

Carpyncho - Utilidades de Machine Learning

Juan B. Cabral



Jornada NOVA en Córdoba

21 de Agosto del 2015



About Me

- Ingeniero en Sistemas egresado de la UTN-FRC
- Doctorando en Informática en la *Universidad Nacional de Rosario* y el *Instituto de Astronomía Teórico y Experimental (IATE)*
- El título de mi plan de trabajo:

Análisis y diseño de procesos de minería de datos astrofísicos sobre catálogos fotométricos múltiple época

Directores:

Dr. Pablo Granitto (CIFASIS-CONICET)

Dr. Sebastian Gurovich (IATE-OAC-UNC)



Objetivos de Investigación

- Mejorar usos de los datos.
- Integrar Aprendizaje automático al proceso de análisis de catálogos astronómicos.
- Hacer "crecer" los catálogos integrando datos de diferentes experimentos.
- Facilitar la realización de experimentos sobre los datos.
- Hacer transparente la distribución de datos, algoritmos y resultados.
- En resumen: Colaborar en el "Data Driven Science"

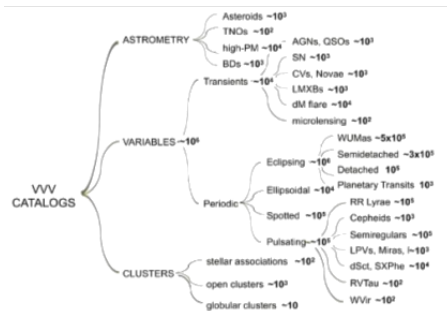
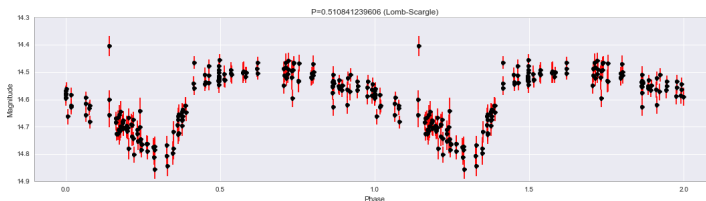


Diagrama adaptado de Minniti et al 2010 que expone el numero esperado de fenómenos astrofísicos que espera detectar VVV en sus catálogos

Que es Carpyncho Actualmente

- Es una herramienta en Python (<http://carpyncho.jbcabral.org/>)
- Es una base de datos que alberga tiles del VVV en un formato util para el análisis.
- Tiene la capacidad de consolidar los datos de los pawprints con los tiles por proximidad (matching por proximidad).
- Calcula estadísticas de las consolidaciones.
- Tiene implementado una primera versión de **cone-search** (con capacidad de paginación automática)
- Calcula estadísticas de magnitud automáticamente.
- Calcula periodos (LS y PDM), componentes de Fourier para las fuentes que se deseen.
- Soporta análisis de "estabilidad" de cálculos de periodos y componentes Fourier a través de monte-carlo.



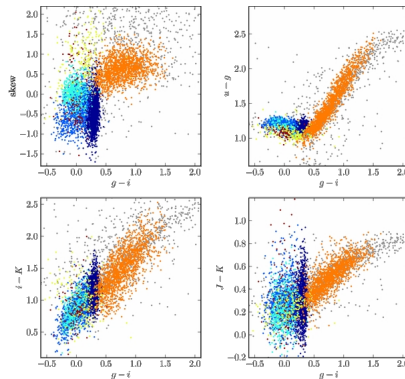
A Futuro: Búsquedas

```
>>> import carpyncho as pyncho
>>> conn = pyncho.Connection()
>>> sources = conn.cone_search(ra, dec, sr)
>>> list(sources)
[source_1, source_2, ...]
>>> src = sources[0] # la primera
>>> src.tags
["rrlyrae_a"]
>>> src.ra
0.1
>>> src.magnitude.std
0.32
>>> src.mjd
[0,1,2,3...]
>>> src.magnitudes
[3,4,5,6...]
```

```
>>> sources = conn.cone_search(ra, dec, sr)
>>> sources.filter(magnitude.average > 23).exclude(sources.ra < 32)
```

A Futuro: Clasificaciones

```
>>> sources = conn.filter(magnitude in [1, 2, 3])
>>> test, training = sources.random_split(3)
>>> model = pyncho.RandomForest(training, test, bla, bla, bla)
>>> model.cmatrix # confusión matrix
>>> model.score(test) # clasificación
```



Adaptado de Lovro Palaversa et al. 2013

A Futuro: Réplicas y Pedidos de integración

```
>>> mysql = pyncho.Storage(  
    "mysql://usr:passwd@localhost:3306/db_name")  
>>> csv = pyncho.Storage("csv:///archivo.csv")  
>>> replica_0 = mysql.write(sources)  
>>> replica_1 = csv.write(sources)  
>>>  
>>> mysql.close()  
>>> csv.close()  
  
# esto ya no se conecta remoto pero tiene los mismos datos  
>>> replica_0.cone_search(...).filter(...)
```

```
>>> sources_raras  
>>> estadisticas_de_sources_raras  
>>> conn.data_integration_request(  
    sources_raras, estadisticas_de_sources_raras, tag="la papa",  
    message="integren esto por que es la verdad de la milanese")  
# en un futuro  
>>> sources = conn.filter(tag.name=="la papa")
```

Mas a Futuro: Aun sin planear

- Integrar selección de features.
- Integrar visores multidimensionales.
- Publicar una acceso directo a la DB subyacentes.
- Permitir ejecutar experimentos en remoto.
- Análisis multicriterio.
- Gráficos!.



¿Preguntas?

- Slides: <https://goo.gl/o643pG>
- **Contacto:**
 - Homepage: jbcabral.org
 - Email: jbc.develop@gmail.com



Referencias

- Minniti, D., Lucas, P. W., Emerson, J. P., Saito, R. K., Hempel, M., Pietrukowicz, P., ... & Padilla, N. (2010). VISTA Variables in the Via Lactea (VVV): The public ESO near-IR variability survey of the Milky Way. *New Astronomy*, 15(5), 433-443.
- Palaversa, L., Ivezić, Ž., Eyer, L., Ruždjak, D., Sudar, D., Galin, M., ... & Oluseyi, H. (2013). Exploring the Variable Sky with LINEAR. III. Classification of Periodic Light Curves. *The Astronomical Journal*, 146(4), 101.