

USO DE JERARQUÍAS LINGÜÍSTICAS EXTENDIDAS PARA MEJORAR LA REPRESENTACIÓN DE INFORMACIÓN LINGÜÍSTICA NO BALANCEADA

M. Espinilla¹ L. Martínez¹

¹ Dpto. de Informática, Universidad de Jaén, 23071 - Jaén, España, {mestevez,martin}@ujaen.es

Resumen

La utilización de variables lingüísticas para modelar la información en problemas implica la realización de procesos de computación con palabras. Existen modelos computacionales para operar con precisión en un conjunto de términos lingüísticos simétricos y uniformemente distribuidos. Sin embargo, podemos encontrar problemas que necesitan de conjuntos de términos lingüísticos no balanceados para representar la información que manejan. Recientemente, se ha presentado una metodología para representar conjuntos de términos lingüísticos no balanceados y operar sobre ellos sin pérdida de información. Sin embargo, esta metodología tiene distintas limitaciones de representación. En esta contribución, revisamos éste método para representar conjuntos de términos lingüísticos no balanceados y presentamos una nueva metodología que supera sus limitaciones de representación.

1. INTRODUCCIÓN

La necesidad de modelar información vaga e imprecisa en problemas reales de Evaluación Sensorial, Toma de Decisiones, Consenso, Sistemas de Recuperación de Información, etc., ha producido que surjan diferentes modelos y técnicas para su tratamiento. En esta contribución nos centramos en el uso del Enfoque Lingüístico Difuso [9] para modelar dicha información mediante el uso de variables lingüísticas que implican la realización de procesos de computacionales con palabras.

En muchos problemas que trabajan con información lingüística se ha modelado dicha información con conjuntos de etiquetas simétricas y uniformemente distribuidas, existiendo modelos computacionales precisos

para este tipo de información [5]. Sin embargo, en distintos problemas [1, 7] aparece la necesidad de trabajar con información lingüística no balanceada. En [3, 4] se presentó una metodología para representar conjuntos de términos lingüísticos no balanceados (CTLNB) y operar de forma precisa sobre los mismos. Para ello, esta metodología hace uso de las Jerarquías Lingüísticas [6] que son una estructura lingüística que permite operar con variables lingüísticas valoradas en conjuntos lingüísticos con distinta granularidad, imponiendo un conjunto de reglas. En [2, 8] se presentó una mejora sobre las Jerarquías Lingüísticas denominada Jerarquías Lingüísticas Extendidas que relaja las reglas de construcción permitiendo una mayor flexibilidad de representación.

En esta contribución, utilizamos las Jerarquías Lingüísticas Extendidas para representar CTLNB y así superar las restricciones de representación debidas al uso de las Jerarquías Lingüísticas. Igualmente, esta nueva propuesta podrá operar con el CTLNB de forma precisa, aunque el modelo computacional no es objeto de estudio en este trabajo. Debido a la limitación de espacio, hacemos una propuesta inicial que muestra las ideas básicas de la nueva metodología y que será extendida en el futuro.

Este trabajo se estructura como sigue: en la Sección 2 revisamos los conceptos básicos necesarios para nuestra propuesta. En la Sección 3, vemos brevemente la metodología actual para representar CTLNB. En la Sección 4, presentamos una nueva metodología para representar CTLNB basada en el uso de las jerarquías lingüísticas extendidas. Finalmente, indicaremos una serie de conclusiones.

2. INFORMACIÓN LINGÜÍSTICA

Hemos indicado el uso de las Jerarquías Lingüísticas y de las Jerarquías Lingüísticas Extendidas como base de las metodologías de representación de CTLNB, por lo que en esta sección, hacemos una breve revisión de

las mismas.

2.1. JERARQUÍAS LINGÜÍSTICAS

Las Jerarquías Lingüísticas se presentaron como una estructura para realizar procesos computacionales con información lingüística multi-granular [6].

Una Jerarquía Lingüística se compone de un conjunto de niveles, donde cada nivel es un conjunto de términos lingüísticos simétrico y uniformemente distribuido con distinta granularidad al resto de niveles de su jerarquía. Cada nivel de la jerarquía es notado como, $l(t, n(t))$, siendo t un número que indica el nivel de la jerarquía y $n(t)$ la granularidad del conjunto lingüístico del nivel t .

Los niveles dentro de una jerarquía están ordenados de acuerdo a su granularidad, es decir, para dos niveles sucesivos t y $t + 1$, $n(t + 1) > n(t)$. Una Jerarquía Lingüística, LH, se define como la unión de todos los niveles t : $LH = \bigcup_t l(t, n(t))$ donde el conjunto de etiquetas del nivel t es representado como $S^{n(t)}$.

La construcción de una Jerarquía Lingüística debe satisfacer las siguientes reglas básicas [6]:

- *Regla 1.* Preservar todos los puntos modales previos de las funciones de pertenencia de cada uno de los términos lingüísticos de cada nivel con respecto a los del nivel siguiente.
- *Regla 2.* Realizar las transiciones entre dos niveles consecutivos suaves. El propósito es construir un nuevo conjunto de términos lingüísticos, $S^{n(t+1)}$, de modo que se añada un nuevo término lingüístico entre cada pareja de términos pertenecientes al conjunto de términos del nivel anterior t . Para realizar esta inserción de nuevos términos, se reduce el soporte de las etiquetas lingüísticas para dejar hueco entre ellas para la nueva etiqueta.

Estas reglas imponen que la granularidad de un conjunto de términos de un nivel $t + 1$ se obtiene de su predecesor t con la siguiente expresión: $l(t, n(t)) \rightarrow l(t + 1, 2 \cdot n(t) - 1)$, lo que implica una limitación en los conjuntos de términos lingüísticos que pueden pertenecer a una LH como se puede observar gráficamente en las Figuras 1 y 2.

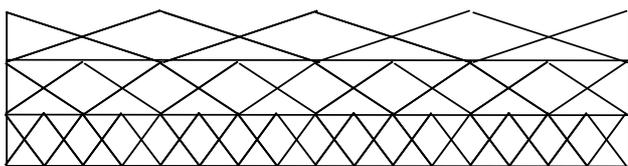


Figura 1: Jerarquía Lingüística de 5, 9 y 17 etiquetas.

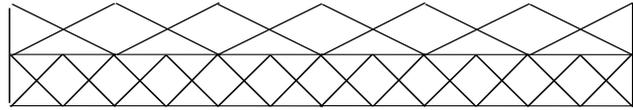


Figura 2: Jerarquía Lingüística de 7 y 13 etiquetas.

Junto a las Jerarquías Lingüísticas se ha desarrollado un modelo computacional que permite operar sin pérdida de información [6].

2.2. JERARQUÍAS LINGÜÍSTICAS EXTENDIDAS

Las Jerarquías Lingüísticas Extendidas se presentaron en [2, 8] para eliminar la limitación de los conjuntos de etiquetas que pueden pertenecer a una Jerarquía Lingüística y seguir operando de forma precisa sobre ellos. Para ello, las Jerarquías Lingüísticas Extendidas relajaron las reglas de construcción de una Jerarquía Lingüística, tal como sigue:

- *Regla 1 Extendida:* Incluir en la Jerarquía Lingüística Extendida, ELH, el conjunto de niveles necesarios ordenados por su granularidad, $l(t, n(t))$, $t = 1, \dots, m$.
- *Regla 2 Extendida:* Añadir un último nivel, $l(t', n(t'))$ con $t' = m + 1$ con la granularidad $n(t') = m.c.m.(n(t) - 1) + 1$, $t = 1, \dots, m$, siendo *m.c.m.* el *mínimo común múltiplo*. Este nivel se utilizará únicamente para realizar los procesos de computación.

Por tanto, la construcción de una Jerarquía Lingüística Extendida implica la inclusión de los m conjuntos de términos lingüísticos a utilizar y la adición final de un nuevo nivel $l(t', n(t'))$, con $t' = m + 1$ y con la granularidad mencionada anteriormente. En este último nivel se llevan a cabo los procesos computacionales de forma precisa [2, 8]. Una Jerarquía Lingüística Extendida puede incluir niveles con 5 y 7 etiquetas lingüísticas (ver Figura 3), niveles que no podían pertenecer a una misma Jerarquía Lingüística (ver Figuras 1 y 2).

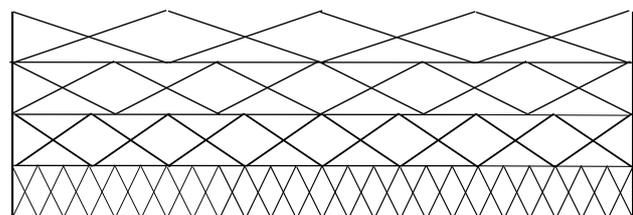


Figura 3: Jerarquía Lingüística Extendida de 5, 7, 9 y 25 etiquetas.

3. REPRESENTACIÓN DE CTLNB. ANTECEDENTES

Para entender nuestra propuesta de representación de CTLNB, en primer lugar haremos una revisión del método de representación presentado en [3, 4].

En el método un CTLNB, S^{nb} , se representa utilizando una Jerarquía Lingüística dada. S^{nb} está caracterizado por una etiqueta central y para representarlo se divide el conjunto en 3 subconjuntos $S^{nb} = S_L \cup S_C \cup S_R$.

- S_L contiene las etiquetas situadas a la izquierda de la etiqueta central. El número de etiquetas del subconjunto es representado como $\#(S_L)$.
- S_C contiene la etiqueta central y,
- S_R contiene las etiquetas situadas a la derecha del término central. El número de etiquetas del subconjunto es representado como $\#(S_R)$.

Básicamente, el método busca en una Jerarquía Lingüística, LH, un nivel con una granularidad igual a la de cada subconjunto, S_L y S_R . Si existe un nivel con dicha granularidad representa el subconjunto de forma uniforme en ese nivel. En el caso de que no exista un nivel adecuado, el subconjunto es representado de forma no uniforme mediante el uso de dos niveles consecutivos, $\{l(t, n(t)), l(t+1, n(t+1))\} \in LH$. De modo que la zona con mayor concentración de etiquetas se representa en el nivel de mayor granularidad, $n(t+1)$, y la zona con una menor concentración de etiquetas en el nivel con menor granularidad, $n(t)$.

A continuación, mostramos brevemente el procedimiento:

- Si S_L o S_R satisface (1) entonces el subconjunto es representado con una distribución uniforme, usando el nivel de la Jerarquía Lingüística que cumple (1).

$$\exists t \in LH, \frac{n(t) - 1}{2} = \#(S_{[L,R]}) \quad (1)$$

- Si S_L o S_R no satisface (1) entonces se buscan dos niveles consecutivos $\{t, t+1\} \in LH$, tal que:

$$\frac{n(t) - 1}{2} < \#(S_{[L,R]}) < \frac{n(t+1) - 1}{2} \quad (2)$$

En este caso, se define la distribución del subconjunto: extrema o central, siendo *extrema* cuando se agrupa un mayor número de etiquetas en el extremo del CTLNB y *centrada* cuando se concentra un mayor número de etiquetas cerca de S_C .

Este método presenta limitaciones debido a que las Jerarquías Lingüísticas están prefijadas, es decir, los conjuntos de términos lingüísticos que pertenecen a LH no pueden modificarse libremente sino que deben de respetar las reglas de construcción. A continuación, mostramos algunas de las limitaciones más evidentes:

1. Cuando existe un subconjunto del CTLNB distribuido de forma uniforme, pero la LH que se proporciona no tiene un nivel con una granularidad adecuada, el método solicita la distribución, centrada o extrema, y representa el subconjunto de forma no uniforme con la distribución señalada.
2. Cuando se quiere representar un subconjunto del CTLNB de forma no uniforme y LH contiene un nivel con una granularidad adecuada para representar ese subconjunto, el método no solicita ninguna distribución y representa el subconjunto de forma uniforme.

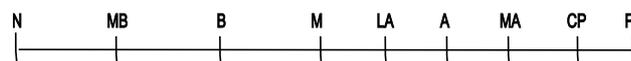


Figura 4: CTLNB de 9 etiquetas.

Por ejemplo, dada la LH de la Figura 1 y el CTLNB mostrado en la Figura 4, el método revisado representaría el CTLNB de la forma que se observa en la Figura 5.

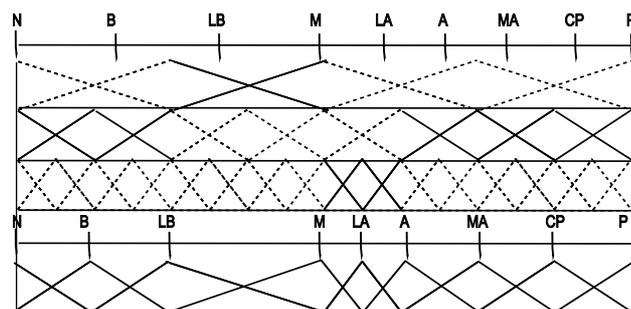


Figura 5: Representación de 9 etiquetas.

Vemos que aunque S_R está distribuido uniformemente, el método lo representa de forma no uniforme, en este caso con una distribución centrada, ya que no hay ningún nivel en LH que sea adecuado para representar el subconjunto. Del mismo modo ha representado S_L que aún siendo uniforme, el método lo ha representado de forma no uniforme.

4. REPRESENTACIÓN DE UN CTLNB CON JERARQUÍAS LINGÜÍSTICAS EXTENDIDAS

En esta sección, presentamos una metodología que representa CTLNB mediante el uso de las Jerarquías Lingüísticas Extendidas, con el objetivo de superar las limitaciones provocadas por el uso de las Jerarquías Lingüísticas de la metodología propuesta en [4].

La idea principal de la metodología que presentamos se fundamenta en la flexibilidad en la construcción de las Jerarquías Lingüísticas Extendidas, ya que permiten incluir cualquier conjunto de términos frente a los conjuntos predefinidos de las Jerarquías Lingüísticas. Para representar un CTLNB, nuestra propuesta aprovecha la flexibilidad construyendo una Jerarquía Lingüística Extendida *ad hoc*, atendiendo a las necesidades de representación del CTLNB y no al revés, como sucede en [4], que es el CTLNB el que debe adaptarse a la Jerarquía Lingüística.

La nueva metodología de representación consiste en dos fases, la primera de construcción de la ELH *ad hoc* y la segunda de representación del CTLNB (ver Figura 6).

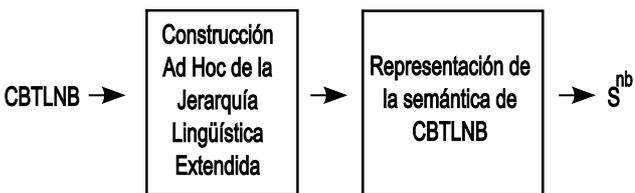


Figura 6: Fases de la nueva metodología.

4.1. CONSTRUCCIÓN DE LA JERARQUÍA LINGÜÍSTICA EXTENDIDA AD HOC

En esta fase se incluyen los niveles necesarios para representar el CTLNB, S^{nb} , en la Jerarquía Lingüística Extendida, ELH, y que son determinados por la distribución de S_L y S_R . Primero, se incluyen los conjuntos para S_L y, a continuación, para S_R en ELH.

Atendiendo a la distribución de etiquetas se pueden dar los siguientes casos:

- *Caso 1. La distribución en S_L es uniforme.* Añadir a la jerarquía un nivel, $l(k, n(k))$ cuya granularidad se determina como sigue:

$$n(k) = 2 \cdot \#(S_L) + 1$$

- *Caso 2. La distribución de S_L es extrema y $\#(S_L)$ es impar.* Añadir dos niveles a la jerarquía,

$l(i, n(i))$ y $l(j, n(j))$, cuyas granularidades son determinadas por las siguientes expresiones:

$$n(j) = (\#S_L - 1) \cdot 2 + 1, \quad n(i) = (2 \cdot n(j)) - 1$$

- *Caso 3. La distribución de S_L es centrada y $\#(S_L)$ es impar.* Se añaden dos niveles, cuyas granularidades son similares al caso 2.

Evidentemente, existe un mayor número de casos de distribución de etiquetas en el CTLNB, pero inicialmente nos vamos a centrar en estos tres para mostrar el funcionamiento general de nuestra propuesta.

Una vez incluidos los niveles para S_L , continuamos con los niveles necesarios para S_R . El procedimiento es similar, salvo que algún nivel necesario para S_R ya hubiese sido incluido para S_L entonces no hace falta añadirlo.

Para finalizar la construcción de la Jerarquía Lingüística Extendida, se incluye un último nivel siguiendo la Regla Extendida 2. En este nivel se llevarán a cabo los procesos de computación con palabras.

4.2. REPRESENTACIÓN DE LA SEMÁNTICA DE CBTLNB

La representación de la semántica de un CTLNB depende de la distribución y del número de las etiquetas de los subconjuntos S_L y S_R . Para representar la semántica de CTLNB es indistinto qué subconjunto, S_L o S_R , se represente primero, ya que el procedimiento para cada subconjunto es independiente. Aquí, describimos el procedimiento para la representación del subconjunto S_L , siendo el procedimiento para representar S_R análogo.

Dado que la representación de la semántica del CTLNB en la Jerarquía Lingüística Extendida depende de la distribución y del número de etiquetas del subconjunto. El primer paso, consiste en conocer las etiquetas que se asignará a cada nivel.

Si la distribución es uniforme se utiliza un solo nivel $l(k, n(k))$ para representar el subconjunto, asignando las etiquetas del subconjunto $\#S_L$ la semántica de las etiquetas del nivel $l(k, n(k))$.

Si la distribución es centrada o extrema, se utilizan para la representación 2 niveles, $l(j, n(j))$ y $l(i, n(i))$, que han sido incluidos en ELH en la fase anterior. El nivel de mayor granularidad, $l(i, n(i))$, representa la zona con mayor concentración de etiquetas y el nivel de menor granularidad, $l(j, n(j))$, representa la zona con menor concentración de etiquetas. Para asignar la semántica, calculamos el número de etiquetas que se representarán en $l(i, n(i))$ y en $l(j, n(j))$:

$$lab_j = \frac{n(i) - 1}{2} - \#S_L, \quad lab_i = \#S_L - lab_j$$

Siendo lab_j el número de etiquetas de S_L representadas en el nivel $l(j, n(j))$ y lab_i en número de etiquetas de S_L representadas en el nivel $l(i, n(i))$. Cabe mencionar que habrá una etiqueta, llamada *punte*, que su semántica se representará utilizando media etiqueta del nivel $l(i, n(i))$ y media etiqueta del nivel $l(j, n(j))$. Para ello, primero asignamos a la etiqueta punte una etiqueta completa en el correspondiente nivel y a continuación, reasignamos media etiqueta con la correspondiente media etiqueta en el otro nivel.

A continuación, atendiendo a la distribución S_L y su número de etiquetas, describimos el procedimiento de representación en cada uno de los casos existentes.

Representación de S_L

- *Caso 1. La distribución de S_L es uniforme.*
 1. A las etiquetas del lado izquierdo de S^{nb} se les asignan todas las etiquetas del lado izquierdo del nivel $l(k, n(k))$.

$$S_L^{nb} \leftarrow S_L^{n(k)}$$

- *Caso 2. La distribución de S_L es extrema, siendo $\#(S_L)$ impar.*
 1. Se asignan lab_i etiquetas del extremo del nivel $l(i, n(i))$ a las lab_i etiquetas del extremo de S^{nb} .

$$\{s_0^{nb}, \dots, s_{lab_i-1}^{nb}\} \leftarrow \{s_0^{n(i)}, \dots, s_{lab_i-1}^{n(i)}\}$$

2. Se asignan el resto de etiquetas del subconjunto en el nivel $l(j, n(j))$.

$$\{s_{lab_i}^{nb}, \dots, s_{\#S_L-1}^{nb}\} \leftarrow \{s_{\frac{lab_i}{2}}^{n(j)}, \dots, s_{\frac{n(j)-1}{2}-1}^{n(j)}\}$$

3. Se realiza el puente modificando la semántica de la parte baja de la etiqueta $s_{lab_i}^{nb}$.

$$\underline{s_{lab_i}^{nb}} \leftarrow \underline{s_{lab_i}^{n(i)}}$$

- *Caso 3. La distribución de S_L es centrada, siendo $\#(S_L)$ impar.*
 1. Se asignan lab_j etiquetas del extremo del nivel $l(j, n(j))$ a las lab_j etiquetas del extremo de S^{nb} .

$$\{s_0^{nb}, \dots, s_{lab_j-1}^{nb}\} \leftarrow \{s_0^{n(j)}, \dots, s_{lab_j-1}^{n(j)}\}$$

2. Se asignan el resto de etiquetas del subconjunto al nivel $l(i, n(i))$.

$$\{s_{lab_j}^{nb}, \dots, s_{\#S_L-1}^{nb}\} \leftarrow \{s_{lab_j, 2}^{n(i)}, \dots, s_{\frac{n(i)-1}{2}-1}^{n(i)}\}$$

- Se realiza el puente la semántica de la etiqueta s_{lab_j} .

$$\underline{s_{lab_j}^{nb}} \leftarrow \underline{s_{lab_j}^{n(j)}}$$

Representación de S_C

- *Caso 1. La distribución de S_L es uniforme.*
 1. A la parte alta de la etiqueta central de S^{nb} se le asigna la parte alta de la etiqueta central del nivel $l(k, n(k))$.

$$\overline{s_c^{nb}} \leftarrow \overline{s_c^{n(k)}}$$

- *Caso 2. La distribución de S_L es extrema, siendo $\#(S_L)$ impar.*
 1. A la parte alta de la etiqueta central de S^{nb} se le asigna la parte alta de la etiqueta central del nivel $l(j, n(j))$.

$$\overline{s_c^{nb}} \leftarrow \overline{s_c^{n(j)}}$$

- *Caso 3. La distribución de S_L es centrada, siendo $\#(S_L)$ impar.*
 1. A la parte alta de la etiqueta central de S^{nb} se le asigna la parte alta de la etiqueta central del nivel $l(i, n(i))$.

$$\overline{s_c^{nb}} \leftarrow \overline{s_c^{n(i)}}$$

A continuación, mostramos distintos ejemplos que ilustran cada uno de los casos.

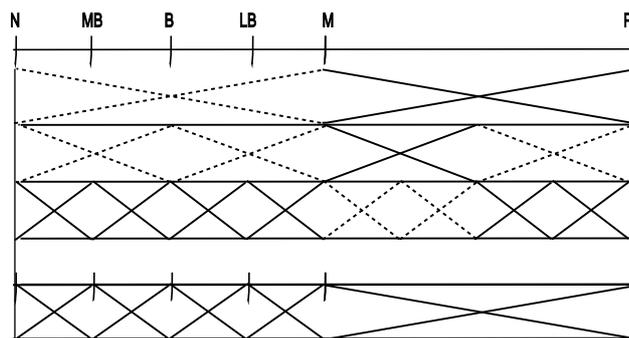


Figura 7: Caso 1. Representación de S_L

Por último, asumiendo el CTLNB de la Figura 4, el nuevo método representa el CTLNB de la forma que se observa en la Figura 10. Podemos observar como el método propuesto ha representado fielmente el CTLNB, utilizando la ELH *ad hoc* construida para dicho CTLNB.

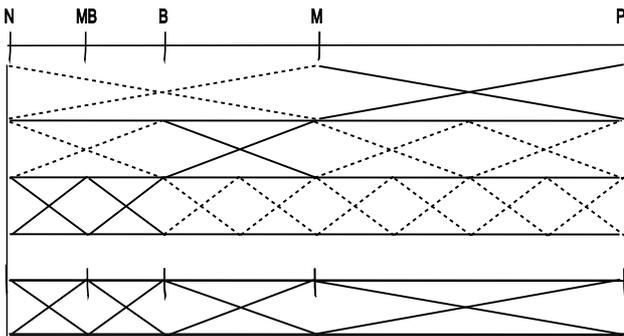
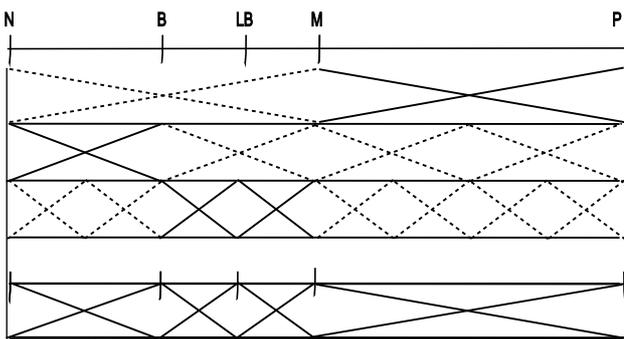
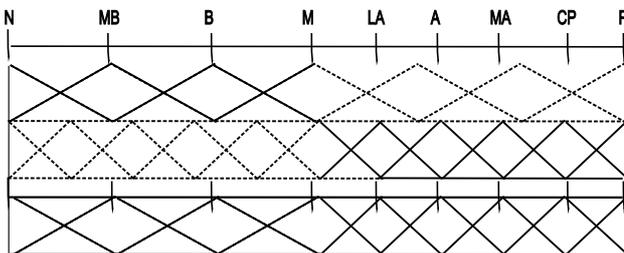
Figura 8: Caso 2. Representación de S_L Figura 9: Caso 3. Representación de S_L 

Figura 10: Representación del CTLNB utilizando el nuevo método

5. COMENTARIOS FINALES

En este trabajo, hemos desarrollado una nueva metodología que representa conjuntos de términos lingüísticos no balanceados, venciendo las limitaciones de la metodología revisada. Para ello, hemos utilizado una nueva estructura denominada extensión de las jerarquías lingüísticas. Nuestras líneas de actuación futuras se encaminan en desarrollar el modelo computacional para poder operar sobre las etiquetas del conjunto de términos lingüísticos no balanceado sin pérdida de información. Para ello, utilizaremos el último nivel de la jerarquía lingüística ampliada.

Agradecimientos

Este artículo ha sido realizado con la ayuda de los

Proyectos P08-TIC-3548, TIN2009-08286 y los Fondos Feder.

Referencias

- [1] F.J. Cabrerizo, S. Alonso, and E. Herrera-Viedma. A consensus model for group decision making problems with unbalanced fuzzy linguistic information. *International Journal of Information Technology and Decision Making*, 8(1):109–131, 2009.
- [2] M. Espinilla. Nuevos Modelos de Evaluación Sensorial con Información Lingüística. PhD thesis, Universidad de Jaén. <http://sinbad2.ujaen.es/sinbad2/files/TesisGrupo/TesisMacarena.pdf>, 2009.
- [3] F. Herrera, E. Herrera-Viedma, and L. Martínez. A hierarchical ordinal model for managing unbalanced linguistic term sets based on the linguistic 2-tuple model. In *Proceedings of the Eurofuse Workshop on Preference Modelling and Applications*, pages 201–206, Granada (Spain), April 2001.
- [4] F. Herrera, E. Herrera-Viedma, and L. Martínez. A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems*, (2), 2008.
- [5] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.
- [6] F. Herrera and L. Martínez. A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, 31(2):227–234, 2001.
- [7] E. Herrera-Viedma and A.G. López-Herrera. A model of information retrieval system with unbalanced fuzzy linguistic information. *International Journal of Intelligent Systems*, 22(11):1197–1214, 2007.
- [8] L. Martínez, M. Espinilla, and J. Liu. Extended linguistic hierarchies. *International Conference on Intelligent System and Knowledge Engineering*, Xiamen (China), 2008.
- [9] L.A. Zadeh. The concept of a linguistic variable and its applications to approximate reasoning. *Information Sciences, Part I, II, III*, 8,8,9:199–249,301–357,43–80, 1975.