

Quantum computing per i sistemi antifrode: i risultati del progetto condotto da TAS con l'Università di Verona

1 Sintesi del progetto

In un'epoca caratterizzata dal proliferare delle transazioni digitali e dalla crescente dipendenza dai metodi di pagamento elettronici, il problema delle frodi mediate da carte di credito appare come una sfida critica per le istituzioni finanziarie, le imprese e i consumatori. **Il progetto di ricerca avviato a Novembre 2022 da TAS e dall'Università di Verona si è proposto di affrontare questo problema sviluppando nuovi potenti algoritmi di contrasto agli illeciti che sfruttano il paradigma di calcolo definito dalla computazione quantistica.**

In particolare sono tre le limitazioni principali che affliggono i sistemi automatizzati basati su algoritmi di machine learning utilizzati oggi per il rilevamento delle frodi:

- un elevato costo computazionale per l'ottimizzazione dei relativi iper-parametri;
- limitazioni nell'individuazione di nuovi pattern di pagamento fraudolenti (nel momento in cui si manifestano) e nell'integrazione di nuova conoscenza mediata dalle ultime transazioni elaborate (online-learning);
- capacità di individuare transazioni fraudolente buona ma non ottimale.

Il progetto si è focalizzato sul terzo punto ossia sul miglioramento delle performance dei modelli di rilevamento delle frodi. L'idea proposta ed oggetto di studio, è stata quella di caratterizzare i pagamenti sfruttando i metodi analitici messi a disposizione dalla **topological data analysis (TDA)**, e già applicati con successo per il rilevamento di anomalie in serie temporali. Questo tipo di analisi risulta assai dispendiosa (in termini di risorse computazionali) se eseguita per mezzo di dispositivi classici, ma può essere velocizzata sfruttando dispositivi e algoritmi quantistici. I risultati dello studio, descritti più approfonditamente di seguito, hanno portato alla definizione di un nuovo approccio all'analisi delle transazioni con carte di credito che può essere sfruttato per analizzare come l'ultimo pagamento effettuato da un dato utente vada ad impattare sulla struttura topologica associata allo storico dei pagamenti effettuati dall'utente stesso fino a quel punto.

La variazione della struttura topologica del dato, catturata tramite funzioni (*Betti curves*) calcolate mediante il computer quantistico, rappresenta un'informazione utile a determinare se l'ultimo pagamento sia fraudolento o meno.

Ulteriori studi, una migliore strumentazione e dataset di test più ampi potranno confermare l'utilità dell'applicazione del quantum computing ai sistemi di individuazione e contrasto delle frodi nei pagamenti.

1.1 Topological Data Analysis e Calcolo Quantistico

In questa sezione verrà descritto ad alto livello il contenuto dell'articolo "*Higher-order topological kernels via quantum computation.*" Incudini et.al *arXiv:2307.07383*. frutto dell'attività di ricerca svolta durante il corso del progetto. **Il documento descrive una possibile**

pipeline di elaborazione dei dati che consente di estrarre features topologiche (o attributi topologici) associate a dati complessi come serie temporali, nuvole di punti, immagini e grafi. Il soggetto principale dell'articolo è, di fatto, la *topological data analysis* (TDA), disciplina a cui fanno capo metodi avanzati usati per l'analisi delle proprietà geometriche e di connettività dei dati. Si basa sulla topologia, un ramo della matematica che esplora le proprietà di oggetti e forme rappresentandoli come spazi topologici e caratterizzandoli tramite descrittori invarianti rispetto a trasformazioni continue (i.e. non cambiano se applichiamo all'oggetto analizzato trasformazioni come rotazioni, traslazioni e rescaling).

Molti sono gli strumenti analitici utili che rientrano nell'ambito della TDA. **Tra questi troviamo l'analisi dei Betti numbers: un metodo algoritmico che consente di caratterizzare un oggetto contando il numero di buchi n-dimensionali contenuti nello spazio topologico ad esso**

associato. Per capire cosa si intende con "buco" di dimensione (o ordine) "n" si possono fare degli esempi pratici. I buchi di ordine 0 sono le componenti connesse dello spazio topologico (es. 2 punti staccati sono 2 componenti connesse distinte o 2 buchi di ordine 0); un buco di ordine 1 è un ciclo (es. una circonferenza è un esempio di buco di ordine 1); lo spazio vuoto interno ad una superficie chiusa in uno spazio tridimensionale corrisponderà infine ad un buco di ordine 2 (es. spazio vuoto racchiuso da una sfera cava). Negli esempi, i 2 punti, la circonferenza, e la sfera cava, sono tutti oggetti associabili a spazi topologici distinti e a loro volta distinguibili contando i corrispondenti Betti numbers di vario ordine.

In questo contesto è possibile effettuare un'analisi ancor più sofisticata considerando una sequenza di spazi topologici e analizzando come cambia la conta dei Betti numbers dall'uno all'altro. Queste variazioni possono essere catturate, in senso matematico, da delle funzioni, le *multivariate Betti curves*, che descrivono l'andamento dei Betti numbers di vario ordine lungo la sequenza di spazi topologici. Nella pratica tali sequenze sono ottenibili a partire da nuvole di punti sfruttando un metodo di elaborazione dei dati detto *filtration*.

Tramite questo tipo di analisi siamo quindi in grado di associare ad una nuvola di punti una funzione che ne cattura le caratteristiche topologiche. Definendo una formula per il calcolo della distanza tra due Betti Curves multi-variate, riusciamo ad ottenere un valore di dissimilarità utilizzabile per definire un *kernel* ossia una funzione matematica che misura quanto i due oggetti di partenza (in questo caso le nuvole di punti) sono simili. I kernel sono strumenti particolarmente utili nel contesto del *machine learning* (quella sotto-branca dell'intelligenza artificiale dove i dati vengono sfruttati per addestrare algoritmi su task come predizione, pattern recognition o anomaly detection). Tramite i kernel è infatti possibile migliorare le performance di molti modelli di machine learning sviluppati per risolvere tasks di classificazione (il modello impara dai dati come associare un elemento ad una categoria) e regressione (il modello impara dai dati ad associare un elemento ad un valore). **In particolare, sfruttando il kernel basato sulle multivariate Betti curves, possiamo allenare un modello in grado di categorizzare oggetti secondo le loro caratteristiche topologiche.**

Un problema rilevante quando si effettua questa analisi è legato alle elevate risorse computazionali che si devono avere a disposizione per poter effettuare tutti i calcoli.

Il conteggio dei Betti numbers è stato dimostrato essere un compito oneroso da risolvere usando metodi classici; risulta quindi conveniente ottenere una stima dei Betti Numbers sfruttando gli appositi algoritmi quantistici definiti in letteratura che consentono di rendere più efficiente l'intero processo.

Esistono vari approcci quantistici per la stima dei Betti number, ciascuno con un grado diverso di complessità e in grado di fornire più o meno elevati speed-up rispetto ai metodi classici. In ogni caso, è bene considerare un importante elemento critico che accomuna tutti questi

metodi. **In generale, questi algoritmi, per essere correttamente eseguiti, hanno bisogno di avere accesso a risorse quantistiche fault-tolerant (i.e. resilienti rispetto alle problematiche che affliggono i moderni sistemi quantistici: rumore, basso tempo di coerenza, ecc.).**

Dispositivi di questo tipo non sono ancora disponibili e ciò rende l'approccio proposto non di facile applicazione. Tuttavia, il problema del calcolo di Betti number, pur rimanendo oneroso, risulta più semplice da risolvere rispetto all'uso dei soli metodi classici.

Il vantaggio che si ottiene con il calcolo quantistico lo si ha in particolare per il calcolo dei Betti number associati a buchi di ordine (dimensione) elevato. **Di conseguenza, considerando i vincoli di tempo propri di ciascuna applicazione pratica, otterremo che l'algoritmo quantistico sarà in grado di calcolare i Betti numbers di ordine superiore, ai quali non avremo accesso se sfruttassimo solo le risorse classiche.**

Riassumendo la pipeline di elaborazione dei dati proposta nello studio prevede il susseguirsi dei seguenti step:

1. Codifica dei dati (immagini, serie temporali, grafi...) di un dataset sotto forma di nuvole di punti.
2. Generazione di sequenze di spazi topologici (una per ogni elemento nel dataset) sfruttando un filtration method.
3. Calcolo delle multivariate Betti curves e della relativa funzione kernel sfruttando gli algoritmi di calcolo quantistici che permettono di analizzare attributi topologici di ordine elevato.
4. Addestramento di un modello di machine learning in grado di fornire output in base alle informazioni topologiche associate ai dati.

La pipeline così definita potrebbe trovare applicazione in svariati campi come pattern recognition, classificazione di immagini e grafi, oppure anomaly detection su time series. È quest'ultimo l'ambito applicativo su cui si sono concentrati gli sforzi del successivo periodo di ricerca essendo questo il problema più vicino a quello della credit card fraud detection.

1.2 TDA e Credit Card Fraud Detection

La pipeline introdotta nella sezione precedente si è detto essere applicabile a diversi problemi a seconda del tipo di dato che ci si trova a dover maneggiare. **Nel caso della fraud detection, i dati non appaiono direttamente nel formato ideale, ma è possibile elaborare il dataset di pagamenti in modo tale da ottenere delle serie temporali associate a ciascuna transazione e analizzabili per mezzo del framework sopra descritto.** In particolare, prendendo in considerazione una specifica transazione, si può considerare lo storico dei pagamenti effettuati dallo stesso utente fino a quel punto e valutare come la struttura topologica di questa serie di valori cambia a fronte dell'ultimo pagamento. **La variazione della struttura topologica del dato, catturata tramite le Betti curves calcolate sfruttando algoritmi quantistici, potrebbe rappresentare un'informazione utile a determinare qualora l'ultimo pagamento sia fraudolento o meno.**

Questo assunto deve ancora essere dimostrato teoricamente e a livello sperimentale. Un primo test di fattibilità è stato portato avanti per stabilire quali sono le condizioni necessarie perchè l'approccio topologico si riveli utile per il caso d'uso d'interesse. In particolare, l'approccio proposto è stato utilizzato per calcolare delle feature topologiche da aggiungere ai normali attributi descrittivi delle transazioni. Questi dati aggiuntivi hanno il compito di



quantificare quanto la struttura topologica del dato sia stata influenzata dal pagamento in analisi. **Più in generale, l'obiettivo è quello di rendere i dati più rappresentativi, informativi e adatti all'analisi. La progettazione accurata del metodo di calcolo delle feature topologiche può portare a migliorare le prestazioni del modello di machine learning utilizzato per la credit card fraud detection.**

1.3 Considerazioni

Ci sono ancora diverse domande aperte e sfide da affrontare per poter confermare l'utilità dell'approccio proposto. **Con il futuro sviluppo di nuove teorie e una migliore strumentazione (dispositivi quantistici fault tolerant) sarà possibile valutare l'effettiva applicabilità di questi metodi.**