Universidad de Granada

SECABA LAB

Quality Evaluation & Information Retrieval

FUZZY LINGUISTIC MODELLING

Enrique Herrera-Viedma

MALAGA, 30 de mayo de 2017

Universidad de Granada

SECABA LAB
Quality Evaluation & Information Retrieval

Outline

- Liguistic variables
- Granularity of uncertainty
- Elements of a linguistic approach
- Models of linguistic preference modelling
- New trends in linguistic preference modelling

Linguistic Variables

- Definition: Variables whose values are not numbers but words or sentences in a natural or artificial language.

- How defining the linguistic values. Each linguistic value is characterized by:
  - *A syntactic value or label*: It is a word or sentence belonging to a linguistic term set .
  - *A semantic value* or *meaning*: It is a fuzzy subset defined in the universe of discourse (problem domain).

Linguistic Variables

- <u>Important aspect:</u>

For defining a term set we have to establish the **Granularity of uncertainty**, i.e., the cardinality of the linguistic term set used to express the information.

Granularity of Uncertainty

- Properties of Granularity of Uncertainty:
  1. It must be small enough so as not to impose useless precision on the users, and
  2. It must be rich enough in order to allow a discrimination of the assessments in a limited number of degrees.
- Typical values of cardinality: They are odd values, such as 7 or 9, with an upper limit of granularity of 11 or no more than 13, where the mid term represents an assessment of "approximately 0.5".

Granularity of Uncertainty

- These classical cardinality values seems to fall in line with Miller's observation about the fact that human beings can reasonably manage to bear in mind seven or so items.

Elements of an Approach

- To apply the linguistic preference modeling in decision making we have to define two elements:
  1. **The linguistic representation model**: It is defined by choosing
     - *the linguistic term set* used to express linguistic assessments and
     - *its semantics.*
  2. **The linguistic computational model**: It is defined by designing *aggregation operators* to develop the processes of Computing with Words that allow to solve GDMP.

Models

- We identify three foundation models of linguistic preference modeling:

1 **The Approximate Linguistic Preference Modeling**

2 **The Ordinal Linguistic Preference Modeling**

3 **The 2-Tuple Linguistic Preference Modeling**

Approximate Linguistic Preference Modeling

1. **Approximate representation model:**
   – *Linguistic term set*: It is defined by means of a context-free grammar
   – *Semantics*: It is represented by
     - fuzzy numbers described by membership functions based on parameters and defined in the [0,1] interval, and
     - by a semantic rule.

2 **Approximate computational model:** It uses fuzzy arithmetic based on the Extension Principle to make linguistic computations and the membership functions associated with the linguistic terms.

Universidad de Granada

SECABA LAB Quality Evaluation & Information Retrieval

Approximate Linguistic Preference Modeling-DRAWBACKS

1. **The complexity of defining a grammar** to represent the linguistic terms.

2 **The complexity of determining the parameters of fuzzy sets** associated to the terms according to all user's attitudes. Example: differents concepts of distribution.

3 In the computational model appears the problem of **linguistic approximation**, i.e., the computation results are fuzzy sets which do not correspond to any label in the original linguistic term set.

Ordinal Linguistic Preference Modeling

1 **Ordinal representation model:**

- *Linguistic term set:* It is defined by means of an ordered structure of linguistic terms distributed on a scale, e.g. the [0,1] internal, with an odd cardinal and the mid term representing an assessment of "approximately 0.5" and with the rest of the terms being placed symmetrically around it.
- *Semantics*: It is established from the ordered structure of the term set by considering that each linguistic term for the label pair (s_i,s_{g-i}) ($g+1$ is the cardinality of linguistic term set) is equally informative.

2 **Ordinal computational model:** It is based on the symbolic computation and acts by direct computation on labels. It uses the index of labels to compute.

Defining the Ordinal Representation Model (1)

1. $S=\{s_i\}, i \in H=\{0,..,g\}$ a finite label set with odd cardinal.
2. Limit of granularity = 11 or 13.
3. All linguistic terms are primary terms, i.e., we do not use a context-free grammar for generating the terms.
4. $s_{g/2}$ represents an assessment of "approximately 0.5".
5. The rest of the terms are placed symmetrically around $s_{g/2}$
6. The terms are distributed symmetrically on a scale on which a total order is defined: $s_i \geq s_j$ if $i \geq j$.

Defining the Ordinal Representation Model (2)

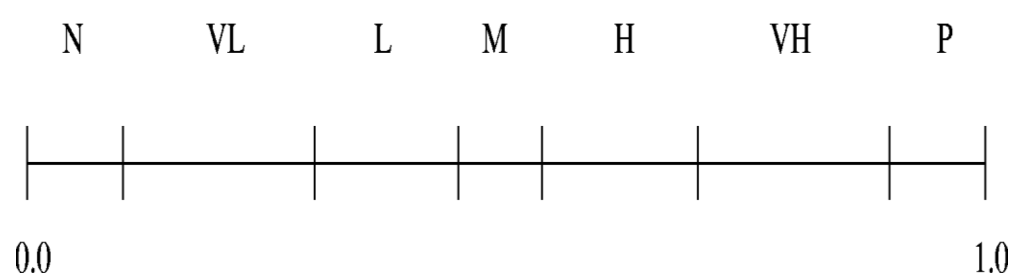
Semantics:

- It is defined by considering that each linguistic term for the pair (s_i, s_{g-i}) is equally informative. That is we assume:

  Negation operator: $\boldsymbol{NEG(s_i) = s_{g-i}}$

  – Example 1. A set of seven labels without membership functions associated:

  S={Perfect,Very-High,High,Medium,Low,Very-Low,Null}

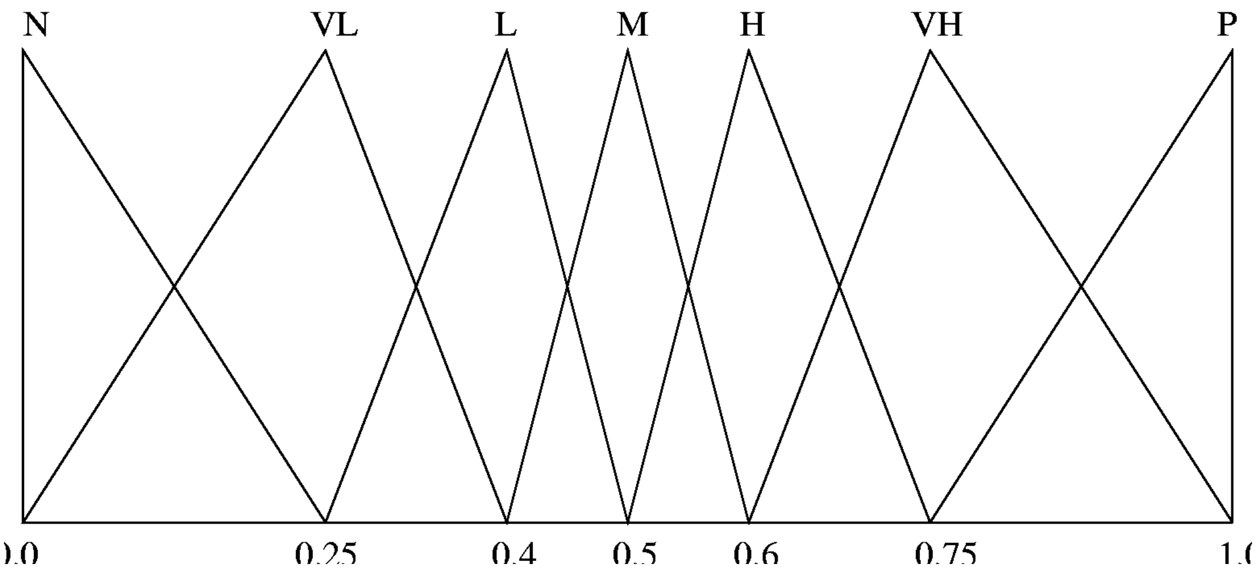


Defining the Ordinal Representation Model (2)

Semantics:

– Sometimes, we can assign to each label a fuzzy number represented by a trapezoidal or triangular membership functions defined on the problem domain, *(a_i, b_i, α_i, β_i)*.

– Example 2:



Defining the Ordinal Computational Model(1)

1 **Comparison operators:**

- Maximum operator: MAX(s_i ,s_j)= s_i if i≥j
- Minimum operator: MIN(s_i ,s_j)= s_i if j≥i

2 **Aggregation operators** are based on **Symbolic Computation:**

- Use the order of linguistic terms in the ordered structure of linguistic terms (index of labels).
- The membership functions associated to the labels are not used.
- Sometimes, a simple approximation process is required:

Basic rounding operation

An Aggregation Operator of Ordinal Values

- **Linguistic Ordered Weighted Averaging (LOWA)**

$$\phi(a_1,\ldots, a_m) = W \bullet B^t = C^m\{w_k, b_k, k= 1,\ldots,m\} =$$
$$w_1 \odot b_1 \oplus (1-w_1) \odot C^{m-1}\{\beta_h, b_h, h = 2,\ldots,m\}.$$

- It is based on the OWA operator (Yager):
  - $A= (a_1,\ldots, a_m)$ set of ordinal values to aggregate,
  - B is the set A ordered
  - $W=(w_1,\ldots,w_m)$ weigthing vector, $\Sigma_k w_k =1, w_k \in [0,1]$.
- It is defined recursively in function of the convex combination of the labels C^m:
  - $\beta_h = w_h / \Sigma^m{}_2 w_k$, h=2,...,m, weights recalculated in the recursive process.
  - $C^2\{w_i, b_i, i=1,2\}=w_1 \odot s_j \oplus (1-w_1) \odot s_i= s_k$,
    $k = \min\{g, i+\mathrm{round}(w_1 \bullet (j-i))\}$.

4. Models of Linguistic Preference Modeling: Ordinal Linguistic Preference Modeling-Drawbacks(1)

1.First Drawback:

Loss of Information in the aggregation process as a consequence of rounding operation

- Problem: In the aggregation of different assessments we obtain the same result. Example, using W=[0.4,0.6] :
  - $\phi(VH,M)=C^2\{VH,M\}=H$, given that $4=3+min(6,round(5\text{-}3)0.4)$
  - $\phi(VH,H)= C^2\{VH,H\}=H$, given that $4=4+min(6,round(5\text{-}4)0.4)$

– Solution: **2-Tuple Linguistic Representation Model**

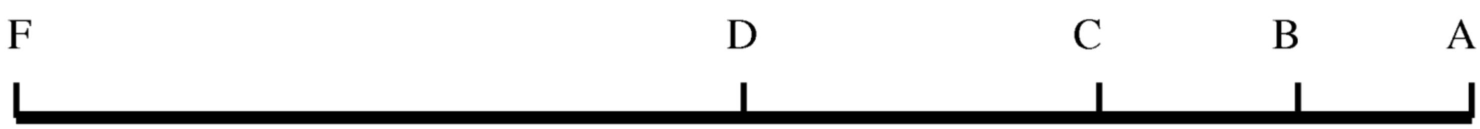
4. Models of Linguistic Preference Modeling: Ordinal Linguistic Preference Modeling-Drawbacks(2)

2 Second Drawback:

The ordinal linguistic approach is not applicable to problems with non-symmetrical linguistic domains, i.e.,

$$NEG(s_i) \neq s_{g-i} \, .$$

Example: Grading system in education.



- Solution: Unbalanced Linguistic Preference Modeling

4. Models of Linguistic Preference Modeling: 2-Tuple Linguistic Preference Modeling

1 **It is based on the ordinal model**

2 **Differences:**

- 2-Tuple representation model.

  The linguistic information is represented by means of a pair of values, (s,α) where s is a linguistic label and α is a numerical value that represents the value of the Symbolic Translation.

- 2-Tuple computational model:

  It is an extension of the ordinal computational model that avoids the loss of information by means of the 2-tuple linguistic representation model

4. Models of Linguistic Preference Modeling: 2-Tuple Representation Model

- $S=\{s_i\}, i \in H=\{0,..,g\}$ a finite label set
- Let $\beta \in [0,g]$ be the result of an aggregation of the indices of a set of labels assessed in a linguistic term set S, i.e., the result of a symbolic aggregation operation.
- The 2-tuple (s_i, α) that expresses the equivalent information to β is obtained by the function Δ:

$$\Delta: [0,g] \rightarrow S \times [-.5,.5)$$

$$\Delta(\beta) = (s_i, \alpha)$$

  being $i=round(\beta)$ and α *the* **symbolic translation value** obtained as $\alpha = \beta - i$.

- There is always Δ^{-1} inverse function:

$$\Delta^{-1}: S \times [-.5,.5) \rightarrow [0,g] \; ; \; \Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

4. Models of Linguistic Preference Modeling: Example of 2-Tuple Linguistic Representation

- If we have any label $s_i \in S$ then the associated 2-tuple is $(s_i, 0)$.
  - *Example: VH→(VH,0).*
- If we use the LOWA operator to aggregate two labels (VH,M) with the W=[0.4,0.6] :
  - $\phi(VH,M)=C^2\{VH,M\}=H$, because $4=min\{6,3+round((5\text{-}3)0.4)\}$;
  - however in 2-tuple linguistic representation the result would be:

$$\beta=3.8 \rightarrow (H,-0.2)$$

  - And therefore, if we aggregate *(VH,M)* with 2-tuple linguistic representation we obtain

    $\phi(VH,H)= C^2\{VH,H\}=H$, because $4=min\{6,4+round((5\text{-}4)0.4)\}$

$$\beta=4.4 \rightarrow (H,0.4)$$

4. Models of Linguistic Preference Modeling: 2-Tuple Computational Model

- The comparison of 2-tuples (s_k,α_1) and (s_l,α_2) linguistic is defined as:

1. if $\alpha_1 = \alpha_2$ then (s_k,α_1), (s_l,α_2) represent the same information
2. if $\alpha_1 < \alpha_2$ then (s_k,α_1) is smaller than (s_l,α_2)
3. if $\alpha_1 > \alpha_2$ then (s_k,α_1) is bigger than (s_l,α_2)

4. Models of Linguistic Preference Modeling: 2-Tuple Computational Model

- The Negation of (s_i, α) : $Neg\,(s_i, \alpha) = \Delta(g - \Delta^{-1}(s_i, \alpha))$

- It is easy to define aggregation operators using Δ and Δ^{-1}

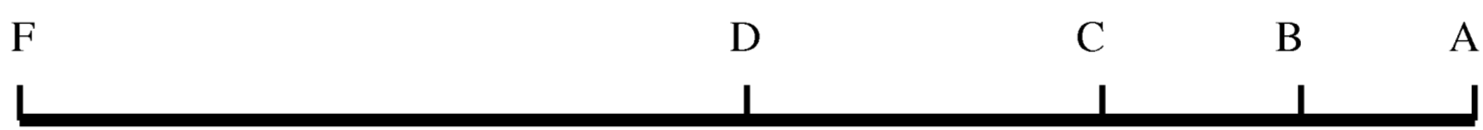
4. Models of Linguistic Preference Modeling: 2-Tuple Preference Modeling-Drawback

- <u>Drawback</u>:

As happens in the ordinal linguistic preference modeling is not applicable to problems with non-symmetrical linguistic domains, i.e.,

$$\boldsymbol{Neg(s_i,\alpha) \neq} \Delta(g-\Delta^{-1}(s_i,\alpha)) \,.$$

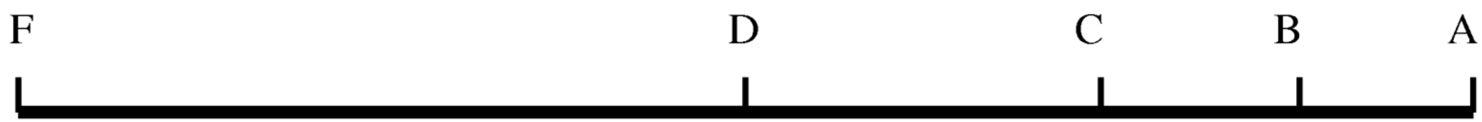
Example: Grading system in education.



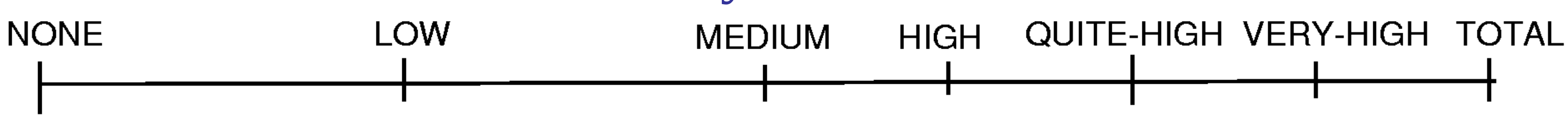
Solution: Unbalanced Linguistic Preference Modeling

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling

- Definition: Unbalanced linguistic term sets are linguistic term sets with different discrimination levels on both sides of mid linguistic term
  - Exmple 1: Grading system in education



  - Example 2: To evaluate the importance and relevance in Information Retrieval Systems



5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling

- Proposal:

  A methodology to manage unbalanced linguistic information which is sensitive to the discourse domain.

- This methodology is based on two elements:

  2-Tuple Linguistic Preference Modeling: To combine the unbalanced linguistic information.

  HIERARCHICAL LINGUISTIC CONTEXTS: To represent the unbalanced linguistic information

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Information-Hierarchical Linguistic Contexts

DEFINING HIERARCHICAL LINGUISTIC CONTEXTS

- A ***Linguistic Hierarchy***, $LH = \cup_t\ l(t,n(t))$ is a set of levels, where each level t represents a linguistic term set $S^{n(t)}$ with different granularity $n(t)$ to the remaining levels.
- In each t the semantics of the linguistic terms is represented by triangular-shaped membership functions, assuming that are symmetrically and uniformly distributed in [0,1] and with an odd value of granularity.
- $t+1$ is a refinement of the previous t, i.e., $n(t+1) > n(t)$.
- The linguistic term set of level $t+1$, $S^{\wedge n(t+1)} = \{s^{n(t+1)}_0,\ldots,s^{n(t+1)}_{n(t+1)-1}\}$ is obtained from its predecessor $S^{n(t)}$ as:

$$l(t,n(t)) \rightarrow l(t+1, 2 \bullet n(t)-1).$$

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Information-Hierarchical Linguistic Contexts

DEFINING HIERARCHICAL LINGUISTIC CONTEXTS(2)

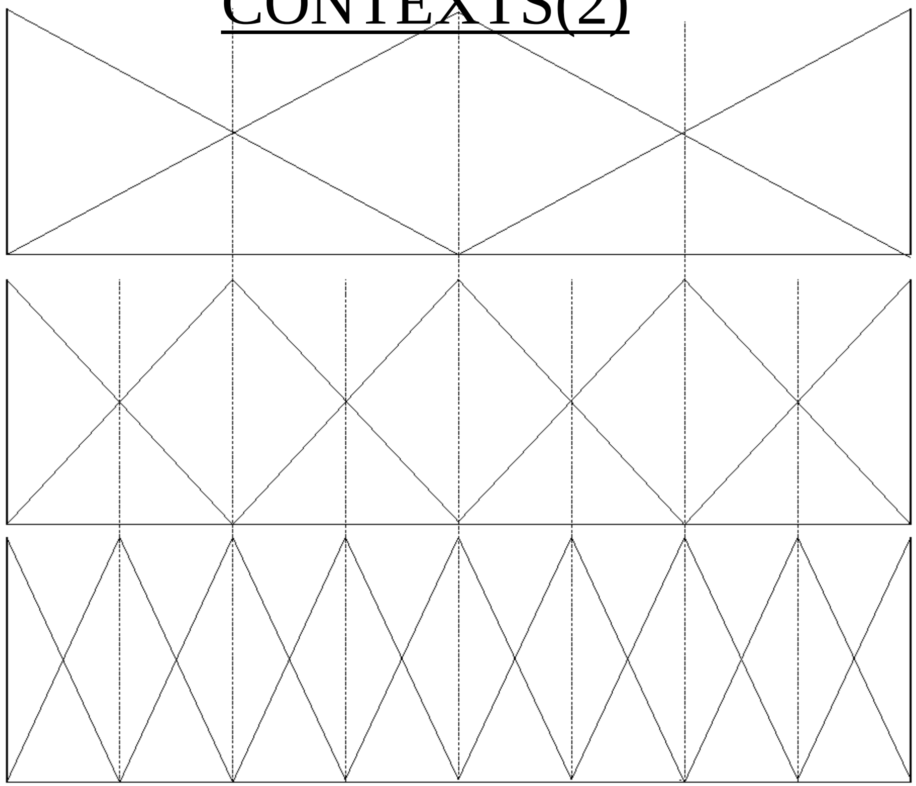
- 

Figure 2: Linguistic Hierarchy of 3, 5 and 9 Labels

DEFINING HIERARCHICAL LINGUISTIC CONTEXTS (3)

- The transformation function from a 2-tuple in level *t* to a 2-tuple in level *t'*:

$$TF_{t'}^{t} : l(t, n(t)) \longrightarrow l(t', n(t'))$$

$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) =$$

$$\Delta_{n(t')}\left(\frac{\Delta_{n(t)}^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t')-1)}{n(t)-1}\right).$$

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling- METHODOLOGY

- To define a methodology to manage unbalanced linguistic information we need:

1 **Unbalanced Representation Model**

2 **Unbalanced Computational Model**

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling- Representation Model

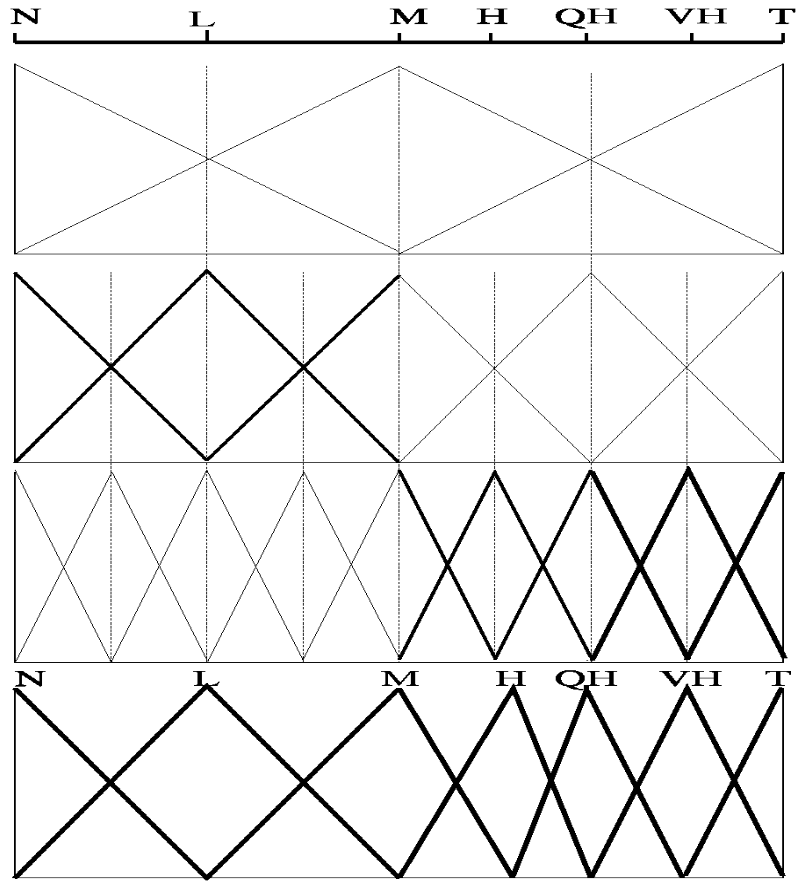
1 Steps to represent the unbalanced linguistic term set S by means of a linguistic hierarchy, *LH*:

1.1 Choose a level t^- with an adequate granularity to represent using the 2-tuple representation model the subset of linguistic terms of S on the left of the mid linguistic term.

1.2 Choose a level t^+ with an adequate granularity to represent using the 2-tuple representation model the subset of linguistic terms of S on the right of the mid linguistic term.

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling- Representation Model

- Graphically:



Representation for an Unbalanced Term Set of 7 Labels

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling- Computational Model

1 Choose a level $t' \in \{t^-, t^+\}$, such that $n(t') = max\{n(t^-), n(t^+)\}$.

2 Define:

- the comparison of two unbalanced linguistic 2-tuples $(s_k^{n(t)}, \alpha_1)$, $t \in \{t^-, t^+\}$, and $(s_l^{n(t)}, \alpha_2)$, $t \in \{t^-, t^+\}$,
- the negation operator of unbalanced linguistic 2-tuple $(s_k^{n(t)}, \alpha_1)$, $t \in \{t^-, t^+\}$,
- aggregation operators of unbalanced linguistic 2-tuples

by using the 2-tuple computational model but acting on the unbalanced linguistic values transformed by means of $TF^{t}_{t'}$ and $TF^{t'}_{t}$, $t \in \{t^-, t^+\}$, $t \neq t'$.

5. New Trends in Linguistic Preference Modeling: Unbalanced Linguistic Preference Modeling

- <u>FUTURE RESEARCH PROBLEM</u>:

  When we need more levels of LH to represent the unbalanced linguistic information

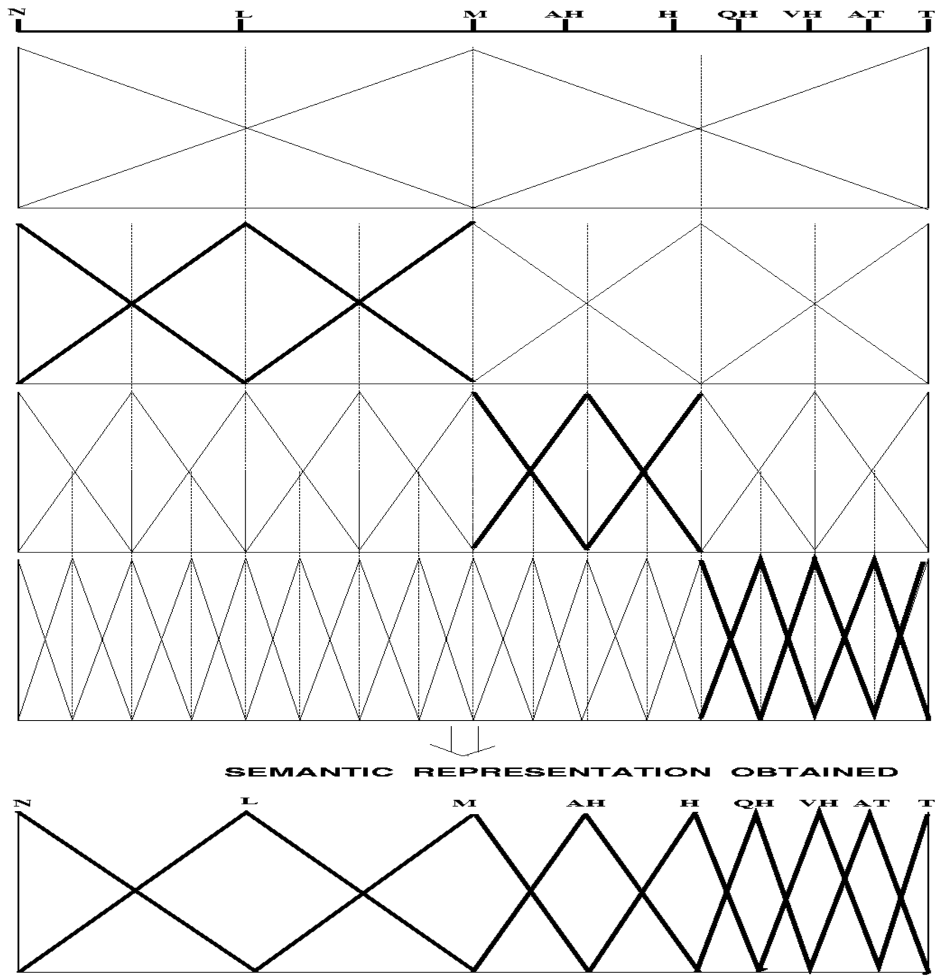


Fig. 16. Representation of $\mathcal{S} = \{N, L, M, AH, H, QH, VH, AT, T\}$ in LH.