Tecnologie e metodi per l'identificazione delle anomalie in sistemi complessi

Rapporto della Borsa di Ricerca

Responsabile scientifico: Prof. Giovanni Stilo
Ricercatore: Dott. Mattia Masci

1. CONTESTO

Negli ultimi anni, la rilevazione delle anomalie si è affermata come una delle aree di ricerca più significative nel campo dell'apprendimento automatico, con applicazioni che spaziano dalla rilevazione delle frodi alla manutenzione predittiva. Le tecniche di "*anomaly detection*" [1] presentano tuttavia limitazioni nel trattare volumi ingenti di dati complessi, soprattutto quando le anomalie risultano rare e i modelli dei dati normali mostrano una considerevole variabilità.

In

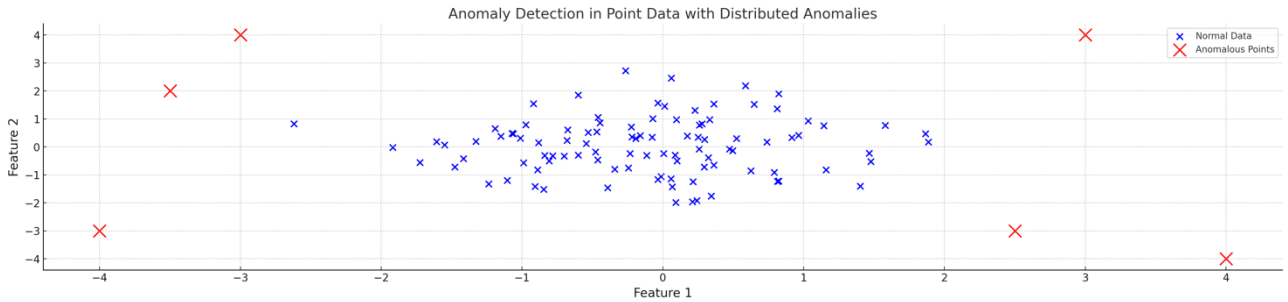

Figura 1 è rappresentato un grafico a dispersione utilizzato per il rilevamento di anomalie, che mostra un insieme di dati bidimensionale. Gli assi del grafico sono etichettati come "Feature 1" e "Feature 2", rappresentando due variabili del set di dati. Le anomalie sono indicate con una "x" rossa di dimensioni maggiori rispetto ai punti normali. Questi punti si trovano lontano dal cluster principale, sparsi nelle regioni periferiche del grafico, distanti dalla concentrazione dei dati normali.

Tra le tecnologie emergenti più promettenti per superare tali difficoltà si annoverano i "*variational autoencoder*" (VAE) [2]. Questi modelli generativi sono capaci di apprendere una rappresentazione compatta (codifica) dei dati, massimizzando una stima della probabilità condizionale del dato osservato. Tuttavia, l'approccio convenzionale dei VAE può risultare inadeguato per la rilevazione delle anomalie, poiché il modello tende a ricostruire anche queste ultime, trattandole come semplici variazioni del dato normale.

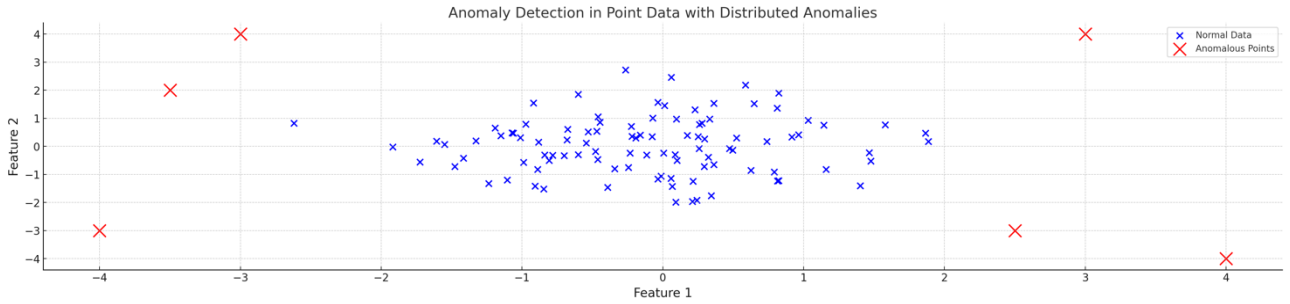

Figura 1. *Esempio di anomalie puntiformi (in rosso).*

Tra le tecnologie emergenti più promettenti per superare tali difficoltà si annoverano i "*variational autoencoder*" (VAE) [2]. Questi modelli generativi sono capaci di apprendere una rappresentazione compatta (codifica) dei dati, massimizzando una stima della probabilità condizionale del dato osservato. Tuttavia, l'approccio convenzionale dei VAE può risultare inadeguato per la rilevazione delle anomalie, poiché il modello tende a ricostruire anche queste ultime, trattandole come semplici variazioni del dato normale.

Per potenziare l'efficacia dei VAE nell'ambito della rilevazione delle anomalie, si propone un approccio innovativo che prevede l'integrazione di esemplari rappresentativi durante le fasi di addestramento e decodifica del modello.

Questo progetto di ricerca mira a indagare l'uso di un "*variational autoencoder*" basato su esemplari (*exemplar-based* VAE, o ex-VAE) [3], con l'obiettivo di migliorare l'efficacia nell'identificazione delle anomalie all'interno di set di dati complessi e squilibrati.

2. FASI DEL PROGETTO

Sviluppo di un Variational Autoencoder basato su esemplari (Exemplar-based VAE)
Il primo obiettivo del progetto è lo sviluppo di un modello VAE che integri esemplari rappresentativi per migliorare la qualità della codifica latente. Gli esemplari agiranno come punti di riferimento nel processo di codifica, permettendo al modello di mappare in modo più preciso le caratteristiche distintive dei dati normali rispetto a quelli anomali.

Ottimizzazione per il task di anomaly detection
Il VAE basato su esemplari verrà specificamente ottimizzato per i compiti di rilevazione delle anomalie. Il modello sarà addestrato per individuare le anomalie come deviazioni dalle rappresentazioni latenti associate agli esemplari normali, migliorando così la capacità del modello di identificare anomalie rare e difficili da rilevare.

Validazione e valutazione delle prestazioni
Il modello sarà sottoposto a valutazione su una serie di dataset di benchmark ampiamente utilizzati nella rilevazione delle anomalie, come KDDCUP99, ALOI e WDBC [4]. L'obiettivo è confrontare le prestazioni del modello proposto con tecniche tradizionali e avanzate di anomaly detection, tra cui Isolation Forest [5], autoencoder convenzionali e altri algoritmi all'avanguardia.

Studio comparativo dell'efficacia
Il progetto prevede un approfondito studio comparativo tra il VAE standard e il modello exemplar-based VAE, utilizzando metriche quali l'Area Under the ROC Curve (AUC) e l'Area Under the Precision-Recall Curve (PR-AUC). L'obiettivo finale è dimostrare che l'inclusione degli esemplari contribuisce in maniera significativa al miglioramento della rilevazione delle anomalie, in particolare su set di dati fortemente sbilanciati.

3. METODOLOGIA

Abbiamo avviato il progetto implementando la struttura del Variational Autoencoder (VAE) in Python, dotandolo di un encoder e un decoder costituiti da strati fully-connected, con dimensioni variabili specificate dinamicamente in funzione della dimensionalità del dataset.

Durante la prima epoca di addestramento, abbiamo effettuato un campionamento uniforme degli esemplari a partire dal dataset disponibile. Per le epoche successive, invece, abbiamo adottato un campionamento ponderato degli esemplari, il cui obiettivo è costruire un insieme di exemplar specifico per ciascuna epoca, basandosi sull'evoluzione dello spazio latente sviluppato fino a quel momento. Il processo di selezione degli esemplari si fonda sul codice latente associato a ogni istanza del dataset, permettendo una variazione dinamica nella scelta degli esemplari man mano che lo spazio latente si raffina. I campioni del dataset vengono campionati con sostituzione, seguendo una funzione di probabilità che assegna un peso maggiore ai campioni con punteggi anomali inferiori nella loro rappresentazione nello spazio latente.

Per assegnare tali probabilità, abbiamo utilizzato l'algoritmo Local Outlier Factor (LOF) [6], sebbene qualsiasi tecnica di outlier scoring possa essere applicata a questo livello. Una rappresentazione dello schema descritto è riportata in **Figura 2**.Una rappresentazione dello schema descritto è riportata in **Figura 2**.

Successivamente, abbiamo sviluppato le routine per l'addestramento e la valutazione, nonché tutte le funzioni essenziali per la lettura e l'elaborazione del dataset, assicurando la piena operatività dei dati all'interno del nostro framework.

Completata l'implementazione delle funzionalità chiave, siamo passati alla valutazione delle prestazioni del nostro modello ex-VAE, impiegando una serie di dataset di benchmark. L'obiettivo era confrontare i risultati con quelli ottenuti da altre tecniche di anomaly detection all'avanguardia (come SAE, OCSVM [7], Hawkins [8], MOGAAL [9], e altre), nonché con un VAE standard.

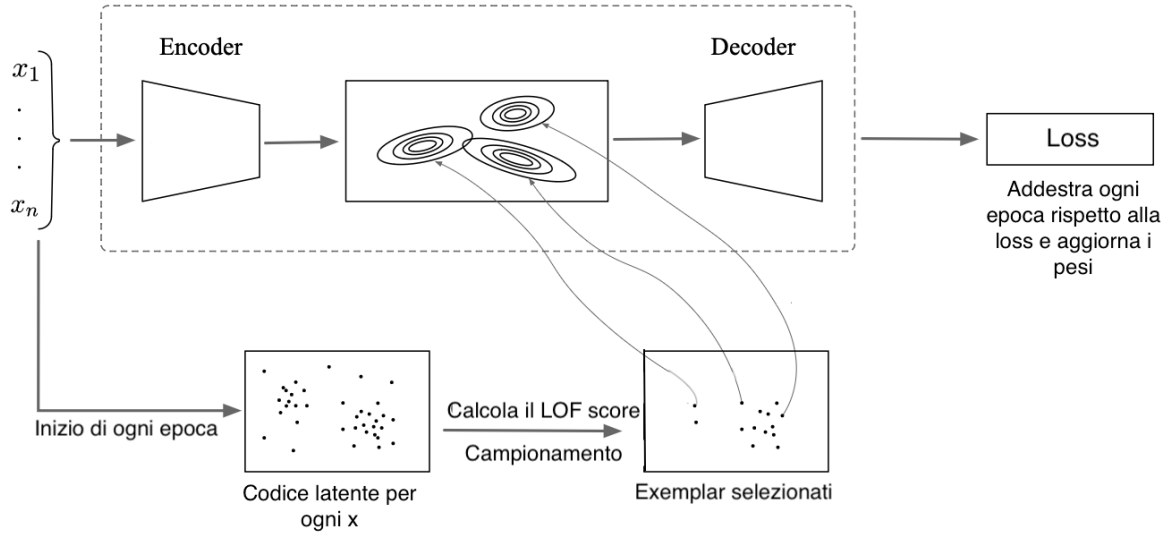

Figura 2. *Selezione degli exemplar nel framework descritto*

Abbiamo addestrato il modello utilizzando diversi dataset e condotto numerose esecuzioni, per ottenere risultati robusti e affidabili. L'addestramento è stato eseguito sempre in modalità non

supervisionata, senza alcuna conoscenza a priori delle istanze inlier o outlier. La valutazione delle prestazioni, invece, è stata condotta in maniera supervisionata, utilizzando metriche quali AUC e PR-AUC.

I risultati della ricerca hanno mostrato che l'ex-VAE è in grado di ottenere prestazioni simili ai migliori modelli di rilevazione delle anomalie, confermandosi una soluzione competitiva. In termini di AUC e PR-AUC, il modello ha raggiunto valori paragonabili a quelli di tecniche avanzate come Isolation Forest, dimostrando un'elevata precisione nell'identificare anomalie e una bassa frequenza di falsi positivi.

L'ex-VAE ha inoltre superato significativamente il VAE standard in entrambe le metriche di valutazione, mostrando una maggiore capacità di rilevare anomalie e riducendo le ricostruzioni errate di dati anomali. L'uso degli esemplari ha permesso al VAE di mantenere una rappresentazione più accurata dei dati normali, riducendo il rischio di considerare le anomalie come semplici variazioni normali, specialmente in contesti in cui le anomalie sono rare.

Per migliorare ulteriormente la capacità della rete neurale di discriminare tra inlier e outlier, abbiamo introdotto uno strato Latent Feature Reconstruction (LFR) [10] prima su un normale autoencoder fully-connected e successivamente sul nostro modello ex-VAE. Tuttavia, nel secondo caso, i test non hanno fornito i risultati sperati e, pertanto, questa soluzione non è stata inclusa nel framework finale.

Oltre ai confronti con altre tecniche di anomaly detection, abbiamo condotto un'analisi di ablazione per valutare l'impatto del numero di esemplari sul processo di addestramento e sulle prestazioni del modello. Abbiamo eseguito esperimenti variando la quantità di esemplari, partendo da percentuali del 5%, 10%, 20%, 30%, 40%, fino al 50%, e testando anche con un numero fisso di 50 esemplari. Infine, abbiamo sperimentato una funzione che calcolava la quantità ottimale di esemplari da selezionare in base al numero di istanze nel dataset.

4. RISULTATI

I risultati della nostra ricerca evidenziano come il metodo proposto, basato sull'utilizzo di esemplari in un opportuno variational autoencoder (exemplar-based VAE), rappresenti un significativo avanzamento nel campo della rilevazione delle anomalie.

Il modello exemplar-based VAE ha dimostrato prestazioni comparabili ai migliori modelli attualmente disponibili nello stato dell'arte e superiori rispetto a un VAE standard, in particolare in contesti caratterizzati da dati sbilanciati. Inoltre, il modello ha evidenziato una maggiore capacità di generalizzazione, riducendo la tendenza a ricostruire erroneamente le anomalie come semplici variazioni dei dati normali.

In termini specifici, il metodo proposto ha raggiunto un valore medio di PR-AUC pari al 27,56%, superando in modo significativo il VAE standard (PR-AUC pari a 20,87%) e risultando comparabile ai metodi basati su Isolation Forest (PR-AUC pari a 28,55%). Tuttavia, va sottolineato che, sebbene l'Isolation Forest si sia dimostrato efficace, esso non beneficia delle rappresentazioni latenti apprese, limitandosi a isolare le anomalie sulla base di strutture dati in alta dimensionalità.

Questi risultati suggeriscono che l'approccio basato su esemplari possa avere un notevole potenziale applicativo in ambiti pratici quali la sicurezza informatica, la rilevazione di frodi e il monitoraggio industriale.

Alla luce dei risultati ottenuti, siamo attualmente impegnati nella redazione di un articolo scientifico, il quale presenterà in dettaglio la nostra metodologia, i risultati sperimentali e un'analisi comparativa

con altri modelli avanzati di anomaly detection. L'articolo sarà sottoposto a revisione per la possibile inclusione nel programma del ACM Symposium on Applied Computing (SAC) 2025, uno degli eventi internazionali di maggiore rilievo nel campo dell'informatica applicata.

5. ULTERIORI ATTIVITA' DI RICERCA

Parallelamente sotto la guida del Prof. Stilo il dottor Mattia Masci, ha affrontato una delle problematiche più rilevanti nel campo dell'intelligenza artificiale applicata: l'ottimizzazione del processo di addestramento delle reti neurali profonde (DNN) [11]. Le DNN, caratterizzate da migliaia di parametri regolabili, richiedono elevate risorse computazionali e lunghi tempi di esecuzione, ostacolando la diffusione e l'adozione di queste tecnologie in applicazioni pratiche. Per risolvere questo problema, si sono proposte due metodologie innovative, denominate Cryo-SC e Cryo-Seq, che permettono di accelerare il processo di training congelando e scongelando dinamicamente gli strati della rete durante l'addestramento, ottimizzando così l'allocazione delle risorse senza compromettere l'accuratezza del modello.

Le tecniche Cryo-SC e Cryo-Seq si basano su strategie di freezing dinamico. Nel caso di Cryo-SC, i layer della rete vengono selezionati casualmente per essere congelati, mentre Cryo-Seq congela sequenzialmente i layer a partire dal primo strato della rete. Entrambi i metodi consentono di ridurre il carico computazionale mantenendo elevate prestazioni. Le tecniche sono state validate su due architetture di reti neurali diverse, Long Short-Term Memory (LSTM) [12] e Visual Geometry Group (VGG) [13], utilizzando due dataset rappresentativi: UCI-HAR [14] per il riconoscimento delle attività umane e CIFAR-10 [15] per la classificazione delle immagini. Gli esperimenti hanno dimostrato riduzioni significative nei tempi di addestramento, rispettivamente del 26,42% per LSTM e dell'8,69% per VGG, con perdite minime in termini di accuratezza, evidenziando una diminuzione massima dell'1,79% rispetto al metodo di training classico.

La ricerca ha quindi incluso uno studio comparativo approfondito tra le nuove tecniche Cryo-SC e Cryo-Seq e i metodi di freezing già presenti in letteratura, come Freezing Rate Schema e LayerOut, utilizzando metriche quali il tempo di esecuzione, l'accuratezza e l'efficienza computazionale. Cryo-Seq, in particolare, ha dimostrato di essere il metodo più efficace in termini di riduzione dei tempi di addestramento, mantenendo al contempo una precisione comparabile. Dalle analisi di ablazione, è emerso che Cryo-Seq offre i maggiori vantaggi nei contesti con grandi dataset e reti neurali profonde, dove i guadagni in termini di riduzione del tempo di addestramento sono particolarmente rilevanti, senza significative perdite di accuratezza.

I risultati raggiunti suggeriscono che queste tecniche potrebbero contribuire significativamente alla democratizzazione dell'intelligenza artificiale, rendendo il deep learning più accessibile e sostenibile, specialmente per organizzazioni con risorse computazionali limitate. L'approccio proposto, inoltre, ha potenziali applicazioni in scenari industriali su larga scala, dove l'efficienza nel training delle reti neurali è cruciale per implementazioni pratiche. Alla luce dei risultati ottenuti, il team ha redatto un articolo scientifico, attualmente in fase di revisione, per la presentazione alla AAAI Conference on Artificial Intelligence 2025, un'importante piattaforma internazionale per discutere scoperte innovative nel campo dell'AI e ricevere i commenti dagli esperti del settore.

BIBLIOGRAFIA

[1] C. S. L. C. A. V. D. H. GUANSONG PANG, «Deep Learning for Anomaly Detection: A Review,» 2020.

[2] M. W. Diederik P Kingma, «Auto-Encoding Variational Bayes,» 2018.

[3] D. J. F. M. N. Sajad Norouzi, «Exemplar VAE: Linking Generative Models, Nearest Neighbor Retrieval, and Data Augmentation,» 2020.

[4] [Online]. Available: https://www.dbs.ifi.lmu.de/research/outlier-evaluation/DAMI/.

[5] K. M. T. Z.-H. Z. Fei Tony Liu, «Isolation Forest,» 2008.

[6] H.-P. K. ,. R. T. N. ,. J. S. Markus M. Breunig, «LOF: Identifying Density-Based Local Outliers,» 2000.

[7] J. C. P. J. S.-T. A. J. S. R. C. W. Bernhard Schölkopf, «Estimating the Support of a High-Dimensional Distribution,» 1999.

[8] H. H. G. W. R. B. Simon Hawkins, «Outlier Detection Using Replicator Neural Networks,» 2002.

[9] Z. L. C. Z. Y. J. J. S. M. W. X. H. Yezheng Liu, «Generative Adversarial Active Learning for Unsupervised Outlier Detection,» 2019.

[10] Y. H. W. X. J. G. J. Z. S. Z. Jinghuang Lin, «Latent feature reconstruction for unsupervised anomaly detection,» 2023.

[11] J. Schmidhuber, «Deep Learning in Neural Networks: An Overview,» 2014.

[12] J. S. Sepp Hochreiter, «Long Short-Term Memory,» 1997.

[13] K. S. &. A. Zisserman, «VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,» 2015.

[14] A. G. L. O. X. P. J. L. R.-O. Davide Anguita, «A public domain dataset for human activity recognition using smartphones,» 2013.

[15] A. Krizhevsky, «Learning Multiple Layers of Features from Tiny Images,» 2009.