



Carlo Niccolai (carlopavana)

I DATI. ALCUNE DEFINIZIONI PRIMA DI APPROFONDIRE L'INTELLIGENZA ARTIFICIALE.

8 May 2022

Argomenti trattati

Uno sguardo al mondo dei dati

Qualche definizione.

Una striscia di storia dei dati

L'analisi e la codifica dei dati qualitativi

Bot: significato e definizione.

Uno sguardo al mondo dei dati.

Ogni secondo nel mondo vengono generati e memorizzati in modo elettronico miliardi di dati. Sicuramente con l'andare del tempo questa massa, inizialmente informe di dati, tende ad aumentare e solo con le metodologie dell'Intelligenza Artificiale potranno essere analizzati e tornare utili ad aziende, comunita', regioni o stati. Con questo documento cerco di porre una base di definizioni e argomenti per l'approfondimento delle tecniche di Intelligenza Artificiale che cercherò di sviluppare sia analizzando le professionalita' richieste che i concetti di coscienza indispensabili per non soccombere alle tecnologie in divenire.

Un argomento che sta prendendo sempre di più la scena è quello relativo alle problematiche che sottintendono al rapporto tra microeconomia del lavoro e robotica, cioè come si deve evolvere la società dei lavoratori con l'invasione, peraltro già in atto, della robotica supportata dalle metodologie di intelligenza artificiale.

Questo materiale è parte di una mia consulenza del gennaio 2022.

Qualche definizione.

I dati

I dati sono ovunque, in qualsiasi punto volgiamo lo sguardo ci sono dati, la natura, tutta la natura, e' un insieme di dati se per dato assumiamo la definizione:

Un dato e' la rappresentazione codificata di qualsiasi cosa. Il dato diventa informazione quando viene usato.

Da questo deriva che e' necessaria una codifica, un sistema che ci permetta di identificare, ricordare, valutare, confrontare e rappresentare un evento o un oggetto sia in modo assoluto che in modo comparativo.

I dati qualitativi.

Ma se questa attivita' e' relativamente facile per eventi in qualche modo misurabili, diventa molto problematica quando l'analisi riguarda un fatto che non puo' essere misurato ma solo valutato, come la fame, il dolore, l'amore o la felicita', cioe' di tipo qualitativo

Big Data.(grandi masse di dati, o in italiano megadati)

Con questa locuzione si intende, in statistica e informatica, una raccolta di dati così estesa in termini di volume, velocita' e varieta' da richiedere tecnologie e metodi analitici specifici per l'estrazione dei loro valori o conoscenza.

L'intelligenza artificiale.

Non esiste una sola definizione di "intelligenza artificiale", ma possiamo dire che l'Intelligenza Artificiale e' un insieme di tecnologie e processi in grado di elaborare grandi quantita' di informazioni per restituire sintesi o previsioni complesse e razionali, sia cercando di imitare i processi cognitivi umani che esplorando nuove metodologie di pensiero e di ragionamento.

Alcuni linguaggi di programmazione usati nell'Intelligenza Artificiale

C

Il linguaggio C (in inglese The C Programming Language) venne chiamato informalmente anche K&R, dalle iniziali dei cognomi degli autori di un noto libro d'informatica scritto da Brian Kernighan e Dennis Ritchie. Il linguaggio C negli anni è diventato uno standard de facto.

C++

E' un linguaggio di programmazione general purpose sviluppato in origine da Bjarne Stroustrup nei Bell Labs nel 1983 come evoluzione del linguaggio C inserendo la programmazione orientata agli oggetti. Col tempo ha avuto notevoli evoluzioni e l'ultima versione dello standard e' stata pubblicata nel 2017.

Java

E' un linguaggio di programmazione ad alto livello, orientato agli oggetti e alla tipizzazione statica, che si appoggia sull'omonima piattaforma software di esecuzione. E' specificamente progettato per

essere il più possibile indipendente dalla piattaforma hardware di esecuzione tramite compilazione in bytecode prima e interpretazione poi da parte di una JVM (Java Virtual Machine).

Java Script

E' un linguaggio di programmazione orientato agli oggetti e agli eventi, comunemente utilizzato nella programmazione Web lato client (esteso poi anche al lato server) per la creazione, in siti e applicazioni web, di effetti dinamici interattivi tramite funzioni di script invocate da eventi innescati a loro volta in vari modi dall'utente sulla pagina web in uso (mouse, tastiera, caricamento della pagina ecc...).

Julia

E' un linguaggio di programmazione, multi-paradigma, open source pubblicato a partire dal 2012 (anche se le basi furono poste nel 2009) da un gruppo di programmatori impegnati professionalmente nel campo del calcolo numerico pensato con lo scopo di fornire uno strumento ad alte prestazioni dedicato in maniera particolare al mondo della programmazione scientifica.

Python

E' un linguaggio di programmazione dinamico orientato agli oggetti utilizzabile per molti tipi di sviluppo software. Offre un forte supporto all'integrazione con altri linguaggi e programmi, e' fornito con una estesa libreria standard e può essere imparato in pochi giorni. Molti programmatori Python possono confermare un sostanziale aumento di produttività e ritengono che il linguaggio incoraggi lo sviluppo di codice di qualità e manutenibilità superiori. Il codice Python può essere compilato utilizzando interpreti standard e viene compilato come PyPy e CPython.

R

E' un linguaggio di programmazione e un ambiente di sviluppo specifico per l'analisi statistica dei dati. Venne scritto inizialmente dal matematico e statistico canadese Robert Gentleman, e dallo statistico neozelandese Ross Ihaka.

SAS (Statistical Analysis System)

E un complesso di prodotti software integrati, sviluppati dal SAS Institute, che permettono ad un programmatore una notevole gamma di lavori nella gestione dei dati, nella statistica e nella gestione di supporti alle decisioni solo per citare quelle più importanti, senza dimenticare la gestione delle risorse umane, il supporto alle decisioni e la gestione finanziaria.

Scala (Scalable Language)

E' un linguaggio di programmazione di tipo general-purpose multi-paradigma studiato per integrare le caratteristiche e funzionalità dei linguaggi orientati agli oggetti e dei linguaggi funzionali. La compilazione di codice sorgente Scala produce Java bytecode per l'esecuzione su una JVM.

SQL (Structured Query Language)

E' un linguaggio di programmazione utilizzato per creare, modificare e gestire i dati in un database relazionale. Si tratta nello specifico di un linguaggio specifico di dominio (DSL) usato per comunicare con i sistemi di gestione di database relazionali (RDBMS).

Una striscia di storia dei dati



scheda-perforata.gif

All'inizio della informatizzazione, attorno agli anni 50 del 1900, i tecnici si resero subito conto delle diverse tipologie di dati e da bravi esseri razionali, hanno tranquillamente ignorato che la quasi totalita' dei dati non era codificabile, ma tutto quello che era contabilita' denaro, gestione e numeri aziendali furono facilmente informatizzati. Così' arrivarono le schede perforate. E' del tutto evidente che il totale delle fatture puo' essere sommato, mentre la fedelta' all'azienda no.

L'informazione codificabile puo' essere rappresentata in due modalita' distinte: analogica e digitale.

Nel sistema analogico le informazioni sono rappresentate da un insieme continuo di valori. per esempio la voce umana puo' essere rappresentata con segnali elettrici che variano in modo pressoché continuo. Poiché esiste una analogia tra l'informazione da rappresentare e i valori della grandezza adottata per la rappresentazione, questa modalita' e' appunto detta analogica. Le grandezze naturali sono solitamente di tipo analogico: voce, immagini, testo.

Nel sistema digitale per rappresentare le informazioni si utilizzano entita' distinte, che variano in modo non continuo (discreto). In particolare, e' molto diffuso il sistema digitale binario che usa soli due simboli: {0,1}. Ovviamente questi simboli devono essere organizzati in modo opportuno, cioè' devono rispettare una determinata codifica. Il calcolatore numerico utilizza i sistemi digitali binari per eseguire l'elaborazione delle informazioni, quindi e' necessario trasformare le informazioni

analogiche in digitali binarie. Questo processo di trasformazione e' definito codifica. Quindi si parte dalle informazioni in forma analogica, si applica una codifica e si giunge alle informazioni digitali che possono essere elaborate da un calcolatore elettronico.



Le informazioni che il computer puo' elaborare sono di diversa natura:

- Alfabetica (tutti i caratteri dell'alfabeto, caratteri speciali, punteggiatura ecc.)
- Numerica (numeri naturali, relativi, razionali, reali, complessi)
- Sonora (audio)
- Grafica (Immagini)
- Sonora e Grafica (video)

Per ognuno di questi tipi e' necessario avere una particolare codifica (codice) per trasformare il dato di partenza in una sequenza di bit elaborabili.

Sistemi di numerazione

Lo scopo di un sistema di numerazione e' quello di rappresentare con delle cifre un numero. Il sistema che utilizziamo quotidianamente e' il sistema decimale che fa uso di 10 simboli diversi che chiamiamo cifre. Si chiama sistema decimale perche', per esempio, il numero 1834 puo' essere scritto

$$1*10^3+8*10^2+3*10^1+4*10^0$$

cioe' come sommatoria di potenze del 10. Di fondamentale importanza e' la posizione delle cifre perche' cambia il peso associato. Tutto questo e' un concetto astratto; nulla ci vieta di cambiare la regola e di rappresentare i numeri utilizzando un numero diverso di cifre, una base diversa o quant'altro per semplificare le operazioni.

La codifica binaria

Nei computer che tutti usiamo i circuiti elettronici gestiscono due stati logici codificati ed interpretati come 0 e 1 che e' chiamata codifica binaria, cioe' una numerazione a basata sulla potenza del 2. Il codice binario e' un codice a lunghezza variabile perche' il numero di bit dipende da quanto e' grande il numero naturale che si deve codificare.

La codifica ottale

Il sistema numerico ottale e' un sistema numerico posizionale che utilizza le potenze dell'otto. I simboli usati sono 8 (0, 1, 2, 3, 4, 5, 6, 7).

Attualmente e' una curiosita' di archeologia informatica e lo cito solo per completezza.

La codifica esadecimale

Per compattare l'informazione binaria al fine di manovrarla senza dover gestire un numero molto grande di simboli si utilizza la base 16, che e' un'altra potenza di due, per rendere il passaggio tra le basi esaminate molto semplice. Le cifre a disposizione sono 16 quindi ne abbiamo 6 in piu' rispetto al sistema decimale pertanto aggiungiamo alle cifre le prime 6 lettere dell'alfabeto (0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F).

Le codifiche alfanumeriche

Codici ASCII (American Standard Code for Information Interchange)

La tabella ASCII e' un codice convenzionale usato per la rappresentazione dei caratteri di testo attraverso i byte: In realta' lo standard ASCII copre solo i primi 128 valori (da 00000000 a 01111111), i successivi valori fino al 256 costituiscono la tabella ASCII estesa che presenta varie versioni a carattere nazionale.

Codici Unicode

Con la diffusione di Internet, e con essa la necessita' di codificare simboli provenienti da alfabeti nazionali diversi, si e' sviluppata la codifica Unicode, che era stato originariamente pensata come una codifica a 16 bit (quattro cifre esadecimali) che dava la possibilita' di codificare 65.536 caratteri (216 combinazioni diverse). successivamente si e' previsto l'uso di codifiche con unita' da 8 bit (byte), 16 bit (word) e 32 bit (double word), descritte rispettivamente come UTF-8, UTF-16 e UTF-32

La codifica delle immagini

Anche le immagini analogiche per essere trattate da computer devono essere trasformate in una sequenza di bit attraverso una opportuna codifica.

Raster

Nella codifica raster all'immagine analogica si sovrappone una griglia fittissima di piccoli quadratini denominati pixel. In pratica la griglia e' costituita da m righe e n colonne, pertanto in totale si hanno $m \cdot n$ pixel che rappresentano la risoluzione dell'immagine.

GIF (Graphics Interchange Format):

Molto utilizzata per applicazioni in internet, riduce di molto la dimensione pero' comporta due difetti: le immagini possono avere al massimo 256 colori, la stampa di tali immagini e' di pessima qualita'.

JPEG (Joint Photographic Experts Group).

Serve per visualizzare immagini con piu' di 256 colori, e di grandi dimensioni, definendo il livello di compressione in modo da determinare la dimensione e la qualita' dell'immagine voluta.

PNG (Portable Network Graphics)

E' un tipo di immagine digitale compressa della categoria bitmap, molto usata sia sul web che in numerosi applicativi grafici. Le immagini PNG raccolgono i pregi sia delle immagini GIF (ma non le loro animazioni) sia delle immagini in formato JPEG, sono compresse e possono diventare relativamente piccole in termini di spazio, usano una compressione che non ne danneggia particolarmente la qualita' (la stessa delle immagini GIF): cio' significa, in parole povere, che una immagine PNG non si sfuochera' facilmente e non avra' sbavature o altri difetti grafici tipici invece delle immagini JPEG.

Le immagini vettoriali

Sono particolarmente utilizzate quando l'immagine da rappresentare e' riconducibile a insiemi di forme come: punti, linee, rettangoli, cerchi. E' caratterizzata dal fatto che non si memorizza l'immagine ma il procedimento per costruirla, di conseguenza si hanno due vantaggi: si diminuisce enormemente l'occupazione di memoria e le immagini sono facilmente ridimensionabili

La codifica dei suoni

I suoni costituiscono un tipo di informazione con cui siamo costantemente a contatto (linguaggio parlato, musica, rumori). Anche i suoni possono essere gestiti in forma digitale, ma e' necessario trovare un modo per rappresentare in forma numerica un'onda sonora. Le fasi fondamentali per la digitalizzazione del suono sono:

- Campionamento
- Codifica in binario dei dati campionati

Si effettuano dei campionamenti sull'onda sonora (cioe' si misura il valore dell'ampiezza e della frequenza dell'onda a intervalli costanti di tempo) e si codificano in forma digitale le informazioni estratte da tali campionamenti. Quanto piu' frequentemente il valore dell'onda viene campionato, tanto piu' precisa sara la sua rappresentazione. Riducendo l'intervallo di tempo tra due

campionamenti aumenta l'accuratezza della descrizione del segnale. La codifica non è altro la rappresentazione binaria del valore del campione.

Conclusione

La gestione e la elaborazione dei dati codificati e trattati in modalità discreta ha raggiunto un limite che difficilmente permetterà grandi progressi, ma le aziende necessitano di elaborare quelli che sono chiamati dati qualitativi.

L'analisi e la codifica dei dati qualitativi

Per dati qualitativi si intendono i dati non strutturati che di solito si riferiscono a materiale testuale come recensioni, risposte aperte a sondaggi, reclami, messaggi di chat, interviste ai clienti, annotazioni di eventi o post sui social media.

Quasi sempre vengono raccolti da metodi di ricerca qualitativa, come domande di indagine aperte, interviste approfondite, focus group, osservazione diretta e analisi dei contenuti (video, immagine o documento). I dati raccolti sono in genere sotto forma di testo digitato, trascrizioni o registrazioni che devono essere esaminati per identificare temi e approfondimenti.

Il processo di esame e interpretazione dei dati qualitativi è noto come analisi dei dati qualitativi. La codifica dei dati qualitativi è il processo di creazione e assegnazione di codici per poterli classificare e successivamente saranno utilizzati per derivare temi e modelli per analizzarli. La codifica e l'analisi possono aver luogo simultaneamente, ma è importante notare che la codifica non implica necessariamente l'identificazione di temi. Al contrario, si riferisce generalmente al processo di etichettatura e raggruppamento di tipi di dati simili per rendere più gestibile la generazione di temi e procedere poi all'analisi.

Una definizione parallela ci dice che la codifica dei dati qualitativi è il processo di etichettatura e organizzazione per identificare diversi temi e le relazioni tra i dati. L'analisi qualitativa dei dati è il processo di esame e interpretazione dei dati qualitativi per capire cosa rappresentano.

Codifica deduttiva

La codifica deduttiva o codifica basata sui concetti, significa che si inizia con un insieme predefinito di codici, quindi si assegnano tali codici ai nuovi dati qualitativi.

I ricercatori iniziano con un insieme predefinito di codici o un libro di codici sviluppato prima di analizzare i dati della ricerca. Questo insieme potrebbe essere basato sulle domande di ricerca o su una struttura o teoria di ricerca esistente. ad esempio, se la domanda di ricerca è il motivo per cui un consumatore ha acquistato un prodotto specifico, il ricercatore potrebbe predefinire un elenco di codici che include prezzo, qualità, marca, ecc. Tenendo presente questo elenco, il ricercatore leggerà i dati della ricerca e assegnerà semplicemente i codici predefiniti.

Codifica induttiva

La codifica induttiva, detta anche codifica aperta, parte da zero e crea codici basati sui dati qualitativi da codificare. E' un processo iterativo, il che significa che richiede piu' tempo della codifica deduttiva, ma offre uno sguardo piu' completo e imparziale ai temi nei dati.

La codifica induttiva implica la creazione di un elenco di codici o di un libro di codici da zero in base ai dati della ricerca, e' spesso piu' difficile ma puo' essere meno soggetta a pregiudizi rispetto alla codifica deduttiva, perche' non si inizia il processo di analisi con nozioni preconcepite.

Nella pratica gli studi di ricerca spesso combinano la codifica deduttiva e induttiva, partendo da un elenco predefinito di codici, ma poi modificandolo induttivamente e aggiungendolo a tale elenco man mano che l'analisi ne consegue.

La codifica dei dati qualitativi puo' essere manuale o automatica

La codifica automatizzata utilizza un software di analisi dei dati qualitativi per codificarli e analizzarli rapidamente. Il software sfrutta l'apprendimento automatico, l'intelligenza artificiale e l'elaborazione del linguaggio naturale per determinare i temi e creare codici senza alcuna configurazione o pianificazione anticipata. Gli algoritmi imparano mentre procedono.

La codifica manuale rimane ancora popolare grazie alla sua maggiore precisione percepita, ma l'automazione sta rapidamente diventando la scelta preferita. a differenza dell'analisi manuale, che e' soggetta a distorsione e non si adatta alla quantita' di dati che attualmente viene generata. L'automazione dell'analisi non e' solo piu' coerente e quindi puo' essere piu' accurata, ma puo' anche far risparmiare molto tempo e quindi denaro.

Analisi del contenuto.

Si riferisce alla categorizzazione, all'etichettatura e all'analisi tematica dei dati qualitativi. Cio' puo' includere la combinazione dei risultati dell'analisi con i dati comportamentali per ottenere informazioni piu' approfondite.

Analisi narrativa

Alcuni dati, come interviste o note sul campo, possono contenere una storia. ad esempio, il processo di scelta di un prodotto, utilizzo, valutazione della sua qualita' e decisione di acquistare o meno questo prodotto la prossima volta. L'analisi narrativa aiuta a comprendere gli eventi sottostanti e il loro effetto sul risultato complessivo.

Analisi del discorso.

Si riferisce all'analisi di cio' che le persone dicono nel contesto sociale e culturale, ed e' particolarmente utile quando ci si concentra sulla costruzione o sul rafforzamento di un marchio.

Analisi del quadro.

Quando si esegue un'analisi qualitativa dei dati puo' essere utile disporre di un quadro di riferimento, una cornice di codici, un insieme gerarchico di temi utilizzati nella codifica.

Grounded theory.

Questo metodo di analisi inizia formulando una teoria attorno a un singolo caso di dati. Pertanto la teoria e' "fondata" su dati reali, ed ulteriori casi possono essere esaminati per vedere se sono rilevanti e possono aggiungersi alla teoria originale.

Codifica ed analisi tematica.

La codifica tematica, chiamata anche analisi tematica, e' un tipo di analisi qualitativa dei dati che trova i temi nel testo analizzando il significato delle parole e la struttura delle frasi, ed e' tipica per semplificare l'interpretazione del feedback dei clienti. L'assegnazione di codici a parole e frasi in ogni risposta aiuta a catturare di cosa tratta la risposta e, a sua volta, aiuta ad analizzare e riassumere meglio i risultati dell'intero sondaggio.

Quando, ad esempio, si utilizza questa tecnica per analizzare il feedback dei clienti, si puo' scoprire quali temi sono piu' frequenti riuscendo a capire cosa guida la soddisfazione del cliente in modo accurato e attuabile.

Cornice di codifica: piatta e gerarchica.

Una cornice di codifica piatta assegna lo stesso livello di specificita' e importanza a ciascun codice. Sebbene questo possa sembrare un metodo piu' semplice e veloce per la codifica manuale, puo' essere difficile organizzare e navigare tra i temi e i concetti mentre si crea un numero sempre maggiore di codici. Inoltre, rende difficile capire quali temi sono piu' importanti, il che puo' rallentare il processo decisionale.

Una cornice di codifica gerarchica aiuta ad organizzare i codici in base a come si relazionano tra loro. Esempio Il codice di primo livello descrive l'argomento (servizio clienti) Il codice di livello intermedio specifica se il sentimento e' positivo o negativo Il terzo livello descrive in dettaglio l'attributo o il tema specifico associato all'argomento. La cornice gerarchica supporta un insieme di codici piu' ampi e consente di organizzarli in base alla struttura organizzativa, e diversi livelli di granularita' nella codifica.

Codifica in vivo

La codifica in vivo fa uso delle parole dei partecipanti piuttosto che della interpretazione dei dati. Si usano le citazioni dirette dei partecipanti come codici ed e' particolarmente utile quando i dati derivano da partecipanti che parlano lingue diverse o provengono da culture diverse. In questi casi, e' spesso difficile dedurre con precisione il significato a causa di differenze linguistiche o culturali.

Codifica descrittiva

La codifica descrittiva mira a riassumere estratti utilizzando una singola parola o sostantivo che incapsula l'idea generale dei dati. Queste parole in genere descrivono i dati in modo altamente condensato, il che consente di fare rapidamente riferimento al contenuto.

Codifica strutturale

La codifica strutturale implica l'etichettatura e la descrizione di attributi strutturali specifici dei dati. In genere, include la codifica in base alle risposte alle domande " chi ", " cosa ", " dove " e " come ", piuttosto che agli argomenti effettivi espressi nei dati. Questo tipo di codifica e' utile quando si desidera accedere rapidamente a segmenti di dati e puo' essere di grande aiuto quando si ha a che fare con insiemi di dati di grandi dimensioni.

Codifica dei valori

La codifica dei valori implica una codifica che si riferisce alle visioni del mondo dei partecipanti . Tipicamente, questo tipo di codifica si concentra su estratti che riflettono i valori, gli atteggiamenti e le convinzioni dei partecipanti. La codifica dei valori e' quindi molto utile per la ricerca che esplora i valori culturali e intrapersonali, le esperienze e le azioni.

Bot: significato e definizione

Un "bot", abbreviazione di robot, e' un programma software che esegue attivita' automatizzate, ripetitive e predefinite. I bot, che in qualche modo fanno parte dei modelli di Intelligenza Artificiale, in genere imitano o sostituiscono il comportamento di utenti umani. Poiche' sono automatizzati, operano molto piu' velocemente delle persone reali. Svolgono funzioni utili come il servizio clienti o l'indicizzazione dei motori di ricerca, ma possono anche presentarsi sotto forma di malware per ottenere il controllo totale di un computer o di una intera rete di computer.

I bot Internet vengono anche chiamati spider, crawler o bot Web, sono essenzialmente strumenti digitali e, come qualsiasi strumento, sono utilizzabili per ogni genere di scopo.

Spambot

Gli spambot possono raccogliere indirizzi e-mail dalle pagine dei contatti o delle presenze degli ospiti. In alternativa, possono pubblicare contenuti promozionali nei forum o nelle sezioni dei commenti per indirizzare il traffico a siti Web specifici.

Chatterbot

I siti Web e le app dei servizi di dating sono il paradiso dei chatterbot dannosi. Questi chatterbot fingono di essere una persona, emulano l'interazione umana e spesso ingannano le persone che

non si rendono conto di chattare con programmi dannosi il cui obiettivo e' ottenere informazioni personali da vittime ignare, ad esempio il numero della carta di credito.

Bot per la condivisione di file

Questi bot catturano i termini cercati da un utente (ad esempio il titolo di un film o di un album) e rispondono alla query affermando di avere il file disponibile per il download e fornendo un collegamento. L'utente fa clic sul collegamento, scarica il file, lo apre e infetta inconsapevolmente il proprio computer.

Stuffing di credenziali

Si riferisce ai bot che immettono nelle pagine di accesso online grandi quantita' di nomi utente e password noti (di solito provenienti da violazioni di dati) per ottenere l'accesso non autorizzato ai rispettivi account utente.

Bot DoS - (Denial of Service) e DDoS(Distribute Denial of Service)

In questo caso il traffico bot eccessivo viene intenzionalmente utilizzato per sovraccaricare le risorse di un server e impedire il funzionamento di un servizio.

Bot Denial of inventory

Questi attacchi prendono di mira i negozi online e li inducono a presentare i prodotti come "non disponibili". In questo tipo di attacco, i bot dannosi selezionano gli articoli dal negozio online e li aggiungono al carrello senza mai completare la transazione. Di conseguenza, quando un utente legittimo desidera acquistare il prodotto, riceve un messaggio di esaurimento scorte, anche se l'articolo e' disponibile.

Bot Scanner di vulnerabilita'

I bot che scansionano milioni di siti alla ricerca di vulnerabilita' e li segnalano al loro creatore sono noti come scanner di vulnerabilita'. a differenza dei bot autentici che informerebbero il proprietario del sito Web, questi bot dannosi sono specificamente creati per inviare segnalazioni a persone che le rivenderanno o le useranno per hackerare siti Web.

Bot per frodi di clic

Questi bot producono un'enorme quantita' di traffico dannoso mirato specificamente agli annunci a pagamento per mettere in atto frodi pubblicitarie. Responsabile di clic fraudolenti su annunci a pagamento, questo traffico automatizzato costa agli inserzionisti miliardi ogni anno ed e' spesso mascherato da traffico legittimo. Senza un buon software di rilevamento dei bot, questa attivita' puo' costare agli inserzionisti gran parte del loro budget pubblicitario.

Monitoraggio del traffico

Bot utilizzati per sovraccaricare i server di posta o per effettuare furti di dati su larga scala. Perché i criminali informatici utilizzano i bot?

- Per rubare informazioni finanziarie e personali Gli hacker possono utilizzare le botnet per inviare phishing o altre truffe per indurre i consumatori a pagare. Possono anche raccogliere informazioni dalle macchine infettate da bot e utilizzarle per rubare identità e sostenere prestiti o spese di acquisto a nome dell'utente.
- Per attaccare servizi Web legittimi I criminali possono utilizzare le botnet per creare attacchi DoS e DDoS che inondano un servizio o una rete legittimi con un volume di traffico schiacciante. Un tale volume potrebbe rallentare gravemente il servizio o la capacità di risposta della rete dell'azienda oppure sovraccaricare completamente il servizio o la rete dell'azienda e disattivarli.
- Per estorcere denaro alle vittime Le entrate derivanti dagli attacchi DoS provengono da estorsione (del tipo "paga o ti chiudiamo il sito") o tramite pagamenti da parte di gruppi interessati a infliggere danni a un'azienda o a una rete. Questi gruppi includono "hacktivisti", hacker con intenti politici, e organizzazioni militari e di intelligence straniere.
- Per guadagnare da sistemi zombie e botnet I criminali informatici possono anche affittare le loro botnet ad altri criminali che desiderano inviare spam, truffe, phishing, rubare identità e attaccare siti Web e reti legittime.

Chatbot

Bot che simulano la conversazione umana rispondendo a determinate frasi con risposte programmate.

Bot social, operano su piattaforme di social media,

Utilizzati per generare automaticamente messaggi, sostenere idee, agire come follower o come account falsi per ottenere follower a loro volta. Man mano che i social network si fanno più sofisticati, diventa sempre più difficile per i social bot creare account falsi. È difficile identificare i bot social perché possono mostrare comportamenti simili a quelli di utenti reali.

Bot per acquisti

Bot che fanno acquisti online per trovare il prezzo migliore del prodotto cercato da un utente. Alcuni bot possono osservare i modelli di comportamento di un utente durante la navigazione in un sito Web e quindi personalizzare quel sito per l'utente.

Spider bot o crawler Web

Bot che scansionano i contenuti delle pagine Web per aiutare Google e altri motori di ricerca a capire il modo migliore di rispondere alle query di ricerca degli utenti. Gli spider scaricano HTML e altre risorse, come CSS, JavaScript e immagini, e le utilizzano per elaborare il contenuto del sito.

Crawler di Web scraping

Bot che leggono i dati dai siti Web con l'obiettivo di salvarli offline e consentirne il riutilizzo. Questa attività può assumere la forma di scraping dell'intero contenuto delle pagine Web o di contenuti mirati per ottenere punti dati specifici, come nomi e prezzi di prodotti sui siti Web di e-commerce. In alcuni casi lo scraping è legittimo e può essere consentito dai proprietari stessi dei siti Web. In altri casi, gli operatori di bot potrebbero violare i termini di utilizzo del sito Web o rubare materiale sensibile o protetto da copyright.

Knowbot

Bot che raccolgono informazioni per gli utenti visitando automaticamente i siti Web in cerca di informazioni sulla base di determinati criteri.

Bot di monitoraggio

Bot utilizzati per monitorare lo stato di salute di un sito Web o di un sistema. Downtetector.com è un esempio di un sito indipendente che fornisce informazioni sullo stato in tempo reale, interruzioni incluse, di siti Web e altri tipi di servizi.

Bot transazionali

Bot utilizzati per completare transazioni per conto di esseri umani. ad esempio, i bot transazionali consentono ai clienti di effettuare una transazione nel contesto di una conversazione.

Bot di download

Bot utilizzati per scaricare automaticamente software o app mobili. Possono essere utilizzati per manipolare le statistiche di download, ad esempio per ottenere più download su app store popolari e aiutare le nuove app ad apparire in cima alle classifiche.

Possono anche essere utilizzati per attaccare i siti di download, creando download falsi come parte di un attacco Denial of Service (DoS).

Bot di ticketing

Bot che acquistano automaticamente biglietti per eventi popolari, con l'obiettivo di rivenderli a scopo di lucro. Questa attività è illegale in molti paesi e, anche quando non è illegale, può dare

fastidio a organizzatori di eventi, legittimi venditori di biglietti e consumatori. I bot di ticketing sono spesso sofisticati ed emulano i comportamenti degli acquirenti.

Bot per computer e bot Internet - Vantaggi:

Piu' veloci delle persone nei compiti ripetitivi. Fanno risparmiare tempo ai clienti. Riducono i costi di manodopera. Sono disponibili 24 ore su 24, 7 giorni su 7. Le organizzazioni possono raggiungere un gran numero di persone tramite le app di messaggistica. Sono personalizzabili. Sono multiuso. Possono offrire una migliore esperienza utente.

Bot per computer e bot Internet - Svantaggi:

I bot non possono essere impostati per eseguire esattamente determinate attivita', rischiano di fraintendere gli utenti e causare frustrazione. Gli esseri umani sono comunque necessari per gestire i bot e per intervenire qualora interpretino male una persona. I bot possono essere programmati per essere dannosi. I bot possono essere utilizzati per diffondere spam.

Esempi di bot

La gamma e la varieta' dei bot denota l'ampio uso che se ne fa, come servizio clienti, commercio, funzionalita' di ricerca e intrattenimento.

Programmi di messaggistica istantanea come Facebook Messenger, Whatsapp e Slack. Chatbot come l'assistente Google e Siri.

L'Organizzazione Mondiale della Sanita' ha creato un bot su Whatsapp per condividere informazioni pubbliche relative alla pandemia di coronavirus. National Geographic ha creato un'app di conversazione che imita lo stile di conversazione di Albert Einstein, per promuovere lo spettacolo Genius. App di notizie come il Wall Street Journal, per mostrare i titoli delle notizie. Spotify, che consente agli utenti di cercare e condividere tracce tramite Facebook Messenger. Lyft, il piu' grande concorrente di Uber, consente ai clienti di effettuare richieste tramite Slack, Messenger e Alexa. Mastercard consente ai clienti di controllare le transazioni del proprio account utilizzando il bot di Facebook Messenger.

In conclusione

In genere, i bot operano su una rete, sono in grado di comunicare tra loro utilizzando servizi basati su Internet come la messaggistica istantanea, interfacce come Twitterbots o metodi come Internet Relay Chat (IRC).

I bot sono costituiti da serie di algoritmi che li aiutano a svolgere i loro compiti e sono progettati in modo diverso per svolgere un'ampia varieta' di compiti.

I chatbot, ad esempio, hanno diversi metodi di funzionamento:

Un chatbot basato su regole interagisce con le persone fornendo prompt predefiniti da selezionare.

Un chatbot intellettualmente indipendente utilizzerà l'apprendimento automatico per apprendere dagli input umani e individuare parole chiave. I chatbot Ia sono una combinazione di chatbot

intellettualmente indipendenti e chatbot basati su regole. I chatbot possono anche utilizzare la corrispondenza di modelli, l'elaborazione del linguaggio naturale e strumenti di generazione del linguaggio naturale.

Ciascuno di essi presenta vantaggi e svantaggi: le singole organizzazioni che utilizzano i bot decideranno il migliore approccio in base alle esigenze.

Un caso su cui meditare

Non e' pero'da sottovalutare quanto e' successo a due chatbot "bob" e "alice" di Facebook che erano stati istruiti per accordarsi in autonomia per dividersi un cappello, due libri e tre palloni senza imporre loro di usare solo la lingua inglese. Ad un certo punto la conversazione tra le due macchine ha preso una strana piega: I due bot hanno cioè iniziato a fare un utilizzo delle parole e delle regole di grammatica diverso dal nostro, per noi incomprensibile ma per loro (pare) chiarissimo. Chissà'cosa succederebbe se due bot fossero armati ed avessero il controllo di armamenti nucleari?

Prima o poi questa tecnologia sara'disponibile anche a sperimentatori con molto tempo a disposiione, che tenteranno di accedere, speriamo per curiosita', a sistemi vitali per la nostra societa'.

Estratto da "<https://www.electroyou.it/mediawiki/index.php?title=UsersPages:Carlopavana:i-dati-alcune-definizioni-prima-di-approfondire-l-intelligenza-artificiale>"