

COMUNICACIÓN DE TRABAJO – CONTRIBUTED PAPER

Be o no Be. Esa es la cuestión

Y. J. Aidelman^{1,2}, F. Ronchetti^{3,4}, F. Quiroga³, A. Granada^{5,6}, C. Escudero², L. Lanzarini³

(1) *Facultad de Ciencias Astronómicas Y Geofísicas - UNLP*

(2) *Instituto de Astrofísica La Plata - CONICET-UNLP*

(3) *Instituto de Investigación en Informática LIDI, Facultad de Informática - UNLP*

(4) *Comisión de Investigaciones Científicas de la Pcia. De Bs. As. - CIC-PBA*

(5) *Universidad Nacional de Río Negro, Sede Andina - CITECCA*

(6) *Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina*

Resumen. Ante la gran cantidad de datos astronómicos disponibles, resulta necesario implementar nuevas estrategias para su análisis. Por ello, las técnicas de aprendizaje automático, en particular las que consisten en aprendizaje supervisado, resultan una herramienta muy útil para tal fin. Para la implementación de estas estrategias es fundamental contar con un conjunto de datos (*data set*) etiquetado de manera confiable con el cual entrenar y evaluar el modelo. En este trabajo presentamos un nuevo *data set* compuesto por las magnitudes u , g , r , $H\alpha$, i , J , H , K , $W1$ y $W2$ para 3 365 790 estrellas de las cuales 3 374 fueron clasificadas en la literatura como Be, 13 214 clasificadas como de tipo espectral B y 2 948 600 con la presencia de emisión en $H\alpha$. De estas últimas, solamente 2 758 estrellas fueron reportadas como Be.

Abstract. Given the large amount of astronomical data available, it is necessary to implement new strategies for its analysis. For this reason, machine learning techniques, particularly those that consist of supervised learning, are a fundamental tool for this purpose. For the implementation of these strategies it is essential to have a reliably labeled data set with which to train and evaluate the model. In this work we present a new data set composed of the magnitudes u , g , r , $H\alpha$, i , J , H , K , $W1$ and $W2$ for 3 365 790 stars, of which 3 374 were classified in the literature as Be, 13 214 classified as spectral type B and 2 948 600 with the presence of emission in $H\alpha$. Of the latter, only 2 758 stars were reported as Be.

1. Introducción

En el último tiempo, se han incrementado notablemente las bases de datos y los relevamientos del cielo. Es por ello que, ante la gran cantidad de datos disponibles, resulta necesario implementar nuevas estrategias para poder manejarlos y analizarlos. En este contexto, las técnicas de Aprendizaje Automático (Chollet, 2017), como son las Redes Neuronales (Goodfellow et al., 2016), permiten

operar un volumen significativo de información, resultando ser una herramienta fundamental para el análisis y procesamiento de los mismos. En particular, las técnicas de aprendizaje automático supervisado consisten en deducir una función (modelo) a partir de un conjunto de datos “etiquetados”, para luego poder utilizarlo sobre otro conjunto de datos sin etiquetar. Es por ello que, para garantizar que el modelo **funcione en forma apropiada** es necesario contar con un conjunto de datos adecuadamente etiquetado.

En este trabajo presentamos una nueva base de datos construida a partir de la información obtenida de relevamientos fotométricos y espectroscópicos, de diversos catálogos de estrellas Be, y demás datos disponibles en la literatura. La concatenación de toda esta información resultó en un *data set* compuesto de 3 365 790 estrellas de tipos espectrales OBA y estrellas con $H\alpha$ en emisión como ser Be, B[e] y Herbig AeBe (HAeBe). Este conjunto de datos está siendo utilizado para entrenar y mejorar una red neuronal que hemos confeccionado con el fin de generar una herramienta que permita, a través de datos fotométricos, identificar de forma rápida y fácil **estrellas candidatas a presentar el fenómeno Be** (ver Aidelman et al., 2020, 2021). Caracterizar mejor la población de estrellas Be en distintos ambientes, es indispensable para comprender los procesos involucrados en la formación de estos objetos en rotación rápida.

2. Metodología

Para construir este *data set* se buscó en la literatura, y en diversas bases de datos, estrellas que hayan sido confirmadas como Be. Es decir, clasificadas como estrellas de tipo espectral B no supergigantes, y se haya registrado alguna vez la línea $H\alpha$ en emisión (Jaschek et al., 1981; Collins 1987). Además, el *data set* se completó con otro tipo de objetos, como estrellas de tipos espectrales O, B y A y demás objetos que también emiten en $H\alpha$ ¹ (p. ej. B[e] y HAeBe).

Las magnitudes aparentes de los objetos se obtuvieron de los relevamientos: VPHAS (en el hemisferio sur), IGAPS (en el hemisferio norte), 2MASS y AllWISE. De este modo, contamos con magnitudes aparentes en los filtros u , g , r , $H\alpha$, i , J , H , K , $W1$ y $W2$, lo que permite cubrir el rango del espectro electromagnético entre el óptico y el infrarrojo cercano. La elección de estos filtros se basa en las características que presentan las estrellas Be (ver Rivinius et al., 2013; Rímulo et al., 2018; McSwain & Gies, 2005a; Mohr-Smith et al., 2015; Raddi et al., 2015; Granada et al., 2018; Aidelman et al., 2021) **por ejemplo, el índice $(u - g)$ da una estimación del salto de Balmer, la magnitud en $H\alpha$ da cuenta de si existe o no emisión en dicha línea y con los filtros infrarrojos se pueden detectar excesos de color anómalos debidos a la presencia de la envoltura circunestelar.**

¹FratTA et al. (2021); Zari et al. (2021); Cochetti et al. (2020); Vioque et al. (2020); Chen et al. (2019); Liu et al. (2019); Aidelman et al. (2018, 2015, 2012); Mohr-Smith et al. (2017, 2015); Hou et al. (2016); Raddi et al. (2015); Huang et al. (2010); McSwain et al. (2009); Mathew et al. (2008); Witham et al. (2008); McSwain & Gies (2005b); Jaschek & Egret (1982); Zorec et al. (2005, 2009), *Be Star Observation Survey* (BeSOS, <http://besos.ifa.uv.cl/#/>), *Be Star Spectra* (BeSS, <http://basebe.obspm.fr/basebe/>), *COncvection ROtation et Transits planétaires* (CoRoT, <http://www.astrosurf.com/buil/us/becat.htm>) y *Large Sky Area Multi-Object Fibre Spectroscopic Telescope* (LAMOST, <http://dr6.lamost.org/>) v6 y SIMBAD.

3. Dataset

Toda la información antes mencionada, se organizó en una tabla que contiene la siguiente información: identificación (col. 1: ID), coordenadas (cols. 2 y 3: RA y DEC J2000.0), tipo espectral publicado (col. 4: SpT), tipo de objeto (col. 5: type: Be, B[e], BSG, etc), emisión en $H\alpha$ reportada (col. 6: EM), si fue reportada como Be (col. 7: Be), referencias (col. 8: Ref), magnitudes (cols. 9-18: u , g , r , $H\alpha$, i , J , H , K , $W1$ y $W2$), probabilidad de ser objeto puntual según VPHAS o IGAPS (col. 19: PStar). Luego, agregamos información de LAMOST DR6 V2 General Catalog: identificación (col. 20: obsid), tipo de objeto (col. 21: OBJType), Tipo de espectro (galaxia o estrella, col 22: Class) y clasificación espectral (col. 23: Subclass).

La correspondencia cruzada de los objetos con los distintos relevamientos se realizó con el programa TOPCAT². Con el objetivo de evitar resultados duplicados se utilizó un radio de búsqueda de 0,2 arcsec. Se eliminaron las estrellas que no tenían la información completa en las magnitudes. De este modo conseguimos una tabla con un total de 3 365 790 objetos puntuales de los cuales 3374 (0,1 %) son estrellas Be confirmadas en la literatura (col. Be = 1).

4. Análisis preliminar

A partir de todos los datos colectados realizamos un análisis preliminar con el objetivo de etiquetar estrellas candidatas a Be. A continuación se describen los criterios utilizados para hallarlas.

1) Estrellas B: se etiquetaron como estrellas de tipo espectral B (col.26: B-TS = 1) a aquellas que fueron clasificadas en la literatura (col. 24: B-TS1) o por LAMOST (col. 25: B-TS2) con este tipo espectral. De este modo, contamos un total de 13 214 (0,4 %) estrellas clasificadas, al menos una vez, como de tipo espectral B, un total de 522 189 (15,5 %) estrellas clasificadas con tipo espectral diferente al B. El resto no tiene clasificación espectral (84 %).

2) Estrellas con emisión en $H\alpha$: se etiquetaron como estrellas con la línea $H\alpha$ en emisión (col.28: EMobj = 1) a aquellas estrellas que fueron reportadas como EM en la literatura o por LAMOST (col. 27: EM1). De esta forma, contamos un total de 2 948 600 (87,6 %) objetos con emisión en $H\alpha$, de los cuales sólo 2758 (0,09 %) fueron reportados como Be (col. Be = 1). De modo que contamos con un total de 2 945 842 estrellas EM que no fueron reportadas como Be.

3) Primera selección de candidatas a Be: se seleccionaron aquellas estrellas etiquetadas como de tipo espectral B (col. B-TS = 1) que no sean supergigantes, que no presenten fenómeno Be (col. Be $\neq$ 1) pero que tengan emisión en $H\alpha$ (col. EMobj = 1). En total encontramos 1991 candidatas a Be (col. 29: BeC1 = 1).

4) Estrellas con exceso de color ($r - H\alpha$): se etiquetaron aquellas estrellas que tienen exceso en $H\alpha$ (col. 30: EM2) de acuerdo al diagrama $Q_{ugrH\alpha}$ vs. Q_{ugri} (DQQ, Fig. 1). De este modo, encontramos 1239 estrellas reportadas como Be que tienen exceso en $H\alpha$ (col. 31: Be_EM). Cabe mencionar, que en el DQQ se

²<http://www.star.bris.ac.uk/~mbt/topcat/>

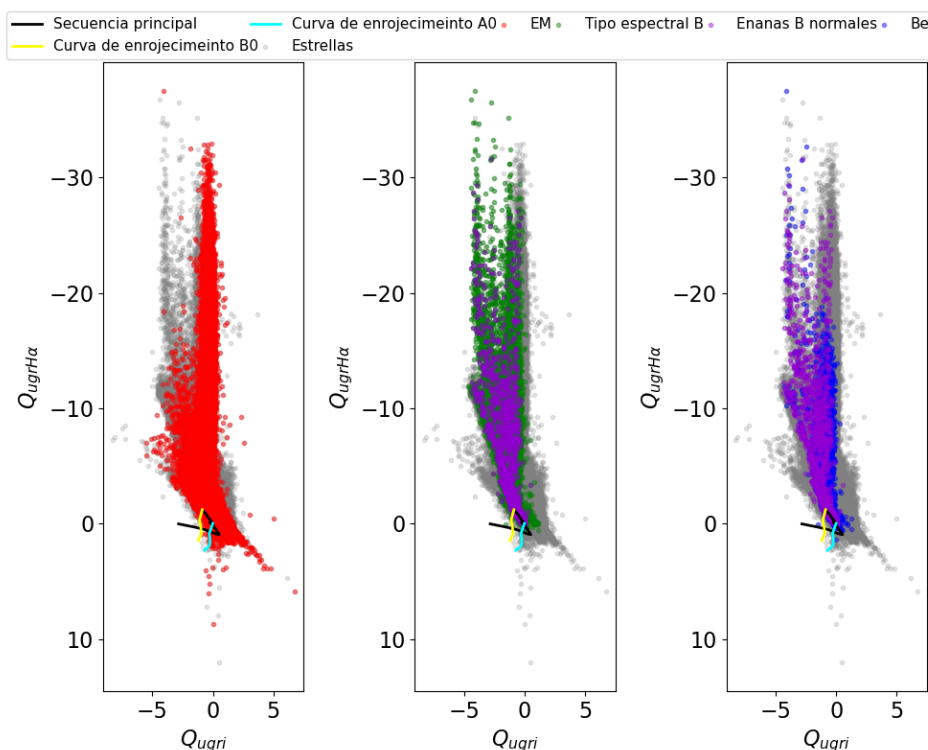


Figura 1. Diagramas QQ: A modo de ejemplo se grafican algunos diagramas QQ. El subíndice de cada eje indica las magnitudes que se utilizaron para calcular el índice $Q = (m_1 - m_2) - (r_1 - r_2)(m_3 - m_4)/(r_3 - r_4)$. Los valores de los coeficientes r_i son los calculados por Yuan et al. (2013).

observa cómo las estrellas con emisión en $H\alpha$ se desparraman en sentido opuesto al enrojecimiento debido al medio interestelar.

5) Segunda selección de candidatas a Be: se seleccionaron estrellas de tipo espectral B (col. B-TS = 1) no Be (col. Be $\neq$ 1) y que tengan exceso de color en $H\alpha$ (de acuerdo al punto 4). Nuevamente se eliminaron aquellas estrellas clasificadas como supergigantes. En total contamos 5954 estrellas que cumplen estas condiciones (col 32: BeC2).

6) Candidatas a Be: Sumando las candidatas del punto 3 y las del punto 5 contamos un total de 7204 (0,2%) estrellas candidatas a Be (col. 33: BeC = 1). Los experimentos realizados previamente (Aidelman et al., 2020 y 2021) dejaron en claro que la estrategia propuesta es adecuada para identificar estrellas candidatas a Be. Con este nuevo dataset, al ser significativamente mayor y conformado por objetos en regiones bien diferentes del cielo, podremos realizar un nuevo entrenamiento de la red neuronal con el objetivo de mejorar los resultados obtenidos para la completitud y la pureza del modelo.

Agradecimientos. Este proyecto ha recibido financiación dentro del marco del Programa de Investigación e Innovación Horizonte 2020 (2014-2020) de la Unión Europea en virtud del Acuerdo de subvención Marie Skłodowska-Curie No. 823734 y de CONICET (PIP 1337). A.G. agradece el apoyo financiero de ANPCyT - PICT 2017-3790 y UNRN - PI2020-40-B-890.

Referencias

- Aidelman Y., et al., 2012, *A&A*, 544, A64
Aidelman Y., et al., 2015, *A&A*, 577, A45
Aidelman Y., et al., 2018, *A&A*, 610, A30
Aidelman Y., et al., 2020, CCIS book series, 1291, 111
Aidelman Y., et al., 2021, BAAA, Argentina, 62, 62
Chen B. Q., et al., 2019, *MNRAS*, 487, 1400
Chollet, Francois, 2017, "Deep Learning with Python", New York, Manning Publications
Cochetti Y.R., et al., 2020, *A&A*, 634, A18
Collins George W. I., 1987, A. Slettebak, T.P. Snow (Eds.), IAU Colloq. 92: Physics of Be Stars, 3
Fratta M., et al., 2021, *MNRAS*, 505, 1135
Goodfellow I., Bengio Y., Courville A., 2016, "Deep Learning", MIT Press
Granada A., et al., 2018, *AJ*, 155, 50
Hou W., et al., 2016, *Research in Astronomy and Astrophysics*, 16, 138
Huang W., Gies D.R., McSwain M.V., 2010, *ApJ*, 722, 605
Jaschek M., Egret D., 1982, Jaschek & Groth (Eds.), *Be Stars*, vol. 98, 261
Jaschek M., Slettebak A., Jaschek C., 1981, *Be Star Newsletter*, 4, 9
Liu Z., et al., 2019, *ApJS*, 241, 32
Mathew B., Subramaniam A., Bhatt B.C., 2008, *MNRAS*, 388, 1879
McSwain M.V., Gies D.R., 2005a, *ApJ*, 622, 1052
McSwain M.V., Gies D.R., 2005b, *ApJS*, 161, 118
McSwain M.V., Huang W., Gies D.R., 2009, *ApJ*, 700, 1216
Mohr-Smith M., et al., 2015, *MNRAS*, 450, 3855
Mohr-Smith M., et al., 2017, *MNRAS*, 465, 1807
Raddi R., et al., 2015, *MNRAS*, 446, 274
Rímulo L.R., et al., 2018, *MNRAS*, 476, 3555
Rivinius T., Carciofi A.C., Martayan C., 2013, *A&A Rev.*, 21, 69
Vioque M., et al., 2020, *A&A*, 638, A21
Witham A.R., et al., 2008, *MNRAS*, 384, 1277
Zari E., et al., 2021, *A&A*, 650, A112
Zorec J., Frémat Y., Cidale L., 2005, *A&A*, 441, 235
Zorec J., et al., 2009, *A&A*, 501, 297