

MIGLIORAMENTO CONTINUO

SOMMARIO

CICLI DI FEEDBACK	2
FASI	2
POSSIBILI PROBLEMI	3
Cause del collasso del modello	4

CICLI DI FEEDBACK

Un ciclo di feedback AI è un processo iterativo in cui le decisioni e gli output di un modello AI vengono continuamente raccolti e utilizzati per migliorare o riaddestrare lo stesso modello, con conseguente apprendimento, sviluppo e miglioramento continui del modello. In questo processo, i dati di addestramento, i parametri del modello e gli algoritmi del sistema di intelligenza artificiale vengono aggiornati e migliorati in base all'input generato dall'interno del sistema. Principalmente ci sono due tipi di cicli di feedback AI: 1. Cicli di feedback AI positivi: Quando i modelli di intelligenza artificiale generano risultati accurati in linea con le aspettative e le preferenze degli utenti, gli utenti forniscono un feedback positivo tramite un ciclo di feedback, che in cambio rafforza l'accuratezza dei risultati futuri. Tale ciclo di feedback è definito positivo. 2. Cicli di feedback AI negativi: Quando i modelli di intelligenza artificiale generano risultati imprecisi, gli utenti segnalano i difetti tramite un ciclo di feedback che in cambio cerca di migliorare la stabilità del sistema correggendo i difetti. Tale ciclo di feedback è definito negativo. Entrambi i tipi di cicli di feedback AI consentono lo sviluppo continuo del modello e il miglioramento delle prestazioni nel tempo. E non vengono utilizzati o applicati in modo isolato. Insieme, aiutano i modelli di intelligenza artificiale implementati in produzione a sapere cosa è giusto o sbagliato.

FASI

Comprendere come funzionano i cicli di feedback dell'IA è importante per sbloccare l'intero potenziale dello sviluppo dell'IA. Esploriamo le varie fasi dei cicli di feedback AI di seguito.

1. Raccolta di feedback: Raccogliere i risultati del modello rilevanti per la valutazione. In genere, gli utenti forniscono il loro feedback sul risultato del modello, che viene quindi utilizzato per il riaddestramento. Oppure possono essere dati esterni dal Web curati per ottimizzare le prestazioni del sistema.

2. Riqualficazione del modello: Utilizzando le informazioni raccolte, il sistema AI viene riaddestrato per fare previsioni migliori, fornire risposte o svolgere attività particolari affinando i parametri o i pesi del modello.
3. Integrazione e test del feedback: Dopo il riaddestramento, il modello viene testato e valutato nuovamente. In questa fase, è incluso anche il feedback degli esperti in materia (PMI) per evidenziare problemi al di là dei dati.
4. distribuzione: Il modello viene ridistribuito dopo aver verificato le modifiche. In questa fase, il modello dovrebbe riportare prestazioni migliori sui nuovi dati del mondo reale, con conseguente miglioramento dell'esperienza utente.
5. Monitoraggio: Il modello viene monitorato continuamente utilizzando metriche per identificare il potenziale deterioramento, come la deriva. E il ciclo di feedback continua. La creazione di solidi sistemi di intelligenza artificiale richiede una comprensione approfondita dei potenziali problemi nei dati di produzione (dati del mondo reale) e nei risultati del modello

POSSIBILI PROBLEMI

Diamo un'occhiata ad alcuni problemi che diventano un ostacolo per garantire l'accuratezza e l'affidabilità dei sistemi di intelligenza artificiale:

1. Deriva dei dati: Si verifica quando il modello inizia a ricevere dati del mondo reale da una distribuzione diversa rispetto alla distribuzione dei dati di training del modello.
2. Deriva del modello: Le capacità predittive e l'efficienza del modello diminuiscono nel tempo a causa del cambiamento degli ambienti del mondo reale. Questo è noto come deriva del modello.
3. Risultati del modello AI rispetto alla decisione del mondo reale: I modelli di intelligenza artificiale producono output imprecisi che non sono in linea con le decisioni delle parti interessate nel mondo reale.

4. Pregiudizio ed equità: I modelli di intelligenza artificiale possono sviluppare problemi di pregiudizio e correttezza. Ad esempio, nell' Intervento TED di Janelle Shane, descrive la decisione di Amazon di smettere di lavorare su un algoritmo di ordinamento dei curriculum a causa della discriminazione di genere.

Sulla scia della rapida adozione dell'IA generativa, i ricercatori hanno studiato un fenomeno noto come Crollo del modello. Definiscono il collasso del modello come: "Processo degenerativo che colpisce generazioni di modelli generativi appresi, dove i dati generati finiscono per inquinare il training set della prossima generazione di modelli; essendo addestrati su dati inquinati, poi percepiscono male la realtà. Model Collapse è costituito da due casi speciali,

- Crollo del primo modello si verifica quando "il modello inizia a perdere informazioni sulle code della distribuzione", ovvero le estremità estreme della distribuzione dei dati di addestramento.
- Crollo tardivo del modello accade quando il "modello intreccia diverse modalità delle distribuzioni originali e converge a una distribuzione che porta una piccola somiglianza con quella originale, spesso con una varianza molto piccola".

Cause del collasso del modello

Affinché i professionisti dell'IA affrontino questo problema, è essenziale comprendere le ragioni del Model Collapse, raggruppate in due categorie principali:

1. Errore di approssimazione statistica: Questo è l'errore principale causato dal numero finito di campioni e scompare man mano che il conteggio dei campioni si avvicina all'infinito.
2. Errore di approssimazione funzionale: Questo errore si verifica quando i modelli, come le reti neurali, non riescono a catturare la vera funzione sottostante che deve essere appresa dai dati. Quando i modelli di intelligenza artificiale si addestrano sui

contenuti generati dall'IA, ha un effetto distruttivo sui cicli di feedback dell'IA e può causare molti problemi ai modelli di intelligenza artificiale riaddestrati, come ad esempio:

- **Modello Comprimi:** Come spiegato sopra, Model Collapse è una probabile possibilità se il ciclo di feedback AI contiene contenuti generati da AI.
- **Dimenticanza catastrofica:** Una tipica sfida nell'apprendimento continuo è che il modello dimentica i campioni precedenti quando apprende nuove informazioni. Questo è noto come oblio catastrofico.
- **Inquinamento dei dati:** Si riferisce all'inserimento di dati sintetici manipolativi nel modello AI per compromettere le prestazioni, spingendolo a produrre output imprecisi. In che modo le aziende possono creare un robusto ciclo di feedback per i loro modelli di intelligenza artificiale? Le aziende possono trarre vantaggio dall'utilizzo di cicli di feedback nei loro flussi di lavoro AI. Segui i tre passaggi principali di seguito per migliorare le prestazioni dei tuoi modelli di intelligenza artificiale.
- **Feedback dagli esperti in materia:** Le PMI sono altamente competenti nel loro dominio e comprendono l'uso dei modelli di intelligenza artificiale. Possono offrire approfondimenti per aumentare l'allineamento del modello con le impostazioni del mondo reale, offrendo maggiori possibilità di risultati corretti. Inoltre, possono governare e gestire meglio i dati generati dall'IA.
- **Scegli le metriche di qualità del modello pertinenti:** La scelta della giusta metrica di valutazione per l'attività giusta e il monitoraggio del modello in produzione sulla base di queste metriche possono garantire la qualità del modello. I professionisti dell'IA utilizzano anche strumenti MLOps per la valutazione e il monitoraggio automatizzati per avvisare tutte le parti interessate se le prestazioni del modello iniziano a deteriorarsi in produzione.
- **Cura rigorosa dei dati:** Poiché i modelli di produzione vengono riaddestrati su nuovi

dati, possono dimenticare le informazioni passate, quindi è fondamentale curare dati di alta qualità che si allineino bene con lo scopo del modello. Questi dati possono essere utilizzati per addestrare nuovamente il modello nelle generazioni successive, insieme al feedback degli utenti per garantire la qualità.