Individuazione delle Anomalie: da problema a opportunità

Alessandro Lazzeri, Responsabile R&D at Deepclever S.r.l. by Polaris Engineering S.r.l.

Le tecniche di individuazione delle anomalie (AD, anomaly detection) appartengono alla disciplina del data mining e permettono di trovare eventi rari che si differenziano sensibilmente dalla maggior parte dei dati in un dataset [1]. Individuare velocemente eventi anomali, che si manifestano raramente o che non si sono mai realizzati nel passato, può consentire un intervento reattivo e accurato anticipando l'evolversi di situazioni dannose o la perdita di opportunità. Queste tecniche sono utilizzate per individuare comportamenti non leciti su internet come gli attacchi hacker, truffe assicurative, frodi bancarie, problemi strutturali, sorveglianza, malfunzionamenti, problemi di salute medica e così via.

Cosa è una anomalia?

La principale difficoltà nell'individuare anomalie è la definizione stessa di anomalia. In generale un evento è anomalo se si discosta dal comportamento atteso. Ad esempio, nel lancio di una moneta l'esito atteso è di ottenere un risultato di testa o croce e una moneta che resta in bilico è un evento alquanto anomalo/improbabile. Il problema è a dir poco complesso perché la definizione stessa di anomalia può essere difficile da dare e può variare da un contesto a un altro: un aumento del traffico di dati attraverso una rete informatica potrebbe indicare un attacco hacker e una violazione della sicurezza dei dati, una notizia o un'immagine potrebbero essere alterate per diffondere notizie false, le transazioni bancarie o su carte di credito a seconda degli importi e dei conti possono aiutare a identificare le frodi, e così via. Talvolta, anziché partire dal definire le anomalie è opportuno cercare di definire qual è la fattispecie normale, ad esempio le caratteristiche di un animale domestico come un cane, e identificare come anomalo tutto ciò che cade al di fuori della definizione. Questa inversione semplifica il problema e apre spiragli per l'utilizzo di numerose strategie di risoluzione, ma non lo risolve completamente perché:

Figura 2: Una piccola anomalia genetica: una coda tra gli occhi di questo cucciolo [4].

Figura 1: Alcuni modelli stimano che la probabilità di ottenere una caduta di taglio sia 1 su 6000 [3].

- Definire il comportamento normale può essere complesso e per alcune caratteristiche potrebbe non essere possibile definire delle soglie precise, ad esempio "le transazioni sopra i *10000 €* sono anomale" è una definizione corretta per le vendite di un negozio di generi alimentari, ma non lo è per una concessionaria di automobili;
- In contesti contrastanti, ad esempio quando un truffatore vuole far apparire normale un evento anomalo, i comportamenti malevoli si adattano e diventa più difficile individuarli;
- In contesti dinamici il concetto di normale è valido in un determinato periodo di tempo e non è detto che sia rappresentativo anche nel futuro;
- I criteri di anomalia cambiano da un dominio applicativo ad un altro: ad esempio una variazione percentuale della temperatura corporea può essere fisiologicamente normale mentre la stessa variazione percentuale del prezzo azionario di un titolo può essere un'anomalia;
- Spesso c'è carenza di annotazioni dei dati e difficoltà nel verificare le capacità dei sistemi di identificazione delle anomalie;

- Il problema può sovrapporsi a due problemi simili: la presenza/rimozione del rumore e il verificarsi di eventi nuovi. Nel primo caso il rumore è una componente dovuta all'osservazione del fenomeno che rende più difficile la corretta identificazione (es., un evento normale osservato in condizioni rumorose potrebbe sembrare anomalo o viceversa); nel secondo caso, un evento raro mai verificatosi nel passato potrebbe sembrare anomalo pur essendo normale;

Caratteristiche del compito

Tre importanti aspetti sono rilevanti per la struttura e l'impostazione dei compiti di AD:

- Natura del dato: il dato è l'insieme di attributi che descrivono un fenomeno o un evento. Gli attributi possono essere di tipo numerico continuo, categorico e binario. In base alla quantità e i tipi di attributi che compongono il dato sono individuate tecniche appropriate per AD. Inoltre, i dati – le osservazioni del fenomeno – possono essere in relazione tra loro, come per esempio avviene nelle serie storiche, oppure nei dati geo-referenziati e nelle strutture a grafo.
- Tipo di anomalia
    - Puntuale: un'anomalia è puntuale se dipende unicamente dal singolo dato, e.g., una temperatura ambientale di 45°C.
    - Contestuale: un'anomalia è contestuale se è condizionata dalla situazione in cui si manifesta, e.g., una temperatura di 28°C il 3 gennaio a Trento.
    - Collettivo: un'anomalia è collettiva se un insieme di osservazioni è anomalo, e.g., una prolungato aumento della temperatura nel tempo come in Figura 3.
- Annotazioni
    - Supervisionato: i dati comprendono osservazioni sia normali sia anomale e sappiamo ciascuna osservazione a quale classe appartiene;
    - Semi-supervisionato: i dati a disposizione contengono solo osservazioni normali.
    - Non supervisionato: non sono a disposizione informazioni sui dati, i quali possono essere solo normali o misti. L'unica assunzione è che i dati normali sono più numerosi di quelli anomali.
- Risultato:
    - Label: il risultato è binario nel dominio normale/anomalo;
    - Punteggio: a ciascun dato è assegnato un punteggio che descrive il grado di anomalia.

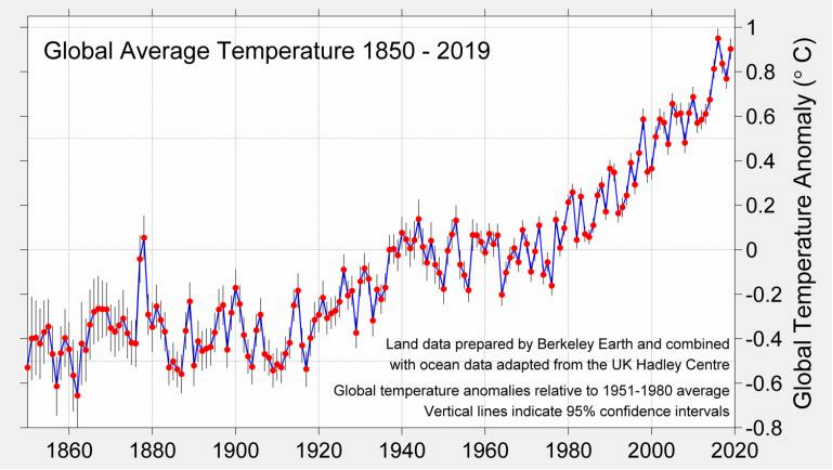

Figura 3: Gli ultimi 40 anni sono caratterizzati da una tendenza anomala di incremento della temperatura media globale [5].

Approcci

- Classificazione: nei compiti di classificazione l'obiettivo è addestrare un modello su un dataset di addestramento, i.e. un insieme di dati normali e anomali sui quali il modello impara a fare distinzione, e testarlo su un dataset di verifica, i.e. dati che il modello classifica normali o anomali, sul quale si valutano le prestazioni del modello.
- Nearest Neighbor: le tecniche di "vicinato" si basano sul concetto di distanza o similarità tra coppie di dati. Si distinguono in: K-nearest neighbor, in cui lo score di anomalia di un nuovo dato dipende dalle *k* istanze più vicine presenti nel dataset; densità relativa, lo score dipende dalla densità nell'intorno del nuovo dato.
- Clustering: il clustering è un insieme di tecniche non supervisionate per raggruppare le istanze di un dataset in gruppi in cui: le istanze appartenenti allo stesso gruppo sono simili tra loro e le istanze in gruppi diversi sono dissimili. In AD si assume che le istanze che cadono all'interno dei gruppi sono considerate normali e le istanze che cadono al di fuori sono anomale.
- Statistiche: in generale le tecniche statistiche si basano sulla stima della funzione di probabilità di osservare un determinato dato in una regione dello spazio. I dati che si osservano in regioni dello spazio con alta probabilità sono considerati normali, mentre i dati che si osservano in regioni con bassa probabilità sono anomali.

Robust Deep Autoencoders

Nell'ambito dell'AD la letteratura scientifica è composta da molte tecniche che possono essere applicate nei contesti precedentemente descritti e descriverle tutte in dettaglio, specificando assunzioni, pregi e difetti, va ben oltre l'obiettivo di questo articolo. Quindi, senza nulla togliere alla validità e all'efficacia di tecniche omesse in questo articolo, nella letteratura recente una tecnica di Deep Learning molto promettente sono i Robust Deep Autoencoder (RDA) [2]. Gli RDA nascono dall'unione dei Deep Autoencoders (AE) e dalla Robust Principal Component Analysis (RPCA).

Deep Autoencoders

Un Deep Autoencoder è una rete multi-strato feed-forward che mappa l'input su sé stesso. Gli autoencoders sono fatti da due sottoreti: l'encoder $E()$e il decoder $D()$. La caratteristica che rende non banale questa architettura è la ridotta dimensione dello strato nascosto di neuroni, fattore che impedisce l'operazione di identità tra input X e output $\bar{X}$, dove:

$$\bar{X} = D(E(X))$$

e l'obiettivo è trovare $E()$e il $D()$ tali che l'errore di ricostruzione sia minimo secondo:

$$\min_{E,D} L(X - \bar{X})$$

con $L()$ in genere la norma-2.

Robust Principal Component Analysis

La RPCA è una generalizzazione dell'analisi delle componenti principali (PCA) che ci prefigge come obiettivo la riduzione della sensibilità in presenza di outliers. Più nel dettaglio, in RPCA il dataset in ingresso X è diviso in due matrici: la matrice L con basso rango e la matrice S sparsa tali che:

$$X = L + S$$

In questo modo, la matrice a basso range L rappresenta la maggior parte di dati nel dataset, mentre la matrice S cattura le anomalie.

La decomposizione delle matrici può essere calcolata risolvendo il seguente problema di ottimizzazione:

$$\min_{L,S} \rho(L) + \lambda \|S\|_0$$

$$s.t.\ \|X - L - S\|_F^2 = 0$$

dove: $\rho(L)$ è il rango di L , $\|S\|_0$ è il numero di valori non-0 in S e $\| \quad \|_F$ è la norma di Frobenius. Sfortunatamente, il problema è non-convesso (NP-completo) e non può essere trattato per matrici di grandi dimensioni, ma può essere rilassato nella seguente forma:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1$$

$$s.t.\ \|X - L - S\|_F^2 = 0$$

dove $\| \quad \|_*$ è la norma matriciale e $\| \quad \|_1$ la norma-1.

Metodo

I RDA dividono i dati come RPCA in due parti L_D e S sfruttando un deep autoencoder per ricostruire con basso errore la porzione maggiore L_D e scartando rumore e anomalie in S. Il problema può essere scritto nella seguente forma:

$$\min_{\theta} \|L_D - D_\theta(E_\theta(L_D))\|_2 + \lambda \|S\|_1$$

$$s.t. X - L_D - S = 0$$

dove $\| \quad \|_2$ è la norma-2 e $\| \quad \|_1$ la norma-1, D_θ e E_θ, sono il decoder e l'encoder, e λ un parametro che regola la sparsità di S: un valore piccolo farà aumentare il numero di outlier individuati e un valore più grande li farà ridurre. La regolarizzazione $\lambda\|S\|_1$ può essere sostituita con $\lambda\|S\|_{2,1}$, dove $\| \quad \|_{2,1}$ è la norm-2,1 così calcolata:

$$\|X\|_{2,1} = \sum_{j=1}^{n} \|x_j\|_2$$

La norma-2,1 ha il vantaggio di poter individuare anomalie sia singolo attributo sia anomalie su gruppi.

Infine, per individuare istanze anomale è possibile impostare il problema nel seguente modo:

$$\min_{\theta} \|L_D - D_\theta(E_\theta(L_D))\|_2 + \lambda \|S^T\|_{2,1}$$

$$s.t. X - L_D - S = 0$$

dove S^T è la trasposizione della matrice S .

Infine, la procedura di addestramento è composta da due parti che si ripetono fino a convergenza: la prima parte in cui $\lambda\|S^T\|_{2,1}$ è costante e l'autoencoder $\|L_D - D_\theta(E_\theta(L_D))\|_2$ è addestrato con back-propagation e la seconda parte in cui l'autoencoder è costante e $\lambda\|S^T\|_{2,1}$ è ottimizzato con proximal gradient ($\lambda\|S^T\|_{2,1}$ non è differenziabile). Si rimanda a [2] per maggiori dettagli.

Applicazioni

Alta incertezza, fenomeni complessi, attività dinamiche e contesti competitivi sono alcune delle caratteristiche che risaltano dagli ambiti di applicazione delle tecniche di AD. Lo studio dei casi storici è spesso fonte di informazioni che permettono di definire regole e policy per prevenire molte situazioni anomale, ma risultano spesso insufficienti e difficili da aggiornare. Per questo motivo è quasi sempre necessario affiancare ai propri sistemi delle tecniche di AD. I seguenti paragrafi illustreranno brevemente alcune applicazioni di AD in diversi settori.

Individuazione di intrusioni in computer e reti

Gli attacchi informatici sono delle procedure messe in atto da soggetti o gruppi di hacker volti al danneggiamento o al furto di informazioni da sistemi informatici. In generale, i pirati informatici sfruttano dei difetti dei sistemi, falle nella sicurezza, sistemi non aggiornati ed errori umani per poter mettere in atto gli attacchi e danneggiare i bersagli designati, che siano aziende o privati. Gli strumenti e le strategie a disposizione dei pirati informatici sono in continuo sviluppo e quando l'illecito è stato compiuto e si può analizzare cosa è accaduto ormai è troppo tardi: l'attacco è stato completato, i sistemi sono stati danneggiati e/o i dati sono stati trafugati all'esterno.

Individuazione di frodi

Le frodi o truffe sono dei comportamenti illeciti volti al conseguimento di profitti e possono coinvolgere diversi settori come le banche, le assicurazioni, tasse, ma anche social network, vendite online e telefoniche. Gli autori delle frodi cercano di trarre in inganno la controparte mascherando un comportamento truffaldino come un comportamento perfettamente lecito. Come per le intrusioni informatiche, molte tecniche possono essere adattate con varianti più o meno consistenti a differenti contesti e per questo motivo è molto difficile individuarle con precisione e velocità.

Sanità

Nel settore medicale l'avanzamento della tecnologia permette di eseguire esami sempre più accurati e dettagliati. Questo enorme passo in avanti consente sia ai medici di avere più informazioni per valutare lo stato dei propri pazienti ed effettuare delle diagnosi più precise sia agli enti e alle case farmaceutiche di pianificare gli investimenti e la ricerca nel medio e lungo periodo. La grande mole di dati a disposizione deve però essere visionata ed analizzata con cura e questo richiede molta attenzione e tempo che purtroppo non sempre è disponibile. In questo settore i sistemi di AD in ambito medico permettono di individuare anomalie nel battito cardiaco e nelle immagini prodotte da radiografie e TAC.

Industriale

In ambito industriale gli strumenti e i macchinari sono sottoposti a continuo stress e logoramento a causa del continuo utilizzo e alle condizioni di funzionamento. Sebbene le linee guida dei produttori permettano di pianificare le operazioni di manutenzione e sostituzione per garantire la sicurezza e l'operatività è sempre presente un grado di incertezza che troppe volte produce perdite economiche e talvolta incidenti sul lavoro.

Analisi di Immagini e Audio

In linea con le necessità di analizzare immagini nel settore medico, gli strumenti di sorveglianza e i social network sono fonti di immagini, video e audio che possono essere analizzati per individuare illeciti, violazioni di norme di comportamento o situazioni pericolose.

Analisi del Testo

L'analisi di anomalie nel testo riguarda principalmente l'individuazione di nuovi argomenti di discussione, opinioni ed eventi in collezioni di documenti. Le collezioni di documenti possono avere natura diversa come: messaggi e post scritti sui social network, forum, pagine web e articoli di giornali quotidiani o riviste, ma anche documentazione aziendale, schede tecniche, bilanci e altro.

Reti di Sensori

Le reti di sensori sono delle reti distribuite di dispositivi elettronici per la raccolta di dati da un determinato ambiente. Le reti di sensori sono ormai diffuse in tutti i settori, dal militare alla sanità, industriale e domestico e così via. L'ambiente in cui sono inserite è spesso complesso e imprevedibile per cui l'individuazione di anomalie è quasi sempre un requisito standard per individuare malfunzionamenti, intrusioni, variazioni dei consumi energetici e così via.

Conclusione

In questo articolo è stato descritto il concetto di anomalia e le difficoltà che occorrono nell'utilizzo di approcci standard per la loro individuazione; purtroppo, talvolta non è possibile conoscere le caratteristiche che identificano le anomalie e ancora più spesso non è conveniente elencare tutte le caratteristiche di normalità. La natura del dato raccolto riguardo un fenomeno è un elemento cruciale per descrivere l'anomalia e, di conseguenza, per poter scegliere una o più tecniche di apprendimento appropriate per risolvere il problema. Una tecnica promettente, basata su deep learning e analisi delle componenti principali, è il Robust Deep Autoencoder che permette di individuare diversi tipi di anomalie in modo non supervisionato. Le tecniche non supervisionate sono una scelta frequente nei compiti di AD proprio per la natura del problema che comporta una carenza di dataset annotati. Infine, le tecniche di apprendimento sono spesso utilizzate trasversalmente, con le dovute accortezze, per l'individuazione di anomalie e i benefici sono stati dimostrati in numerose di applicazioni, come ad esempio industriale, informatica, medico, sorveglianza e molti altri ancora.

Riferimenti

[1] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 1-58.

[2] Zhou, C., & Paffenroth, R. C. (2017, August). Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 665-674).

[3] https://en.wikipedia.org/wiki/Coin_flipping

[4] Picture of the Dog with a tail on his head: https://www.irishtimes.com/news/offbeat/puppy-with-second-tail-on-his-head-found-wandering-streets-1.4082714

[5] Global warming: http://berkeleyearth.org/archive/2019-temperatures/