

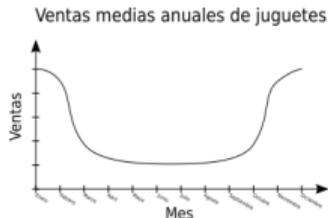
Detección de concept drift en minería de procesos mediante agrupamiento jerárquico de trazas

Víctor J. Gallego Fontenla, Juan C. Vidal Aguiar, Manuel Lama Penín

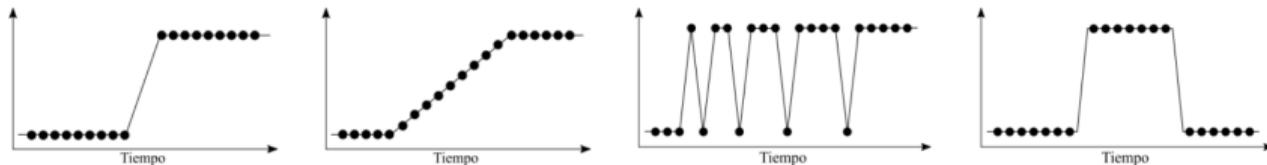
Octubre 2017

Concept Drift

Cambios en la **distribución conjunta** de los valores de las variables predictoras y predichas.



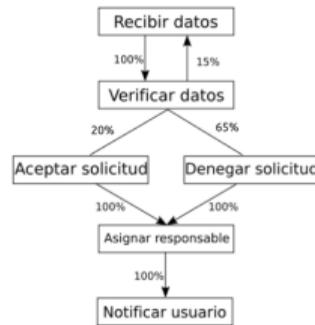
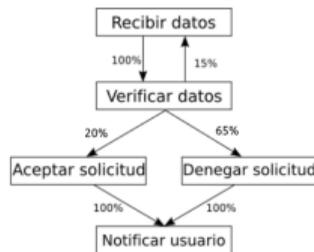
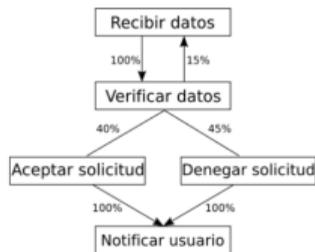
Cambios **abruptos**, **incrementales**, **graduales** y **recurrentes**.



Concept Drift

Importante en minería de procesos para **optimizar** el funcionamiento de las organizaciones. Las **técnicas tradicionales no son aplicables**, las trazas representan **datos complejos**, lo que dificulta el tratamiento.

Buscamos detectar **cambios estructurales**, no de comportamiento.



Concept Drift

Tres problemas a resolver:

- **Detección** del cambio

En el registro se ha detectado un cambio

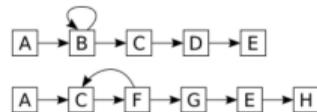
- **Localización** y **caracterización** del cambio.

El cambio es de tipo abrupto entre las trazas 6 y 7

- **Evolución** del proceso.

La evolución del proceso ha sido la siguiente

1	ABCDE
2	ABBBCDE
3	ABBCDE
4	ABCDE
5	ABCDE
6	ABBBBBBCDE
7	ACFGEH
8	ACFCFGEH
9	ACFGEH
10	ACFCFCFCFGEH
11	ACFGEH
12	ACFGEH



Algoritmo

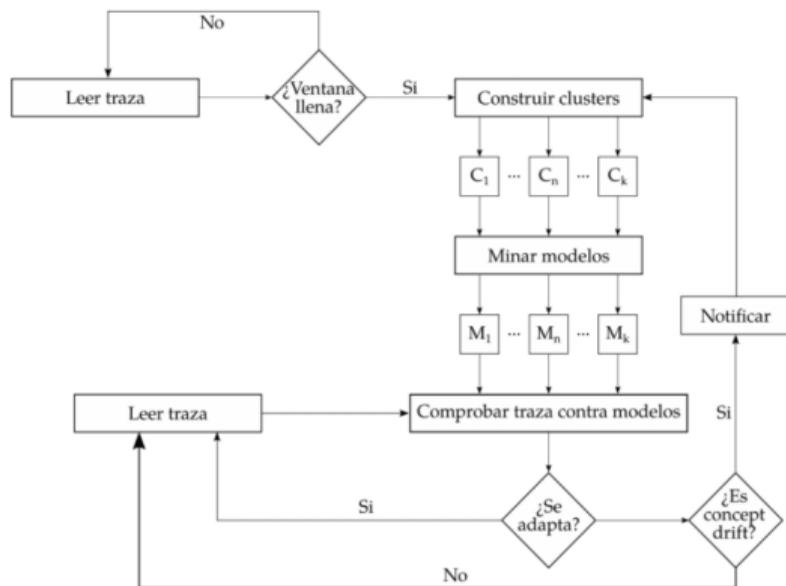
Objetivo

Detectar momento en el que un cierto **porcentaje de trazas mayor que un umbral** no se adapta a los modelos subyacentes hasta el momento.

Ventana deslizante con las últimas n trazas del registro.

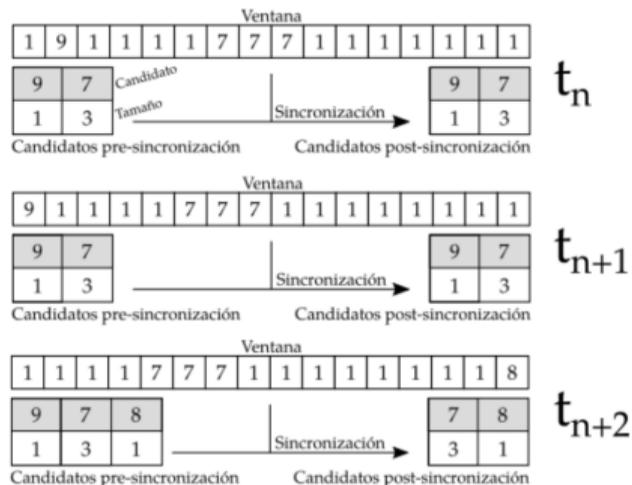
Clustering sobre la ventana para **generar modelos**.

Comprobación de las trazas sobre los modelos generados.



Algoritmo - Sincronización

Eliminar **candidatos a concept drift** que ya **no aparecen en la ventana** cuando se añade un dato nuevo.

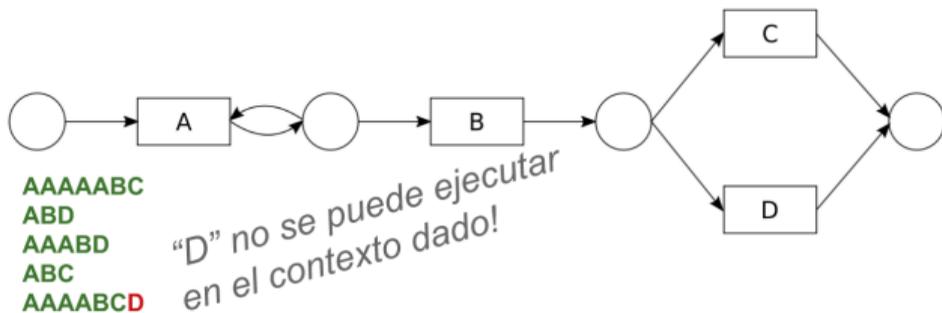


Algoritmo - Comprobación de trazas

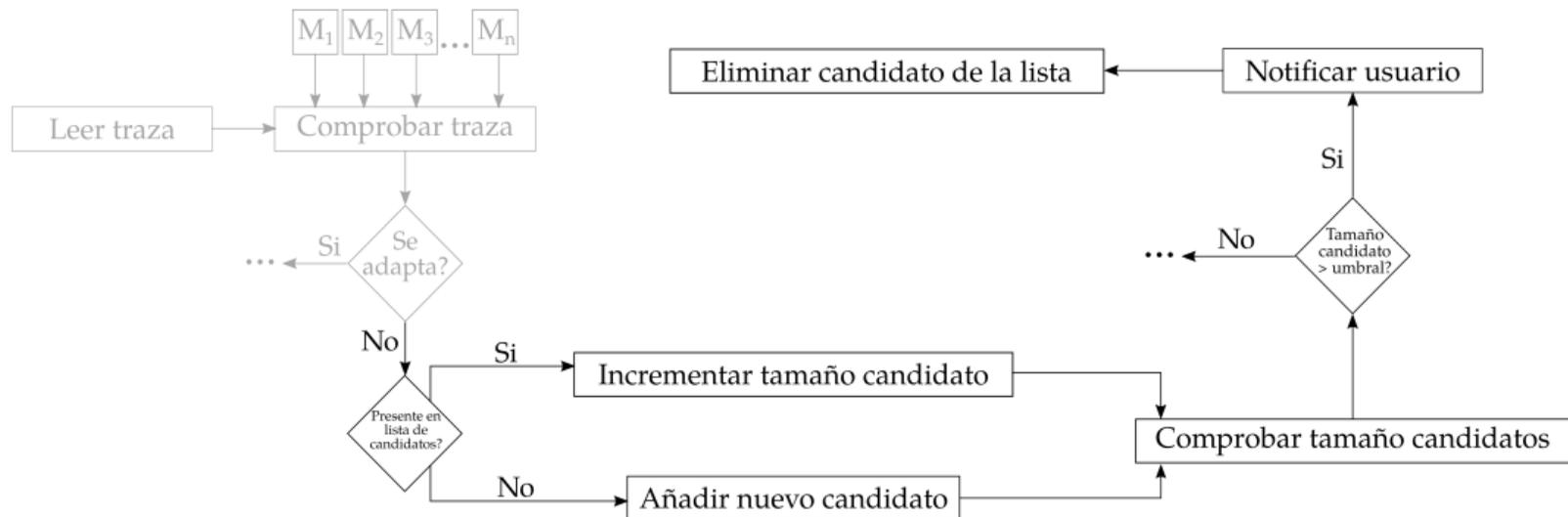
Reejecución de las **trazas** con los **modelos descubiertos previamente**.

Consideramos que una traza es **válida** si se puede **ejecutar completamente** en **al menos uno** de los **modelos** generados.

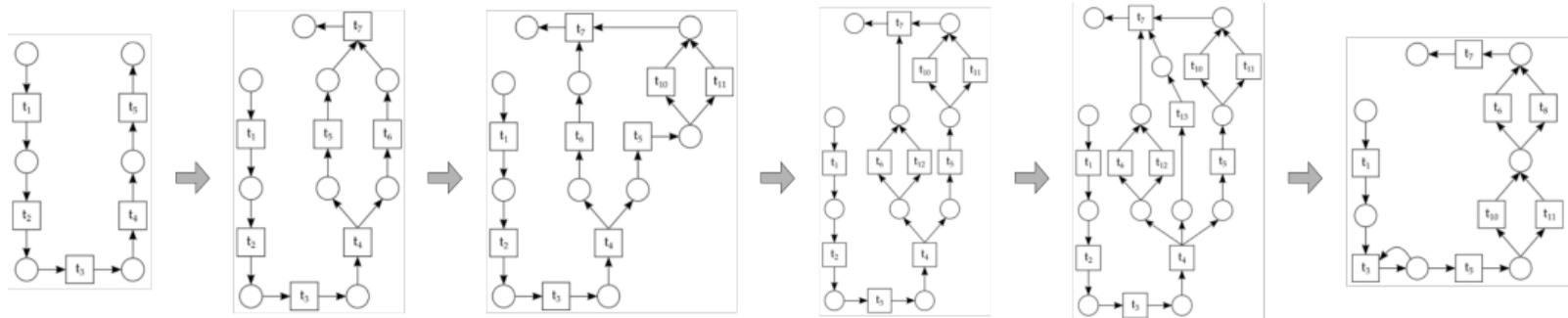
Modelos con **fitness replay** < 1 dan lugar a **falsos positivos!**



Algoritmo - Detección de Concept Drift



Caso de estudio



6 evoluciones. 12000 trazas, 2000 para cada evolución. 12 actividades. 87851 eventos.

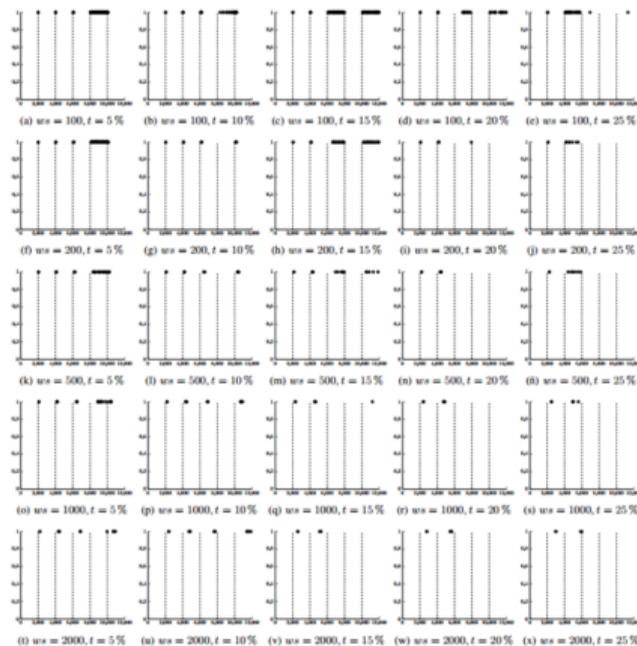
Tamaños de **ventana** de **100, 200, 500, 1000** y **2000** trazas.

Umbrales de detección de **5 %, 10 %, 15 %, 20 %** y **25 %**.

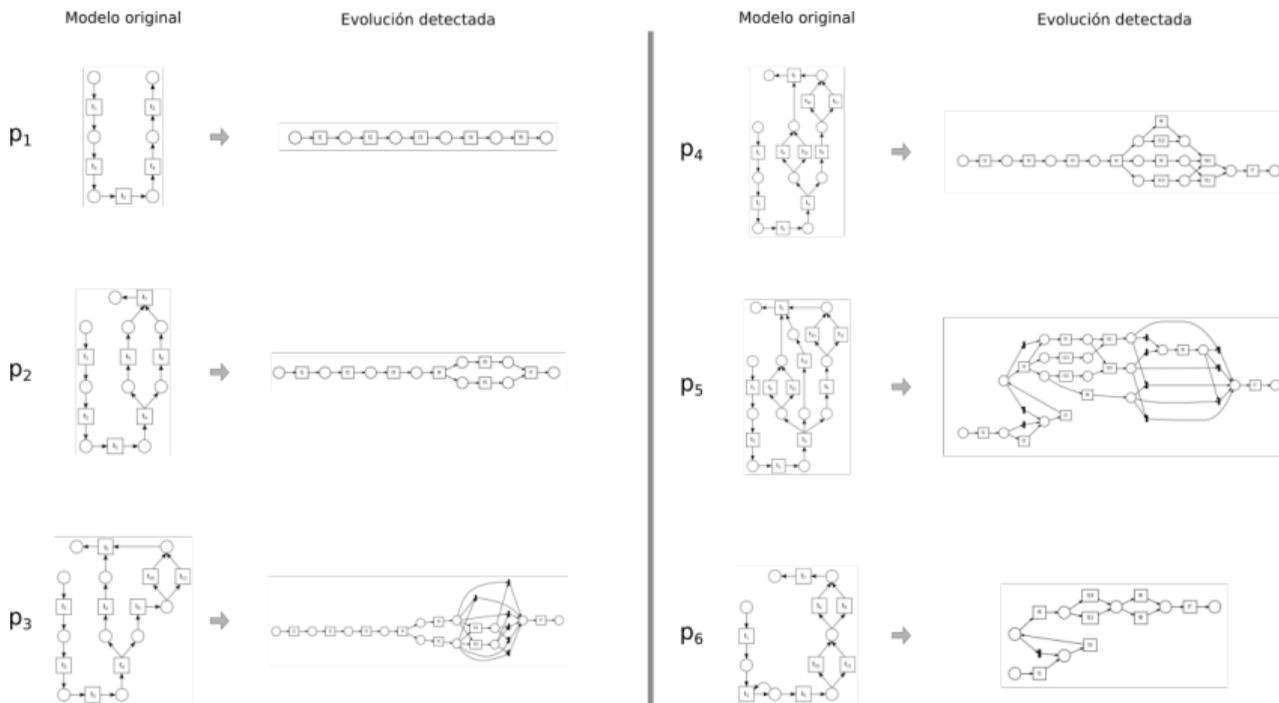
Resultados

Falsos positivos para la cuarta modificación (*fitness replay* de los modelos < 1).

Los **tiempos se disparan** para **tamaños de ventana grandes** debido a la complejidad del clustering y del descubrimiento.



Resultados



Conclusiones

Se ha planteado una **primera aproximación** a la **detección y tratamiento de concept drift en minería de procesos**, la cual proporciona soporte al tratamiento **online** y **offline** de los registros, e incluye capacidades de **detección** y **localización** del cambio.

Se deben **mejorar** los **resultados** para los **registros complejos** en los que los algoritmos de minado no proporcionan un *fitness replay* completo.

Se proporciona una **aproximación inicial** a la **caracterización** del cambio y a la **evolución** del proceso.

No se han podido realizar **pruebas** con registros de ejecución **reales**.

Trabajos futuros

Optimización del clustering para reducir los tiempos de ejecución.

Mejora del **tratamiento de outliers**.

Pruebas con **registros reales**.

Aplicación de **técnicas de Big Data y Streaming** para tratamiento de grandes registros y tratamiento online masivo.

¿Preguntas?

victorjose.gallego@usc.es



UNIÓN EUROPEA



XUNTA
DE GALICIA

galicia

Fondo Europeo de Desenvolvemento Rexional
“Unha maneira de facer Europa”