



Corso di Laurea in Informatica

TESI DI LAUREA

Rilevamento Continuo di Anomalie nel Contesto
Satellitare: una Validazione Empirica

Relatore:
Vincenzo Lomonaco

Candidato:
Francesco Fiaschi

Correlatore:
Lampeï Li

ANNO ACCADEMICO 2024/2025

Sommario

bgnbrtgnbrtgnt

Indice

1	Introduzione	7
1.1	Rilevamento di Anomalie nei Satelliti	7
1.2	NASA ed ESA	8
1.3	Obbiettivo	8
2	Aquisizione dei Dati	11
2.1	Dataset ESA	11
2.2	Motivazioni	12
2.3	OP-SAT Dataset	13
3	Misure di Valutazione	17
4	Valutazione dei Modelli	21
5	Panoramica sull'Efficientamento	23
A	Questa è un'appendice	25

Elenco delle figure

2.1	Satellite OP-SAT. Fonte: European Space Agency	14
3.1	Curva ROC	19

1 Introduzione

1.1 Rilevamento di Anomalie nei Satelliti

Negli anni si è fatto sempre più presente il bisogno di esplorare lo spazio puntando sempre più lontano, studiando e comprendendo le regole e le strutture dell'universo partendo dall'osservazione di quest'ultimo. Con l'avanzare della tecnologia abbiamo potuto avvicinarci allo spazio tramite i satelliti posizionati in orbita intorno alla terra; questi sono circa 3500 attivi e si è fatta sempre più importante la richiesta di rilevare le anomalie di tali satelliti e comunicandole alla stazione di controllo, portando un aumento nello scambio di informazioni con un costo non irrilevante data la banda limitata. Prima dell'avvento dell'intelligenza artificiale la rilevazione delle anomalie era lasciata ad un'analisi statica delle componenti, controllo di qualità con regole e soglie fisse, questo ovviamente portava ad avere un alta probabilità di ricevere tante segnalazioni di anomalie non rilevanti o date da cambiamenti nei riferimenti che creano un diverso ambiente vanificando i controlli statici. Con l'intelligenza artificiale siamo riusciti ad avere un rilevamento adattivo in base all'ambiente che portava ad avere meno false anomalie ma anche questo approccio portava con se delle problematiche: l'algoritmo di intelligenza artificiale veniva allenato nel centro di controllo su dati reali e poi spedito sul satellite, ogni volta che bisognava cambiare i parametri dell'algoritmo bisognava ripetere tutto il processo dato che i satelliti all'interno hanno un processore molto poco prestante e quindi non sufficiente ad eseguire l'allenamento degli algoritmi, sprecando così tanta banda e molto tempo di trasmissione.

1.2 NASA ed ESA

Negli anni sia la NASA (National Aeronautics and Space Administration) che l'ESA (Agenzia Spaziale Europea) hanno pubblicato molteplici dataset contenenti dati reali di missioni e relativi benchmark sull'esecuzione di algoritmi di intelligenza artificiale per trovare il giusto compromesso tra efficacia e efficienza di un algoritmo, focalizzandosi maggiormente sul trovare l'algoritmo migliore per il rilevamento delle anomalie.

I dataset sono stati resi pubblici per incentivare la comunità a contribuire alla ricerca di nuove tecniche di monitoraggio e rilevamento delle anomalie, potendo usare dati reali.

1.3 Obbiettivo

Partendo dai dati proposti nei dataset e benchmark di ESA e NASA vogliamo analizzare quali test sono stati effettuati per capire come poter fare un ulteriore passo in avanti nel rilevamento continuo di anomalie cercando di rendere più efficiente un algoritmo già efficace in modo da poterlo allenare direttamente sul satellite, limitando così lo scambio di comunicazioni e restringendo ancora di più il rilevamento di false anomalie.

Tutto questo processo è dedito a trovare un algoritmo che abbia un giusto compromesso tra efficienza ed efficacia così che possa rilevare in modo corretto la maggioranza delle anomalie senza però avere un costo molto alto in termini di consumo di banda e di risorse del satellite.

Successivamente come prima cosa analizzeremo i dati a nostra disposizione relativi all'efficacia ed efficienza in termini di, rispettivamente, anomalie rilevate e tempo di esecuzione. Dopo di ciò procederemo a scendere più nello specifico nell'implementazione di un **algoritmo "da scrivere il nome"** provando a renderlo più efficiente così da poter effettuare il training sui satelliti. Per validare l'implementazione fatta portiamo a sostegno risultati ottenuti in primo luogo effettuando test e confronti tra un'implementazione standard e la nostra proposta

mettendo a paragone le metriche di valutazione; in secondo luogo procederemo ad eseguire l'algoritmo su una macchina RaspberryPi, ovvero un mini-computer con un processore molto simile a quello di un satellite standard con poca capacità di calcolo, per avere un riscontro simile ad un'applicazione reale.

2 Aquisizione dei Dati

Procediamo ora a descrivere i dataset contenenti dati di telemetrie registrati negli anni e resi pubblici in modo da poter sperimentare e migliorare gli algoritmi già presenti.

2.1 Dataset ESA

Il dataset di riferimento più importante per la rilevazione delle anomalie è quello fornito dall'ESA (European Space Agency) che conta dati di tre missioni. Di queste solo i dati di due vengono utilizzati per creare il benchmark, per le caratteristiche dell'insieme di dati; infatti in "*Mission 3*" abbiamo: poche anomalie e per lo più banali, un gran numero di buchi di comunicazione e segmenti non validi.

Nell'articolo in questione i dati satellitari grezzi di "*Mission 1*" e "*Mission 2*" vengono preprocessati per renderli uniformi e quindi utilizzabili con la maggior parte degli algoritmi.

In questa fase utilizziamo OXI^[1] per l'etichettatura collaborativa dei dati da cui si è potuto estrarre delle telemetrie che rappresentano periodi di funzionamento nominale e anomalo. Tutti i dati sono stati sottoposti ad una doppia fase di etichettatura e controllo.

I canali sono divisi tra "target" e "non target" dove quest'ultimi sono usati per gli algoritmi solo come informazioni aggiuntive. Un canale target, quello usato per rilevare le anomalie, è diviso in gruppi di dati con caratteristiche simili così da rendere più facile per l'algoritmo processarli ed interpretarli e, nel caso, allenarlo solo sul quel gruppo specifico.

	Mission 1	Mission 2
Channels	76	100
Target/Non target	58/18	47/53
Telecommands	698	123
Annotate Events	200	644
Anomalies	118	31
Rare Nominal Events	78	613
Univariate/Multivariate	32/164	9/635

Tabella 2.1: Configurazione Dataset ESA

Possiamo osservare dalla Tabella 2.1 che la densità di anomalie in termini di punti di dati annotati, varia tra 0,57% per la "Mission 2" e 1,80% per la "Mission 1" questo va a confermare un'impronta più realistica rispetto ai dataset meno recenti che avevano una densità di anomalie estremamente più alta ed irrealistica.

2.2 Motivazioni

Il dataset ESA precedentemente descritto porta alla risoluzione di vari problemi noti nell'ambito della rilevazione di anomalie. Il primo problema come si evidenzia nei dataset *NASA Soil Moisture Active Passive* (SMAP) e *Mars Science Laboratory* (MSL) i quali offrono brevi frammenti di segnali e comandi correlati da 55 e 27 parametri di telemetria, rispettivamente, con un totale di 105 anomalie annotate; infatti abbiamo una densità di anomalie irrealistica, alto numero di anomalie banali, verità di base etichettate in maniera errate e una mancanza di correlazione tra comandi e canali. Quindi è stato deciso che questi dataset non andrebbero usati per il benchmarking del rilevamento delle anomalie. Il secondo problema invece rappresenta la mancanza di annotazione di eventi anomali, alcuni esempi sono gli insiemi di dati di *Mars Express* e *NASA WebTCAD*.

Il dataset ESA risolve i problemi elencati ma nasce come dataset di missioni su larga scala, le quali sono molto complesse e stabili portando con se problemi non più relativi alla distribuzione delle anomalie ma improntati verso problemi di esecuzione e potenza di calcolo.

Il dataset che introdurremo dopo è concettualmente diverso, infatti affronta

una missione ESA OP-SAT molto semplificata al fine di rendere più accessibile l'uso, oltre ad essere di dimensione considerevolmente minore.

Questo processo di apertura è reso anche tramite la sostituzione giornaliera dell'intero sistema software fino al sistema operativo del satellite così da consentire alle persone che sperimentano di caricare il proprio software a bordo e a terra. Per implementare questa funzionalità venne usato FDIR^[2]

Le telemetrie grezze sono caratterizzate da molte lacune nei dati ed altre imprecisioni, queste vengono curate da ingegneri ed esperti nei modelli di apprendimento automatico per rendere l'insieme di dati fruibile alla creazione di tecniche di rilevamento delle anomalie basate sui dati. Una di queste modifiche è la seguente:

Selezione e annotazione di verità di base (ground-truth) di 2123 brevi frammenti di telemetria a canale singolo ossia serie temporali univariate^[3], registrate in 9 canali di telemetria.

Infine questo dataset con il relativo benchmark e tutti i dati al suo interno disponibili è stato messo a disposizione per aiutare la comunità nella ricerca di nuovi approcci per la rilevazione delle anomalie, confrontandosi tra di loro in modo imparziale ed equo così da affrontare anche il problema della riproducibilità nell'ambiente dell'apprendimento automatico (dato che i dataset sono uniformati, l'ambiente di esecuzione diverso e varie metriche usate)

Tutto ciò che è stato sviluppato con l'aiuto di questo dataset verrà convalidato con il lancio alla fine del 2025 del satellite successore OP-SAT VOLT (dopo averli distribuiti a bordo del satellite).

2.3 OP-SAT Dataset

Nel dataset OP-SAT sono contenuti i dati delle telemetrie del satellite OP-SAT dell'ESA (Figura 2.1). Questo satellite di tipo CubeSat aveva dimensione 3 unità (3U dove $1U=10\text{cm}^3$), esso ormai non è più in orbita, era stato lanciato a Dicembre del 2019 con lo scopo di dimostrare l'elaborazione dei dati in orbita e di generare dati utili come immagini satellitari e telemetrie.



Figura 2.1: Satellite OP-SAT. Fonte: European Space Agency

Come nel dataset precedente anche qui i dati hanno bisogno di essere preprocessati per renderli fruibili alla maggior parte degli algoritmi (OXI^[1]). In questo caso sono state progettate 18 caratteristiche per l'attività di rilevamento delle anomalie, ossia sono stati estratti tratti significativi dai dati. Questo processo è chiamato Feature Extraction e serve per ridurre la complessità dei dati in ingresso rendendoli più significativi.

Tutti i segmenti ricavati rappresentano le sfide che dobbiamo affrontare con i dati delle telemetrie, ognuno di questi è composto da:

- `<timestamp>`: rappresenta il marcatore temporale nel momento della registrazione;
- `<channel>`: è il nome del canale;
- `<value>`: è il valore del segnale acquisito;
- `<label>`: rappresentano le annotazioni certe, quelle annotate manualmente;
- `<segment>`: rappresenta il numero consecutivo del segmento;
- `<sampling>`: tasso di campionamento;
- `<train>`: che indica se il segmento appartiene al training set.

Il dataset è stato diviso in una parte di allenamento, Training Set (T), ed una di test, Test Set (Ψ). Questi rappresentano rispettivamente i dati usati per

l'allenamento del modello e i dati usati per fare le valutazioni delle performance dell'algoritmo.

Classi	Training Set (T)	Test Set (Ψ)	Total
Nominali	1273	416	1689
Anomalie	321	529	434

Tabella 2.2: Composizione Dataset OP-SAT

Le classi, nominali e anomalie, rappresentano rispettivamente segmenti che rispettano valori normali o attesi e segmenti invece che superano questi valori o discordano dai valori attesi. Partendo dai benchmark effettuati su questi due dataset, analizzeremo le metriche e le prestazioni utilizzando in modo più specifico il secondo dataset (OP-SAT) che contiene dati più rilevanti per il nostro obiettivo.

3 Misure di Valutazione

Analizzeremo ora le misure che ci permettono di valutare le prestazioni di un modello allenato sui dati del training set, quindi con i relativi pesi calcolati. I modelli vengono provati sul test set ed otteniamo le seguenti metriche:

- **Accuratezza:** rappresenta la percentuale di previsioni corrette (TP - True Positive) su tutti i casi possibili, tiene in considerazione anche i veri negativi (TN). Questa misura ci indica quanto ci stiamo avvicinando ai dati di allenamento.

Formula:

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

- **Precisione:** rappresenta la percentuale di anomalie vere rilevate in confronto a tutte le anomalie segnalate quindi la percentuale di veri positivi.

Formula:

$$\frac{TP}{TP + FP} \quad (3.2)$$

Nel nostro caso questa metrica è particolarmente importante dato che un valore troppo basso significherebbe un'alta probabilità di avere falsi positivi portando quindi uno spreco di banda ed energia per mandare i messaggi ed un allarme che richiede un intervento non necessario.

- **Richiamo:** rappresenta la percentuale di rilevare tutte le anomalie vere tenendo quindi in considerazione anche le anomalie non rilevate (FN). Questa misura è anche detta sensibilità.

Formula:

$$\frac{TP}{TP + FN} \quad (3.3)$$

Nel contesto dei satelliti questa metrica è cruciale dato che con un valore basso avremmo un grande numero di falsi negativi (FN) che quindi senza intervento potrebbero portare a conseguenze molto gravi.

- **F_1 score:** rappresenta una media armonica tra precisione e recupero dove il massimo si ottiene con il valore uno e il minimo a zero.

Formula:

$$F_1 = \frac{2 * TP}{2 * TP + FP + FN} \quad (3.4)$$

Questa misura nel nostro caso che cerchiamo un buon compromesso tra precisione e richiamo per non avere un alto numero né di falsi negativi né di falsi positivi. Questo porta ad un equilibrio tra precisione e capacità di rilevazione.

- **Coefficiente di correlazione di Matthews (MCC)**^[4]: rappresenta una misura della qualità del modello con dati molto variabili.

Formula:

$$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + TP) \cdot (TN + FN)}} \quad (3.5)$$

Questa misura è particolarmente indicata per il rilevamento delle anomalie dato che concede la stessa importanza a veri positivi, falsi positivi, veri negativi e falsi negativi.

- **L'area sotto la curva ROC (AUC_{ROC})**^{[5][6]}: rappresenta il rapporto tra il tasso di veri positivi e il tasso di falsi positivi (Figura: 3.1) Permette di osservare come varia il richiamo in funzione della metrica di precisione. Può anche essere usato per scegliere il modello migliore tra due guardando semplicemente l'area sotto al grafico, quella con l'area più grande è generalmente quello migliore.

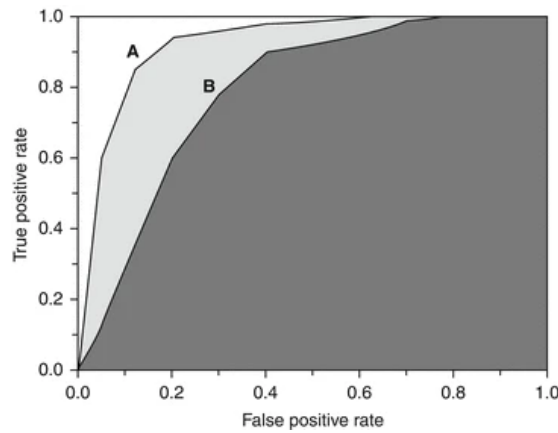


Figura 3.1: Curva ROC

- **Area sotto la curva Precisione-Richiamo (AUC-PR):** rappresenta un semplice modo per sintetizzare le prestazioni generali di un modello, più alto è il valore più avrà un numero di predizioni alto. Non vengono considerati i veri positivi nella curva precisione-richiamo.

In queste formule per il calcolo delle metriche abbiamo usato TP per rappresentare i veri positivi ossia i segmenti delle telemetrie correttamente identificati come anomalie, TN per i veri negativi cioè segmenti correttamente identificati come nominali ossia regolari), FP per i falsi positivi ossia i segmenti erratamente classificati anomalie e FN per i falsi negativi che rappresentano i segmenti erratamente classificati come non anomalie. Tutte le metriche descritte possono assumere valori tra $[0, 1]$ tranne MCC che assume valori tra $[-1, 1]$. Tutte le metriche però più si avvicinano ad uno e più il modello testato risulta migliore rispetto ad uno con valori inferiori delle stesse.

Le metriche viste precedentemente ci serviranno successivamente per valutare come poter migliorare in termini di efficienza mantenendo livelli buoni delle metriche fondamentali per il rilevamento delle anomalie, non andando ad intaccare il funzionamento degli algoritmi.

4 Valutazione dei Modelli

5 Panoramica sull'Efficientamento

Efficientare un algoritmo significa renderlo eseguibile in maniera ottimizzata anche con macchine con caratteristiche molto restringenti, come nel caso dei satelliti, con poche risorse hardware come ad esempio la CPU poco performante. Nel nostro caso effettuiamo un Transfer Learning, ossia il riuso di un modello già addestrato su un dataset, riadattandolo al nuovo compito o dataset che vogliamo utilizzare. Questo processo può essere effettuato in diversi modi:

- Estrazione delle Caratteristiche: consiste nell'estrarre caratteristiche utili dai dati senza modificare i pesi calcolati nel training precedente;
- Fine-Tuning:
 - Addestramento sull'intera rete: in questo caso aggiorniamo i pesi eseguendo di nuovo il training sul nuovo dataset;
 - Addestramento del Classificatore Finale: aggiorniamo solo gli stati più profondi (finali) della rete mentre i pesi dei primi rimangono congelati;
 - Addestramento a Blocchi: i pesi dei blocchi vengono aggiornati singolarmente effettuando l'addestramento blocco per blocco.

Per il nostro scopo possiamo effettuare diversi tipi di fine-tuning:

1. Precisione dei Pesi: i pesi derivati dall'addestramento della rete hanno una precisione, ossia il numero di bit che vengono usati per rappresentare il numero, diminuendola impiegheremo meno memoria necessaria e velocizzeremo la velocità di calcolo;

-
2. Riducendo la Complessità del Modello: possiamo eliminare nodi poco significativi tramite la tecnica detta pruning, diminuendo anche la quantità di operazioni necessarie;
 3. Compromesso Concorrenza Aggiornamento pesi: il batch size ossia il numero di pesi che si attende prima di aggiornarli evitando di farlo ogni volta per non sprecare risorse di calcolo, incentivando così la concorrenza;
 4. Ridurre tempo di addestramento: riducendo il numero di epoche, ossia il numero di volte in cui il dataset viene passato attraverso il modello durante la fase di allenamento.

A Questa è un'appendice

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Questa è un'altra sezione, ma non viene inserita nell'indice

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Bibliografia

- [1] J. Andrzejewski C. Haskamp B. Ruszczak, K. Kotowski and J. Nalepa. Oxi: An online tool for visualization and annotation of satellite time series data. Jul. 2023.
- [2] D. J. Evans. Ops-sat: Fdir design on a mission that expects bugs - and lots of them. *SpaceOps Conferences*, SpaceOps 2016 Conference, American Institute of Aeronautics and Astronautics, 2016.
- [3] Bogdan Ruszczak, Krzysztof Kotowski, Jacek Andrzejewski, Alicja Musiał, David Evans, Vladimir Zelenevskiy, Sam Bammens, Rodrigo Laurinovics, and Jakub Nalepa. Machine learning detects anomalies in ops-sat telemetry. In Jiří Mikyška, Clélia de Mulatier, Maciej Paszynski, Valeria V. Krzhizhanovskaya, Jack J. Dongarra, and Peter M.A. Sloot, editors, *Computational Science – ICCS 2023*, pages 295–306, Cham, 2023. Springer Nature Switzerland.
- [4] Brunak S. Chauvin Y. Andersen C. A. F. Nielsen H. Baldi, P. Assessing the accuracy of prediction algorithms for classification: an overview. <https://academic.oup.com/bioinformatics/article-pdf/16/5/412/48836094/bioinformatics165412.pdf>, *Bioinformatics*16, 2000.
- [5] Google. Classificazione: Roc e auc. <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>.
- [6] McNeil B. J. Hanley, J. A. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143, 1982.