

Generación automática de descripciones lingüísticas de datos en meteorología usando metaheurísticas

Andrea Cascallar Fuentes

Alejandro Ramos Soto

Alberto José Bugarín Diz

Centro Singular de Investigación en Tecnoloxías da Información

Universidade de Santiago de Compostela

II Machine Learning Workshop Galicia

citius.usc.es

Tabla de contenidos

- 1 Introducción
- 2 Trabajos relacionados
- 3 Objetivo
- 4 Descripciones de observación meteorológica en tiempo real
- 5 Evaluación
- 6 Conclusiones y trabajo futuro



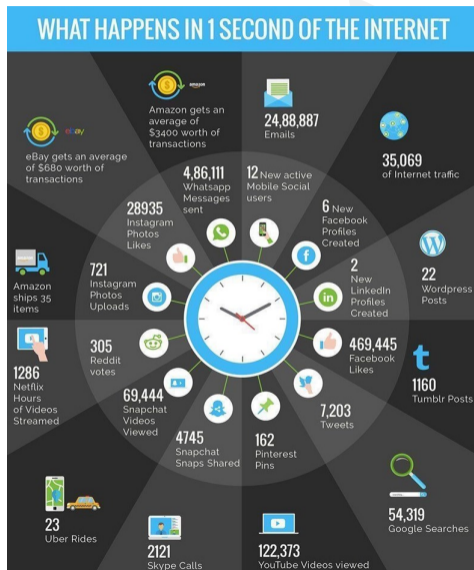
Tabla de contenidos

- 1 **Introducción**
- 2 Trabajos relacionados
- 3 Objetivo
- 4 Descripciones de observación meteorológica en tiempo real
- 5 Evaluación
- 6 Conclusiones y trabajo futuro



Introducción

- Grandes cantidades de datos generados en cada instante
 - ▷ 2.5 billones de bytes generados cada día
 - ▷ En 2009 se generó tanta cantidad de datos como la que había hasta ese momento



Introducción

- Gran valor en la toma de decisiones
- Necesario explotarlos
- Herramientas de visualización ofrecen resultados difíciles de interpretar
- Inteligencia Artificial
 - ▷ Herramientas de análisis de datos
 - ▷ Herramientas de procesamiento del lenguaje



Integrate



Analyze



Visualize

Introducción

■ Generación del Lenguaje Natural (NLG)

▷ Data-To-Text

- Generan texto a partir de datos simbólicos o numéricos
- Capaces de procesar grandes cantidades de datos

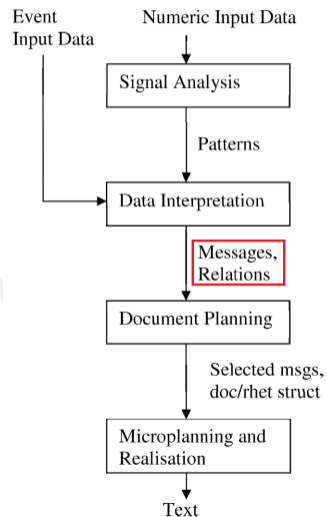


Tabla de contenidos

- 1 Introducción
- 2 Trabajos relacionados**
- 3 Objetivo
- 4 Descripciones de observación meteorológica en tiempo real
- 5 Evaluación
- 6 Conclusiones y trabajo futuro



Trabajos relacionados

■ Generación de lenguaje natural

▷ Proyecto BabyTalk

- Sistema BT-45

▷ Sistema de generación de predicciones meteorológicas de la MetOffice británica

- 97 % consideraron las predicciones textuales útiles para comprender mejor la predicción simbólica

▷ SUMTIME-MOUSAM

▷ RoadSafe

▷ TEMSIS

■ Descripciones lingüísticas de datos

▷ Resumen del consumo energético en los hogares

▷ Soporte en la toma de decisiones en inversiones

▷ Resumen de la monitorización de la actividad física diaria de una persona

The screenshot displays the NICU Reporter interface. On the left is a dark sidebar with navigation options: PATIENTS, ALERTS, MESSAGES, and NOTES. The main content area is titled 'PATIENT STATUS UPDATE' for '10 FEBRUARY 2015 - 08:50'. It lists the current doctor as 'DR. R. DINESH' and the nurse on duty as 'M. DALE'. A text block describes a patient event: 'Between 16:40 and 17:04 there were a number of successive desaturations down to 0. Fraction of Inspired Oxygen (FIO₂) was raised to 100%. There were 3 successive bradycardias down to 60. Noncpuff ventilation was given to the baby a number of times, and the baby was re-intubated successfully.' Below this, another text block states: 'The baby was resuscitated and this caused bruising to the skin. Blood gas results received at 16:45 showed that: PH = 7.3, PO₂ = 5, PCO₂ = 6.9, BE = -0.7'. A third text block notes: 'At 17:15, FIO₂ was lowered to 33%. TcPO₂ had rapidly decreased to 6.8. Previously T1 had rapidly increased to 35.0.' At the bottom of the main area are buttons for 'NEW REPORT' and 'PRINT REPORT'. On the right side, there is a patient profile card for 'DAVID WILLIAMS' (DOB: 01/02/2015, GENDER: MALE, DATE ADMITTED: 11/01/2015) and a 'KEY HEALTH STATISTICS' table.

KEY HEALTH STATISTICS	
HR	95
T1	35.7
T2	34.5
SpO ₂	100%

Trabajos relacionados

- GALiWeather: predicciones meteorológicas en lenguaje natural
- Proyecto PLASTIC: inteligencia de negocio
- MonitorSI-Text: monitorización de la calidad de servicios informáticos
- SLAR: e-Learning
- MeteoAvisos-Text: generación de avisos meteorológicos
- PREBENCAR: e-Health

Predicción a curto prazo para Santiago de Compostela

	martes, 17 de outubro			mércores, 18 de outubro			xoves, 19 de outubro			venres, 20 de outubro		
	Mañá	Tarde	Noite	Mañá	Tarde	Noite	Mañá	Tarde	Noite	Mañá	Tarde	Noite
Estado do ceo												
Vento	↗	↑	↑	↑	↗	↗	↑	↗	↗	↗	↗	↗
Ocorrencia chuvia	40%	40%	80%	60%	40%	40%	70%	85%	80%	80%	80%	90%
Temperatura(°C)	MIN MAX 12* 18*			MIN MAX 12* 18*			MIN MAX 12* 18*			MIN MAX 17* 19*		
Ver evolución												
Avisos Meteo	■			■			■			-		
Calidade aire	■			■			■			-		
Comentario	<p>Agárdase para os vindeiros días unha alternancia de períodos de ceos moi anubrados con outros con nubes e claros. Agárdanse precipitacións tódolos días. As temperaturas serán altas para as mínimas e normais para as máximas respecto ó habitual nesta época do ano, con mínimas en ascenso moderado e máximas sen cambios.</p> <p>En canto ó estado da calidade do aire, manterase bo en termos xerais.</p>											

Descripciones lingüísticas de datos

Utilización de algoritmos de búsqueda para la extracción de información

- Soluciones heurísticas
 - ▷ GALiWeather
- Soluciones metaheurísticas: algoritmos genéticos
 - ▷ Descripción de series temporales
 - ▷ Resúmenes lingüísticos a partir de datos financieros

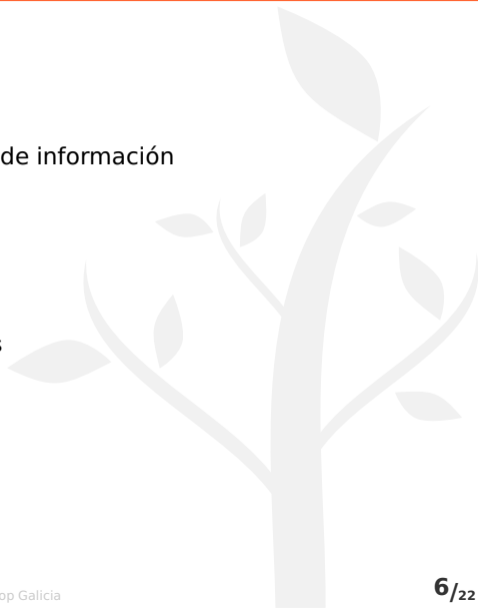


Tabla de contenidos

- 1 Introducción
- 2 Trabajos relacionados
- 3 Objetivo**
- 4 Descripciones de observación meteorológica en tiempo real
- 5 Evaluación
- 6 Conclusiones y trabajo futuro



Objetivo

Generación de descripciones lingüísticas de **datos de observación meteorológica** en tiempo real

- Métodos basados en lógica borrosa
- Métodos de búsqueda metaheurística
- Restricción temporal de 5 minutos
- Agencia Meteorológica de la Xunta de Galicia (MeteoGalicia)
 - ▷ Predicciones a corto plazo
 - Estado general de la comunidad
 - Situación para cada ayuntamiento



Tabla de contenidos

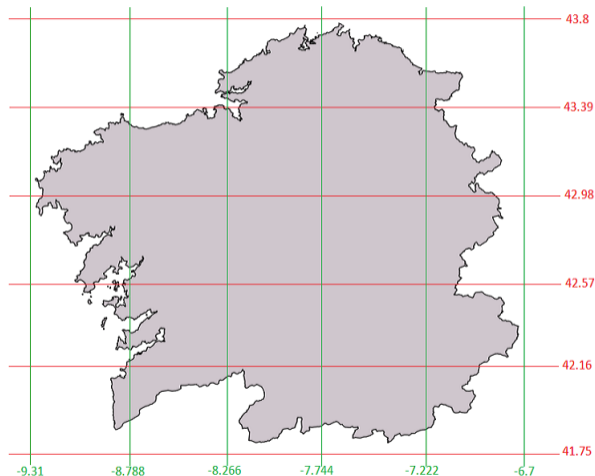
- 1 Introducción
- 2 Trabajos relacionados
- 3 Objetivo
- 4 Descripciones de observación meteorológica en tiempo real**
- 5 Evaluación
- 6 Conclusiones y trabajo futuro



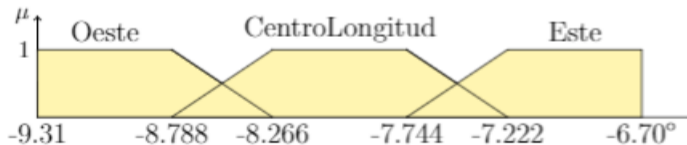
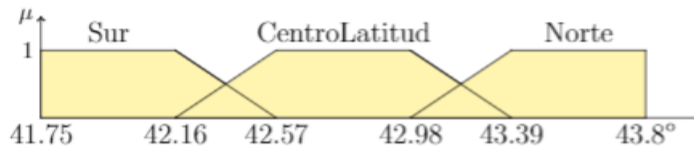
Descriptores geográficos

Norte, Sur, Centro, Este, Oeste

- Basados en longitud y latitud
- Definidos de forma no precisa



Descriptores geográficos



Variables de entrada

Estado del cielo



Variables de entrada

Viento

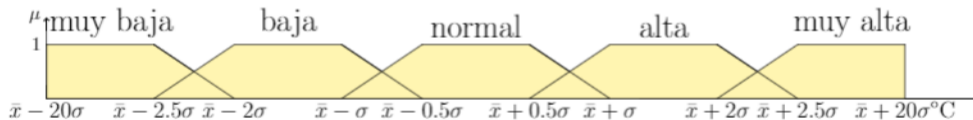
	VAR										

Temperatura

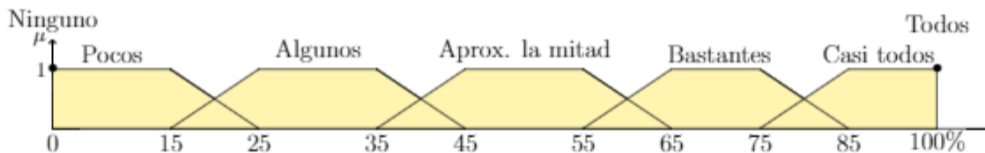
- Grados centígrados

Variables lingüísticas

- Estado del cielo: misma definición
- Viento: misma definición
- Temperatura: TMAX y TMIN



Cuantificadores



Generación de sentencias

Tipo I: En pocos ayuntamientos el estado del cielo es soleado

- Versión inicial
 - ▷ Se generan **todas las descripciones posibles**
 - ▷ 63.875 descripciones
 - ▷ 25 minutos
- Optimización 1
 - ▷ **Eliminar valores no presentes** en el estado actual
 - ▷ Aproximadamente 2.000 descripciones
 - ▷ 50-60 segundos

Tipo II: En algunos ayuntamientos del Norte el estado del cielo es soleado

- Versión inicial
 - ▷ Se generan **todas las descripciones posibles**
 - ▷ 1.094.170 descripciones
 - ▷ Aproximadamente 3 días
- Optimización 1
 - ▷ Eliminar **valores no presentes** en el estado actual
 - ▷ Aproximadamente 13.000 descripciones
 - ▷ Aproximadamente 8 horas

Descripciones de tipo II: Algoritmo de búsqueda metaheurística

- Optimización 1
 - ▷ Coste temporal muy elevado
- Solución metaheurística
 - ▷ Algoritmos genéticos
 - ▷ Algoritmos de búsqueda local
 - ▷ Búsqueda Tabú
 - ▷ **Temple simulado**



Mejores soluciones

Eliminar descripciones "Ninguno"

Umbral en el grado de cumplimiento = 0,5

Ordenación de las descripciones

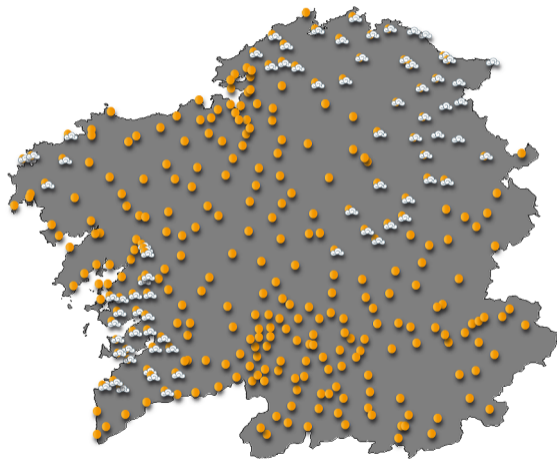
Tipo I	Tipo II
Máximo 100 descripciones	Aproximadamente 400 descripciones
50-60 segundos	Aproximadamente 2 minutos

Tabla de contenidos

- 1 Introducción
- 2 Trabajos relacionados
- 3 Objetivo
- 4 Descripciones de observación meteorológica en tiempo real
- 5 Evaluación**
- 6 Conclusiones y trabajo futuro



Estado del cielo del 20 de septiembre a las 18:30



Tipo I

Posición	Descripción
2	En bastantes ayuntamientos el estado del cielo es despejado

Tipo II

Posición	Descripción
16	En algunos ayuntamientos del Oeste el estado del cielo es nubes y claros
47	En pocos ayuntamientos del Norte el estado del cielo es nubes y claros
54	En pocos ayuntamientos del Norte el estado del cielo es muy nublado

Estado del viento del 20 de septiembre a las 18:30



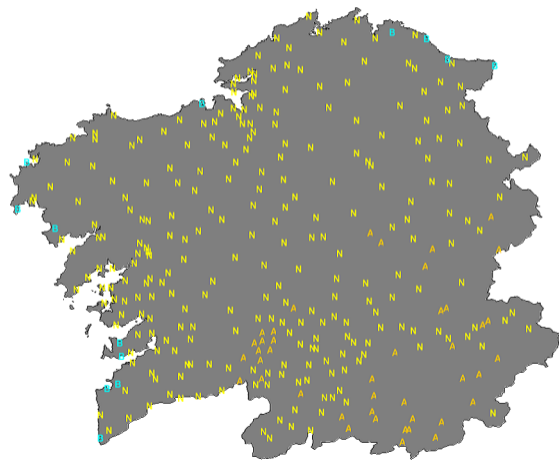
Tipo I

Posición	Descripción
1	En aproximadamente la mitad de ayuntamientos el viento tiene dirección Oeste e intensidad baja

Tipo II

Posición	Descripción
102	En pocos ayuntamientos del Oeste el viento tiene dirección Sudoeste e intensidad baja
105	En pocos ayuntamientos del Norte el viento tiene dirección Norte e intensidad baja
110	En pocos ayuntamientos del Este el viento tiene dirección Este e intensidad baja

Temperaturas máximas del 20 de septiembre a las 18:30



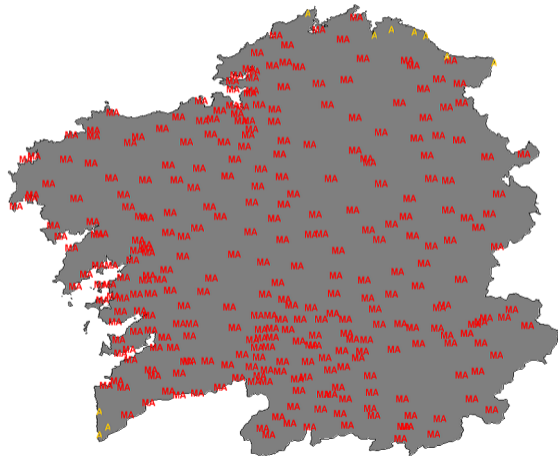
Tipo I

Posición	Descripción
6	En casi todos los ayuntamientos la temperatura es normal con respecto a las máximas y muy alta con respecto a las mínimas

Tipo II

Posición	Descripción
12	En algunos ayuntamientos del Sur la temperatura es alta con respecto a las máximas y muy alta con respecto a las mínimas
120	En pocos ayuntamientos del Oeste la temperatura es baja con respecto a las máximas y muy alta con respecto a las mínimas

Temperaturas mínimas del 20 de septiembre a las 18:30



Tipo I

Posición	Descripción
6	En casi todos los ayuntamientos la temperatura es normal con respecto a las máximas y muy alta con respecto a las mínimas

Tipo II

Posición	Descripción
12	En algunos ayuntamientos del Sur la temperatura es alta con respecto a las máximas y muy alta con respecto a las mínimas
120	En pocos ayuntamientos del Oeste la temperatura es baja con respecto a las máximas y muy alta con respecto a las mínimas

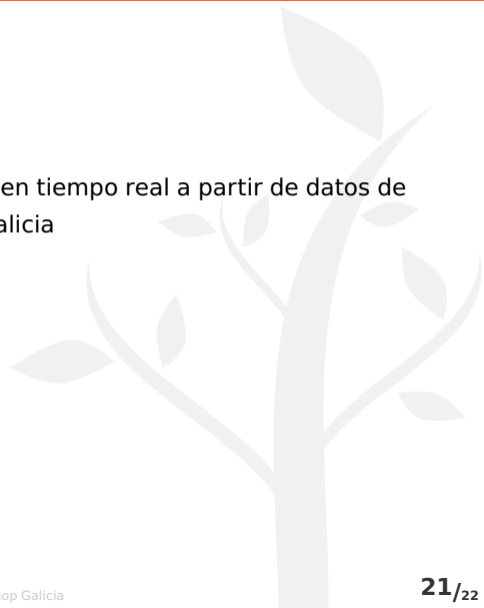
Tabla de contenidos

- 1 Introducción
- 2 Trabajos relacionados
- 3 Objetivo
- 4 Descripciones de observación meteorológica en tiempo real
- 5 Evaluación
- 6 Conclusiones y trabajo futuro**



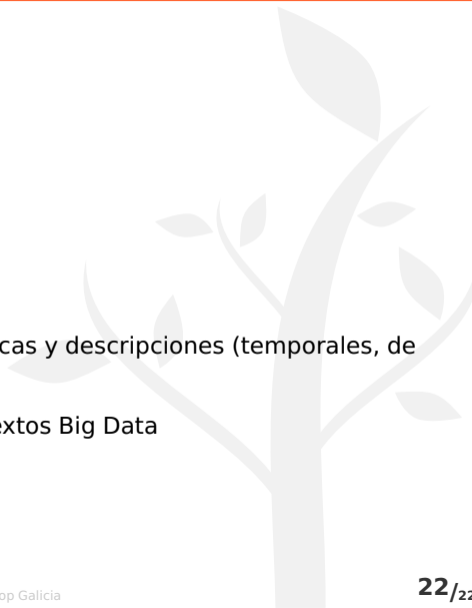
Conclusiones

- Método de generación de descripciones lingüísticas en tiempo real a partir de datos de observación meteorológica para la comunidad de Galicia
- Aplicación de solución metaheurística
- Ejecución total de aproximadamente 3 minutos



Trabajo futuro

- Nuevos descriptores geográficos
 - ▷ Noroeste, Sudeste,...
- Comparación entre algoritmos de búsqueda
- Estudio del dominio
- Desarrollo de un sistema NLG completo
- Aplicación en ámbitos que demanden nuevas analíticas y descripciones (temporales, de relación, ...)
- Nuevas metaheurísticas de búsqueda en otros contextos Big Data



Gracias por su atención

andrea.cascallar.fuentes@usc.es

citius.usc.es



XUNTA
DE GALICIA



UNIÓN EUROPEA

galicia

Cofinanciado pola Unión Europea

Fondo Europeo de Desenvolvemento Rexional "Unha maneira de facer Europa"