

# Minería de Datos basada en Agentes y Artefactos (*JaCa-DDM*)

Dr. Alejandro Guerra-Hernández

**Universidad Veracruzana**

Centro de Investigación en Inteligencia Artificial

*Sebastián Camancho No. 5, Centro*

*Xalapa, Ver., México 91000*

<mailto:aguerra@uv.mx>

<http://www.uv.mx/personal/aguerra>

Reunión Universidad Veracruzana - INAOE

Análisis de Imágenes Médicas

10 de Octubre, 2016

# JaCa-DDM [4]

- ▶ Disponible en:

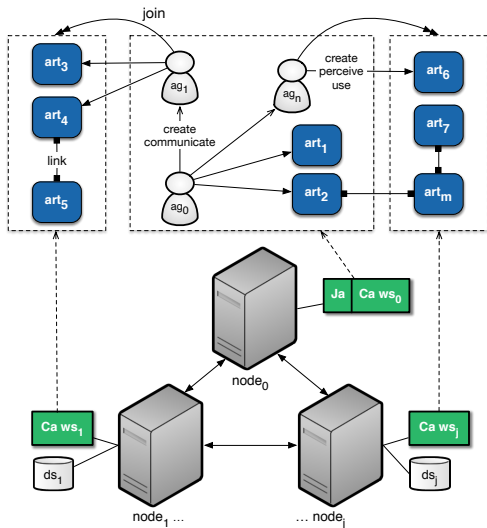
<https://sourceforge.net/projects/jacaddm/>

- ▶ Tiene dos componentes conceptuales:

**Estrategia JaCa-DDM.** Flujo de trabajo basado en **agentes Jason** [3] y **artefactos CArtAgO** [6], basados en Weka [8] y MOA [2].

**Sistema de Despliegue JaCa-DDM.** La manera en que los componentes del sistema se ubican en la arquitectura de cómputo distribuida donde se ejecutará la estrategia. Basada en **nodos CArtAgO**.

# La arquitectura JaCa-DDM

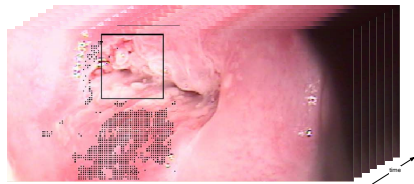


- ▶ 2 procesadores Xeon a 2.40 GHz, con 4 cores, dos hilos c/u.
- ▶ 24 GB of RAM.
- ▶ GPU Tesla C2050 con 448 cores, 6 GB de memoria.

# Segmentación basada en píxeles

Secuencias de imágenes de colposcopia, presentando posibles lesiones cervicales pre-cancerosas [1]:

- ▶ 38 pacientes.
- ▶ Pre-procesamiento con FIJI [7].
- ▶ 1,434,060 píxeles de entrenamiento.
- ▶ 30 atributos.
- ▶ 6 clases, agrupadas en 2.



# Windowing (C4.5)

## Algoritmo

```
1: function Windowing(Exs)
2:   Window  $\leftarrow$  sample(Exs)
3:   Exs  $\leftarrow$  Exs - Window
4:   repeat
5:     stopCond  $\leftarrow$  true
6:     model  $\leftarrow$  induce(Window)
7:     for ex  $\in$  Exs do
8:       if classify(model, ex)  $\neq$  class(ex) then
9:         Window  $\leftarrow$  Window  $\cup$  {ex}
10:        Exs  $\leftarrow$  Exs - {ex}
11:        stopCond  $\leftarrow$  false
12:      end if
13:    end for
14:  until stopCond
15:  return model
16: end function
```

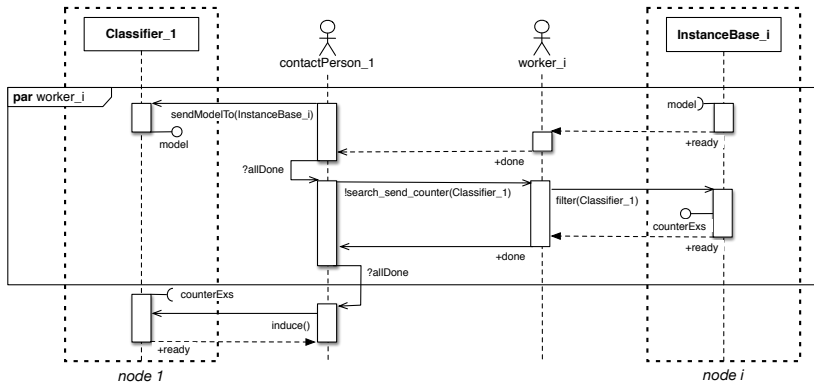
## Ventajas

- ▶ Reduce ejemplos.
- ▶ Mantiene precisión.

## Desventajas

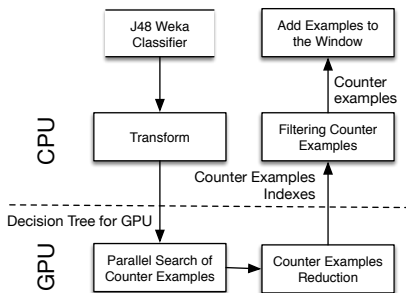
- ▶ Costo inducción.
- ▶ Costo clasificación.

# Estrategia *Parallel Counter GPU Extra*

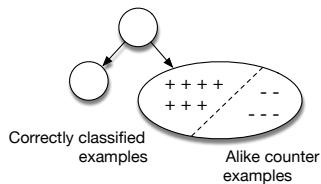


# El GPU y el extra [5]

## Clasificación en GPU



## Reducción de la ventana



# Precisión, sensibilidad y especificidad

Estrategia	Precisión		Wilcoxon test	Sen	Esp
<b>Parallel Counter GPU Extra</b>	67.61 ±	19.32	–	60.96	64.83
Weka Centralized	63.68 ±	18.44	Lost	60.80	61.60
Centralizing VFDT	53.34 ±	20.58	Lost	53.10	58.51
Bagging	64.25 ±	21.78	Lost	65.40	59.16
Random Forest	58.88 ±	23.71	Lost	68.78	49.34
Acosta2009	67.00 ±	n/a	n/a	71.00	59.00



# Número de ejemplos y tiempo

Estrategia	% Ejemplos		Tiempo (Seg.)	
<b>Parallel Counter GPU Extra</b>	37.00 ±	3.52	3782.26 ±	1094.21
Weka Centralized	100.00 ±	0.00	6436.64 ±	923.16
Centralizing VFDT	100.00 ±	0.00	32.03 ±	2.61
Bagging	100.00 ±	0.00	1138.83 ±	108.83
Random Forest	100.00 ±	0.00	1817.10 ±	179.18
Acosta2009	n/a		n/a	

# Referencias I



H-G Acosta-Mesa, N Cruz-Ramírez y R Hernández-Jiménez. "Aceto-white temporal pattern classification using k-NN to identify precancerous cervical lesion in colposcopic images". En: *Computers in biology and medicine* 39.9 (2009), págs. 778-784.



A Bifet y col. "MOA: Massive online analysis". En: *The Journal of Machine Learning Research* 11 (2010), págs. 1601-1604.



RH Bordini, JF Hübner y M Wooldridge. *Programming Multi-Agent Systems in Agent-Speak using Jason*. John Wiley & Sons Ltd, 2007.



X Limón y col. "An Agents & Artifacts approach to Distributed Data Mining". En: *MICAI 2013: Eleventh Mexican International Conference on Artificial Intelligence*. Ed. por F Castro, A Gelbukh y MG Mendoza. Vol. 8266. Lecture Notes in Artificial Intelligence. Berlin Heidelberg: Springer Verlag, 2013, págs. 338-349.



X Limón y col. "A windowing based GPU optimized strategy for the induction of Decision Trees". En: *Artificial Intelligence Research and Development*. Ed. por E Armengol, D Boixader y F Grimaldo. Vol. 277. Frontiers in Artificial Intelligence and Applications. Amsterdam, NL: IOS Press, 2015, págs. 100-109.

# Referencias II



A Ricci, M Pianti y M Viroli. "Environment programming in multi-agent systems: an artifact-based perspective". En: *Autonomous Agents and Multi-Agent Systems* 23.2 (2011), págs. 158-192.



J Schindelin y col. "Fiji: an open-source platform for biological-image analysis". En: *Nature methods* 9.7 (2012), págs. 676-682.



IH Witten, E Frank y MA Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA., USA: Morgan Kaufmann Publishers, 2011.