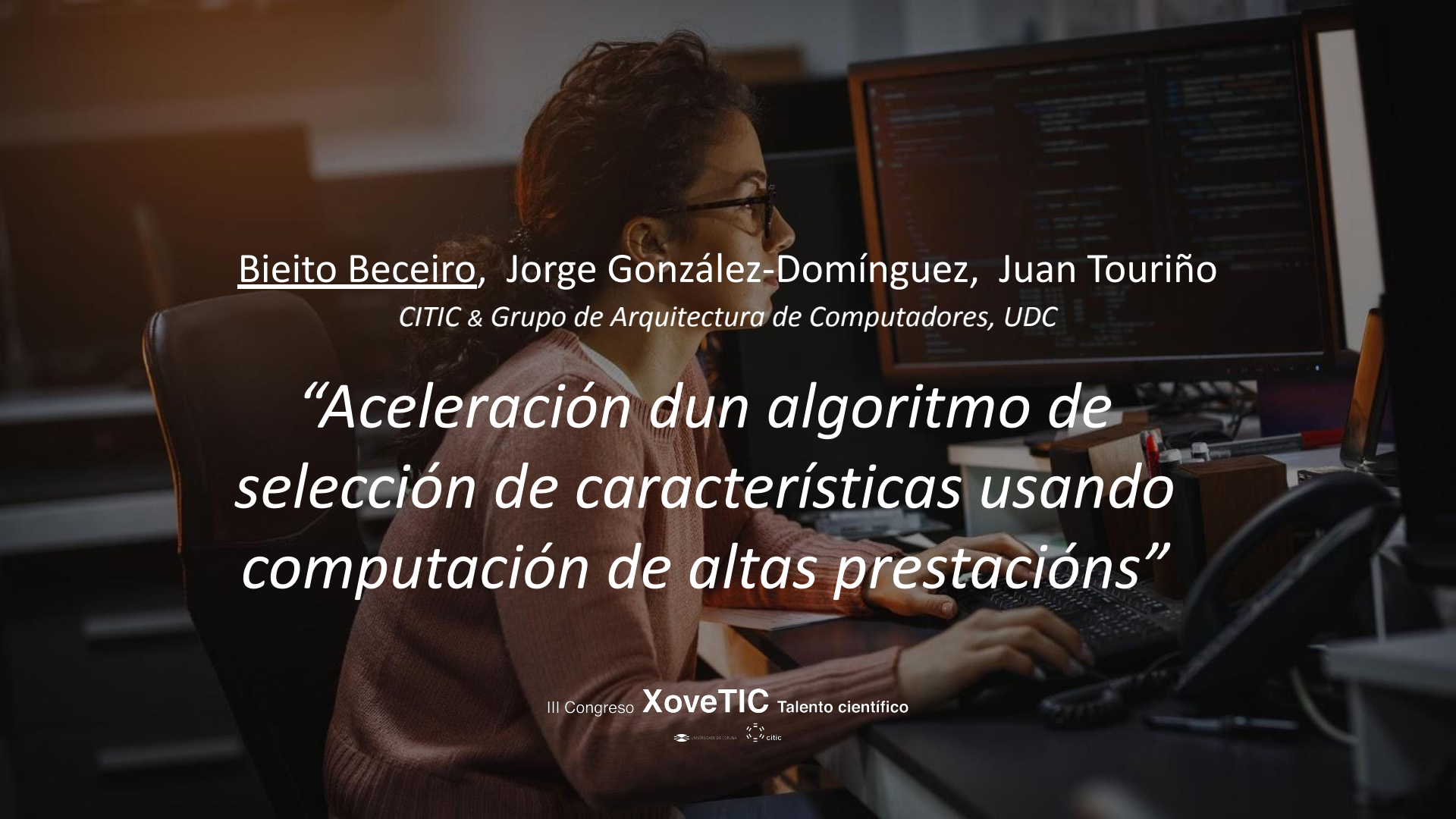


Bieito Beceiro, Jorge González-Domínguez, Juan Touriño

*CITIC & Grupo de Arquitectura de Computadores, UDC*

# *"Aceleración dun algoritmo de selección de características usando computación de altas prestacións"*

III Congreso **XoveTIC** Talento científico

## 1. Introdución

Contexto e impacto do problema

## 2. Desenvolvemento

Deseño e implementación da solución

## 3. Probas de rendemento

Estudo da eficacia da solución proposta

## 4. Conclusións

Impacto dos resultados

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

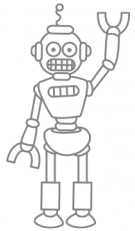
# Contexto

*Disciplinas en auxe na actualidade*

VIVENDAS INTELIXENTES

VISIÓN ARTIFICIAL

PROCESAMENTO DE LINGUAXE



ROBÓTICA

SAÚDE

CONDUCIÓN AUTÓNOMA

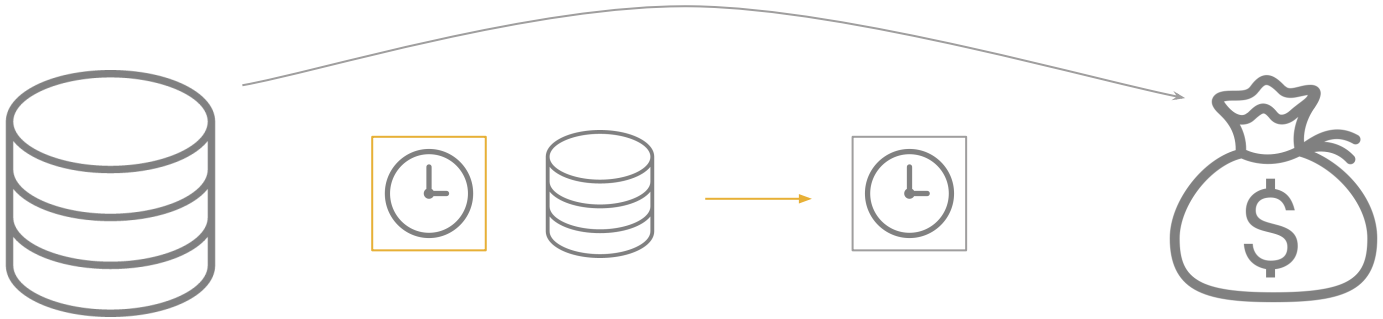
MÁRKETING E BUSINESS INTELLIGENCE

galicia citic UNIVERSIDADE DA CORUÑA Xacobeo 2021 UNIÓN EUROPEA XUNTA DE GALICIA

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Contexto

*Redución da dimensionalidade*




FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Contexto

*FEAST: A Feature Selection Toolbox*

- Selección de características mediante filtrado
  - Obtemos o subconxunto que mellor describe ao dataset orixinal como paso previo á análise
- Ganancia de información como métrica
  - Maximizar relevancia intentando reducir redundancia
- **Método JMI** (*Joint Mutual Information*)
  - Mellores resultados respecto a precisión, estabilidade e flexibilidade
  - Custo computacional moi alto
    - → Potenciado por problemas "*big p, small n*", cada vez máis habituais

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

1. Introdución

Contexto e impacto do problema

2. Desenvolvemento

Deseño e implementación da solución

3. Probas de rendemento

Estudo da eficacia da solución proposta

4. Conclusións

Impacto dos resultados

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Desenvolvemento

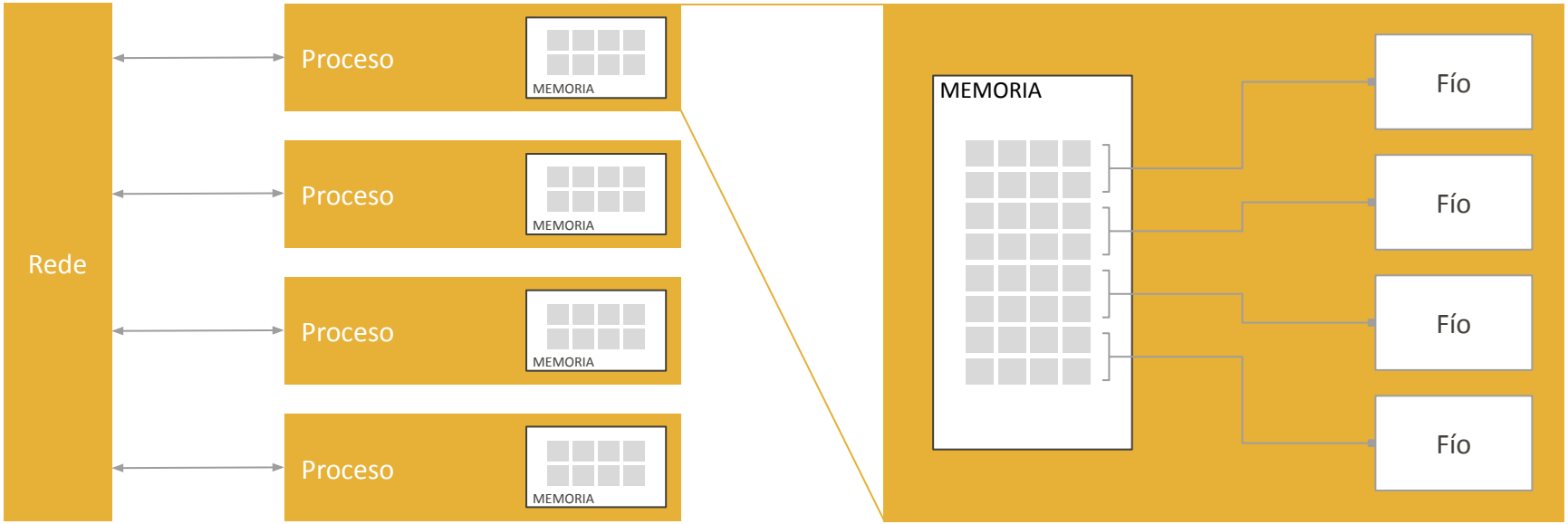
*Visión xeral*

- Implementación paralela híbrida
  - Paralelismo explícito e implícito con descomposición de dominio
- Lectura semi-distribuída
- Compresión de rango

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Desenvolvemento

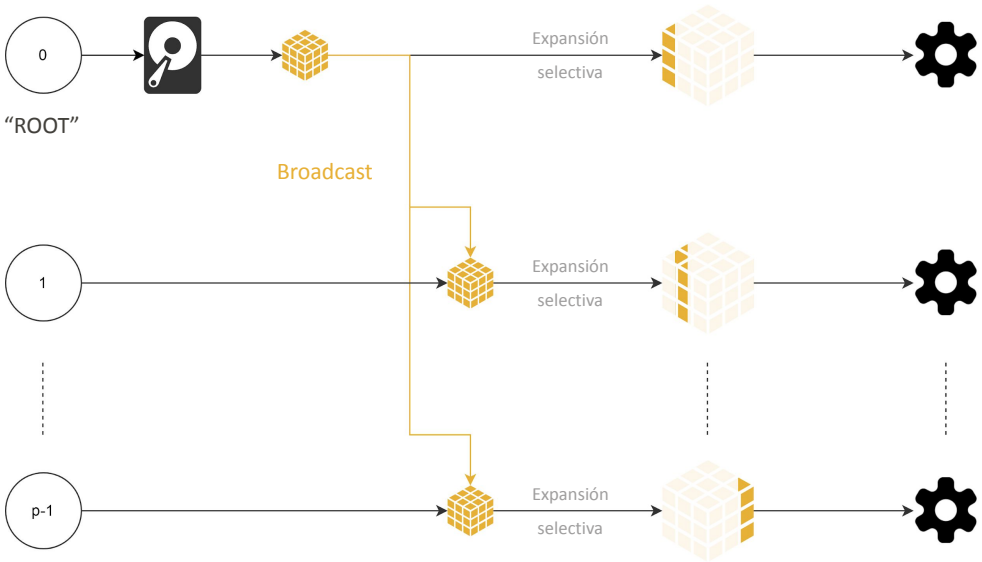
*Paralelismo explícito e implícito*




FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Desenvolvemento

*Lectura semi-distribuída*



# Desenvolvemento

*Compresión de rango*

Característica orixinal

| 2000 | 1000 | 1000 | 0 | 1000 |
|---|---|---|---|---|

Rango: [0, 2000]

2001 valores posibles
3 valores utilizados

| | |
|---|---|
| 2000 | → 0 |
| 1000 | → 1 |
| 0 | → 2 |

Característica transformada

| 0 | 1 | 1 | 2 | 1 |
|---|---|---|---|---|

Rango: [0, 2]

3 valores posibles
3 valores utilizados

1. Introdución

Contexto e impacto do problema

2. Desenvolvemento

Deseño e implementación da solución

3. Probas de rendemento

Estudo da eficacia da solución proposta

4. Conclusións

Impacto dos resultados

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Probas de rendemento

*Entorno de probas*

- Clúster privado de 16 nodos
  - CPU: 2 x octa-core con Hyperthreading (16 núcleos / 32 fíos lóxicos)
  - RAM: 64 GB
  - **En total:** 256 núcleos (ata 512 fíos lóxicos) e 1024GB de memoria RAM
- Uso de datasets representativos
  - Ratio de mostras-características
  - Bi-clase ou multiclase
  - E un dataset de gran tamaño (~512GB)

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Probas de rendemento

*Resultados*

- Método *Joint Mutual Information*
  - Aceleracións de ata 198x para 256 elementos de procesado (eficiencia > 75%)
- Lectura semi-distribuída
  - Posibilidade de analizar un dataset que previamente era imposible
- Compresión de rango
  - Aforro de memoria e tempo de cómputo de JMI cun preprocesado mínimo

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Probas de rendemento

*Nun caso real...*

Seleccionar 200 características de News20

| FEAST-JMI | Parallel-JMI |
|---|---|
| ~30 minutos | ~10 segundos |
| ~a duración de 3 exposicións | ~o tempo que tardei en dicir esta oración |

galicia citic UNIVERSIDADE DA CORUÑA Xacobeo 2021 UNIÓN EUROPEA XUNTA DE GALICIA

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Probas de rendemento

*Nun caso real...*

Seleccionar 200 características de E2006

| FEAST-JMI | Parallel-JMI |
| --- | --- |
| imposible | ~13 minutos |
| o dataset non cabe na memoria do nodo | |

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

1. Introdución

Contexto e impacto do problema

2. Desenvolvemento

Deseño e implementación da solución

3. Probas de rendemento

Estudo da eficacia da solución proposta

4. Conclusións

Impacto dos resultados

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Conclusións

- O algoritmo forma parte dunha libraría *open-source* que se publicará proximamente
- Boa aceleración e escalabilidade
- Posibilidade de analizar datasets de tamaño superior á memoria dun nodo
- Moi bos resultados para a optimización de compresión de rango

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad

# Grazas!

**Bieito Beceiro Fernández**

Outubro 2020

FONDO EUROPEO DE DESARROLLO REGIONAL "Una manera de hacer Europa" Promover el desarrollo tecnológico, la innovación y una investigación de calidad