



TRABAJO FIN DE MÁSTER

MÁSTER DATCOM: CIENCIA DE DATOS

Árboles de clasificación monotónica sobre flujos de datos.

Autor

Carlos Manuel Sequí Sánchez(alumno)

Directores

Salvador García López(tutor)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, septiembre de 2019



ugr

Universidad
de Granada

Árboles de clasificación monotónica sobre flujos de datos.

Autor

Carlos Manuel Sequí Sánchez

Directores

Salvador García López



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E I.A.

Árboles de clasificación monotónica sobre flujos de datos.

Carlos Manuel Sequí Sánchez(alumno)

Palabras clave: Monotonic, data streams, decision tree, Hoeffding tree, classification, MOA

Resumen

Con el fin de trabajar con temas de mi agrado, crear algo novedoso y aprovechar los conocimientos adquiridos a lo largo del máster, decidí aceptar la propuesta de mi tutor Salvador, y realizar un algoritmo mediante el uso de árboles de clasificación con información subyacente del problema basada en restricciones monotónicas sobre flujos de datos.

Monotonic classification trees on data streams.

Carlos Manuel Sequí Sánchez(student)

Keywords: Monotonic, data streams, decision tree, Hoeffding tree, classification, MOA

Abstract

In order to work with topics that I like, create something new and take advantage of the knowledge acquired throughout the master, I decided to accept the proposal of my tutor Salvador, and make an algorithm by using classification trees with underlying information of the problem based on monotonic restrictions on data streams.

Yo, **Carlos Manuel Sequí Sánchez**, alumno de la titulación Máster DATCOM: Ciencia de Datos de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación**, con DNI 20486926K, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Máster en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Carlos Manuel Sequí Sánchez

Granada a 1 de septiembre de 2019.

D. **Salvador García López**(tutor1), Profesor del Departamento de Ciencias de la Computación e I.A. de la Universidad de Granada.

Informa:

Que el presente trabajo, titulado *Árboles de clasificación monotónica sobre flujos de datos.*, ha sido realizado bajo su supervisión por **Carlos Manuel Sequí Sánchez**, y autoriza la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expide y firma el presente informe en Granada a 1 de Septiembre de 2019 .

El director:

Salvador García López

Agradecimientos

Llegado a este punto, agradezco la paciencia e interés puesto en mi aprendizaje a todos los profesores que han formado parte, durante todo este año, de poner en mis manos la semilla de conocimiento que me servirá para lanzarme al mundo profesional, así como a mi tutor Salvador García López, quien se ha encargado de ayudarme y supervisar este TFG. Agradezco a toda mi familia y, con mayor énfasis a mis padres y a mi hermano, el interés y el apoyo que me han ofrecido desde el primer momento, aunque no entiendan del todo las "letras raras" en la pantalla de mi ordenador cuando trabajo, o que no nos enseñen a "hackear" cosas. Por último, agradezco el haber prolongado el contacto en el ámbito académico con los amigos que hice durante el grado, con buena compañía todo ha sido más sencillo, ya sabéis.

Índice general

1. Capítulo 1. Introducción y objetivos	1
2. Capítulo 2. Background en problemas	3
2.1. Data streaming classification	3
2.2. Ordinal and monotonic classification	3
3. Capítulo 3. Background en algoritmos de árboles de decisión	5
3.1. Fundamentos de árboles de decisión	5
3.2. Hoeffding Trees y otros algoritmos de data streaming	5
3.3. Árboles de decisión monotónicos	5
4. Capítulo 4. Propuesta	7
5. Capítulo 5. Software desarrollado y uso	9
6. Capítulo 6. Experimentos	11
6.1. Framework	11
6.2. Resultados	11
6.3. Análisis	11
7. Capítulo 7. Conclusiones y trabajo futuro	13

Índice de figuras

Capítulo 1

Introducción y objetivos

El objetivo principal de este documento, primeramente, es dotar al lector de los conocimientos necesarios sobre las distintas técnicas a utilizar para el desarrollo del algoritmo, con el fin de llegar a entender el propósito de este proyecto.

En este punto se tratarán los temas de clasificación con restricciones monotónicas, la clasificación sobre flujos de datos, y el background correspondiente a la clasificación mediante el uso de árboles de decisión, así como la combinación de algunas de estas técnicas entre sí, tales como la del uso de árboles de clasificación sobre flujos de datos.

Capítulo 2

Background en problemas

2.1. Data streaming classification

2.2. Ordinal and monotonic classification

Capítulo 3

Background en algoritmos de árboles de decisión

- 3.1. Fundamentos de árboles de decisión
- 3.2. Hoeffding Trees y otros algoritmos de data streaming
- 3.3. Árboles de decisión monotónicos

Capítulo 4

Propuesta

abcssssssssssssssssssssssss

Capítulo 5

Software desarrollado y uso

aqui softwareee

Capítulo 6

Experimentos

6.1. Framework

6.2. Resultados

6.3. Análisis

Capítulo 7

Conclusiones y trabajo futuro

Finalmente, una vez explicado el desarrollo completo del proyecto, es el momento de mirar hacia atrás para cerciorarme de las asignaturas del grado cursadas sin las que no hubiera sido capaz de desenvolverme de manera tan efectiva para la realización del proyecto, así como darme cuenta de los conocimientos que he adquirido durante esta etapa de mi carrera y, por último, realizar una valoración de todo ello.

Bibliografía

