XIX Congreso Español sobre Tecnologías y Lógica Fuzzy (XIX ESTYLF)

ESTYLF 7:
APLICACIONES II

Estudio empírico sobre el tratamiento de la sinonimia en los razonamientos vagos

Gael Velasco Benito
Dpto. de Filosofía y Antropología
Universidad de Santiago de Compostela
gael.velasco@gmail.com

Alberto Bugarín
Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS)
Universidad de Santiago de Compostela
alberto.bugarin.diz@usc.es

Alejandro Sobrino
Dpto. de Filosofía y Antropología
Universidad de Santiago de Compostela
alejandro.sobrino@usc.es

***Resumen*—Existen dos alternativas relacionadas con la manera en la que se debe formalizar los términos vagos de nuestro lenguaje para su correcta implementación en un lenguaje de programación lógica. Una opción es hacer preciso el contenido semántico (es decir, eliminar la vaguedad de un término lingüístico estableciendo límites claros), y otra opción es mantener las propiedades vagas del lenguaje (a pesar de las muchas complicaciones que surgen a raíz de ello). En este artículo ofrecemos un argumento empírico a favor de esta última postura, y analizamos la posibilidad de desarrollar una aproximación al razonamiento impreciso de naturaleza más bien lingüística que numérica. Los resultados obtenidos apoyan esta posibilidad, y ofrecen evidencias según las cuales algunos de los problemas más básicos que pueden presentarse al implementar un método de resolución vago en computación pueden resolverse haciendo uso de la conocida como "aproximación de grado".**

***Palabras clave*—vaguedad; aproximación de grado; razonamiento natural; computación con palabras.**

I. Introducción

En el campo de la computación con lenguaje natural parece existir un acuerdo general sobre el hecho de que el lenguaje humano es, esencialmente, vago, y que dicha vaguedad influye de manera directa en nuestra manera de razonar. Este comportamiento difiere del procesamiento del lenguaje natural clásico, que usa algoritmos precisos. Tradicionalmente, la vaguedad se ha contrapuesto a la precisión: *"We are able to conceive precision; indeed, if we could not do so, we could not conceive vagueness, which is merely the contrary of precisión"* [19:5]. Esta diferencia es la que ha generado más dificultades a la hora de desarrollar un sistema automático computacional que sea capaz de procesar y manipular de manera correcta los términos vagos.

Frente a este problema, una primera aproximación consiste en tratar la vaguedad como un defecto que debe ser eliminado de nuestro lenguaje [19]. Los límites entre las extensiones positivas y negativas de un término vago (es decir, los parámetros que permiten clasificar un elemento como perteneciente o no a la categoría establecida por una palabra) podrían ser descubiertos y aplicados de manera precisa, obteniendo así un formalismo lógico menos imperfecto que el lenguaje natural y que podría llegar a sustituirlo en cualquier ámbito. Las aproximaciones más comunes a esta postura hacen uso de la lógica bivalente o de dos valores (algo es bello o no lo es, una persona es joven o no lo es, etc.), al considerarla como la lógica más sencilla y que mejor representa la realidad [18]. En esta tradición semántica supervaluacionista de Fine [6] y Dummett [3] se propone la posibilidad de que a cualquier término lingüístico P se le puede (y debe) aplicar una serie de precisiones (*precisifications* en inglés) que permitan señalar qué elementos pertenecen a las extensiones positivas y negativas de P.

Existen varios argumentos en contra de esta aproximación, señalando la obtención de resultados contraintuitivos y poco naturales como su principal argumento, e indicando que existen otras lógicas fuera de la lógica bivalente mejores para el estudio formal de la vaguedad [21]. La lógica multivalorada aporta resultados más próximos a la realidad, utilizando una serie de grados numéricos que permiten una clasificación más fina de los elementos pertenecientes al dominio de una palabra. El problema está en que si estos grados no se utilizan de manera aproximada [4], el modelo que se obtiene es un modelo que también trata con precisión la vaguedad (aunque con varios —o infinitos— límites en lugar de solo uno), arrastrando consigo todos los problemas inherentes a estas aproximaciones. Sería, pues, necesario buscar un formalismo gradual no numérico, capaz de representar el significado de los términos vagos sin, para ello, modificar su forma de uso o su contenido semántico.

Tomando como punto de partida la opinión de que la acción de precisar los términos vagos es una solución que, aunque funciona como aproximación general al problema, en ciertos casos puede llevar a resultados poco intuitivos, en este trabajo estudiamos la posibilidad de una alternativa no precisa basada en las relaciones semánticas ya existentes en nuestro lenguaje.

La organización del artículo es la siguiente. En la Sección II introducimos los fundamentos teóricos del problema, así como la conveniencia práctica de utilizar la aproximación de grado como base. En las Secciones III y IV presentamos el estudio empírico llevado a cabo en el ámbito de los razonamientos vagos. En la Sección V se discuten los resultados obtenidos, los cuales sugieren una posible correlación entre los juicios valorativos humanos en un razonamiento y la distancia entre grados en una escala.

II. Tratamiento gradual no preciso de los razonamientos vagos

Uno de los intentos más importantes y exitosos de tratar la vaguedad desde una perspectiva computacional es la conocida como Computación con Palabras [25], [26], un sistema que pretende simplificar la computación de problemas complejos que involucran lenguaje impreciso o vago y facilitar su implementación. Su principal aportación es la introducción del concepto de variable lingüística, así como la idea de que en computación se puede trabajar directamente con palabras (o percepciones), y no únicamente con valores numéricos o medidas. Siguiendo sus directrices, un problema tratado mediante la Computación con Palabras estará expresado en lenguaje natural, de manera que los resultados obtenidos serán, a su vez, proposiciones o palabras.

El inconveniente reside en que, aunque las bases de la Computación con Palabras son claras, el proceso de resolución no es lingüístico, sino numérico, dado que reduce las proposiciones de entrada a expresiones propias de la Lógica Difusa y las manipula hasta conformar un resultado preciso que, en un paso posterior, es transformado de nuevo a lenguaje natural. Inevitablemente, el contenido semántico de las expresiones vagas se volverá preciso, y por ende modificado. A pesar de las múltiples ventajas propias de esta aproximación, consideramos que en el caso de los razonamientos vagos sería conveniente buscar un método alternativo que nos permitiera conservar el significado de las palabras tal y como lo percibimos los seres humanos durante todos los pasos. Creemos que, de ser esto posible, las conclusiones obtenidas serían más coherentes y naturales, y que el proceso de resolución sería más sencillo de comprender para aquellas personas no versadas en la computación o en la Lógica.

En la bibliografía pueden encontrarse diversos argumentos en contra de la acción de precisar de los términos vagos, todos los cuales coinciden al defender que la vaguedad no es un defecto superficial que pueda corregirse trazando límites precisos, sino una característica semántica de nuestro lenguaje (una primera aproximación al respecto puede encontrarse en [23]). Por solo citar algunos casos, en [24] la acción de volver preciso lo que es vago es criticada al argumentar que, aunque pudiésemos definir reglas explícitas que formalizasen la vaguedad, estas no servirían para explicar el significado de un término vago a alguien que no esté familiarizado con él (es decir, estas reglas no serían "informativas"). De forma similar, en [1] se apoya esta postura al argumentar que las condiciones de verdad de las palabras vagas solo pueden ser gobernadas por reglas igualmente vagas. Y otros autores como Edgington [4], aunque apoyan el uso de la Lógica Difusa en el desarrollo de una teoría formal de la vaguedad, advierte que ésta nunca podría ser utilizada de manera precisa, sino como valores aproximados que permitan construir un modelo aproximado del problema y no den lugar a infinitos estados precisos.

En el presente artículo partiremos de las ideas contenidas en [17], [22], donde se propone como alternativa la utilización de las relaciones lingüísticas ya existentes en nuestro lenguaje, las también conocidas como "relaciones semánticas". Algunos ejemplos serían la sinonimia (igualdad de significado, como por ejemplo "escaso" y "poco"), la antonimia (significado opuesto o contradictorio, como es el caso de "afilado" y "romo") o la hiperonimia (inclusión de significados, como es el caso de "libro", que es un hiperónimo de "novela").

Introducir este tipo de relaciones en un lenguaje de programación lógico sería una ventaja enriquecedora y sustancial, dado que la mayoría de ellos se limita a interpretar relaciones sintácticas, y no semánticas. Utilizando como ejemplo el lenguaje Prolog, supongamos que tenemos un programa con los siguientes hechos y reglas donde $P(x)$ denota "x posee la propiedad P" (es decir, "la pasta está rica", "el sushi está bueno", etc.) y $menu(X)$:- $rico(X)$ representa una norma según la cual "el elemento designado por X estará en el menú si X cumple con un requisito, a saber: que X cumpla la propiedad de estar rico":

rico(pasta).

bueno(sushi).

delicioso(paella).

menu(X) :- *rico*(X).

En este caso, la respuesta a la pregunta "¿Qué hay en el Menú?" es clara: "$X = pasta$".

La solución, aunque válida, se nos antoja incompleta dado que aunque el sushi y la paella, según el programa, no estén "ricos", no es del todo coherente afirmar que no cumplen con el requisito necesario para aparecer en el Menú. Pero tampoco podemos afirmar lo contrario (que el sushi y la paella cumplen el requisito al mismo nivel que la pasta).

Lo que haría falta sería que el programa supiese que estos tres adjetivos están semánticamente relacionados entre sí, y cuál es el nivel o tipo de sinonimia de cada uno con respecto a los otros dos. De esta forma, podría presentar más de una respuesta, aunque matizando que algunas de ellas tienen un grado de validez superior al de las otras.

A este respecto se han realizado varios intentos basados en un tratamiento preciso de la vaguedad ([7], [11], [13], entre otros). Muchos de estos métodos parten de la idea de asignar un valor numérico a cada propiedad definida (por ejemplo, asignando los valores 1 y 0,8 a "pasta" y "sushi", respectivamente, con respecto a la propiedad "estar rico"), y utilizar esa clasificación para ordenar las respuestas de más a menos válida, cada una de ellas señalada con un número entre 0 y 1 que determina cuán válida es con respecto a la pregunta hecha al programa. Estos trabajos logran con éxito aumentar la capacidad expresiva de lenguajes como Prolog, pero como ya hemos mencionado anteriormente, esta manera de tratar la vaguedad tiene sus problemas, siendo el principal de todos que asignan valores numéricos arbitrarios sin una base sólida que los justifique, los cuales forman a su vez una serie de infinitos estados precisos que simulan un comportamiento semántico muy distinto al de las palabras vagas de nuestro lenguaje.

Las relaciones lingüísticas no presentarían estas dificultades, aunque un método así será más difícil de implementar a causa

de las diferencias inherentes entre seres humanos y sistemas computacionales. Las relaciones semánticas más útiles en el caso de los razonamientos vagos son las de sinonimia y antonimia a causa de sus paralelismos directos con los conceptos de igualdad y negación implicados en los procesos resolutivos de lenguajes de programación Prolog; es decir, el emparejamiento o *matching* y la resolución por refutación.

Una posible solución se basa en la aproximación conocida como "aproximación de grado" [8]–[10]. Según esta alternativa, los elementos pertenecientes ciertas categorías gramaticales como la de los adjetivos estarían asociados a una entidad abstracta, el grado d, que representaría el valor o cantidad de la propiedad dada por esos elementos. Estos grados estarían a su vez ordenados en escalas, elementos lingüísticamente relevantes que pueden definirse por medio de tres parámetros: un conjunto de grados, una relación de orden que ordena dicho conjunto, y una "dimensión" (también llamada "medida") que establece cuál es la propiedad real representada por la escala (peso, tamaño, edad, etc.). Así pues, todos los elementos de la escala están relacionados, de forma que dos elementos cualesquiera que pertenezcan a ella podrán compararse entre sí (siendo uno mayor que el otro). No hay que perder de vista que la aproximación de grado, en su forma clásica, impone en sus escalas un determinado grado "estándar de comparación" o "normal" que sirve como límite claro entre aquellos grados que representan de manera positiva la dimensión de la escala y aquellos que no. Dado el tratamiento no vago de este formalismo, en el estudio no adoptaremos dicha medida, centrándonos únicamente en la existencia de los grados y de una relación de orden total que nos permite ordenarlos de mayor a menor dentro de una escala.

III. Hipótesis de trabajo

La principal hipótesis sobre la que trabajamos tiene su apoyo teórico en [15], donde el contenido semántico de un término lingüístico se subdivide en dos categorías: significado "cognitivo" y significado emotivo o "no cognitivo". En términos generales, esta clasificación respondería a dos facultades psicológicas distinguibles en el uso del lenguaje: una intelectual, lógica, y otra intuitiva o emotiva. Sin embargo, muchos de los factores que definen el significado no cognitivo de una palabra no están realmente relacionados con lo que se suele considerar como "significado" o "contenido semántico", sino con otros elementos de carácter pragmático tales como la aceptabilidad estilística o social (por ejemplo, el significado no cognitivo determinaría el carácter formal o coloquial de una palabra, o si existe algún contexto en que dicha palabra no puede o no debe ser usada). Por ese motivo, Lyons determina que el tratamiento formal de la sinonimia, así como su aplicación a cualquier proposición natural, debe restringirse únicamente a su contenido cognitivo.

Volviendo a la aproximación gradual, el significado de una palabra, una vez dejado a un lado cualquier implicación no lógica y centrándonos únicamente en el significado cognitivo, estaría representado por los grados. En el caso de un lenguaje de programación Prolog, estos grados estarían representados por las mismas palabras, las cuales conformarían una escala lingüística (en el sentido dado en [12] de "un conjunto de alternantes lingüísticos, o expresiones contrastivas de la misma categoría gramatical, que pueden ordenarse en un orden lineal por grados de informatividad o fuerza semántica"). De esta manera, si prescindimos del formalismo bivalente, inherente de las aproximaciones de grado clásicas, serían pues estas palabras, y su relación total con el resto de palabras de la misma escala, los únicos elementos necesarios para representar el contenido lógico de los términos y su influencia en las condiciones de verdad de un razonamiento vago.

En consecuencia, si Prolog u otros lenguajes de programación pudiesen interpretar de manera correcta estos grados, y a su vez pudiesen mostrarlos al usuario mediante el uso del lenguaje natural, podrían alcanzarse soluciones no precisas para razonamientos en los que intervienen términos vagos, ya que las escalas son elementos que Prolog y otros lenguajes interpretan y manejan con facilidad.

La pregunta que se pretende estudiar es esta: si tenemos una serie de premisas conectadas entre sí por sinónimos de las cuales se infiere una conclusión, ¿existe una relación directa entre el juicio de valor que, de media, realizaría un ser humano acerca de dicha conclusión, y la separación que existe entre los grados de los sinónimos implicados en el razonamiento?

IV. Análisis empírico del problema

El experimento antes mencionado tiene como finalidad estudiar, a partir de la aproximación gradual al problema de la vaguedad, un método con el cual introducir juicios coherentes no precisos de razonamientos vagos en un sistema automático no supervisado. Más concretamente, el experimento explora la posible correlación existente entre la evaluación lingüística de la conclusión de un razonamiento básico y la separación existente entre los adjetivos valorativos implicados dentro de una escala. El objetivo no es tanto establecer una prueba indiscutible sobre la validez de la aproximación gradual como servir de argumento a favor de una teoría de la vaguedad empírica, la cual podría ser sistematizada sin necesidad de aplicar un método numérico.

IV-A. Metodología

El estudio se ha realizado mediante un formulario sobre el que se ha consultado a estudiantes, profesores e investigadores pertenecientes al ámbito universitario. En total, los resultados aquí mostrados se corresponden a las respuestas dadas por 38 participantes. En una primera parte se les hizo una serie de preguntas básicas, a saber: su edad, nivel de estudios y nivel de conocimiento del castellano. Como resultado, la edad media de los participantes ha sido de 25,2 años, y dado que todos han señalado el castellano como su lengua madre (o, en su defecto, un avanzado conocimiento de dicho idioma), y no se han detectado irregularidades en la totalidad de sus respuestas, no ha sido necesario excluir ninguno de los cuestionarios recibidos.

IV-B. Materiales y procedimiento

Los adjetivos utilizados para este experimento han sido cuatro, dos asociados a un parámetro medible (edad para "anciano", temperatura para "caliente") y dos asociados a uno que no puede ser medido (al menos de manera numérica), siendo "guapo" y "divertido" los elegidos para este último caso[1].

El cuestionario se divide en dos partes bien diferenciadas. En la primera, al participante se le presentaron 12 textos breves ordenados al azar, cada uno de los cuales se correspondía con una serie de premisas, enriquecidas de forma escueta para proveerles de un contexto que facilitara su comprensión. A continuación se les mostraba una conclusión posible para cada fragmento, y finalmente se les pedía que estableciesen su grado de conformidad con dicha conclusión en una escala Likert[2] de siete puntos [14].

Un ejemplo es el siguiente:

TEXTO: "El director de casting de una película pone en varias páginas web el siguiente anuncio: "se necesita actor maduro para una nueva campaña de perfumes". Al día siguiente su asistente llega a su despacho y le dice que hay un candidato esperando en la sala. Cuando el director le pregunta sobre el aspecto de dicho candidato, el ayudante le dice que es un anciano."

PREGUNTA: "¿Estás de acuerdo con la afirmación "el candidato cumple con los requisitos del anuncio"?"

OPCIONES: 1) Totalmente de acuerdo, 2) Bastante de acuerdo, 3) Algo de acuerdo, 4) Ni de acuerdo ni en desacuerdo, 5) Algo en desacuerdo, 6) Bastante en desacuerdo, 7) Totalmente en desacuerdo.

La peculiaridad de las premisas aquí presentes es que el punto de unión entre ellas no viene dado por una misma palabra, sino por dos sinónimos, uno de los cuales es siempre uno de los cuatro adjetivos-base antes mencionados ("anciano", "caliente", etc.) y el otro uno de sus sinónimos.

Fundamentándonos en la clasificación dada por Cruse [2], Palmer [16] y Lyons [15], existirían tres tipos de sinónimos básicos: sinónimos absolutos (palabras con idéntico significado, como son "computadora" y ordenador" donde la definición de ambas viene sustentada por las mismas acepciones), sinónimos cognitivos (palabras que, a pesar de tener un contenido semántico formal o "intelectual" idéntico, se diferencian en la intención, actitud o carga emotiva de cada una, como son "delgado" y "flaco", donde "delgado" tiene una carga neutra mientras "flaco" suele utilizarse en sentido peyorativo) y "plesiónimos" o cuasi-sinónimos (palabras entre las que existe una diferencia semántica tal que si una fuera sustituida por otra en una proposición, las condiciones de verdad de dicha proposición cambiarían, como es el caso de "amplio" y "vasto" donde, aunque ambos sirven para designar un espacio grande, "vasto" hace referencia a un espacio mayor que "amplio").

El cuestionario pretende estudiar la relación de cada uno de los cuatro adjetivos-base con un sinónimo cognitivo y dos cuasi-sinónimos, todos ellos recogidos en WordNet [5] (ver Tabla I)[3].

Tabla I: Clasificación semántica de los adjetivos utilizados

	Sin. Cognitivo	Cuasi-sinónimo 1	Cuasi-sinónimo 2
Anciano	Viejo	Senil	Maduro
Caliente	Cálido	Ardiente	Tibio
Guapo	Bonito	Precioso	Agraciado
Divertido	Entretenido	Alegre	Animado

La segunda y tercera columna de cuasi-sinónimos se corresponden con los adjetivos que hacen referencia a un grado superior e inferior, respectivamente, de la dimensión a la que están asociados (la temperatura a la que hace referencia "ardiente" es superior a la que hace referencia "caliente", mientras que la temperatura a la que hace referencia "tibio" es inferior a la que hace referencia "caliente", etc.). Estos adjetivos han sido elegidos con el fin de que todas las categorías sinonímicas existentes estuviesen representadas en el estudio (cognitivos y cuasi-sinónimos tanto en un grado superior como inferior), así como dos ejemplos claros de adjetivos asociados a medidas (la edad de un "anciano" puede medirse en años, y la temperatura de algo "caliente" en grados centígrados) y otros dos que no pueden ser reducidos a una sola medida (no existe ninguna unidad capaz de indicar de manera explícita el nivel de belleza y/o de diversión).

La segunda parte del cuestionario consiste en darle al participante un conjunto de palabras ordenadas al azar y pedirle que las coloque en una escala de diez posiciones en los lugares que considerase más correctos (el número de posiciones ha sido elegido con el fin de que existiera espacio suficiente para que cada participante situase la distancia que considerase oportuna entre cada par de palabras sin que el tamaño de la escala influyese en su decisión). Cada conjunto se corresponde a uno de los adjetivos-base, sus tres sinónimos utilizados en la primera parte (tal y como aparecen en la Tabla I) y uno o dos sinónimos nuevos más (ver Tabla II).

Por último, a los participantes se les pregunta si alguno de los cuatro conjuntos de palabras les ha resultado especialmente difícil de ordenar.

Tabla II: Grupos de adjetivos

Edad	maduro, envejecido, canoso, senil, viejo, anciano
Temperatura	tibio, ardiente, cálido, templado, caliente
Belleza	guapo, precioso, bonito, cautivador, agraciado, atractivo
Diversión	divertido, alegre, animado, entretenido, ameno

IV-C. Predicciones y resultados

El cuestionario fue desarrollado para comprobar si los razonamientos lingüísticos que realizarán los participantes en

[1]La categoría gramatical de los adjetivos ha sido la elegida para el estudio porque a) en la mayoría de los casos son los adjetivos quienes, junto a los adverbios, introducen la vaguedad en nuestro lenguaje, y b) "la gramática tradicional ha reconocido el grado como uno de los accidentes gramaticales del adjetivo", ya que "expresan propiedades, características o dimensiones que, sin variar de manera cualitativa, puede, en cambio, variar cuantitativamente" [20:6-8].

[2]Una escala Likert es una escala psicométrica utilizada en cuestionarios donde al participante se le presenta una serie de afirmaciones y este debe elegir aquellas opciones que mejor representen su nivel de acuerdo o desacuerdo en cada una. Nos hemos decantado por este tipo de cuestionarios gracias al paralelismo existente entre las escalas Likert (cuyos grados están expresados en lenguaje natural) y las escalas lingüísticas de la segunda parte del experimento.

[3]La ausencia de sinónimos absolutos en el estudio se debe a que su existencia no es del todo clara, y el número de casos en el que puede sugerirse su existencia dentro de la categoría de los adjetivos es mínimo [2].

la primera parte tienen algún tipo de correlación con su ordenamiento en las escalas de la segunda parte. De ser así, los resultados darían una pista sobre cómo debería implementarse un razonador lingüístico que tuviese en cuenta el contenido semántico vago de los términos en un lenguaje de programación lógica (como veremos a continuación).

Los resultados de la primera parte arrojaron luz sobre el primer punto a tener en cuenta: que las respuestas no están uniformemente distribuidas entre los siete puntos de la escala Likert, sino que se agrupaban en torno a un valor central, formando una curva de tipo distribución normal o Campana de Gauss (ver Fig. 1)[4]. Esto parece indicar que el uso de

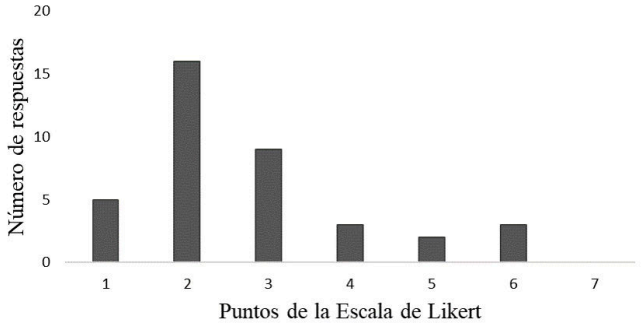


Fig. 1: Distribución de las respuestas para el caso de los sinónimos "anciano" y "senil" en una escala Likert de siete puntos.

sinónimos en un silogismo sí modifica las condiciones de verdad del razonamiento, comportamiento del que no puede dar cuenta un tratamiento de la vaguedad bivalente.

El segundo punto a comentar es que, efectivamente, tal y como se había predicho, las palabras asociadas a una magnitud numérica (edad, temperatura) resultaron más fáciles de ordenar que las asociadas a una magnitud no numérica (belleza, diversión). Esto debe tenerse en cuenta, pues puede ser, a su vez, otra razón que explique por qué las condiciones de verdad de una conclusión se ven modificadas.

Por último, el resultado más relevante de todos puede verse referido en la Fig. 2. El eje de las abscisas está formado por los siete puntos de la escala Likert de la primera parte (igual que en la Fig. 1), y el eje de las ordenadas por un valor numérico entre 0 y 5 que señala la distancia existente entre un par de palabras situadas en la misma escala de la segunda parte (0 = situados en la misma posición, 1 = situados en una posición adyacente, 2 = situados en la escala con una posición de diferencia, etc.). A su vez, cada uno de los doce puntos se corresponde a los doce pares de palabras estudiados en el cuestionario, cuyas coordenadas x-y han sido obtenida mediante los valores medios de los resultados obtenidos en la primera y segunda parte, respectivamente. Las barras de error están representadas a partir de los valores de la segunda y cuarta columna de la Tabla III y representan sus intervalos de confianza (para un valor de confianza del 95 %). Teniendo

[4]El eje de las abscisas va de 1 a 7, valores numéricos que se corresponden a los siete puntos Likert de la escala:

1 = "Totalmente de acuerdo"
2 = "Bastante de acuerdo"
3 = "Algo de acuerdo"
4 = "Ni de acuerdo ni en desacuerdo"
5 = "Algo en desacuerdo"
6 = "Bastante en desacuerdo"
7 = "Totalmente en desacuerdo"

Tabla III: Valores representados en la Fig. 2

Par de palabras	Valor_P1	Conf_P1	Valor_P2	Conf_P2
Anciano-Senil	2.74	0.46	1.58	0.42
Anciano-Viejo	18.7	0.40	1.05	0.30
Anciano-Maduro	5.76	0.44	4.45	0.55
Caliente-Tibio	6.34	0.18	4.21	0.51
Caliente-Cálido	2.82	0.60	1.32	0.28
Caliente-Ardiente	6.05	0.52	2.24	0.34
Guapo-Bonito	2.37	0.47	1.47	0.36
Guapo-Agraciado	3.34	0.49	2.34	0.47
Guapo-Precioso	4.42	0.65	2.26	0.39
Divertido-Entretenido	4.55	0.57	2.79	0.59
Divertido-Animado	3.13	0.50	2.34	0.45
Divertido-Alegre	3.68	0.47	2.08	0.45

todos estos valores en consideración, la recta representada en la gráfica es una estimación obtenida mediante la técnica de mínimos cuadrados con un coeficiente de determinación o R^2 de 0,70, valor que refleja la muy buena calidad de este ajuste.

V. Conclusión y discusión general

Los resultados antes expuestos nos permiten concluir dos hechos relevantes, siendo el primero que la sustitución de una palabra A por un sinónimo $SinA$ en un razonamiento natural casi siempre se percibe como una pérdida de las condiciones de verdad de la conclusión (de no ser así, todas o casi todas las respuestas se corresponderían al primer punto de la escala Likert). En segundo lugar, nos ha permitido comprobar que, si la palabra sustituida A dentro del razonamiento es un sinónimo valorativo graduable, la modificación de las condiciones de verdad percibida por un ser humano se relaciona directamente con la separación existente entre los grados representados por A y $SinA$ dentro de una escala lingüística obtenida de manera empírica.

Si las implicaciones de nuestro experimento resultan correctas, en estos casos ya no sería necesario reducir el contenido semántico de una palabra a valores numéricos mediante la

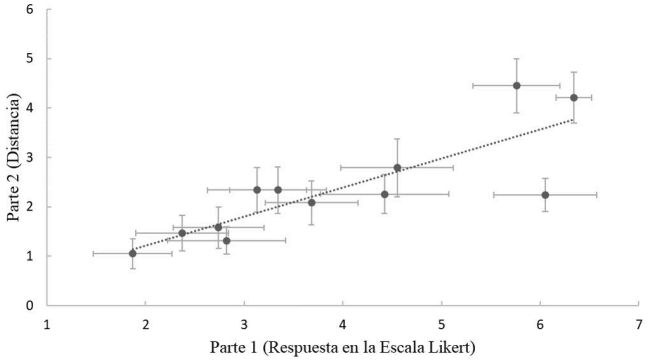


Fig. 2: Representación de los datos del estudio (Tabla III) y su ajuste por mínimos cuadrados.

lógica difusa para su tratamiento no supervisado. Entonces un lenguaje de programación lógica como Prolog podría arrojar más soluciones que las aportadas actualmente, valorando el grado de verdad de cada conclusión de manera lingüística midiendo únicamente la distancia entre los grados asociados a cada término dentro de una escala.

Por supuesto, quedan por resolver algunas cuestiones antes de considerar plausible esta solución, siendo el problema más relevante el de la desambiguación de los términos implicados en un silogismo. Aun limitándonos a tratar únicamente el significado cognitivo (o puramente lógico) de los términos, todavía quedaría determinar la dimensión o magnitud a la que están haciendo referencia en el razonamiento. Y es que, por ejemplo, no es lo mismo utilizar el adjetivo "senil" para hacer referencia a la avanzada edad de un ser humano que a su deterioro cognitivo, ya que en ambos contextos el adjetivo hará referencia a dos grados diferentes, cada uno perteneciente a su propia escala.

Con lo que respecta a los datos tomados, la estimación por mínimos cuadrados nos muestra de manera bastante precisa que cada punto de la escala Likert de la primera parte encuentra una correspondencia directa con una de las distancias establecidas por los participantes en la segunda parte. Si el objetivo es diseñar un método resolutivo que emule la manera de juzgar una conclusión de los seres humanos, de acuerdo a lo aquí expuesto la información podría quedar recogida en escalas ordenadas de manera empírica por el suficiente número de usuarios, lo que simplificaría las inferencias aproximadas o vagas realizadas al estilo Prolog. Por ejemplo, en el caso de un razonamiento tipo Prolog, bastaría con modificar el proceso de unificación con una nueva regla según la cual el emparejamiento entre dos términos se produciría también si los grados asociados a ambos términos en el contexto dado pertenecen a la misma escala, añadiendo en esos casos un determinado "grado de coherencia" al emparejamiento por medio de la distancia existente entre ambos grados.

Para concluir es importante señalar que, aunque es sorprendente la precisión con la que los datos señalan una correlación entre las distancias que existen entre grados y juicios valorativos, los resultados en sí no lo son. Al fin y al cabo, si aceptamos la premisa principal de la aproximación gradual según la cual el contenido semántico cognitivo de los adjetivos se expresa mediante los grados a los que estos hacen referencia, es evidente que cuanto más alejados estén los grados de A y $SinA$ dentro de una misma escala, mayores serán los cambios introducidos en las condiciones de verdad de la conclusión del silogismo al sustituir un término medio por otro.

En todo lo discutido anteriormente, hay que tener en cuenta que para comprobar la validez de estos resultados sería necesario repetir el experimento para un número mayor de participantes y palabras. De igual forma, en futuros estudios cabría la posibilidad de estudiar si este tratamiento de la vaguedad puede extenderse a otras categorías gramaticales sensibles de poseer un contenido semántico vago (tales como los adverbios o, en cierta medida, los sustantivos o incluso los verbos) o si, por el contrario, requieren de un enfoque distinto, no fundamentado en el uso de escalas y grados.

Agradecimientos

Este trabajo ha sido funanciado por el Ministerio de Economía y Competitividad (proyectos TIN2014-56633-C3-1-R y TIN2017-84796-C2-1-R) y la Consellería de Educación de la Xunta de Galicia (proyectos GRC2014/030 y Acreditación 2016-2019, ED431G/08"). Todos los proyectos fueron cofinanciados por el programa FEDER.

Referencias

[1] L. C. Burns, "Vagueness: an investigation into natural languages and the Sorites paradox", Philosophy and Phenomenological Research 54(2), 1994, pp. 483-486.
[2] D. A. Cruse, Lexical Semantics. Cambridge University Press, 1986.
[3] M. Dummett, "Wang's paradox", Synthese, 30, 1975, pp. 301–24.
[4] D. Edgington, Vagueness by Degrees. In Rosanna Keefe and Peter Smith (eds.), Vagueness: A Reader. MIT Press, 1997.
[5] C. Fellbaum, WordNet. An Electronic Lexical Database. Cambridge, MA: MIT Press, 1998. Disponible en: http://wordnetweb.princeton.edu/perl/webwn. [Consulta: 16 junio 2018].
[6] K. Fine, "Vagueness, truth and logic", Synthese 30(3-4), 1975, pp.265-300.
[7] M. Ishizuka, y N. Kanai," Prolog-ELF incorporating fuzzy logic", New Generation Computing, 3(4),1985, pp.479-486.
[8] C. Kennedy, Projecting the adjective: The syntax and semantics of gradability and comparison. New York: Garland, 1999.
[9] C. Kennedy, "Polar opposition and the ontology od 'degrees' ". Linguistics and Philosophy, 24, 2001, pp. 33-70.
[10] C. Kennedy, y L. McNally, "Scale structure and the semantic typology of gradable predicates". Language, 81(2), 2005, pp. 345-381.
[11] R. C. T. Lee, "Fuzzy logic and the Resolution Principle". Journal of the Association for Computing Machinery, 19(1), 1972, pp. 109-119.
[12] S. C. Levinson, Pragmática. Barcelona: Teide, 1989.
[13] D. Li, and D. Liu, A Fuzzy Prolog Database System. Research Studies Press and John Wiley and Sons, 1990.
[14] R. Likert, "A technique for the measurement of attitudes". Archives of Psychology, 140, 1932, pp. 1–55.
[15] J. Lyons, Introduction to Theoretical Linguistics. Cambridge University Press, 1968.
[16] F. R. Palmer, Semantics: A new outline. Cambridge University Press, 1976.
[17] M. Pereira-Fariña, A. Sobrino, and A. Bugarín, "A proposal of fuzzy chained syllogism based on the concept of synonymy". Fuzzy Systems (FUZZ-IEEE), 2013, pp. 1-8.
[18] W. V. Quine, "What price bivalence?", Journal of Philosophy 78 (2), 1981, pp. 90-95.
[19] B. Russell, "Vagueness". Australasian Journal of Philosophy 1 (2), 1923, pp. 84-92.
[20] C. Sánchez López, El grado de adjetivos y adverbios. Madrid: Arco/Libros, 2006.
[21] D. H. Sanford, "Competing semantics of vagueness: Many values versus super-truth". Synthese 33 (2-4), 1976, pp. 195-210.
[22] A. Sobrino, "The Role of Synonymy and Antonymy in Natural Fuzzy Prolog". Soft Computing in Humanities and Social Sciences. R. Seising, V. Sanz (eds.), Springer Berlin Heidelberg, 2012, pp. 209-236.
[23] L. Wittgenstein, Philosophical Investigations, trad. Anscombe, G.E.M. Wiley-Blackwell, 1953.
[24] C. Wright, "On the coherence of vague predicates", Synthese 30 (3-4), 1975, pp. 325-65.
[25] L.A. Zadeh, "Fuzzy logic = computing with words". IEEE Transactions of Fuzzy Systems, 4(2), 1996, pp. 103-111.
[26] L.A. Zadeh, "From computing with numbers to computing with words: From manipulation of measurements to manipulation of perceptions". Circuits and Systems I: Fundamental Theory and Applications, IEEE, 46(1), 1999, pp. 105-119.

Uso de Expresiones Lingüísticas Comparativas en AFRYCA 3.0

1º Álvaro Labella
Dpto. de Informática
Universidad de Jaén
Jaén, España
alabella@ujaen.es

2º Rosa M. Rodríguez
Dpto. de CC de la Computación e IA
Universidad de Granada
Granada, España
rosam.rodriguez@decsai.ugr.es

3º Luis Martínez
Dpto. de Informática
Universidad de Jaén
Jaén, España
martin@ujaen.es

***Resumen*—La mayor parte de los problemas de Toma de Decisión en Grupo (TDG) del mundo real se definen en contextos de incertidumbre, en los cuales hay que modelar y operar con información vaga e imprecisa. El uso Enfoque Lingüístico Difuso ha proporcionado excelentes resultados en el tratamiento de dicha incertidumbre. Sin embargo, dicho enfoque presenta limitaciones en la expresión de la información debido al uso de vocabularios limitados y cerrados, lo que demanda expresiones lingüísticas más flexibles y no limitadas a priori. Para ello, se introdujo el concepto de *Conjunto de Términos Lingüísticos Difusos Dudosos* (CTLDD) que permite modelar Expresiones Lingüísticas Comparativas (ELCs). En los problemas de TDG, los Procesos de Alcance de Consenso (PAC) son uno de los principales campos de investigación, dando lugar a numerosas propuestas de PAC. Como consecuencia, la labor de seleccionar y aplicar el PAC más adecuado para un problema de TDG dado puede resultar compleja. Por esta razón se diseñó AFRYCA, un framework para el estudio y análisis de problemas de TDG. Esta contribución presenta nuevas características de la versión 3.0 de AFRYCA para el modelado de incertidumbre mediante información lingüística difusa dudosa y ELCs en PAC.**

***Palabras clave*—AFRYCA, toma de decisión en grupo, consenso, expresiones lingüísticas comparativas**

I. Introducción

La Toma de Decisiones (TD) es habitual en nuestras actividades diarias en las que, a partir de un conjunto de posibles opciones o alternativas, se debe escoger la mejor. En la Toma de Decisión en Grupo (TDG), un grupo de expertos son los encargados de decidir cuál es la mejor solución para un problema de TD [4], [8]. Los problemas de TDG normalmente se definen en contextos bajo incertidumbre donde la información es vaga e imprecisa. Existen diferentes enfoques que manejan este tipo de información [20]–[22], entre ellos, el *Enfoque Lingüístico Difuso* [22] ha proporcionado buenos resultados en la resolución de problemas de TDG definidos bajo incertidumbre. Hasta ahora estos enfoques modelaban la información lingüística utilizando un único termino lingüístico, lo que limita la expresión del conocimiento de los expertos [14]. Varias propuestas se han presentado en la literatura para intentar superar esta limitación y permitir utilizar más de un término lingüístico. En esta contribución se introduce el concepto de *Conjunto de Términos Lingüísticos Difusos Dudosos* (CTLDD) [15]. Un CTLDD permite modelar la información lingüística mediante el uso de varios términos lingüísticos, permitiendo a los expertos expresar la duda entre ellos. Las *Expresiones Lingüísticas Comparativas* (ELCs) hacen uso de los CTLDDs y permiten a los expertos expresar sus opiniones de una forma más cercana a su modelo cognitivo habitual.

Las necesidades en la resolución de los problemas de TDG han evolucionado a lo largo del tiempo. En ocasiones, ya no solo se requiere encontrar la mejor solución, sino que además, esa solución debe de ser satisfactoria para todo el grupo. Tradicionalmente, los problemas de TDG se resolvían aplicando un proceso de selección de alternativas [2], donde se tenían en cuenta las opiniones individuales de los expertos, pero no el acuerdo entre ellos [16]. Por lo tanto, la decisión final podía no ser del agrado de todos los expertos, provocando que parte de ellos se sintieran excluidos del proceso de decisión. Por esta razón surgieron los Procesos de Alcance de Consenso (PAC) como una fase adicional en la resolución de problemas de TDG [17]. En un PAC, los expertos revisan, discuten y modifican sus opiniones con el objetivo de alcanzar una solución que satisfaga al grupo. Los PAC se han convertido en un pilar fundamental dentro del marco de la TDG y, como consecuencia, se pueden encontrar numerosas propuestas de modelos de PAC, también denominados *modelos de consenso* [1], [3], [5]. Debido a esto, identificar y aplicar el modelo de consenso que mejor se adapte a un problema de TDG puede resultar una tarea ardua, compleja y que demanda mucho tiempo.

AFRYCA, *A FR*amework for the anal*Y*sis of *C*onsensus *A*pproaches, es un framework que permite el análisis y estudio de la resolución de problemas de TDG mediante la simulación de PAC [10]. Los principales objetivos de AFRYCA son: identificar ventajas y desventajas de los diferentes modelos de consenso, establecer comparaciones entre modelos de consenso, identificar cuándo un modelo de consenso es más apropiado para un problema de TDG en particular y visualizar la evolución de las preferencias de los expertos a lo largo del PAC. AFRYCA permite llevar a cabo estas tareas mediante la simulación de PAC, lo que conlleva simular los comportamientos de los expertos que participan en el proceso. Desde su inicio, AFRYCA ha evolucionado tanto en tecnologías como en herramientas para PAC [7]. En esta

Esta contribución está parcialmente financiada por el proyecto de investigación TIN2015-66524-P.

contribución se presenta una nueva versión de AFRYCA que incluye nuevas capacidades para el tratamiento de información lingüística difusa dudosa. Estas funcionalidades son: definición de problemas de TDG donde los expertos expresan sus preferencias mediante el uso de ELCs, soporte para la visualización de dichas preferencias e incorporación de un nuevo modelo de consenso para este tipo de información.

Esta contribución se estructura de la siguiente forma: en la Sección II se introducen brevemente algunos conceptos preliminares. En la Sección III se presenta AFRYCA 3.0 y sus nuevas funcionalidades, mientras que en la Sección IV se muestra un ejemplo ilustrativo de dichas funcionalidades. Finalmente, la Sección V expone algunas conclusiones y trabajos futuros.

II. Preliminares

Aquí se revisan distintos conceptos necesarios para entender AFRYCA y el uso de CTLDDs en dicho framework.

II-A. Toma de Decisión en Grupo

Un problema de TDG está compuesto por un conjunto de expertos, $E = \{e_1, e_2, \ldots, e_m\}$, que opinan sobre un conjunto de alternativas $X = \{x_1, x_2, \ldots, x_n\}$ evaluadas en base a un conjunto de criterios $C = \{c_1, c_2, \ldots, c_k\}$, con el objetivo de encontrar la mejor solución [8]. Es habitual que las opiniones de los expertos sobre las alternativas se representen mediante relaciones de preferencia. Una relación de preferencia asociada con el experto e_i se nota como $P_i = (p_i^{lk})_{n\times m}$ y se representa en (1) tal que:

$$P_i = \begin{pmatrix} - & \ldots & p_i^{ln} \\ \vdots & \ddots & \vdots \\ p_i^{n1} & \ldots & - \end{pmatrix} \quad (1)$$

Existen distintos tipos relaciones de preferencia según la información que utilicen, las más comunes son:

- *Relación de preferencia difusa* [9]: En una Relación de Preferencia Difusa (RPD), cada valoración $p_i^{lj} = \mu_{P_i}(x_l, x_j) \in [0, 1]$ representa el grado de preferencia de e_i sobre x_l con respecto a x_j, $l, k \in \{1, \ldots, n\}, l \neq j$.
- *Relación de preferencia lingüística* [11]: En una Relación de Preferencia Lingüística (RPL), cada valoración $p_i^{lj} \subseteq X \times X$, con un grado de preferencia $\mu_p : X \times X \to S$, donde S es un conjunto de etiquetas predefinido y $\mu_p(x_l, x_j) = p_{lj}$ representa el grado de preferencia lingüístico de la alternativa x_l sobre x_j.
- *Relación de preferencia dudosa* [19]: En una Relación de Preferencia Dudosa (RPDD), cada valoración $p_i^{lj} \subseteq X \times X$, donde $p_{lj} = \{p_{lj}^{\beta}, \beta = 1, 2, \ldots, \#p_{lj}\}$ ($\#p_{lj}$ es el número de valores en p_{lj}), es un elemento difuso dudoso que indica todos los posibles grados de preferencia de la alternativa x_l sobre x_j.
- *Relación de preferencia lingüísitica dudosa* [13], [14]: En una Relación de Preferencia Lingüística Dudosa (RPLD), cada valoración $p_i^{lj} \subseteq X \times X \to S$, donde S es un conjunto de etiquetas predefinido y $p_{lj} = \{p_{lj}^{\beta} \mid \beta = 1, 2, \ldots, \#p_{lj}\}$ ($\#p_{lj}$ es el número de términos lingüísticos en p_{lj}), es un conjunto de términos lingüísticos difusos que indica todos los posibles grados de preferencia de la alternativa x_l sobre x_j.

II-B. Procesos de Alcance de Consenso

Los PAC son procesos iterativos y dinámicos, en los que los expertos modifican sus opiniones iniciales con el objetivo de acercar posturas y alcanzar un alto nivel de acuerdo sobre las preferencias del problema de TDG después de varias rondas de debate [17]. Un PAC es normalmente guiado por un *moderador* que supervisa el proceso. En problemas de TDG bajo incertidumbre, la lógica difusa desempeña un papel importante al proporcionar modelos de representación y procesamiento de información en opiniones vagas e imprecisas.

Los PAC se incluyen como una fase adicional dentro del esquema de resolución de problemas de TDG, con el objetivo de obtener una solución consensuada que satisfaga a los expertos involucrados en la toma de decisiones. El concepto de *consenso* ha sido definido desde diferentes puntos de vista: (i) A veces como la *unanimidad*, aunque esta visión es poco realista y prácticamente imposible de alcanzar en problemas de TDG reales. (ii) Otras veces se han aportado visiones más flexibles, como la de *soft consensus* o *consenso suave*, basada en el concepto de mayoría difusa definida por Kacprzyk [4], más cercana a la percepción que el ser humano tiene del consenso. Según este concepto, el consenso es alcanzado cuando *la mayoría de los individuos están de acuerdo en la mayoría de las cuestiones más relevantes*.

Las fases que componen un PAC son:

1. *Recolección de preferencias*: Se recoge la opinión de cada experto sobre las alternativas.
2. *Cálculo del nivel de consenso*: Se calcula el nivel de consenso alcanzado por el grupo en la ronda de consenso.
3. *Control de consenso*: Se compara el nivel de consenso obtenido con un valor umbral predefinido que representa el mínimo nivel de consenso que debe de ser alcanzado en el PAC. Si el nivel de consenso alcanzado por el grupo es mayor que el umbral, se inicia el proceso de selección de la mejor alternativa, en caso contrario se iniciará una nueva ronda de consenso.
4. *Generación de recomendaciones*: Cuando no se alcanza el umbral de consenso, es necesario iniciar un proceso en el cual las opiniones de los expertos deben modificarse para así acercar posturas entre ellos. Este proceso será llevado a cabo por el moderador [10], que debe proporcionar sugerencias a los expertos involucrados para que modifiquen sus opiniones iniciales.

II-C. Conjuntos de Términos Lingüísticos Difusos Dudosos

El concepto de CTLDD se introdujo para facilitar la labor de los expertos a la hora de dar sus valoraciones empleando expresiones lingüísticas, especialmente en aquellos casos en los que dudan entre varios términos lingüísticos.

Definición 1: [15] Sea $S = \{s_0, \ldots, s_g\}$ un conjunto de términos lingüísticos, un CTLDD, H_S, se define como un subconjunto finito ordenado de términos lingüísticos consecutivos de S.

$$H_S = \{s_i, s_{i+1}, \ldots, s_j\}, \text{tal que}, s_k \in S, k \in \{i, \ldots, j\} \quad (2)$$

El concepto de CTLDD facilita el modelado de las opiniones de los expertos mediante múltiples términos lingüísticos, sin embargo, no es parecida a cómo los seres humanos expresan sus opiniones. Por esta razón, se presentaron las ELCs [15], dichas expresiones se generan mediante una gramática libre de contexto y son próximas, en términos de expresividad, a la forma de expresarse de los seres humanos.

Definición 2: [15] Sea G_H una gramática libre de contexto y $S = \{s_0, \ldots, s_g\}$ un conjunto de términos lingüísticos. Los elementos de $G_H = (V_N, V_T, I, P)$ son definidos como:

$V_N = \{(\text{término primario}), (\text{término compuesto}),$
$(\text{relación unaria}), (\text{relación binaria}),$
$(\text{conjunción})\}$
$V_T = \{\text{menor que}, \text{mayor que}, \text{entre}, \text{y}, s_0, s_1, \ldots, s_g\}$
$I \in V_N$

Las reglas de producción son definidas mediante la forma extendida de Backus Naur, en la que la utilización de paréntesis indica elementos opcionales y el símbolo | indica elementos alternativos.

$P = \{I ::= (\text{término primario})|(\text{término compuesto})$
$(\text{término compuesto}) ::= (\text{relación unaria})$
$(\text{término primario})|(\text{relación binaria})(\text{término primario})$
$(\text{conjunción})(\text{término primario})$
$(\text{término primario}) ::= s_0|s_1|\ldots|s_g$
$(\text{relación unaria}) ::= \text{menor que}|\text{mayor que}$
$(\text{relación binaria}) ::= \text{entre}$
$(\text{conjunción}) ::= \text{y}\}$

Dichas expresiones lingüísticas comparativas pueden ser representadas mediante CTLDD utilizando la función E_{G_H}.

Definición 3: [15] Sea E_{G_H} una función de transformación entre expresiones lingüísticas ll, $ll \in G_H$, y H_s, siendo S el conjunto de términos lingüísticos utilizados por G_H:

$$E_{G_H} : ll \rightarrow H_S$$

Dependiendo del tipo de ELC, su transformación en CTLDD se lleva a cabo de diferentes formas:

- $E_{G_H}(s_i) = \{s_i / s_i \in S\}$
- $E_{G_H}(\text{menor que } s_i) = \{s_j / s_j \in S \text{ y } s_j \leq s_i\}$
- $E_{G_H}(\text{mayor que } s_i) = \{s_i / s_i \in S \text{ y } s_j \geq s_i\}$
- $E_{G_H}(\text{entre } s_i \text{ y } s_j) = \{s_i / s_i \in S \text{ y } s_i \leq s_k \leq s_j\}$

III. Expresiones Lingüísticas Comparativas en AFRYCA 3.0

En esta sección se muestran las nuevas funcionalidades de la versión 3.0 de AFRYCA relacionadas con el modelado de información lingüística difusa dudosa mediante ELCs. Primero se presenta la arquitectura software de AFRYCA 3.0, y después el esquema de resolución del framework. Por último, se analizan las nuevas funcionalidades en detalle.

III-A. Arquitectura de AFRYCA 3.0

AFRYCA es una aplicación software desarrollada bajo la plataforma *Eclipse Rich Client Platform* (RCP), una plataforma para construir y desarrollar aplicaciones multiplataforma ricas de escritorio fáciles de mantener y extender. Estas características vienen determinadas por el tipo de arquitectura software sobre la que se construye el framework. Dicha arquitectura *basada en componentes* está diseñada para resolver los problemas más comunes en el desarrollo software, como son el mantenimiento, reutilización, ampliación y modificación. Los componentes, también denominados *plug-ins*, son pequeñas porciones de software que se conectan entre si formando una aplicación software.
Anteriores versiones de AFRYCA se desarrollaron empleando la rama estable de Eclipse RCP, 3.X. El uso de esta rama proporcionaba ciertas ventajas, como una mayor estabilidad, documentación y compatibilidad con componentes de terceros. Sin embargo, en términos tecnológicos, AFRYCA no presentaba ninguna evolución. Este es el motivo por el que AFRYCA 3.0 ha sido desarrollada bajo la nueva rama 4.X de Eclipse RCP, e4. Esta evolución permite que AFRYCA 3.0 se beneficie de nuevas tecnologías orientadas al desarrollo de aplicaciones bajo una arquitectura basada en componentes como la inyección de dependencias o el uso de servicios declarativos.
AFRYCA 3.0 está compuesta por más de 100 componentes, todos desarrollados bajo la rama 4.X (Fig. 1).

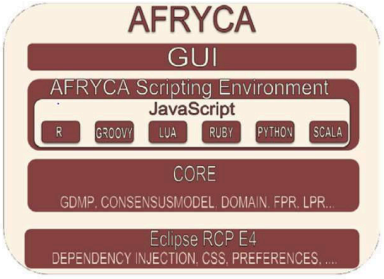


Figura 1. Arquitectura de AFRYCA 3.0

Visitar la web de AFRYCA para más información[1].

III-B. Esquema de Resolución de AFRYCA 3.0

El esquema de resolución de AFRYCA para problemas de TDG se basa en diferentes fases que son descritas a continuación:

[1]http://sinbad2.ujaen.es/afryca/es

1. *Definición del problema de TDG*: Aquí se define el problema de TDG, incluyendo *expertos, alternativas* y *criterios*. Además, también se recogen las opiniones de los expertos, que pueden emplear diferentes tipos de relaciones de preferencia. AFRYCA 3.0 da soporte a RPD, RPL, RPDD y RPLD.
2. *Selección del modelo de consenso*: El siguiente paso es seleccionar el modelo de consenso que se va a emplear en la simulación del PAC. AFRYCA 3.0 tiene incorporados un conjunto de modelos de consenso separados en dos grandes grupos, aquellos que utilizan un proceso de generación de recomendaciones y los que no.
3. *Configuración de los parámetros del modelo de consenso*: Una vez seleccionado el modelo de consenso, es indispensable configurar sus parámetros. Los parámetros de un modelo de consenso influyen directamente en el resultado de la simulación y deben ser ajustados en base al problema de TDG que se esté tratando.
4. *Selección del comportamiento de expertos*: Para llevar a cabo la simulación del PAC, AFRYCA simula el comportamiento de los expertos que participan en el proceso. Actualmente, AFRYCA 3.0 incluye dos tipos de patrones de comportamientos, *estándar* y *estándar con oposición*. En el patrón de comportamiento estándar los expertos pueden *aceptar* o *rechazar* las recomendaciones. En el caso del patrón de comportamiento estándar con oposición; los expertos pueden *aceptar*, *rechazar* o *ir en contra* de las recomendaciones sugeridas.
5. *Configuración de los parámetros del comportamiento*: Al igual que en los modelos de consenso, es posible configurar los parámetros de un patrón de comportamiento. Los patrones de comportamiento se simulan siguiendo una distribución de probabilidad binomial. La configuración de los parámetros de la distribución determinan el comportamiento de los expertos, por ejemplo, definiendo la probabilidad de que un experto acepte una recomendación. Es obvio que la evolución de un PAC será muy distinta si el patrón de comportamiento seleccionado define una probabilidad muy alta de aceptar la recomendación frente a una muy baja.

Nota: En el caso de seleccionarse un modelo de consenso que no aplique un proceso de generación de recomendaciones, no es necesario llevar a cabo los dos últimos pasos para comenzar la simulación.

III-C. Soporte de Información Lingüística Difusa Dudosa en AFRYCA

Posibilitar el modelado de información lingüística dudosa en AFRYCA y el uso de ELCs, ha requerido del desarrollo de nuevos componentes software que han sido incorporados y conectados con el resto de componentes que componen el framework. Estos nuevos componentes han dotado a AFRYCA de nueva funcionalidad que será explicada de forma detallada a continuación.

- *Soporte de conjuntos de términos lingüísticos difusos*: AFRYCA 3.0 permite la creación de conjuntos de términos lingüísticos difusos [21]. Los términos lingüísticos de los que se componen el conjunto, son representados mediante etiquetas lingüísticas representadas por funciones de pertenencia triangulares. AFRYCA 3.0 permite la creación de este tipo de conjuntos con cualquier número de etiquetas lingüísticas. Un ejemplo de creación de un conjunto de este tipo se muestra en la Fig. 2.

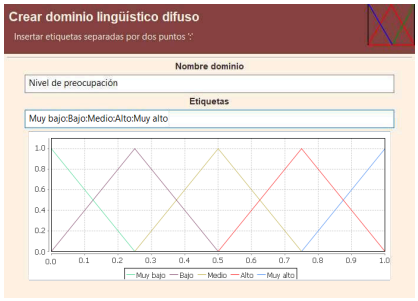


Figura 2. Creación de un conjunto de términos lingüísticos difusos

- *Soporte de relaciones de preferencia lingüísticas dudosas*: AFRYCA modela RPLD cuyos valores están representados por ELCs (ver Fig. 3). Para la generación de las ELCs es indispensable previamente definir un conjunto de términos lingüísticos difusos con el que formar las expresiones. Además, AFRYCA también puede llevar a cabo la transformación de ELCs a CTLDD.

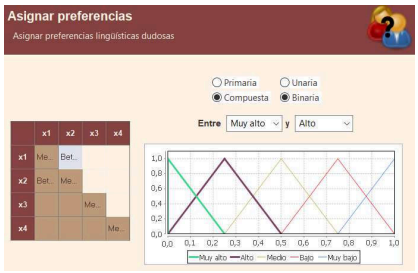


Figura 3. Relación de preferencia lingüística difusa dudosa

- *Modelos difusos de consenso*: La capacidad de tratar información lingüística difusa dudosa en AFRYCA abre un abanico de nuevas posibilidades en cuanto modelos de consenso se refiere. Ahora es posible incluir modelos de consenso que lleven a cabo operaciones con ELCs, lo que permite realizar nuevas comparaciones entre modelos y estudios sobre su funcionamiento. En relación a esto, se ha incorporado un nuevo modelo de consenso [12] que emplea ELCs.
- *Visualización de RPLD*: Una de las principales características de AFRYCA es la visualización de las preferencias de los expertos a los largo del PAC simulado. El incorporar un nuevo tipo de información y por tanto de relación de preferencia, conlleva a incluir nuevas formas de visualización. AFRYCA emplea dos técnicas para visualizar las preferencias de los expertos que son el *escalamiento multidimensional* [6] y la técnica de *análisis*

de componentes principales [18]. Estas dos técnicas han tenido que ser adaptadas para representar la información lingüística difusa dudosa a partir de RPLD.

IV. Ejemplo Ilustrativo

Con el objetivo de mostrar el funcionamiento de las nuevas funcionalidades de AFRYCA 3.0, relacionadas con la representación y computación de ELCs, esta sección presenta un problema de TDG en el cual los expertos expresan sus opiniones a través de ELCs y que empleará un modelo difuso de consenso orientado a trabajar con ELCs [12].

Supongamos el entrenador de la Selección Nacional de Fútbol que debe realizar la convocatoria de aquellos jugadores que participarán en la Copa del Mundo de 2018 que se celebrará en Rusia. Al ser una decisión tan compleja y delicada, el entrenador no la tomará solo y será aconsejado por sus tres asistentes, tomando la decisión final un total de 4 profesionales, $E = \{e_1, e_2, e_3, e_4\}$. Después de disputar todos los partidos de clasificación a la Copa del Mundo el seleccionador y sus asistentes tienen clara la mayor parte de la convocatoria, únicamente queda una plaza vacante que la disputan 4 jugadores, $A = \{a_1, a_2, a_3, a_4\}$. De los jugadores únicamente será evaluada su técnica con el balón $C = \{c_1\}$. Debido a la complejidad de la decisión y a la duda que genera, el entrenador y sus asistentes proporcionan sus preferencias sobre los jugadores usando RPLD cuyos valoraciones están representadas por ELCs. Los técnicos usarán el conjunto de términos lingüísticos representado en la Fig. 4 para generar las ELCs y valorar la calidad de los futbolistas.

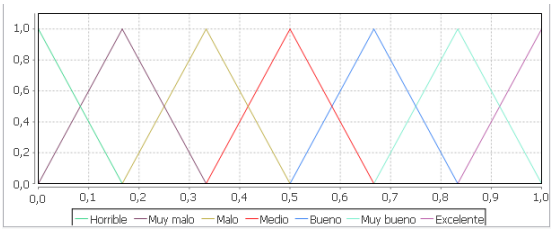


Figura 4. Conjunto de términos lingüísticos

Para resolver de forma consensuada el problema de TDG planteado, seguiremos el esquema de resolución de AFRYCA introducido en la Sección III-B. Se define el problema de TDG junto con todos sus elementos, el conjunto de expertos $E = \{e_1, e_2, e_3, e_4\}$, alternativas $A = \{a_1, a_2, a_3, a_4\}$ y criterios, que en este caso solo es uno que representa la calidad del futbolista $C = \{c_1\}$. También se definen las preferencias de los expertos mediante RPLD. La Fig. 5 ilustra la definición del problema empleando AFRYCA.

- *Definición del problema de TDG*
- *Selección del modelo de consenso* Una vez definido el problema, el siguiente paso es seleccionar el modelo de consenso. Para este problema se selecciona el modelo de consenso difuso presentado por Rodríguez et al. [12] (ver Fig 6), centrado en el tratamiento de CTLDDs donde los expertos expresan sus opiniones a partir de ELCs.

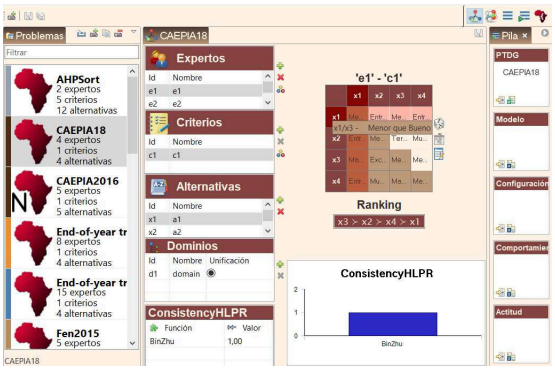


Figura 5. Definición del problema

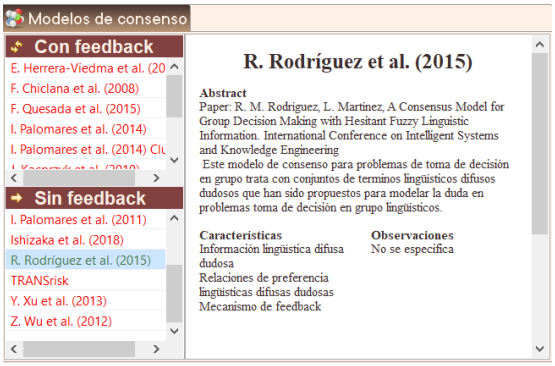


Figura 6. Selección del modelo de consenso

- *Configuración de los parámetros del modelo de consenso* Cuando se selecciona un modelo de consenso, se deben configurar sus parámetros junto con dos adicionales que son incluidos en todos los modelos de consenso, éstos son el umbral de consenso y el máximo número de rondas permitidas en el PAC (mu y h_{max} en Fig. 7). La configuración de los parámetros del modelo seleccionado se representa en la Fig. 7.

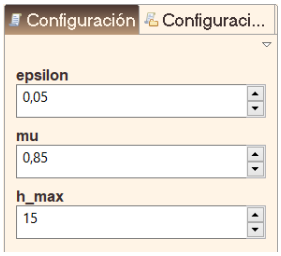


Figura 7. Configuración de parámetros

Ya que el modelo de consenso seleccionado no emplea un mecanismo de generación de recomendaciones, no es necesario llevar a cabo ningún paso más para comenzar con la simulación del PAC y obtener una solución del problema.

Una vez se han completado todos los pasos previamente definidos, se lleva a cabo la simulación del PAC. El resultado

se muestra en la Fig. 8. Además, AFRYCA permite visualizar la evolución de las preferencias de los expertos a lo largo del PAC. Esta visualización, empleando para ello la técnica de escalamiento multidimensional, se muestra en la Fig. 9. Notar que en este tipo de visualización siempre se representa la opinión colectiva del grupo en el centro del eje de coordenadas.

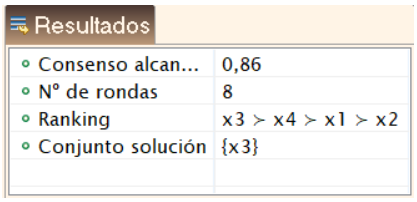


Figura 8. Resultado de la simulación

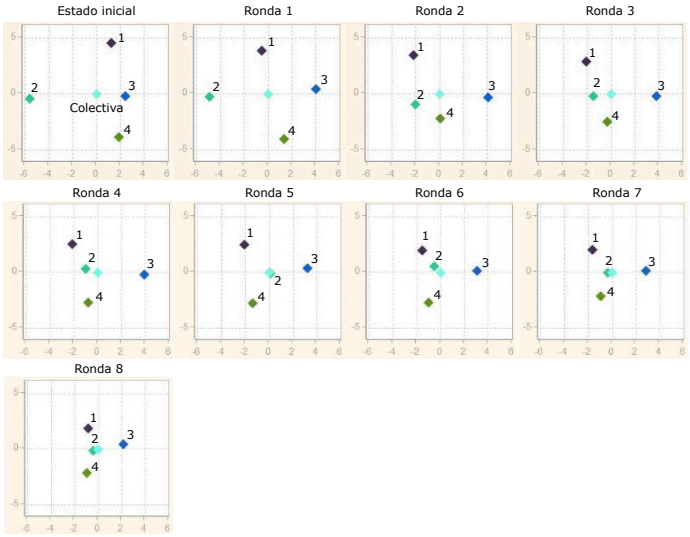


Figura 9. Evolución de las preferencias de los expertos

En la Fig. 8 se muestra que el consenso final alcanzado por los expertos es 0.86, y se han necesitado de un total de 8 rondas para alcanzar el grado de consenso fijado. La Fig. 9 también nos permite ver la evolución de los expertos a lo largo del PAC y como van acercando posturas en las sucesivas rondas de debate. Este acercamiento de posturas puede apreciarse fácilmente, los expertos modifican sus preferencias con el objetivo de llegar a un consenso con el resto del grupo, lo que conlleva a que su opinión esté cada vez más próxima a la opinión colectiva del grupo y su posición en el gráfico tienda hacia el centro. Finalmente, si se aplica un proceso de selección a este consenso, el jugador que debería ser convocado es x_3.

V. Conclusiones

AFRYCA es una herramienta software que permite la resolución de problemas de TDG mediante la simulación de PAC, diseñada bajo una arquitectura basada en componentes cuyas principales ventajas son la reusabilidad y la inclusión de nuevas características. En esta contribución se han presentado nuevas funcionalidades de AFRYCA relacionadas con el manejo de información lingüística difusa dudosa y más concretamente con ELCs. Además se ha descrito tanto la arquitectura como el esquema de resolución del framework. Este conjunto de nuevas funcionalidades abre camino a futuras mejoras en AFRYCA como la inclusión de nuevos modelos difusos de consenso que empleen información lingüística y la incorporación de nuevos tipos de expresiones lingüísticas.

Referencias

[1] S. Alonso, I.J. Pérez, F.J. Cabrerizo, and E. Herrera-Viedma. A linguistic consensus model for web 2.0 communities. *Applied Soft Computing*, 13(1):149–157, 2013.

[2] F. Herrera, E. Herrera-Viedma, and J.L. Verdegay. A sequential selection process in group decision making with linguistic assessments. *Information Sciences*, 85(4):223–239, 1995.

[3] E. Herrera-Viedma, F.J. Cabrerizo, J. Kacprzyk, and W. Pedrycz. A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17:4–13, 2014.

[4] J. Kacprzyk. Group decision making with a fuzzy linguistic majority. *Fuzzy Sets and Systems*, 18(2):105–118, 1986.

[5] Janusz Kacprzyk and Sławomir Zadrożny. Supporting consensus reaching processes under fuzzy preferences and a fuzzy majority via linguistic summaries. In *Preferences and Decisions*, pages 261–279. Springer, 2010.

[6] J. B. Kruskal and M. Wish. *Multidimensional scaling*, volume 11. Sage, 1978.

[7] Á. Labella, F. J. Estrella, and L. Martínez. Afryca 2.0: an improved analysis framework for consensus reaching processes. *Progress in Artificial Intelligence*, pages 1–14, 2017.

[8] J. Lu, G. Zhang, D. Ruan, and F. Wu. *Multi-Objective Group Decision Making*. Imperial College Press, 2006.

[9] S.A. Orlovsky. Decision-making with a fuzzy preference relation. *Fuzzy Sets and Systems*, 1(3):155–167, July 1978.

[10] I. Palomares, F.J. Estrella, L. Martínez, and F. Herrera. Consensus under a fuzzy context: Taxonomy, analysis framework AFRYCA and experimental case of study. *Information Fusion*, 20:252–271, 2014.

[11] R. M. Rodríguez, M. Espinilla, P. J. Sánchez, and L. Martínez. Using linguistic incomplete preference relations to cold start recommendations. *Internet Research*, 20(3):296–315, 2010.

[12] R. M. Rodríguez and L. Martínez. A consensus model for group decision making with hesitant fuzzy linguistic information. In *Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on*, pages 540–545. IEEE, 2015.

[13] R. M. Rodríguez, L. Martínez, and F. Herrera. A group decision making model dealing with comparative linguistic expressions based on hesitant fuzzy linguistic term sets. *Information Sciences*, 241(1):28–42, 2013.

[14] R.M. Rodríguez, A. Labella, and L. Martínez. An overview on fuzzy modelling of complex linguistic preferences in decision making. *International Journal of Computational Intelligence Systems*, 9:81–94, 2016.

[15] R.M. Rodriguez, L. Martinez, and F. Herrera. Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, 20(1):109–119, 2012.

[16] M. Roubens. Fuzzy sets and decision analysis. *Fuzzy Sets and Systems*, 90(2):199–206, 1997.

[17] S. Saint and J. R. Lawson. *Rules for Reaching Consensus. A Modern Approach to Decision Making*. Jossey-Bass, 1994.

[18] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[19] M. Xia and Z. Xu. Managing hesitant information in GDM problems under fuzzy and multiplicative preference relations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 21(06):865–897, 2013.

[20] J. Yang, J. Liu, J. Wang, H. Sii, and W. Hong-Wei. Belief rule-base inference methodology using the evidential reasoning approach-rimer. *IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans*, 36(2):266–285, 2006.

[21] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[22] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning—i. *Information sciences*, 8(3):199–249, 1975.

Ajuste a una progresión hipotética de aprendizaje

Estudio de 3 casos de estudiantes de Magisterio

Itziar García-Honrado
Statistics, O. R. and Mathematics Didactics,
University of Oviedo,
Oviedo, Spain
garciaitziar@uniovi.es

***Abstract*— Esta comunicación presenta un modelo matemático a través del uso de técnicas borrosas para describir el ajuste de los alumnos a progresiones hipotéticas de aprendizaje, parte fundamental de las conocidas Trayectorias Hipotéticas de Aprendizaje. En concreto, se hace uso del modelo granular lingüístico de un fenómeno. Para su implementación se distinguen distintos niveles de consecución de cada una de las etapas indicadas en la progresión hipotética de aprendizaje.**

Keywords*— *evaluación; resumen lingüístico; trayectoria hipotética de aprendizaje

I. INTRODUCCIÓN

El constructo de Trayectoria Hipotética de Aprendizaje (THA) pertenece al ámbito de la educación matemática, nace de la mano de Simon en 1995 [4], y permite al profesor de matemáticas modelar las distintas escenarios por los que sus estudiantes hipotéticamente trascurren para la consecución de un objetivo de aprendizaje fijado a priori. Para ello, una vez fijado el objetivo, se diseña unas tareas y se muestran los efectos esperados en los estudiantes hasta conseguir el objetivo. Dichos efectos, divididos en distintas etapas, constituirán una progresión hipotética de aprendizaje.

En posteriores estudios [5], Simon y Tzur experimentan en un aula el ajuste de los alumnos a la progresión hipotética de aprendizaje, mostrando cuando un alumno cumple o no cumple el efecto hipotético esperado en cada una de las etapas.

En la presente comunicación, proponemos realizar un estudio de este ajuste en términos imprecisos. Para ello, mostraremos el diseño de una THA y lo implementaremos en el aula, como se ha hecho en [1]. Por otro lado, diseñaremos unas rúbricas de evaluación que nos permita distinguir distintos niveles de consecución de la progresión hipotética de aprendizaje, así como el diseño de los elementos del modelo granular lingüístico que permitirán obtener una descripción lingüística del ajuste a la progresión hipotética de aprendizaje diseñada a priori de los alumnos. Finalmente, a modo de ejemplo, se mostrará el modelo desarrollado para valorar el ajuste de 3 alumnos a la progresión hipotética de aprendizaje.

II. CONCEPTOS BÁSICOS

A. Progresión hipotética de aprendizaje

En una THA, se distinguen los siguientes componentes: el objetivo de aprendizaje, las tareas de aprendizaje y el proceso de aprendizaje hipotético; es decir, la predicción de cómo el pensamiento y la comprensión de los estudiantes evolucionan en el contexto de las actividades de aprendizaje [4]. Se destaca especialmente la importancia de la tarea, su diseño, los objetivos de aprendizaje, los conocimientos previos de los estudiantes y la progresión hipotética de aprendizaje. Así mismo, se considera relevante reconocer los efectos de la secuencia de tareas en la progresión de aprendizaje de la futura maestra [5].

A través de la THA, se muestran diferentes escenarios por los que los estudiantes pueden pasar transformando sus ideas intuitivas a una comprensión más formal de los conceptos matemáticos. Los niveles cada vez más sofisticados de razonamiento matemático evidencian aspectos de la progresión de la comprensión de un contenido específico. Por todo lo anterior, se considera que las THA constituyen un recurso instructivo muy eficaz para la enseñanza. La THA proporciona al profesorado recursos relevantes para acompañar a cada estudiante en el logro de su aprendizaje. Indican también el estado actual de aprendizaje del alumnado a través de una secuencia de estados de aprendizaje.

B. Modelo granular lingüístico de un fenómeno (GLMP)

Este modelo se sustenta dentro del campo de la lógica fuzzy. En consecuencia, no solo permite determinar si existe o no el ajuste de un alumno a la progresión hipotética existe, sino que permite asignar distintos grados de cumplimiento del ajuste.

Para determinar el ajuste utilizaremos el concepto de variable lingüística [8], entendida como aquella que puede tomar valores lingüísticos como Muy bajo / Bajo / Medio / Alto / Muy alto. Cada uno de ellos se representará, en nuestro caso, por un conjunto borroso continuo triangular,

debido a su simplicidad en la función (lineal definida por trozos) y a la posibilidad de ser definida por sus vértices. Al conjunto de los conjuntos borrosos que representa los valores que puede tomar la variable se le llama partición borrosa (Figura 1).

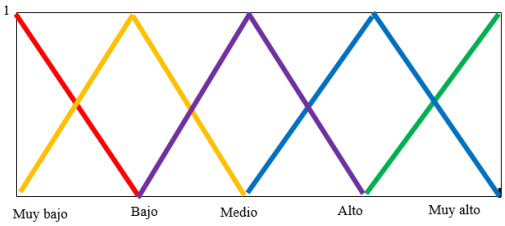


Figura 1: Conjuntos borrosos modelando las etiquetas lingüísticas

Combinando los problemas de control borroso con el actual auge de la conocida computación con palabras [9], surge el paradigma Modelo Granular Lingüítico de un fenómeno, o, en inglés, *Granular Linguistic Model of a Phenomenon* (GLMP) [6], el cual consiste en hacer resúmenes lingüísticos de fenómenos complejos como puede ser el ajuste de un estudiante a la progresión hipotética de aprovechamiento. En la literatura se pueden encontrar otras aplicaciones de este paradigma en la evaluación del aprendizaje de los alumnos [3].

III. Metodología

A. *Diseño de la THA.*

En esta investigación hemos construido la siguiente THA cuyo objetivo de aprendizaje es reconocer la generalización de patrones en la construcción de la secuencia geométrica de las figuras que aparecen en la obra artística "Composición aritmética" de Theo van Doesburg que puede observarse en la Figura 2. La actividad tiene un carácter competencial aunando arte y matemáticas, en concreto se eligió dicha obra, por las regularidades matemáticas que en ella aparecen [2,7] idóneas la búsqueda de patrones gemétrico y su generalización.

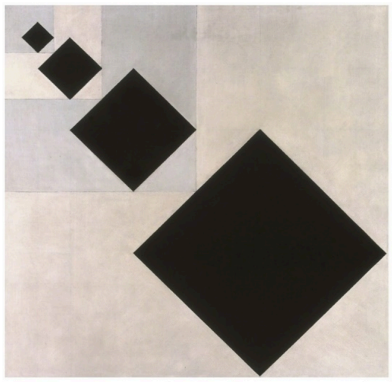


Figura 2. *Arithmetic Composition*. Theo van Doesburg, 1929-1930.

Las tareas y la progresión hipotética de aprendizaje se describen en la Tabla 1 y 2, respectivamente.

Tabla1. Tareas de la THA

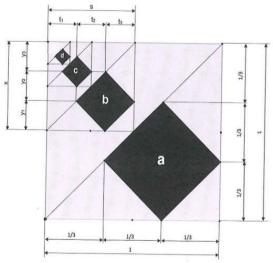
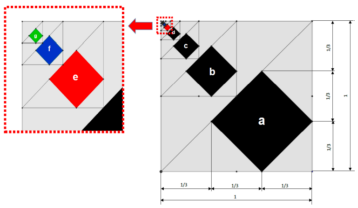
Tarea de aprendizaje	
Tarea 1. (Relacionada con la Etapa 1)	A partir de la obra de arte de la Figura 1, se ha construido la imagen que se muestra a continuación (la información que se incluye puede ayudar a resolver las tareas 5, 6 y 7).  Completa los datos que faltan en la figura, encontrando los valores de x, y1, y2, y3, s, t1, t2, t3. Justifica tu respuesta.
Tarea 2. (Relacionada con la Etapa 2)	(a) En el cuadro gris grande de la Figura 1, ¿cuántas figuras "a" pueden colocarse sin superponerse? (b) Dentro de la figura negra "a", ¿cuántas figuras negras "b" podrías colocar?
Tarea 3. (Relacionada con la Etapa 3)	En la Figura 1, (a) Calcula el área de la figura negra "a". (b) Calcula el área de la figura negra "b". (c) Calcula el área de las figuras negras "c" y "d".
Tarea 4. (Relacionada con la Etapa 4)	Si continuáramos la secuencia de las figuras negras podríamos dibujar otras, tal y como se muestra en la ampliación siguiente.  ¿Sabrías calcular las áreas de las nuevas figuras "e", "f" y "g"? Explica cómo lo has hecho.

Tabla 2. Progresión hipotética de aprendizaje.

Progresión hipotética de aprendizaje	
Etapa 1	Coordinación de la estructura espacial y la estructura numérica (a partir de las cotas) en la identificación de la secuencia de longitudes de los lados de los cuadrados grises involucrados en la obra artística.

Etapa 2	Coordinación de la estructura espacial y la numérica de la secuencia de los cuadrados, estableciendo relaciones entre las áreas de los cuadrados de la Figura 1.
Etapa 3	Cálculo del área de la sucesión de los cuadrados negros, relacionando el valor del área de un cuadrado con el valor del área del anterior, sigue un proceso de generalización cercana [Estadio1 de [10]].
Etapa 4	Abstracción de la coordinación espacial y numérica para el establecimiento de la relación funcional entre el área de los cuadrados negros [Estadio 2 de [10]].

B. Diseño de la rúbrica de evaluación

Distinguiremos distintos niveles en el logro de cada una de las etapas de la progresión hipotética de aprendizaje: no se logra (N0), se logra a nivel bajo (N1), a nivel medio (N2) o a nivel alto (N3).

Para ello se elabora una rúbrica de evaluación en la que se describe cada uno de estos niveles en las distintas etapas.

- Respecto a la Etapa 1:
  - N0: No identifica las longitudes de los lados de los cuadrados grises involucrados en el cuadro.
  - N1: Identifica parcialmente las longitudes de los lados de los cuadrados grises involucrados en el cuadro.
  - N2: Identifica la secuencia de longitudes de los lados de los cuadrados grises, pero no muestra coordinación entre la estructura espacial y la numérica.
  - N3: Identifica la secuencia de longitudes de los lados de los cuadrados grises mostrando coordinación entre la estructura espacial y la numérica.
- Respecto a la Etapa 2:
  - N0: No establece relaciones entre el área de los cuadrados negros del cuadro.
  - N1: Establece relaciones, únicamente, entre las áreas de una pareja de cuadrados del cuadro.
  - N2: Establece relaciones entre las áreas de los cuadrados del cuadro, pero no muestra coordinación entre la estructura espacial y la numérica.
  - N3: Establece relaciones entre las áreas de los cuadrados del cuadro mostrando coordinación entre la estructura espacial y la numérica.
- Respecto a la Etapa 3:
  - N0: No evidencia el cálculo de ningún área ni la relación que le permite obtener el área de uno a partir del área del anterior.
  - N1: Calcula las áreas de algunos de los cuadrados de la sucesión.
  - N2: Calcula las áreas de los cuadrados de la sucesión y muestra la relación que permitirá obtener el proceso de generalización cercana.
  - N3: Calcula las áreas de los cuadrados de la sucesión, consiguiendo el valor del área de un cuadrado con el valor del área del anterior, es decir desarrolla el proceso de generalización cercana.
- Respecto a la Etapa 4:
  - N0: No consigue llegar a una relación entre la posición que ocupan los cuadrados y su área.
  - N1: Llega a una relación recursiva entre la posición que ocupa el cuadro y su área, es decir, basándose exclusivamente en el área del anterior.
  - N2: Consigue una obtener una relación entre la posición que posición que ocupan los cuadrados y su área, aunque no se utilice lenguaje matemático.
  - N3: Establece una relación funcional entre la posición que ocupan los cuadrados y su área.

Es posible que no siempre sepamos con certeza que un alumno está en un determinado nivel en una etapa, podemos permitir que un alumno verifique características de dos o más niveles y demos como resultado final una media ponderada del nivel en el que se encuentra. Los niveles mostrados en la rúbrica se representan por conjuntos fuzzy, en este caso, triangulares. Asimilaremos N0 a la etiqueta lingüística nulo, N1 a bajo, N2 a medio y N3 a alto.

C. Diseño del modelo granular lingüístico

Las rúbricas diseñadas formarán parte de los inputs de los sistemas de reglas fuzzy que se definen con el objetivo de conseguir el resumen lingüístico del ajuste de un alumno a la progresión hipotética de aprendizaje.

A partir de las rúbricas se definirán las percepciones de primer orden que serán las Etapas: 1, 2, 3 y 4.
Las etapas 1 y 2 se agregarán mediante un sistema de reglas, formando la percepción de segundo orden: Ajuste a los conceptos previos para el acercamiento a la relación entre los cuadrados de una sucesión. Dicha variable tomará 5 etiquetas lingüísticas Muy bajo / Bajo / Medio / Alto / Muy alto. Concretamente, el sistema se compondrá de 8 reglas:

1. Si en la Etapa 1 el nivel alcanzado por el alumno es nulo, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es muy bajo.

2. Si en la Etapa 2 el nivel alcanzado por el alumno es nulo, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es muy bajo.
3. Si en la Etapa 1 el nivel alcanzado por el alumno es bajo y en la Etapa 2 el nivel alcanzado por el alumno es bajo, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es bajo.
4. Si en la Etapa 1 el nivel alcanzado por el alumno es bajo y en la Etapa 2 el nivel alcanzado por el alumno es medio, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es medio.
5. Si en la Etapa 1 el nivel alcanzado por el alumno es medio y en la Etapa 2 el nivel alcanzado por el alumno es bajo, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es medio.
6. Si en la Etapa 1 el nivel alcanzado por el alumno es medio y en la Etapa 2 el nivel alcanzado por el alumno es medio, el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es alto.
7. Si en la Etapa 1 el nivel alcanzado por el alumno es alto, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es muy alto.
8. Si en la Etapa 2 el nivel alcanzado por el alumno es alto, entonces el ajuste de los conceptos previos a la progresión hipotética de aprendizaje es muy alto.

El método de resolución del sistema de reglas será el método Mamdani aplicando el producto a las sentencias Si/Entonces. La Figura 3 muestra la superficie de control del sistema de reglas.

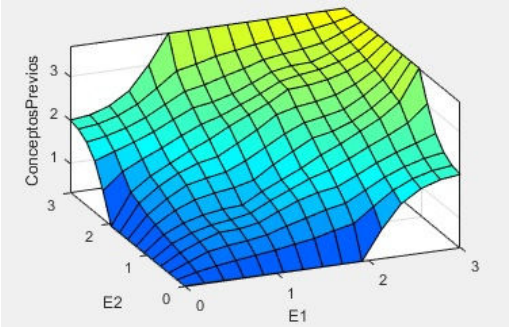


Fig.3. Superficie de control para el sistema de reglas descrito.

Las etapas 3 y 4 se agregan mediante un sistema de 8 reglas, análogo al anterior, para originar la percepción de segundo orden: Ajuste al proceso de generalización según Zapatera y Callejo [10] (cercana y funcional). Dicha variable se define a través de 5 etiquetas lingüísticas del mismo modo que la anterior.

Finalmente, a través de otro sistema de 13 reglas, similar a los anteriores, a partir de las dos percepciones de segundo orden generadas, se definirá la percepción general sobre el ajuste a la progresión hipotética de aprendizaje de la progresión que haya seguido el alumno.

Todos los sistemas de reglas involucrados se han implementado en Matlab. Ese engranaje permite la construcción de una descripción lingüística del ajuste de cada alumno a la progresión hipotética de aprendizaje, del siguiente modo:

El alumno A se ha ajustado a un nivel Muy bajo/Bajo/ Medio/Alto/Muy alto a la progresión hipotética de aprendizaje en las tareas encaminadas a trabajar los conceptos previos para el acercamiento a la relación entre los cuadrados de una sucesión. En el caso del proceso de generalización, el ajuste con la progresión hipotética de aprendizaje ha sido Muy bajo/Bajo/ Medio/Alto/Muy alto. Con todo ello, podemos hablar de un ajuste global a la progresión hipotética de aprendizaje Muy bajo/Bajo/ Medio/Alto/Muy alto.

Además, se asignará un valor numérico al ajuste global en el continuo [0,4] donde 0 representará el menor y 4 el mayor ajuste posible.

IV. Contexto de la experimentación

El experimento de enseñanza se llevó a cabo en una sesión de 2 horas en un aula con estudiantes del Grado de Maestro el pasado curso, 2017-2018, en la Universidad de Oviedo.

Las tareas que se recoge en este trabajo pertenecen a la segunda sesión y parte central de una secuencia de actividades relacionadas con la obra de arte de Theo van Doesburg realizadas a lo largo de tres sesiones.

Participaron 33 alumnos en el experimento de los cuales se seleccionaron tres casos distintos en los que podemos ver distintos niveles de ajuste a la progresión hipotética de aprendizaje. Nos referiremos a los casos de los alumnos con pseudónimo: Ángela, Andrea y José.

V. Casos seleccionados

A. *Ajuste a la progresión hipotética de aprendizaje de Ángela*

Ángela realiza exclusivamente las primeras actividades de la tarea. Dado el alto nivel alcanzado en las mismas, la falta de las demás tareas podría estar motivada por un inadecuado ajuste del tiempo de realización.

En la Etapa 1, Ángela se encuentra en un nivel alto ya que justifica las medidas pedidas a través de la imagen del cuadro.

Respecto a la Etapa 2, Ángela muestra las relaciones pedidas mostrando una coordinación entre la estructura gráfica y numérica pues se apoya en la imagen para calcular las relaciones como se muestra en la Figura 4. Por lo tanto, el nivel alcanzado en esta etapa es alto.

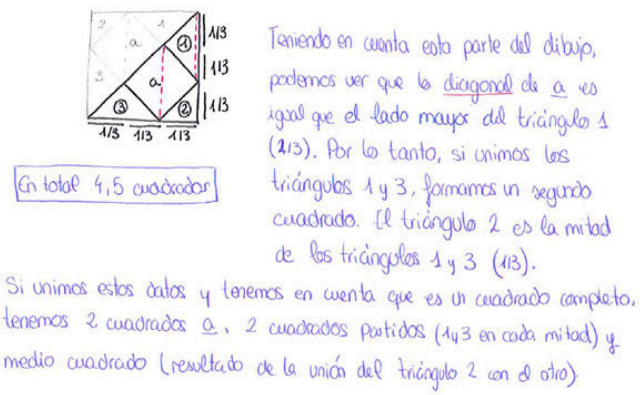


Figura 4. Explicación dada por Ángela justificando la relación entre la figura "a" y el cuadro gris grande.

Sin embrago, en la Etapa 3 se encuentra en un nivel entre bajo y medio, ya que simplemente calcula las áreas de algunos cuadrados negros de la sucesión, no obstante, en este cálculo, muestra la relación que permite obtener el proceso de generalización cercana, ya que relaciona estas respuestas con la relación entre dos cuadrados negros consecutivos. El nivel alcanzado en la Etapa 4, es nulo, pues la alumna no realiza la tarea.

Tras el análisis de la resolución de Ángela, las percepciones de primer nivel en este caso serán: 3 (E1), 3 (E2), 1.5 (E3), y 0(E4). A partir de estos datos, se implementan los sistemas de reglas involucrados. En la Figura 5, se recoge el sistema de reglas que permite obtener la percepción global.

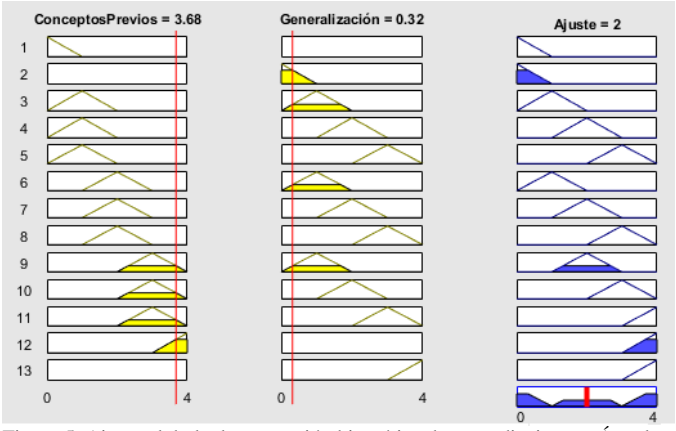


Figura 5. Ajuste global a la progresión hipotética de aprendizaje para Ángela

El resumen lingüístico para el caso de Ángela sería:

Ángela se ha ajustado a un nivel muy alto a la progresión hipotética de aprendizaje en las tareas encaminadas a trabajar los conceptos previos para el acercamiento a la relación entre los cuadrados de una sucesión. En el caso del proceso de generalización, el ajuste con la progresión hipotética de aprendizaje ha sido entre Bajo y Medio. Con todo ello, podemos hablar de un ajuste global a la progresión hipotética de aprendizaje Medio. (2/4)

B. Ajuste a la progresión hipotética de aprendizaje Andrea

Andrea, en la Etapa 1, está en un nivel entre medio y alto, ya que escribe directamente el resultado justificándolo por: "Estos datos los obtengo de observar los datos que ya sabemos". Así que, aunque no hay evidencias de las operaciones que hace para la obtención de los datos, interpretamos que puede existir una coordinación entre la estructura espacial y numérica.

En la Etapa 2, establece la relación que hay entre las áreas del cuadrado "a" y el cuadrado gris grande, a través de operaciones del cálculo de áreas y sumas repetidas del área de "a" hasta aproximarse a 1, obteniendo 4.5. Averigua la relación entre los cuadrados negros a través de razonamientos con la medida de los lados de los mismos. Por lo tanto, establece las relaciones, aunque interpretamos que, a través de la estructura numérica, no de la espacial. Entonces, el nivel es medio.

En la Etapa 3, se asigna un nivel entre Bajo y Medio, ya que calcula las áreas de los cuadrados, pero no muestra la relación que le permite obtener la generalización cercana.

Finalmente, en la Etapa 4, realiza una generalización cercana evidenciando la relación entre cuadrados consecutivos de la sucesión. Por consiguiente, llega a una relación recursiva entre la posición que ocupa el cuadro y su área. Así pues, está en un nivel Bajo.

Con todo esto, las percepciones de primer nivel en este caso serán: 2.5 (E1), 2 (E2), 1.5 (E3), y 1(E4). En la Figura 6, se muestra el último sistema de reglas que permite generar el ajuste global a la progresión hipotética de aprendizaje.

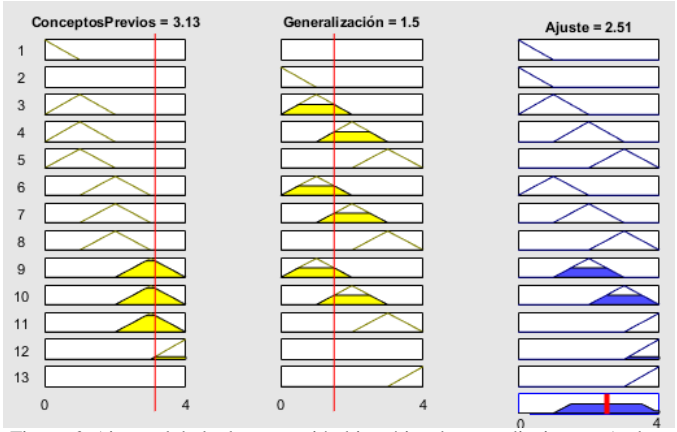


Figura 6. Ajuste global a la progresión hipotética de aprendizaje para Andrea

El resumen lingüístico para el caso de Andrea sería:

Andrea se ha ajustado a un nivel alto a la progresión hipotética de aprendizaje en las tareas previas encaminadas a trabajar los conceptos para el acercamiento a la relación entre los cuadrados de una sucesión. En el caso del proceso

de generalización, el ajuste con la progresión hipotética de aprendizaje ha sido entre Bajo y Medio. Con todo ello, podemos hablar de un ajuste global a la progresión hipotética de aprendizaje entre Medio y Alto. (2.51/4

C. Ajuste a la progresión hipotética de aprendizaje de José

Según la rúbrica de evaluación José alcanza un nivel muy alto en todas las etapas. En la Etapa 1, además de identificar la secuencia de longitudes, coordina las estructuras espacial y numérica justificando gráfica y numéricamente los resultados: "todas son 1/3 de 1/2, por lo tanto 1/6". La forma de proceder en la Etapa 2 es similar a la resolución de Ángela, realiza explicaciones gráficas muy similares a su compañera y las coordina con la estructura numérica, por tanto, se encuentra en un nivel alto.

Respecto a las etapas 2 y 3, José muestra un nivel alto. En la Etapa 3, utiliza la relación entre cuadrados negros para calcular el área del siguiente de la sucesión, por lo tanto desarrolla el proceso de generalización cercana y alcanza un nivel alto. En la Etapa 4, establece la relación funcional entre la posición que ocupan los cuadrados en la sucesión y su área, llegando a la fórmula general que se muestra en la Figura 7.

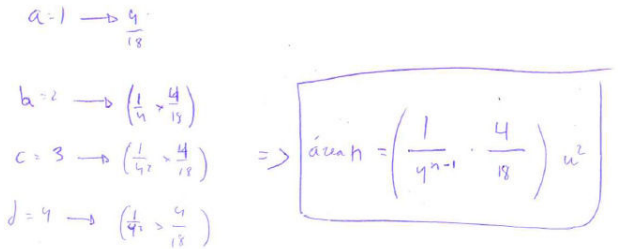


Figura 7. Relación funcional establecida por José.

Así pues, las percepciones de primer nivel en este caso serán: 3 (E1), 3 (E2), 3 (E3), y 3(E4), de las que se infieren las percepciones de segundo nivel (ambas con centroide del nivel: 3.68) y la percepción global con un nivel muy alto (3.65) como puede apreciarse en la Figura 8.

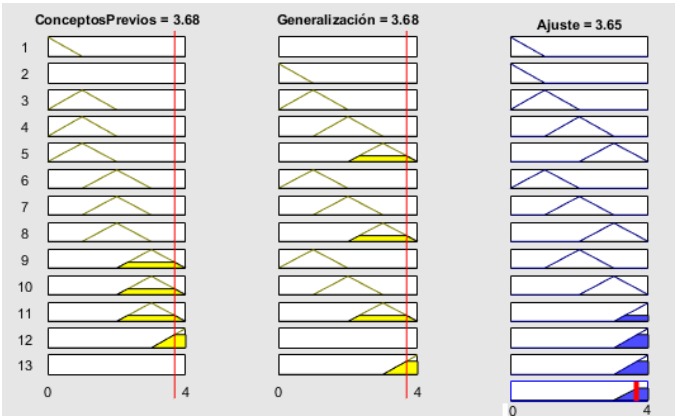


Figura 8. Ajuste global de la progresión hipotética de aprendizaje para José

El resumen lingüístico para este caso es:

José se ha ajustado a un nivel muy alto a la progresión hipotética de aprendizaje en las tareas encaminadas a trabajar los conceptos previos para el acercamiento a la relación entre los cuadrados de una sucesión. En el caso del proceso de generalización, el ajuste con la progresión hipotética de aprendizaje ha sido muy alto. Con todo ello, podemos hablar de un ajuste global a la progresión hipotética de aprendizaje muy alto. (3.65/4)

VI. CONCLUSIONES

En este trabajo, se muestra cómo valorar las resoluciones de alumnos a tareas enmarcadas dentro del marco de las Trayectorias Hipotéticas de Aprendizaje y mostrar el ajuste de las mismas a progresiones hipotéticas de aprendizaje diseñadas a priori. El ajuste, se expresa a partir de términos lingüísticos modelando los distintos modos de ajuste a través de etiquetas lingüísticas. Esto resulta beneficioso en comparación con estudios previos en los que sólo se recogía si había o no ajuste simplificando la riqueza de las resoluciones de los alumnos y no valorando los distintos niveles de adquisición.

AGRADECIMIENTOS

El estudio forma parte del Proyecto EDU2015-65378-P, MINECO.

REFERENCIAS

[1] García-Honrado, I., Clemente, F., Vanegas, Y., Badillo, E. y Fortuny, J.M. "Análisis de la progresión de aprendizaje de una futura maestra" Actas del Congreso XXII SEIEM, aceptado para su publicación, Septiembre 2018.

[2] Pim, D. "Some Notes on Theo van Doesburg (1883-1931) and his Arithmetic Composition 1" *For the learning of Mathematics*, 2001, vol. 21(2), pp. 31-36.

[3] Sanchez-Torrubia, M. G., Torres-Blanc, C. y Triviño, G. "Modelo lingüístico del aprendizaje para la evaluación automática basada en criterios" En F. Bobillo, H. Bustince, F. J. Fernández, y E. Herrera-Viedma (Eds.), *Actas del XVII ESTYLF*, pp. 417-422, Zaragoza, 2014.

[4] Simon, M. A. "Reconstructing mathematics pedagogy from a constructivist perspective" *Journal for Research in Mathematics Education*,1995, vol. 26, pp.114-145.

[5] Simon M.A. y Tzur, R. "Explicating the Role of Mathematical Tasks in Conceptual Learning: An Elaboration of the Hypothetical Learning Trajectory" *Mathematical thinking and learning*, 2004, vol. 6(2), pp.91–104.

[6] Triviño, G. y Sugeno, M "Towards linguistic descriptions of phenomena" *International Journal of Approximate Reasoning*, 2013, vol. *54*(1), pp.22-34.

[7] Walter, M. "Looking at a Painting with a mathematical Eye" *For the learning of Mathematics*,2001, vol. 21(2), pp.26-30.

[8] Zadeh, L. A. " The concept of a linguistic variable and its application to approximate reasoning – III". *Information Sciences,1975, vol. 9*(1), pp. 43-80.

[9] Zadeh, L. A. "Fuzzy logic= computing with words" *Fuzzy Systems, IEEE Transactions on fuzzy systems*, 1996, vol. *4*(2), 103-111.

[10] Zapatera, A. y Callejo, M.L. "Prospective primary teachers' noticing of students' understanding of pattern generalization" J. Math Teacher Education, Springer, 2017, vol. 20, pp. 309-333.

Caracterizando a los usuarios a través de sus opiniones

Jesus Serrano-Guerrero, Francisco P. Romero, Jose Ruiz-Morilla, Jose A. Olivas

Departamento de Tecnologías y Sistemas de Información
Escuela Superior de Informática (UCLM), Ciudad Real, 13071, España
jesus.serrano@uclm.es, joseangel.olivas@uclm.es, franciscop.romero@uclm.es

***Resumen*—El análisis de sentimientos basados en aspectos permite detectar comportamientos de ciertos usuarios. Estos comportamientos necesitan ser modelado para poder entenderse la opinión final sobre un producto o servicio. Este artículo propone el modelado de cada opinión final en base a la agregación borrosa de las distintas opiniones sobre los distintos aspectos que conforman el producto opinado. Igualmente propone algunas aplicaciones para dicho aplicaciones.**

Palabras clave—análisis de sentimientos basado en aspectos; operadores OWA; perfil de usuario

I. Introduccion

Cada vez es más frecuente el uso de servicios, especialmente en Internet gracias a plataformas como Amazon o Tripadvisor, que permiten expresar opiniones a los usuarios sobres distintos productos. Detrás de estas opiniones se pueden ocultar distintos comportamientos que es necesario que sean caracterizados con el fin de entender los distintos usuarios y las situaciones en las cuales hacen un comentario determinado u otro.

El campo del Análisis de Sentimientos se centra principalmente en detectar posibles expresiones que sean susceptibles de ser interpretadas como opiniones y cuantificar el grado en el que cada opinión sea más o menos positiva, entre otras tareas. La detección de las opiniones se puede realizar a nivel de sentencia, de aspecto o de documento.

Este trabajo está más centrado en la detección a nivel de aspecto, porque permite ver cómo los usuarios pueden expresar sus preferencias acerca de determinadas características de un producto o servicio descartando otros. Por ejemplo, si un usuario que acaba de adquirir un teléfono hace una crítica feroz acerca de su batería, esto puede suponer que adquirió el teléfono pensado que la batería era uno de sus puntos fuertes y a la vez, se podría inferir que una de las preferencias del usuario era dicha batería. Sin embargo, si el usuario a pesar de hacer una queja sobre un aspecto como podría ser la cámara frontal del móvil no es demasiado efusivo, podría interpretarse que el producto no es bueno, pero tampoco era un detalle importante para el usuario a la hora de tomar su decisión final sobre comprar un móvil u otro.

Todos estos detalles hacen que para entender la opinión final de un usuario sobre un producto o servicio, sea necesario tener en cuenta la importancia de cada aspecto y el grado de positividad o negatividad emitido hacia cada uno de ellos. Es por ello, que se propone el modelado de una opinión final como la agregación borrosa de las opiniones parciales emitidas sobre cada aspecto.

La principal contribución de este artículo es una visión acerca de algunos posibles comportamientos de los usuarios cuando expresan opiniones, especialmente, sobre productos y servicios que compran o utilizan, así como una posible aproximación borrosa que permitiría el modelado de dichos comportamientos, y sus posibles aplicaciones.

El artículo está organizado en las siguientes secciones: la sección II se corresponde con el estado del arte, la sección III presenta la principal motivación del artículo, la sección IV presenta cuál sería la metodología para agregar las distintas opiniones y dar una conclusión final al usuario sobre un producto, la sección V presentará algunas aplicaciones de uso y finalmente, algunas conclusiones serán remarcadas.

II. Estado del arte

Según Cambria, el Análisis de Sentimientos puede ser considerado como un problema de Procesamiento de Lenguaje Natural (PLN) muy restringido, en el que sólo es necesario entender los sentimientos positivos o negativos respecto a cada frase y/o las entidades o temas objetivo [1]. Sin embargo, a pesar de ser un problema restringido, todos los trabajos en este campo, así como todos los trabajos en Recuperación de Información, siempre luchan con problemas de PLN no resueltos (manejo de la negación, reconocimiento de entidades con nombre, desambiguación del sentido de la palabra,...) que son esenciales para detectar claves del lenguaje como la ironía o el sarcasmo [2], y en consecuencia, para encontrar y valorar sentimientos.

Uno de los principales aspectos que debe abordar la PLN son los diferentes niveles de análisis. Dependiendo de si el objetivo del estudio es un texto completo o un documento, una o varias frases enlazadas, o una o varias entidades o aspectos de esas entidades, se pueden realizar diferentes tareas de PLN y Análisis de Sentimientos. Por lo tanto, es necesario distinguir

tres niveles de análisis: (i) nivel de documento, (ii) nivel de frase y (iii) nivel de entidad/aspecto.

A nivel de documento se considera que un documento es una opinión sobre una entidad o aspecto de la misma. Este nivel está asociado con la tarea llamada clasificación de sentimientos a nivel de documento [3]–[6]. Sin embargo, si un documento presenta varias frases que tratan de diferentes aspectos o entidades, entonces el nivel de oración es más adecuado. El nivel de la oración está relacionado con la clasificación de la subjetividad de la tarea; considera cada oración como una opinión positiva, negativa o neutral [7]–[10]. Y finalmente, cuando se necesita información más precisa, entonces surge el nivel de entidad/aspecto. Es el nivel de grano más fino, considera un objetivo sobre el que el ponente de opinión expresa una opinión positiva o negativa. Este último nivel es posiblemente el más complejo, ya que es necesario extraer con gran precisión muchas características, como las fechas o el lapso de tiempo, las diferentes características/espectáculos y entidades a tener en cuenta, así como las relaciones entre ellas, los formadores de opinión y sus características, etc. Está estrechamente relacionado con tareas como Opinion Mining y Opinion Summarization [9], [10].

Muchas tareas surgen vinculadas al Análisis de Sentimientos. Algunas de ellas están estrechamente relacionadas y es difícil separarlas claramente porque comparten muchos aspectos. Los más importantes son:

1. Clasificación de sentimientos: también llamada orientación de sentimientos, orientación de opinión, orientación semántica o polaridad de sentimientos [11]. Se basa en la idea de que un documento/texto expresa la opinión de un titular sobre la entidad y trata de medir el sentimiento de ese titular hacia la entidad. Por lo tanto, consiste principalmente en clasificar las opiniones en tres categorías principales: positivas, negativas o neutras. Parece una tarea simple; sin embargo, es una tarea compleja, especialmente cuando las opiniones provienen de múltiples dominios o idiomas [12], [13]. Esta tarea está estrechamente relacionada con la predicción de la valoración de los sentimientos, que consiste en medir la intensidad de cada sentimiento.

2. Clasificación de subjetividad. Consiste principalmente en detectar si una frase dada es subjetiva o no. Una frase objetiva expresa información objetiva, mientras que una frase subjetiva puede expresar otro tipo de información personal como opiniones, evaluaciones, emociones, creencias, etc. Además, las frases subjetivas pueden expresar sentimientos positivos o negativos, pero no todas lo hacen. Esta tarea puede ser vista como un paso previo a la clasificación de los sentimientos. Una buena clasificación de la subjetividad puede asegurar una mejor clasificación de los sentimientos [14]–[17].

3. Resumen de opiniones. Se centra especialmente en extraer las características principales de una entidad compartida dentro de uno o varios documentos y los sentimientos al respecto [18]. Por lo tanto, se pueden distinguir dos perspectivas en esta tarea: la integración de un solo documento y la integración de varios documentos. La integración de un solo documento consiste en analizar hechos internos presentes en el documento analizado, por ejemplo, cambios en la orientación de los sentimientos a lo largo del documento o vínculos entre las diferentes entidades/características encontradas, y principalmente mostrar aquellos textos que mejor los describen. Por otro lado, en la integración multidocumento, una vez detectadas las características y entidades, el sistema debe agrupar y/o ordenar las diferentes frases que expresan sentimientos relacionados con dichas entidades o características. El resumen final puede presentarse en forma de gráfico o texto que muestre las principales características/entidades y cuantifique el sentimiento con respecto a cada una de ellas de alguna manera, por ejemplo, agregando intensidades de sentimientos o contando el número de frases positivas o negativas [19], [20].

4. Recuperación de opiniones. Intenta recuperar documentos que expresan una opinión sobre una consulta determinada. En este tipo de sistemas, se requieren dos puntajes para cada documento, el puntaje de relevancia frente a la consulta y el puntaje de opinión sobre la consulta, y ambos se utilizan generalmente para clasificar los documentos [21], [22].

5. Sarcasmo e ironía. Se centra en detectar afirmaciones con contenido irónico y sarcástico. Esta es una de las tareas más complicadas en este campo, especialmente debido a la falta de acuerdo entre los investigadores sobre cómo se puede definir formalmente la ironía o el sarcasmo [2], [23], [24]

6. Otros. Además de las actividades anteriormente mencionadas, existen otras tareas relacionadas con el Análisis de Sentimientos, como por ejemplo, la detección de género o autoría, que trata de determinar el género o la persona que ha escrito un texto/opinión [25], [26], la detección de spam de opinión, que trata de detectar opiniones o reseñas que contienen contenidos no confiables publicados para distorsionar la opinión pública hacia personas, empresas o productos [27]–[29], o la identificación de temáticas en función de los términos referidos a sentimientos [30].

III. Motivación

Son muchas las aplicaciones y páginas web que permiten al usuario expresar sus opiniones. Analizando detenidamente las distintas opiniones de forma general, se puede observar que los comportamientos de los usuarios pueden ser muy diversos, pudiendo establecerse distintos tipos de usuario dependiendo de las formas en las que emiten sus opiniones.

De igual manera, son muchas las formas en las que un usuario puede expresar su opinión. Por ejemplo, entre las más típicas podemos encontrar el uso de conjuntos de estrellas como puede verse en muchas páginas webs de venta de productos como Amazon[1] o Joom[2]:

Fig. 1. Puntuación mediante estrellas

o mediante una barra que permita establecer el grado de conformidad con respecto a un aspecto concreto:

Fig. 2. Barra de puntuación

entre otros posibles mecanismos. Como puede observarse, estos mecanismos permiten de una manera fácil, rápida e intuitiva expresar una opinión, sin embargo, carecen de cierta expresividad que si provee el lenguaje natural. Por ejemplo, las estrellas vistas anteriormente podrían ir acompañadas de distintas etiquetas lingüísticas {No satisfecho, Poco satisfecho, Medianamente satisfecho, Satisfecho, Muy satisfecho}, que pueden resultar más intuitivas e interpretables que las estrellas o su representación numérica: 1 estrella, 2 estrellas, 3 estrellas, etc.

Esto se puede ilustrar mediante el siguiente ejemplo (Figura 3) sacado de la página web TripAdvisor[3]:

Fig. 3. Opinión con puntuación poco precisa

Si se observan las puntuaciones, la estancia parece haber sido perfecta porque son máximas respecto a todos los aspectos posibles (valor, localización, habitaciones, limpieza, ….). Sin embargo, si se observan los comentarios esto no es así. El tamaño de la habitación era bueno pero no excelente. Lo mismo ocurre con la calidad del sueño, fue buena pero no excelente. Se podría pensar que este usuario utilizaría la etiqueta lingüística "bueno" como símbolo para representar la máxima calidad, sin embargo, si se lee la primera frase, la localización era "excelente", por lo que gradación de las puntuaciones podría llegar hasta ese extremo.

A. *Algunos posibles tipos de usuarios*

Ya hemos visto algún tipo de usuario como puede ser el usuario poco preciso de la imagen anterior, pero podemos ver más posibles usuarios. Por ejemplo, si miramos a la siguiente opinión que expresa de forma fehaciente que el hotel es fantástico:

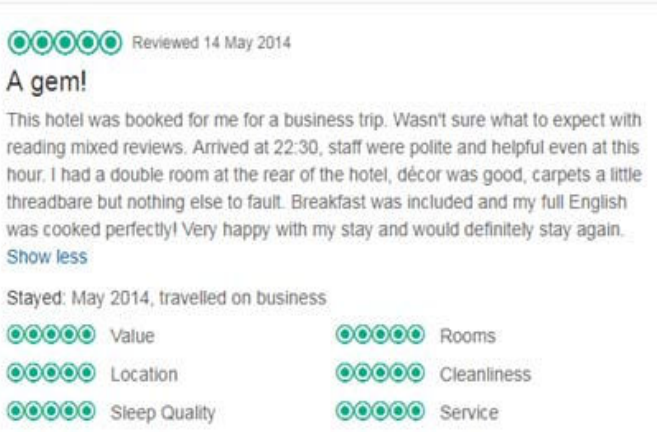


Fig. 4. Opinión excesivamente positiva

contrasta con la siguiente opinión emitida con una semana de diferencia, donde el usuario afirma que el mismo hotel es horrible:

[1] www.amazon.com
[2] www.joom.com
[3] www.tripadvisor.com

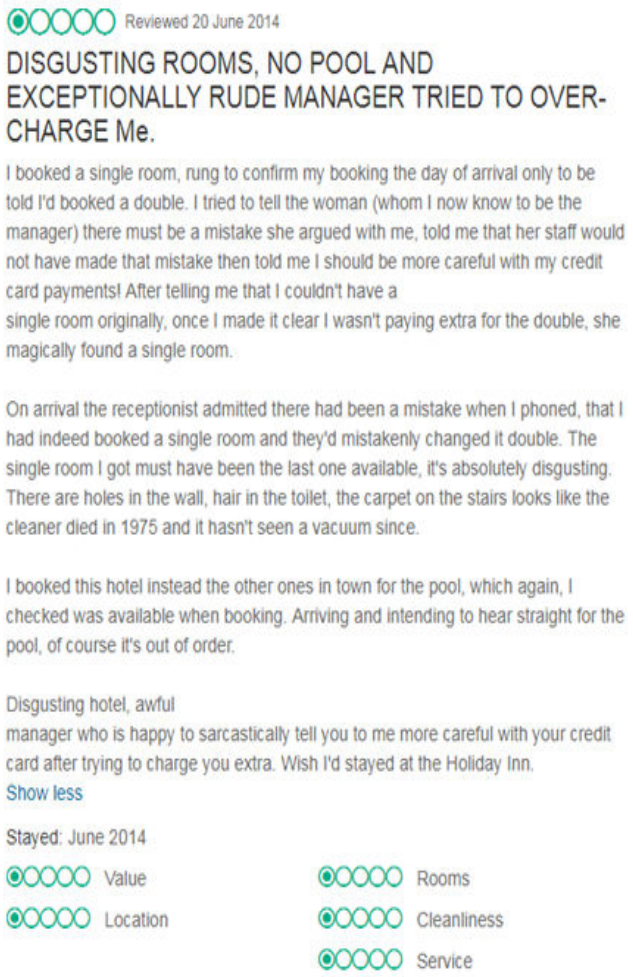


Fig. 5. Opinión excesivamente negativa

Leyendo detenidamente ambos comentarios, puede verse cómo se comparan cosas como el personal o las instalaciones con perspectivas absolutamente distintas. En este caso, podría pensarse que los usuarios pueden ser falsos, es decir, un usuario expresa su opinión simplemente para denigrar la calidad de un hotel, lo que podría conocerse como un "hater", o un usuario podría intentar mejorar la calidad de un hotel bien porque tiene algún interés escondido, en este caso sería un "lover". Igualmente, podría ser, simplemente, que el estado de ánimo con el que se emitieron las opiniones no permite expresar de forma excesivamente realista cuál fue la situación real del hotel, o sí.

IV. PROPUESTA

Como puede observarse, modelar el comportamiento de los usuarios puede ser complejo y hay distintos factores a tener en cuenta. Entre estos factores pueden destacarse:

- la importancia que tiene para cada usuario un aspecto determinado dentro del producto del que se está opinando.
- La opinión que tiene el usuario de cada uno de los aspectos que conforman un objeto.
- La opinión global de un producto o servicio vista como el resultado de la agregación de cada uno de los aspectos en función de la importancia.

A. *Opinión sobre un aspecto*

Así pues, matemáticamente un producto podría definirse como un conjunto A de N aspectos:

$$A = \{A_1, A_2, A_3, \ldots, A_N\}$$

Sobre cada aspecto es posible emitir una opinión que puede ser representada mediante un conjunto de etiquetas lingüísticas $\{S_1, S_2, S_3 \ldots\}$. Estas etiquetas lingüísticas podrían ser definidas mediante número triangulares borrosos NTB, donde S_i se puede representar por el número $NTB_i = (a_i, b_i, c_i)$. La distribución de los parámetros puede dar lugar a un conjunto de etiquetas balanceado o no. Esto puede permitir tener más capacidad a la hora de filtrar a los usuarios y a los vendedores de los productos.

Véase el siguiente ejemplo. Si la valoración de las opiniones de cada aspecto fuera usando el siguiente conjunto balanceado de 5 etiquetas:

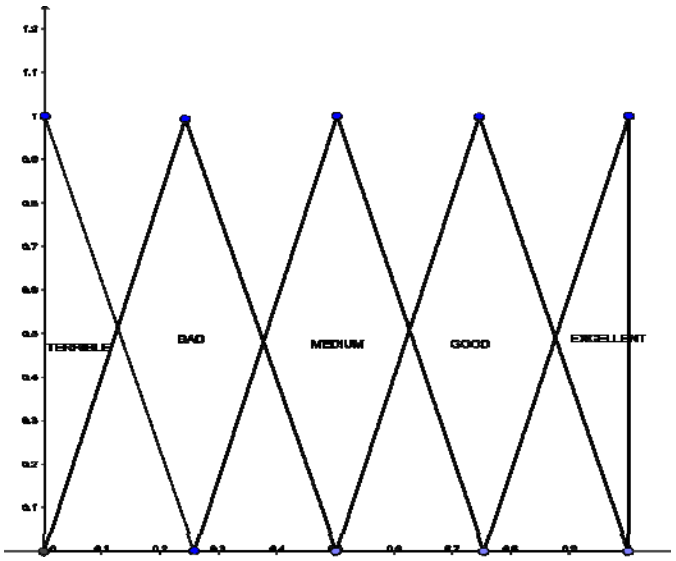


Fig. 6. Distribución de etiquetas balanceada

Podría corresponder intuitivamente con el sistema de estrellas visto en plataformas como Amazon (ver Figura 1), donde una opinión con una única estrella representa la peor valoración, es decir, el primer conjunto de la figura (TERRIBLE), y 5 estrellas representa la mejor valoración, en este caso el último conjunto (EXCELENTE).

Sin embargo, si se utilizara un conjunto de etiquetas como el visto en la siguiente figura, el usuario tendría un mayor poder discriminación, primero porque existen más etiquetas con las que valorar una opinión y segundo porque el soporte de las etiquetas que representan opiniones positivas es mucho menor, por lo que el usuario sería mucho más exigente a la

hora de decir si un producto es BUENO, MUY BUENO o EXCELENTE.

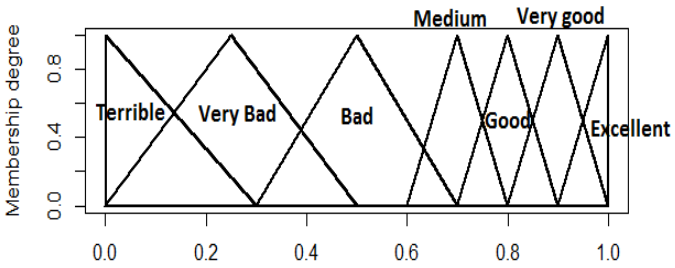


Fig. 7. Distribución de etiquetas no balanceada

B. *Importancia de cada aspecto*

Cada usuario puede considerar cada aspecto según su propio criterio. Suponer un teléfono móvil, ¿cuál es el aspecto más importante a la hora de comprarlo? Cada usuario pensará uno distinto, unos querrán que la cámara sea de alta definición para hacer fotos, a otros la cámara no les importará y preferirán que la batería le dure más tiempo, otros preferirán que tenga mucha memoria para almacenar muchos archivos, o la que conexión sea 5G, etc.

Por tanto, a la hora de emitir la opinión general sobre un producto, cada aspecto no tiene por qué pesar lo mismo. Así, es necesario modelar el peso de cada aspecto según el usuario.

Este peso puede ser calculado de muchas maneras, por ejemplo, a través de un valor numérico dentro de una escala [0,1]. Siguiendo con el ejemplo del teléfono se podría dar las siguientes importancias a estos posibles aspectos:

- Batería: 0.5
- Pantalla: 0.2
- Conectividad 5G: 0.1
- Cámara: 0.2

En este caso, como puede verse, todos los pesos suman 1, y el más importante sería en este caso el aspecto de la batería, mientras que la conectividad 5G no parece ser nada relevante para este usuario.

Siguiendo una aproximación más intuitiva, la importancia podría ser una variable representada mediante un conjunto de etiquetas $\{S_1, S_2, S_3\ldots\}$, las cuales nuevamente podrían ser definidas mediante número triangulares borrosos NTB, donde S_i se puede representar por el número $NTB_i = (a_i, b_i, c_i)$. La distribución puede ser igual balanceada o no, y como ejemplo podría darse el siguiente conjunto, en el cual el usuario podría seleccionar si un aspecto para él es irrelevante, simplemente importante o de vital importancia:

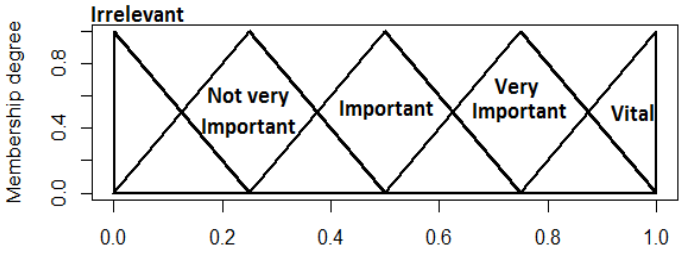


Fig. 8. Ejemplo de etiquetas para medir la importancia de un aspecto

Dependiendo de la forma en la que se mida la importancia, determinará cómo se calcula la opinión general que puede tener un usuario respecto a un producto.

C. *Opinión general: Agregación de aspectos*

Una vez que el usuario ha determinado cuál es la importancia de cada aspecto, y ha opinado sobre cada uno de ellos, es necesario agregar todas las opiniones individuales pero, teniendo en cuenta cuál es más importante que la otra.

Para ello, se puede proponer el uso de los operadores OWA [31]. Dependiendo de cómo se hayan obtenidos los distintos pesos y las valoraciones de cada aspecto, se podrán utilizar distintas familias de operadores OWA.

En el caso, por ejemplo, de medir a través de la calificación mediante estrellas (ver Figura 1) y que los pesos sean introducidos como valores en el rango [0,1] sumando en su conjunto 1, al ser todos los valores de naturaleza crisp, se podría utilizar un operador OWA clásico. Sin embargo, en el caso de que se utilizaran conjuntos borrosos para obtener los parámetros necesarios, serían necesarios acceder a familias como los OWA de Tipo-1 [32].

V. APLICACIONES

A través de la propuesta anterior, es posible modelar de una forma más fehaciente, cuál puede ser la opinión general de un usuario cuando valora un producto. Entre las aplicaciones que podría encontrarse a esta propuesta podrían mencionarse:

- Detección y clasificación de usuarios: Poder clasificar a los usuarios según sus opiniones puede resultar interesante desde distintos puntos vista. Por ejemplo, un usuario que siempre puntúa positivamente los productos de una marca y negativamente los de la competencia, no tiene por qué implicar que sus opiniones no sean válidas, simplemente es su forma de pensar. Así, un sistema de recomendaciones tendría más información para saber qué productos recomendarle y cuáles no.
- Asignación del grado de confianza de un usuario: Como consecuencia de lo anterior, las opiniones de todos los usuarios no parecen ser igualmente fiables, por lo que la descripción de cada usuario a través de un grado de confiabilidad podría permitir descartar las opiniones de ciertos usuarios. Este proceso hoy es imposible en webs como Amazon o Trivago.

VI. Conclusiones

En este artículo se pone de manifiesto la necesidad de representar las opiniones de los usuarios de forma distinta a la utilizada por muchas plataformas web debido a su falta de expresividad en muchas ocasiones. El uso de aproximaciones borrosas puede ayudar a paliar estos defectos como se ha comentado a lo largo del artículo.

Siguiendo el modelo propuesto, sería más fácil descartar opiniones que pueden no tener valor para un usuario, bien porque en sí pudieran ser contradictorias o de poco valor, o porque estuvieran emitidas por usuarios con el propósito de influir en otros usuarios de forma malintencionada.

Agradecimientos

Este artículo ha sido financiado por FEDER y la Agencia Española de Investigaciones del Ministerio de Economía y Competitividad bajo el proyecto TIN2016-76843-C4-2-R (AEI/FEDER, UE).

Bibliografía

[1] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE Intell. Syst., vol. 28, no. 2, pp. 15–21, 2013.

[2] A. Reyes, P. Rosso, and D. Buscaldi, "From humor recognition to irony detection: The figurative language of social media," Data Knowl. Eng., vol. 74, pp. 1–12, 2012.

[3] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment analysis of Chinese documents: From sentence to document level," J. Am. Soc. Inf. Sci. Technol., vol. 60, no. 12, pp. 2474–2487, 2009.

[4] R. Moraes, F. Valiati, J, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," Expert Syst. Appl., vol. 40, no. 2, pp. 621–633, 2013.

[5] A. Duric and F. Song, "Feature selection for sentiment analysis based on content and syntax models," Decis. Support Syst., vol. 53, no. 4, pp. 704–711, 2012.

[6] Y. He and D. Zhou, "Self-training from labeled features for sentiment analysis," Inf. Process. Manag., vol. 47, no. 4, pp. 606–616, Jul. 2011.

[7] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), 2005, pp. 347–354.

[8] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis," Comput. Linguist., vol. 35, no. 3, pp. 399–433, 2009.

[9] A. Agarwal, F. Biadsy, and K. R. McKeown, "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic N-grams," in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL '09), 2009, pp. 24–32.

[10] R. Remus and C. Hänig, "Towards well-grounded phrase-level polarity analysis," in Proceedings of the 12th international conference on Computational linguistics and intelligent text processing (CICLing'11), 2011, pp. 380–392.

[11] L.-C.- Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," Knowledge-Based Syst., vol. 41, pp. 89–97, 2013.

[12] Y. He, C. Lin, and H. Alani, "Automatically extracting polarity-bearing topics for cross-domain sentiment classification," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11), 2011, pp. 123–131.

[13] M. Dragoni and G. Petrucci, "A fuzzy-based strategy for multi-domain sentiment analysis," Int. J. Approx. Reason., vol. 93, pp. 59–73, 2018.

[14] S. Raaijmakers and W. Kraaij, "A Shallow Approach to Subjectivity Classification," in Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM '08), 2008, pp. 216–217.

[15] K. Sarvabhotla, P. Pingali, and V. Varma, "Sentiment classification: a lexical similarity based approach for extracting subjectivity in documents," Inf. Retr. Boston., vol. 14, no. 3, pp. 337–353, 2011.

[16] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," Decis. Support Syst., vol. 53, no. 4, pp. 675–679, 2012.

[17] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," Decis. Support Syst., vol. 53, no. 4, pp. 680–688, 2012.

[18] D. Wang, S. Zhu, and T. Li, "SumView: A Web-based engine for summarizing product reviews and customer opinions," Expert Syst. Appl., vol. 40, no. 1, pp. 27–33, 2013.

[19] K.-M. Park, H. Park, H.-G. Kim, and H. Ko, "Review summarization based on linguistic knowledge," in Proceedings of the 17th international conference on Database Systems for Advanced Applications (DASFAA'12), 2012, vol. 7240, pp. 105–114.

[20] K. Ganesan, C. Zhai, and E. Viegas, "Micropinion generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions," in Proceedings of the 21st international conference on World Wide Web - WWW '12, 2012, pp. 869–878.

[21] S.-W. Lee, Y.-I. Song, J.-T. Lee, K.-S. Han, and H.-C. Rim, "A new generative opinion retrieval model integrating multiple ranking factors," J. Intell. Inf. Syst., vol. 38, no. 2, pp. 487–505, May 2011.

[22] L. Guo and X. Wan, "Exploiting syntactic and semantic relationships between terms for opinion retrieval," J. Am. Soc. Inf. Sci. Technol., vol. 63, no. 11, pp. 2269–2282, 2012.

[23] E. Filatova, "Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing," in Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), 2012, pp. 392–398.

[24] A. Reyes and P. Rosso, "Making objective decisions from subjective data: Detecting irony in customer reviews," Decis. Support Syst., vol. 53, no. 4, pp. 754–760, 2012.

[25] J. Savoy, "Authorship Attribution Based on Specific Vocabulary," ACM Trans. Inf. Syst., vol. 30, no. 2, pp. 1–30, 2012.

[26] M. Montesi and T. Navarrete, "Classifying web genres in context: A case study documenting the web genres used by a software engineer," Inf. Process. Manag., vol. 44, no. 4, pp. 1410–1430, 2008.

[27] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in Proceedings of the 21st international conference on World Wide Web (WWW '12), 2012, pp. 191–200.

[28] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 823–831.

[29] G. Wang, S. Xie, B. Liu, and P. S. Yu, "Identify Online Store Review Spammers via Social Review Graph," ACM Trans. Intell. Syst. Technol., vol. 3, no. 4, pp. 1–21, 2012.

[30] K. Gutiérrez-Batista, J. R. Campaña, M.-A. Vila, and M. J. Martin-Bautista, "Fuzzy Analysis of Sentiment Terms for Topic Detection Process in Social Networks," in International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2018, pp. 3–14.

[31] R. R. Yager, "Families of OWA operators," Fuzzy Sets Syst., vol. 59, no. 2, pp. 125–148, 1993.

[32] S.-M. Zhou, F. Chiclana, R. I. John, and J. M. Garibaldi, "Type-1 OWA operators for aggregating uncertain information with uncertain weights induced by type-2 linguistic quantifiers," Fuzzy Sets Syst., vol. 159, no. 24, pp. 3281–3296, Dec. 2008.

Engine health monitoring with fuzzy data: lessons learned from aircraft industry

Luciano Sánchez
Universidad de Oviedo
Departamento de Informática
Gijón, Asturias, Spain
luciano@uniovi.es

Inés Couso
Universidad de Oviedo
Departamento de Estadística
Gijón, Asturias, Spain
couso@uniovi.es

***Abstract*—The use of data-driven techniques for health monitoring and prognosis of engines is often plagued by the lack of information about failure scenarios. This problem is aggravated when there are missing or partially missing records in the log data. Unluckily, the chance that a sequence of events leading to a failure is not properly stored is higher when the system exhibits an abnormal behaviour. In absence of patterns that can be related to deterioration, failure detection is often guided to find novelties in the records, or uncommon combinations of variables that may indicate an anomalous performance. Finding novelties in sequences of low quality data is a complex problem where numerous ambiguities must be resolved. Aircraft industry is a paradigmatic case in that it involves monitoring and prognosis with less than perfect data, partly because many different control and security subsystems prevent that failure conditions are actually reached. The use of fuzzy technologies for modelling and diagnosing aero-engines is discussed in this paper, along with the most challenging issues and future research lines, according to our own experience in this field.**

***Index Terms*—Fuzzy data, Engine health monitoring, Fault diagnosis, Fuzzy model**

I. Introduction

The main purpose of engine data is to monitor engine parameters in order to avoid running the engine under undesired conditions. Engine instrumentation is configured to trigger alerts related to operator action, maintainance action or, if a significant condition is found, shutting down the engine.

The development over time of the engine data is also monitored in a process that is called Engine Health Monitoring (EHM). EHM management systems have diagnostic and prognostic purposes, as not only individual working conditions but also the trend over time are examined in order to identify rapid levels of deterioration. EHM management systems estimate the Remaining Useful Life (RUL) of an engine, anticipating certain events or findings and therefore reducing the number and degree of engine refurbishments [1]. In the simplest case, EHM management systems compare engine data against those parameters identified to be characteristic of known engine conditions or against design limits [2]. However, predicting the engine parameter deterioration levels over time is complex. There are multiple methods of EHM data assessment developed, whose range of application depends on the type of engine and the amount and quality of the engine data.

This study concerns the application of fuzzy technologies for EHM in aircrafts, where multi-sensor information is used to report failures and predict the RUL of commercial turbofan engines. Aircraft industry has specific requirements for EHM systems, derived from the fact that gas turbines are inherently fail-safe and the reliability of the turbine may be higher than that of its sensors [3]. Sensor degradation cannot be tolerated because deviations in sensor readings can be mistaken for engine degradations and this can cause secondary failures because of wrong control decisions [4]. Hence, aircraft engines tend to be under-sensorized and a substantial uncertainty in the knowledge of the operating point of the engine is accepted. Uncertainty in EHM data is accounted for via mathematical models, statistical or intelligent techniques, and in many cases with the help of fuzzy technologies, which will be reviewed in the forthcoming sections.

The structure of this paper is as follows: Section II reviews the use of fuzzy technologies in aircraft EHM, discusses the strengths are weakness of these techniques, and states some open problems. A new family of fuzzy diagnostic tools is proposed in Section III. An explanatory application of the new tools is developed in Section IV. The conclusions of the paper are presented in Section V.

II. Review of the use of fuzzy technologies in aeroengine health monitoring

Most of the aircraft EHM assessment methods are based around Gas Path Analysis (GPA) [2]. The gas path components are all air-washed parts within the engine gas path: the compressors, the combustor and the turbines. The gas path components are susceptible to distinct different issues, such as worn seals, excessive tip clearances, burning, cracking or missing parts or sections of parts, etc. Changes in the internal working conditions of the engine are detected either by direct observation of EHM parameters or indirectly, through a suitable transform of the EHM data.

This work was supported by the Science and Innovation Spanish Ministry under Projects TIN2014-56967-R and TIN2017-84804-R

A. Direct methods: detection of novelties with pattern recognition

The most common GPA methodology consists in detecting changes or *novelties* [5] in the internal working conditions of the engine as early as possible, by finding specific combination of values or *signatures* of the different defects. Fuzzy expert systems have been used to identify engine trends and step changes [6]. These algorithms rely on an experience database comprising GPA variables sampled at engines with known faults. Recent approaches learn fuzzy models from clean engine data and deteriorations are detected as novelties in the EHM data or in the residuals between the measurements and the model predictions [7], [8].

The diagnosis of an engine is formulated as a supervised pattern recognition problem, often via a case-based reasoning algorithm in which the measurements are compared to the known signatures of defects that have been observed in the past [9]. Thus, pattern recognition techniques are only efficient if the experience database is complete; otherwise, unseen defects will remain undetected. This is a requirement hard to fulfil, as most of times the Full Authority Digital Engine Control (FADEC) and other safety mechanisms will prevent that the engine reaches these abnormal conditions. Also, there are faults that do not have a well defined signature: for instance, a higher than normal turbine temperature may have different causes, such as an incipient deterioration or a higher load of the engine. Pattern recognition cannot tell apart these root causes if temperature, fuel flow, pressures and thurst are treated in isolation.

B. Subspace methods and manifold learning

Since most of the deteriorations alter more than one GPA variable, some authors consider that each of these variables must be regarded as a mix of different factors [10]. In this respect, subspace methods project EHM variables into a low-dimensional factor space and the signatures of the defects are sought in this projection [11], [12]. In other cases, a non-linear transformation is used; for instance, neural-network based autoencoders condense the EHM information into a reduced number of variables [13]. Recently, other manifold learning techniques have been extended to uncertain data and applied to this problem. In [14], Blind Source Separation (BSS) for interval-valued data data has been applied to the EHM problem thus the independence among the factor is maximized. This last reference addresses the presence of epistemic uncertainty in the data. In any case, note that although the dimensionality reduction simplifies the aforementioned pattern recognition problem, the need for an experience database is still present.

C. Multivariate trend analysis of delta signals

One of the major sources of uncertainty in EHM is related to the scarcity of sensors. Measurements are taken under different flying conditions. Hence, certain changes in the monitored variables may be caused by either a deterioration in the engine or a change in the flight conditions. Furthermore, there is a wide variability between GPA related magnitudes for different engines, as the monitored variables of two engines flying in the same conditions might not be concordant from flight to flight. Following [15], this second variability is solved if defects are not sought in GPA variables but in the gradient of the so-called "delta" functions, which are the differences between GPA variables and the theoretical values of a single reference engine. The gradient of these delta functions are used, because useful information about the deterioration speed is conveyed by the slopes of these curves. However, since GPA variables combine measurements taken in different flying conditions, delta signals are very noisy. This poses a complex computational problem, as determining the gradient of a noisy signal is strongly dependent on the properties of the noise filter. A fuzzy-bandwidth filter was used to compute these slopes in [16], where a fuzzy rule-based classifier was subsequently used to monitor the engine on the basis of the multivariate gradient signal, and also in [17] fuzzy estimations of the RULs of compressor and turbine were produced for prognosis purposes. These methods still make use of a sample of engines with a known condition, that is used for learning classifier and RUL models. In contrast, the experience database with the signatures of the known defects are not needed anymore, as the monitoring and prognosis decisions are taken by data-driven fuzzy rule-based systems.

D. Sequence mining

Multivariate trend analysis transforms variable-length series of gradients into fixed-length feature vectors that are the inputs to a fuzzy classifier or model. This feature selection process discards the sequence order of the deteriorations. For example, suppose that two different motors #1 and #2 have bad compressors and turbines: the compressor of #1 failed first and its turbine was subsequently damaged because of the compressor bad state. In contrast, the turbine in #2 failed first and its compressor was damaged later. It may happen that their feature vectors in trend analysis are identical. In reference [18] the fixed-length transform is avoided, and specific sequential pattern mining-based classifiers (that can operate with variable-length fuzzy inputs) were leveraged. Sequence mining discovers conditions that are not to be found with multivariate trend analysis. In any case, other drawbacks exist: a sample of engines with known conditions is needed (although in this case the decision system does not require an experience database). Furthermore, frequent pattern matching techniques are used for identifying the signatures of the deffects. Hence, isolated deteriorations will not be found.

Summarizing, the conclusions of this review in fuzzy EHM for aircrafts are:

- Direct methods operate with experience databases comprising GPA values of engines with known deteriorations. These methods are sensible to uncertainties originated in the variability from engine to engine, the changing flight conditions and the lack of sensors.
- Fuzzy subspace methods alleviate the uncertainty caused by the lack of sensors, as different imprecise signals are

combined to extract independent factors. The experience database is still needed.
- Multivariate trend analysis relax the requisite of an experience database to a list of engines with known RULs, thus not-yet-seen defects can be discovered. These methods cannot detect the precise time when a defect appeared.
- Sequence mining cannot find defects with a low occurrence, because the signatures are obtained via fuzzy frequent pattern matching.

III. Isotonic fuzzy health models

Diagnosis and prognosis methodologies either depend on an experience database, a sample of engines with known RULs or a model of the engine whose residual can be monitored. The main advantage of model-dependent assessments lies in the possibility of finding defects not seen before, but being dependent on an engine model is a stringent requisite. In this section a new isotonic fuzzy health model of an aeroengine is proposed, whose residual can be related to the location of the deteriorations. This method is novel because it makes use of *the lowest possible amount of domain knowledge*, which is (i) certain variables in the engine are comonotonic, and (ii) the state of health of the engine decreases with time.

The input variables of the proposed model are the same "delta" variables introduced in Section II-C. In the particular case of the aeroengines concerned by this study (see Figure 1), these variables are called ΔP30, ΔT30, ΔTGT, ΔFF and ΔN2. The list of points of interest in the engine are:

- Station 3: This is the High Pressure Compressor (HPC) exit and the entry into the combustion system. The conditions at this point are key for the correct functioning of the engine. The main variables measured at this station are P30 (pressure) and T30 (temperature).
- Station 4: This is the combustion chamber exit and High Pressure Turbine (HPT) entry. The temperature at this point is one of the main engine parameters. T4, may also be known as Turbine Gas Temperature (TGT)
- Station 5: This is the Low Pressure Turbine (LPT) exit. The main variable at this station is P50. This pressure is used to define EPR, which is subsequently used to determine the overall engine thrust. EPR is the relation of P50 to P20.

The Low Pressure (LP) system is the combination of the fan and the LPT. The speed at which the LP system turns is defined as N1. The High Pressure (HP) system is the combination of the HPC and the HPT. The speed at which the HP system turns is known as N2. In addition, the amount of fuel consumed is also monitored through Fuel Flow (FF).

EPR is the relation of P50 (pressure at the Low Pressure Turbine exit) to P20 (pressure at the fan inlet). In two shaft high bypass ratio turbo fans, the thrust is performed by the air compressed by the fan blades and pushed through the engine bypass. The air pushed through the core of the engine is solely used to turn the fan. This is, the air is compressed by the high pressure compressor (HPC) so that the optimum conditions are reached within the combustion chamber to subsequently turn the high pressure turbine (HPT) to maintain the high pressure (HP) system and subsequently turn the low pressure turbine (LPT) which moves the fan and produces the engine thrust.

A. Proposed isotonic fuzzy model

Let the thrust of the considered engine be described by the following function:

$$\mathrm{epr}(\Delta\mathrm{P30}, \Delta\mathrm{T30}, \Delta\mathrm{TGT}, \Delta\mathrm{FF}, \Delta\mathrm{N2}, \mathbf{fc}) \tag{1}$$

where $\mathbf{fc} \in \mathrm{FC}$ is a vector defining the flight conditions, ΔP30 is the difference between the gas pressure at the entry of the HPC and that of the reference engine, etc. Note that the thrust of the reference motor is $\mathrm{epr}_0(0,0,0,0,0,\mathbf{fc})$. For engines that are in a good condition,

- $\mathrm{epr}(\cdot)$ is comonotonical with ΔP30, ΔFF, ΔTGT, ΔN2 (the higher the compressor pressure, the consumed fuel, the turbine temperature and the turbine speed, then the higher the thrust is)
- $\mathrm{epr}(\cdot)$ is antimonotonical with ΔT30 (the higher the compressor temperature, the less dense the air is, hence the thrust is lower)

Thus, the following fuzzy-valued model is proposed:

$$\mu_{\widetilde{\mathrm{EPR}}}(\Delta\mathrm{P30}, \Delta\mathrm{T30}, \Delta\mathrm{TGT}, \Delta\mathrm{FF}, \Delta\mathrm{N2})(e) = \sup_{\alpha}\{e \in [\overline{\mathrm{EPR}}]_\alpha(\Delta\mathrm{P30}, \Delta\mathrm{T30}, \Delta\mathrm{TGT}, \Delta\mathrm{FF}, \Delta\mathrm{N2})\} \tag{2}$$

where the set $[\overline{\mathrm{EPR}}]_\alpha$ is a confidence interval of the thrusts of the engine,

$$[\overline{\mathrm{EPR}}]_\alpha = \{\mathrm{epr}_i(\Delta\mathrm{P30}, \Delta\mathrm{T30}, \Delta\mathrm{TGT}, \Delta\mathrm{FF}, \Delta\mathrm{N2}, \mathbf{fc}) : \mathbf{fc} \in \mathrm{FC}_\alpha\} \tag{3}$$

and FC_α is the smallest subset of FC such that $\mathrm{P}(\mathbf{fc} \in \mathrm{FC}) \geq 1-\alpha$ and $\mathrm{FC}_\alpha \subseteq \mathrm{FC}_\beta$ for $\alpha > \beta$. Lastly, let the parametric definition of $\mathrm{epr}(\cdot)$ be

$$\mathrm{epr}(\Delta\mathrm{P30}, \Delta\mathrm{T30}, \Delta\mathrm{TGT}, \Delta\mathrm{FF}, \Delta\mathrm{N2}, \mathbf{fc}) = f(\kappa_1 \cdot \Delta\mathrm{P30} - \kappa_2 \cdot \Delta\mathrm{T30} + \kappa_3 \cdot \Delta\mathrm{TGT} + \kappa_4 \cdot \Delta\mathrm{FF} + \kappa_5 \cdot \Delta\mathrm{N2}) + g(\mathbf{fc}) \tag{4}$$

and

$$[\overline{\mathrm{EPR}}]_\alpha(\Delta\mathrm{P30}, \Delta\mathrm{T30}, \Delta\mathrm{TGT}, \Delta\mathrm{FF}, \Delta\mathrm{N2}) = f(\kappa_1 \cdot \Delta\mathrm{P30} - \kappa_2 \cdot \Delta\mathrm{T30} + \kappa_3 \cdot \Delta\mathrm{TGT} + \kappa_4 \cdot \Delta\mathrm{FF} + \kappa_5 \cdot \Delta\mathrm{N2}) + [\gamma - \mu(1-\alpha), \gamma + \mu(1-\alpha)] \tag{5}$$

with $\kappa_i \geq 0$, $g(\mathbf{fc}) >= 0$ for all $\mathbf{fc} \in \mathrm{FC}$ and $f(x) \geq f(y)$ for $x > y$.

It can be safely assumed that engines are without deffects in the first cycles after a shop visit, and $\kappa_1, \ldots, \kappa_5$ and γ, μ can be found by maximizing the interval-valued extension of Kendall's tau [19] between $[\overline{\mathrm{EPR}}]_\alpha$ and the actual thrust values

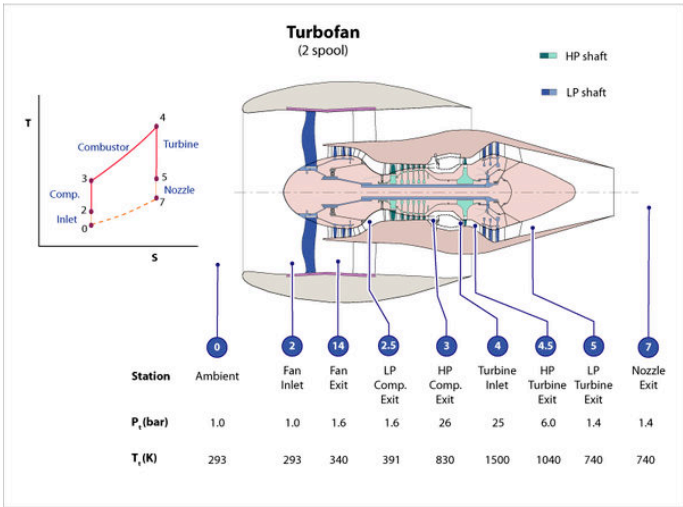


Fig. 1. Outline of a two shaft high bypass ratio turbo fan

EPR^t of the engine at the first $t = 1 \dots N$ cycles, constrained by $\kappa_i \geq 0$, $\gamma > 0$, $\mu > 0$ and

$$\#\{\text{EPR}^t \in [\overline{\text{EPR}}]_\alpha(\Delta\text{P30}^t, \Delta\text{T30}^t, \Delta\text{TGT}^t, \Delta\text{FF}^t, \Delta\text{N2}^t)\} \geq (1-\alpha)\cdot N \quad (6)$$

The expression of the function f is obtained by interpolating the pairs $(\kappa_1 \cdot \Delta\text{P30}^t - \kappa_2 \cdot \Delta\text{T30}^t + \kappa_3 \cdot \Delta\text{TGT}^t + \kappa_4 \cdot \Delta\text{FF}^t + \kappa_5 \cdot \Delta\text{N2}^t + \gamma, \text{EPR}^t)$ with a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP).

Lastly, once the membership function of the fuzzy model is known,

$$\mu_{\widetilde{\text{EPR}^t}}(e) = \mu_{\widetilde{\text{EPR}}}(\Delta\text{P30}^t, \Delta\text{T30}^t, \Delta\text{TGT}^t, \Delta\text{FF}^t, \Delta\text{N2}^t)(e) \quad (7)$$

the residuals of this model can be regarded as a health indicator, as will be shown in an illustrative example in the next Section:

$$\mu_{\widetilde{\text{HEALTH}^t}}(e) = \mu_{\widetilde{\text{EPR}^t}}(e - \text{EPR}^t). \quad (8)$$

In words, the residuals of the proposed model will be centered at zero if the engine is not deteriored, but as soon as a degradation happens (decrease of ΔP30, alone or in combination of an increase of ΔT30, ΔTGT, etc.) the predicted thrust will be lower than that of the non-degraded engine, because of the signs of the coefficients and the monotonicity of the function f. Given that the health of an engine cannot be increased unless it undergones maintainance, isotonic regression techniques are also used to estimate the centerpoints of the aforementioned residuals.

IV. Illustrative example

The new method has been validated on a sample of 330 turbofan engines. A representative case has been chosen that illustrates the properties of this method.

Figure 2 depicts the GPA delta variables of an engine with a light deterioration in the HPT that cannot be found with direct methods, manifold learning, trend analysis neither sequence mining. A sixth variable DN2VIB has been added to the graph; this variable is not being used by the model but serves for validation purposes.

The isotonic fuzzy model has been fitted to the first $N = 1000$ cycles. The whole data spans more than 4000 cycles. Each GPA value is colored according to the EPR (between 1.20 and 1.50). The synthetic HEALTH signal is obtained by fitting a line with isotonic regression to the centerpoints of the residual of the fuzzy model proposed in the preceding section. Observe that the health tends to increase in the first half of the monitored period, but the isotonic regression forces that the signal is monotonically decreasing. Hence, the HEALTH signal is constant in the initial cycles. In the second part of the diagnostic, the residual is decreased and a deterioration in the second half of the figure becomes evident. This signal has been built without supplying the model any kind of information about the signature of a degradation in a similar engine: the only domain knowledge is the monotonicity or antimonotonicity of the GPA delta signals and the thrust. As an additional validation of the method, note that the signal DN2VIB measures the vibrations in the high speed shaft. This information has not been used by the isotonic model, that was nonetheless able to detect a deterioration immediately before the vibrations were measurable.

V. Concluding remarks and future work

Most of times, fault detection are regarded as pattern recognition methods, that can be solved with data-driven techniques with a different amount of domain knowledge. If a large database of engines with different degrees of deterioration is available, direct and subspace methods are the most common fuzzy techniques, however these cannot detect defects that are not in the experience database.

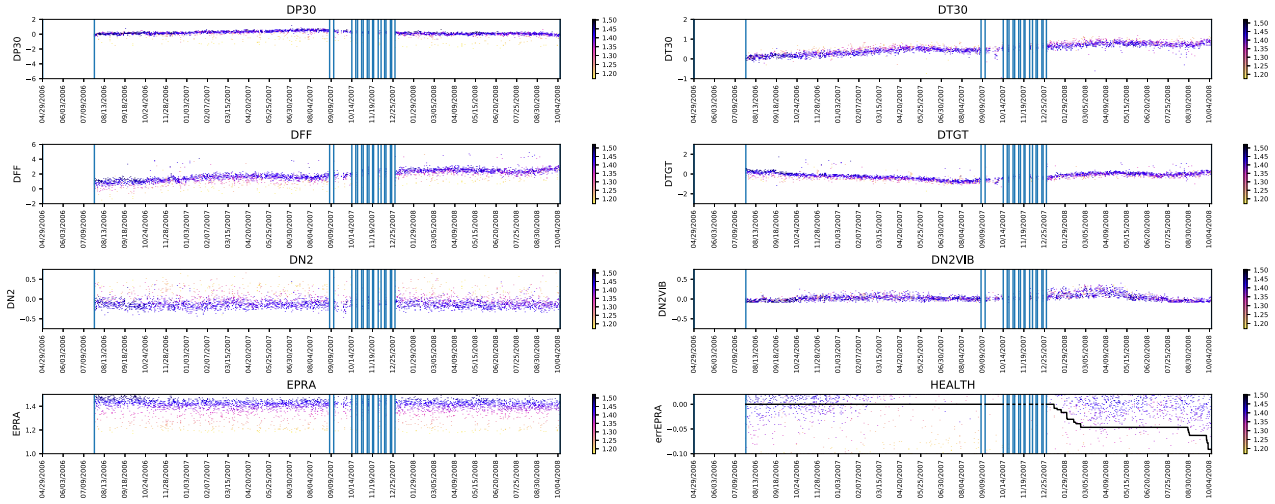


Fig. 2. Example of the application of the isotonic fuzzy model to an actual aeroengine. From left to right, top to bottom: Deltas of P30, T30, FF, TGT, N2, N2VIB (vibrations of the engine, not contemplated in this model) and synthetic HEALTH signal formed by the centerpoints of the isotonic fuzzy model. Observe that the model detects a deterioration of the engine in the last cycles, and this prediction is consistent with the independently measured vibration N2VIB.

The use of delta variables with respect to a reference engine allows applying machine learning algorithms to learn an EHM management system from data, where fuzzy technologies are used for designing the noise filters. The main drawbacks of these methods are the impossibility of detecting the order of the events and the precise cycle when the deterioration initiates. The use of sequence mining solves the first issue, and the isotonic algorithm introduced in this study solves the second problem. Besides its simplicity, the use of a model with a minimal amount of domain knowledge reduces the systematic error of specific models, and at the same time its generalization capabilities improve that of black boxes.

This is a work in progress; there are some aspects of the application of isotonic learning to EHM monitoring that are still unsolved. In particular, it is not yet clear how to attribute the descent in the thrust to the turbine or the compressor. Once this issue is solved, the next step in this research will be to assess the dependence between the RUL of the engine and the jumps in the health signal.

References

[1] Y. Li, "Gas turbine performance and health status estimation using adaptive gas path analysis," *Journal of Engineering for Gas Turbines and Power*, vol. 132, no. 4, p. 041701, 2010.

[2] L. A. Urban, "Gas Path Analysis Applied to Turbine Engine Condition Monitoring," *Journal of Aircraft*, vol. 10, no. 7, 2012.

[3] A. R. Behbahani, "Need for robust sensors for inherently fail-safe gas turbine," in *52nd International Instrumentation Symposium: Engine Controls, Monitoring, and Prognostics. Cleveland, OH*, 2006.

[4] ATSB, "Australian transport safety bureau: Power plant failures in turboprop-powered aircraft 2012 to 2016," https://www.atsb.gov.au/publications/2017/ar-2017-017/, accessed: 2018-06-06.

[5] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.

[6] S. Demirci, C. Hajiyev, and A. Schwenke, "Fuzzy logic-based automated engine health monitoring for commercial aircraft," *Aircraft Engineering and Aerospace Technology*, vol. 80, no. 5, pp. 516–525, 2008.

[7] A. Benyounes, A. Hafaifa, A. Kouzou, and M. Guemana, "Gas turbine modeling using adaptive fuzzy neural network approach based on measured data classification," *Mathematics-in-Industry Case Studies*, vol. 7, no. 4, 2017.

[8] Z. Li, S.-S. Zhong, and L. Lin, "Novel gas turbine fault diagnosis method based on performance deviation model," *Journal of Propulsion and Power*, pp. 1–10, 2016.

[9] C. H. Fontes and O. Pereira, "Pattern recognition in multivariate time series–a case study applied to fault detection in a gas turbine," *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 10–18, 2016.

[10] Z. Chen, H. Fang, and Y. Chang, "Weighted data-driven fault detection and isolation: A subspace-based approach and algorithms," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3290–3298, 2016.

[11] R. Baklouti, M. Mansouri, M. Nounou, H. Nounou, and A. B. Hamida, "Iterated robust kernel fuzzy principal component analysis and application to fault detection," *Journal of Computational Science*, vol. 15, pp. 34–49, 2016.

[12] I. Gueddi, O. Nasri, K. Benothman, and P. Dague, "Fault detection and isolation of spacecraft thrusters using an extended principal component analysis to interval data," *International Journal of Control, Automation and Systems*, vol. 15, no. 2, pp. 776–789, 2017.

[13] W. Yan and L. Yu, "On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach," in *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, 2015.

[14] A. Martinez, L. Sanchez, and I. Couso, "Interval-valued blind source separation applied to ai-based prognostic fault detection of aircraft engines." *Journal of Multiple-Valued Logic & Soft Computing*, vol. 22, 2014.

[15] A. Volponi and B. Wood, "Engine health management for aircraft propulsion systems," in *Proceedings of the Forum on Integrated System Health Engineering and Management (ISHEM) in Aerospace, Napa, CA, November*, 2005, pp. 7–10.

[16] A. Martínez, L. Sánchez, and I. Couso, "Engine health monitoring for engine fleets using fuzzy radviz," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–8.

[17] A. Martinez, L. Sánchez, and I. Couso, "Aeroengine prognosis through genetic distal learning applied to uncertain engine health monitoring data," in *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1945–1952.

[18] A. Palacios, A. Martínez, L. Sánchez, and I. Couso, "Sequential pattern mining applied to aeroengine condition monitoring with uncertain health data," *Engineering Applications of Artificial Intelligence*, vol. 44, pp. 10–24, 2015.

[19] O. Hryniewicz and K. Opara, "Efficient calculation of kendall's tau for interval data," in *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*. Springer, 2013, pp. 203–210.

Análisis de distintas propuestas de sistemas de recomendaciones lingüísticos difusos para personalizar información en bibliotecas digitales

Carlos Porcel
dept. Informática
Universidad de Jaén
Jaén, Spain
cporcel@ujaen.es

Alberto Ching-López
dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, Spain
alch@decsai.ugr.es

Álvaro Tejeda-Lorente
dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, Spain
atejeda@decsai.ugr.es

Juan Bernabé-Moreno
dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, Spain
jbernabemoreno@gmail.com

Enrique Herrera-Viedma
dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, Spain
viedma@decsai.ugr.es

Francisco Javier Cabrerizo
dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, Spain
cabrerizo@decsai.ugr.es

Sergio Alonso
dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, Spain
zerjioi@ugr.es

Los grandes avances que se están produciendo en los últimos años en lo que respecta a las tecnologías de la información y las telecomunicaciones, está cambiando la forma en la que accedemos a la información. Internet es la principal fuente de información y por tanto, está afectando al desarrollo del resto de medios. Tal es el caso de las bibliotecas digitales, que tienen que hacer frente a nuevos problemas provocados por el gran cremiento de contenido digital y aplicaciones basadas en Web [2]. Efectivamente, los usuarios tienen serias dificultades a la hora de acceder a la información deseada de forma sencilla y en el momento oportuno. Es lo que se conoce como problema de sobrecarga de información. El uso de herramientas automáticas ayuda a afrontar y gestionar la gran cantidad de información disponible online en formato digital. También ayuda al personal de la biblioteca en la difusión de información de forma personalizada [1], [2].

Entonces, las herramientas a usar en bibliotecas digitales deben adoptar un papel más activo y ofrecer a los usuarios nuevos servicios que faciliten el acceso y personalización de la información, facilitando además, el que puedan compartir conocimiento. En este sentido, sería muy interesante que las bibliotecas digitales puedan anticipar las necesidades y preferencias de los usuarios, para suministrarles recursos que sean de su interés, incluso aunque no supieran de sus existencia. Pero debido al problema de sobrecarga de información que antes comentábamos, encontramos serias dificultades para conseguir ese objetivo.

Para ello, planteamos el uso de sistemas de recomendaciones, que han demostrado su utilidad en numerosos entornos, para facilitar y personalizar el acceso a la información [1]. Un *sistema de recomendaciones* ayuda a los usuarios a descubrir información de interés y de forma personalizada. Por ello, son especialmente útiles para facilitar a los usuarios el acceso a recursos que antes no conocían. Una de las claves del éxito de estos sistemas es que consiguen una correcta personalización en el sentido en que tratan a cada usuario de forma distinta. Para ello, el sistema debe contar con algún tipo de información sobre los usuarios, como por ejemplo, las valoraciones sobre los recursos que hayan explorado con anterioridad. Con ello, el sistema establece y gestiona sus perfiles, donde quedan representadas las preferencias y necesidades de cada uno. Y hay distintas formas de adquirir ese conocimiento, pudiendo hacerse de forma implícita (es decir, sin intereacción con el usuario) o solicitando que los usuarios expresen sus preferencias de forma explícita.

En este trabajo, versión resumida del trabajo publicado en [3], analizamos distintas propuestas de aplicación de técnicas de recomendación para favorecer la difusión de información en bibliotecas digitales. Pero, dada la gran variedad de formas de representar y evaluar la información cuando los usuarios intervienen (como es el caso que nos ocupa), nos centramos en propuestas basadas en lógica difusa. En concreto, analizamos cuatro propuestas que incorporan el modelado lingüístico difuso multigranular:

1. La primera propone un sistema de recomendaciones

lingüístico difuso que sugiere tanto recursos especializados según el área de interés de cada usuario, como recursos complementarios que podrían ser interesantes para ayudar en la formación de grupos multidisciplinares [6].

2. La segunda presenta un nuevo método para obtener información sobre las preferencias de los usuarios, reduciendo de esta manera el gran esfuerzo que tenían que realizar en propuestas previas y favoreciendo, por tanto, la adquisición de perfiles de usuarios. Los usuarios proporcionan sus preferencias sobre los principales recursos de investigación, usando relaciones de preferencia lingüísticas difusas incompletas. A partir de esta información el sistema construye sus respectivos vectores de preferencias sobre los distintos temas de interés y con ello, genera las recomendaciones [4].
3. En el tercer caso, se plantea una mejora de las propuestas anteriores, basándose en la incorporación de una memoria que ayude a paliar el problema de sobrecarga de información, aún persistente en bibliotecas digitales. La idea consiste en tener en cuenta las preferencias de los usuarios en cuanto al número de recomendaciones de recursos que quieren recibir, y entonces, usar recursos previamente seleccionados aunque no recomendados (por dichas restricciones de cantidad) para formar una nueva selección de recursos en nuevas rondas de recomendación [5].
4. La idea principal desarrollada en la última de las propuestas analizadas, consiste en considerar el proceso de generación de recomendaciones como una tarea con dos elementos distintos. Se trata de encontrar, por un lado, recursos de investigación que sean relevantes para los usuarios, pero por otro lado, encontrar recursos que sean válidos desde el punto de vista de la calidad de los mismos. Entonces, se hace una estimación tanto de la relevancia de los recursos como de su calidad, y ambos valores, se agregan en una nueva métrica que se tiene en cuenta para recomendar o no un recurso [7].

En todo este panorama, vemos que efectivamente en los últimos años se han ido consiguiendo bastantes logros gracias a la aplicación y rápido avance de las tecnologías de la información y las comunicaciones. Se puede destacar la posibilidad de trabajar con enormes cantidades de información, y no sólo con información basada en texto, sino incluyendo imágenes, audio, vídeo y material multimedia. Pero es crucial, no sólo trabajar con esas cantidades y distintos formatos de información, sino hacerlo de forma personalizada. Hasta ahora se han venido aplicando numerosas técnicas pero en determinadas ocasiones son los propios usuarios los que deben gestionar la información de forma explícita [6]. En este contexto, podría ayudar la aplicación de técnicas de aprendizaje automático que posibiliten automatizar la construcción de modelos de usuarios. También se han hecho algunos avances en este sentido [4], [5], [7], pero a pesar de los logros conseguidos, su principal limitación es la necesidad de interaccionar con el personal de la biblioteca para establecer las representaciones internas de los recursos. Por ello, encontramos en este punto un reto significativo al que se deberá hacer frente, así como seguir adaptándose a las nuevas casuísticas de la sociedad actual, como puede ser el auge experimentado en el uso de redes sociales.

Como conclusión, creemos que a pesar de los grances avances llevados a cabo, las oportunidades y retos para las futuras bibliotecas digitales pasan por esta triple perspectiva:

- Dado que la información generada de forma electrónica continúa creciendo y se preveé un crecimiento aún mayor, habrá que hacer frente a ello con nuevas herramientas o técnicas. La idea es que favorezcan la personalización de los servicios y la difusión de información de forma personalizada, por lo que habría que seguir mejorando en el establecimiento automático de los perfiles de usuarios y en el estudio y aplicación de técnicas automáticas que permitan obtener la representación interna de los recursos.
- Considerar el entono social de los usuarios. La idea sería explotar información diversa extraída de los distintos medios sociales en los que participan los usuarios, lo que permitiría ayudar a la construcción automática de perfiles de usuarios.
- Proponer e implantar sistemas reales que faciliten las labores de difusión de información en bibliotecas digitales, y que con su uso lo más extendido posible, permitan obtener datos suficientes que faciliten la evaluación de las distintas propuestas que puedan ir surgiendo.

Agradecimientos

Este trabajo ha sido desarrollado gracias a la financiación FEDER del Proyecto TIN2016-75850-R.

Referencias

[1] M. Franke, A. Geyer-Schulz, and A. Neumann, “Recommender services in scientific digital libraries”, in Multimedia Services in Intelligent Environments. Heidelberg: Springer, pp. 377-417, 2008.

[2] R. D. Montoya, “Boundary objects/boundary staff: supporting digital scholarship in academic libraries”, Journal of Academic Librarianship, vol. 43(3), pp. 216-223, 2017.

[3] C. Porcel, A. Ching-López, J. Bernabé-Moreno, A. Tejeda-Lorente, and E. Herrera-Viedma, “Fuzzy Linguistic Recommender Systems for the Selective Diffusion of Information in Digital Libraries”, J. Inf. Process Syst, Vol.13(4), pp. 653-667, 2017.

[4] C. Porcel, and E. Herrera-Viedma, “Dealing with Incomplete Information in a Fuzzy Linguistic Recommender System to Disseminate Information in University Digital Libraries”, Knowledge-Based Systems, vol. 23, pp. 32-39, 2010.

[5] C. Porcel, J.M. Morales-del-Castillo, M.J. Cobo, A.A.. Ruíz, and E. Herrera-Viedma, “An Improved Recommender System to Avoid the Persistent Information Overload in a University Digital Library”, Control and Cybernetics-ICAE Journal Special Issue, vol. 39(4), pp. 899-924, 2010.

[6] C. Porcel, J.M. Moreno, and E. Herrera-Viedma, “A Multi-Disciplinar Recommender System to Advice Research Resources in University Digital Libraries”, Expert Systems with Applications, vol. 36(10), pp. 12520-12528, 2009.

[7] A. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz, and E. Herrera-Viedma, “A quality based recommender system to disseminate information in a University Digital Library”, Information Science, vol. 261, pp.52-69, 2014.