

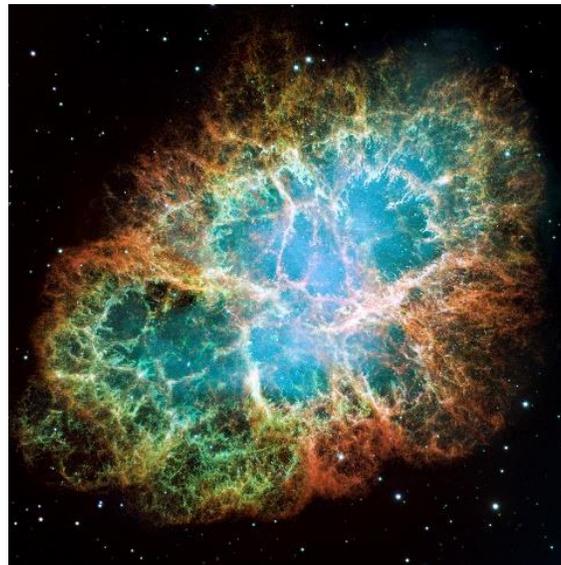
UNIVERSITA' DEGLI STUDI DI NAPOLI
FEDERICO II



SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
AREA DIDATTICA DI SCIENZE MM. FF. NN.
Dipartimento di Fisica

Corso di Laurea in Astrofisica e Scienze dello Spazio
A.A. 2013/2014

Classification of transients with machine learning methods



Candidato
Antonio D'Isanto N91/16

Relatori
Ch.mo Prof. Giuseppe Longo
Dr. Massimo Brescia
Dr. Stefano Cavuoti

Data burst

- In molti campi scientifici, una nuova generazione di strumenti sta generando una crescita esponenziale della mole di dati disponibili, rendendo possibili nuove importanti scoperte.
- In molti casi, è pressante la richiesta di un'analisi rapida ed efficace che, assieme alla grande quantità di dati, implica la **necessità di una rilevazione automatizzata dei fenomeni, analisi e classificazione.**



Nuovi strumenti

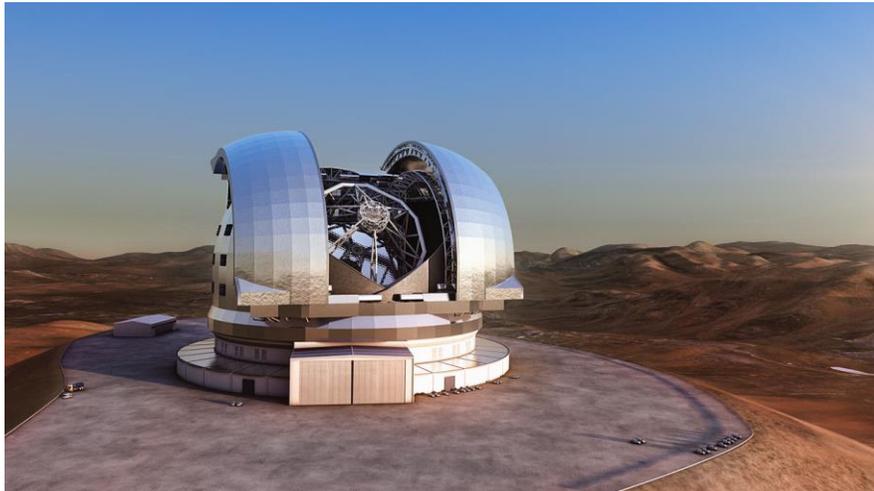
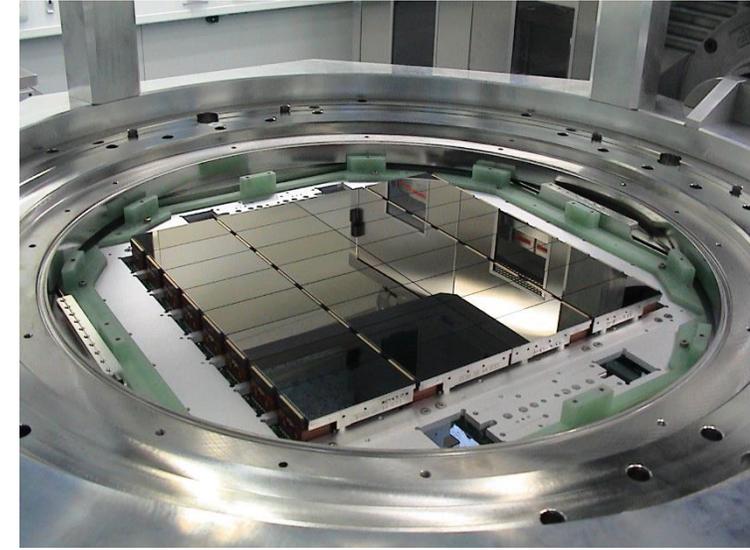
LSST



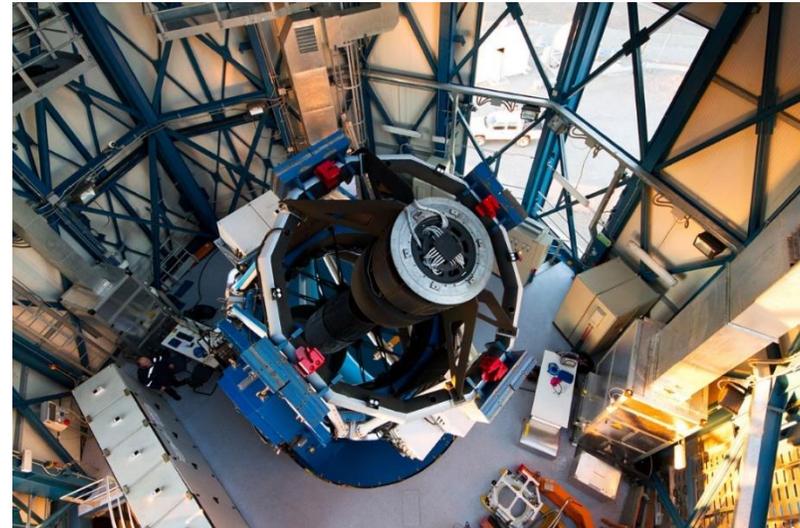
SKA



Omega-Cam



ELT



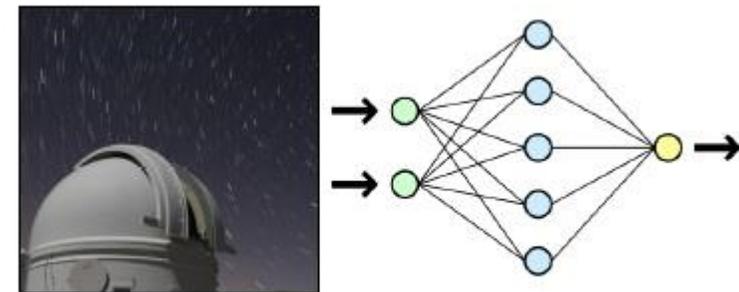
VST

Nuove frontiere della ricerca

- Survey sinottiche (multibanda, multiepoca)
- Time Domain Astronomy
- Nuova scienza ma anche grandi problemi tecnologici e computazionali

LSST: produrrà circa 10^7 transienti/notte

Rivelazione, classificazione real-time per definizione di follow-up spettro-fotometrici



Time Domain Astronomy – Un settore in grande fermento...



Astronomy Data Mining in Action

Exploring the Time Domain

Mega-Flares on normal Sun-like stars = a star like our Sun increased in brightness 300X one night! ... say what??

| | |
|---------------|---------------|
| 1988.3697 | 1988.4487 |
| 1991.2723 | 1994.3679 |
| 1990.1793 | 1997.3408 |

2008 NV

Time Domain Astronomy with the Liverpool Telescope

Chris Davis

<http://telescope.livjm.ac.uk/>
<http://www.facebook.com/liverpooltelescope>

Multi-Messenger Time Domain Astronomy

Time Domain Astrophysics with Swi

Clemson University - October 24-26, 2011

HOT WIRING THE TRANSIENT UNIVERSE 2

Santa Cruz, California • April 28-April 30, 2009

the eventful universe

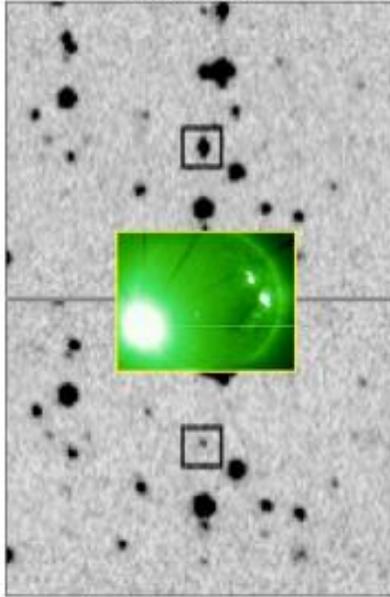
Observing Transient and Variable Phenomena in the Universe from Planetary Transits to Gamma-Ray Bursts

March 17-20 2010 Tucson Arizona
www.noao.edu/meetings/eventful-universe/

Scoprire i transienti

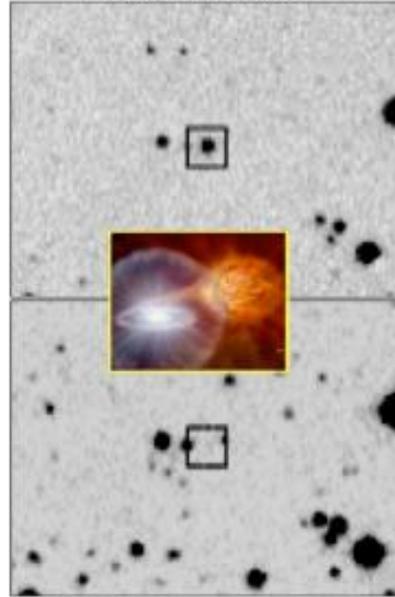
CSS090429:135125-075714

Flare star



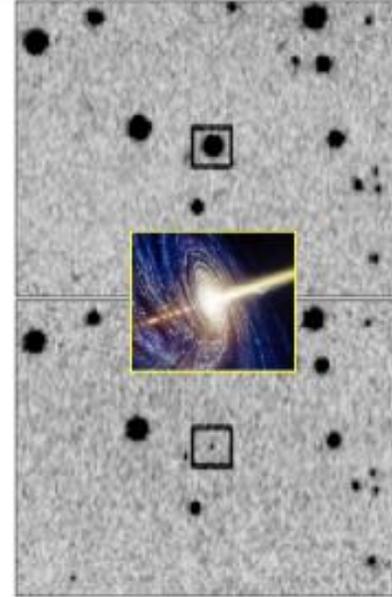
CSS090430:095623-093615

Dwarf Nova



CSS090426:074240+544425

Blazar, 2EG J0744+5438



Appena scoperti, tutti i transienti appaiono uguali, ma in realtà possono rappresentare una vasta gamma di differenti fenomeni fisici.

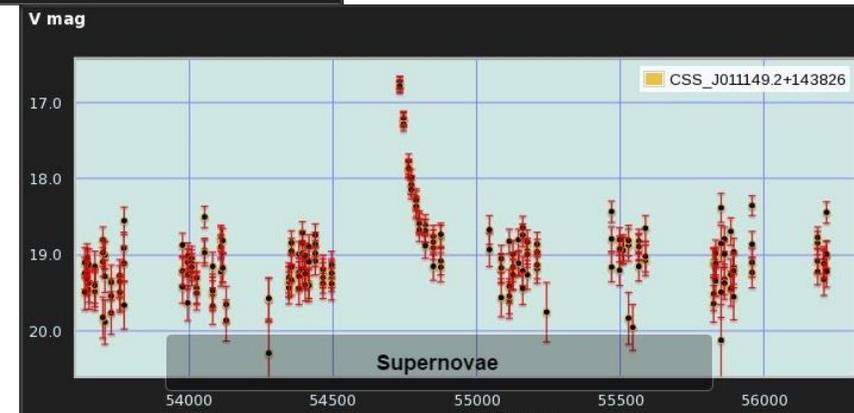
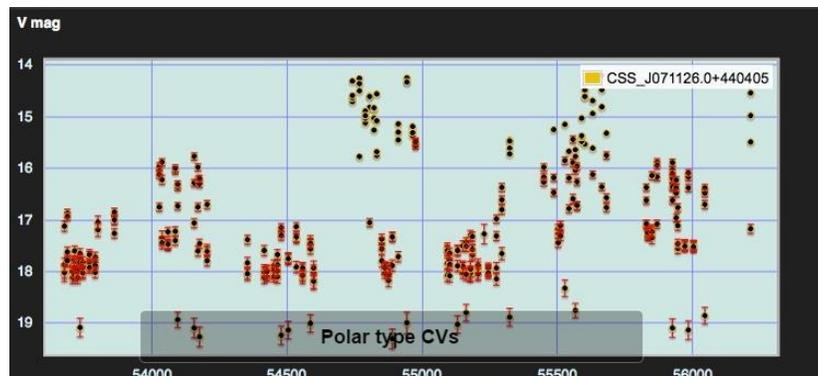
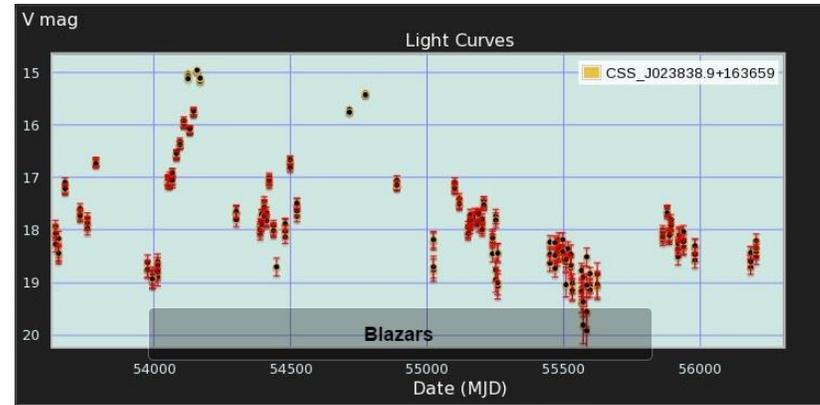
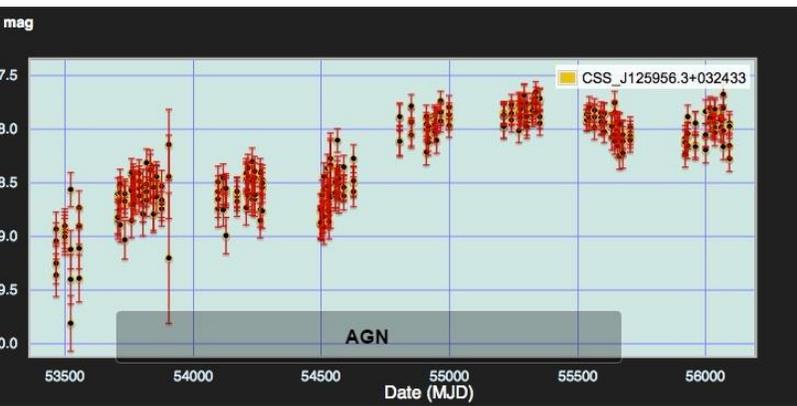
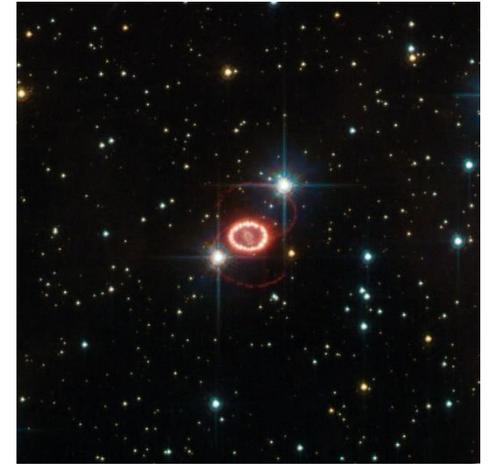
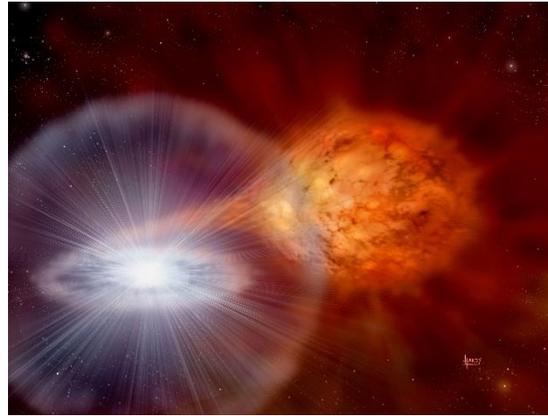
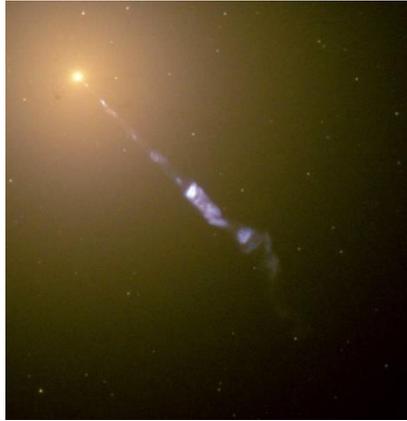
In questo lavoro di tesi (secondo la definizione della CRTS) si intende per transiente un oggetto variabile che mostra una variazione in magnitudine di:

$$\Delta m > 2 \text{ mag}$$

Ricerca dei transienti:

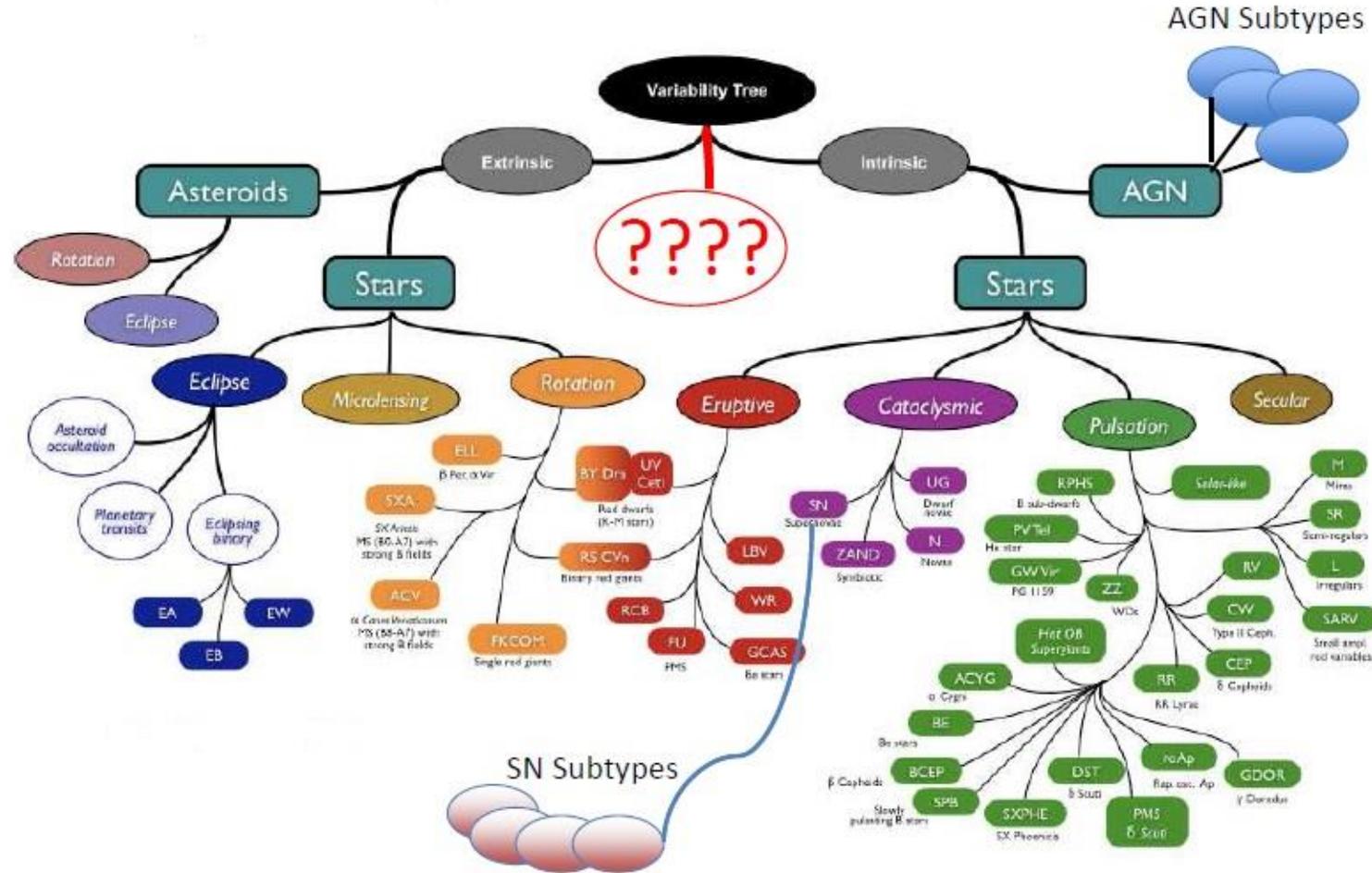
- comparare nuove immagini con quelle d'archivio;
- cercare gli oggetti che hanno subito modifiche;
- minimizzare i falsi positivi;
- classificare i transienti.

Alcuni oggetti variabili...



Time Domain Astronomy

Si definisce Time Domain Astronomy quella branca dell'Astrofisica che si occupa dello studio degli oggetti variabili.



Albero semantico degli oggetti variabili secondo lo schema del Virtual Observatory

Aspetti osservativi/metodologici

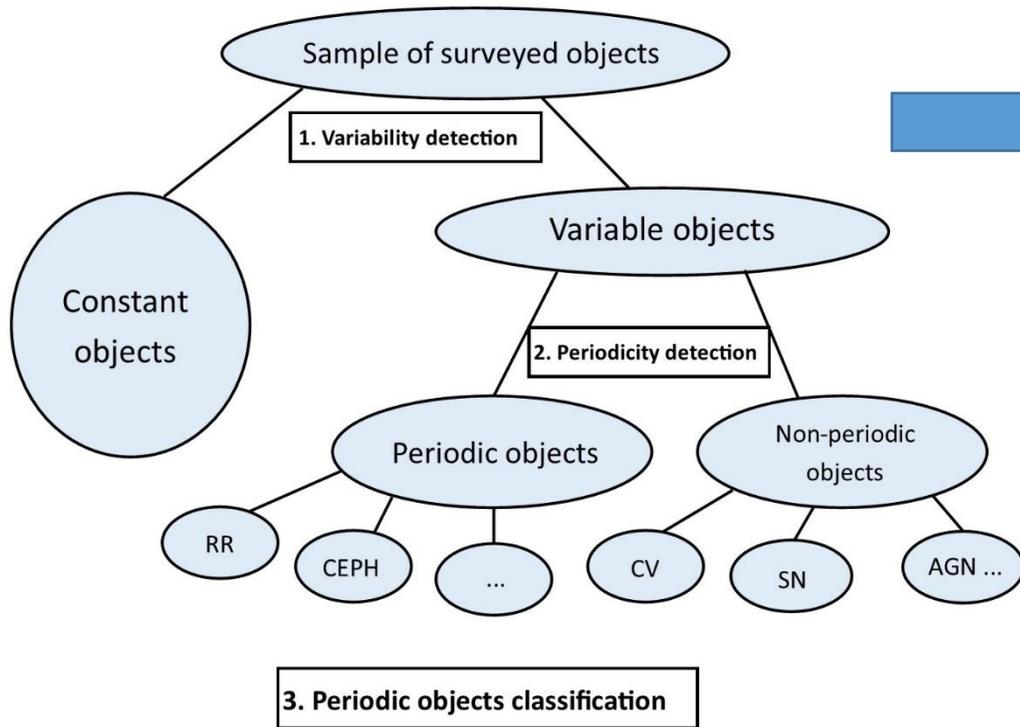
Possiamo distinguere due tipologie fondamentali di ricerca nel campo della TDA:

Online TDA: rivelazione e caratterizzazione in tempo reale dei transienti, per mezzo di sistemi automatizzati.

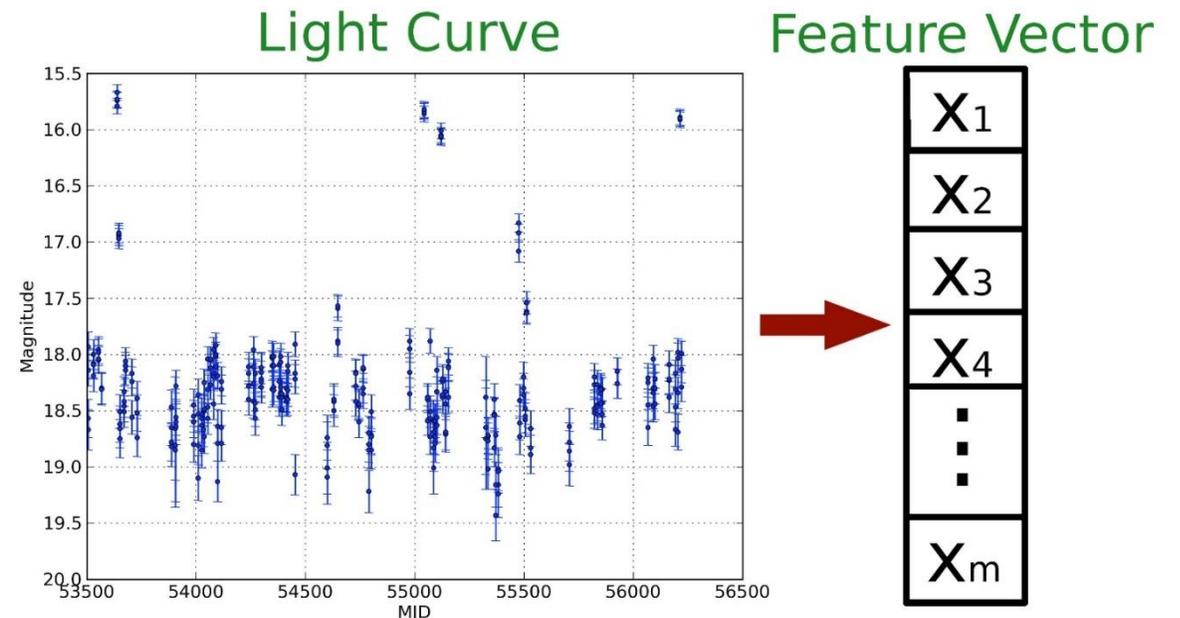
Offline TDA: analisi dell'Universo variabile a partire dal gran numero di curve di luce prodotte dalle survey moderne e conservate negli archivi digitali.

Un'accurata classificazione dei fenomeni transienti costituisce la chiave per determinare le procedure di follow-up ed ottenere una maggiore comprensione fisica degli oggetti.

Classificazione automatizzata dei transienti



Il problema è complicato dal campionamento non uniforme, dal basso rapporto S/N, ecc. I dati vengono perciò omogeneizzati trasformando ogni curva di luce in una serie di parametri (features) statistici.



Features fotometriche

Utilizzando il **Caltech Time Series Characterization Service (CTSCS)** è stato possibile calcolare le 25 features fotometriche utilizzate nel corso degli esperimenti di classificazione.

Alcuni esempi:

- amplitude: media aritmetica tra magnitudine massima e minima
- skew: indice di asimmetria della distribuzione
- linear trend: pendenza della curva di luce
- Lomb-Scargle: periodo ottenuto attraverso il periodogramma, in base al metodo di Lomb-Scargle (Scargle J.D., ApJ, 263, 835, 1982)

$$P_x(\omega) = \frac{1}{2} \left(\frac{[\sum X_j \cos \omega(t_j - \tau)]^2}{\sum \cos^2 \omega(t_j - \tau)} \right) + \left(\frac{[\sum X_j \sin \omega(t_j - \tau)]^2}{\sum \sin^2 \omega(t_j - \tau)} \right)$$

In particolare, il metodo di Lomb-Scargle serve anche per discriminare tra oggetti periodici e non periodici, secondo lo schema di classificazione visto prima.

I dati: Catalina Real Time Transient Survey

La CRTS (<http://crts.caltech.edu/>) è una survey sinottica che copre circa 33mila gradi quadrati del cielo, con lo scopo di scoprire fenomeni transienti. Giunta ormai alla Data Release 2, dà accesso a 500 milioni di curve di luce attraverso un database completamente pubblico.



| Survey | CRTS |
|-------------------------------|----------------------|
| Coverage | 33000 deg^2 |
| Coverage per night | 2200 $deg^2/night$ |
| Field of View | 8 deg^2 |
| Declination | $-75 < dec < 70$ |
| RA | / |
| Galactic latitude | $ b > 15$ |
| Nr of passes/field/night | 4 |
| f _{open} | 0.7 |
| Effective collecting area | 2.326 m^2 |
| t _{exp} | 20 – 30 sec |
| Overall instrument efficiency | 0.7 |
| Full Width Half Maximum | 3 |
| Merit figure | 5470 |
| Limiting magnitude | 21.5 (V) |
| Transient detected | 7500 (CSDR2) |
| Time baseline | From 10 min to 6 yrs |
| Public data release | CSDR2 |
| Number of objects | 500 million |
| Magnitude interval | $11.5 < V < 21.5$ |
| Reference for data | [19] |
| Included surveys | CSS, MLS, SSS |

Reti neurali

Una rete neurale è uno strumento di analisi modellato sulla struttura del cervello umano, inserito in un supporto informatico. Può essere costituita sia da software che da hardware dedicato.

Tre elementi fondamentali caratterizzano la rete:

- L'architettura o topologia
- La funzione di attivazione o trasferimento
- L'algoritmo utilizzato durante la fase di addestramento



Il processo di addestramento si ottiene modificando i pesi relativi alle connessioni tra i neuroni, in modo da garantire la capacità di risolvere un problema specifico. Le tipologie di addestramento sono: supervisionato, non supervisionato e per rinforzo.

Multilayer Perceptron (MLP)

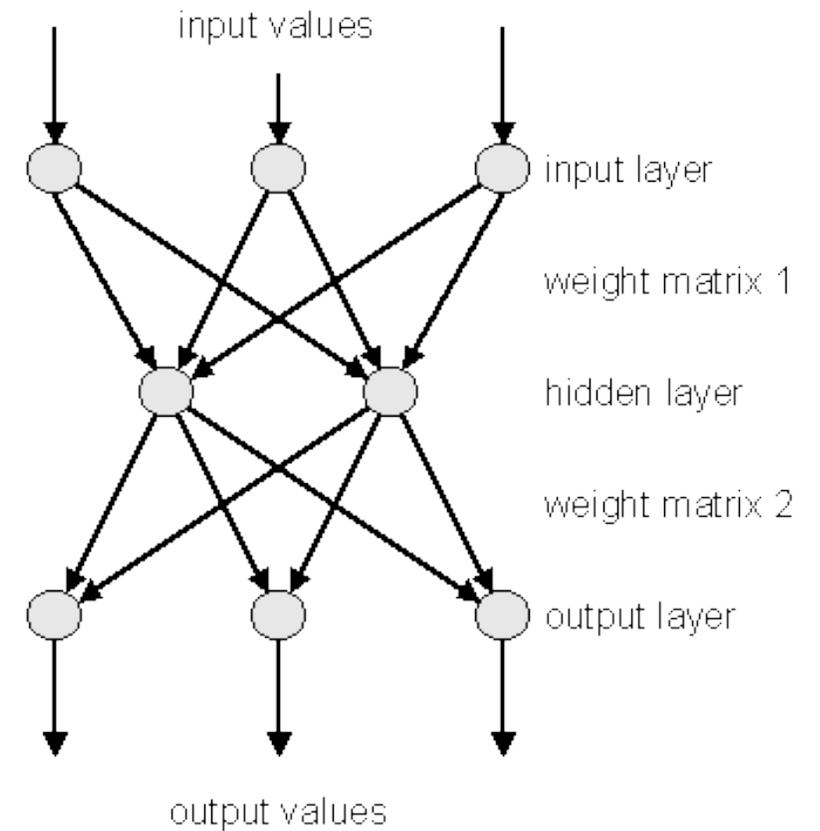
$\bar{x} = \{x_1, x_2, \dots, x_d\}$ input values

$$a_j^{(1)} = \sum_{i=1}^d w_{ji}^{(1)} x_i + b_j^{(1)}, \quad j = 1, \dots, M$$

$$z_j = f(a_j^{(1)}), \quad j = 1, \dots, M$$

$$a_k^{(2)} = \sum_{j=1}^M w_{kj}^{(2)} z_j + b_k^{(2)}, \quad k = 1, \dots, c$$

$$y_k = a_k^{(2)}$$



Il principale algoritmo di addestramento di un MLP è quello della retropropagazione (backpropagation). La funzione d'errore può essere a somma quadratica o cross-entropy, a seconda del particolare tipo di problema in esame (regressione o classificazione).

MLPQNA - I

L'MLPQNA è un tradizionale MLP che implementa come algoritmo di addestramento il modello della Quasi Newton Approximation (QNA).

Il metodo di Newton utilizza il calcolo delle derivate seconde, quindi dell'Hessiano, nella determinazione del minimo dell'errore. Tuttavia in molti casi tale calcolo risulta computazionalmente troppo complesso.

La QNA è un'ottimizzazione dell'algoritmo di addestramento basata su di un'approssimazione statistica dell'Hessiano, attraverso un calcolo ciclico del gradiente, che è alla base del metodo di retropropagazione.

Il QNA, anziché calcolare la matrice Hessiana, effettua una serie di passaggi intermedi, di minore costo computazionale, al fine di generare una sequenza di matrici che risulteranno un'approssimazione via via sempre più accurata di H .

MLPQNA – II

I parametri fondamentali dell'MLPQNA sono:

- Neuroni di input
- Hidden layers e numero di neuroni relativi
- Neuroni di output
- W-step
- Restarts
- Decay
- MaxIts
- K-fold Cross Validation

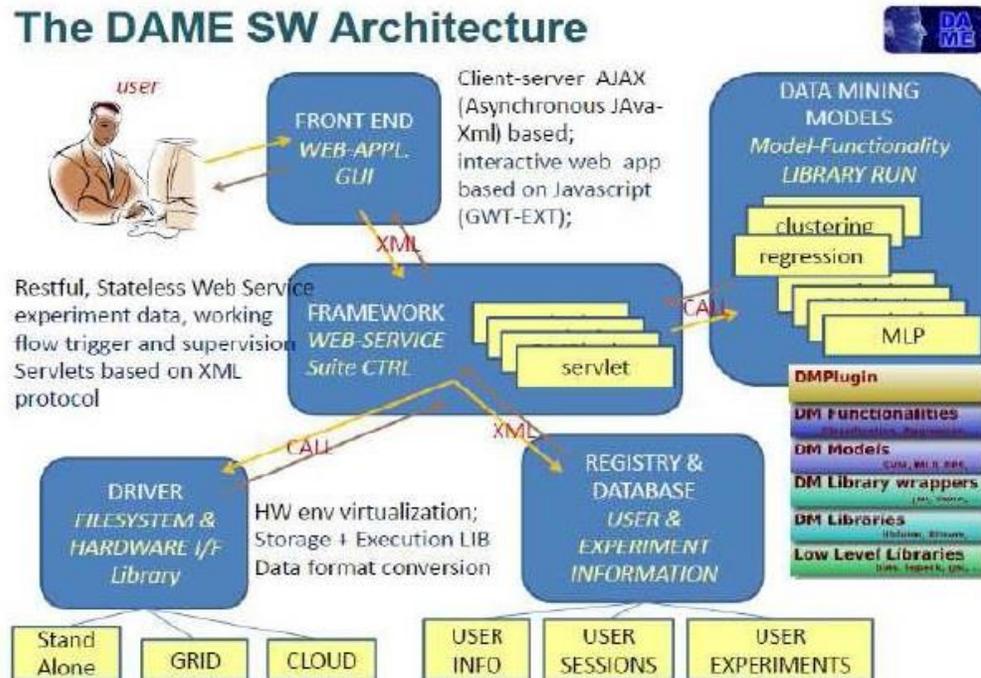
Nel caso della classificazione si ottiene come risultato una matrice di confusione, insieme ai parametri caratterizzanti di efficienza, completezza, purezza e contaminazione per le rispettive classi.

DAMEWARE

DAMEWARE (Data Mining & Exploration Web Application REsource) è un'innovativa piattaforma web-based, totalmente sviluppata a Napoli in collaborazione con il Caltech, per il data mining e l'analisi dati in grandi database attraverso metodi di machine learning.

- Multi-disciplinare
- Accessibile via browser
- Estendibile tramite plug-in e proprio codice

The DAME SW Architecture



| Model | Name | Category | Functionality |
|--------------|---|--------------|----------------------------|
| MLPBP | Multi Layer Perceptron with Back Propagation | Supervised | Classification, regression |
| FMLPGA | Fast MLP trained by Genetic Algorithm | Supervised | Classification, regression |
| MLPQNA | MLP with Quasi Newton Approximation | Supervised | Classification, regression |
| MLPLEMON | MLP with Levenberg-Marquardt Optimization Network | Supervised | Classification, regression |
| SVM | Support Vector Machine | Supervised | Classification, regression |
| RandomForest | Random Forest Algorithm | Supervised | Classification, regression |
| ESOM | Evolving Self Organizing Maps | Unsupervised | Clustering |
| K-Means | | Unsupervised | Clustering |
| SOFM | Self Organizing Feature Maps | Unsupervised | Clustering |
| SOM | Self Organizing Maps | Unsupervised | Clustering |
| PPS | Probabilistic Principal Surfaces | Unsupervised | Feature Extraction |

Base di conoscenza ed esperimenti

Il catalogo finale è costituito da 1619 oggetti e 29 colonne (name, ra, dec, 25 features fotometriche ottenute dalle curve di luce e la classe):

- 461 Cataclismic Variables
- 536 Supernovae
- 124 Blazar
- 140 AGN
- 66 Flare Stars
- 292 RRLyrae

Abbiamo effettuato tre tipologie differenti di esperimenti (classificazione a due classi):

- CV vs ALL
- EXTRA-GALACTIC vs GALACTIC (Blazar+AGN vs CV+SN+Flare)
- SN vs ALL

Esperimenti

Ogni esperimento prevede una classificazione a due classi. Di volta in volta il catalogo da utilizzare veniva preparato selezionando opportunamente le features fotometriche, oltre alla colonna con la classe nota a priori dei patterns (espressa da un numero binario). Il catalogo veniva quindi diviso in due parti, una per il training e l'altra per il test della rete, contenenti rispettivamente l'80% e il 20% degli oggetti.

| amplitude | beyond1std | fpr_mid50 | fpr_mid65 | std | target |
|-----------|------------|-----------|-----------|----------|--------|
| 1.5 | 0.571429 | 0.61319 | 0.806252 | 1.082759 | 1 |
| 0.47817 | 0.33333 | 0.2247 | 0.41031 | 0.29104 | 0 |
| 1.33274 | 0.45454 | 0.71078 | 0.83724 | 0.95925 | 0 |
| 1.36 | 0.310954 | 0.279479 | 0.42138 | 0.50272 | 1 |

STRATEGIA:

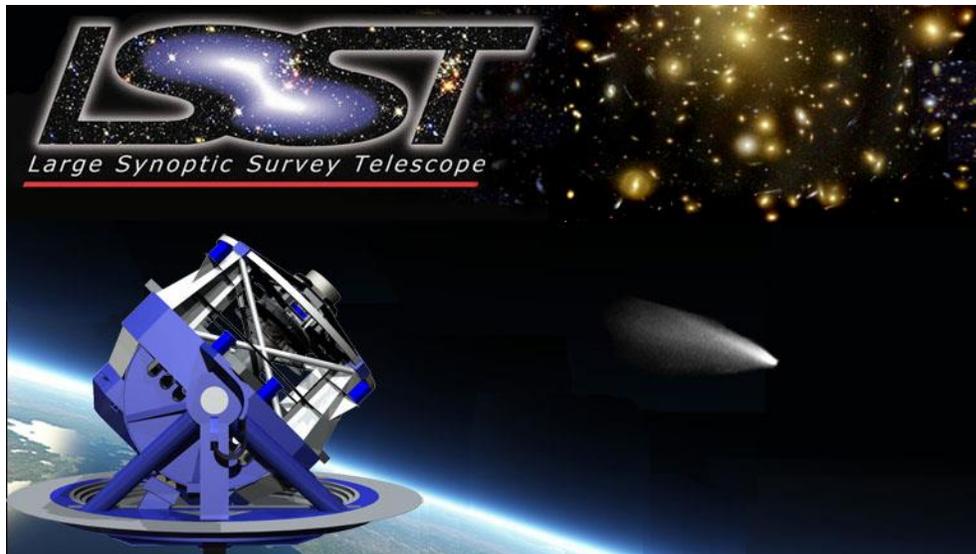
1. Setting iniziale della rete ed esperimenti CV vs ALL
2. Pruning delle features e definizione dei nuclei
3. Esperimenti EXTRA-GALACTIC vs GALACTIC
4. Esperimenti SN vs ALL e confronto con i risultati di Donalek et al. (2013)

Gli esperimenti sono stati ripetuti anche con cataloghi bilanciati (le classi in questione sono spesso molto sbilanciate), per studiare la dipendenza dei risultati dal numero di patterns e dalle features utilizzate.

Lo scopo finale del lavoro è quello di massimizzare l'efficienza negli esperimenti di classificazione!

Ricordiamo che:

Una differenza anche solo dell'1% in efficienza significa classificare correttamente 10^5 oggetti, nel caso di LSST...



10^7 transienti/notte!!!

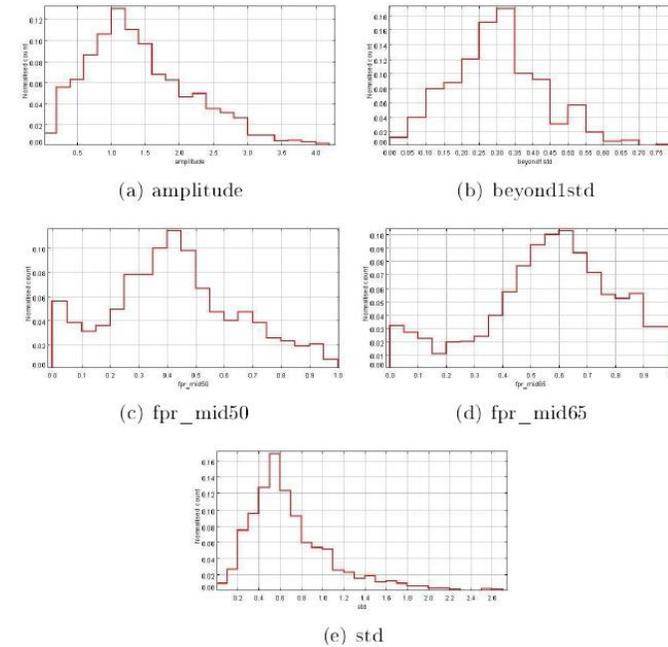
CV vs ALL - I

NUCLEO 1

- amplitude
- beyond1st
- percent_amplitude
- skew
- kurtosis

NUCLEO 2

- amplitude
- beyond1st
- fpr_mid50
- fpr_mid65
- std



| | Hidden layers | Decay | Wstep | Eff (%) | Pur1 (%) | Pur2 (%) | Comp1 (%) | Comp2 (%) |
|-------|---------------|-------|-------|---------|----------|----------|-----------|-----------|
| Test1 | 1 | 0.01 | 0.001 | 78 | 75 | 80 | 65 | 86 |
| Test2 | 1 | 0.001 | 0.001 | 73 | 68 | 75 | 56 | 83 |
| Test3 | 2 | 0.01 | 0.001 | 74 | 70 | 75 | 55 | 85 |
| Test4 | 2 | 0.001 | 0.001 | 73 | 69 | 75 | 54 | 85 |

I primi esperimenti di test, fatti utilizzando il Nucleo 1, servivano a trovare una configurazione di partenza per la rete.

Casi migliori



Configurazioni da usare negli esperimenti successivi

CV vs ALL - II

Nucleo 1:

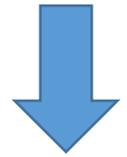
1 Hidden Layer

| N | Feature | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|---|------------------|-------|-------|-------|-------|-------|-------|-------|
| 6 | linear_trend | 79.70 | 67.68 | 86.83 | 75.28 | 81.92 | 24.72 | 18.08 |
| 7 | med_buf_range | 80.83 | 69.77 | 86.11 | 70.59 | 85.63 | 29.41 | 14.36 |
| 8 | pair_slope_trend | 83.46 | 81.63 | 84.52 | 75.47 | 88.75 | 24.53 | 11.25 |

2 Hidden Layer

| N | Feature | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|---|---------------|-------|-------|-------|-------|-------|-------|-------|
| 6 | linear_trend | 78.19 | 66.67 | 85.03 | 72.53 | 81.14 | 27.47 | 18.86 |
| 7 | med_buf_range | 82.33 | 70.93 | 87.78 | 73.49 | 86.34 | 26.51 | 13.66 |
| 8 | fpr_mid65 | 81.95 | 80.64 | 82.66 | 71.43 | 88.82 | 28.57 | 11.18 |

Evidente sbilanciamento



E' emersa tuttavia una fluttuazione nei valori di efficienza effettuando esperimenti ripetuti.

Cataloghi bilanciati per ridurre lo sbilanciamento tra le classi

CV vs ALL - III

Nucleo 2:

1 Hidden Layer

Esperimenti ripetuti
tre volte da questo
momento in avanti.

| Nucleus2 | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-------------------|------------|--------------|--------------|-------------|-------------|--------------|--------------|
| <i>Exp1</i> | 75.94 | 57.83 | 84.15 | 62.34 | 81.48 | 37.66 | 18.52 |
| <i>Exp2</i> | 77.07 | 57.83 | 85.79 | 64.86 | 81.77 | 35.13 | 18.23 |
| <i>Exp3</i> | 75.19 | 53.01 | 85.24 | 61.97 | 80.00 | 38.03 | 20.00 |
| 6 Features | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 77.44 | 64.83 | 84.00 | 67.82 | 82.12 | 32.18 | 17.88 |
| <i>Exp2</i> | 76.31 | 61.54 | 84.00 | 66.67 | 80.77 | 33.33 | 19.23 |
| <i>Exp3</i> | 76.31 | 70.33 | 79.43 | 64.00 | 83.73 | 36.00 | 16.26 |

Esperimenti senza
Cross Validation



Eliminata in quanto identificata come responsabile delle
fluttuazioni, a causa dei pochi dati a disposizione.

EXTRA-GALACTIC vs GALACTIC

Decay 0.01

| Nucleus1 | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-----------------|------------|--------------|--------------|-------------|-------------|--------------|--------------|
| <i>Exp1</i> | 83.46 | 61.40 | 89.47 | 61.40 | 89.47 | 38.60 | 10.53 |
| <i>Exp2</i> | 84.21 | 56.14 | 91.87 | 65.31 | 88.48 | 34.69 | 11.52 |
| <i>Exp3</i> | 84.59 | 57.89 | 91.87 | 66.00 | 88.89 | 34.00 | 11.11 |
| Nucleus2 | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 85.34 | 65.57 | 91.22 | 68.96 | 89.90 | 31.03 | 10.10 |
| <i>Exp2</i> | 87.22 | 67.21 | 93.17 | 74.54 | 90.52 | 25.45 | 9.48 |
| <i>Exp3</i> | 89.47 | 70.49 | 95.12 | 81.13 | 91.55 | 18.87 | 8.45 |

Risultati notevolmente migliori a causa della profonda differenza nella fisica esistente tra le due classi

Decay 0.5 (dopo il pruning)

| Nucleus1 | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-----------------|------------|--------------|--------------|-------------|-------------|--------------|--------------|
| <i>Exp1</i> | 86.09 | 52.63 | 95.21 | 75.00 | 88.05 | 25.00 | 11.95 |
| <i>Exp2</i> | 86.09 | 52.63 | 95.21 | 75.00 | 88.05 | 25.00 | 11.95 |
| <i>Exp3</i> | 86.09 | 52.63 | 95.21 | 75.00 | 88.05 | 25.00 | 11.95 |
| Nucleus2 | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 89.85 | 73.77 | 94.63 | 80.36 | 92.38 | 19.64 | 7.62 |
| <i>Exp2</i> | 89.85 | 72.13 | 95.12 | 81.48 | 91.98 | 18.52 | 8.02 |
| <i>Exp3</i> | 89.85 | 72.13 | 95.12 | 81.48 | 91.98 | 18.52 | 8.02 |

SN vs ALL - I

| Nucleus2 | Eff | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-----------------|------------|--------------|--------------|-------------|-------------|--------------|--------------|
| <i>Exp1</i> | 79.70 | 72.75 | 84.97 | 78.09 | 80.74 | 21.90 | 19.25 |
| <i>Exp2</i> | 81.20 | 76.99 | 84.31 | 78.38 | 83.22 | 21.62 | 16.77 |
| <i>Exp3</i> | 81.20 | 76.99 | 84.31 | 78.38 | 83.22 | 21.62 | 16.77 |

[Donalek et al., "Feature selection strategies for classifying high dimensional astronomical data sets", ArXiv:1310.1976, 2013](#)

| relieff | cfs/fcbf | mcfs | fdr |
|----------------|-------------------|-------------------|--------------|
| amplitude | beyond1std | max_slope | fpr_mid50 |
| beyond1std | linear_trend | percent_amplitude | linear_trend |
| fpr_mid80 | percent_amplitude | pdfp | pdfp |
| skew | | kurtosis | skew |
| std | | | kurtosis |
| magratio | | | std |

Nell'articolo di confronto i vari gruppi di features utilizzati sullo stesso catalogo sono stati ricavati attraverso metodi di selezione automatizzati. Si è così deciso di utilizzare gli stessi gruppi per effettuare la classificazione con l'MLPQNA, mentre i classificatori usati da Donalek et al. sono:

- K-nearest-neighbor (KNN)
- Decision Trees (DT)

SN vs ALL - II

Sia i risultati di Donalek che quelli dei nostri esperimenti ottenuti con due di questi gruppi di features superano i risultati ottenuti con il Nucleo 2.



Forte dipendenza dalle features usate

I risultati sono comunque simili ed in alcuni casi migliori!

| Feature Selection Strategy | KNN Misclass | DT Misclass |
|----------------------------|--------------|-------------|
| ReliefF | 22% | 15% |
| CFS | 24% | 17% |
| FCBF | 24% | 17% |
| MCFS | 32% | 19% |
| FDR | 22% | 16% |

| ReliefF | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-------------|-------|----------|-------|-------|-------|-------|-------|-------|
| <i>Exp1</i> | 77.16 | 22.84 | 60.91 | 85.51 | 68.37 | 80.97 | 31.63 | 19.03 |
| <i>Exp2</i> | 77.47 | 22.53 | 66.36 | 83.18 | 66.97 | 82.79 | 33.03 | 17.21 |
| <i>Exp3</i> | 77.47 | 22.53 | 58.18 | 87.38 | 70.33 | 80.26 | 29.67 | 19.74 |
| cfs/fcbf | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 84.88 | 15.12 | 54.74 | 97.38 | 89.65 | 83.83 | 10.34 | 16.16 |
| <i>Exp2</i> | 85.49 | 14.51 | 56.84 | 97.38 | 90.00 | 84.47 | 10.00 | 15.53 |
| <i>Exp3</i> | 85.49 | 14.51 | 56.84 | 97.38 | 90.00 | 84.47 | 10.00 | 15.53 |
| mcfs | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 64.51 | 35.49 | 29.46 | 83.02 | 47.83 | 69.02 | 52.17 | 30.98 |
| <i>Exp2</i> | 64.51 | 35.49 | 29.46 | 83.02 | 47.83 | 69.02 | 52.17 | 30.98 |
| <i>Exp3</i> | 64.51 | 35.49 | 29.46 | 83.02 | 47.83 | 69.02 | 52.17 | 30.98 |
| fdi | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 84.88 | 15.12 | 71.17 | 92.02 | 82.29 | 85.96 | 17.71 | 14.03 |
| <i>Exp2</i> | 85.80 | 14.20 | 71.17 | 93.43 | 84.95 | 86.15 | 15.05 | 13.85 |
| <i>Exp3</i> | 86.11 | 13.89 | 70.27 | 94.37 | 86.67 | 85.90 | 13.33 | 14.10 |

SN vs ALL - III

cfs/fcbf

| Decay 0.5 | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-------------|-------|----------|-------|-------|-------|-------|-------|-------|
| <i>Exp1</i> | 84.88 | 15.12 | 54.74 | 97.38 | 89.65 | 83.83 | 10.34 | 16.16 |
| <i>Exp2</i> | 85.49 | 14.51 | 56.84 | 97.38 | 90.00 | 84.47 | 10.00 | 15.53 |
| <i>Exp3</i> | 85.49 | 14.51 | 56.84 | 97.38 | 90.00 | 84.47 | 10.00 | 15.53 |
| Decay 0.05 | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 87.34 | 12.66 | 67.37 | 95.63 | 86.49 | 87.60 | 13.51 | 12.40 |
| <i>Exp2</i> | 87.65 | 12.35 | 67.37 | 96.07 | 87.67 | 87.65 | 12.33 | 12.35 |
| <i>Exp3</i> | 87.65 | 12.35 | 67.37 | 96.07 | 87.67 | 87.65 | 12.33 | 12.35 |

Il pruning sul Decay
migliora notevolmente i
risultati ottenuti.

fdr

| Decay 0.5 | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
|-------------|-------|----------|-------|-------|-------|-------|-------|-------|
| <i>Exp1</i> | 84.88 | 15.12 | 71.17 | 92.02 | 82.29 | 85.96 | 17.71 | 14.03 |
| <i>Exp2</i> | 85.80 | 14.20 | 71.17 | 93.43 | 84.95 | 86.15 | 15.05 | 13.85 |
| <i>Exp3</i> | 86.11 | 13.89 | 70.27 | 94.37 | 86.67 | 85.90 | 13.33 | 14.10 |
| Decay 0.005 | Eff | Misclass | Comp1 | Comp2 | Pur1 | Pur2 | Cont1 | Cont2 |
| <i>Exp1</i> | 86.11 | 13.89 | 78.38 | 90.14 | 80.55 | 88.89 | 19.44 | 11.11 |
| <i>Exp2</i> | 86.11 | 13.89 | 78.38 | 90.14 | 80.55 | 88.89 | 19.44 | 11.11 |
| <i>Exp3</i> | 86.11 | 13.89 | 79.28 | 89.67 | 80.00 | 89.25 | 20.00 | 10.75 |

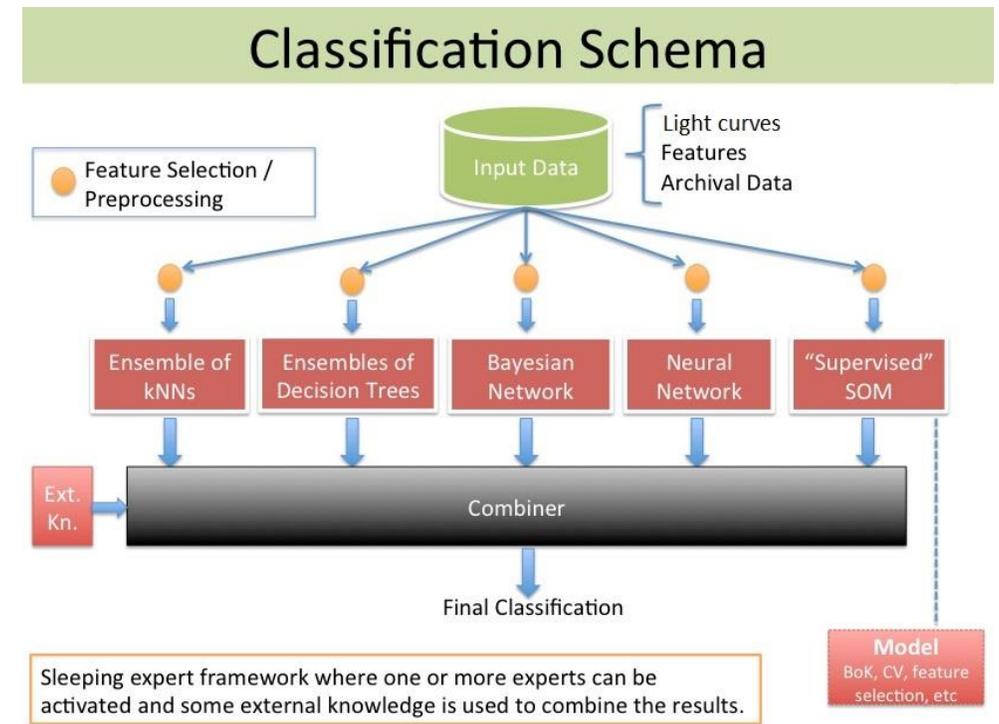
Conclusioni

In definitiva, otteniamo una corretta classificazione, nei casi migliori, con efficienza di:

- 83% negli esperimenti CV vs ALL
- 90% negli esperimenti EXTRA-GALACTIC vs GALACTIC
- 87% negli esperimenti SN vs ALL

Inoltre i risultati ottenuti sono simili a quelli di Donalek et al. e confermano l'efficacia del metodo.

Scopo ultimo, del quale questo lavoro costituisce soltanto un gradino, è quello della costruzione di un intero processo che a partire dai dati grezzi, attraverso una fase di pre-processing e di feature selection, permetta di effettuare la classificazione attraverso l'utilizzo di differenti metodi automatizzati da combinare assieme ad una base di conoscenza esterna.

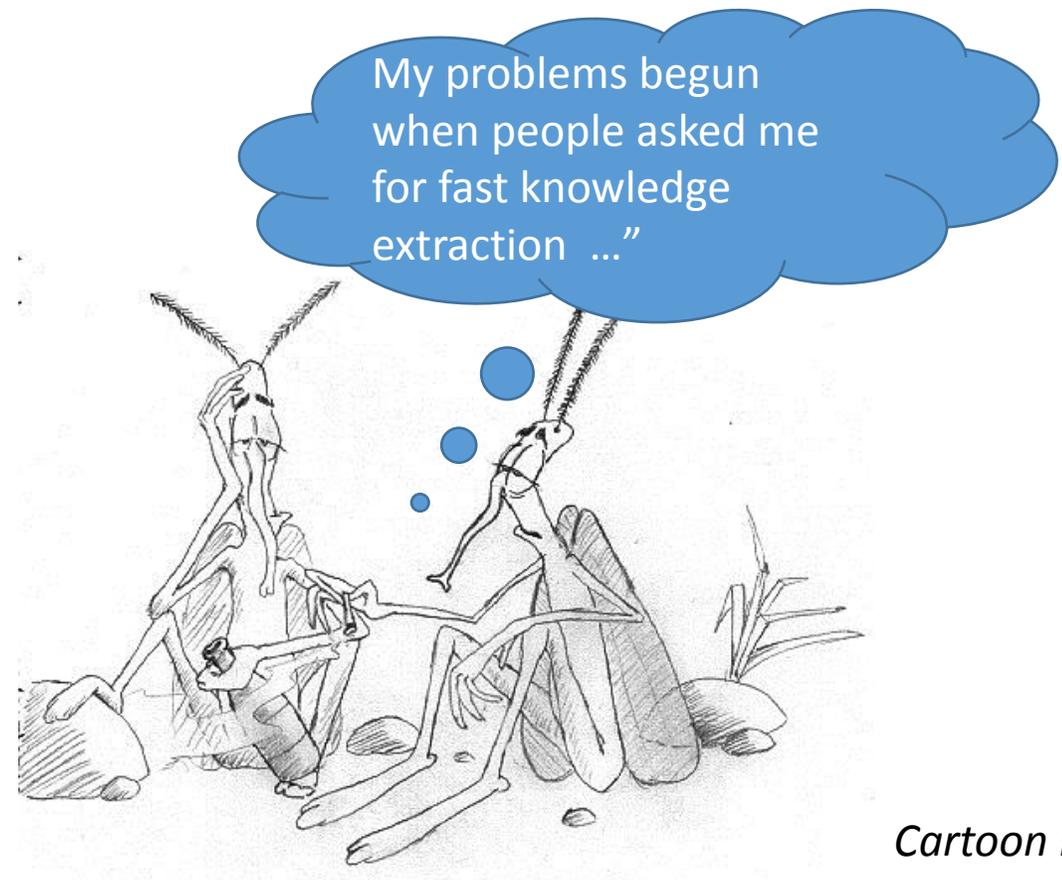


Credit: Donalek et al.

Sviluppi futuri

Gli obiettivi prossimi da raggiungere, allo scopo di completare lo schema di classificazione, sono i seguenti:

- Aumentare la base di conoscenza
- Studiare classificatori ottimali per altri nodi del Decision Tree
- Come combinare i classificatori rimane, tuttora, un problema aperto...



Cartoon by D. Vinkovic



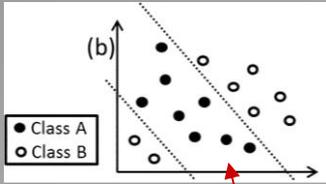
“Ma una cosa è certa: l'uomo, se non altro, è musica e gli stesso. Un tema coraggioso, che rende musicale anche il suo vasto accompagnamento, la sua matrice di uragani e di stelle.”

Olaf Stapledon - Infinito (Last and First Man)

MLP trained by Quasi Newton Rule

$$\min_w E(w) = \frac{1}{2P} \sum_{p=1}^P E_p(w) = \frac{1}{2P} \sum_{p=1}^P (y(x^p; w) - d^p)^2$$

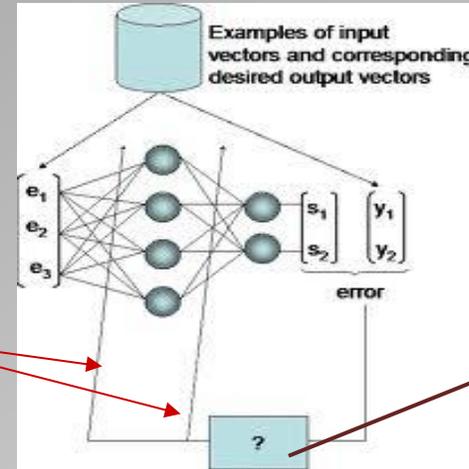
E_p is a measure of the error related to the p -th pattern



$$w^{k+1} = w^k + \alpha^k d^k$$

$d^k \in \mathbb{R}^N$ **DIRECTION OF SEARCH**

$\alpha^k \in \mathbb{R}$



$$\nabla^2 E(w^k) d^k \approx -\nabla E(w^k)$$

Hessian approx. (QNA)

The implemented MLPQNA model uses Tikhonov regularization (AKA weight decay).

When the regularization factor is accurately chosen, then generalization error of the trained neural network can be improved, and training can be accelerated.

If it is unknown what **Decay** regularization value to choose (as usual), it could be possible to experiment the values within the range of 0.001 (weak regularization) up to 100 (very strong regularization)

