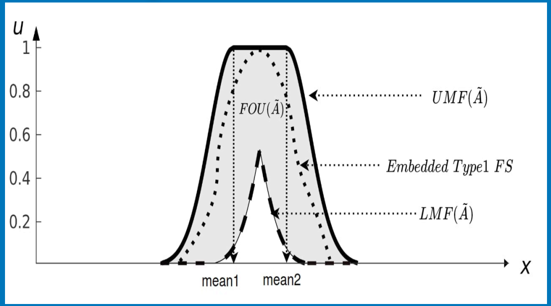
IT2 FS



- As defined by Mendel [1]:

$$\tilde{A} = \{((x,u),\ \mu_{\tilde{A}}(x,u) = 1) \mid \forall x \in X,\ \forall u \in J_x \subseteq [0,1]\}$$

- ***X*** is the primary variable and ***u*** is the primary membership variable.
- $\mu_{\tilde{A}}(x,u)$ represents a secondary membership function. It represents the weight function over the primary membership function defined in the **(*X*; *u*)** plane. For IT2FS secondary membership is unity hence third dimension is hidden for simplicity.

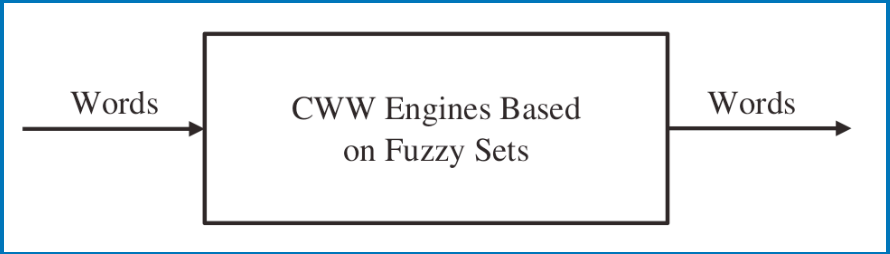


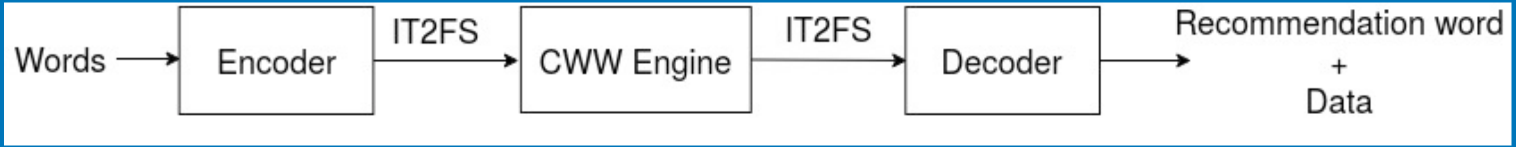
[1] Mendel, J.M., John, R.I., Liu, F.: Interval type-2 fuzzy logic systems made simple. IEEE transactions on fuzzy systems 14(6), 808{821 (2006)

Introduction to Computing With Words

- The information given through humans are mostly words.
- Number of words carries uncertainty and imprecision.
- Let’s take an example in case of Dialog System:
  ‘I want to buy a big house’.
- Here the understanding of big is changing from person to person (and also the country (s)he is residing).

Computing with Words (CWW) [2]





[2] Zadeh, Lotfi A. "From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions." IEEE Transactions on circuits and systems I: fundamental theory and applications 46.1 (1999): 105-119.

Application Description

- Objective: Present a framework of a system that suggest the suitability of restaurant depending upon person's subjective importance given to selection criteria [4].

- Computing with word (CWW) task is performed on the restaurant application. Four categories of restaurants were taken under consideration:
  - (R1) Posh Restaurant
  - (R2) Mall Restaurants
  - (R3) Normal Restaurants
  - (R4) Road side small canteens('Dhabas')

[4] Mishra, R., Barnwal, S.K., Malviya, S., Singh, V., Singh, P., Singh, S., Tiwary, U.S.: Computing with words through interval type-2 fuzzy sets for decision making environment. In: International Conference on Intelligent Human Computer Interaction, pp. 112{123. Springer (2019)

Selection criteria for restaurant

Selection criteria (or linguistic variables) are:

- Cost
- Time
- Food Quality

Linguistic ratings (or Linguistic values) for selection criteria are:

- Cost: Very High(***VH***), High (***H***), Low(***L***), etc.
- Time: Low(***L***), Medium(***M***), Less High(***LH***), etc.
- Food Quality: Very Good(***VG***), Good(***G***), Fair(***F***), etc.

Linguistic Ratings of restaurants for selection criteria

	Selection Criteria		
Restaurant Category	**Cost (C)**	**Time (T)**	**Food Quality (F)**
(R1) Posh Restaurants	*VH*	*L*	*VG*
(R2) Mall Restaurants	*H*	*M*	*G*
(R3) Normal Restaurants	*LH*	*LH*	*F*
(R4) Road side small canteens('Dhabas')	*VL*	*L*	*B*

Linguistic weights of Person for Selection criteria

Linguistic weights (or Linguistic values) of person for selection criteria are:

Cost: Unimportant (***U***), More Less Unimportant(***MLU***), Very Important (***VI***), etc.

Time: Unimportant (***U***), More Less Unimportant(***MLU***), Very Important (***VI***), etc.

Food Quality: Unimportant (***U***), More Less Unimportant(***MLU***), Very Important (***VI***), etc.

Linguistic weights of persons for selection criteria

	Selection Criteria		
Person	**C (Cost)**	**T (Time)**	**F (Food Quality)**
Rich Person	*U*	*VI*	*VI*
Middle Person	*MLI*	*MLU*	*MLI*
Poor Person	*VI*	*U*	*U*

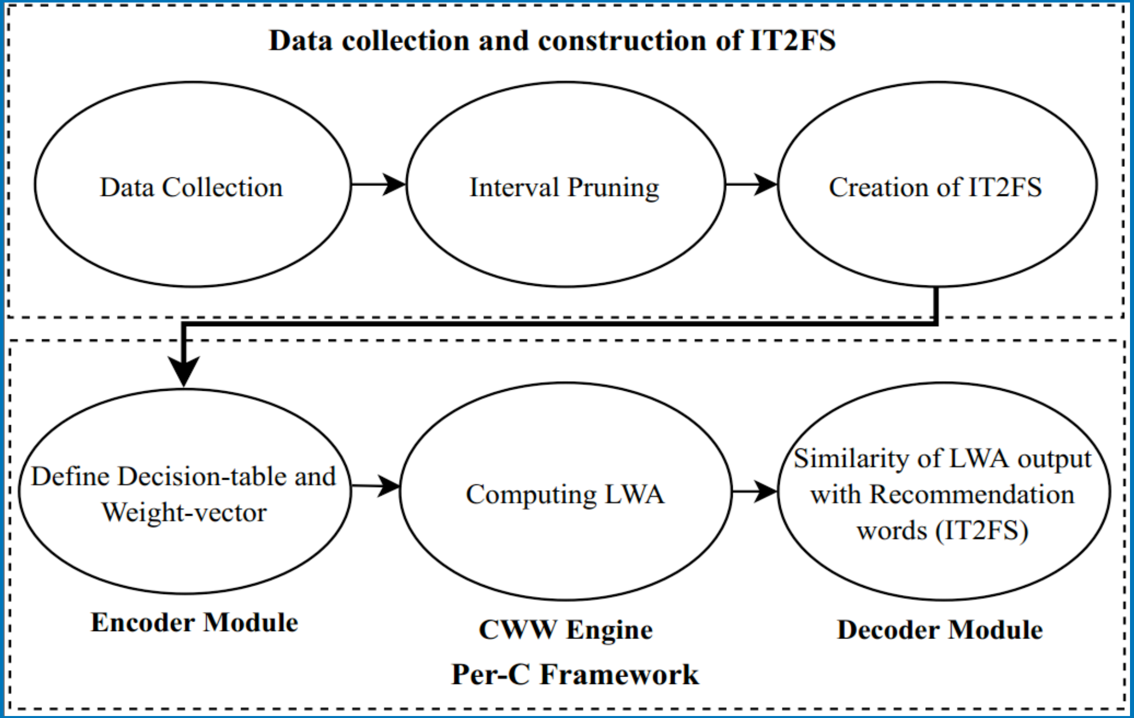
Recommended Words

The list of recommended words are as follows:

Output Words :

- *'Very Low',*
- *'Low',*
- *'Little Low',*
- *'Medium',*
- *'Little High',*
- *'High',*
- *'Extreme'*

Methodology



Vocabulary Construction

Step 1: Data Collection

- It include the collection of interval endpoints of linguistic terms.
- Linguistic terms chosen for corresponding linguistic variable as follows:
  - **Cost**: 'Very Less (VL)', 'Less (L)', 'Fair (F)', 'Little High (LH)', 'High (H)', 'Very High (VH)'
  - **Time**: 'Very Low (VL)', 'Low (L)', 'Little Low (LL)', 'Medium (M)', 'Little High (LH)', 'High (H)', 'Very High (VH)' .
  - **Quality**: 'Bad (B)', 'Somewhat bad (SB)', 'Fair (F)', 'Good (G)', 'Very Good (VG)'.
- The range for interval chosen in between (0-10).
- Data interval were collected from 300 randomly selected people from age group 15 to 60.
- Example: i^{th} person gives the range for 'Less' as [1.5, 3.8]. $[a^{(i)}, b^{(i)}]$

Step 2: Interval Pruning

It involves steps as follows [7]:

Bad Data removal:

$$\left.\begin{aligned} a^{(i)} &\in [0, 10] \\ b^{(i)} &\in [0, 10] \\ b^{(i)} &\geq a^{(i)} \end{aligned}\right\}, \qquad i = 1, \ldots, n$$

Outlier Processing:

$$a^{(i)} \in [Q_a(0.25) - 1.5IQR_a, Q_a(0.75) + 1.5IQR_a]$$
$$b^{(i)} \in [Q_b(0.25) - 1.5IQR_b, Q_b(0.75) + 1.5IQR_b]$$
$$L^{(i)} \in [Q_L(0.25) - 1.5IQR_L, Q_L(0.75) + 1.5IQR_L]$$

- Here $a^{(i)}$ is the lower limit of the range given by i^{th} person , $b^{(i)}$ is upper limit of the range and

$$L^{(i)} = b^{(i)} - a^{(i)}$$

[7] Liu, F., and Mendel, J. M.:Encoding words into interval type-2 fuzzy sets using an interval approach. IEEE transactions on fuzzy systems 16(6),1503-1521 (2008)

Outlier Processing

The Box and Whisker method is applied to remove the outlier data intervals. The method first involves the computation of Lower Quartile (i.e. $Q_*(0.25)$), Upper Quartile (i.e. $Q_*(0.75)$), Inter- Quartile Range (IQR_*) and then the endpoints of the intervals should lie within $[Q_*(0.25)-1.5IQR_*, Q_*(0.75)+1.5IQR_*]$. The check is also applied to interval length ($L^{(i)} = b^{(i)} - a^{(i)}$). Only those intervals are accepted that qualifies all the conditions defined in Equation in the previous slide.

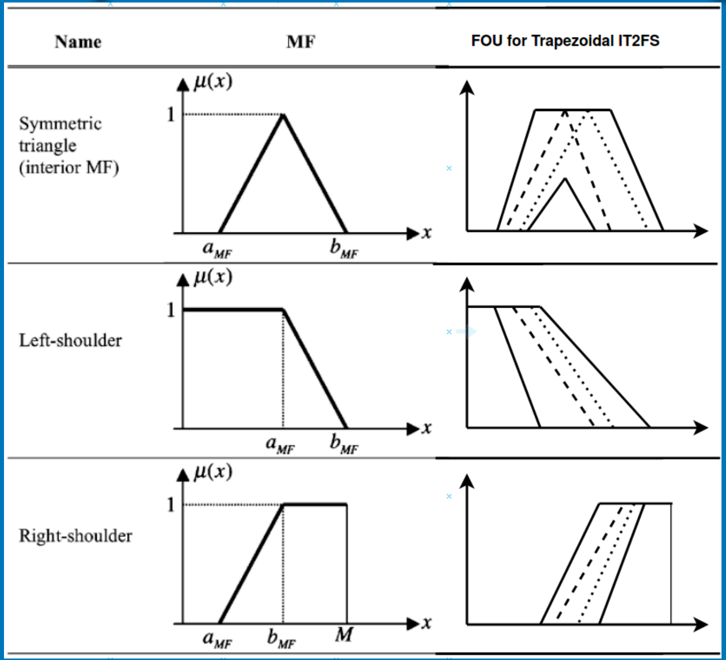
Step 3: Constructing the IT2FS

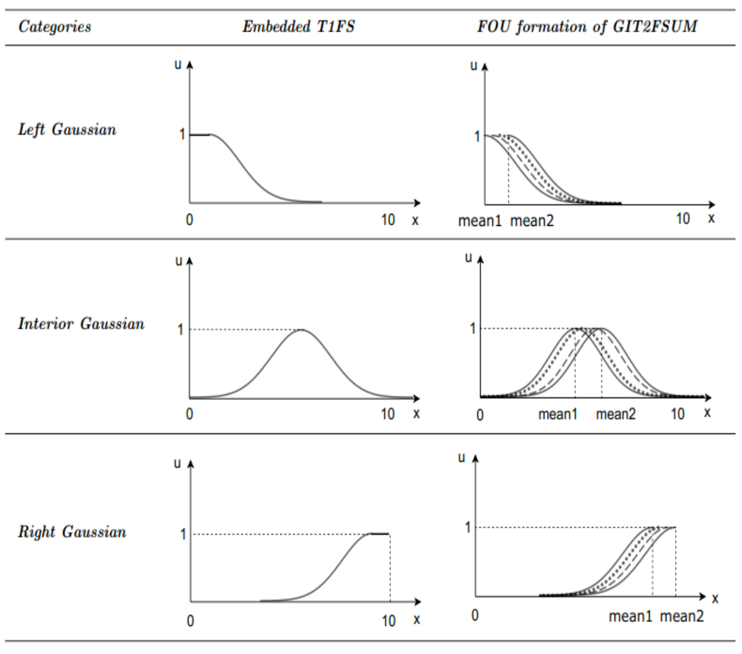
It involves following substeps:

- **Choose Probability distribution and Embedded T1FS model:** Probability distribution chosen is uniform distribution when no information is given.
  - **Embedded Interior T1FS**
  - **Embedded Left T1FS**
  - **Embedded Right T1FS**
- **Mapping the parameters of distribution and Embedded T1FS :** It involves estimation of parameters such as : mean and standard deviation in terms of the data interval parameter.
- **Classification of each interval into Embedded T1FS classes followed by aggregation of all the Embedded T1FS**

Gaussian IT2FS with uncertain mean formed using Gaussian Embedded T1FS

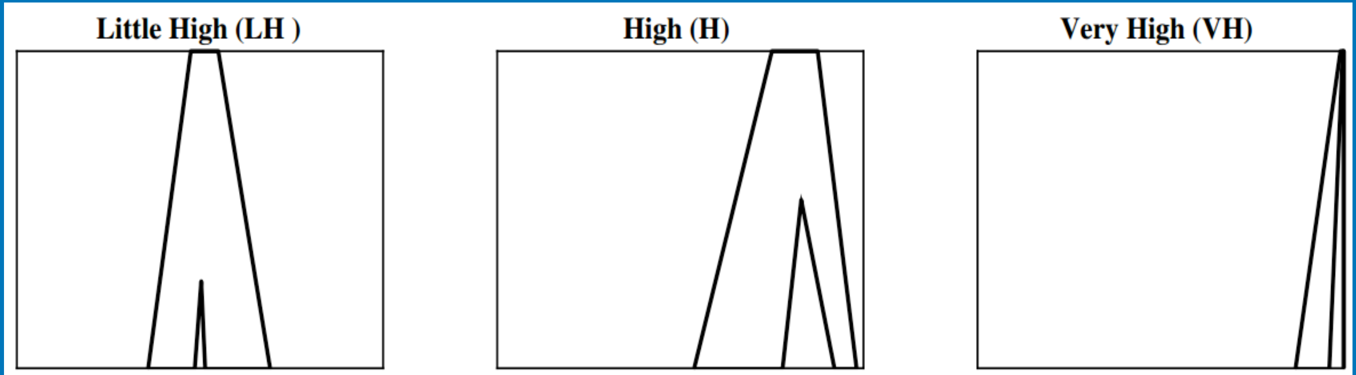
Trapezoidal IT2FS formed using Triangular Embedded T1FS





Resultant IT2FS Vocabulary

Some of the Trapezoidal IT2FS of the Vocabulary are represented



Step 4: Define the Selection Criteria [Encoder Module]

	Selection Criteria		
Restaurant Category	Cost (C)	Time (T)	Food Quality (F)
(R1) Posh Restaurants	*VH*	*L*	*VG*
(R2) Mall Restaurants	*H*	*M*	*G*
(R3) Normal Restaurants	*LH*	*LH*	*F*
(R4) Road side small canteens('Dhabas')	*VL*	*L*	*B*

	Selection Criteria		
Person	C (Cost)	T (Time)	F (Food Quality)
Rich Person	*U*	*VI*	*VI*
Middle Person	*MLI*	*MLU*	*MLI*
Poor Person	*VI*	*U*	*U*

Step 5: Computing Linguistic Weighted Average (LWA) [5] [CWW Engine]

$$\tilde{Y}_{LWA}(A_i) = \frac{\sum_{j=1}^{M} \tilde{x}_{ij}\tilde{w}_j}{\sum_{j=1}^{M} \tilde{w}_j} \quad i = 1, \ldots, N$$

- $\tilde{x}_{ij}$ are IT2FSs that represent the linguistic ratings of the selection criteria for the restaurant whereas $\tilde{W}_{ij}$ are the linguistic weights corresponding to selection criteria for the person. [Note: If $\tilde{x}_{ij}$ is negatively connotated (i.e. linguistic rating for ‘cost’, ‘time’) then should be reversed before LWA]
- **N**: represents total number of restaurants type .
- **M**: represents total number of selection criteria for linguistic weights.

[5] Wu, D., Mendel, J.M.: The linguistic weighted average. In: 2006 IEEE International Conference on Fuzzy Systems, pp. 566{573. IEEE (2006)

Step 6: Similarity of Recommended word with LWA Output [Decoder Module]

- The output word suggested is done based on ***Jaccard similarity measure*** *[6]* between IT2FS of output words in the codebook and LWA output.

$$sim(\tilde{Y},\tilde{Z}) = \frac{\sum_{i=1}^{N} min(\bar{\mu}_{\tilde{Y}}(x_i),\bar{\mu}_{\tilde{Z}}(x_i)) + \sum_{i=1}^{N} min(\underline{\mu}_{\tilde{Y}}(x_i),\underline{\mu}_{\tilde{Z}}(x_i))}{\sum_{i=1}^{N} max(\bar{\mu}_{\tilde{Y}}(x_i),\bar{\mu}_{\tilde{Z}}(x_i)) + \sum_{i=1}^{N} max(\underline{\mu}_{\tilde{Y}}(x_i),\underline{\mu}_{\tilde{Z}}(x_i))}$$

[6] *Wu, D., Mendel, J.M.: A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. Information Sciences 179(8), 1169{1192 (2009)*

LWA output of High - Income Person

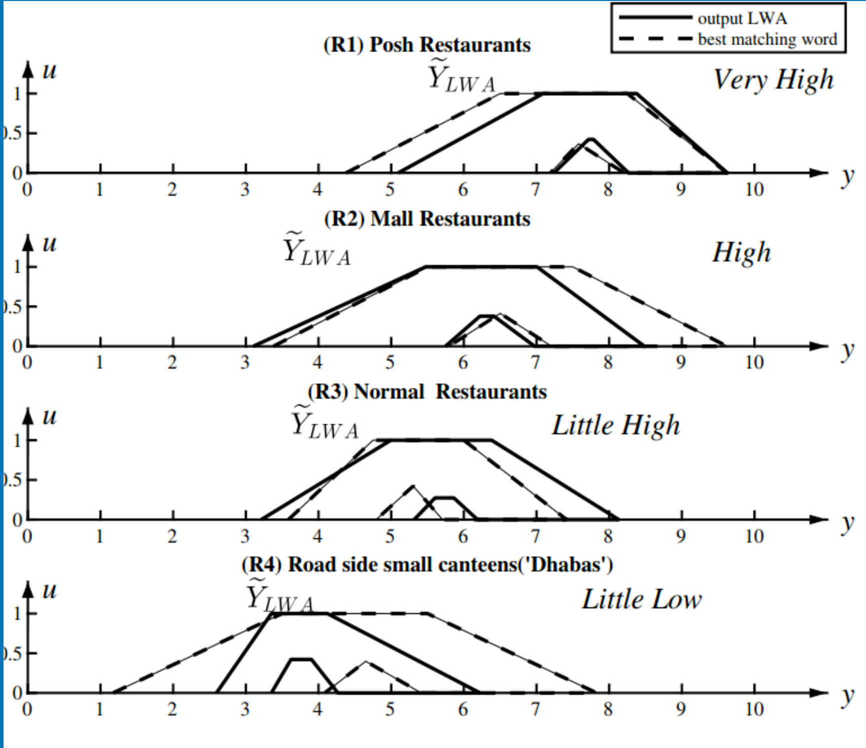


Fig. 7. LWA output for rich person

Indian Institute of Information Technology , Allahabad, India

LWA output of Middle - Income Person

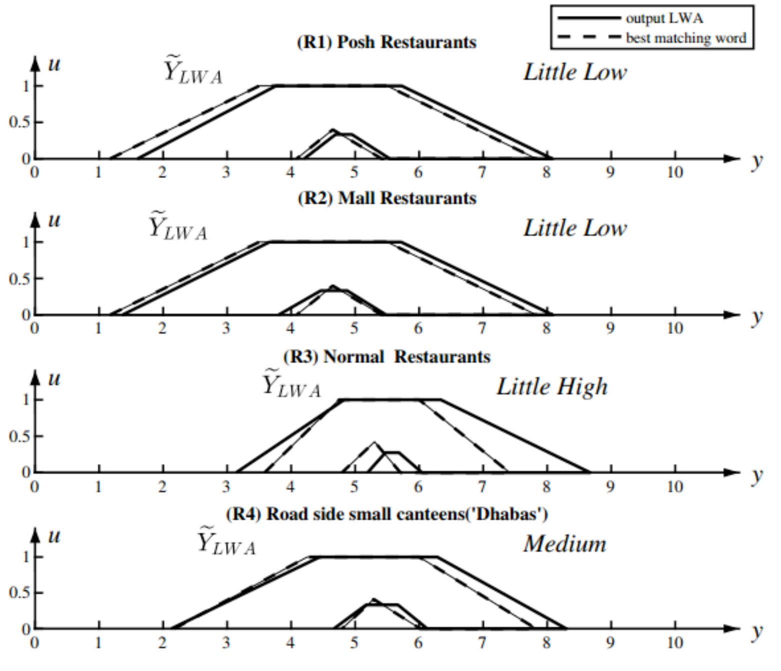


Fig. 8. LWA output for middle person

LWA output of Low -Income Person

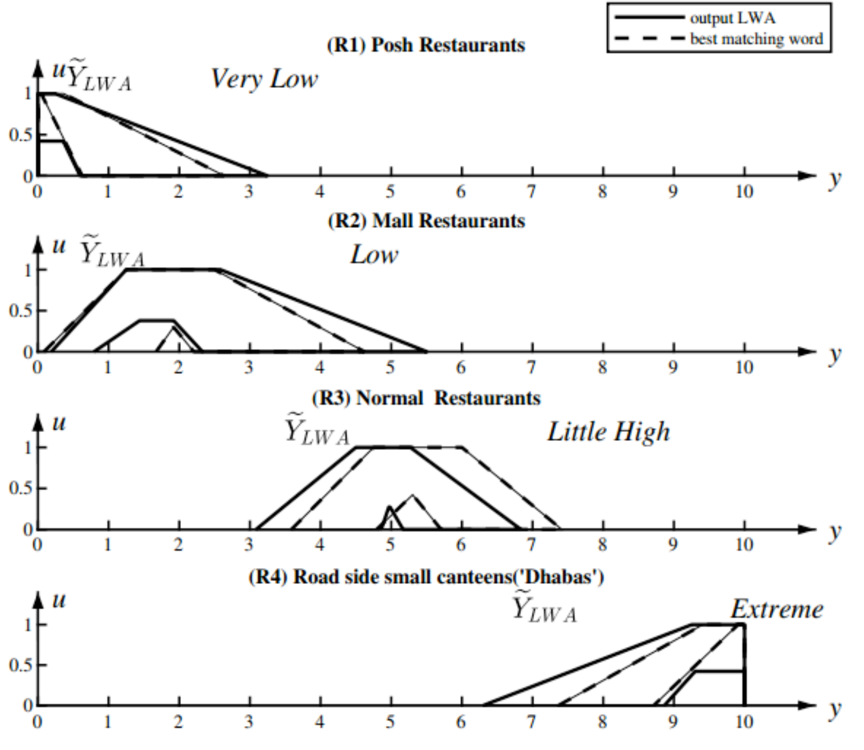


Fig. 9. LWA output for poor person

Indian Institute of Information Technology , Allahabad, India

Remarks

- In the experiment it is clearly demonstrated linguistic ratings and linguistic weights carry both level of uncertainties which are well captured using IT2FSs.
- The similarity between the system output of LWA and output word in the codebook recommends the best suggested word for the restaurant to the person.
- The recommended words are IT2FS that also have uncertainties associated with words similar to the case as shown by humans.

References

[1] Mendel, J.M., John, R.I., Liu, F.: Interval type-2 fuzzy logic systems made simple. IEEE transactions on fuzzy systems 14(6), 808{821 (2006)

[2] Zadeh, Lotfi A. "From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions." IEEE Transactions on circuits and systems I: fundamental theory and applications 46.1 (1999): 105-119.

[3] Mendel, J., Wu, D.: Perceptual computing: Aiding people in making subjective judgments, vol. 13. John Wiley & Sons (2010)

[4] Mishra, R., Barnwal, S.K., Malviya, S., Singh, V., Singh, P., Singh, S., Tiwary, U.S.: Computing with words through interval type-2 fuzzy sets for decision making environment. In: International Conference on Intelligent Human Computer Interaction, pp. 112{123. Springer (2019)

[5] Wu, D., Mendel, J.M.: The linguistic weighted average. In: 2006 IEEE International Conference on Fuzzy Systems, pp. 566{573. IEEE (2006)

[6] Wu, D., Mendel, J.M.: A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. Information Sciences 179(8), 1169{1192 (2009)

[7] Liu, F., and Mendel, J. M.:Encoding words into interval type-2 fuzzy sets using an interval approach. IEEE transactions on fuzzy systems 16(6),1503-1521 (2008)

[8] Wu D., and Mendel J.M.:The linguistic weighted average. In: Proc.FUZZ-IEEE, pp. 566{57. IEEE, Vancouver, BC, Canada(2006).

[9] Wu D. and Mendel J.M.:Aggregation using the linguistic weighted average and interval type-2 fuzzy sets. IEEE Trans. Fuzzy Syst. 15(4), 1145{1161 (2007)

[10] Zadeh, Lotfi A.:Fuzzy logic= computing with words, Computing with Words in Information/Intelligent Systems 1. 3-23, Physica, Heidelberg, (1999).

[11] Mendel, J.M., John, R.I., & Liu, F.: Interval type-2 fuzzy logic systems made simple. IEEE transactions on fuzzy systems, 14(6), 808-821 (2006)