

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

ESTYLF 11: SESIÓN ESPECIAL  
SOFT COMPUTING Y  
GENERACIÓN DEL  
LENGUAJE NATURAL I

**Organizadores:**

ALBERTO BUGARÍN, NICOLÁS MARÍN,  
ALEJANDRO RAMOS, DANIEL SÁNCHEZ







# Modelado borroso de referencias geográficas textuales sobre datos de expertos

A. Ramos-Soto<sup>\*†</sup>, Jose M. Alonso<sup>\*</sup>, Ehud Reiter<sup>†</sup>, Kees van Deemter<sup>†‡</sup>, Albert Gatt<sup>§</sup>

<sup>\*</sup> Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), Universidade de Santiago de Compostela  
alejandro.ramos@usc.es  
josemaria.alonso.moral@usc.es

<sup>†</sup> Department of Computing Science, University of Aberdeen  
alejandrosoto@abdn.ac.uk  
e.reiter@abdn.ac.uk  
k.vdeemter@abdn.ac.uk

<sup>‡</sup> Department of Information and Computing Sciences, Utrecht University  
k.vandeemter@uu.nl

<sup>§</sup> Institute of Linguistics and Language Technology, University of Malta  
albert.gatt@um.edu.mt

**Resumen**—Describimos una metodología para la construcción de modelos borrosos de expresiones geográficas sobre interpretaciones individuales de dichas expresiones por parte de expertos. Esta metodología aborda una tarea de modelado de lenguaje encuadrado dentro del desarrollo de un sistema *data-to-text* que generará descripciones textuales sobre mapas con información meteorológica en tiempo real. Para ello, recogimos datos proporcionados por meteorólogos en una encuesta y, sobre los mismos, creamos modelos borrosos consistentes que agregan las diferencias de interpretación de los expertos. Estos modelos permitirán generar expresiones de referencia geográfica sobre eventos meteorológicos georreferenciados, que formarán parte de los textos generados por el sistema *data-to-text* que desarrollaremos en un trabajo futuro.

**Términos clave**—vagueza, generación de lenguaje natural, conjuntos borrosos, modelado de lenguaje, sistemas de información geográfica, descripciones lingüísticas de datos

## I. INTRODUCCIÓN

La ingente cantidad de datos que se producen en todo tipo de ámbitos ha favorecido la aparición de varias disciplinas centradas en investigar cómo proporcionar a las personas la información relevante que permanece latente en dichos datos. Una de estas disciplinas es la generación de lenguaje natural (NLG), que estudia el problema de cómo generar textos a partir de datos, que puedan resultar útiles a lectores humanos [1], [2]. Dentro de este campo, los sistemas que generan textos a partir de datos no lingüísticos se conocen como *data-to-text* (D2T). En los últimos tiempos este tipo de sistemas gozan de cierta popularidad, gracias a su amplio uso comercial [2], [3] en una gran variedad de dominios.

Gran parte de los sistemas D2T proporcionan textos o informes que describen series de datos temporales, y ciertamente se pueden encontrar muchos ejemplos en el estado del arte, e.g., en predicción meteorológica [4]–[7], salud [8]–[10], o industria [11], entre muchos otros [1], [12]. Así mismo, también existen sistemas D2T que tratan datos caracterizados geográficamente, aunque su número es mucho más reducido respecto a los alimentados por series temporales de datos [13].

Una de las tareas esenciales en la concepción de un sistema D2T es el modelado del lenguaje, i.e., definir la semántica de los términos y expresiones a usar para describir los datos [14], [15]. Existen distintas formas de resolver esta tarea, tales como usar algoritmos heurísticos o de aprendizaje máquina sobre un

conjunto de *corpus* paralelo de texto y datos [16] para crear modelos de las expresiones de interés [17], conseguir que los expertos proporcionen dichos modelos, o recolectar datos de escritores o lectores que puedan ser usados para la aplicación de algoritmos de mapeado.

Por ejemplo, los sistemas D2T que generan textos a partir de series temporales de datos incluyen generalmente expresiones temporales para referirse a eventos o patrones relevantes encontrados en los datos. Realizar la tarea de modelado de lenguaje sobre expresiones temporales en este tipo de sistemas permite asegurarnos de que los textos generados incluyen términos cuyo significado se alinea con la interpretación de los expertos o las expectativas de los lectores, [17], [18]. Del mismo modo, disponer de un buen modelo geográfico en sistemas que generan textos sobre datos geográficos es esencial para generar expresiones que se refieran a localizaciones concretas y a regiones de interés.

Al mismo tiempo, es bastante frecuente que las expresiones temporales y geográficas que deben ser incluidas en los textos generados por sistemas D2T sean vagas, como “por la tarde” [4] o “áreas del suroeste” [19]. En situaciones donde la vaguedad aparece (y con ella, casos fronterizos y conceptos graduales), se ha propuesto el uso de conjuntos borrosos para modelar expresiones lingüísticas en sistemas D2T [2], [20], [21]. Sin embargo, no existen sistemas de este tipo en el estado del arte que hagan uso de dichas técnicas, con la excepción de GALiWeather [5], que hace un uso básico de conjuntos borrosos para modelar expresiones temporales y cuantificadores.

En este contexto, este trabajo describe la metodología que hemos seguido para llevar a cabo una tarea de modelado de lenguaje de expresiones geográficas vagas. Esta tarea forma parte de un proyecto más amplio cuyo objetivo es la creación de un sistema D2T que genere descripciones textuales de mapas con información meteorológica en tiempo real. Nuestra aproximación se compone de una tarea de recogida de datos, proporcionados por expertos, y de un algoritmo heurístico que agrega dichos datos para crear modelos borrosos.

## II. TRABAJO RELACIONADO

El campo de NLG es amplio y existen gran cantidad de tipos de sistemas existentes con distintos propósitos como la gen-

eración de informes a partir de datos, creación de resúmenes a partir de distintas fuentes textuales, o generación de diálogo, narrativas e incluso poesía [1]. En nuestro caso, nos centramos en un problema muy específico, aquellos sistemas D2T cuyos datos de entrada están caracterizados geográficamente y cuyos textos a generar incluyen expresiones geográficas que se refieren a la ocurrencia de ciertos eventos registrados en los datos (e.g., “llovía en el norte de España”, “inundaciones en la Costa del Sol” o “vientos fuertes en el noroeste de Escocia”).

Aunque este tipo de referencias fueron introducidas hace décadas en los textos generados por el sistema FoG [6], RoadSafe es quizás el ejemplo más representativo y reciente de este tipo de sistemas [19], [22]. RoadSafe usaba datos de predicción meteorológica para generar predicciones textuales orientadas al mantenimiento de carreteras. Estos informes incluían expresiones temporales y geográficas para ayudar a identificar dónde y cuándo ciertos fenómenos relevantes tendrían lugar, con el objetivo de ayudar a los equipos de mantenimiento a mantener las condiciones de las vías afectadas en buen estado (ver ejemplos de dichas expresiones en Fig. 1). Así pues, desarrollar RoadSafe supuso también un estudio profundo de cómo generar buenas expresiones geográficas que se refiriesen de forma adecuada a la geografía subyacente a la información relevante extraída de los datos de entrada.

Concretamente, la aproximación de RoadSafe para modelar la generación de expresiones de referencia geográficas se basaba en técnicas estándar del campo de sistemas de información geográfica (GIS), en el que particionan la geografía subyacente de los eventos utilizando distintos esquemas, o *marcos de referencia* espaciales [23], que a su vez se componen de particiones no solapables (conocidas como *descriptores*). Por ejemplo, el marco de referencia *Dirección* se compone de los descriptores “nordeste”, “suroeste”, etc., y el marco “Proximidad Costera” se compone de los descriptores “costa” e “interior”.

Una vez que los límites numéricos de cada descriptor se definen para cada marco usando coordenadas de latitud-longitud, cada punto de datos puede ser caracterizado por un conjunto de descriptores (ej. “suroeste” y “costa”) y el generador de expresiones de referencia geográficas se encarga de seleccionar el mejor conjunto de descriptores que describan el área formada por el subconjunto de puntos que representan el evento.

En trabajos más recientes, el modelo geográfico utilizado en

Road surface temperatures will fall below zero **during the late evening and tonight** except in **areas below 100M**.

SW 10-25 gusts **this afternoon in southwestern areas**, veering WSW and increasing 15-35 after midnight, gusts 55-60 **during the evening and tonight** except in **areas above 500M**, increasing 20-45 then veering W **by early morning**, gusts 70-75 **tomorrow morning in most southern and central places**.

Wintery precipitation will affect most routes at first, falling as snow flurries in **some places above 300M** at first. Snow spreading **throughout the forecast period** to **all areas** and persisting in **some places above 300M** until end of period.

Fig. 1. Ejemplos de textos generados por RoadSafe [19]

RoadSafe fue ampliado mediante la inclusión de referencias espaciales de nombres propios, que según un estudio de varios conjuntos de *corpus* en distintos dominios, son las más predominantes [13]. Además, se plantea el desarrollo de un algoritmo de generación de expresiones de referencia geográficas que integre el marco de referencia de nombres propios con los ya existentes en RoadSafe. Por otro lado, en [24] se tratan las diferencias entre referencias absolutas y relativas y se proporciona un modelo basado en la mereología, en la que los descriptores de nombre propio se combinan con descriptores de otros marcos de referencia.

Las referencias descritas hasta el momento proporcionan buenas aproximaciones que permiten generar expresiones de referencia geográfica apropiadas sobre un conjunto de marcos de referencia. Sin embargo, en todos los casos los modelos descritos fueron desarrollados basados en un particionado nítido de la geografía a tratar. Tomar límites exactos entre descriptores geográficos puede considerarse una asunción poco intuitiva, especialmente si tenemos en cuenta cómo las personas entendemos y usamos incluso las referencias geográficas más simples como “norte” u “oeste”, en los que realmente las fronteras no suelen estar bien definidas, sino que son vagas. Por ejemplo, usando cualquiera de las aproximaciones anteriores, si una aldea se encuentra situada en la frontera entre dos descriptores, dependiendo de la granularidad espacial de nuestros datos es posible asignar descriptores opuestos (como “norte” y “sur”) a dos puntos dentro de la misma localidad.

Por ello, las limitaciones que en este sentido presentan los modelos descritos anteriormente suponen un fuerte incentivo a la hora de buscar otro tipo de aproximaciones que permitan modelar la imprecisión o la incerteza en el uso de referencias geográficas vagas. De hecho, el problema del tratamiento de la vaguedad en referencias geográficas no se limita únicamente a sistemas D2T, sino que existe una discusión más amplia dentro del ámbito de GIS, desde hace ya varias décadas [25]. En este sentido, la teoría de conjuntos borrosos ha sido aplicada en numerosos casos para tratar la vaguedad en conceptos geográficos y relaciones espaciales [26], [27].

Del mismo modo, como se ha descrito en la Sec. I, de una manera más general también se ha propuesto el uso de conjuntos borrosos en sistemas D2T para modelar términos vagos [2], [20], [21]. A día de hoy, el único sistema D2T desplegado en un entorno real que hace un uso básico de este tipo de técnicas es GALiWeather [5], si bien existe un número importante de casos de uso de aplicación de conjuntos borrosos para extracción de información lingüística (descripciones lingüísticas de datos), que en ocasiones se acompaña de generación textual basada en plantillas [2], [12].

### III. MOTIVACIÓN

Partiendo de las limitaciones que presentan los desarrollos previos de sistemas D2T que generan expresiones de referencia geográfica, la principal motivación de este trabajo es mejorar el modelado de conceptos geográficos vagos para fines de generación de lenguaje natural. Concretamente, nuestro objetivo es establecer una metodología de creación de modelos de



referencias geográficas vagas y de algoritmos de generación de expresiones de referencia sobre dichos modelos, basados en el uso de técnicas de conjuntos borrosos.

Puesto que D2T es un campo eminentemente aplicado, en el que los avances a nivel de investigación vienen dados por una necesidad real, en nuestro caso la metodología que proponemos se enmarca dentro del desarrollo de un sistema D2T para la generación de descripciones del estado meteorológico en tiempo real, sobre datos proporcionados por la Agencia de Meteorología de la Xunta de Galicia, MeteoGalicia [28]. Dichas descripciones incluirán expresiones de referencia geográfica que permitirán identificar fenómenos meteorológicos relevantes en el mapa, tales como temperatura, viento y estado del cielo.

En el marco del desarrollo del sistema D2T propuesto, este trabajo describe la tarea de modelado de lenguaje de las expresiones geográficas de interés a incluir en los textos generados por el sistema. Por otro lado, la metodología aquí descrita se basa en las ideas propuestas en [29] y [30], lo que permitirá consolidar nuevas formas de aplicar la teoría de conjuntos borrosos en sistemas D2T de forma práctica.

IV. RECOGIDA DE DATOS DE EXPERTOS

Aunque es corriente realizar el modelado de lenguaje partiendo de un conjunto paralelo de textos y datos, con el fin de analizar el significado de las palabras y expresiones a modelar, en nuestro caso este tipo de recurso no se encontraba disponible. Por un lado, no disponíamos de acceso a un conjunto de datos extenso (ej. predicciones textuales y datos de predicción) en los que realizar tal análisis. Por otro, nuestro plan es desarrollar un nuevo sistema D2T para proporcionar descripciones textuales de datos meteorológicos en tiempo real. Así pues, la aproximación que tomamos consistió en interactuar de forma directa con los expertos.

A. La encuesta

Dado que nuestro propósito es el modelado de expresiones geográficas, pedimos directamente al director del departa-

Marco de referencia		Descriptor
Dirección cardinal		Norte de Galicia, Sur de Galicia, Oeste de Galicia, Este de Galicia, Tercio norte, Extremo norte, Noroeste de Galicia, Noreste de Galicia, Suroeste de Galicia, Sureste de Galicia
Costa / Interior		Interior de Galicia
Nombre propio		Rías Baixas, Comarcas atlánticas
Mixto	Nombre y dirección	Oeste de A Coruña, Oeste de Ourense, Sur de Ourense, Sur de Lugo
	Nombre y costa	Litoral Atlántico, Litoral Cantábrico, Litoral norte, Interior de Coruña, Interior de Pontevedra
	Nombre y elevación	Áreas de montaña de Lugo, Áreas de montaña de Ourense

TABLA I  
LISTA DE REFERENCIAS GEOGRÁFICAS DE LA ENCUESTA.

mento de predicción de *MeteoGalicia* una lista con las expresiones geográficas más usadas por los meteorólogos a la hora de escribir predicciones textuales. Tomando esta lista como base, preparamos una encuesta web que fue distribuida entre los expertos de la agencia meteorológica. En dicha encuesta, se pidió a los participantes que, sobre un mapa de la región de Galicia (mostrada bajo una proyección *Mercator*), dibujasen un polígono que representase una referencia geográfica dada (ver la Figura 2).

En la encuesta se proporcionó a los participantes una lista de 24 descriptores, que aparecían en orden aleatorio. En esta lista, 20 de los 24 descriptores componen la lista original proporcionada por el director de predicción operativa, e incluyen direcciones cardinales, nombres propios y otro tipo de referencias como áreas montañosas, partes de provincias, etc. (ver la Tabla I para una taxonomía completa de los descriptores). Los restantes cuatro fueron añadidos para estudiar la combinación de direcciones cardinales mediante intersección (ej. explorar formas de combinar “norte” y “oeste” para obtener un modelo similar a “noroeste”), pero para nuestro propósito actual nos interesan sólo aquellos utilizados por los meteorólogos a la hora de escribir predicciones.

B. Resultados

La encuesta fue contestada por ocho expertos, obteniéndose 192 polígonos en total (160 sin considerar las intersecciones cardinales). A un nivel general, habíamos hipotetizado que los expertos serían bastante consistentes, dada su experiencia profesional. Así mismo, esperábamos también cierta variación entre las distintas respuestas.

Observamos que dichas hipótesis se han cumplido con claridad; los polígonos dibujados por los expertos se encuentran bastante concentrados y por tanto existe un gran acuerdo entre ellos. Por ejemplo, la Fig. 3 muestra una representación de las respuestas dadas por los meteorólogos para la dirección cardinal “oeste de Galicia” y un mapa de contornos que ilustra el porcentaje de respuestas que se solapan.

V. CREACIÓN DE DESCRIPTORES GEOGRÁFICOS BORROSOS

En la Figura 3, la gráfica de contornos puede tomarse como la base de la semántica de su expresión correspondiente,



Fig. 2. Captura de pantalla de la encuesta realizada por los meteorólogos.



Fig. 3. Representación gráfica de los polígonos dibujados por los expertos y gráfica de contornos para “oeste de Galicia”.

con una región nuclear que es aceptada por la mayoría, y un decaimiento gradual a medida que se avanza hacia la periferia exterior de las líneas de contorno. Así pues, en nuestro caso, la imprecisión surge de las diferencias interpersonales entre los meteorólogos.

Siguiendo esta noción, hemos creado modelos borrosos que agregan las opiniones de los expertos para cada descriptor. El método que usamos para esta tarea de modelado es una mejora del algoritmo heurístico descrito en [30], que producía modelos básicos basado en un muestreo de puntos y el conteo de intersecciones de polígonos, sin la inclusión de ninguna condición previa. Nuestro algoritmo va precedido además por un filtrado simple de los polígonos.

#### A. Filtrado de datos atípicos

Como comentamos anteriormente, los polígonos dibujados por los meteorólogos son muy consistentes visualmente, pero en algunos casos hemos observado pequeñas inconsistencias. Para mantener la elevada consistencia entre las respuestas obtenidas para los descriptores, aplicamos un filtrado simple de los trazados, consistente en descartar respuestas fuera del intervalo [media  $\pm$  2\*desviación típica] en términos de tamaño y localización del centroide.

#### B. Caracterización de un descriptor geográfico borroso

Usando los datos filtrados, pretendemos construir descriptores geográficos borrosos que sean simples y consistentes.

**Definition 1.** Descriptor geográfico borroso,  $G$ :

$$G = \{S, K, \mu_G\} \quad (1)$$

Formalmente, definimos un descriptor geográfico borroso  $G$  (ej. “sur de Galicia”) como un conjunto de 3 elementos: un área de soporte  $S$ , un área de núcleo  $K$ , y una función de pertenencia borrosa  $\mu_G$ , que evalúa el grado en el que un punto en un mapa ( $p = (x, y) \mid x, y \in \mathbb{R}$ ) puede considerarse parte de  $G$ :

$$\mu_G : \{\mathbb{R}, \mathbb{R}\} \rightarrow [0, 1] \quad (2)$$

Tomando  $\mu_G^1$  como base,  $K$  y  $S$  pueden definirse como:

$$K = \{p\} \mid \mu_G(p) = 1 \quad (3)$$

$$S = \{p\} \mid \mu_G(p) > 0 \quad (4)$$

Así pues,  $K$  es el conjunto de puntos (o región) cuyos grados de pertenencia son máximos con respecto a  $G$ , mientras que el soporte incluye a todos los puntos con un grado de pertenencia mayor que cero. Sin embargo, para la consecución de modelos consistentes, necesitamos aplicar las siguientes restricciones a  $G$ :

$$\forall G, K \subseteq S \quad (5)$$

$$\forall \{p_i, p_j\} \mid p_i, p_j \in S, p_i, p_j \notin K \text{ y } p_i \neq p_j \quad (6)$$

$$d(p_i, K) > d(p_j, K) \iff \mu_G(p_i) < \mu_G(p_j)$$

Estas condiciones aseguran que los modelos borrosos sean consistentes, al evitar la posibilidad de obtener  $K$ s y  $S$ s disjuntos, y asegurar la monotonicidad para  $\mu_G$ , donde  $d(p, K)$  es la distancia euclídea a  $K$  desde un punto  $p$  en  $S$ .

#### C. Construyendo descriptores sobre los datos de expertos

Un descriptor geográfico borroso  $G$ , como “norte de Galicia”, se modela de acuerdo con los polígonos dibujados por los expertos para dicha expresión. Formalmente, la colección de polígonos dibujados para un  $G$  específico se define como:

$$R_G = \{P_1, P_2, \dots, P_n\} \quad (7)$$

Cada  $P$  representa un polígono, y  $n$  es el número total de polígonos restantes tras el filtrado inicial. Cada polígono se compone de un conjunto de vértices, definidos bajo una proyección *plate carrée* (pares de valores de longitud y latitud).

La primera tarea para modelar un  $G$  dado consiste en determinar sus constituyentes ( $S$  y  $K$ ), puesto que calcular ambos nos permitirá caracterizar  $\mu_G$  posteriormente. Para ello, en primer lugar se transforman las coordenadas de los polígonos en  $R_G$  desde una proyección *plate carrée* a una proyección *Mercator*. A continuación, se define una malla de puntos equidistantes sobre ambos ejes cartesianos,  $D = (p_1, \dots, p_i, \dots, p_{|D|})$ , que se encuentra delimitada por la extensión máxima de la geografía subyacente (en nuestro caso, la región de Galicia). La distancia entre los puntos de la malla viene determinada por un parámetro  $\delta$ , que especifica un porcentaje de la extensión total del mapa. Por ejemplo,  $\delta = 1$  significa que la distancia entre un par de puntos de la malla contiguos es igual al 1% de la distancia entre límites opuestos del map sobre uno de los ejes.

Como se especifica en el Algoritmo 1, usando  $D$  y la colección de trazados expertos  $R_G$ , calculamos el número de veces que cada  $p$  en  $D$  está contenido en un polígono  $P$  en  $R_G$ , para determinar el porcentaje de polígonos que se solapan en un  $p$  dado. Basándonos en los porcentajes calculados para todos los puntos en  $D$ , determinamos  $K$  usando una aproximación de mayoría simple. Por tanto,  $PK$  está compuesto de

<sup>1</sup>Por simplicidad, nos referiremos a  $\mu_G(p)$  en vez de  $\mu_G(x, y)$ .


**Algorithm 1** Cálculo de  $K$  y  $S$  para un descriptor  $G$ 


---

**Entrada:**  $D, R_G$   
**Salida:**  $K, S$

```

1:  $PCS \leftarrow ()$ 
2: for all  $p_i \in D$  do
3:    $count \leftarrow 0$ 
4:   for all  $P_j \in R_G$  do
5:     if  $p_i \in P_j$  then
6:        $count \leftarrow count + 1$ 
7:     end if
8:   end for
9:    $pcs_i \leftarrow count / |R_G|$ 
10:   $PCS \leftarrow PCS \cup pcs_i$ 
11: end for
12:  $PK \leftarrow \{p_i \in D \mid pcs_i > 0.5\}$ 
13:  $PS \leftarrow \{p_i \in D \mid pcs_i > 0\}$ 
14:  $K \leftarrow ConvexHull(PK)$ 
15:  $S \leftarrow ConvexHull(PS)$ 
16:  $bp \leftarrow \operatorname{argmax}_{p_i \in S}(d(p_i, K))$ 
17:  $op \leftarrow \operatorname{argmin}_{p_i \in D \text{ y } p_i \notin S}(d(p_i, bp))$ 
18:  $od \leftarrow d(K, op)$ 
19: return  $K, S, od$ 

```

---

aquellos puntos cuyos porcentajes son  $> 50\%$ , y  $PS$  cubre todos los puntos donde el porcentaje es  $> 0\%$ . En vez de considerar  $S$  y  $K$  como colecciones de puntos en  $D$  (lo que  $PK$  y  $PS$  son), calculamos sus envolventes convexas, esto es, los polígonos convexas que delimitan sus áreas. Este proceso elimina cualquier área disjunta perteneciente a  $K$  que pueda aparecer por divergencias entre los polígonos dibujados, ya que el envolvente las agrega bajo una única área. Finalmente, calculamos  $od$ , la suma de la distancia del vértice de  $S$  más lejano a  $K$  ( $bp$ ) y la distancia mínima de ese punto a otro externo a  $S$  ( $op$ ), que será usada en la definición de  $\mu_G$ , junto con  $K$  y  $S$ .

#### D. Evaluación de un descriptor geográfico borroso

Tomando como base los tres elementos devueltos por el Algoritmo 1, la función  $\mu_G$  que evalúa un punto  $p$  se define como:

$$\mu_G(p) = \begin{cases} 1 & \text{si } p \in K \\ 1 - d(p, K)/od & \text{si } p \in S \text{ y } p \notin K \\ 0 & \text{si } p \notin S \end{cases} \quad (8)$$

La función de pertenencia  $\mu_G$  se define siguiendo la condición de monotonicidad. Bajo dicha definición, todos los puntos en  $S$  tienen grados de pertenencia  $> 0$  y este grado decrece a medida que el punto evaluado se aleja de  $K$ . En estas condiciones, la caracterización de un descriptor geográfico borroso  $G$  mediante la definición de su soporte y su núcleo nos permite crear modelos simples que son consistentes y fáciles de interpretar.

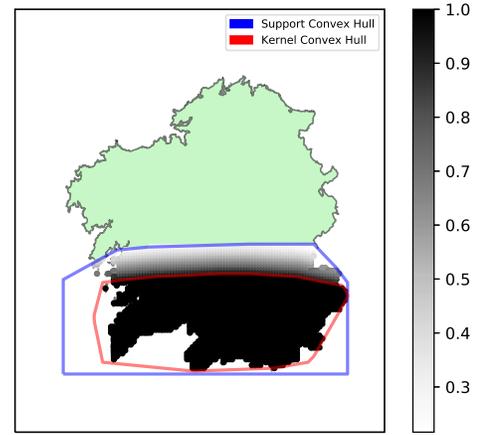


Fig. 4. Representación gráfica del descriptor geográfico borroso “Sur de Galicia”.

#### E. Ejemplos

Con el fin de ilustrar los resultados del algoritmo de modelado y explicar algunas de las decisiones integradas en el Algoritmo 1 y en la definición de  $\mu_G$ , describimos a continuación dos modelos borrosos distintos que resultan de la agregación de los polígonos dibujados por los meteorólogos para dos expresiones geográficas bajo distintas categorías o marcos de referencia en la Tabla I.

El primer descriptor, mostrado en la Figura 4, modela la expresión “sur de Galicia”. Este modelo puede considerarse muy regular, tanto para  $K$  como para  $S$ . La condición de mayoría simple asegura además un mayor consenso a la hora de definir el significado de “sur”. Este descriptor proporciona un área extensa en el que los puntos interiores pueden considerarse como parte del “sur de Galicia” en distintos grados.

Otros descriptores adoptan distintas formas, donde  $S$  varía en ancho respecto a  $K$ . Este es el caso, por ejemplo de “áreas de montaña de Lugo”. Como muestra la Figura 5, la forma de este descriptor es en cierto modo elipsoidal y la distancia

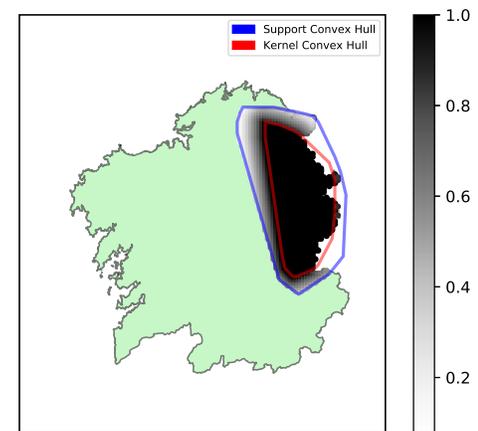


Fig. 5. Representación gráfica del descriptor geográfico borroso “Áreas montañosas de Lugo”.

entre los límites de  $S$  y  $K$  no es constante, comparada con la Figura 4. Este caso ilustra bien cómo algunos puntos en  $S$  más cercanos a su periferia tienen grados de pertenencia más elevados que otros más lejanos, debido a su cercanía a  $K$ .

## VI. CONCLUSIONES

En este trabajo hemos descrito una metodología basada en la aplicación de conjuntos borrosos para la realización de una tarea de modelado de lenguaje, en el contexto del desarrollo de un sistema D2T que generará descripciones meteorológicas georreferenciadas. Esta metodología incluye la recogida de datos expertos sobre referencias geográficas y su agregación mediante la construcción de descriptores borrosos. Como trabajo futuro, desarrollaremos un algoritmo de generación de expresiones de referencia geográficas que utilice los descriptores descritos en este trabajo, que constituirá además el núcleo del sistema D2T.

## AGRADECIMIENTOS

A. Ramos-Soto es investigador postdoctoral financiado por la “Consellería de Cultura, Educación e Ordenación Universitaria” (481B 2017/030) y J.M. Alonso es Investigador Ramón y Cajal (RYC-2016-19802). Además, este trabajo está parcialmente financiado por los proyectos TIN2017-90773-REDT (iGLN), TIN2017-84796-C2-1-R (BIGBISC), TIN2014-56633-C3-1-R (BAI4SOW) y TIN2014-56633-C3-3-R (ABS4SOW) cofinanciados por el “Ministerio de Economía y Competitividad.” También reconocemos el apoyo de la Xunta de Galicia (Centro singular de investigación de Galicia acreditación 2016-2019) y la Unión Europea (Fondo FEDER - European Regional Development Fund - ERDF).

## REFERENCIAS

- [1] A. Gatt and E. Kraemer, “Survey of the state of the art in natural language generation: Core tasks, applications and evaluation,” *Journal of Artificial Intelligence Research*, vol. 61, pp. 65–170, 2018.
- [2] A. Ramos-Soto, A. Bugarín, and S. Barro, “On the role of linguistic descriptions of data in the building of natural language generation systems,” *Fuzzy Sets and Systems*, vol. 285, pp. 31–51, 2016.
- [3] Gartner, “Neural Networks and Modern BI Platforms Will Evolve Data and Analytics,” <http://www.gartner.com/smarterwithgartner/neural-networks-and-modern-bi-platforms-will-evolve-data-and-analytics/>, accessed: 2017-03-14.
- [4] S. Sripada, E. Reiter, and I. Davy, “Sumtime-mousam: Configurable marine weather forecast generator,” *Expert Update*, vol. 6, no. 3, pp. 4–10, 2003.
- [5] A. Ramos-Soto, A. Bugarín, S. Barro, and J. Taboada, “Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data,” *Fuzzy Systems, IEEE Transactions on*, vol. 23, no. 1, pp. 44–57, Feb 2015.
- [6] E. Goldberg, N. Driedger, and R. Kittredge, “Using natural-language processing to produce weather forecasts,” *IEEE Expert*, vol. 9, no. 2, pp. 45–53, 1994.
- [7] J. Coch, “Interactive generation and knowledge administration in multi-meteo,” in *Proceedings of the Ninth International Workshop on Natural Language Generation, Niagara-on-the-lake, Ontario, Canada, 1998*, pp. 300–303, software demonstration.
- [8] J. Hunter, Y. Freer, A. Gatt, E. Reiter, S. Sripada, and C. Sykes, “Automatic generation of natural language nursing shift summaries in neonatal intensive care: Bt-nurse,” *Artificial Intelligence in Medicine*, vol. 56, no. 3, pp. 157 – 172, 2012.
- [9] E. Reiter, R. Robertson, and L. Osman, “Types of knowledge required to personalise smoking cessation letters,” in *Artificial Intelligence and Medicine: Proceedings of AIMDM-1999*, W. Horn, Ed. Berlin, New York: Springer, 1999, pp. 398–399.
- [10] A. Goldstein and Y. Shahar, “An automated knowledge-based textual summarization system for longitudinal, multivariate clinical data,” *Journal of biomedical informatics*, vol. 61, pp. 159–175, 2016.
- [11] J. Yu, E. Reiter, J. Hunter, and S. Sripada, “Sumtime-turbine: A knowledge-based system to communicate gas turbine time-series data,” in *Developments in Applied Artificial Intelligence*, ser. Lecture Notes in Computer Science, P. Chung, C. Hinde, and M. Ali, Eds. Springer Berlin Heidelberg, 2003, vol. 2718, pp. 379–384.
- [12] N. Marin and D. Sánchez, “On generating linguistic descriptions of time series,” *Fuzzy Sets and Systems*, vol. 285, pp. 6 – 30, 2016, special Issue on Linguistic Description of Time Series.
- [13] R. de Oliveira, Y. Sripada, and E. Reiter, *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*. Association for Computational Linguistics, 2015, ch. Designing an Algorithm for Generating Named Spatial References, pp. 127–135.
- [14] D. Roy and E. Reiter, “Connecting language to the world,” *Artificial Intelligence*, vol. 167, no. 1-2, pp. 1–12, 2005.
- [15] E. Reiter, “An architecture for data-to-text systems,” in *Proceedings of the 11th European Workshop on Natural Language Generation*, S. Busemann, Ed., 2007, pp. 97–104.
- [16] J. Novikova, O. Dušek, and V. Rieser, “The E2E dataset: New challenges for end-to-end generation,” in *Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Saarbrücken, Germany, 2017, arXiv:1706.09254. [Online]. Available: <https://arxiv.org/abs/1706.09254>
- [17] E. Reiter, S. Sripada, J. Hunter, and I. Davy, “Choosing words in computer-generated weather forecasts,” *Artificial Intelligence*, vol. 167, pp. 137–169, 2005.
- [18] E. Reiter and S. Sripada, “Should corpora texts be gold standards for nlg?” in *Proceedings of the International Natural Language Generation Conference*, 2002, pp. 97–104.
- [19] R. Turner, S. Sripada, E. Reiter, and I. P. Davy, “Selecting the content of textual descriptions of geographically located events in spatio-temporal weather data,” *Applications and Innovations in Intelligent Systems*, vol. XV, pp. 75–88, 2007.
- [20] J. Kacprzyk, “Computing with words is an implementable paradigm: Fuzzy queries, linguistic data summaries, and natural-language generation,” *IEEE Trans. Fuzzy Systems*, pp. 451–472, 2010.
- [21] A. Ramos-Soto, A. Bugarín, and S. Barro, “Fuzzy sets across the natural language generation pipeline,” *Progress in Artificial Intelligence*, pp. 1–16, 2016.
- [22] R. Turner, S. Sripada, E. Reiter, and I. P. D. Davy, “Using spatial reference frames to generate grounded textual summaries of georeferenced data,” in *Proceedings of the 2008 International Conference on Natural Language Generation (INLG08)*, Salt Fork, Ohio, 12-14 June 2008.
- [23] S. C. Levinson, *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press, 2003, vol. 5.
- [24] R. de Oliveira, S. Sripada, and E. Reiter, “Absolute and relative properties in geographic referring expressions,” in *Proceedings of the 9th International Natural Language Generation conference*, 2016, pp. 256–264.
- [25] P. Fisher, “Sorites paradox and vague geographies,” *Fuzzy sets and systems*, vol. 113, no. 1, pp. 7–18, 2000.
- [26] V. B. Robinson, “A perspective on the fundamentals of fuzzy sets and their use in geographic information systems,” *Transactions in GIS*, vol. 7, no. 1, pp. 3–30, 2003.
- [27] P. Fisher, A. Comber, and R. Wadsworth, “Approaches to uncertainty in spatial data,” *Fundamentals of spatial data quality*, pp. 43–59, 2006.
- [28] Meteogalicia, “Meteogalicia’s web site.” <http://www.meteogalicia.es>, 2018.
- [29] A. Ramos-Soto, N. Tintarev, R. de Oliveira, E. Reiter, and K. van Deemter, “Natural language generation and fuzzy sets: An exploratory study on geographical referring expression generation,” in *IEEE World Congress on Computational Intelligence, 2016 IEEE International Conference on Fuzzy Systems*, 2016.
- [30] A. Ramos-Soto, J. M. Alonso, E. Reiter, K. van Deemter, and A. Gatt, “An empirical approach for modeling fuzzy geographical descriptors,” in *Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.



# Indistinguibilidad y Modelado de Tamaño en Sistemas *Data-to-text*

Nicolás Marín, Gustavo Rivas-Gervilla y Daniel Sánchez,  
 Departamento de Ciencias de la Computación e Inteligencia Artificial  
 Universidad de Granada, 18071 Granada, España  
 Email: {nicm, griger, daniel}@decsai.ugr.es

**Abstract**—En este trabajo extendemos nuestra propuesta previa de modelado de tamaño en base al contexto, incluyendo el tratamiento de la indistinguibilidad en la percepción del tamaño. Ilustramos la importancia de este aspecto con una experimentación que, adicionalmente, muestra una forma de determinar experimentalmente los parámetros del método.

**Index Terms**—Tamaño, modelado difuso, descripción lingüística de datos, sistemas *data-to-text*

## I. INTRODUCCIÓN

El lenguaje natural es una herramienta cada vez más imprescindible en el desarrollo de interfaces de comunicación hombre-máquina amigables. Prueba de ello es la importancia, cada vez más marcada, de la investigación en el ámbito de Generación de Lenguaje Natural, en general, y en el ámbito de los sistemas *data-to-text* en particular. Este tipo de sistemas tienen como objetivo elaborar información textual a partir del análisis de datos [1]. Particularmente, se pretende que dichos textos sean similares a los que escribiría un experto humano.

Uno de los problemas más importantes en este ámbito es determinar la semántica de los términos lingüísticos empleados en los textos, así como el grado de correspondencia entre dichos términos y los datos. Dado que dichos términos suelen representar conceptos difusos, es natural considerar conjuntos difusos como modelos de dicha semántica [2]–[5]. Estos conjuntos difieren de los conjuntos clásicos (o *crisp*) en que cada elemento del dominio pertenece al conjunto con un grado en el intervalo  $[0, 1]$ . La función que asocia a cada elemento este grado se denomina función de pertenencia. Uno de los usos más importantes de estos términos es la identificación de objetos en un conjunto, para lo cual se utilizan *expresiones de referencia*, que son sintagmas nominales formados habitualmente por una conjunción de términos que expresan propiedades de los objetos.

En [6] hemos propuesto un enfoque para la determinación de las funciones de pertenencia que modelan la semántica de conceptos difusos relativos al tamaño. Dicha propuesta considera funciones que son dependientes y condicionadas por el contexto de la aplicación donde se van a usar los términos lingüísticos.

Este trabajo ha sido financiado en parte por el Gobierno de España y el Fondo Europeo de Desarrollo Regional - FEDER a través del proyecto TIN2014-58227-P. Así como por la beca del Ministerio de Educación, Cultura y Deporte de España, FPU16/05199. Corresponding author: G. Rivas-Gervilla.

En este trabajo avanzamos un paso más en nuestra propuesta incluyendo en la misma el tratamiento de la indistinguibilidad. Este problema surge debido a que las computadoras son capaces de almacenar información sobre los objetos (sobre su tamaño en este caso) de manera más precisa de lo que nosotros somos capaces de percibir visualmente, lo cual afecta a nuestro método.

En la siguiente sección describimos la importancia del contexto en el modelado de términos relativos al tamaño y resumimos la propuesta presentada en [6]. En la sección III describimos el problema de la indistinguibilidad y cómo afecta a nuestro método, y detallamos los cambios en la misma que nos permitirán adaptarnos al grado de indistinguibilidad del tamaño de los objetos en un contexto concreto. En la sección IV se presenta un experimento con usuarios para ilustrar nuestra discusión. En esa misma sección se muestra una posible forma de determinar experimentalmente uno de los parámetros clave en nuestra propuesta. Finalmente, las dos últimas secciones contienen una breve discusión de los resultados y las conclusiones de nuestro trabajo, respectivamente.

## II. MODELADO DEL TAMAÑO

### A. Modelado y contexto

Como ya hemos indicado, uno de los aspectos clave en el desarrollo de sistemas *data-to-text* es la definición de la semántica de los términos lingüísticos que se van a emplear en los textos generados. En el caso particular de conceptos relativos a propiedades de los objetos, dicha semántica debe establecer una correspondencia entre características y valores de los datos asociados a dichos objetos y el cumplimiento de la propiedad. En el caso particular de propiedades difusas (como puede ser el tamaño, ya que un objeto puede ser grande en un cierto grado), este objetivo se cumple mediante la definición de las funciones de pertenencia asociadas a dichos conceptos difusos, como es bien conocido. Determinar las funciones de pertenencia es uno de los problemas clave en el uso de conjuntos difusos en éste y otros dominios de aplicación.

Sin embargo, la determinación de la función de pertenencia no plantea únicamente el problema de cómo obtener la asignación de grados a elementos de un referencial. En muchas ocasiones ocurre que la función de pertenencia no es única, sino que depende del contexto de aplicación (datos concretos, usuario, objetivos, etc.). Es decir, la definición de la función de pertenencia es relativa al contexto.

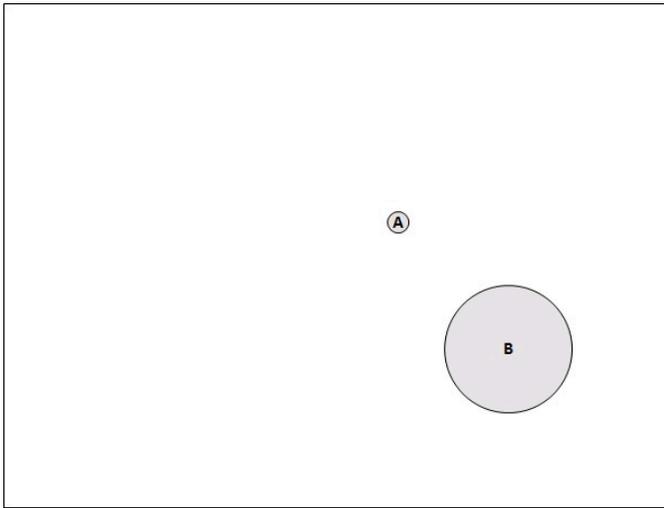


Fig. 1. Escena con objetos A y B

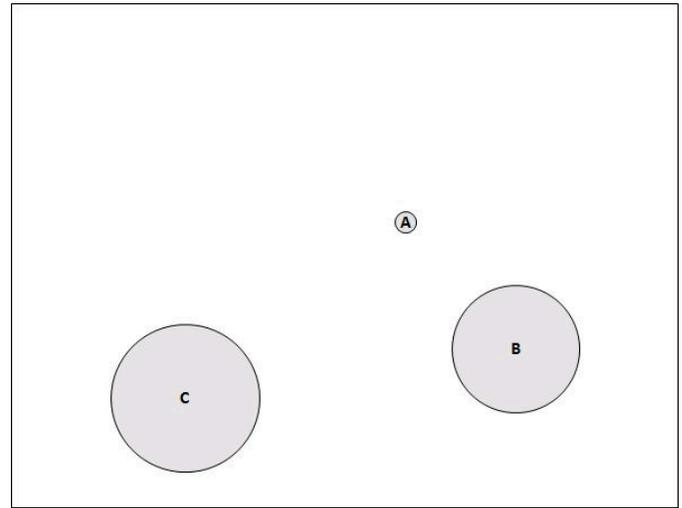


Fig. 2. Escena con objetos A, B y C

Para ilustrar esta situación en el contexto de sistemas *data-to-text*, vamos a considerar el ejemplo de las Figuras 1 y 2, extraído de [6]. En la escena de la Figura 1, para referirnos al objeto B es suficiente decir *el objeto grande*. Sin embargo, en la escena de la Figura 2, que contiene los mismos objetos de la Figura 1 más un objeto adicional, la referencia *el objeto grande* no llevaría a un usuario hipotético a señalar el objeto B, sino el C. Es decir, el contexto (en este caso, la aparición de nuevos datos, concretamente el objeto C) hace que el cumplimiento de la propiedad “tamaño” cambie, particularmente porque el rango de valores de tamaño que consideramos ha cambiado con la aparición de un nuevo tamaño (el tamaño de C). En este caso, el objeto B pasa a encajar mejor con la etiqueta *mediano*, siendo por tanto más adecuado en la Figura 2 referirse a B como *el objeto mediano* de cara a conseguir que un hipotético usuario lo identifique correctamente. Por lo tanto, queda claro que la definición de las funciones de pertenencia para las etiquetas *grande* y *mediano* es dependiente del contexto, por lo que fijar un conjunto de funciones de pertenencia independiente del contexto a priori no nos permite interactuar correctamente con el usuario en la mayoría de los casos, como demostramos en [6].

### B. Una propuesta de modelado basada en contexto

En [6] aportamos una primera aproximación a la solución de este problema. La propuesta consiste en determinar las funciones de pertenencia para las etiquetas *grande*, *mediano* y *pequeño* asociadas a la propiedad “tamaño” de forma dependiente del contexto asociado a los datos, teniendo en cuenta los distintos tamaños precisos de los objetos presentes en el conjunto de datos/escena. Asumiendo que el conjunto de tamaños es  $S = \{s_1, \dots, s_n\}$ ,  $n > 1$ , con  $s_i \in \mathbb{R}^+$  y  $0 < s_1 \leq \dots \leq s_n \forall 1 \leq i \leq n$ , la propuesta puede resumirse como sigue:

- En primer lugar, determinamos los tamaños que van a ser totalmente representativos de cada una de las eti-

quetas consideradas, es decir, van a tener grado 1 de pertenencia a las mismas. De manera natural establecemos por definición  $pequeño(s_1) = grande(s_n) = 1$  y  $pequeño(s_n) = grande(s_1) = mediano(s_1) = mediano(s_n) = 0$ . Nótese que esta definición es independiente de los valores concretos de  $s_1$  y  $s_n$ , así como de la diferencia  $|s_1 - s_n|$  entre ambos (volveremos sobre este último punto más adelante, puesto que es una de las motivaciones del presente trabajo). En el caso  $n > 2$  tendremos también representantes de la propiedad *mediano*. Si  $n$  es impar hay un único representante:  $s' = s'' = s_{(\lceil n/2 \rceil)}$ . En caso contrario tenemos dos representantes:  $s' = s_{(n/2)}$  y  $s'' = s_{(n/2+1)}$ . Para todo representante  $s \in \{s', s''\}$  de mediano tenemos  $mediano(s) = 1$  y  $pequeño(s) = grande(s) = 0$  por definición.

- El grado de pertenencia a las tres etiquetas de los tamaños en  $S$  que no son representativos de ninguna de ellas se determina mediante un proceso de *clustering*, centrado en los valores representativos de cada etiqueta, y en base a las distancias de dichos tamaños con respecto a los representativos. Como método de *clustering* empleamos el método definido en [7] en base a representaciones por niveles [8]. Este método toma como punto de partida una relación difusa reflexiva y simétrica de semejanza en el conjunto de tamaños  $S$ . En el caso particular de los tamaños, esta relación se obtiene en base a las distancias a través de un parámetro  $d_m$  que define la máxima distancia a la que pueden estar dos valores para que su semejanza sea mayor que 0. Sean  $s_i, s_j \in S$  y sea  $d(s_i, s_j) = |s_i - s_j|$  la distancia Euclidea entre ellos. La semejanza entre  $s_i$  y  $s_j$  en base al parámetro  $d_m$ ,  $R^{d_m}(s_i, s_j)$ , se calcula mediante la siguiente expresión:

$$R^{d_m}(s_i, s_j) = \max \left\{ 1 - \frac{d(s_i, s_j)}{d_m}, 0 \right\} \quad (1)$$



Como se demuestra en [6], para garantizar las condiciones que hemos expuesto anteriormente sobre los grados de pertenencia de  $s_1$ ,  $s'$ ,  $s''$  y  $s_n$  a las tres etiquetas de tamaño, es suficiente con elegir un  $d_m$  que satisfaga las siguientes condiciones:

- En el caso  $n = 2$  se requiere  $d_m \leq s_n - s_1$ .
- En el caso  $n > 2$  se requiere  $0 < d_m \leq \min(d(s_1, s'), d(s'', s_n))$ . Esta condición implica asimismo  $d_m \leq s_n - s_1$  en este caso.

A partir de una relación de semejanza difusa como ésta, el proceso de *clustering* basado en representación por niveles se basa en la realización de un *clustering crisp* en los diferentes niveles  $\Lambda(R^{d_m}) = \{R^{d_m}(s_i, s_j) \mid s_i, s_j \in S\} \setminus \{0\}$ . Este conjunto es finito,  $\Lambda(R^{d_m}) = \{\alpha_1, \dots, \alpha_k\}$  con  $k \geq 1$  y  $1 = \alpha_1 > \dots > \alpha_{k+1} = 0$ . Para cada  $\alpha_k \in \Lambda(R^{d_m})$  puede obtenerse un *clustering crisp* basado en recubrimientos de  $S$  de la siguiente forma:

- 1) Calcular el  $\alpha_k$ -corte de  $R^{d_m}$ , que es una relación *crisp* reflexiva y simétrica.
- 2) Calcular el *clustering* de  $S$  a nivel  $\alpha_k$  como el conjunto de cliques maximales del  $\alpha_k$ -corte de  $R^{d_m}$ . Sea  $C_{\alpha_k} \subset \{0, 1\}^S$  el conjunto de *clusters* obtenidos, cada uno de ellos asociado biunívocamente a un clique maximal. Nótese que estos *clusters* pueden solaparse, formando un recubrimiento de  $S$ .

Como consecuencia de las restricciones impuestas sobre  $d_m$  se da  $R^{d_m}(s_1, s_n) = 0$  para todo  $d_m$  que las cumpla, y por tanto  $|C_{\alpha_k}| \geq 2$ , dado que  $s_1$  y  $s_n$  estarán siempre en distintos *clusters*. Además es fácil demostrar que existe un solo *cluster*  $C_{\alpha_k}^p \in C_{\alpha_k}$ , verificando  $s_1 \in C_{\alpha_k}^p$ , y un solo *cluster*  $C_{\alpha_k}^g \in C_{\alpha_k}$ , cumpliéndose  $s_n \in C_{\alpha_k}^g$ , siendo como ya hemos comentado  $C_{\alpha_k}^p \neq C_{\alpha_k}^g$ . Además, en el caso  $n > 2$  definimos:

$$C_{\alpha_k}^m = \bigcup \{C_{\alpha_k}^i \mid C_{\alpha_k}^i \cap \{s', s''\} \neq \emptyset\} \quad (2)$$

Es decir,  $C_{\alpha_k}^m$  es la unión de todos los *clusters* que contienen al menos a uno de los representantes de la categoría *mediano*. Nótese que  $C_{\alpha_k}^m$  puede ser vacío, algo que ocurre en todos los niveles particularmente cuando  $n = 2$ .

A partir del *clustering* en cada nivel, los *clusters* difusos que modelan las funciones de pertenencia para las tres propiedades *pequeño*, *mediano* y *grande* se obtienen, siguiendo las ideas expuestas en [7], [8], como sigue:  $\forall s_i \in S$ ,

$$pequeño(s_i) = \sum_{\alpha_k \mid s_i \in C_{\alpha_k}^p} (\alpha_k - \alpha_{k+1}) \quad (3)$$

$$mediano(s_i) = \sum_{\alpha_k \mid s_i \in C_{\alpha_k}^m} (\alpha_k - \alpha_{k+1}) \quad (4)$$

$$grande(s_i) = \sum_{\alpha_k \mid s_i \in C_{\alpha_k}^g} (\alpha_k - \alpha_{k+1}) \quad (5)$$

$\alpha_k$	$C_{\alpha_k}^p$	$C_{\alpha_k}^m$	$C_{\alpha_k}^g$
1	{10}	{25}	{40}
2/3	{10}	{20, 25, 30}	{40}
1/3	{10, 20}	{20, 25, 30}	{30, 40}

TABLE I  
Clusters crisp POR NIVELES PARA  $d_m = 15$ .

Como ejemplo, sea  $S = \{10, 20, 25, 30, 40\}$  ( $n = 5$ ). En este contexto tenemos que  $s_1 = 10$  es el tamaño representativo para *pequeño*,  $s_5 = 40$  para *grande*, y  $s_3 = 25$  es el (único) representante para *mediano*. Sea  $d_m = 15$ , que es el máximo valor posible en este ejemplo. Dado que el conjunto de distancias entre tamaños es  $\{0, 5, 10, 15, 20, 30\}$  tenemos que  $\Lambda(R^{15}) = \{1, 2/3, 1/3\}$ . La Tabla I muestra los *clusters crisp* en cada nivel. En base a esta información y las Ecuaciones (3)-(5), se tiene finalmente:

$$\begin{aligned} pequeño &= 1/10 + (1/3)/20 \\ mediano &= (2/3)/20 + 1/25 + (2/3)/30 \\ grande &= (1/3)/30 + 1/40 \end{aligned}$$

### III. INDISTINGUIBILIDAD Y MODELADO

#### A. Influencia de la indistinguibilidad

Uno de los grandes problemas en la interacción con datos a través de las computadoras es lo que se denomina *hueco semántico*<sup>1</sup>. La principal forma en que se manifiesta este problema es que, mientras que los ordenadores almacenan datos descritos mediante valores de variables en dominios que suelen ser muy precisos y contener muchos valores distintos (por ejemplo, mediante números reales), los seres humanos solemos utilizar información descrita en base a conceptos expresados de forma lingüística. Hemos podido ver distintos ejemplos claros de esta diferencia en las secciones anteriores: mientras que para el ordenador el tamaño de un objeto se describe mediante un valor real numérico, los seres humanos utilizamos términos lingüísticos como *grande*, etc. Cada uno de estos términos se corresponde con un grupo o intervalo de valores precisos, grupo o intervalo que suele tener una frontera imprecisa, y por tanto una forma adecuada de modelar dicha correspondencia es mediante el uso de conjuntos difusos, como hemos visto y es bien conocido en nuestro ámbito.

Sin embargo, existe un problema adicional asociado al *hueco semántico* que recibe menos atención en la literatura, y que tiene que ver con la distinguibilidad. Este problema es típico en, aunque no exclusivo de, aplicaciones que involucran la percepción humana, como lo es el problema del tamaño que nos ocupa en este trabajo, y puede expresarse de manera informal como sigue: el ser humano no es capaz de percibir diferencias en el valor de propiedades de objetos al mismo nivel que una computadora.

Podemos ilustrar el problema que supone la indistinguibilidad para nuestro método de modelado sensible al contexto con

<sup>1</sup>Del inglés *semantic gap*.

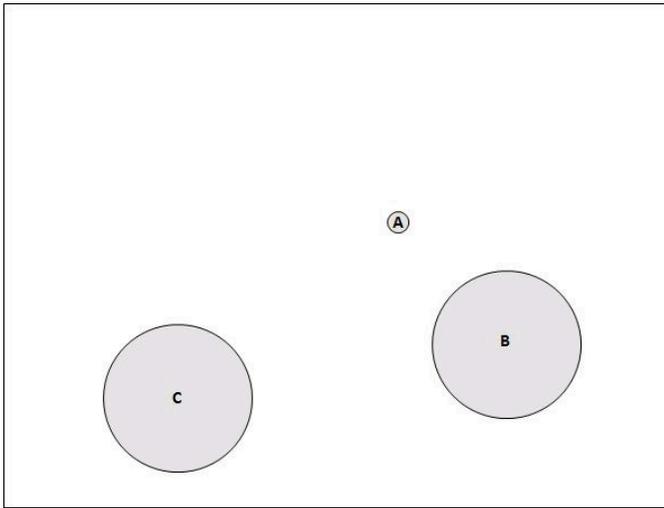


Fig. 3. Modificación de la escena de la Figura 2: tamaño del objeto B distinto pero indistinguible del tamaño del objeto C

un ejemplo. Consideremos la escena de la Figura 3, que se ha obtenido a partir de la escena de la Figura 2 aumentando el tamaño del objeto B hasta ser indistinguible (pero no igual) del tamaño del objeto C. En este caso, dado que para una computadora los tamaños de C y B son distintos, nuestro método diría que el objeto B es *mediano* con grado 1, mientras que los objetos A y C son medianos con grado 0. En base a este modelado sensible al contexto, un algoritmo de generación de expresiones de referencia propondría *el objeto mediano* como una forma de identificar claramente al objeto B. Sin embargo, al ser los tamaños de B y C indistinguibles visualmente para el usuario, éste no tendría claro a cuál de los dos objetos (B ó C) estamos haciendo referencia con dicha expresión (dado que A es claramente el pequeño en este caso).

Para solventar este problema es necesario modificar nuestra propuesta para tener en cuenta la distinguibilidad de objetos. Para ello vamos a considerar las siguientes hipótesis:

- Es posible determinar un umbral de distancia  $d_s$  tal que los tamaños  $s_i$  y  $s_j$  son distinguibles si, y solo si,  $d(s_i, s_j) > d_s$ .
- El umbral  $d_s$  puede ser sensible a distintos aspectos del contexto, tales como distancia entre los objetos, etc. Más adelante prestaremos atención a este problema.

En la siguiente sección detallamos la modificación que proponemos a nuestro método de modelado sensible al contexto para resolver el problema de la indistinguibilidad de objetos.

### B. Método propuesto

Nuestra propuesta consta de los siguientes pasos:

- 1) Determinación del umbral de distinguibilidad  $d_s$  adecuado para el modelado en el contexto concreto de aplicación. Es posible obtener  $d_s$  de distintas formas. En la Sección IV veremos un sencillo ejemplo de obtención de  $d_s$  de forma experimental bajo ciertos supuestos. El



Fig. 4. Ejemplo de pregunta del cuestionario.

estudio y propuesta de otras técnicas se plantea como trabajo futuro.

- 2) Determinación del umbral de similitud  $d_m$  para el contexto concreto de aplicación, teniendo en cuenta el valor de  $d_s$  en dicho contexto. Es importante determinar  $d_m$  después de  $d_s$  dado que el valor de este último afecta a las condiciones que se imponen sobre  $d_m$ . En primer lugar, por la semántica de ambos umbrales, es evidente que debe cumplirse  $d_s < d_m$ , dado que dos tamaños son indistinguibles cuando la distancia entre los mismos está por debajo de  $d_s$ , mientras que son absolutamente diferentes cuando la distancia está por encima de  $d_m$ . Pero además, el valor de  $d_s$  afecta a la forma en que se obtiene  $d_m$ , dado que, como hemos visto, este último viene restringido por condiciones que tienen que ver con la distancia entre objetos representativos de cada una de las etiquetas para los cuales, tras la discusión y el ejemplo que hemos visto en la sección anterior, debemos exigir que sean distinguibles. Más concretamente:

- Si  $s_1$  y  $s_n$  son indistinguibles entonces no podemos modelar tamaños de forma sensible al contexto.
- Supongamos  $s_1$  y  $s_n$  distinguibles y sea  $S_m \subset S$  el subconjunto de tamaños que son distinguibles tanto de  $s_1$  como de  $s_n$  (obviamente  $s_1, s_n \notin S_m$ ). Entonces:
  - Si  $S_m = \emptyset$  solo podemos modelar dos tamaños, *pequeño* y *grande*, siendo sus representantes  $s_1$  y  $s_n$ , respectivamente.
  - Si  $S_m \neq \emptyset$  entonces podemos modelar también *mediano*, siendo su(s) representante(s) el valor o valores centrales de  $S_m$  (dependiendo de que  $S_m$  tenga un número impar o par de valores).

Una vez determinados los representantes de las etiquetas se seleccionará un valor de  $d_m$  adecuado que cumpla con los requisitos ya expuestos anteriormente:  $d_m \leq s_n - s_1$  para dos etiquetas, y  $0 < d_m \leq$



$\min(d(s_1, s'), d(s'', s_n))$  para las tres.

- 3) Aplicar nuestra técnica de *clustering* para determinar las funciones de pertenencia para las tres etiquetas. Para ello utilizaremos una relación de semejanza que introduce una modificación con respecto a la de la Ecuación (1), dado que queremos que cuando la distancia entre dos tamaños sea menor o igual a  $d_s$ , su semejanza sea 1. La nueva relación de semejanza viene dada por

$$R^{d_m}(s_i, s_j) = \begin{cases} 1 & d(s_i, s_j) \leq d_s \\ 1 - \frac{d(s_i, s_j)}{d_m} & d_s < d(s_i, s_j) < d_m \\ 0 & d_m \leq d(s_i, s_j) \end{cases} \quad (6)$$

Es fácil demostrar que cuando  $d_s = 0$  (lo cual implica que todo par de tamaños distintos son distinguibles), nuestra propuesta se reduce a la propuesta expuesta en la Sección II-B.

#### IV. EJEMPLO DE APLICACIÓN

Para ilustrar la influencia de la indistinguibilidad y su dependencia del contexto se ha realizado una sencilla experimentación. Concretamente, se ha elaborado una encuesta con diferentes escenas en las que cada escena consta de dos objetos de distinto tamaño a una distancia determinada. Al usuario se le pregunta que identifique el más pequeño de los dos, permitiéndole elegir entre tres respuestas posibles: *el que está a la derecha, el que está a la izquierda y no sabría decirte*. La figura 4 muestra un ejemplo de escena de la encuesta (aparece marcada la respuesta correcta).

Para elaborar el conjunto de escenas de la encuesta, se ha trabajado con tres granos de tamaño  $C < B < A$  (elegidos en los extremos y centro del referencial), con cinco niveles de distancia (0, 0.3, 0.5, 0.7 y 1), también relativos, y configurando ejemplos en el que el ratio de parecido en tamaño va desde 0.95 hasta 0.995. La experimentación se ha realizado en la misma computadora con seis usuarios distintos a los que se les presenta el mismo conjunto de escenas con orden aleatorio.

El objetivo de la encuesta es señalar la posible dependencia del contexto de la indistinguibilidad, comprobando la variabilidad de respuestas entre usuarios, tamaños y distancias, e ilustrar una forma experimental para fijar el umbral  $d_s$  para una configuración de contexto dada.

##### A. Resultados

La tabla II muestra, para cada pareja distancia/grano de tamaño, los umbrales de ratio entre tamaños a partir de los cuales se ha obtenido una respuesta distinta de la correcta. Se resalta en negrita el valor mínimo obtenido en cada columna. Las distancias se muestran en valor relativo, representando el 1 la máxima distancia utilizada y el 0 la mínima distancia utilizada.

La figura 5 ilustra los datos de la tabla de manera gráfica. Como se puede observar, en la experimentación realizada, conforme descende el tamaño, en general, baja también el umbral, es decir, el usuario necesita una mayor distancia en

TABLE II  
UMBRALES AGREGANDO CON EL MÍNIMO

Distancia (rel.)	Tamaño		
	A	B	C
1	<b>98</b>	98,5	<b>95</b>
0,7	98,5	98,5	96
0,5	98	<b>98</b>	95
0,3	98	98,5	97
0	98,5	98	96

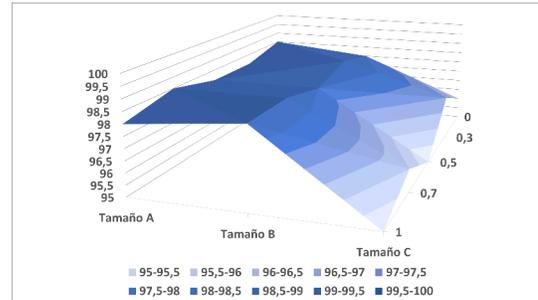


Fig. 5. Umbrales por tamaño y distancia, agregando con el mínimo.

tamaño relativo para acertar. No se aprecia dicha tendencia en el caso de la distancia. Los mismos datos, agregando los valores de umbral obtenidos en cada usuario con una media en lugar de con el mínimo, se muestran en la tabla III.

La figura 6 ilustra los datos de la tabla de manera gráfica. Se observan similares tendencias, pero menos acusadas, dado que la media limita el efecto de los usuarios con menor precisión en la percepción del tamaño.

Por último, se muestran los datos de dos usuarios a título de ejemplo. La tabla IV y la figura 7, muestran los datos de un usuario con alta precisión en la percepción del tamaño, mientras que la tabla V y la figura 8 muestran los datos de un usuario con menor precisión.

#### V. DISCUSIÓN

A la vista de la anterior experimentación, se pone de manifiesto que la *indistinguibilidad* es un elemento variable altamente dependiente del contexto. Por tanto, es importante que el sistema de generación de expresiones de referencia sea capaz de ajustar su funcionamiento a la capacidad de distinguibilidad propia del contexto en el que se ejecuta.

La experimentación que hemos desarrollado aquí, aunque simple, sirve para ilustrar que el ajuste de la indistinguibilidad es complejo. Nótese que la experimentación realizada utiliza solo tres variables para caracterizar el contexto, a saber, el *usuario*, el *tamaño* del objeto y la *distancia* entre ellos. El resto de variables se fijan: el *tipo* de objeto, el *color*, la *posición* en la pantalla, el *dispositivo*, ... por enumerar solo algunas variables de una lista que puede ser muy extensa dependiendo del sistema y del tipo de escenas con las que trabaje. Por citar solo un ejemplo, piénsese que se varía el color de los objetos: es bien conocido por todos que el color influye en la percepción del tamaño de los objetos.

TABLE III  
 UMBRALES AGREGANDO CON LA MEDIA

Distancia (rel.)	Tamaño		
	A	B	C
1	<b>98,7</b>	99,1	98,0
0,7	99,2	99,3	<b>97,8</b>
0,5	99,2	<b>99,0</b>	97,8
0,3	98,7	99,3	98,5
0	99,2	99,0	98,1

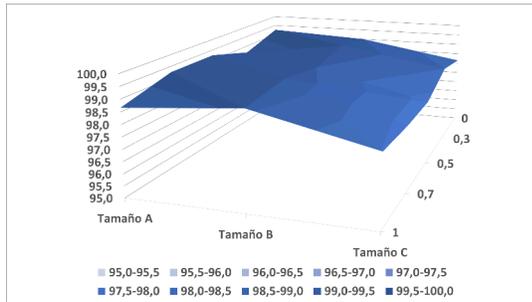


Fig. 6. Umbrales por tamaño y distancia, agregando con la media.

Por tanto, un sistema de generación de expresiones de referencia será mejor cuanto mayor sea el ajuste al contexto que tiene en cuenta para producir sus expresiones.

## VI. CONCLUSIONES

Una de las contribuciones más reconocidas de la Teoría de Conjuntos Difusos en el ámbito de los sistemas *data-to-text* es el modelado de propiedades difusas. Este trabajo contribuye a facilitar dicha tarea proponiendo una metodología para determinar de forma automática la semántica de propiedades difusas en base al contexto y teniendo en cuenta la distinguibilidad de las propiedades de los objetos, que es uno de los problemas más arduos que surgen en este tipo de sistemas.

## REFERENCES

- [1] E. Reiter and R. Dale, *Building Natural Language Generation Systems*. Cambridge, UK: Cambridge University Press, 2000.
- [2] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets and Systems*, vol. 285, pp. 6–30, 2016.
- [3] A. Ramos-Soto, A. Bugarín, and S. Barro, "On the role of linguistic descriptions of data in the building of natural language generation systems," *Fuzzy Sets and Systems*, vol. 285, pp. 31–51, 2016.
- [4] K. van Deemter, *Not Exactly: In Praise of Vagueness*. New York, NY, USA: Oxford University Press, Inc., 2012.
- [5] A. Gatt, N. Marín, F. Portet, and D. Sánchez, "The role of graduality for referring expression generation in visual scenes," in *Proceedings IPMU 2016, Part I, CCIS 610*, J. Carvalho, M.-J. Lesot, U. Kaymak, S. Vieira, B. Bouchon-Meunier, and R. R. Yager, Eds. Springer, 2016, pp. 191–203.
- [6] N. Marín, G. Rivas-Gervilla, and D. Sánchez, "An approximation to context-aware size modeling for referring expression generation," in *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, 2018.
- [7] D. Dubois and D. Sanchez, "Fuzzy Clustering based on Coverings," in *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, ser. Studies in Fuzziness and Soft Computing, C. Borgelt, M. A. Gil, J. M. C. Sousa, and M. Verleysen, Eds. Springer, 2013, vol. 285, pp. 319–330.
- [8] D. Sánchez, M. Delgado, M. Vila, and J. Chamorro-Martínez, "On a non-nested level-based representation of fuzziness," *Fuzzy Sets and Systems*, vol. 192, pp. 159–175, 2012.

 TABLE IV  
 UMBRALES PARA USUARIO CON ALTA PRECISIÓN EN LA PERCEPCIÓN DEL TAMAÑO

Distancia (rel.)	Tamaño		
	A	B	C
1	<b>98,7</b>	99,1	98,0
0,7	99,2	99,3	<b>97,8</b>
0,5	99,2	<b>99,0</b>	97,8
0,3	98,7	99,3	98,5
0	99,2	99,0	98,1

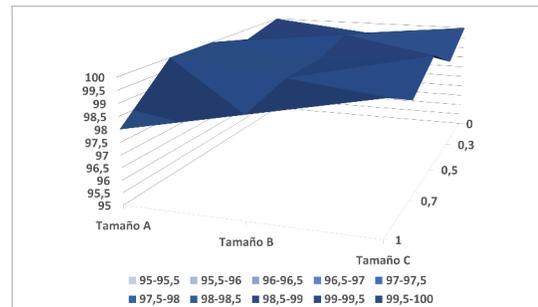


Fig. 7. Umbrales para usuario con alta precisión en la percepción del tamaño.

 TABLE V  
 UMBRALES PARA USUARIO CON MENOR PRECISIÓN EN LA PERCEPCIÓN DEL TAMAÑO

Distancia (rel.)	Tamaño		
	A	B	C
1	<b>98,7</b>	99,1	98,0
0,7	99,2	99,3	<b>97,8</b>
0,5	99,2	<b>99,0</b>	97,8
0,3	98,7	99,3	98,5
0	99,2	99,0	98,1

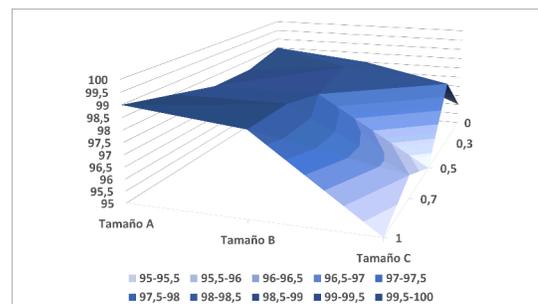


Fig. 8. Umbrales para usuario con menor precisión en la percepción del tamaño.



# Método para medir la credibilidad de los agentes en videojuegos empleando Descripción Lingüística de Fenómenos Complejos y su aplicación como herramienta de apoyo en el proceso de enseñanza-aprendizaje en un curso de Inteligencia Artificial \*†‡

Clemente Rubio-Manzano, Tomás Lermenda, Alejandra Segura, Claudia Martínez-Araneda y Christian Vidal  
SOMOS Research Group - Universidad del Bío-Bío & MCIS Research Group - Universidad de Cádiz

## Resumen

El objetivo de este trabajo es proponer un nuevo método para medir la credibilidad de los agentes artificiales que actúan en un mundo virtual empleando la Descripción Lingüística de Fenómenos Complejos y las Funciones de Equivalencia Restringidas. La combinación de ambas técnicas computacionales permite generar y comparar informes de comportamiento en lenguaje natural generados automáticamente. Con el fin de mostrar y explorar las posibilidades de este nuevo método, se desarrolló una aplicación web que puede emplearse como herramienta de apoyo a la docencia de Inteligencia Artificial y donde cada alumno puede evaluar por sí mismo la calidad de su implementación en los respectivos proyectos.

**Palabras Clave:** Inteligencia Artificial, Videojuegos, Descripción Lingüística de Datos, Funciones de Equivalencia Restringidas, Lógica Difusa.

## I. Introducción

La Inteligencia Artificial (IA) tiene como objetivo lograr que una máquina tenga una inteligencia similar a la humana [7]. Sin embargo, históricamente, han coexistido dos visiones (IA fuerte e IA débil) que se corresponden con las siguientes definiciones dadas por Marvin Minsky y John McCarthy, respectivamente: “IA es la ciencia de

*crear máquinas capaces de realizar tareas que requerirían inteligencia si las tuviera que hacer un humano*”; “IA es la ciencia y la ingeniería de crear máquinas inteligentes”. Como se indica en [7]: “Todos los avances logrados hasta ahora en el campo de la IA son manifestaciones de la IA débil”, es decir, dotar a las máquinas de la capacidad de realizar tareas específicas mejor que las personas se puede realizar con éxito, incluso, una IA específica puede superar a la pericia humana.

Por otro lado, un aspecto fundamental en el campo de la IA es determinar si las máquinas son realmente inteligentes, entonces dicha evaluación dependerá de la visión de IA que se tenga. En el caso de la IA fuerte, Alan Turing propuso su famoso test para determinar si un computador exhibía inteligencia humana. En el caso de la IA débil se optó por chequear si las máquinas realizan correctamente las tareas para las cuales fueron diseñadas y programadas.

La IA débil se ha aplicado con éxito en numerosas disciplinas. En el ámbito de los videojuegos, la IA se centra en crear entidades software (creadas mediante algoritmos/programas de computador que actúan sobre un entorno virtual) que produzcan en el espectador la ilusión de inteligencia. Por otro lado, se ha comprobado que los agentes (también llamados NPCs, por sus siglas en inglés Non-player Characters, o Bots) más divertidos y “jugables” son aquellos que se comportan como los jugadores humanos [13]. Surge así el desafío de tratar de desarrollar NPCs que posean un buen grado de imitación y, al mismo tiempo, diseñar e implementar instrumentos que permitan medirlo y evaluarlo [6]. No solo se trata en evaluar que un NPC realice las tareas específicas como un jugador humano, además, queremos que la IA se parezca a la IA del jugador humano lo más posible, debe ser creíble, cuanto más, mejor.

Esta característica se conoce como “credibilidad” (“believability”) y habitualmente se evalúa empleando una extensión el mencionado Test de Turing [5]: “Supongamos que estamos jugando e interactuamos con algún agente, ¿se podría distinguir observándolo al agente si es un

\* Clemente Rubio-Manzano. Departamento de Matemáticas, Universidad de Cádiz, España

† Clemente Rubio-Manzano, Tomás Lermenda, Christian Vidal y Alejandra Segura, Departamento de Sistemas de Información, Universidad del Bío-Bío, Chile

‡ Claudia Martínez-Araneda. Departamento de Ingeniería Informática, Universidad Católica Santísima Concepción, Chile

*jugador humano o un bot? Si no se puede distinguir, entonces se considera que el bot ha pasado el test* ". Esta prueba se diseñó para evaluar las habilidades de los bots para suplantar a un jugador humano.

El test anterior se puede considerar una evaluación basada en el chequeo visual de una determinada funcionalidad. Este hecho tiene como principal limitación a la subjetividad del proceso porque la observación visual es una tarea compleja al requerir procesar una gran cantidad de información generada durante las partidas. Incluso, cabe la posibilidad, de pérdida información debido a la rapidez de ejecución de los algoritmos de IA.

Nuestro trabajo consiste en enriquecer estos test alternativos para reducir la subjetividad mediante la captura de todos los datos posibles generados durante las partidas y entregar automáticamente un conjunto de descripciones lingüísticas en lenguaje natural que proporcionen información sobre el comportamiento de los agentes. Para ello, emplearemos la Descripción Lingüística de Fenómenos Complejos (DLFC)[14] y las Funciones de Equivalencia Restringidas (REFs, de sus siglas en inglés) [1]. En este sentido, el "comportamiento de un agente" está relacionado directamente con el comportamiento del algoritmo IA elegido para su implementación, en este caso, algoritmos de búsqueda en un espacio de estados.

La estructura del trabajo es como sigue, en la Sección II se realiza una introducción desde el punto de vista del proceso de enseñanza-aprendizaje al ser éste el fin último de su aplicación. La Sección III detalla un método para medir la credibilidad de los agentes implementados y comparar el comportamiento de los jugadores humanos con los agentes. Posteriormente, en la Sección IV se explica la experimentación y la evaluación implementadas sobre los proyectos de los estudiantes. Finalmente, la Sección de conclusiones y trabajo futuro proporciona algunos comentarios finales sobre el presente trabajo y el futuro del mismo.

## II. Motivación

Desde el año 2011 se ha implementado una nueva metodología de enseñanza-aprendizaje basada en videojuegos en el curso de IA de la Universidad del Bío-Bío (Chile). Su objetivo es proporcionar a los estudiantes una herramienta de ayuda en el proceso de comprensión de los diferentes algoritmos heurísticos que se enseñan en las clases de teoría.

En particular, en el año 2017, el proyecto consistió en desarrollar un videojuego cuyo agente protagonista tenía que capturar un conjunto de recompensas (que proporcionan energía) distribuidas por el escenario. El agente, que pierde energía cada cinco segundos, debía mantenerse en el escenario durante el mayor tiempo posible sin agotar

energía. Tres agentes adversarios (también programados por los estudiantes) tratarán de robarle la energía, en el caso de que ésta se agote la partida termina.

En este contexto, un agente (y su algoritmo de comportamiento asociado) está correctamente implementado cuando es capaz de imitar a un jugador experto humano (ver sección III) . El proceso de evaluación consiste en chequear la calidad de la imitación desde dos puntos de vista: visualmente, observando el comportamiento del agente artificial en una partida; formalmente, analizando el código del proyecto del alumno. Este proceso tiene algunas limitaciones:

- 1) Tiempo excesivo que se requiere para chequear de forma visual la funcionalidad de los proyectos. Se podría convertir en un problema cuando el número de estudiantes es alto y hay un único profesor.
- 2) Tarea compleja debido a la dificultad de evaluar un gran número parámetros importantes en el proyecto: cantidad (memoria utilizada, iteraciones realizadas, estructuras de datos usadas) y calidad (como es la IA, es decir, que habilidad tiene a la hora de realizar ciertas acciones)
- 3) Se pueden perder numerosos detalles en la inspección visual al transcurrir las partidas muy rápidamente: el depurador genera mucha información que es difícil de entender; gran cantidad de datos que se generan en la ejecución.
- 4) Imposibilidad (dificultad) de realizar una evaluación individual

Para superar estas limitaciones se propone incorporar la DLFC al proceso de enseñanza-aprendizaje de la asignatura IA y, en particular, a su evaluación basada en proyectos. Utilizando DLFC los estudiantes pueden obtener retroalimentación inmediata, técnica y personalizada sobre los fallos cometidos durante el desarrollo de sus proyectos y conseguir más información sobre todos los detalles de ejecución de los algoritmos heurísticos empleados para programar los agentes inteligentes. En la literatura, existen trabajos que han estudiado la aplicación de la DLFC (y en la algunos casos de la Descripción Lingüística de Datos-LDD) al proceso de enseñanza-aprendizaje [9], [3], [2] y [12].

Por otro lado, esta propuesta es muy beneficiosa para los profesores ya que les permitirá:

- Ahorrar tiempo al permitirles centrarse en otros aspectos a evaluar lo que implicará un mejor entendimiento de la implementación de los algoritmos.
- Mejorar el proceso clásico de evaluación proporcionando retroalimentación personalizada a los estudiantes.
- Apoyar el desarrollo individual de proyectos al permitir un seguimiento más cercano y la oportunidad



de centrarse en los aspectos más debilidades de los estudiantes.

Con el fin de mostrar y explorar las posibilidades de esta herramienta se ha desarrollado una plataforma web y se ha incorporado al proceso de enseñanza aprendizaje de la asignatura. Ahora cada estudiante puede obtener, en tiempo real, retroalimentación sobre su proyecto, además establecer su propio plan de aprendizaje. Adicionalmente, los perfiles de comportamiento para tales agentes y jugadores humanos se pueden utilizar para medir la calidad de las IA mediante una adaptación del test Turing al contexto de los videojuegos.

### III. Un método para medir la credibilidad de los agentes y compararlo con jugadores humanos

En esta sección se presenta un método para medir la credibilidad de los agentes basado en cuatro pasos.

#### Paso 1. Generación de la traza de ejecución.

Se ha empleado una traza de ejecución para capturar y almacenar los datos generados a partir de los movimientos de los humanos, agentes y oponentes. Una traza de ejecución es una técnica empleada habitualmente para capturar y almacenar los valores de las estructuras de datos empleadas en una aplicación. Se emplea, también, como herramienta de depuración y análisis del rendimiento.

Concretamente, la traza de ejecución implica la detección y almacenamiento de eventos relevantes durante la ejecución de un programa para posteriormente realizar un análisis más detallado. En particular, empleamos las trazas de ejecución para almacenar los valores de las métricas a partir de los cuales se pueden analizar el comportamiento de las entidades. Las métricas son las siguientes:

- Distancia entre jugador ( $P$ ) y adversarios ( $O$ ) calculada a partir de sus posiciones.
- Distancia entre jugador, adversarios y la recompensa más cercana ( $R^*$ ) calculada a partir de sus posiciones.
- Tiempo en llegar a la recompensa.
- Energía del jugador en cada momento.
- Memoria ocupada e iteraciones realizadas.

#### Paso 2. Definición de la red de percepciones computacionales

Ampliamos la red de percepciones computacionales presentada en [10]. En este caso, la energía, memoria ocupada e iteraciones realizadas se incorporan como nuevas variables. Se extiende la red, reglas y las plantillas de generación de lenguaje natural para soportar los nuevos requisitos.

Las variables lingüísticas se definen a partir de las métricas establecidas previamente. Mediante la agregación de los términos lingüísticos que componen cada una de las variables se definen las reglas de comportamiento. Por ejemplo, la percepción computacional (CP, de sus siglas en inglés) de situación ( $CP_{Situation}$ ) quedaría definida de la siguiente forma:

$$CP_{Situation} = CP_{Protection}(P, O), CP_{Distance}(P, O), CP_{Energy}(P).$$

Esta ecuación establece que la CP de situación se calcula en función de la protección, la distancia entre el jugador-oponente y la energía del jugador. Los valores correspondientes para esta CP se calculan usando las siguientes reglas:

Consecuente (Situación)	Antecedente (Protección, Distancia, Energía)
Risky	Intermediate, Close, Normal
Dangerous	Low, Close, Normal
Safe	Intermediate, Normal, Normal
Easy	Low, Normal, Normal
Dangerous	Low, Normal, Low
Dangerous	Normal, Close, Low
Dangerous	Normal, Normal, Low

El resto de CPs se definen de forma análoga:

- $CP_{Attitude} = CP_{Distance}(P, R^*), CP_{Distance}(O, R^*)$
- $CP_{Movement} = CP_{Distance}(P, R^*), CP_{Distance}(P, O), CP_{Energy}(P)$
- $CP_{Ability} = CP_{Attitude}, CP_{Movement}, CP_{Time}$
- $CP_{Skill} = CP_{Attitude}, CP_{Movement}, CP_{Situation}$
- $CP_{Resources} = CP_{Iterations}, CP_{Memory}$

#### Paso 3. Generación automática de los perfiles de comportamiento.

En cada periodo de tiempo (habitualmente cada segundo) se calculan los valores para cada una de las CPs instantáneas. Con el fin de agruparlas y resumirlas se emplea el concepto de CP Resumen, denotada por  $\Sigma CP$  y definida como  $\Sigma CP = ((a_1, w_1), \dots, (a_n, w_n))$ . El objetivo de cada  $\Sigma CP$  es sintetizar en un vector lingüístico difuso el resultado global obtenido tras la partida para una CP. En nuestro caso, este tipo de CP nos permite obtener el número de veces que un valor  $(a_1, \dots, a_n)$  ocurrió durante la partida.

Estos tipos de CP nos proporcionan un conjunto de variables, su valor asociado y el grado  $\alpha$ , que indica la media difusa para un valor determinado. Por ejemplo, un valor para la CP de Situación podría ser *Safe* con un grado de 0.8 en un instante  $i$ ,  $X = Safe$  con 0.7 en un instante  $i + 1$ , y etc. Por tanto, al final de la ejecución, tendremos que  $a_i$  (en el ejemplo “safe”) se ha dado  $N$  veces con  $M$  grados diferentes  $\beta_1, \dots, \beta_m$  (por supuesto, algunos de

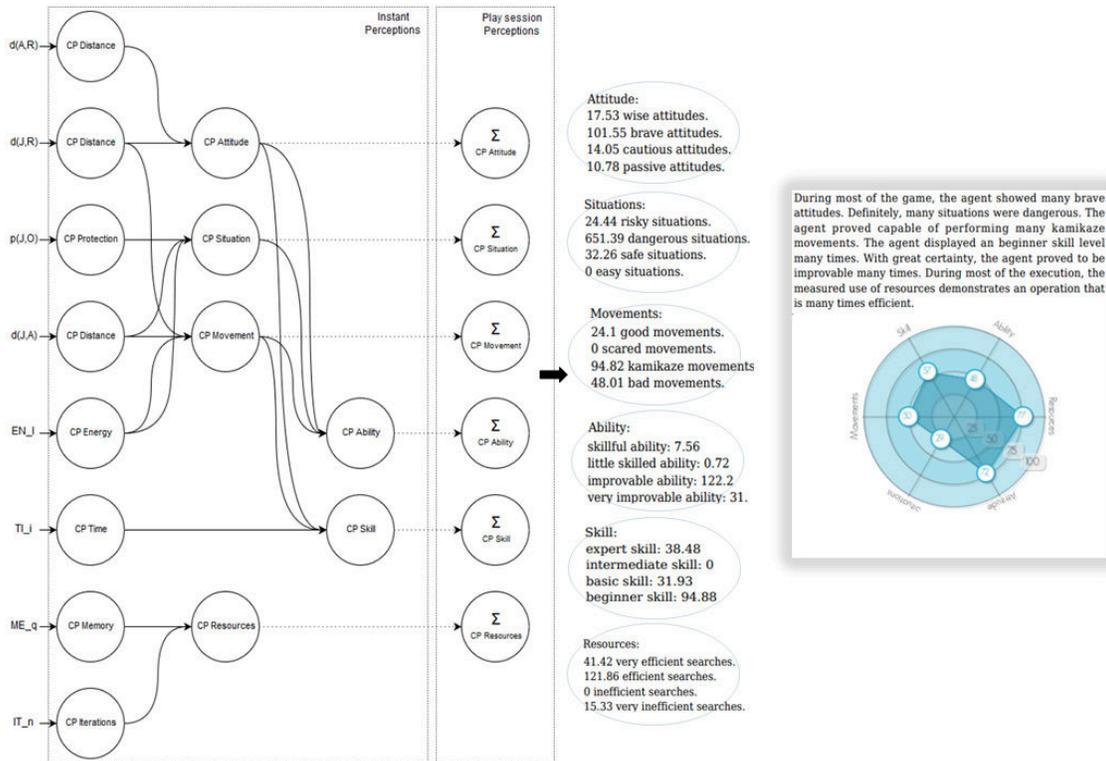


Fig. 1. Generación automática de informe de comportamiento a partir de los valores capturados en la traza de ejecución

estos grados podrían ser iguales). El grado final se calcula como sigue:  $\alpha_i = ((\beta_1 + \dots + \beta_n)/N)$ .

Por ejemplo, los siguientes resúmenes pueden obtenerse a partir de diferentes  $\Sigma CP$  (ver Figura 1). La generación del perfil de comportamiento se realiza empleando el conjunto de  $\Sigma CP$ . Para cada CP se crea una descripción lingüística en función del par  $(a_i, w_i) \in \Sigma CP$ . Los porcentajes se calculan para cada  $\Sigma CP$ . El porcentaje  $p_i$  se transforma en un término lingüístico de cantidad como sigue: "few" cuando  $p_i \in [0, 1/3]$ ; "several" cuando  $p_i \in [1/3, 2/3]$  o "many" cuando  $p_i \in [2/3, 1]$ . Entonces, vamos a considerar cuatro casos:

- 1) Existe un par  $(a_i, p_i) \in \Sigma CP$  cuyo  $p_i$  es mayor o igual a 66 por ciento.
- 2) Existe un par  $(a_i, p_i) \in \Sigma CP$  cuyo  $p_i$  es mayor o igual a 33 por ciento.
- 3) Existe dos pares  $(a_1, p_1), (a_2, p_2) \in \Sigma CP$  cuyos  $p_i$  son mayor o igual a 33 por ciento.
- 4) No existe ningún par  $(a_i, p_i) \in \Sigma CP$  cuyos  $p_i$  son mayor a 33 por ciento.

Y generar las siguientes descripciones en función de ellos. Mostramos aquí la tabla de CP de situación. El informe completo se puede ver en la Figura 2.

Casos	Descripción Lingüística
1	Definitely, <i>degree</i> situations were <i>value</i>
2	<i>degree</i> situations were <i>value</i>
3	<i>degree</i> situations were <i>value</i> <sub>1</sub> , although <i>degree</i> situation also were <i>value</i> <sub>2</sub>
4	Diverse situations were detected during the most part of the play session

#### Paso 4. Método para comparar perfiles de comportamiento entre jugadores humanos y agentes.

Como se ha mencionado en la introducción, uno de los objetivos más importantes en IA en el ámbito de los videojuegos es crear agentes que imiten la habilidad de un jugador humano. Por lo tanto, debemos definir la referencia de mejor jugador humano analizando las partidas y obteniendo un perfil de comportamiento referencia. Tras este proceso se obtuvo que:

- La actitud del jugador es valiente durante la mayor parte de la partida;
- Las situaciones por las que pasa son principalmente seguras;
- Los movimientos realizados son principalmente buenos;
- Se obtiene que es experto y hábil con un uso eficiente de tiempo y espacio.

Ahora, se puede proceder a la comparación de los mismos



mediante la definición de una medida de similitud basada en las funciones de equivalencia restringidas (REFs) entre las CPs Resumen obtenidas para cada caso.

**Definición 3.1:** Una REF,  $f$ , es una correspondencia  $[0, 1]^2 \rightarrow [0, 1]$  que satisface las siguientes condiciones:

- 1)  $f(x, y) = f(y, x)$  para todo  $x, y \in [0, 1]$
- 2)  $f(x, y) = 1$  si y sólo si  $x = y$
- 3)  $f(x, y) = 0$  si y sólo si  $x = 1$  y  $y = 0$  o  $x = 0$  y  $y = 1$
- 4)  $f(x, y) = f(c(x), c(y))$  para todo  $x, y \in [0, 1]$ ,  $c$  siendo una negación fuerte.
- 5) Para todo  $x, y, z \in [0, 1]$ , si  $x \leq y \leq z$ , entonces  $f(x, y) \geq f(x, z)$  and  $f(y, z) \geq f(x, z)$

Por ejemplo,  $g(x, y) = 1 - |x - y|$  satisface las condiciones (1)-(5) con  $c(x) = 1 - x$  para todo  $x \in [0, 1]$ . Una medida de similitud basada en REFs entre dos términos lingüísticos para ampliar el motor de inferencia del lenguaje Bousi Prolog se propuso en [11].

Un algoritmo de IA, usado para implementar un agente, tiene un comportamiento cercano al jugador humano cuando sus respectivos perfiles de comportamiento son similares. La puntuación final (de 1 a 7) se calcula mediante una medida de similitud para cada CP. La ecuación para calcular el grado de credibilidad es la siguiente:

$$FG = G_{Min} + S(Actitud) + S(Situacin) + S(Movimiento) + S(Habilidad) + S(Competencia) + S(Eficiencia)$$

$G_{Min} = 1$  punto (mínimo puntuación que un alumno en la Universidad del Bío-Bío puede obtener). Donde  $S(Attitude) = S_{REF}(\Sigma CP_{Attitude}^{Human}, \Sigma CP_{Attitude}^{Bot})$  es la similitud entre la actitud del jugador humano y la actitud del agente. El resto de relaciones de similitud se definen de forma análoga. La siguiente definición formaliza esta media:

**Definición 3.2:** Dadas dos percepciones computacionales  $\Sigma CP_i$ ,  $\Sigma CP_j$  cuyos vectores de porcentajes lingüísticos son  $\{(a_1, p_1) \dots, (a_n, p_n)\}$  y  $\{(b_1, q_1) \dots, (b_n, q_n)\}$  respectivamente. Una medida de similitud entre  $\Sigma CP_i$  y  $\Sigma CP_j$  se define como

$$S_{REF}(\Sigma CP_i, \Sigma CP_j) = \sum_{i=0}^n (REF(p_i, q_i)) / n$$

siendo  $REF(p_i, q_i) = 1 - |p_i - q_i|$

Por ejemplo, dadas dos CP Resumen  $CP_{Actitud}^{Humano}$ ,  $CP_{Actitud}^{Agente}$ , entonces  $\Sigma CP_{Actitud}^{Humano} = \{(wise, 122.35), (brave, 289), (cautious, 87.59), (passive, 8.75)\}$  y  $\Sigma CP_{Actitud}^{Agente} = \{(wise, 17.53), (brave, 101.55), (cautious, 14.05), (passive, 10.78)\}$ . Entonces el vector de porcentaje lingüístico se calcula para cada  $\Sigma CP$  empleando los totales:

$Total_{\Sigma CP_{Actitud}^{Humano}} (507.69)$   $Total_{\Sigma CP_{Actitud}^{Agente}} (143.61)$ , respectivamente:

- $\Sigma CP_{Actitud}^{Human} = \{(wise, 0.240), (brave, 0.569), (cau-$

$tious, 0.172), (passive, 0.017)\}$

- $\Sigma CP_{Actitud}^{Bot} = \{(wise, 0.122), (brave, 0.709), (cautious, 0.097), (passive, 0.075)\}$

Ahora, la similitud  $S_{REF}(\Sigma CP_{Actitud}^{Human}, \Sigma CP_{Actitud}^{Bot})$  se calcula con:  $REF(0.240, 0.122) = 1 - |0.240 - 0.122| = 0.882$ ,  $REF(0.569, 0.172) = 1 - |0.569 - 0.172| = 0.882$ ,  $REF(0.172, 0.097) = 1 - |0.172 - 0.097| = 0.925$  y  $REF(0.017, 0.075) = 1 - |0.017 - 0.075| = 0.942$  Por tanto,  $S_{REF}(\Sigma CP_{Actitud}^{Human}, \Sigma CP_{Actitud}^{Bot}) = \frac{3.402}{4} = 0.838$ . El resto de similitudes se calcula de forma similar. La puntuación final, junto al perfil de comportamiento obtenido para el jugador humano y el agente, proporciona al alumno una referencia sobre la calidad de los algoritmos implementados para los agentes actuando en el videojuego.

#### IV. Hacia una herramienta de apoyo a la docencia y el aprendizaje de algoritmos

Para llevar a cabo las ideas planteadas y definidas en las secciones anteriores se optó por el desarrollo una plataforma web que permite a cada estudiante obtener un informe de evaluación de forma automática que le proporcionará retroalimentación de forma inmediata sobre la calidad de implementación de sus agentes. La herramienta permite realizar una comparación entre el comportamiento de un jugador humano y los agentes desarrollados en los proyectos. De esta forma el alumno obtiene una nota de referencia de credibilidad de los mismos.

Esto puede verse como un test de turing permitiéndole testear a partir de las trazas de ejecución <sup>1</sup> que genera su proyecto un informe sobre el perfil de comportamiento de sus agentes inteligentes. A la herramienta se accede a través de la siguiente URL:

<http://www.youractionsdefineyou.com/assess>

La ventana principal muestra dos opciones: log in y register. El registro de un usuario consiste en introducir el correo electrónico, nombre de usuario, RUT (el equivalente al DNI español) y una clave. Para confirmar que el registro en la aplicación se ha realizado correctamente se enviará un correo electrónico al usuario. Una vez realizado el registro, el usuario puede ingresar introduciendo su usuario y clave. Una vez dentro de la aplicación, el informe de perfil de comportamiento se puede obtener seleccionando y cargando un archivo de traza de ejecución, cuando este archivo se carga el informe se obtiene automáticamente.

#### V. Agradecimientos

Al proyecto interno de la Universidad del Bío-Bío 170915

<sup>1</sup>[http://youractionsdefineyou.com/assess/web/examples\\_traces](http://youractionsdefineyou.com/assess/web/examples_traces)

## HUMAN BEHAVIOR PROFILE

The human player showed several brave attitudes. Definitely, many situations were safe. The human player proved to be capable of performing several good movements. The human player displayed an expert skill level several times. The human player proved to be skillful several times. During the most of the execution, the measured use of resources demonstrates an operation that is many times very efficient.

Grade: 7



## BOT BEHAVIOR PROFILE

During the most of the game, the bot showed many brave attitudes. Definitely, many situations were dangerous. The bot proved to be capable of performing several kamikaze movements. The bot displayed an beginner skill level several times. With great certainty, the bot proved to be improvable many times. During the most of the execution, the measured use of resources demonstrates an operation that is many times efficient.

Grade: 5.15



Fig. 2. Similitud entre los informes de comportamiento generados para el jugador humano y los agentes

2/R. Al grupo de investigación SOMOS (Software-MOdeling-Science) con referencia GI170315/EF

## VI. Conclusiones y trabajo futuro

En este trabajo se ha tratado el tema de la evaluación de los sistemas de Inteligencia Artificial (fuerte y débil) en el ámbito de los videojuegos. Se ha planteado la idea de evaluar la “inteligencia humana” de los agentes artificiales y se ha presentado un método para medir la “credibilidad” de los mismos analizando su comportamiento y comparándolo con el de los jugadores humanos. El método permite obtener de forma automática un informe sobre el comportamiento del jugador humano que puede compararse con la del agente. Para ello, se ha definido una medida de similitud entre las percepciones computacionales resumen encargadas de capturar y sintetizar los eventos ocurridos durante las partidas.

Esta es una primera aproximación hacia un marco más amplio y aún queda mucho trabajo por hacer. Nos gustaría incorporar en un futuro más detalles que enriquezcan la evaluación de los proyectos. Permitir que el sistema pueda entregar recomendaciones más ricas y detalladas sobre los fallos cometidos por los alumnos durante el diseño e implementación de los algoritmos de Inteligencia Artificial.

## References

- [1] Bustice, H., Barrenechea, E., Pagola, M. Restricted Equivalence Functions. *Fuzzy Sets and Systems* 157, pp. 2333-2346 (2006).
- [2] García-Honrado, I., Fortuny, J. M., Ferrer, M., Morera, L. (2016). Análisis del aprovechamiento de oportunidades de aprendizaje generadas en la discusión en gran grupo de un problema de transformaciones geométricas *Investigación en Educación Matemática XX*, 253-264.
- [3] Gkatzia, D., Hastie, H., Janarthnam, S., & Lemon, O. Generating student feedback from time-series data using Reinforcement Learning. In *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 115-124).
- [4] Hernández-Orallo, J. (2017). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review*, 48(3), 397-447.
- [5] Hingston, P. (2009). A turing test for computer game bots. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(3), 169-186.
- [6] Livingstone, D. (2006). Turing’s test and believable AI in games. *Computers in Entertainment (CIE)*, 4(1), 6.
- [7] López de Mantaras Badia, R., Mesenguer González, P. (2017) ¿Qué sabemos de? *Inteligencia Artificial Editorial: CSIC y Catarata*
- [8] Phedes 2017 Phedes Lab: Linguistic Description of Complex Phenomena. <http://phedes.com/>
- [9] Ramos-Soto, A., Vazquez-Barreiros, B., Bugarín, A., Gewerc, A., & Barro, S. Evaluation of a Data To Text System for Verbalizing a Learning Analytics Dashboard. *International Journal of Intelligent Systems*, 32(2), 177-193.
- [10] Rubio-Manzano, C., & Trivino, G. Improving player experience in Computer Games by using players’ behavior analysis and linguistic descriptions. *International Journal of Human-Computer Studies*, 95, 27-38.
- [11] Rubio-Manzano, C. Similarity measure between linguistic terms by using restricted equivalence functions and its application to expert systems 9th European Symposium on Computational Intelligence and Mathematics. *Faro (Portugal)*, October 4th-7th, 2017.
- [12] Sánchez-Torrubia, M. G., Torres-Blanc, C., & Trivino, G. (2012). An approach to automatic learning assessment based on the computational theory of perceptions. *Expert Systems with Applications*, 39(15), 12177-12191.
- [13] Soni, P. & Hingston, P. (2008, June). Bots trained to play like a human are more fun. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence)*. IEEE International Joint Conference on (pp. 363-369). IEEE.
- [14] Trivino, G & Sugeno, M. Towards linguistic descriptions of phenomena. *International Journal of Approximate Reasoning*, 54(1), 22-34.



# Statistical Language Modelling for Automatic Story Generation

Marta Vicente, Cristina Barros and Elena Lloret

University of Alicante

{mvicente,cbarros,elloret}@dlsi.ua.es

**Abstract**—This abstract summarises the research work published in the *Journal of Intelligent & Fuzzy Systems*, vol. 34 (5), 2018. In this publication, we proposed and analysed an end-to-end Natural Language Generation approach to automatically create fiction stories using statistical language models, both in the macroplanning and surface realisation stages. The results obtained from the evaluation conducted showed that the combination of the aforementioned stages in an end-to-end approach was appropriate and had positive effects in the resulting generated text. Specifically, the use of language models allows the generation task to be more flexible, as far as the adaptation of the approach to different languages, domains and textual genres is concerned.

## I. INTRODUCTION

The generation of stories is a difficult task in which complex cognitive processes are involved. The process of creating a story requires both knowledge of the world and ability to perform tasks such as planning events or resolving situations [8]. Natural Language Generation (NLG), as a subarea of Natural Language Processing, can provide, in turn, useful mechanisms to create a story automatically. In order to generate a text, the responsibilities of NLG include, to determine “*what to say*” and “*how to say it*”. This dual purpose can be addressed as a pipeline of two stages: macroplanning and surface realisation. Macroplanning is in charge of determining the structure of the text to be produced, as well as deciding the content it will have. Surface realisation is focused on taking that content as input, and transforming it into human language, using the appropriate vocabulary, syntactic structures, etc.

In this research work, we studied the problem of whether a story could be automatically generated from scratch without having any clue about the characters, topics, or facts to describe, just having some previous information from other existing stories that could serve as an inspiration source.

To achieve this, a statistical perspective was adopted. Specifically, macroplanning was addressed using Positional Language Models (PLM) while, in the surface realisation stage, Factored Language Models (FLM) were employed. As output, the approach produced a set of sentences in the form of a story, inspired by the structure and linguistic elements of an existing one, but with a different realisation (different events, actions and vocabulary).

The preliminary results showed that statistical language models can definitely contribute to the development of more adaptive and flexible story generation systems, thus providing

mechanisms that can be extended to other kind of domains, languages and textual genres.

## II. BACKGROUND

Recent work in narrative and storytelling has been focused on regenerating stories from graphs of intentions [4] or approaching the task working with discourse and story planning simultaneously to differentiate levels of narrative [9]. Regarding the statistical models used in this research, on the one hand, PLMs have been successfully employed in some language-related areas such as summarisation [2] or information retrieval [5]. However, they have not been directly implemented within the generation framework. Also, in recent years, FLM have been employed for NLG, such in BAGEL [6] or in [7]. The main novelty of our approach is that human intervention in the NLG process is not necessary at any stage, so we avoid the hand-coding of the story constraints, and we increase the automation of the generation process.

## III. END-TO-END STORY GENERATION APPROACH

The NLG approach consists of a macroplanning and a surface realisation stage that are sequentially executed.

### A. Macroplanning.

In the NLG process, macroplanning is responsible for both selecting the content and providing the structure that articulates the output.

Different from the common bag-of-words perspective, PLMs are able to take into account the positions of words together with the number of their occurrences. On this basis, for each position  $i$  of a text, we computed a model  $P(w | i)$  by which every word  $w$  of the vocabulary gets a value for that location  $i$  conditioned by the distance to other appearances of that word within the text.

Using this model, we could obtain the distribution of the elements along the text, considering those with higher values the most relevant ones. We translated this information into a document plan.

### B. Surface realisation.

From the document plan provided by the macroplanning stage, the realisation of the final output was performed relying on FLMs. For this type of models, a word  $w$  is seen as a collection of  $K$  parallel factors, they can be any linguistic feature related to the word, from whose combination a statistical model is built.

We used over-generation and ranking, where a set of candidate sentences is generated and ranked according to the probabilities given by the model. We took into account a simple grammar and prioritised the information specified in the document plan to select the words belonging to each candidate.

#### IV. EXPERIMENTAL SCENARIO

Some experiments were conducted to analyse the performance of the approach. Each stage was empirically adjusted to achieve an optimal configuration and afterwards, a subset of the corpora was used to produce the document plans. The surface realisation module generated from them the final stories. In this research work, from each document plan only one story was generated.

Regarding the macroplanning, some decisions were made about the vocabulary and the design of the document plan itself. The vocabulary was formed by synsets in order to incorporate grammatical and semantic features to the process and, to conform the document plan, we produced one line per each sentence of the source text, containing synsets of verbs, nouns, adjectives and adverbs.

Once the document plans were created, the surface realisation was in charge of generating the final stories following the guidelines the plans had established while being also consistent with the grammar.

We worked with a collection of English children stories, including the Lobo and Matos corpus [3] and automatically gathered stories<sup>1</sup> from *Bedtime stories* and *Hans Christian Andersen: Fairy Tales and Stories*.

#### V. EVALUATION, RESULTS AND DISCUSSION

In order to evaluate how the document plan impacted in the generated text as well as the quality of the generated fiction story, we performed two tests.

On the one hand, we evaluated how the synsets and their distribution in the document plan were reflected in the resulting stories. To estimate the influence of the document plan in the generation of the new tales, we analysed the relation between both documents, measuring how the elements were shared.

On the other hand, a user evaluation was performed to analyse the general problems and errors of the approach. We used a point score scale, and three users manually read and analysed a set of stories in order to assess the system performance and provide feedback based on the detection of problems. They shed light on the limitations of our approach.

At a word level, the users highlighted the appropriate variety and richness of the vocabulary. Even though, for some examples they remarked that it would be adequate to use more synonyms, in order to prevent finding consecutive sentences repeating exactly the same terms.

At a sentence level, the evaluation revealed that, independently of the relation with their neighbours, sentences, in general, would become more meaningful once inflected. To

understand why some excerpts present an oddly shape, we should attend to the grammar beneath, on the one side, and to the semantic aspect, on the other.

Thirdly, at the level of discourse, adapted to our defined framework, the users were able to identify themes and guiding threads. They also noted the presence of characters as being involved in possible actions. They stated that the repetition of elements was essential to detect those underlying features. This reinforces our initial consideration regarding the relevance of a proper distribution of the elements through the text. Along with these comments, the users indicated that in some of the examples the absence of enough information impeded the assumption of any of those elements (themes, characters), but mostly this difficulty was related to the semantics of the generated tales.

#### VI. CONCLUSION

This paper proposed and evaluated a novel end-to-end NLG approach for generating fiction stories that integrates the macroplanning and the surface realisation stages within a statistical framework.

Evaluation showed that all the sentences generated by the approach contained elements from the document plan, 81% of the sentences having at least one of the highest rated elements. Furthermore, as there were not only one type of element within the document plan, on average, 83% of the content of each generated tale came from the document plan. Concerning the results of the user evaluation, we received positive feedback from 47% of the generated stories. We consider these results to be an indicator of the positive effects of combining the two stages (i.e. macroplanning and surface realisation).

Although the results are still preliminary, there are some issues which need to be improved and some interesting research lines opening new directions for future work.

#### ACKNOWLEDGMENT

Funded by the Generalitat Valenciana (ACIF/2016/501) and Spanish Government (TIN2015-65100-R, TIN2015-65136-C2-2-R).

#### REFERENCES

- [1] C. Baker, C.J. Fillmore, and J.B. Lowe, *The Berkeley FrameNet Project*, in: *Procs. of the 36th Meeting of the ACL – Volume 1*, 1998, 86–90.
- [2] S. Liu et al, *Positional Language Modelling for Extractive Broadcast News Speech Summarization*, in: *INTERSPEECH*, 2015, 2729–33.
- [3] P.V. Lobo, and D.M. De Matos, *Fairy Tale Corpus Organization Using Latent Semantic Mapping and an Item-to-Item Top-N Recommendation Algorithm*, in: *LREC*, 2010.
- [4] S.M. Lukin, L.I. Reed and M.A. Walker, *Generating Sentence Planning Variations for Story Telling*, in: *16th SIGdial*, 2015, 188.
- [5] Y. Lv, Yuanhua, and C. Zhai, *Positional Language Models for Information Retrieval*, in: *Procs. of the 32Nd ACM SIGIR*, New York, 2009, 299–306.
- [6] F. Mairesse and S. Young, *Stochastic Language Generation in Dialogue Using Factored Language Models*, *Computational Linguistics*, 40(4), 2014, 763–99.
- [7] E.M. Novais, and I. Paraboni, *Portuguese Text Generation Using Factored Language Models*, *Journal of the Brazilian Computer Society* 19(2), 2012, 135–46.
- [8] S.R. Turner, *Minstrel: A Computer Model of Creativity and Storytelling*, Ph.D. Dissertation. University of California at Los Angeles, 1993.
- [9] D.R. Winer and R.M. Young, *Discourse-Driven Narrative Generation with Bipartite Planning*, in: *The 9th INLG*, 2016, 11–20.

<sup>1</sup><https://freestoriesforkids.com/>, <http://hca.gilead.org.il/>