

DetECCIÓN, CARACTERIZACIÓN Y USO DE CÚMULOS ABIERTOS EN UN CONTEXTO GALÁCTICO EN UN ENTORNO *Big Data*

Autor: Alfred Castro Ginard

Tesis doctoral dirigida por: Xavier Luri y

Carne Jordi

Centro: Universitat de Barcelona

Fecha de lectura: 30 de abril de 2021

Los cúmulos estelares abiertos son conjuntos de estrellas, ligadas gravitacionalmente, nacidas de la misma nube de gas molecular que tienen posiciones, cinemática, edades y metalicidades similares. Los métodos tradicionales para la búsqueda de cúmulos abiertos se basan en la inspección visual de regiones del cielo buscando sobredensidades posicionales y comprobando que sus estrellas siguen una isócrona en un diagrama de color-magnitud. El segundo lanzamiento de datos de *Gaia*, con más de 1300 millones de estrellas con medidas de paralaje y movimiento propio junto con fotometría en tres bandas, ha catapultado el desarrollo de novedosas técnicas basadas en *machine learning* para automatizar la detección de cúmulos abiertos, usando información tanto astrométrica como fotométrica.

La caracterización de los cúmulos abiertos nos permite conocer propiedades del disco Galáctico tales como la estructura y evolución de los brazos espirales, o la comprobación de estudios sobre evolución estelar, ya que los parámetros astrofísicos de los cúmulos se pueden estimar con una precisión mayor que para estrellas de campo individuales. Por esto, una buena caracterización de la población de cúmulos abiertos en la Vía Láctea es clave para estudios de arqueología Galáctica.

Nuestro objetivo para esta tesis es transformar los métodos clásicos para la detección de diferentes tipos de patrones en datos astronómicos, que mayoritariamente dependen de la inspección visual, a métodos automáticos de minería de datos para extraer información útil de catálogos estelares. También usaremos el resultado de la aplicación de técnicas de *machine learning* a los datos de *Gaia* en un contexto Galáctico más amplio.

Hemos desarrollado una metodología de minería de datos para la búsqueda a ciegas de cúmulos abiertos en el disco Galáctico. Primero, usamos un algoritmo basado en densidad, DBSCAN, para buscar sobredensidades estelares en el espacio astrométrico de cinco dimensiones de *Gaia*. El despliegue del método de detección en un entorno de *Big Data*, en el superordenador *MareNostrum* colocado en el *Barcelona Supercomputing Center*, evita que la búsqueda esté limitada por restricciones computacionales. Después, las sobredensidades detectadas son clasificadas en simples sobredensidades estadísticas o estructuras físicas usando una red neuronal artificial entrenada para reconocer las isócronas que los cúmulos abiertos forman en un diagrama de color-magnitud.

Para toda la población de cúmulos abiertos, estimamos los parámetros astrofísicos tales como edad, distancia y extinción en la línea de visión usando una red neuronal artificial entrenada en cúmulos abiertos bien estudiados. Usamos esta

información, junto con velocidades radiales medias para cada cúmulo, para trazar la estructura espiral de nuestra Galaxia usando *GaussianMixtureModels* para asociar los cúmulos más jóvenes (< 30 My) al brazo espiral donde se han formado. También describimos la evolución de los brazos espirales durante los últimos 80 My para dar nuevas pistas sobre la naturaleza de la estructura espiral de la Vía Láctea.

La automatización del procedimiento de detección de cúmulos abiertos, junto con su despliegue en un entorno de *Big Data*, ha resultado en más de 650 nuevos cúmulos abiertos detectados con este método. Los nuevos cúmulos UBC representan un tercio del censo total de cúmulos abiertos actualmente conocido y es la mayor contribución individual al catálogo de cúmulos abiertos.

Somos capaces de añadir 264 cúmulos abiertos jóvenes a las 84 regiones de formación estelar de alta masa que se han usado tradicionalmente para trazar los brazos espirales, para incrementar el rango en azimut Galáctico donde los brazos espirales están definidos y estimar mejor sus parámetros. Analizando la distribución de cúmulos abiertos a través de los brazos espirales y calculando las diferentes velocidades de estos brazos somos capaces de desfavorecer las clásicas ondas de densidad como el mecanismo principal de formación de los brazos espirales de la Vía Láctea, favoreciendo un comportamiento transitorio.

Con esta tesis hemos mostrado que el uso de *machine learning*, con un tratamiento correcto de los recursos computacionales, tiene un largo viaje por delante en un futuro de la Astronomía dominada por los datos.

Tesis disponible en: <https://www.tesisenred.net/handle/10803/671790#page=1>