

Hacia una sistematización del proceso de detección de Cúmulos Estelares Abiertos

Towards a systematization of the process of detecting Open Star Clusters

Doctoranda:

Gabriela Soledad Elizalde Caviglia

Universidad Tecnológica Nacional. Facultad Regional Concepción del Uruguay. Grupo de Investigación en Inteligencia Computacional y Optimización de Sistemas. Concepción del Uruguay. Argentina
g.elizalde.caviglia@gmail.com

Directoras:

Maria Fabiana Piccoli, Mariela Corti

Resumen

La clasificación de objetos estelares constituye una etapa fundamental en la caracterización de cúmulos estelares y regiones de formación estelar. Los métodos tradicionales se basan en criterios empíricos definidos mediante umbrales sobre atributos fotométricos obtenidos de catálogos; sin embargo, dichos umbrales presentan variaciones entre distintos estudios, lo que introduce incertidumbre y dificulta la identificación consistente de las diferentes clases de objetos. En este trabajo se propone un modelo de clasificación basado en teoría de conjuntos difusos que reformula las reglas de clasificación binarias utilizadas en la literatura mediante funciones de pertenencia y operadores de agregación difusos. El modelo permite representar el grado de compatibilidad de una fuente estelar con cada categoría, reemplazando la decisión binaria por una evaluación gradual, la cual incorpora la incertidumbre inherentes a los criterios de clasificación. La propuesta preserva el conocimiento contenido en los esquemas tradicionales y proporciona un marco flexible para integrar información proveniente de distintos catálogos. Este trabajo constituye una base metodológica para el desarrollo de sistemas de clasificación más robustos y para la posterior incorporación de técnicas de análisis y caracterización de cúmulos estelares.

Palabras clave: Cúmulos Estelares Abiertos; Agrupamiento Espacial; Lógica Difusa; Cómputo de Altas Prestaciones

Abstract

The classification of stellar objects is a fundamental step in the characterization of star clusters and star-forming regions. Traditional methods are based on empirical criteria defined by thresholds on photometric attributes obtained from catalogs; however, these thresholds vary between different studies, which introduces uncertainty and makes it difficult to consistently identify the different classes of objects. This paper proposes a classification model based on fuzzy set theory that reformulates the binary classification rules used in the literature by means of fuzzy membership functions and aggregation operators. The model allows representing the degree of compatibility of a source to a given class and preserves the knowledge contained in traditional schemes and provides a flexible framework for integrating information from different catalogs. This work constitutes a methodological basis for the development of more robust classification systems and for the subsequent incorporation of classification analysis techniques and characterization of star clusters.

Keywords: Open Star Clusters; Spatial Clustering; Fuzzy Logic; High-Performance Computing.

Introducción

Los cúmulos estelares son conjuntos de estrellas surgidas a partir de una misma nube molecular. Estas estrellas poseen una misma historia evolutiva, teniendo edades similares pero diferenciándose por otras propiedades físicas. El estudio de estos cúmulos es crucial para entender el origen y la evolución de nuestra galaxia, la Vía Láctea (Dobbs & Pringle, 2010).

Distintas misiones, como por ejemplo GAIA del National Aeronautics and Space Administration (NASA) de Estados Unidos, han recolectado datos no estructurados provenientes de distintas regiones de nuestra galaxia, entre las que se incluyen valores de frecuencia en múltiple longitudes de onda de objetos espaciales de diversos tipos. Estos datos, son susceptibles a problemas de calidad debido a deficiencias y limitaciones de los instrumentos utilizados para capturarlos, y por interferencias del medio interestelar. (Cutri et al., 2012)

A partir de estos datos, el descubrimiento de cúmulos estelares se realiza mediante un proceso que involucra tareas de filtrado, análisis y clasificación de las fuentes estelares evaluadas. Este proceso, si bien hace uso de algunas herramientas de machine learning, no es un proceso automatizado y en el mismo intervienen múltiples decisiones atravesadas por la subjetividad de los analistas, introduciendo así sesgos que pueden ocasionar resultados inconsistentes.

Por otro lado, los algoritmos de machine learning utilizados son implementados no consideran en su funcionamiento las particularidades de los datos utilizados. Por ejemplo el algoritmo utilizado para el agrupamiento de los objetos estelares candidatos para la conformación de un cúmulo estelar, utiliza funciones de distancia euclidiana sobre datos como coordenadas celestes sin realizar las transformaciones apropiadas, y no incorporan datos de fotométricos al análisis. (Hunt & Reffert, 2023)

Además, la cantidad de datos obliga a los analistas a segmentar el espacio de búsqueda en regiones manejables de manera manual, por lo que resultan necesarias estrategias adecuadas para poder procesar tal cantidad de datos de una forma eficiente.

Por todo lo expuesto, se propone un proceso computacional sistematizado para el descubrimiento de cúmulos estelares abiertos que considere lo antes expuesto. En este artículo se presenta una primera propuesta para solucionar el problema de la falta de consenso en los criterios de clasificación utilizados para la identificación de miembros de cúmulos estelares abiertos, utilizando teoría de conjuntos difusos.

Desarrollo

La detección de cúmulos estelares abiertos se realiza actualmente utilizando datos disponibles en distintos catálogos de acceso libre, contruidos tras misiones de recolección de datos como GAIA. Estos datos constan de atributos astrométricos (paralaje, movimiento propio, etc) de estrellas y otros objetos estelares, así como también datos fotométricos en múltiples longitudes de onda y atributos sobre la calidad de la medición. (Castro-Ginard et al., 2020; Hunt & Reffert, 2023)

El proceso de detección parte de la selección de los datos de regiones pequeñas del espacio celeste para reducir la numerosidad de las bases de datos y volver el proceso manejable de manera manual. Posteriormente, estos datos son procesados para normalizar los atributos y luego se utilizan algoritmos de agrupamiento sobre los datos astrométricos para poder determinar conjuntos de estrellas que forman parte de un mismo cúmulo estelar. Estos datos luego son analizados mediante distintos criterios para descartar falsos positivos y para clasificar a los miembros en distintas categorías, tales como Objeto Estelar Joven Clase I (YSO I), Objeto Estelar Joven Clase II (YSO II), entre otros. (Corti et al., s. f.; Paíz et al., 2025)

Este tipo de caracterización se realiza mediante una clasificación de las fuentes según el grado de enrojecimiento, entre otros detalles, para el cual se requiere de los atributos de la fuente en el espectro del infrarrojo (IR). Existen diversos catálogos que aportan este tipo de información, para los cuáles se asocia un conjunto de criterios que permiten asignar una clasificación del Objeto Estelar. Si bien estos criterios se asocian a condiciones estrictas, las distancias a las cuáles se obtienen las magnitudes fotométricas, el ruido y el mismo medio interestelar aporta un grado de incertidumbre que dificulta la clasificación. Asimismo, algunos analistas difieren en los límites admitidos en estas condiciones, es decir, no se obtiene un único consenso de análisis. (Fischer et al., 2016; Koenig & Leisawitz, 2014)

Uno de los casos estudiados en este trabajo, resulta de las diferencias entre los estudios realizados en la región de Canis Major, y el esquema de clasificación de Koenig y Leisawitz, donde ambos estudios utilizan datos del IR del catálogo AllWISE.

Las dificultades para definir regiones estrictas, así como las diferencias de los límites motivan realizar un esquema de clasificación flexible que se adapte a dichos análisis, por lo que se presenta un esquema de clasificación basado en conjuntos difusos, permitiendo modelar las transiciones graduales y determinar el grado de pertenencia que un objeto presenta en relación con la categoría, con el objetivo de no descartar por completo aquellas fuentes que se encuentran en los límites de las regiones, sino que sean foco de análisis más profundos.

Considerando lo anterior, se plantea como solución computacional un modelo difuso de criterios de clasificación de objetos estelares, esto es reemplazar las condiciones de clasificación CRISP establecidas en Koenig, utilizando teoría de conjuntos difusos para obtener el grado de pertenencia asociada a cada categoría. Con este enfoque se define dichos conjuntos con funciones de pertenencia de la forma: $\mu : U \rightarrow [0, 1]$, donde U representa el universo de discurso asociado al atributo y la función devuelve un valor comprendido entre 0 y 1, representando el grado de pertenencia del elemento conjunto difuso (Klir & Yuan, 1995). De esta forma, se admiten valores que demuestren la compatibilidad de la fuente al conjunto de la categoría.

Se considera entonces una fuente estelar d_i , la cual para clasificarla como un Objeto Estelar Joven Clase I (YSO I), debe cumplir con las reglas CRISP y con el predicado compuesto tal como determina en Koenig.

$$YSO_I(d_i) = f_1(d_i) > 2,0 \wedge f_2(d_i) > 0 \wedge f_3(d_i) > 0 \wedge f_4(d_i) < 4,5$$

de donde $f_1(d_i) = w_{i,2} - w_{i,3}$; $f_2(d_i) = w_{i,1} - 0,58 \times w_{i,2} - 0,42 \times w_{i,3} - 2,2$; $f_3(d_i) = w_{i,1} - 1,46 \times w_{i,2} + 0,46 \times w_{i,3} + 0,9$; $f_4(d_i) = w_{i,2} - w_{i,3}$; y $w_{i,k}$, $1 \leq k \leq 3$, representan las magnitudes fotométricas correspondientes a la fuente estelar d_i .

Siguiendo dichos criterios y considerando los umbrales de las inecuaciones previamente detalladas, se propone utilizar los siguientes valores difusos (Tanaka & Werners, 1997): $\mu_1(f_1(d_i)) = \{(1,8, 0), (2, 1)\}$, $\mu_2(f_2(d_i)) = \{(-0,2, 0), (0, 1)\}$, $\mu_3(f_3(d_i)) = \{(-0,2, 0), (0, 1)\}$, $\mu_4(f_4(d_i)) = \{(4,5, 1), (4,7, 0)\}$. Y luego, para la función de pertenencia a la categoría YSO Clase I, se define la conjunción difusa de los números difusos correspondientes, mediante una T-norma (T) apropiada para el criterio a modelar, siguiendo la forma:

$$\mu_{YSO_I}(d_i) = \mu_1(f_1(d_i)) T \mu_2(f_2(d_i)) T \mu_3(f_3(d_i)) T \mu_4(f_4(d_i))$$

De esta forma las funciones de pertenencia de cada categoría se definen a través de conocimiento experto, considerando criterios establecidos en la literatura, mediante distribuciones observadas de las poblaciones (Tanaka & Werners, 1997), entre otras. Este procedimiento se realiza para cada una de las categorías propuestas, obteniendo así los grados de pertenencia de una fuente estelar respecto a la misma.

Resultados

Una vez planteada la solución a dicha problemática, se buscó determinar la viabilidad mediante la aplicación del método sobre dos investigaciones previas, cuyo objetivo fue la clasificación de miembros de cúmulos estelares para la identificación de objetos YSO Clase I y II. Para aplicar las reglas de clasificación, se codificaron los predicados difusos en Python mediante el uso de la librería Fuzzy Predicates. En ambos casos de prueba, se utilizó la T-norma y S-norma de Lukasiewicz considerando ser menos drástica en la evaluación de la verdad, a comparación con las clásicas Min y Product, siendo más intuitiva en la interpretación (Tanaka & Werners, 1997). En ambos casos se utiliza el catálogo del IR AllWISE, de donde se consideran los atributos de la Tabla 1.

Atributos	Descripción
-----------	-------------

$w_{i\text{sigmpro}}$	Incertidumbre promedio de ajuste de la banda i ($i = 1 \cdots 4$)
$w_{i\text{snr}}$	Relación ruido-senal de la banda i ($i = 1 \cdots 4$)
$w_{i\text{rchi2}}$	Chi2 reducido de la banda i ($i = 1 \cdots 4$)
$w_{i\text{mpro}}$	Magnitud fotométrica de la banda i ($i = 1 \cdots 4$)
J	Magnitud fotométrica de la banda J
H	Magnitud fotométrica banda H
K_s	Magnitud fotométrica banda K_s

Tabla 1: Atributos del catálogo AllWISE utilizados para la caracterización de fuentes estelares

El primer caso de estudio: ESFERA, se realiza con 37 fuentes estelares previamente analizadas siguiendo el conjunto de criterios CRISP para evaluar la identificación de YSOs, tal como se detalla para los datos AllWISE. Al aplicar el modelado de criterios difusos se obtienen 3 fuentes como YSOs Clase II, estos resultados son compatibles con los hallados previamente en la primera parte del análisis de los primeros 5 cúmulos estelares estudiados. Dichos valores se representan en la Tabla 2.

i	Clase CRISP	μ_{ClaseI}	$\mu_{ClaseII}$
1	Clase II	0	1
2	-	0	0
3	-	0	0
4	-	0	0
⋮	⋮	⋮	⋮
30	Clase II	0	1
31	-	0	0
32	Clase II	0	1
⋮	⋮	⋮	⋮

Tabla 2: Clasificación de objetos estelares d_i para el caso de estudio ESFERA

Para la evaluación del segundo caso de estudio, se analiza la misma región planteada en x, tomando como centro de coordenadas en AllWISE a Canis Major con un radio de 2° debido a las limitaciones del catálogo. Con las fuentes obtenidas se realiza la categorización de 205 fuentes estelares, siendo estas un subconjunto del original. Se aplica la configuración previamente mencionada y se acotan los atributos, a los asociados a la magnitud fotométrica, incluyendo $w_{i\text{mpro}}$, J, H, K_s . Los resultados se corresponden con los realizados en Fischer, y se logra observar particulares en los valores obtenidos, referidos en la Tabla 3, donde al aplicar las reglas difusas, se permite observar fuentes que al encontrarse en los límites de corte, obtienen grados de pertenencia en ambas categorías: Clase I y Clase II, permitiendo ahondar en la información de dicha fuente, previo a realizar una categorización “fija”.

i	Clase CRISP	μ_{ClaseI}	$\mu_{ClaseII}$
-----	-------------	----------------	-----------------

1	Clase II	0	1
2	Clase I	1	0
.	.	.	.
.	.	.	.
.	.	.	.
11	Clase I	1	0.275
12	Clase II	0	1
13	Clase II	0	1
14	Clase II	0	1
15	Clase II	0.6833	1
.	.	.	.
.	.	.	.
.	.	.	.

Tabla 3: Clasificación de objetos estelares d_i para el caso de estudio Canis Major

Conclusiones

El modelo propuesto permite reemplazar las decisiones binarias por funciones difusas de pertenencia, permitiendo de esta forma representar el grado de compatibilidad entre la fuente estelar y la categoría asignada, este enfoque permite flexibilizar la caracterización, incorporando de forma explícita la incertidumbre asociada en los umbrales, permitiendo obtener información sobre aquellas fuentes que se encuentran en los límites de corte, permitiendo de esta forma, investigar casos particulares en profundidad en lugar de descartarlos o dar una clasificación “fija”.

En cuanto a las líneas futuras líneas de trabajo se encuentra aplicar este modelo para distintos catálogos, permitiendo integrar la información previo a las categorizaciones finales. Asimismo, se busca extender el enfoque, considerando el proceso de análisis de cúmulos abiertos, esto es, incorporar algoritmos de agrupamiento que incluyan información astrométrica y fotométrica, así como definir metodologías que permitan estimar la edad de los cúmulos bajo el ajuste de isocronas.

Referencias

- Castro-Ginard, A., Jordi, C., Luri, X., Cid-Fuentes, J. Á., Casamiquela, L., Anders, F., Cantat-Gaudin, T., Monguió, M., Balaguer-Nuñez, L., Solà, S., & others. (2020). Hunting for open clusters in Gaia DR2: 582 new open clusters in the Galactic disc. *Astronomy & Astrophysics*, 635, A45.
- Corti, M. A., Rizzo, L., Paíz, L. G., & Caviglia, G. S. E. (s. f.). Multiwavelength study of the Carina–Sagittarius Arm—I. Astrometric and photometric search for new open clusters in the $320^\circ \leq l \leq 325^\circ$ region | *Astronomy & Astrophysics (A&A)*. *Astronomy & Astrophysics*, 712, A18.

Cutri, R. M., Wright, E. L., Conrow, T., Bauer, J., Benford, D., Brandenburg, H., Dailey, J., Eisenhardt, P. R. M., Evans, T., Fajardo-Acosta, S., Fowler, J., Gelino, C., Grillmair, C., Harbut, M., Hoffman, D., Jarrett, T., Kirkpatrick, J. D., Leisawitz, D., Liu, W., ... Wittman, M. (2012, marzo). *Explanatory supplement to the WISE all-sky data release products* (p. 1).

Dobbs, C. L., & Pringle, J. E. (2010). Age distributions of star clusters in spiral and barred galaxies as a test for theories of spiral structure. *Monthly Notices of the Royal Astronomical Society*, 409(1), 396-404.

<https://doi.org/10.1111/j.1365-2966.2010.17323.x>

Fischer, W. J., Padgett, D. L., Stapelfeldt, K. L., & Sewilo, M. (2016). A WISE CENSUS OF YOUNG STELLAR OBJECTS IN CANIS MAJOR. *The Astrophysical Journal*, 827(2), 96.

<https://doi.org/10.3847/0004-637X/827/2/96>

Hunt, E. L., & Reffert, S. (2023). Improving the open cluster census. II. An all-sky cluster catalogue with Gaia DR3. *Astronomy & Astrophysics*, 673, A114. <https://doi.org/10.1051/0004-6361/202346285>

Klir, G., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic* (Vol. 4). Prentice hall New Jersey.

Koenig, X. P., & Leisawitz, D. T. (2014). A CLASSIFICATION SCHEME for YOUNG STELLAR OBJECTS USING the WIDE-FIELD INFRARED SURVEY EXPLORER ALLWISE CATALOG: REVEALING LOW-DENSITY STAR FORMATION in the OUTER GALAXY. *The Astrophysical Journal*, 791(2), 131. <https://doi.org/10.1088/0004-637X/791/2/131>

Paíz, L. G., De Biasi, M., Orellana, R. B., & Corti, M. A. (2025). A New Open Cluster Census in the Region of Cygnus OB2 with Gaia Data. *International Journal of Astronomy and Astrophysics*, 15(2), 171.

<https://doi.org/10.4236/ijaa.2025.152013>

Tanaka, K., & Werners, B. (1997). *An introduction to fuzzy logic for practical applications* (Vol. 1). Springer.