UNIVERSITÀ
DEGLI STUDI
DI PALERMO

Progetto e sviluppo di un sistema software per l'analisi di fake news

Tesi di Laurea Magistrale in Ingegneria Informatica

F. Bafumo
Relatore: Prof. Giuseppe Lo Re
Correlatore: Prof. Marco Morana

Progetto e sviluppo di un sistema software per l'analisi di fake news

Tesi di Laurea di
Dott. Francesco Bafumo

Relatore
Ch.mo Prof. Giuseppe Lo Re

Correlatore
Prof. Marco Morana

Sommario

La rapida diffusione degli Online Social Network, piattaforme attraverso cui gli utenti hanno la possibilità di condividere informazioni relative a svariati argomenti, ha favorito la nascita di nuovi modelli di comunicazione. Tutti i social, e Twitter in particolare, costituiscono un potente mezzo attraverso cui gli utenti possono pubblicare istantaneamente i loro pensieri e le loro opinioni; d'altro canto, questo accesso immediato e la libertà di condivisione che ne deriva hanno negli anni comportato anche un aumento dei fenomeni di disinformazione basati sulla divulgazione di informazioni false, note come "*fake news*".

Nel presente lavoro, vengono studiati algoritmi intelligenti per il rilevamento di *fake news* e viene proposto un sistema automatico per la creazione di *dataset* annotati tramite cui tali algoritmi possono essere valutati e testati. Tale processo è realizzato a partire dalla raccolta massiva di *tweet* attraverso parole chiave; questi vengono analizzati e raggruppati in argomenti rilevanti (*topic detection*) che sono successivamente sintetizzati (*topic summarization*) al fine di potere essere correlati con fatti noti. Quest'ultima fase è realizzata interrogando i più noti siti di *fact-checking* al fine di determinare il valore di verità delle affermazioni raccolte.

Dataset così ottenuti possono essere utilizzati per la progettazione di approcci non supervisionati di rilevamento di *fake news* che, tipicamente, considerano una combinazione di *features* basate sul contenuto ed informazioni relative ai modelli di propagazione per determinare la veridicità delle osservazioni riportate. In questo lavoro, la valutazione sperimentale ha riguardato quattro algoritmi, e precisamente: Expectation Maximization, Expectation Maximization with Multiple Features, Penalized Expectation Maximization e Constrained Expectation Maximization.

I test condotti mostrano la bontà del processo di annotazione così come le criticità e punti di forza degli approcci considerati.

Indice

Elenco delle tabelle

Elenco delle figure

Capitolo 1

Introduzione

Negli ultimi anni, il tema della divulgazione delle informazioni false, meglio note come fake news, è divenuto un argomento di centrale importanza nell'interesse della società, a causa delle preoccupazioni derivate dalla capacità di questo fenomeno di impattare negativamente sull'opinione pubblica e sugli eventi. La recente e rapida diffusione delle piattaforme di social media, che offrono agli utenti registrati la possibilità di condividere informazioni di qualunque tipologia istantaneamente, e i mancati o non sufficienti sforzi da parte delle piattaforme nel tentativo di controllare il contenuto dei post, hanno comportato una crescita dei fenomeni di disinformazione. Infatti, i social network rappresentano uno strumento potente che consente agli utenti sia la condivisione di notizie, pensieri ed opinioni, sia l'acquisizione di informazioni su questioni sociali, politiche ed economiche. L'utilizzo di queste piattaforme, che costituiscono ormai la più ampia fonte di informazioni, permette agli individui di comunicare in maniera istantanea, annullando la distanza fisica che li separa; l'accesso immediato, congiuntamente a non sufficienti controlli da parte delle piattaforme stesse, incrementano i fenomeni di disinformazione e misinformazione. Nei social network utilizzati nelle reti veicolari, per esempio, è possibile far credere agli utenti che in una certa zona ci sia traffico quando in realtà non è così [2].

Rumors e fake news sono ormai considerati delle minacce alla democrazia, al giornalismo e alla libertà di espressione; infatti, diminuiscono la fiducia degli utenti nei confronti dei governi e delle notizie vere in circolazione. È stato dimostrato dalla ricerca che post con contenuto disinformativo si diffondono tipicamente in maniera più rapida rispetto ai post che riportano informazioni vere, soprattutto quelli riguardanti la politica. Ad esempio, un'analisi delle elezioni presidenziali statunitensi nel 2016 effettuata dagli economisti Allcott e Gentzkow [1] ha rivelato che sono state ampiamente condivise fake news durante i 3 mesi precedenti le elezioni con circa 30 milioni di condivisioni Facebook totali di cui circa 22 milioni di post riguardanti false storie pro-Trump e 7,6 milioni di post riguardanti 41 note false storie pro-Clinton.

La politica però non è la sola vittima di questo fenomeno: fake news e rumors influenzano infatti anche l'economia e i mercati azionari. Nel 2013, un tweet di Associated Press, agenzia di stampa internazione, secondo cui Barak Obama sarebbe stato ferito da una esplosione, ha causato una perdita di 130 miliardi di dollari del valore delle azioni, nonostante AP abbia poi affermato che il suo account Twitter era stato violato [3]. Per fare un altro esempio, nel settembre 2008, un articolo di sei anni addietro che riportava il fallimento della società madre della United Airlines

nel 2002 è riemerso su Internet e si è creduto erroneamente che riportasse una nuova dichiarazione di fallimento da parte della società. Questo episodio ha causato un calo del prezzo delle azioni della società fino al 76% in pochi minuti. Dopo che la "notizia" è stata identificata come falsa, il prezzo del titolo è rimbalzato, ma ha comunque chiuso la giornata dell'11,2% al di sotto della chiusura precedente [4]. Inoltre, spesso emergono false informazioni inerenti a disastri naturali, come il terremoto in Giappone nel 2011 [5] e l'uragano Sandy nel 2012 [6], che hanno l'intento di causare un aumento del panico e del disordine sociale.

È stato evidenziato che i fattori psicologici e sociali influenzano il processo di diffusione dei fenomeni di disinformazione: la ricerca ha mostrato che le notizie dove ci si aspetta autenticità e obiettività ottengono più facilmente la fiducia del pubblico; in più, si tende a fidarsi di una determinata notizia se questa rispecchia le proprie convinzioni (bias di conferma [7]) oppure se la notizia viene ripetutamente riscontrata (effetto di validità [8]). L'incapacità dell'individuo di distinguere tra notizie vere e notizie false è stata attribuita alla propria capacità cognitiva e ai pregiudizi ideologici: ciò è dimostrato dal sondaggio di YouGov [9] rivolto a 1.684 adulti britannici a cui sono stati mostrati sei singoli articoli di notizie, tre veri e tre falsi, e solo il 4% è stato in grado di identificarli tutti in modo corretto, mentre il 49% degli intervistati pensava che almeno una delle storie false fosse vera.

Questa tendenza dei social (in primis Twitter) a far sorgere disinformazione motiva lo sviluppo di sistemi di classificazione e rilevamento di fake news basati su tecniche di machine learning per limitarne la diffusione. Infatti, grazie ai progressi dell'intelligenza artificiale, l'apprendimento automatico svolge un ruolo importante per la gestione della crescente quantità di notizie. Nonostante rumors e fake news siano per definizione due concetti leggermente diversi, le tecniche di rilevamento e di classificazione adottate sono applicabili indistintamente ad entrambi i fenomeni. Brevemente, le fake news possono essere considerate come rumors che sono stati ritenuti infondati a valle di una classificazione. In generale, lo scopo del rilevamento è quello di distinguere informazioni reali da informazioni false e quindi stabilire se un messaggio è vero, falso o non verificato, considerando alcune caratteristiche come il contenuto, il tempo di pubblicazione e l'utente del messaggio [15].

1.1 Definizioni e Storia

Il termine "Rumors" non ha un sostanziale corrispondente nella lingua italiana: si potrebbe fare ricorso a termini come dicerie, chiacchiere, indiscrezioni, ma non rappresentano dei sinonimi perfetti. La parola deriva dal latino "rumores", ovvero voci, e per definizione costituiscono informazioni non verificate che possono rivelarsi vere, false, parzialmente false o rimanere irrisolte. Sono quindi informazioni il cui stato di veridicità deve ancora essere verificato al momento della pubblicazione. Gli psicologi Nicholas Di Fonzo e Prashant Bordia, nel saggio intitolato *Rumor Psychology: Social and Organizational Approaches*, definiscono i rumors notizie non verificate che nascono in contesti di incertezza, pericolo o potenziale minaccia e che hanno la funzione di aiutare le persone a dare un senso alle cose, a gestire il rischio. La storia dei rumors è antica come la storia dell'uomo. Da sempre, fin nella remota antichità, le voci, le dicerie, i pettegolezzi, le indiscrezioni, venivano messe in circolazione per

Figura 1.1: Dea Fama sul tetto dell'Università di Arti Visive di Dresda.

influenzare le vicende umane, sia intenzionalmente che in buona fede. Nell'antica Roma, i romani veneravano una Dea dedicata ai rumors: la Dea Fama. La divinità rappresentava la messaggera di Giove e aveva il compito di diffonderne gli annunci. Era raffigurata sempre come una donna in moto che gridava continuamente diffondendo notizie buone e cattive, ed era figurata giovane e irruente con ali cosparse di occhi, di bocche e di lingue, in atto di suonare una tromba, oppure due, una per la verità, l'altra per la menzogna. Quella lunga era dedicata a diffondere la fama e la corta a diffondere false voci. Allegoricamente, questa divinità alata rappresentava appunto le dicerie che nascono, si diffondono e acquistano credibilità alterando a piacimento i fatti, senza alcuna distinzione tra vero e falso.

A differenza dei rumors, le fake news sono sempre false: sono un tipo specifico di bufala, una falsa storia divulgata per mascherare la verità, poiché sono informazioni false diffuse attraverso organi di informazione che mirano ad avere vantaggi politici e finanziari. Tuttavia ad oggi non viene fornita una definizione universale di fake news; infatti a volte vengono definite come articoli di notizie o messaggi diffusi attraverso i media, che riportano informazioni errate indipendentemente dalle motivazioni, o ancora sono considerate notizie intenzionalmente false pubblicate da una fonte di notizie. Per notizie si intendono articoli, dichiarazioni, affermazioni, post che possono essere messi in circolazioni da giornalisti o da non giornalisti. È bene notare che

rumors e fake news diffusi intenzionalmente sono più difficili da rilevare perché cercano di ingannare il pubblico sulla verità dell'informazione e dunque tendono ad essere indistinguibili dalle notizie vere. Uno degli esempi più famosi di fake news risale al 1814, in pieno periodo napoleonico. Il 21 Febbraio di quell'anno, un uomo vestito come un ufficiale britannico si presentò in una locanda a Dover spacciandosi per il Colonnello Du Bourg e affermò di essere a conoscenza della notizia più importante degli ultimi vent'anni: la morte di Napoleone. In realtà, Napoleone era ancora vivo, si trattava di una notizia falsa, ma piuttosto verisimile viste le difficoltà dell'imperatore. Il finto Colonnello chiese una carrozza per raggiungere Londra e, una volta ottenuta, iniziò a diffondere la notizia fermandosi in ogni locanda che incontrava nel tragitto. All'alba, il Colonnello e la bufala erano entrambe giunte a destinazione: il primo fece perdere le sue tracce, la seconda diventò in breve il principale argomento discusso in città. Nonostante mancavano ancora notizie ufficiali, i piccoli azionisti iniziarono ad investire contando sul fatto che l'imperatore francese fosse morto e i titoli di stato decollarono. Dopo qualche ora, arrivò in città una carrozza sfarzosa dal quale alcuni uomini ribadivano la falsa storia. Solo nel pomeriggio fu scoperta la verità e i presunti cospiratori furono incriminati e condannati. Quale fu il motivo della divulgazione della bufala? Quella mattina sei persone avevano venduto titoli governativi, acquisiti poco tempo prima, per oltre un milione di sterline. Tra loro anche Sir Thomas Cochrane, detto Lord Cochrane, politico radicale ed eroe di guerra, visto anche in compagnia di Charles Random de Berenger, cioè l'impostore che si era spacciato per il colonnello Du Bourg.

Un altro esempio degno di nota fu il caso della *Guerra dei mondi*, trasmissione radiofonica di Orson Welles del 1938. La trasmissione, fu mandata in onda in modo da sembrare una serie di comunicati tra importanti autorità statunitensi. Lo scopo non era quello di diffondere una fake news, tanto che Welles ribadisce ciò sia all'inizio che alla fine della trasmissione. Nonostante la buona fede del conduttore radiofonico, i telespettatori credettero che si trattasse di notizie reali, soprattutto quelli che si sintonizzavano a programma già iniziato dopo la premessa iniziale di Welles. Dopo il malinteso, Welles preparò una dichiarazione in cui sottolineava l'intento principale della trasmissione, che era quello di intrattenere il pubblico la vigilia della notte di Halloween, ma la stampa iniziò ad accusare duramente la radio di irresponsabilità definendola come uno dei più pericolosi eventi della cultura moderna. Il caso della *Guerra dei Mondi* dimostra come i mezzi di comunicazione principali possono essere manipolati e strumentalizzati con lo scopo di diffondere fake news.

In questo contesto vengono anche utilizzati i termini disinformazione e misinformazione, che hanno una definizione più ampia: la disinformazione si riferisce ad informazioni deliberatamente false, la misinformazione è costituita da informazioni false che circolano a causa di un errore onesto e che vengono divulgate non in maniera intenzionale.

Capitolo 2

Stato dell'Arte

Nel presente Capitolo, sarà riportata una breve panoramica riguardante i metodi di fake news detection utilizzati in letteratura. Inizialmente saranno discusse alcune teorie comportamentali che forniscono informazioni utili al rilevamento delle notizie false, e le tipologie di features che sono considerate nei problemi di classificazione di questo tipo. In seguito, saranno presentati gli approcci di rilevamento più utilizzati tra cui metodi tradizionali di apprendimento automatico e metodi di deep learning. Infine, saranno illustrati i principali dataset utilizzati nei modelli, insieme ai siti di fact-checking più diffusi, e saranno discussi i problemi legati alla valutazione dei metodi e alla necessità di avere dataset annotati.

2.1 Teorie fondamentali per il rilevamento

Recenti studi hanno mostrato che alcune teorie comportamentali sviluppate nelle scienze sociali ed economiche, forniscono importanti informazioni per il rilevamento di rumors e/o fake news. Pertanto, possono nascere nuove opportunità di ricerca in questo campo che porterebbero alla realizzazione di nuovi modelli di rilevamento. Secondo un'indagine, le teorie comportamentali che devono essere considerate a tal scopo sono: teorie relative alle notizie e teorie relative agli utenti.

Le **teorie relative alle notizie** riguardano le possibili caratteristiche che permettono di distinguere le notizie reali dalle fake news. Queste features comprendono lo stile di scrittura, la qualità di scrittura, il conteggio delle parole (secondo la teoria della manipolazione delle informazioni) [13], e i sentimenti espressi. Queste teorie sono valide per il rilevamento di disinformazione e di affermazioni che hanno lo scopo di ingannare; dunque non riguardano esplicitamente le fake news o i rumors. Pertanto, un'opportunità di ricerca potrebbe consistere nel verificare se queste caratteristiche sono compatibili anche con i modelli di rumors e fake news detection. Le principali teorie relative alle notizie sono riportate nella Tabella 2.1.

Le **teorie relative agli utenti** riguardano le caratteristiche degli utenti che pubblicano informazioni e che quindi contribuiscono alla diffusione di fake news. Queste caratteristiche comprendono like, commenti, post, followers e seguiti. I principali utenti dannosi sono i social bot, che diffondono informazioni errate intenzionalmente, ma bisogna notare che le fake news coinvolgono anche utenti onesti e non a conoscenza del fatto che stanno pubblicando una notizia falsa. Questa vulnerabili-

News-related Theories	
Theory	**Phenomenon**
Undeutsch hypothesis [Undeutsch 1967][10]	A statement based on a factual experience differs in content style and quality from that of fantasy
Reality monitoring [Johnson and Raye 1981][11]	Actual events are characterized by higher levels of sensory-perceptual information
Four-factor theory [Zuckerman et al. 1981][12]	Lies are expressed diffrerently in terms of arousal, behavior control, emotion, and thinking from truth
Information manipulation theory [McCornack et al. 2014][13]	Extreme information quantity often exists in deception

Tabella 2.1: Teorie relative alle notizie.

tà degli utenti deriva psicologicamente dagli impatti sociali e dalle teorie riportate nella Tabella 2.2. La fiducia e la diffusione delle fake news derivano dunque sia dall'eccessiva esposizione che da fattori psicologici e convinzioni.

2.2 Features

Le features svolgono un ruolo cruciale nei problemi di classificazione in quanto da esse dipende la qualità dei risultati del modello. Ne esistono di tanti tipi tra cui features linguististiche, temporali, basate sulle informazioni degli utenti e basate sulle interazioni. Nel rilevamento delle fake news, l'utilizzo di differenti features porta a differenti risultati. In generale, i ricercatori considerano le features in base ai modelli di apprendimento che intendono utilizzare per ottenere i risultati che desiderano. Ad esempio, Rosas et al. [32] hanno scoperto che per argomenti seri come l'istruzione e la politica, occorre prestare maggiore attenzione alle features linguistiche nel messaggio, e per le notizie sulle celebrità, occorre porre più attenzione alle emozioni o alle opinioni dell'autore. Le principali features sono:

- **Features Linguistiche** - Le features linguistiche sono tra le più importanti per il rilevamento delle fake news. Sono utili per comprendere la polarità del messaggio che svolge un ruolo significativo in questo contesto. Kwon et al. [33] hanno classificato i messaggi in diverse categorie contando la proporzione delle parole psicologiche attraverso l'utilizzo di uno strumento di sentiment analysis denominato Linguistic Inquiry and Word Count (LIWC). Inoltre, Zeng et al. [34] hanno utilizzato TF-IDF per classificare la polarità del sentiment dei messaggi. Un'altra tecnica per ottenere le features linguistiche è il word embedding. Zhou et al. hanno utilizzato questo approccio per convertire il messaggio in un vettore che può essere utilizzato direttamente come input per i classificatori (ad esempio SVM). Il word embedding può anche essere integrato in architetture di reti neurali (soprattutto le CNN) per rilevare fake news all'interno di un framework di deep learning [35, 36, 37, 38, 39, 40], reti neurali ricorrenti (RNN) [41, 42, 43] e transformer [44, 45]. Oltre alle features linguistiche utilizzate per il word embedding (definite latenti), vi sono features testuali generali utilizzate nei frameworks di machine learning tradizionali che descrivono il contenuto da quattro livelli linguistici: lessico, sintassi, discorso, semantica.

User-related Theories		
Social Impacts	Conservatism bias [Basu 1997][14]	The tendency to revise one's belief insufficiently when presented with new evidenze
	Semmelweis reflex [Balint and Balint 2009][16]	Individual tend to reject new evidence because it contradicts with established norms and beliefs
	Echo chamber effect [Jamieson and Cappella 2008][17]	Beliefs are amplified or reinforced by communication and repetition within a closed system
	Attentional bias entional bias [MacLeod et al. 1986][18]	An individual's perception is is affected by his or her recurring thoughts at the time
	Validity effect [Boehm 1994][8]	Individuals tend to believe informartion is correct after repeated exposures
	Bandwagon effect [Leibenstein 1950][19]	Individuals do something primarily because other are doing it
	Normative influence theory [Deutsch and Gerard 1955][20]	The influence of other leading us to conform to be like and accepted by them
	Social identity theory [Ashforth and Mael 1989][21]	An individual's self-concept derives from perceived membership in a relevant social group
	Availability cascade [Kuran and Sunstein 1999][22]	Individuals tend to adopt insights expressed by others when such insights are gaining more popularity within their social circles
Self-Impacts	Confirmation bias [Nickerson 1998][7]	Individuals tend to trust information that confirms their preexisting beliefs or hypotheses
	Selective exposure [Freedman and Sears 1965] [23]	Individual prefer information that confirm their preexiting attitudes
	Desirability bias [Fisher 1993][24]	Individuals are inclined to accept information that pleases them
	Illusion of asymmetric insight [Pronin et al. 2001][25]	Individuals perceive their knowledge to surpass that of others
	Naive realism [Ward et al. 1997][27]	The senses provide us with direct awareness of objects as they really are
	Overconfidence effect [Dunning et al. 1990][28]	A person's subjective confidence in his judgments is reliably grater than the objective ones
Benefits	Prospect theory [Kahneman and Tversky 2013][29]	People make decisions based on the value oflosses and gains rather than the outcome
	Contrast effect [Hovland et al. 1957][30]	The enhancement or diminishment of cognition due to successive or simultaneous exposure to a stimulus oflesser or greater value in the same dimension
	Valence effect [Frijda 1986][31]	People tend to overestimate the likelihood of good things happening rather than bad things

Tabella 2.2: Teorie relative all'utente.

- **Features Temporali** - Anche queste features sono degne di considerazione in quanto le fake news hanno diverse proprietà temporali durante il processo di propagazione rispetto alle notizie reali. Uno strumento efficace utilizzato per apprendere le features temporali sono le reti neurali ricorrenti (RNN). Ruchansky et al. [46] hanno catturato il numero di utenti coinvolti in un articolo e anche ottenuto come il coinvolgimento è stato spaziato nel tempo. In seguito a questi risultati, il loro modello fu in grado di classificare le fake news dalle proprietà temporali.

- **Features basate sulle informazioni degli utenti** - Le informazioni riguardanti gli utenti coinvolti nei processi di diffusione delle fake news possono essere utili al rilevamento di quest'ultime. Features di questo tipo comprendono l'età di registrazione, il numero di follower, il numero di tweet scritti in passato che permettono di stabilire dei livelli di credibilità degli utenti.

- **Features basate sulle interazioni** - Nelle piattaforme di social media, in particolare Twitter, gli utenti possono non solo commentare i tweets di altri commenti ma anche retwittarli allo scopo di diffondere l'informazione più velocemente. Sono stati studiati i pattern di interazione tra gli utenti e i processi di propagazione per trovare la relazione con la diffusione delle fake news. Ad esempio, Liu et al. [47] hanno utilizzato una rete neurale convoluzionale (CNN) per apprendere la rappresentazione di ciascun percorso di propagazione; Ma et al. [48] hanno utilizzato architetture top-down e bottom-up per descrivere il processo di propagazione, hanno costruito alberi di propagazione di tweets e calcolato la somiglianza tra alberi per evidenziare le differenze tra processi di propagazione di fake news e notizie vere.

2.3 Approcci

Gli approcci relativi al rilevamento di rumors e fake news erano costituiti in passato esclusivamente da tecniche di machine learning. Negli ultimi anni, invece, i progressi relativi al deep learning hanno comportato l'utilizzo di metodi più sofisticati per il rilevamento di essi. Inoltre, i sistemi di rilevamento possono ulteriormente classificarsi in sistemi basati sulla conoscenza, sistemi basati sulla propagazione, e sistemi basati sulla credibilità della fonte.

2.3.1 Metodi tradizionali di apprendimento automatico

I metodi tradizionali di apprendimento automatico comprendono Decision Tree, Random Forest, Regressione Lineare, algoritmo di Bayes ed SVM (Support Vector Machine).

Gli approcci basati su alberi comprendono alberi decisionali e Random Forest. Castillo et al [49] hanno utilizzato alberi decisionali J48 nel loro modello e hanno ottenuto una buona accuratezza tra molti metodi di apprendimento automatico. Know et al. [33] hanno invece utilizzato Random Forest con un meccanismo di voto di più alberi decisionali di classificazione di rumors allo scopo di ridurre l'overfitting e hanno ottenuto una maggiore precisione rispetto agli alberi decisionali. Gli approcci basati

sulla Support Vector Machine sono i più accurati e molto utilizzati dai ricercatori e questo spiega il fatto che la SVM è considerata il classificatore di base da confrontare con altri classificatori [51]. Infine, gli approcci basati sull'algoritmo di Bayes sono tra i più efficienti e molto utilizzati nella classificazione del testo e nel rilevamento di spam, in quanto si basano su solide basi matematiche. Mihalcea e Strapparava [52] hanno addestrato classificatori Naive Bayes ed SVM con l'intento di distinguere fake news da notizie vere. L'accuratezza delle classificazioni era circa il 70% nell'identificare le bugie delle persone sulle loro convinzioni e il 75% nell'identificare bugie sui loro sentimenti. Un'analisi dettagliata dell'uso delle parole ha rivelato che in tutte le affermazioni ingannevoli mancavano connessioni con il sé ed erano dominanti altre classi di parole legate all'uomo ("Tu, gli altri, gli umani"). Inoltre, è stato riscontrato che erano frequenti anche le parole relative alla "certezza", il che è probabilmente spiegato dalla necessità per il parlante di utilizzare esplicitamente parole legate alla verità come mezzo per rendere più credibili le loro false dichiarazioni [53].

2.3.2 Metodi di deep learning

Nel campo di rilevamento dei rumors e delle fake news, il deep learning sta diventando sempre più dominante in quanto può adattarsi facilmente a diverse applicazioni. L'impiego di tecniche di deep learning rende non necessaria l'ingegnerizzazione delle features, quindi non prevede le impegnative fasi di progettazione delle features, e permette il raggiungimento di prestazioni eccellenti. Gli approcci più utilizzati sono CNN, RNN, Long Short Term Memory (LSTM), Attention Mechanism e Joint Learning. Le reti neurali convoluzionali sono molto utilizzate per il rilevamento di fake news anche grazie al word embedding che consente di trasformare frasi in matrici da passare in input al modello. Sarkar et al. [54] hanno proposto un modello di CNN per rilevare notizie di satira, per catturare informazioni sia a livello di frase che di documento. Tuttavia un singolo strato di CNN consente di costruire solo rappresentazioni da parole vicine. Questo ha spinto Qian et al. [55] a realizzare un modello di CNN a due livelli (TCNN) per catturare informazioni semantiche profonde. L'impiego di reti neurali ricorrenti si basa invece sul fatto che per estrarre meglio le caratteristiche di una frase, occorre concentrarsi sulla posizione occupata dalla parola all'interno della frase. In questo tipo di modelli, l'input consiste in una sorta di serie temporali che vengono trasmesse in modo circolare nel modello. Inoltre, al fine di trattenere distanti e separate le informazioni, questi modelli vengono arricchiti di un meccanismo di LSTM. Rashkin et al. [56] hanno usato sequenze di parole come input per prevedere la valutazione dei post. Infine, il modello LSTM ha sovraperformato l'altro modello quando si utilizzava solo il testo come input, il che significava che tali informazioni lessicali erano ridondanti rispetto a ciò che il modello aveva già imparato dal testo. Popat et al. [57] hanno utilizzato l'LSTM bidirezionale al posto dell'LSTM standard, che potrebbe catturare sia le caratteristiche del passato che quelle future degli stati. Uno svantaggio dei modelli LSTM è che l'input è codificato in un vettore di lunghezza fissa e la decodifica è dunque limitata a questa rappresentazione vettoriale. Ciò causa una diminuzione delle prestazioni quando l'input del modello è grande. Vaswani et al. [58] hanno proposto un Attention Mechanism che permette di assegnare pesi diversi a ciascuna parte dell'input e dunque estrarre informazioni importanti. Popat et al. [57] hanno utilizzato questo

meccanismo per estrarre le parole salienti in un articolo: ad esempio, alcune parole come "a malapena vero" e "prove imprecise" ha aiutato il sistema a identificare l'affermazione come non credibile. D'altra parte, parole come "rivela" e "documentato" ha aiutato il sistema a valutare l'affermazione come credibile. Infine, un altro approccio utilizzato per il rilevamento dei rumors è il Joint Learning. L'idea è che se ci si concentra solo su un singolo modello potrebbero ignorarsi informazioni utili al rilevamento. Shu et al. [59] hanno proposto un framework che ha dimostrato che, considerando simultaneamente l'utente che pubblica la notizia e il coinvolgimento degli utenti, si può migliorare il rilevamento di fake news. Volkova et al. [60] hanno costruito una rete neurale che ha congiuntamente appreso dai contenuti del tweet e dalle interazioni sociali per classificare notizie vere e false.

2.3.3 Rilevamento basato sulla conoscenza

I modelli di rilevamento basati sulla conoscenza utilizzano un processo noto come la verifica dei fatti, che mira a valutare la verità di una notizia confrontando la conoscenza estratta dal contenuto delle notizie con fatti noti. Questa verifica può essere manuale o automatica.

La verifica manuale può a sua volta essere suddivisa in verifica basata su esperti e verifica basata sul crowdsourcing. Nel primo caso, la verifica si basa su esperti di dominio altamente credibili che verificano i contenuti delle notizie in circolazione. Il rilevamento in questione è semplice da gestire (i siti di fact-checking facilitano questi controlli, alcune fake news sono inoltre disponibili pubblicamente, ad esempio LIAR), ma non è scalabile con l'aumento del volume delle notizie da controllare. Il controllo basato sul crowdsourcing si basa invece su una vasta popolazione di individui normali che agiscono come verificatori di fatti. Questo tipo di controllo è però difficile da gestire a causa dei pregiudizi politici dei verificatori e potrebbe succedere che siano presenti annotazioni contrastanti; infatti è necessario un meccanismo di risoluzione dei risultati in conflitto. I siti di fact-checking per il crowdsourcing sono ancora in fase di sviluppo; tra questi è presente Fiskkit che consente agli utenti di caricare articoli e fornire delle valutazioni a questi.

É chiaro che nei sistemi di verifica manuale, la scalabilità è il principale vincolo. Ecco perché sono state ideate tecniche di verifica automatica di fatti. In questo contesto, è necessario fornire una rappresentazione standard di conoscenza: la conoscenza è definita SPO (soggetto, predicato, oggetto); ad esempio, la conoscenza all'interno della frase "Donald Trump è il presidente degli Stati Uniti" può essere (Donald Trump, Professione, Presidente). Inoltre sono utili ulteriori definizioni come:

- fatto, ovvero una conoscenza verificata come verità;
- base di conoscenza (Knowledge Base, KB), ovvero un insieme di fatti;
- grafo di conoscenza (Knowledge Graph, KG), ovvero una struttura a grafo che rappresenta le triple SPO di una base di conoscenza.

Per formare una base di conoscenza, e quindi un grafo di conoscenza, occorre un processo noto come estrazione di conoscenza, che può avvenire tramite una sola fonte o tramite più fonti. L'estrazione da un'unica fonte (ad esempio Wikipedia) è più efficiente ma la conoscenza può risultare incompleta; viceversa, l'estrazione da più fonti

(open-source) è sicuramente meno efficiente ma porta ad una migliore conoscenza. Dopo l'estrazione dei fatti, per realizzare un KG sono necessarie ulteriori operazioni: innanzitutto, bisogna risolvere triple SPO equivalenti (processo noto come *risoluzione dell'entità*) [61]; in seguito, si deve verificare la correttezza delle triple in base al loro intervallo di validità (processo noto come *invalidità*); si devono risolvere le triple contrastanti (*risoluzione dei conflitti*), e si utilizzano in questo caso metodi decisionali multicriterio (MCDM) [62, 63, 64, 65]; successivamente, occorre verificare la credibilità della fonte, ed esistono sistemi come NewsGuard18 o MediaRank che permettono di verificare la credibilità dei siti web [66]; infine, bisogna dedurre in modo affidabile ulteriore conoscenza sulla basa di quella esistente (processo noto come *completamento KG*).

Piuttosto che costruire un KG da zero, è possibile fare affidamento su modelli esistenti, come YAGO [67, 68], Freebase [70], NELL [71], PATTY [72], DBpedia [73], Elementary/DeepDive [74] e Knowledge Vault [75]. Costruita una base di conoscenza, è possibile confrontare la conoscenza estratta dalle notizie con la base di conoscenza. Dunque, la strategia di verifica dei fatti per una tripla SPO consiste nell'individuazione dell'entità, e quindi il soggetto viene abbinato ad un nodo del KG che rappresenta la stessa entità del soggetto, e nella verifica della relazione, e dunque verificare se nel grafo è presente il predicato tra il soggetto e l'oggetto in questione. In caso positivo, la tripla SPO viene considerata una verità.

2.3.4 Rilevamento basato sulla propagazione

I sistemi di rilevamento basati sulla propagazione utilizzano le informazioni relative alla diffusione di fake news. L'input per i modelli di rilevamento in questione può essere costituito da una cascata di notizie e cioè una rappresentazione diretta della propagazione delle notizie, o un grafo auto-definito, ovvero una rappresentazione indiretta che comprende più informazioni sulla propagazione. Una cascata di notizie è un albero o una struttura simile ad un albero che acquisisce direttamente la propagazione di un determinato articolo di notizie su un social network. Il nodo radice rappresenta l'utente che ha diffuso la notizia per primo, mentre gli altri nodi costituiscono gli utenti tramite cui si è propagata la notizia pubblicata dall'utente del nodo radice. Le cascate di notizie possono essere di due tipi: basate su hop (e quindi sul numero di passaggi cha la notizia ha percorso) e basate sul tempo. Su quelle basate su hop è possibile considerare tre grandezze:

- Profondità - il numero massimo di passi che ha percorso la notizia;
- Ampiezza - il numero di utenti che hanno diffuso la notizia ad un determinato passo;
- Dimensione - il numero di utenti coinvolti nella propagazione della notizia.

Allo stesso modo, nelle cascate di notizie basate sul tempo si possono considerare:

- Lifetime - l'intervallo più lungo in cui la notizia è stata propagata;
- Real-time heat - il numero di utenti che diffondono la notizia in un determinato istante di tempo;

- Overall heat - il numero totale di utenti coinvolti nella propagazione della notizia.

In questo caso le tecniche di rilevamento si basano su modelli di machine learning tradizionali e su modelli di deep learning. Nel primo caso, le cascate di notizie vengono rappresentate come un insieme di features e vengono adottati sistemi di apprendimento supervisionato come SVM, alberi decisionali, e RF. È stato evidenziato che ci sono differenze nella propagazione tra notizie vere e notizie false: infatti, le fake news si diffondono più velocemente, più ampiamente e più lontano (ampiezza e profondità delle cascate di notizie false sono maggiori rispetto ad ampiezza e profondità delle cascate di notizie reali e a parità di tempo, la dimensione delle cascate di notizie false è maggiore di quella delle notizie vere). Inoltre, la ricerca ha dimostrato che le fake news politiche si diffondono più velocemente rispetto a quelle riguardanti altri domini come affari, terrorismo, scienza ed intrattenimento [76]. Le tecniche di deep learning, invece, si basano soprattutto sulle reti neurali, in cui una funzione softmax funge da classificatore. Ad esempio, Ma et al. svilupparono reti neurali ricorsive (RvNNs), una rete neurale strutturata ad albero, basata su cascate di notizie [77, 78].

Nei sistemi di rilevamento basati su grafi di propagazione auto-definiti, si utilizzano reti che catturano in modo flessibile la propagazione di notizie false e si dividono in omogenee, eterogenee e gerarchiche. Le reti omogenee contengono solo un tipo di nodo. Ad esempio la Spreader Net [79] mostra le relazioni tra più utenti coinvolti nella propagazione di un articolo di notizie. Ogni nodo della rete rappresenta un utente che diffonde la notizia e i collegamenti tra nodi rappresentano le relazioni tra gli utenti. Classificare un articolo corrisponde a classificare una rete. Un altro esempio di rete omogenea è costituito dalla Stance Net [80], in cui i nodi rappresentano post da parte degli utenti e i collegamenti rappresentano relazioni di supporto o opposizione del punto di vista sulla notizia. La somiglianza tra ogni coppia di host può essere calcolata tramite la divergenza di Jensen-Shannon [80] o distanza di Jaccard [81]. In questo caso, il rilevamento di fake news si riduce alla valutazione della credibilità dei post relativi alle notizie. Le reti eterogenee hanno invece diversi tipi di nodi e collegamenti. Un esempio è costituito dalla rete users-tweets-news_events in cui Gupta et al. [82] valutano la credibilità delle notizie attraverso un algoritmo simile al PageRank [84]. L'idea alla base dell'algoritmo è che gli utenti che hanno bassa credibilità tendono a pubblicare tweet spesso inerenti a notizie false. Infine, nelle reti gerarchiche, i nodi formano una gerarchia. Un esempio è la rete news-tweet-retweet-reply che è un'estensione della cascata di notizie [85]. Possono quindi essere considerate le features come ampiezza, dimensione, profondità per utilizzare questa rete gerarchica in un framework tradizionale di ML.

I metodi appena discussi sono robusti alla manipolazione dello stile di scrittura da parte di autori di fake news; tuttavia, i sistemi di rilevamento basati sulla propagazione risultano inefficienti in quanto non permettono di rilevare la notizia prima che questa si diffonda.

2.3.5 Rilevamento basato sulla fonte

I modelli di rilevamento basati sulla fonte valutano le notizie analizzando la credibilità della fonte. Quest'ultima è costituita da fonti che creano la notizia, come gli autori

delle notizie, fonti che pubblicano la notizia, come gli editori di notizie, e fonti che la diffondono, come gli utenti dei social media. I metodi di rilevamento in oggetto ipotizzano quindi che fonti poco credibili e inaffidabili diffondano fake news nonostante non sia improbabile che queste sorgenti pubblichino notizie vere. Nonostante questo presupposto arbitrario, tecniche di rilevamento del genere risultano essere molto efficienti [87]. Per quanto riguarda gli autori e gli editori delle notizie, è stato dimostrato che le reti che si creano dalla collaborazione di questi presentano omogeneità: nello specifico, Sitaula et al. [88] hanno costruito una rete di autori/editori basandosi sul Dataset FakeNewsNet [89]; ogni nodo rappresenta un autore e gli edge indicano una collaborazione per la realizzazione di uno o più articoli di notizie. I nodi della rete comprendono: autori di notizie vere, autori di fake news e autori che pubblicano sia fake news che notizie vere. La rete è omogenea in quanto gli autori dello stesso tipo di nodo risultano più densamente connessi rispetto ad autori di altri tipi di nodo.

Il rilevamento di fonti inaffidabili può essere ridotto al rilevamento di siti web poco credibili, in quanto gli editori spesso pubblicano le notizie nei propri siti web. Esistono degli algoritmi che valutano la credibilità di siti web come PageRank [84] e HITS [90] soprattutto utili a migliorare la risposta dei motori di ricerca in seguito ad una ricerca di un utente. Un'altra risorsa è NewsGuard che valuta la credibilità della fonte attraverso alcuni criteri tra cui: se pubblica ripetutamente contenuti falsi, se presenta informazioni in modo responsabile, se corregge regolarmente o chiarisce gli errori, se gestisce responsabilmente la differenza tra notizie e opinioni, se evita di ingannare con i titoli, se etichetta chiaramente la pubblicità, se fornisce informazioni sui creatori di contenuti.

Per quanto riguarda i diffusori di fake news sui social, è intuitivo che gli utenti con scarsa credibilità pubblicano notizie false con maggiore probabilità. La principale minaccia in questo contesto è rappresentata dai social bot, ovvero applicazioni software che eseguono su internet attività automatizzate. I social bot sono utilizzati sia per fornire servizi utili, ma anche per ingannare gli utenti e diffondere fake news. Uno studio recente ha dimostrato che il 9-15% di account su Twitter siano bot attivi [91]. Si pensa anche che milioni di social bot abbiamo partecipato a discussioni online sulle elezioni presidenziali statunitensi del 2016 e sulle elezioni francesi del 2017 [93]. Shao et al. [94] hanno dimostrato che i bot diffondono notizie inaffidabili in modo più veloce rispetto agli utenti con lo scopo di diffondere in modo virale la notizia. Infatti, gli utenti retwittano i post pubblicati da bot quasi quanto quelli pubblicati da normali utenti e questo evidenzia le vulnerabilità umane alle informazioni online non affidabili.

2.4 Problemi legati ai Dataset

OMISSISS

Lo sviluppo di nuove soluzioni per il rilevamento di fake news e rumors è stato limitato dalla qualità dei dati [26, 83]. Il tipo di informazioni raccolte dai dataset dipendono dallo scopo dell'applicazione e potrebbero variare significativamente tra un dataset e l'altro. Ad esempio, alcuni dataset si concentrano su fatti di gossip mentre altri comprendono dichiarazioni politiche. Inoltre, si differenziano anche in base al tipo di contenuti che vengono inclusi (ad esempio risposte degli utenti, fonte

dell'affermazione ecc.), e alle etichette che vengono fornite [50]. I dataset sono spesso usati come modelli di addestramento o di convalida. Ciò significa che la quantità e la qualità di dati nel dataset e il numero di labels influenzano la classificazione degli algoritmi di fake news detection. Ad esempio, KaggleFN comprende articoli di notizie contrassegnati come falsi dall'applicazione BS-Detector e non comprende verifica giornalistica o umana, quindi addestrare un modello ad apprendere le labels in questione significa essenzialmente addestrare un modello per imitare BS-Detector [96]. In seguito a queste considerazioni, che evidenziano da un lato la significativa influenza dei dataset sugli algoritmi di rilevamento di fake news, e dall'altro la limitata qualità delle informazioni contenute, è stato realizzato un sistema di raccolta massiva dei tweets con annotazione semi-automatica, descritto nel Capitolo 3.

Capitolo 3

Sistema per la creazione di dataset annotati

I problemi legati alla valutazione dei metodi di rilevamento di fake news e alla necessità di avere dataset annotati, evidenziati nella Sezione 2.4, hanno posto l'attenzione verso la sperimentazione di un sistema di raccolta massiva dei tweets, con annotazione semi-automatica.

In questo Capitolo sarà illustrata la struttura del sistema proposto e saranno descritte dettagliatamente le componenti che lo costituiscono.

Il sistema dovrà prevedere una fase di raccolta dei tweets, con tutte le relative informazioni, ed inoltre sarà costituito da una fase che consenta di attribuire un valore di verità ai tweets raccolti.

3.1 Struttura generale

OMISSIS

3.2 Ottenimento dei tweets

L'input del sistema è costituito da un insieme di parole chiave relative a determinati argomenti, tra cui politica, cronaca, sport e gossip. Questo insieme di parole, o keywords, rappresenta di conseguenza l'input per la prima fase dell'architettura, ovvero l'*Ottenimento dei Tweets*.

Il set di keywords sarà necessario per la raccolta dei tweets; infatti, come sarà descritto, è possibile ottenere dei tweets attraverso delle API messe a disposizione dalla piattaforma Twitter. In dettaglio, è possibile effettuare una ricerca per parole. Per ciascuna parola dell'insieme di keywords, sarà effettuata una richiesta e saranno restituiti i tweets che contengono la keyword. Di conseguenza, le parole chiave devono essere scelte coerentemente alle tematiche di cui si intende raccogliere i tweets.

La prima fase del sistema proposto consiste nell'ottenimento dei tweets a partire dall'insieme di keywords di input. La piattaforma Twitter mette a disposizione delle API attraverso cui è possibile ottenere determinati tweets in modo semplice ed unico. Per potere utilizzare queste API, è necessario richiedere un account Twitter per

sviluppatori [97]. Una volta accettata la richiesta, saranno associate all'account quattro chiavi:

- consumer_key;
- consumer_secret;
- access_token;
- access_token_secret.

Lo scopo di queste chiavi è quello di autorizzare le richieste alla piattaforma. Le API forniscono accesso ad un set di metodi che consentono di ottenere un gran numero di informazioni inerenti ai tweets. La risposta è infatti costituita da un oggetto json, rappresentativo del tweet, contenente tutti gli attributi relativi, tra cui: testo del tweet, id del tweet, data di creazione, url o menzioni presenti nel tweet, numero di like, numero di retweets, media allegati al tweet, hashtag e tanto altro. Inoltre sono presenti anche informazioni riguardanti l'utente che ha pubblicato il tweet: nome dell'utente, id dell'utente, data di creazione dell'account, numero di followers, numero di friends, timeline dell'utente ecc.

I tweets possono essere ottenuti o a partire dal loro identificativo, o a partire da una parola presente nel tweet. In questo caso i tweets saranno ottenuti a partire da un insieme di keywords, sia perché non si hanno a disposizione id dei tweets, sia perché in questo modo si possono ottenere tweets relativi ad un determinato argomento.

Dunque, per ogni keyword presente nell'insieme di input, bisogna effettuare una richiesta alla piattaforma Twitter tramite il relativo metodo delle API. Per la raccolta dei dati, è stata utilizzata la libreria python *tweepy* che consente di accedere all'API di Twitter in modo semplice [98]. L'autenticazione è gestita dalla classe *tweepy.OAuthHandler()*

```
1
2  auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
3  auth.set_access_token(access_token, access_token_secret)
4  api = tweepy.API(auth)
5
```

Una volta eseguita l'autenticazione, la variabile *api* può essere utilizzata per effettuare richieste tramite i metodi di cui dispone. In questo caso, è stato utilizzato il metodo *api.search()* che appunto consente di ottenere i tweets che contengono la parola passata come argomento alla funzione. La risposta è costituita da una lista di tweets. In questa fase, vengono considerati tutti gli attributi più importanti richiesti dagli algoritmi di rilevamento di fake news (informazioni sui tweets, sui retweets, sugli utenti ecc.). A tal scopo i tweets vengono raccolti in tre file:

- *Tweets.json*, che comprende i tweets ottenuti (non i retweets);
- *Retweets.json*, che comprende i retweets;
- *Original_ Tweets.json*, che comprende i tweets originali che sono stati retwittati.

Successivamente, tutti i tweets presenti nei tre file vengono concatenati all'interno di un nuovo file: ***All_ Tweets.json***. Il file costituisce un oggetto che presenta due campi:

- "ALL_TWEETS_COUNT", che rappresenta il numero di tweets totali raccolti con le API Twitter;
- "ALL_TWEETS", che rappresenta la lista degli oggetti rappresentativi dei tweets raccolti.

Di seguito è riportata la rappresentazione di ogni tweet in formato json.

```
{
  "TEXT": "testo_del_tweet",
  "ID": 1470554*******32403,
  "ID_STR": "1470554*******32403",
  "IS_RETWEET": "False",
  "ORIGINAL_TWEET_ID": "-",
  "CREATED_AT": "aaaa-mm-gg hh:mm:ss",
  "LANG": "en",
  "SOURCE_URL": "https://******",
  "RETWEET_COUNT": ***,
  "FAVORITE_COUNT": ***,
  "ENTITIES": {
    "HASHTAGS": [],
    "USER_MENTIONS": [],
    "MEDIA": [],
    "URLS": []
  },
  "USER": {
    "ID": 2475****94,
    "ID_STR": "2475****94",
    "NAME": "name_of_user",
    "SCREEN_NAME": "screen_name_of_user",
    "CREATED_AT": "aaaa-mm-gg hh:mm:ss",
    "LANG": null,
    "URL": "https://******",
    "PROTECTED": false,
    "FOLLOWERS_COUNT": ****,
    "FRIENDS_COUNT": ****,
    "FAVOURITES_COUNT": ****
  }
}
```

Il dataset iniziale ottenuto costituirà l'input della seconda fase dell'architettura, ovvero la *Topic Detection*, e da questo seguiranno tutte le procedure necessarie alla determinazione della ground truth.

3.3 Topic detection

OMISSIS

3.4 Topic summarization

Dopo che i tweets sono stati raggruppati in cluster, lo step successivo consiste nel determinare il claim che è rappresentato da ogni singolo cluster ottenuto.

In generale, la summarization consiste nel trasformare uno o più testi in un testo più breve che li riassuma. In questo contesto, consiste quindi nel trasformare un insieme di tweet che costituiscono un cluster in un claim. I metodi di summarization possono essere raggruppati in due categorie principali [100]:

- metodi estrattivi, cioè metodi tradizionali in cui l'obiettivo principale è quello di identificare le frasi più significative del testo ed aggiungerle alla sintesi. Ne sono esempi TextRank, LexRank, LSA, Luhn e KL-Sum;
- metodi astrattivi, cioè metodi più avanzati in cui l'obiettivo è quello di identificare le sezioni più importanti del testo, interpretare il contesto e riprodurlo in un modo nuovo. Alcuni modelli sono i transformers Bart, T5, GPT-2 ecc.

OMISSIS

3.5 Estrazione delle keywords

OMISSIS

3.5.1 Natural Language Processing

A differenza dei linguaggi di programmazione, che seguono regole ben precise e sono facilmente interpretabili dalle macchine, la lingua utilizzata dagli umani non è facilmente rappresentabile in quanto è meno strutturata rispetto ad un linguaggio di programmazione [101]. La linguistica computazionale si concentra sullo studio del funzionamento del linguaggio naturale in modo da elaborare programmi eseguibili dalle macchine. Il Natural Language Processing (NLP) si occupa principalmente di testi, intesi come sequenze di parole che esprimono uno o più messaggi (es. email, tweet, pagine web). I task di NLP sono numerosi e comprendono il riconoscimento della lingua, l'analisi semantica e l'analisi del sentiment [102]. L'estrazione delle keywords significative dai claims è stata eseguita per mezzo di una libreria gratuita per Natural Language Processing (denominata spaCy) con molte funzionalità integrate [103]. Il modulo comprende diversi tipi di modelli: ad esempio, il modello predefenito della lingua inglese è "en_core_web_sm".

```
1
2   import spacy
3   nlp = spacy.load("en_core_web_sm")
4   text = "Hello, ..."
5   obj = nlp(text)
6
```

Il codice riportato mostra come utilizzare questa libreria. L'oggetto *nlp* è un'istanza del modello di linguaggio; la variabile *text* rappresenta una stringa contenente il testo che si intende processare; infine, la variabile *obj* rappresenta l'oggetto in cui viene convertito il testo, ovvero una lista di token caratterizzati da relativi attributi

necessari all'elaborazione. In questo caso, poichè l'obiettivo è quello di considerare solo le parole significative di una frase, vengono utilizzati principalmente tre attributi:

OMISSIS

3.6 Ottenimento dei fatti dai siti di fact-checking

OMISSIS

3.7 Determinazione della Ground Truth

OMISSIS

3.8 OUTPUT: Dataset annotato

A questo punto, ad ogni claim del dataset di partenza, e quindi ad ogni tweet appartenente al relativo claim, è assegnato un valore di verità tra Vero, Falso, Undetermined e Unknow. I tweets con attributo Undetermined e Unknow vengono scartati e vengono considerati solo quelli con l'attributo Vero o Falso, in modo da potere valutare le prestazioni degli algoritmi di fake news detection eseguiti sul dataset. Il dataset finale è rappresentato nel file ***Dataset.json***, in cui sono contenuti gli oggetti json rappresentativi dei tweets annotati, con i relativi attributi. L'oggetto json comprende tre campi:

OMISSIS

Gli oggetti rappresentativi dei tweets hanno i seguenti attributi:

OMISSIS

Oltre al dataset contenente i tweets, è stato raccolto un ulteriore set di dati che comprende, per ogni utente dei tweets, una lista di seguiti, in modo da potere utilizzare il dataset anche in algoritmi che tengono conto delle relazioni tra gli utenti. Questa fase di ottenimento dei seguiti non è considerata una vera e propria componente dell'architettura del sistema di raccolta dei tweets in quanto non è utile alla determinazione della ground truth, ma solo agli algoritmi che tengono conto di come gli utenti di Twitter si influenzano tra loro. A tal scopo, come per la raccolta dei tweets, è stata utilizzata la libreria python *tweepy* [98]. In questo caso, è stato utilizzato il metodo *api.friends()* che appunto consente di ottenere i 20 seguiti più recenti dall'utente con id specificato come argomento del metodo. Il file è denominato ***Friends.json*** ed è costituito da una lista di oggetti json con i seguenti attributi:

OMISSIS

Il dataset ottenuto dal sistema in questione sarà utilizzato come input per gli algoritmi di fake news detection descritti nel Capitolo 4 e le prestazioni saranno valutate e discusse nel Capitolo 5.

Capitolo 4

Algoritmi di Fake News Detection

Nel presente Capitolo, saranno descritti dei modelli di rilevamento di notizie false che circolano nei social media, precisamente su **Twitter**. Si tratta di algoritmi già esistenti in letteratura, che sono stati in questo caso implementati al fine di potere utilizzare il sistema di raccolta discusso nel Capitolo 3. Gli algoritmi di truth-finding che saranno presentati sono di tipo non supervisionato; sono dunque tecniche di apprendimento automatico (Machine Learning) in cui vengono forniti in input solo esempi non annotati, in quanto le etichette non sono note a priori e vengono apprese automaticamente dai modelli. Gli approcci che saranno presentati hanno il fine di determinare:

- il valore di verità delle affermazioni;
- il significato di alcuni indicatori di veridicità, come l'attendibilità della fonte.

Il Capitolo sarà strutturato come segue: nella prima parte sarà presentato il tipo di dato necessario all'esecuzione dei modelli, ovvero le matrici di dati che costituiranno l'input degli approcci; in seguito, sarà presentata una spiegazione intuitiva dell'algoritmo statistico di **Expectation Maximization** che rappresenterà il fulcro dei modelli, e che sarà utile per introdurre gli approcci veri e propri di fake news detection nelle successive Sezioni.

4.1 Caratteristiche comuni

Gli approcci di truth-finding qui considerati sfruttano informazioni relative alle associazioni tra "soggetti e claims" e "features e claims", nonché la relazione tra l'autore del claim ed il suo antenato nel grafo sociale. Tali caratteristiche vengono rappresentate mediante l'utilizzo di tre matrici: **SC**, **FC**, e **D**.

4.1.1 Associazione tra "soggetti e claims"

OMISSIS

4.1.2 Associazioni tra "features e claims"

Per quanto riguarda la realizzazione della matrice FC, bisogna stabilire quali features considerare negli approcci. In questi modelli, le K caratteristiche che entrano in gioco sono tre:

OMISSIS

Per quanto riguarda la terza feature (claim riportato da almeno due fonti indipendenti), occorrono ulteriori osservazioni. Gli utenti iscritti nei social media, Twitter nel caso in questione, possono essere rappresentati da un grafico di influenza o grafo sociale. Gli utenti costituiscono i nodi del grafo mentre gli archi indicano la relazione che li lega. Gli archi sono orientati e rappresentano appunto l'influenza che ha un utente sull'altro. Precisamente, gli archi stabiliscono una relazione antenato-discendente tra gli utenti coinvolti. Se due utenti, rappresentati da due nodi nel grafo sociale, non sono collegati da un arco, allora i due utenti sono **indipendenti**. Invece, se esiste un arco da $S1$ a $S2$ nel grafo sociale, allora $S1$ rappresenta un antenato per $S2$, che per lo stesso motivo, è un discendente di $S1$. Il grafico d'influenza illustra quindi queste relazioni tra gli utenti e può essere realizzato in tre modi, specifici per Twitter [106]:

- La prima modalità di realizzazione si basa sulla relazione followers-seguiti. Se un utente $S1$ segue un utente $S2$, e quindi è follower di $S2$, allora esiste un arco nel grafo sociale che parte da $S2$ e punta su $S1$, in quanto l'utente $S2$ "influenza" $S1$ e quindi rappresenta per lui un antenato.

- La seconda modalità di realizzazione si basa sul comportamento di retweet degli utenti. In questo caso, esiste un arco da $S1$ a $S2$ se l'utente $S2$ retwitta alcuni tweets della fonte $S1$. Anche in questo caso $S1$ rappresenta quindi un antenato di $S2$.

- La terza modalità combina le due precedenti e quindi esiste un arco da $S1$ a $S2$ nel grafo se $S2$ è follower di $S1$ e retwitta alcuni suoi tweets.

Pertanto, anche la terza feature è un'informazione binaria, e indica se un determinato claim è riportato da almeno due fonti indipendenti. Ad esempio, se il claim rappresentato dalla quarta colonna della matrice FC è riportato da due autori indipendenti, allora la cella $F2C3$ sarà settata a 1; viceversa, avrà valore 0.

4.1.3 Relazione tra autore ed antenati

La terza ed ultima matrice (matrice D) ha le stesse dimensioni della matrice SC, poiché riporta in maniera analoga sulle righe gli autori e sulle colonne i claims. In questo caso però l'informazione contenuta non è la diretta relazione tra autore e asserzione, bensì la relazione tra antenato dell'autore e claim. Infatti, la cella della matrice D corrispondente alla fonte Si e all'asserzione Cj avrà valore 1 nel caso in cui esiste almeno un antenato di Si che riporta l'affermazione Cj. Al contrario, quando nessun antenato di Si riporta il claim Cj, allora si avrà che $Dij = 0$. Da notare che, se $SiCj = 1$ e $Dij = 1$, la sorgente Si potrebbe non agire in modo indipendente nell'affermare Cj. Piuttosto, potrebbe semplicemente ripetere Cj perché un suo

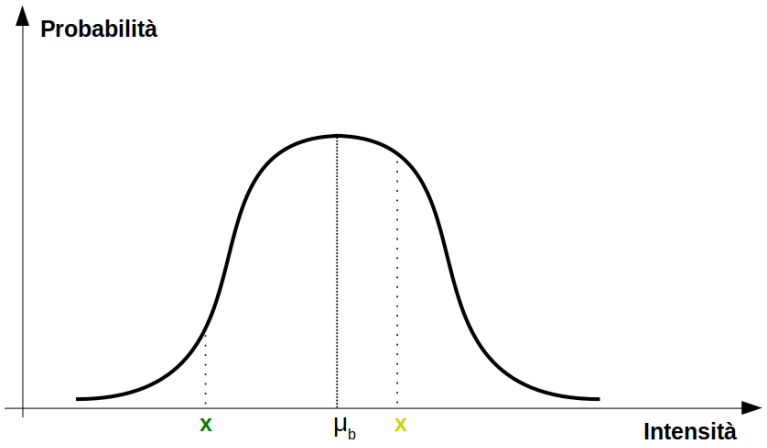

Figura 4.1: Distribuzione di probabilità delle stelle blu.

antenato ha fatto la stessa affermazione. Ciò ha implicazioni sulla veridicità delle affermazioni.

4.2 Expectation Maximization - Spiegazione intuitiva

In questa sezione, sarà spiegato il funzionamento dell'algoritmo di Expectation Maximization (EM) attraverso un esempio intuitivo, che riguarda la classificazione di alcune tipologie di stelle. L'unica informazione disponibile è che esistono diverse tipologie di stelle ma non è possibile distinguerle a priori. L'obiettivo del task è quello di classificare i punti di intensità e caratterizzare le categorie a cui appartengono; in altre parole, il fine consiste nel classificare ogni stella considerata e ricavare la distribuzione di probabilità di ogni categoria. Si ipotizza che esistono soltanto due tipi di stelle:

- stelle rosse;
- stelle blu.

Avendo un insieme di stelle $X = X1, X2, ..., XN$, lo scopo dell'algoritmo è di stabilire se ogni stella appartenente all'insieme X è una stella rossa o una stella blu, ed inoltre deve ottenere la distribuzione di probabilità sia delle stelle rosse che delle stelle blu, al variare dell'intensità di esse. Ad esempio, osservando la Figura 4.1, che illustra la distribuzione di probabilità delle stelle blu, la stella raffigurata in giallo è molto probabile che sia una stella blu poiché è vicina alla media, mentre la stella in verde è poco probabile che sia classificata come una stella blu in quanto è lontana dal valore medio della distribuzione. Poiché l'unica informazione disponibile è la conoscenza del fatto che esistono due categorie di stelle, e non si sa nulla nè sull'appartenenza delle stelle alle categorie, nè sulla distribuzione di probabilità delle categorie, sembrerebbe un problema impossibile da risolvere. Quest'ultimo può essere scomposto in due parti, spiegate di seguito.

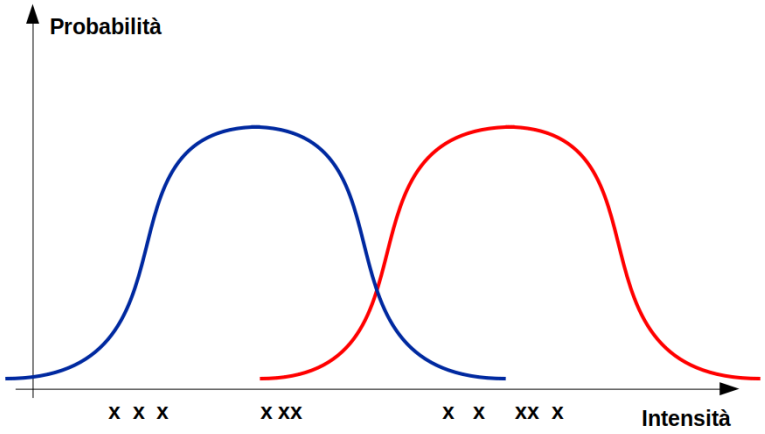

Figura 4.2: Distribuzioni di probabilità note a priori.

4.2.1 Caso 1 - Distribuzioni di probabilità note

Se è nota la distribuzione di probabilità delle categorie, classificare le stelle diventa banale (Figura 4.2). Assumendo che le distribuzioni in questione sono Gaussiane, sono caratterizzate da una propria media e da una propria deviazione standard. Precisamente, saranno indicate con μ_r e σ_r rispettivamente la media e la deviazione standard della distribuzione relativa alle stelle rosse, mentre μ_b e σ_b rappresenteranno rispettivamente la media e la deviazione standard della distribuzione relativa alle stelle blu. Per ogni stella appartenente all'insieme di dati disponibili, occorre calcolare la probabilità che sia una stella rossa e la probabilità che sia una stella blu. Indicando Dr la distribuzione delle stelle rosse, e con Db la distribuzione delle stelle blu, si avrà che:

$$Pr(Xi \in Dr) = \alpha e^{-(Xi-\mu_r)^2/\sigma_r^2} \tag{4.1}$$

$$Pr(Xi \in Db) = \alpha e^{-(Xi-\mu_b)^2/\sigma_b^2} \tag{4.2}$$

Quindi, la probabilità che una stella Xi sia una stella rossa è proporzionale ad e elevato ad una quantità che dipende dai parametri della distribuzione (μ_r e σ_r) e dal valore di Xi. Allo stesso modo la probabilità che la stella Xi sia una stella blu è proporzionale ad e elevato ad una quantità che dipende da Xi e dai valori di μ_b e σ_b della distribuzione Gaussiana relativa alle stelle blu. Se la probabilità $Pr(Xi \in Dr)$ è maggiore di $Pr(Xi \in Db)$, la stella sarà classificata come una stella rossa; al contrario, se $Pr(Xi \in Db)$ è maggiore di $Pr(Xi \in Dr)$, la stella sarà classificata come una stella blu. Quindi, se sono note le distribuzioni di probabilità che caratterizzano le categorie, la classificazione diventa un'operazione abbastanza semplice.

4.2.2 Caso 2 - Tipologia delle stelle nota

Se non si è a conoscenza delle distribuzioni di probabilità che descrivono le categorie, ma si conosce la tipologia di ogni stella dell'insieme di dati X, come mostrato dalla Figura 4.3, la caratterizzazione delle categorie diventa, anche in questo caso, piuttosto banale. Infatti, occorre calcolare μ_b e σ_b, e cioè la media e la deviazione

Figura 4.3: Tipologia delle stelle nota a priori.

standard dei punti classificati come stelle blu, e analogamente, calcolare μ_r e σ_r dei punti classificati come stelle rosse. Essendo N_b il numero di stelle blu, e N_r il numero di stelle rosse, i parametri possono essere così calcolati:

$$\mu_b = \sum_{i=1}^{N_b} Xi/N_b \tag{4.3}$$

$$\sigma_b = \sqrt{\sum_{i=1}^{N_b} (Xi - \mu_b)^2/N_b} \tag{4.4}$$

$$\mu_r = \sum_{i=1}^{N_r} Xi/N_r \tag{4.5}$$

$$\sigma_r = \sqrt{\sum_{i=1}^{N_r} (Xi - \mu_r)^2/N_r} \tag{4.6}$$

A questo punto, diventa semplice ottenere le due distribuzioni:

$$Db = 1/\sqrt{2\pi\sigma_b^2} * e^{-(x-\mu_b)^2/2\sigma_b^2} \tag{4.7}$$

$$Dr = 1/\sqrt{2\pi\sigma_r^2} * e^{-(x-\mu_r)^2/2\sigma_r^2} \tag{4.8}$$

Quindi, se si hanno a disposizione i punti classificati, determinare le distribuzioni di probabilità delle tipologie delle stelle è un operazione elementare.

4.2.3 Expectation Maximization

In questa sottosezione, sarà descritto l'algoritmo di Expectation Maximization che consentirà di ottenere sia la classificazione delle stelle, sia la distribuzione di probabilità delle tipologie di stelle, non note a priori. La procedura si divide in due step:

- Expectation (E-Step): è nota la distribuzione di probabilità delle categorie e si deve determinare la tipologia di ogni punto dell'insieme X (Caso 1);
- Maximization (M-Step): è nota la categoria di ogni stella e si devono determinare le distribuzioni di probabilità che caratterizzano le tipologie di stelle (Caso 2).

Si suppone di avere un insieme di tanti punti che variano di intensità, unica variabile nota. Il primo passo è quello di inizializzare le distribuzioni di probabilità in maniera random, in quanto queste non sono note a priori (Figura 4.4). Dunque occorre scegliere dei valori casuali di μ_b, σ_b, μ_r e σ_r. Pertanto, avendo stabilito le distribuzioni,

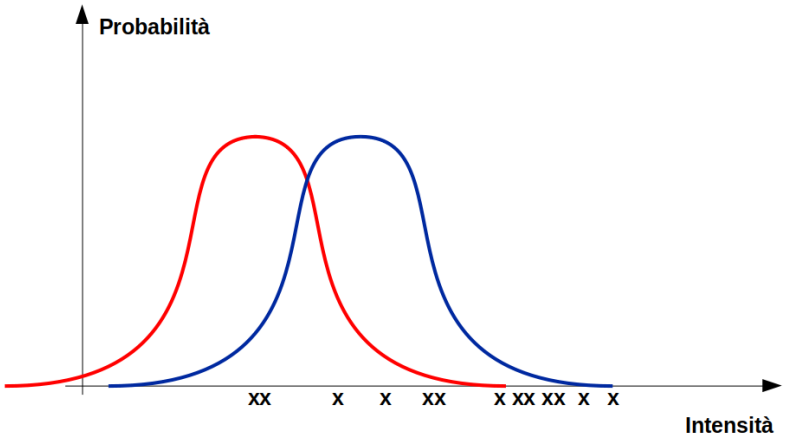

Figura 4.4: Inizializzazione delle distribuzioni di probabilità.

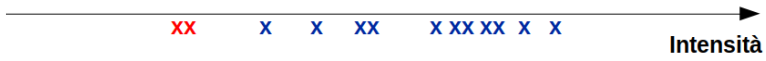

Figura 4.5: Tipologia delle stelle determinata con l'E-Step.

l'**E-Step** prevede di effettuare la classificazione e determinare la tipologia di ogni stella, ripetendo la procedura descritta nel Caso 1. L'output è mostrato nella Figura 4.5. Successivamente, con i punti classificati nel passo di Expectation, si effettua una ricaratterizzazione delle tipologie di stelle, tramite l'**M-Step**, e quindi si ridetermi-nano le due distribuzioni di probabilità, ricalcolando i parametri come descritto nel Caso 2 e mostrato in Figura 4.6. L'Expectation Maximization è iterativo e quindi consiste in ripetute procedure di E-Step ed M-step: ottenute le distribuzioni, que-ste vengono utilizzate per effettuare la classificazione nel passo di Expectation della successiva iterazione. L'algoritmo termina quando le distribuzioni convergono, non mutano tra un'iterazione e l'altra. Dalla Figura 4.6 si può osservare che il punto x_3, che al passo di Expectation era classificato come stella blu, si trova a cavallo tra le due distribuzioni ottenute nel passo di Maximization; inoltre, nella successiva iterazione dell'algoritmo (non illustrata) il punto sarà classificato come stella rossa, poichè si trova più vicino a μ_r rispetto a μ_b e quindi ciò comporterà una distribuzio-ne di stelle rosse spostata verso destra. La procedura continua fino a quando le due distribuzioni non variano più.

4.3 Approccio a massima verosimiglianza

In questa Sezione, sarà descritto il primo dei modelli presentati nel presente lavoro per il rilevamento di fake news nei social media, in particolare su Twitter. Sarà trattata una versione minimalista dell'algoritmo mentre la versione più estesa sarà discussa insieme alle sue varianti nelle successive Sezioni. L'approccio presentato è stato proposto da Dong Wang, Lance Kaplan, Hieu Le e Tarek Abdelzaher [107]. L'articolo che propongono gli autori appena menzionati tratta della scoperta della verità dai dati rumorosi di social sensing. Il social sensing si riferisce all'uso di utenti

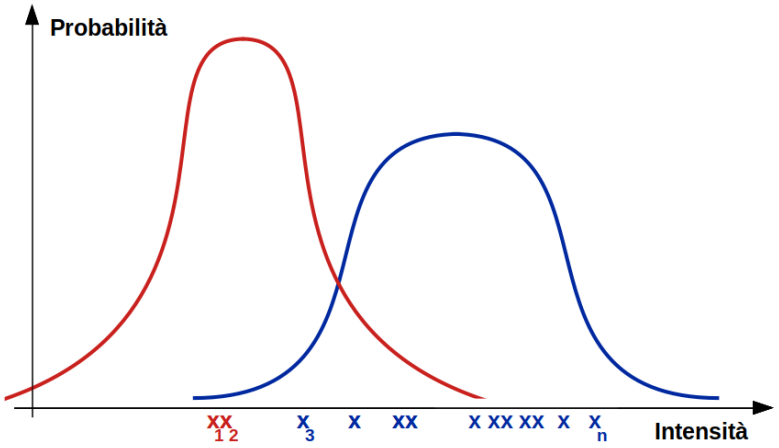

Figura 4.6: Determinazione delle distribuzioni di probabilità.

sui social media come "sensori" che riportano osservazioni sul mondo fisico. Infatti, i social sono ormai diventati la principale fonte di informazioni in cui gli utenti riportano "misurazioni" e cioè osservazioni su tutto ciò che li circonda, come eventi sportivi, catastrofi naturali, notizie di gossip, aggiornamenti politici, economici, sanitari, incidenti e tanto altro.

La principale sfida consiste nel riuscire a sviluppare dei sistemi che siano in grado sia di determinare la correttezza delle affermazioni fatte dagli autori (in letteratura *claims*), sia di determinare l'affidabilità degli utenti. Infatti, a differenza dei sensori fisici che effettuano misurazioni accurate in quanto possono essere calibrati e ben testati, i sensori umani non sono così perfetti e dunque le misurazioni che effettuano sono "rumorose", la probabilità che le misurazioni delle fonti siano corrette è sconosciuta. Gli autori, nel loro articolo, dimostrano come usare queste misurazioni incerte, tramite un approccio di stima a massima verosimiglianza, affinché si possa determinare sia l'affidabilità delle fonti che la correttezza delle loro misurazioni (osservazioni).

Nel caso dei social media, le misurazioni tenute in considerazione sono binarie: le fonti possono riportare se una determinata affermazione è vera o se è falsa. Si noti che un modo immediato per risolvere i problemi di truth-finding potrebbe essere quello di considerare come vere solo quelle affermazioni che sono riportate da un numero sufficiente di fonti; il problema di questo approccio, simile ad un sistema di voto, non è efficace poiché non riesce a determinare l'affidabilità delle sorgenti e dunque potrebbe accadere che le affermazioni riportate da tante sorgenti poco affidabili vengono considerate vere, mentre vengono considerati falsi i claims riportati da pochi autori ma affidabili. In questo caso invece si stimano in modo congiunto le due variabili incognite (affidabilità sorgenti e correttezza affermazioni). Il modo più adatto a risolvere queste sfide è l'algoritmo di **Expectation Maximization**, già spiegato in maniera intuitiva nella Sezione 4.2. Infatti, nel caso dell'esempio delle stelle, si doveva determinare congiuntamente la distribuzione di probabilità relativa alla tipologia di stelle e la categoria di ogni stella, avendo a disposizione soltanto l'intensità di ogni stella; nel caso del social sensing invece si devono determinare la tipologia dei claims (veri o falsi) e l'affidabilità degli utenti che riportano i claims in

questione.

4.3.1 Dati di input e considerazioni iniziali

Nell'algoritmo in questione, l'input è costituito esclusivamente dalla matrice delle osservazioni SC, già definita nella Sezione 4.1.1. Brevemente, è una matrice composta da 1 e 0 che riporta sulle righe le sorgenti e sulle colonne i claims; se la fonte i-esima (Si) riporta l'affermazione j-esima (Cj), la cella $SiCj$ della matrice SC sarà settata a 1; viceversa, sarà settata a 0. Dunque, si è a disposizione di un insieme di N utenti ($S = S1, S2, ..., SN$) che effettuano delle osservazioni individuali su un insieme di M variabili misurate nel loro ambiente ($C = C1, C2, ..., CM$). Ad esempio, dei conducenti potrebbero segnalare dei punti di un'autostrada che necessitano di riparazione oppure un gruppo di cittadini potrebbe indicare dei graffiti offensivi nel loro quartiere. Anche in questo caso quindi, l'informazione sarà binaria e indicherà la presenza o l'assenza della condizione di pericolo, nel primo esempio, o di offesa, nel secondo.

Assumendo che ogni utente può riportare un sottoinsieme di tutte le misurazioni effettuate, e che la stessa osservazione può essere segnalata da più sorgenti, il problema consiste nel determinare quali sono le osservazioni vere e false, non avendo alcuna informazione circa l'affidabilità delle fonti.

L'approccio sarà adesso spiegato mediante delle rappresentazioni formulari e osservazioni statistiche. Come già accennato, la sorgente i-esima è indicata con Si, mentre l'osservazione j-esima con Cj. $P(SiCj)$ rappresenta la probabilità che l'utente i-esimo riporta il j-esimo claim, e cioè il *reporting rate*; $P(Cj^t)$ costituisce la probabilità che il j-esimo claim sia vero, invece $P(Cj^f)$ indica la probabilità che sia falso. Lo scopo dell'algoritmo è determinare il valore di vertà di ogni claim Cj, non essendo nota a priori l'affidabilità della fonte, indicata con:

$$ti = P(Cj^t|SiCj) \tag{4.9}$$

Inoltre definiamo altri due parametri relativi ad ogni sorgente, ovvero ai che rappresenta la probabilità che l'utente riporta il j-esimo claim quando questo è vero, e bi che indica la probabilità che l'utente riporta il j-esimo claim quando invece è falso:

$$ai = P(SiCj|Cj^t) \tag{4.10}$$

$$bi = P(SiCj|Cj^f) \tag{4.11}$$

Applicando il teorema di Bayes, ai e bi sono equivalenti a:

$$ai = P(SiCj|Cj^t) = \frac{P(SiCj, Cj^t)}{P(Cj^t)} = \frac{P(Cj^t|SiCj)P(SiCj)}{P(Cj^t)} \tag{4.12}$$

$$bi = P(SiCj|Cj^f) = \frac{P(SiCj, Cj^f)}{P(Cj^f)} = \frac{P(Cj^f|SiCj)P(SiCj)}{P(Cj^f)} \tag{4.13}$$

$$ai = \frac{ti \cdot si}{d} \tag{4.14}$$

$$bi = \frac{(1 - ti) \cdot si}{1 - d} \tag{4.15}$$

dove d è la probabilità a priori che una j-esima affermazione sia vera, mentre si rappresenta il reporting rate ed è calcolabile dalla matrice SC come la frazione tra il numero di osservazioni riportate dall'utente in questione e il totale delle osservazioni effettuate da tutte le fonti. Dal punto di vista matematico, il problema consiste nel determinare la probabilità di correttezza di ogni claim e l'affidabilità di ogni fonte in modo da massimizzare la probabilità dei dati; in altre parole, essendo hj la stima di ogni variabile misurata Cj e ei l'affidabilità della sorgente i-esima, formalmente si devono determinare i vettori H ed E tali che:

$$< H^*, E^* >= argmax_{<H,E>} P(SC|H, E) \tag{4.16}$$

L'Expectation Maximization è quindi un algoritmo per trovare la stima di massima verosimiglianza dei parametri in un modello statistico.

4.3.2 Formulazione del problema

OMISSIS

4.3.3 Pseudocodice

OMISSIS

4.3.4 Inizializzazione di θ con *ai* e *bi* uguali

OMISSIS

4.4 Scoperta della verità con dati multimodali

In questa Sezione, sarà descritto un ulteriore modello di rilevamento di fake news su Twitter, che costituisce un estensione dell'approccio descritto nella precedente. L'algoritmo in questione è stato presentato in un articolo intitolato *Truth Discovery with Multi-modal Data in Social Sensing* [108]. Anche in questo caso, si tratta di un algoritmo non supervisionato che sfrutta le features di contenuto per valutare la veridicità delle osservazioni fatte dagli utenti nei social media. Infatti, a differenza del precedente approccio di Expectation Maximization della Sezione 4.3 (EM), non utilizza soltanto l'affidabilità della fonte per stimare il valore di verità delle affermazioni effettuate, ma utilizza ulteriori indicatori chiavi di veridicità, o features.

Nell'algoritmo di EM discusso precedentemente, l'unica informazione e quindi l'unico input era costituito dalla matrice SC, che evidenzia la relazione tra autore e claim riportato. Non si tiene quindi in considerazione il legame tra le fonti, l'influenza che può avere un autore nei confronti di un altro autore. Quando una stessa affermazione viene riportata da più autori che sono fra di loro indipendenti, la probabilità che l'affermazione sia vera è più alta rispetto ad un'affermazione riportata da due fonti dipendenti. Inoltre, se un utente Twitter pubblica un tweet su un determinato evento e include un link o un'immagine che testimoniano l'evento, la probabilità che l'evento sia vero è più alta rispetto ad un evento riportato in un tweet che non include nè link nè immagini. Queste considerazioni sono alla base della realizzazione del

nuovo modello di fake news detection. Precisamente, saranno illustrati tre approcci, simili tra loro, che estendono l'algoritmo di Expectation Maximization, ovvero:

- Expectation Maximization with Multiple Features (EM-MultiF);
- Penalized Expectation Maximization (PEM);
- Constrained Expectation Maximization (CEM).

4.4.1 Dati di input e considerazioni iniziali

OMISSIS

4.4.2 Expectation Maximization with Multiple Features

In questa sottosezione sarà descritto il primo approccio di truth-finding, ovvero *Expectation Maximization with Multiple Features*. Per prima cosa, occorre definire una funzione di verosimiglianza che esprime la verosimiglianza delle osservazioni ricevute come funzione dei parametri e variabili da stimare; esattamente, la funzione di likelihood è così definita:

$$L = ln(P(SC, FC; D, \theta)) = ln(\sum_{C \in \{0,1\}} P(SC, FC|C; D, \theta)P(C; D, \theta)) \tag{4.17}$$

Data la log likelihood, il passo di Expectation consiste in:

$$Q(\theta|\theta^t) = \sum_{C \in \{0,1\}} P(C|SC, FC; D, \theta^t) \cdot ln(P(SC, FC|C; D, \theta)P(C; D, \theta))$$

$$Q(\theta|\theta^t) = \sum_{C \in \{0,1\}} P(C|SC, FC; D, \theta^t) \cdot$$

$$ln(P(FC|SC, C; D, \theta)P(SC|C; D, \theta)P(C; D, \theta))$$

$$Q(\theta|\theta^t) = \sum_{j=1}^{M} \sum_{Cj \in \{0,1\}} P(Cj|FCj, SCj; D, \theta^t)\{lnP(Cj; D, \theta) +$$

$$ln(P(FCj|SCj, Cj; D, \theta)P(SCj|Cj; D, \theta))\} \tag{4.18}$$

Nell'Equazione 4.18, la probabilità a posteriori della variabile Cj, può essere riscritta nella seguente maniera:

$$Z(j,t) = P(Cj^t|SCj, FCj; D, \theta^t) = \frac{T(j,t)}{T(j,t) + F(j,t)} \tag{4.19}$$

$$T(j,t) = P(FCj|SCj, Cj^t; D, \theta^t)P(SCj|Cj^t; D, \theta^t)P(Cj^t; D, \theta^t)$$

$$F(j,t) = P(FCj|SCj, Cj^f; D, \theta^t)P(SCj|Cj^f; D, \theta^t)P(Cj^f; D, \theta^t)$$

in cui Cj^t indica la situazione in cui il j-esimo claim è vero, mentre Cj^f indica la situazione in cui il j-esimo claim è falso. La probabilità $P(SCj|Cj; D, \theta^t)$ consiste nella produttoria delle probabilità di reporting di ogni utente, in base ai valori di Cj,

D e θ, mentre $P(FCj|SCj, Cj; D, \theta^t)$ consiste nella produttoria delle probabilità di ogni feature:

$$P(SCj|Cj^t; D, \theta^t) = \prod_{i=1}^{N} (ai^{(t)SiCj}(1 - ai^{(t)})^{(1-SiCj)})^{1-Dij} + (fi^{(t)SiCj}(1 - fi^{(t)})^{(1-SiCj)})^{Dij} \tag{4.20}$$

$$P(SCj|Cj^f; D, \theta^t) = \prod_{i=1}^{N} (bi^{(t)SiCj}(1 - bi^{(t)})^{(1-SiCj)})^{1-Dij} + (gi^{(t)SiCj}(1 - gi^{(t)})^{(1-SiCj)})^{Dij} \tag{4.21}$$

$$P(FCj|SCj, Cj^t; D, \theta^t) = \prod_{k=1}^{K} (pk^{(t)FkCj}(1 - pk^{(t)})^{(1-FkCj)}) \tag{4.22}$$

$$P(FCj|SCj, Cj^f; D, \theta^t) = \prod_{k=1}^{K} (qk^{(t)FkCj}(1 - qk^{(t)})^{(1-FkCj)}) \tag{4.23}$$

dove N indica il numero di autori e K il numero di fetaures. Il passo di Expectation consiste dunque nel calcolare, per ogni claim Cj, la probabilità a posteriori $Z(j,t)$ come mostrato nell'Equazione 4.19. Per quanto riguarda invece il passo di Maximization, questo consiste nel determinare θ^* che massimizza $Q(\theta|\theta^{(t)})$ e utilizzarlo nell'iterazione successiva. Occorre dunque calcolare le derivate parziali di $Q(\theta|\theta^{(t)})$ rispetto a ai, bi, fi, gi, pk, qk e d e porle uguali a 0, ottenendo le seguenti equazioni:

$$ai^{t+1} = \frac{\sum_{Cj \in SiC_1^{D_0}} Z(j,t)}{\sum_{Cj \in SiC_1^{D_0} \cup SiC_0^{D_0}} Z(j,t)} \tag{4.24}$$

$$bi^{t+1} = \frac{\sum_{Cj \in SiC_1^{D_0}} 1 - Z(j,t)}{\sum_{Cj \in SiC_1^{D_0} \cup SiC_0^{D_0}} 1 - Z(j,t)} \tag{4.25}$$

$$fi^{t+1} = \frac{\sum_{Cj \in SiC_1^{D_1}} Z(j,t)}{\sum_{Cj \in SiC_1^{D_1} \cup SiC_0^{D_1}} Z(j,t)} \tag{4.26}$$

$$gi^{t+1} = \frac{\sum_{Cj \in SiC_1^{D_1}} 1 - Z(j,t)}{\sum_{Cj \in SiC_1^{D_1} \cup SiC_0^{D_1}} 1 - Z(j,t)} \tag{4.27}$$

$$pk^{t+1} = \frac{\sum_{Cj \in FkC_1} Z(j,t)}{\sum_{Cj \in FkC_1 \cup FkC_0} Z(j,t)} \tag{4.28}$$

$$qk^{t+1} = \frac{\sum_{Cj \in FkC_1} 1 - Z(j,t)}{\sum_{Cj \in FkC_1 \cup FkC_0} 1 - Z(j,t)} \tag{4.29}$$

$$d^{t+1} = \frac{\sum_{j=1}^{M} Z(j,t)}{M} \tag{4.30}$$

in cui $SiC_1^{D_0} : SiCj = 1$ & $Dij = 0$, $SiC_0^{D_0} : SiCj = 0$ & $Dij = 0$, $SiC_1^{D_1} : SiCj = 1$ & $Dij = 1$, $SiC_0^{D_1} : SiCj = 0$ & $Dij = 1$, $FkC_1 : FkCj = 1$ e $FkC_0 : FkCj = 0$.

4.4.3 Penalized Expectation Maximization

L'idea che sta alla base del *Penalized Expectation Maximization* è che un tweet che contiene più indicatori di verità, ovvero comprende immagini o url che supportano ciò che si sta riportando, è più probabile che sia vero rispetto ad un tweet che non comprende indicatori di verità. Per tale ragione, in questa nuova soluzione si penalizza la probabilità a posteriori di Cj quando sono presenti delle features, nella situazione in cui si ipotizza che Cj è falsa; in questo modo, è più probabile che un'affermazione sia vera quando è supportata da più features. Il fattore di penalità dipende dal numero di features presenti ed è così definito:

$$Penalty\ Factor = \alpha^{n_j(1-Cj)} \tag{4.31}$$

in cui α è un numero compreso tra 0 e 1 e n_j rappresenta il numero di features che sono presenti nel claim Cj. L'Equazione 4.17 può essere così riformulata:

$$L_p = ln(\sum_{Cj\in\{0,1\}} \prod_{j=1}^{M} \prod_{k=1}^{K} \alpha^{n_j(1-Cj)} P(FkCj|SCj, Cj; D, \theta) \cdot$$

$$P(SCj|Cj; D, \theta) P(Cj; D, \theta)) \tag{4.32}$$

Dall'Equazione 4.32 si può constatare che quando $Cj = 1$, il fattore di penalità diventa 1 e dunque indica nessuna penalità; nel caso in cui invece $Cj = 0$, il termine è diverso da 1 e dunque influenza la probabilità che l'affermazione è falsa quando è supportata da features. Più è alto il numero di indicatori di verità presenti, più è bassa la probabilità che il claim sia falso (essendo α compreso tra 0 e 1). Il passo di Expectation consiste dunque nel calcolare la probabilità di ogni affermazione Cj nel seguente modo:

$$Z(j,t) = P(Cj^t|SCj, FCj; D, \theta^t) = \frac{T(j,t)}{T(j,t) + \alpha^{n_j(1-Cj)} F(j,t)} \tag{4.33}$$

$$T(j,t) = P(FCj|SCj, Cj^t; D, \theta^t) P(SCj|Cj^t; D, \theta^t) P(Cj^t; D, \theta^t)$$

$$F(j,t) = P(FCj|SCj, Cj^f; D, \theta^t) P(SCj|Cj^f; D, \theta^t) P(Cj^f; D, \theta^t)$$

in cui Cj^t indica la situazione in cui il j-esimo claim è vero, mentre Cj^f indica la situazione in cui il j-esimo claim è falso. Per quanto riguarda invece il passo di Maximization, questo consiste nel determinare θ^* che massimizza $Q(\theta|\theta^{(t)})$ e utilizzarlo nell'iterazione successiva. Occorre dunque calcolare le derivate parziali di $Q(\theta|\theta^{(t)})$ rispetto a ai, bi, fi, gi, pk, qk e d e porle uguali a 0, esattamente come nell'algoritmo *Expectation Maximization with Multiple Features*. L'M-Step porta agli stessi aggiornamenti già mostrati nelle Equazioni [4.24 - 4.30] e dunque non vengono mostrati per evitare ridondanza.

4.4.4 Constrained Expectation Maximization

Questo algoritmo è l'estensione di un altro algoritmo CEM [110]. Tuttavia, quest'ultimo tiene conto soltanto del comportamento degli utenti e non prende in considerazione le features del contenuto del tweet. L'idea che sta alla base di questo nuovo

approccio è che un affermazione è più probabile che sia vera quando è supportata da più features indipendenti. In particolare, assumiamo che un utente Si effettua una determinata asserzione Cj che comprende più features di contenuto; successivamente, un altro utente Sl indipendente da Si riporta la stessa affermazione Cj supportata da più features. Poiché i due autori sono indipendenti, anche le features di supporto lo sono e dunque la probabilità che Cj sia vera è più alta rispetto alla probabilità che sia falsa. Dunque, si introduce un vincolo che dipende dal numero di features indipendenti, così definito:

$$Constraint = 1 - \lambda^{n_j - 1} \tag{4.34}$$

dove λ è un numero compreso tra 0 e 1 e n_j rappresenta il numero di features di contenuto indipendenti. Il parametro n_j è relativo ad ogni affermazione; per determinarlo, occorre individuare il primo utente $S0$ che la riporta e successivamente bisogna contare, tra tutti gli altri utenti che effettuano l'affermazione riportando features di supporto, quelli che sono indipendenti da $S0$. L'algoritmo di CEM è simile all'algoritmo di EM-MultiF. La differenza consiste nel fatto che, una volta calcolate le probabilità a posteriori $Z(j,t)$ dell'E-Step per ogni claim, bisogna verificare che queste siano maggiori del vincolo:

$$1 - \lambda^{n_j - 1} \geq Z(j,t) \geq 1 \tag{4.35}$$

Dunque, occorre prima calcolare le probabilità a posteriori come nell'Equazione 4.19 e nel caso in cui viene fatta un'affermazione Cj con più features indipendenti ma $Z(j,t)$ è minore del vincolo, si imposta $Z(j,t)$ uguale al vincolo, rispettando la condizione 4.35. Se invece $Z(j,t)$ è maggiore del vincolo, l'aggiornamento non è necessario.

4.4.5 Pseudocodice

OMISSIS

Capitolo 5

Analisi Sperimentale

Nel presente Capitolo, saranno discussi i risultati dell'implementazione degli algoritmi di fake news detection. Dopo una breve spiegazione delle metriche che vengono utilizzate in questo contesto per la valutazione degli approcci, sarà descritta la raccolta di dataset in qualche modo esistenti insieme alla valutazione degli algoritmi su di essi; successivamente sarà presente anche un'analisi ed una valutazione delle prestazioni dei modelli di fake news detection che utilizzano come input il dataset ottenuto dal Sistema di Raccolta dei Dati descritto nel Capitolo 3.

5.1 Metriche di valutazione

Per valutare le prestazioni dei modelli di classificazione e quindi di rilevamento delle fake news, vengono utilizzate le seguenti metriche: Accuracy, Precision, Recall, F1-Score:

- L'Accuracy rappresenta il rapporto tra le previsioni corrette e il totale delle previsioni.
- La Precision invece misura la frazione di notizie false tra tutte quelle annotate come false.
- La Recall consiste nella percentuale di fake news rilevate tra tutte quelle presenti nel dataset.
- L'F1-Score è una metrica più completa in quanto comprende sia Precision che Recall e costituisce una media armonica tra le due.

Poiché, nella maggioranza dei casi, la distribuzione delle fake news è non uniforme e il numero di fake news è spesso minore del numero delle affermazioni reali, occorrono metodi per eliminare questo sbilanciamento che potrebbe influenzare le metriche di valutazione. Il primo è il sotto-campionamento, che consiste nel sotto-campionare le informazioni vere nel training set, rimuovendo alcuni campioni veri dal training set in modo tale da avere un numero simile di fake news e notizie vere. Il secondo metodo è il sovra-campionamento, che consiste nel sovra-campionare i campioni di fake news nel dataset. Entrambi i metodi presentano il difetto dovuto al fatto che alterano le

		True Class	
		Positive	**Negative**
Hypothetised Class	**Positive**	True positives	False positives
	Negative	False negatives	True negatives

Tabella 5.1: Matrice di Confusione.

informazioni del training set e ciò potrebbe comportare mancanza di informazioni o overfitting.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{5.4}$$

Dunque, le 4 metriche di valutazione appena descritte sono quelle che saranno utilizzate per testare gli algoritmi sia su dataset esistenti che sul dataset ottenuto dal Sistema di Raccolta dei Dati.

5.2 Analisi su dataset pubblici

In questa Sezione, sarà discussa l'implementazione degli algoritmi di fake news detection sui dataset pubblici utilizzati a tal scopo. Inizialmente, sarà illustrata la procedura di raccolta dei dataset e successivamente saranno riportate le valutazioni delle prestazioni degli algoritmi, con una descrizione dettagliata sulla configurazione dei parametri e sui risultati ottenuti.

5.2.1 Raccolta dei dataset

OMISSIS

Raccolta Dataset 1

OMISSIS

Raccolta Dataset 2

OMISSIS

Comparazione tra i dataset

OMISSIS

	DATASET	
	Dataset 1	**Dataset 2**
Users	495	1554
Tweets	682	4285
Claims Totali	236	806
Claims Veri	101	168
Claims Falsi	135	638
Features	3	3

Tabella 5.2: Informazioni sui due dataset raccolti.

	DIM MATRIX	
	Dataset 1	**Dataset 2**
SC	495 × 236	1554 × 806
FC	3 × 236	3 × 806
D	495 × 236	1554 × 806

Tabella 5.3: Dimensioni delle Matrici.

	DENSITY MATRIX		
	SC	**FC**	**D**
Numero di 1 - Dataset 1	591	303	0
Numero di 0 - Dataset 1	116229	405	116820
Numero di 1 - Dataset 2	3246	1235	0
Numero di 0 - Dataset 2	1249278	1183	1252524

Tabella 5.4: Densità delle Matrici.

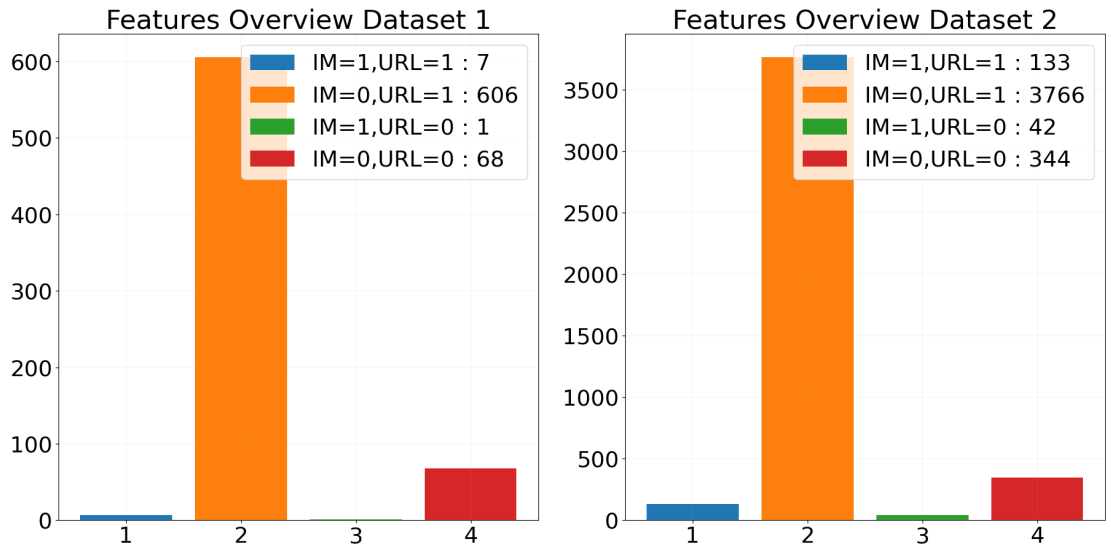

Figura 5.1: Frequenza di immagini e url nei due dataset.

5.2.2 Valutazione degli approcci

In questa sezione, sarà descritta l'implementazione degli approcci illustrati precedentemente. In particolare, nella prima parte saranno spiegate alcune scelte relative all'implementazione, come la determinazione degli iperparametri dei modelli; nella seconda parte saranno invece valutati e discussi i risultati ottenuti dagli algoritmi attraverso il calcolo delle metriche di valutazione, ovvero Accuracy, Precision, Recall ed F1-Score.

Scelte d'implementazione

OMISSIS

Determinazione della frazione

OMISSIS

Determinazione di α e λ

OMISSIS

Risultati ottenuti

Dopo avere effettuato la determinazione dei parametri attraverso una valutazione delle prestazioni degli approcci, eseguiti sul Dataset 1, si può procedere con l'esecuzione dei medesimi sul secondo set di dati, descritto nella Sezione 5.2.1. Inoltre, invece di considerare soltanto l'intero dataset, gli algoritmi vengono testati anche su sottoparti di questo, in modo da analizzare come i dati di input influenzano la classificazione delle fake news. Inizialmente, sarà descritto come vengono estratte le sottoparti dal dataset completo; successivamente saranno discussi i risultati della classificazione degli algoritmi di fake news detection proposti.

Sottoparti del dataset

Il Dataset 2, utilizzato come test set, viene diviso in quattro sottoinsiemi, affinché gli approcci possano essere eseguiti e valutati su ogni sottoparte estratta dal dataset di partenza. Precisamente, vengono create delle sottomatrici SC, FC, D, a partire da quelle relative all'intero set di dati. A tal scopo, l'insieme dei claims che lo costituiscono viene diviso in due sottoinsiemi, così come l'insieme degli autori; in altre parole, avendo due insiemi C e S rappresentanti rispettivamente claims e autori così costituiti:

$$C = \{C_0, C_1, C_2, ..., C_{M-1}\} \tag{5.5}$$

$$S = \{S_0, S_1, S_2, ..., S_{N-1}\} \tag{5.6}$$

dove M è il numero totale di affermazioni, ed N il numero totale di fonti, vengono realizzati da questi altri quattro insiemi:

$$C1 = \{C_0, C_1, C_2, ..., C_{\frac{M}{2}}\} \tag{5.7}$$

GROUND TRUTH DATASET 2					
	Parte 1	**Parte 2**	**Parte 3**	**Parte 4**	**Completo**
Claims Veri	156	12	156	12	168
Claims Falsi	247	391	247	391	638

Tabella 5.5: Ground Truth Dataset 2.

$$C2 = \{C_{\frac{M}{2}+1}, C_{\frac{M}{2}+2}, C_{\frac{M}{2}+3}, ..., C_{M-1}\} \tag{5.8}$$

$$S1 = \{S_0, S_1, S_2, ..., S_{\frac{N}{2}}\} \tag{5.9}$$

$$S2 = \{S_{\frac{N}{2}+1}, S_{\frac{N}{2}+2}, S_{\frac{N}{2}+3}, ..., S_{N-1}\} \tag{5.10}$$

In questo modo, l'insieme $C1$ rappresenta la prima metà di C, e $C2$ la seconda metà; allo stesso modo, $S1$ rappresenta la prima metà di S, mentre $S2$ costituisce la seconda metà.

Combinando gli insiemi $C1$ e $C2$ con gli insiemi $S1$ ed $S2$, si ottengono quattro matrici che equivalgono a sottoparti della matrice SC:

- Parte 1 - combinando $C1$ con $S1$, si ottiene la prima sottomatrice, ovvero quella in alto a sinistra.
- Parte 2 - combinando $C2$ con $S1$, si ottiene la seconda sottomatrice, ovvero quella in alto a destra.
- Parte 3 - combinando $C1$ con $S2$, si ottiene la terza sottomatrice, ovvero quella in basso a sinistra.
- Parte 4 - combinando $C2$ con $S2$, si ottiene la quarta sottomatrice, ovvero quella in basso a destra.

Di conseguenza, anche la Matrice D sarà suddivisa alla stessa maniera, visto che riporta sulle righe gli autori e sulle colonne i claims, così come la Matrice SC. Per quanto riguarda la matrice FC, che riporta sulle righe le features e sulle colonne i claims, la suddivisione riguarda soltanto le colonne e quindi si ottengono solo due sottoparti, piuttosto che quattro: nel caso della prima e della terza parte viene considerata la sottomatrice FC di sinistra, mentre nella seconda e quarta parte viene considerata quella di destra. La Figura 5.2 mostra il numero di claims veri e falsi nelle sottoparti del Dataset 2 estratte, mentre la Figura 5.3 illustra la ground truth dell'intero Dataset 2.

Come si nota anche nella Tabella 5.5, la Parte-1 e la Parte-3, così come Parte-2 e Parte-4, comprendono lo stesso numero di claims veri e claims falsi e ciò è dovuto al fatto che questi subsets derivano dai due sottoinsiemi dell'insieme C (Equazione 5.5) rappresentati nelle Equazioni 5.7 e 5.8. Inoltre, si osserva che l'insieme $C2$ è fortemente sbilanciato, in quanto comprende soltanto 12 affermazioni vere su 403 totali; infine, anche il set di dati completo risulta non essere bilanciato, poichè è costituito solamente da 168 claims veri su 806 totali.

Valutazione dei modelli sui subsets

OMISSIS

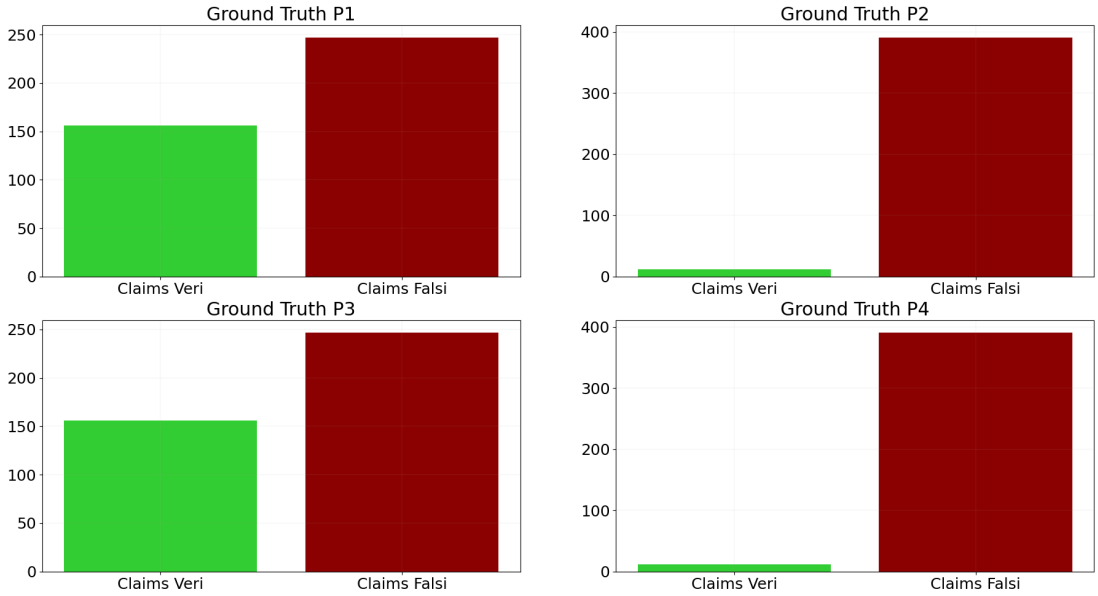

Figura 5.2: Ground Truth delle sottoparti del Dataset 2.

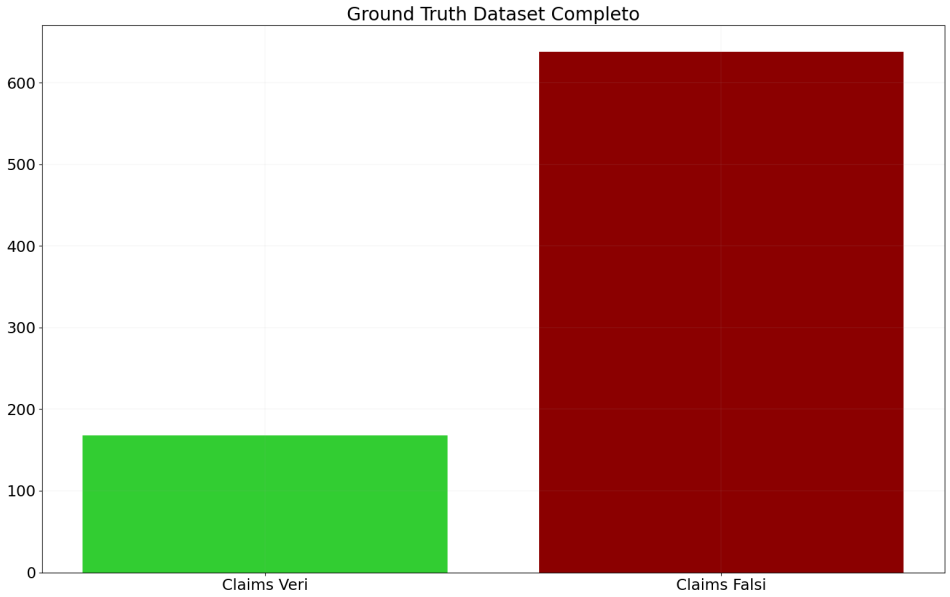

Figura 5.3: Ground Truth Dataset 2.

Valutazione dei modelli sul dataset completo

OMISSIS

5.3 Analisi sul dataset ottenuto dal Sistema

OMISSIS

5.3.1 Valutazione degli approcci

OMISSIS

Capitolo 6

Conclusioni

Il lavoro di tesi proposto riguarda dunque l'analisi delle fake news che circolano sui social network. Infatti, la rapida diffusione delle piattaforme di social media avvenuta negli ultimi anni sta stravolgendo il modo di comunicare e confrontarsi degli utenti. I social network consentono alle persone registrate di essere in contatto 24 ore su 24, abbattendo dunque tutte le barriere costituite dalla distanza fisica che le separa. È possibile pubblicare tutto ciò che si intende condividere con altra gente: opinioni, pensieri, commenti su eventi di qualsiasi genere, idee su fatti politici, religiosi, di cronaca, informazioni e punti di vista sulla situazione sanitaria e tanto altro. Dunque, i social costituiscono di fatto la principale fonte di informazioni, in cui gli utenti riescono ad essere sempre aggiornati sulle ultime vicende. D'altro canto, questa completa ed immediata libertà d'accesso a tutte le informazioni messe a disposizione dagli utenti provoca un aumento dei fenomeni di disinformazione e divulgazione delle fake news.

La piattaforma social su cui si diffondono principalmente le notizie false è Twitter, in quanto costituisce il mezzo più utilizzato per condividere pensieri ed opinioni, soprattutto per quanto riguarda gli avvenimenti politici, economici e sulla situazione sanitaria.

Le fake news, considerate delle minacce alla democrazia, influenzano negativamente la politica, l'economia, i mercati azionari, il giornalismo e diminuiscono la fiducia degli utenti nei confronti delle istituzioni e delle notizie vere. Ecco perchè sono stati sviluppati dei sistemi di rilevamento di fake news con lo scopo di distinguere le informazioni reali da informazioni false. In letteratura esistono svariati metodi che consentono di rilevare le notizie false in circolazione tra cui: metodi tradizionali di apprendimento automatico, metodi di deep learning, metodi di rilevamento basato sulla conoscenza, metodi di rilevamento basato sulla propagazione, metodi di rilevamento basato sulla fonte, che considerano alcune caratteristiche di tipo linguistiche, temporali, basate sulle informazioni degli utenti e basate sulle interazioni.

Il principale problema è costituito dalla qualità dei dati a disposizione. Infatti, il tipo di informazioni raccolte dai dataset dipendono dall'obiettivo dell'applicazione, e possono variare tra un dataset e l'altro, così come le etichette e i tipi di contenuti che vengono forniti. Inoltre, alcuni dataset non comprendono delle etichette assegnate manualmente, tramite verifica giornalistica o umana, bensì assegnate in maniera automatica da determinate applicazioni e dunque le prestazioni degli algoritmi dipendono dalla qualità delle applicazioni in questione.

6.1 Lavoro proposto

OMISSIS

6.2 Risultati e possibili miglioramenti

I risultati ottenuti sia dagli approcci di fake news detection sia dalla progettazione del sistema sono soddisfacenti ma sicuramente migliorabili, soprattutto per quanto riguarda la sperimentazione del sistema di raccolta dei tweets. In entrambi sono presenti dei limiti, alcuni di natura concettuale e altri di natura tecnica che dunque non escludono degli eventuali miglioramenti futuri.

6.2.1 Sistema progettato

OMISSIS

6.2.2 Modelli di fake news detection

Dalle esecuzioni degli algortimi di fake news detection è emerso che le prestazioni migliori sono quelle dell'Expectation Maximization, in termini delle metriche prima citate. In particolare, è stato evidenziato come le performance di quest'ultimo sono molto simili a quelle dell'Expectation Maximization with Multiple Features; considerando che questi due approcci differiscono per la Matrice FC, che è assente nell'EM e presente nell'EM-MultiF, il quale rappresenta la relazione tra le features considerate e i claims, ne consegue che queste caratteristiche utilizzate per la realizzazione della matrice non sono dei buoni indicatori di verità in quanto la loro considerazione non migliora significativamente le prestazioni dell'algoritmo in questione. Anche gli altri due approcci di fake news detection, ovvero PEM e CEM, risultano essere meno performanti rispetto all'EM, in quanto si fondano su delle ipotesi riguardanti proprio le features. Considerando ad esempio il caso del PEM, il concetto che sta alla base di quest'approccio è che un tweet che comprende un'immagine o un url ha più probabilità di essere vero rispetto ad un tweet che è costituito esclusivamente dal testo, quindi senza altre entità. L'idea è che i link o le immagini contenuti nel tweet costituiscono una testimonianza di ciò che si sta affermando, danno supporto e dimostrazione che quello che l'utente sta riportando è vero. Nell'approccio in questione, però, viene utilizzata solo un'informazione binaria, che indica la presenza o l'assenza delle features nel tweet. In base ai risultati ottenuti, questa informazione risulta essere inadeguata o comunque insufficiente, in quanto i claims che comprendono tali features possono essere veri quanto falsi.

Allo stesso modo, anche il CEM si fonda su un'ipotesi risultata sbagliata, secondo cui la probabilità che un claim sia vero aumenta in base al numero di features indipendenti. Anche in questo caso, l'unica informazione necessaria è di tipo binario e riguarda semplicemente la presenza o l'assenza delle features in determinati tweets.

Un eventuale lavoro futuro potrebbe, oltre a considerare le features in questione, effettuare una valutazione e un controllo più accurato sul contenuto delle immagini

e dei link presenti nei tweets, in modo da verificare che tale contenuto sia coerente o di supporto a ciò che si sta pubblicando.

Un altro possibile miglioramento, che riguarda non solo PEM e CEM ma anche EM ed EM-MultiF, è l'inclusione nei modelli di un'informazione temporale. Infatti, tutti i approcci descritti non considerano alcun tipo di variabile temporale; si limitano a valutare un set di claims riportati da un set di autori, senza considerare le date in cui i tweets sono pubblicati. Dunque si potrebbe integrare nei modelli una feature che riguarda questo tipo di informazione e capire come condiziona le prestazioni di questi. Infine, poichè è scaturito che le Matrici D, ottenute dai dati a disposizione, sono costituite interamente da 0, ne deriva che la loro presenza negli approcci presentati è del tutto ininfluente. Un eventuale miglioramento potrebbe consistere nel rimuovere questa dai modelli ed includere invece una matrice che costituisca un'informazione più facile da reperire, come informazioni sugli utenti, informazioni relative al numero di like dei tweets o eventualmente rilassare il modello di propagazione della Matrice D e quindi includere un'informazione meno restringente: ad esempio, piuttosto che settare ad 1 la cella $SiCj$ se un antenato dell'utente Si afferma il claim Cj, si potrebbe settare ad 1 se un antenato ha messo un like al relativo tweet, senza quindi la necessità di riportarlo.

6.2.3 Considerazioni finali

In conclusione, il presente lavoro evidenzia come le fake news hanno comportato un alto grado di interesse da parte dei ricercatori nello sviluppo di modelli e sistemi di rilevamento, con lo scopo di riuscire a distinguere informazioni reali da fake news o rumors e cercare dunque di diminuire le conseguenze provocate dalla divulgazione di queste, soprattutto in ambito politico, economico e finanziario. La continua diffusione di informazioni false e infondate dimostra che le attuali soluzioni e i mezzi di rilevamento non risultano al 100% efficaci e comprendono margini di miglioramento. La speranza è che lo sviluppo tecnologico e i progressi raggiunti, soprattutto in ambito dell'intelligenza artificiale, portino alla progettazione di algoritmi e sistemi sempre più accurati che riescano a limitare questo fenomeno.

Bibliografia

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. J. Econ. Perspect. 31, 2 (2017), 211–36.

[2] F. Concone, F. De Vita, A. Pratap, D. Bruneo, G. Lo Re and S. K. Das, "A Novel Recruitment Policy to Defend against Sybils in Vehicular Crowdsourcing," 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 2021, pp. 105-112, doi: 10.1109/SMARTCOMP52413.2021.00035.

[3] Rapoza, K. (2021, June 30). Can 'fake news' impact the stock market? Forbes. Retrieved December 30, 2021, from https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/?sh=7a089e112fac.

[4] Carlos Carvalho, Nicholas Klagge, and Emanuel Moench. 2011. The persistent effects of a false news shock. J. Empir. Finance 18, 4 (2011), 597–615.

[5] Misako Takayasu, Kazuya Sato, Yukie Sano, Kenta Yamada, Wataru Miura, and Hideki Takayasu. 2015. Rumor diffusion and convergence during the 3.11 earthquake: A Twitter case study. PLoS ONE 10, 4 (2015), e0121443.

[6] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of the 22nd International Conference on World Wide Web. ACM, 729–736.

[7] Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology 2, 2 (1998), 175.

[8] Lawrence E. Boehm. 1994. The validity effect: A search for mediating variables. Personality and Social Psychology Bulletin 20, 3 (1994), 285–293.

[9] YouGov. 2017. C4 study reveals only 4% surveyed can identify true or fake news. Retrieved from http://www.channel4.com/info/press/news/c4-study-reveals-only-4-surveyed-can-identify-true-or-fake-news.

[10] Udo Undeutsch. 1967. Beurteilung der glaubhaftigkeit von aussagen. Handbuch der Psychologie 11 (1967), 26–181.

[11] Marcia K. Johnson and Carol L. Raye. 1981. Reality monitoring. Psychological Review 88, 1 (1981), 67.

[12] Miron Zuckerman, Bella M. DePaulo, and Robert Rosenthal. 1981. Verbal and nonverbal communication of deception. In Advances in Experimental Social Psychology. Vol. 14. Elsevier, 1–59.

[13] Steven A. McCornack, Kelly Morrison, Jihyun Esther Paik, Amy M.Wisner, and Xun Zhu. 2014. Information manipulation theory 2: A propositional theory of deceptive discourse production. Journal of Language and Social Psychology 33, 4 (2014), 348–377.

[14] Sudipta Basu. 1997. The conservatism principle and the asymmetric timeliness of earnings. Journal of Accounting and Economics 24, 1 (1997), 3–37.

[15] F. Concone, A. De Paola, G. Lo Re and M. Morana, "Twitter analysis for real-time malware discovery," 2017 AEIT International Annual Conference, 2017, pp. 1-6, doi: 10.23919/AEIT.2017.8240551.

[16] Péter Balint and Géza Balint. 2009. The Semmelweis-reflex. Orvosi Hetilap 150, 30 (2009), 1430.

[17] Kathleen Hall Jamieson and Joseph N. Cappella. 2008. Echo Chamber: Rush Limbaugh and the Conservative Media Establishment. Oxford University Press.

[18] Colin MacLeod, Andrew Mathews, and Philip Tata. 1986. Attentional bias in emotional disorders.Journal of Abnormal Psychology 95, 1 (1986), 15.

[19] Harvey Leibenstein. 1950. Bandwagon, snob, and Veblen effects in the theory of consumers' demand. Quarterly Journal of Economics 64, 2 (1950), 183–207.

[20] Morton Deutsch and Harold B. Gerard. 1955. A study of normative and informational social influences upon individual judgment.Journal of Abnormal and Social Psychology 51, 3 (1955), 629.

[21] Blake E. Ashforth and Fred Mael. 1989. Social identity theory and the organization. Academy of Management Review 14, 1 (1989), 20–39.

[22] Timur Kuran and Cass R. Sunstein. 1999. Availability cascades and risk regulation. Stanford Law Review 51, 4 (1999), 683–768.

[23] Jonathan L. Freedman and David O. Sears. 1965. Selective exposure. In Advances in Experimental Social Psychology. Vol. 2. Elsevier, 57–97.

[24] Robert J. Fisher. 1993. Social desirability bias and the validity of indirect questioning. Journal of Consumer Research 20, 2 (1993), 303–315.

[25] Emily Pronin, Justin Kruger, Kenneth Savtisky, and Lee Ross. 2001. You don't know me, but I know you: The illusion of asymmetric insight.Journal of Personality and Social Psychology 81, 4 (2001), 639.

[26] F. Concone, G. Lo Re, M. Morana, and C. Ruocco, "Twitter Spam Account Detection by Effective Labeling", 2019 ITASEC Italian Conference on Cybersecurity, 2019.

[27] Andrew Ward, L. Ross, E. Reed, E. Turiel, and T. Brown. 1997. Naive realism in everyday life: Implications for social conflict and misunderstanding. In Values and Knowledge, E. S. Reed, E. Turiel, and T. Brown (Eds.). The Jean Piaget Symposium Series. Lawrence Erlbaum Associates, 103–135.

[28] David Dunning, Dale W. Griffin, James D. Milojkovic, and Lee Ross. 1990. The overconfidence effect in social prediction. Journal of Personality and Social Psychology 58, 4 (1990), 568.

[29] Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In Handbook of the Fundamentals of Financial Decision Making: Part I. World Scientific, 99–127.

[30] Carl I. Hovland, O. J. Harvey, and Muzafer Sherif. 1957. Assimilation and contrast effects in reactions to communication and attitude change.Journal of Abnormal and Social Psychology 55, 2 (1957), 244.

[31] Nico H. Frijda. 1986. The Emotions. Cambridge University Press.

[32] Pérez-Rosas V, Kleinberg B, Lefevre A, et al. Automatic detection of fake news[J]. arXiv preprint arXiv:1708.07104, 2017.

[33] Kwon S, Cha M, Jung K, et al. Prominent features of rumor propagation in online social media[C]//2013 IEEE 13th International Conference on Data Mining. IEEE, 2013: 1103-1108.

[34] Zeng L, Starbird K, Spiro E S. Classifying rumor stance in crisisrelated social media messages[C]//Tenth International AAAI Conference on Web and Social Media. 2016.

[35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 770–778.

[36] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 4700–4708.

[37] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems. 1097–1105.

[38] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. 1989. Backpropagation applied to handwritten zip code recognition. Neural Computation 1, 4 (1989), 541–551.

[39] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. ArXiv:1409.1556.

[40] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1–9.

[41] Kyunghyun Cho, Bart Van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. ArXiv:1406.1078.

[42] Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long short-term memory. Neural Computation 9, 8 (1997), 1735–1780.

[43] Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing 45, 11 (1997), 2673–2681.

[44] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv:1810.04805.

[45] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in Neural Information Processing Systems. 5998–6008.

[46] Ruchansky N, Seo S, Liu Y. Csi: A hybrid deep model for fake news detection[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 797-806.

[47] Liu Y, Wu Y F B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[48] Ma J, Gao W, Wong K F. Rumor detection on twitter with tree-structured recursive neural networks[C]. Association for Computational Linguistics, 2018.

[49] Castillo C, Mendoza M, Poblete B. Information credibility on twitter[C]//Proceedings of the 20th international conference on World wide web. 2011: 675-684.

[50] S. Gaglio, G. Lo Re and M. Morana, "A framework for real-time Twitter data analysis", Computer Communications, Volume 73, Part B, 2016, pp. 236-242, ISSN 0140-3664, doi: 10.1016/j.comcom.2015.09.021.

[51] Kwon S, Cha M, Jung K, et al. Prominent features of rumor propagation in online social media[C]//2013 IEEE 13th International Conference on Data Mining. IEEE, 2013: 1103-1108.

[52] Rada Mihalcea and Carlo Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Association for Computational Linguistics, 309–312.

[53] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Trans. Intell. Syst. Technol. 10, 3, Article 21 (April 2019), 42 pages.

[54] De Sarkar S, Yang F, Mukherjee A. Attending sentences to detect satirical fake news[C]//Proceedings of the 27th International Conference on Computational Linguistics. 2018: 3371-3380.

[55] Qian F, Gong C, Sharma K, et al. Neural User Response Generator: Fake News Detection with Collective User Intelligence[C]//IJCAI. 2018: 3834-3840.

[56] Rashkin H, Choi E, Jang J Y, et al. Truth of varying shades: Analyzing language in fake news and political fact-checking[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 2931-2937.

[57] Popat K, Mukherjee S, Yates A, et al. DeClarE: Debunking fake news and false claims using evidence-aware deep learning[J]. arXiv preprint arXiv:1809.06416, 2018.

[58] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.

[59] Shu K, Wang S, Liu H. Beyond news contents: The role of social context for fake news detection[C]// Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. 2019: 312-320.

[60] Volkova S, Shaffer K, Jang J Y, et al. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2017: 647-653.

[61] Lise Getoor and Ashwin Machanavajjhala. 2012. Entity resolution: Theory, practice & open challenges. Proceedings of the VLDB Endowment 5, 12 (2012), 2018–2019.

[62] Yong Deng. 2015. Generalized evidence theory. Applied Intelligence 43, 3 (2015), 530–543.

[63] Bingyi Kang and Yong Deng. 2019. The maximum Deng entropy. IEEE Access 7, 1 (2019), 120758– 120765.

[64] Gabriella Pasi, Marco Viviani, and Alexandre Carton. 2019.Amulti-criteria decisionmaking approach based on the Choquet integral for assessing the credibility of user-generated content. Information Sciences 503 (2019), 574–588.

[65] Marco Viviani and Gabriella Pasi. 2016. A multi-criteria decision making approach for the assessment of information credibility in social media. In Proceedings of the International Workshop on Fuzzy Logic and Applications. 197–207.

[66] Junting Ye and Steven Skiena. 2019. MediaRank: Computational ranking of online news sources.arXiv:1903.07581.

[67] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence 194 (2013), 28–61.

[68] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: A core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web. ACM, New York, NY, 697–706.

[69] F. Concone, S. Gaglio, G. Lo Re and M. Morana. Smartphone data analysis for human activity recognition. In Conference of the Italian Association for Artificial Intelligence (pp. 58-71). Springer, Cham.

[70] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. ACM, New York, NY, 1247–1250.

[71] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Toward an architecture for never-ending language learning. In Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAA'10), Vol. 5. 3.

[72] Ndapandula Nakashole, Gerhard Weikum, and Fabian Suchanek. 2012. PATTY: A taxonomy of relational patterns with semantic types. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 1135–1145.

[73] Soren Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A nucleus for a web of open data. In The Semantic Web. Springer, 722–735.

[74] Feng Niu, Ce Zhang, Christopher Ré, and Jude Shavlik. 2012. Elementary: Large-scale knowledge- base construction via machine learning and statistical inference. International Journal on Semantic Web and Information Systems 8, 3 (2012), 42–73.

[75] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz,Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge Vault: A web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, 601–610.

[76] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science 359, 6380 (2018), 1146–1151.

[77] Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. ArXiv:2001.06362.

[78] Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on Twitter with tree-structured recursive neural networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 1980–1989.

[79] Reza Zafarani, Xinyi Zhou, Kai Shu, and Huan Liu. 2019. Fake news research: Theories, detection strategies, and open problems. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, 3207–3208.

[80] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16). 2972–2978.

[81] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. 2014. News credibility evaluation on microblog with a hierarchical propagation model. In Proceedings of the IEEE International Conference on Data Mining (ICDM'14). IEEE, Los Alamitos, CA, 230–239.

[82] Manish Gupta, Peixiang Zhao, and Jiawei Han. 2012. Evaluating event credibility on Twitter. In Proceedings of the 2012 SIAM International Conference on Data Mining. 153–164.

[83] F. Concone, G. Lo Re, M. Morana and C. Ruocco, "Assisted Labeling for Spam Account Detection on Twitter," 2019 IEEE International Conference on Smart Computing (SMARTCOMP), 2019, pp. 359-366, doi: 10.1109/SMARTCOMP.2019.00073.

[84] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab.

[85] Kai Shu, Deepak Mahudeswaran, Suhang Wang, and Huan Liu. 2019b. Hierarchical propagation networks for fake news detection: Investigation and exploitation. ArXiv:1903.09196.

[86] F. Concone, G. Lo Re and M. Morana. A fog-based application for human activity recognition using personal smart devices. ACM Transactions on Internet Technology (TOIT), 19(2), 1-20.

[87] Jeppe Norregaard, Benjamin D. Horne, and Sibel Adalı. 2019. NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In Proceedings of the International AAAI Conference on Web and Social Media, Vol. 13. 630–638.

[88] Niraj Sitaula, Chilukuri K. Mohan, Jennifer Grygiel, Xinyi Zhou, and Reza Zafarani. 2020. Credibility-based fake news detection. In Disinformation, Misinformation and Fake News in Social Media: Emerging Research Challenges and Opportunities. Springer.

[89] Kai Shu, Deepak Mahudeswaran, SuhangWang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media.arXiv:1809.01286.

[90] Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM 46, 5 (1999), 604–632.

[91] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. In Proceedings of the 11th International AAAI Conference on Web and Social Media.

[92] F. Concone, G. Lo Re and M. Morana. SMCP: a Secure Mobile Crowdsensing Protocol for fog-based applications. Human-centric Computing and Information Sciences, 10(1), 1-23.

[93] Emilio Ferrara. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. First Monday 22, 8 (2017). Available at http://firstmonday.org/ojs/index.php/fm/article/view/8005/6516.

[94] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. Nature Communications 9, 1 (2018), 4787.

[95] Ma J, Gao W, Wong K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]. Association for Computational Linguistics, 2017.

[96] Kai Shu, Deepak Mahudeswaran, SuhangWang, Dongwon Lee, andHuan Liu. 2018. FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286 (2018).

[97] Twitter. (n.d.). Twitter API documentation | docs | twitter developer platform. Twitter. Retrieved December 17, 2021, from https://developer.twitter.com/en/docs/twitter-api.

[98] Tweepy. (n.d.). Retrieved December 17, 2021, from https://www.tweepy.org.

[99] Learn. scikit. (n.d.). Retrieved February 3, 2022, from https://scikit-learn.org/stable/

[100] Shrivarsheni. (2021, December 19). Text summarization approaches for NLP - practical guide with generative examples. Machine Learning Plus. Retrieved January 31, 2022, from https://www.machinelearningplus.com/nlp/text-summarization-approaches-nlp-example/

[101] S. Gaglio, G. Lo Re and M. Morana. "Real-time detection of Twitter social events from the user's perspective." 2015 IEEE International Conference on Communications (ICC). IEEE, 2015.

[102] Natural language processing (NLP): Come funziona L'elaborazione del Linguaggio Naturale. lineup. (n.d.). Retrieved February 16, 2022, from https://blog.osservatori.net/it_it/natural-language-processing-nlp-come-funziona-lelaborazione-del-linguaggio-na

[103] Real Python. (2021, March 19). Natural language processing with spacy in python. Real Python. Retrieved February 8, 2022, from https://realpython.com/natural-language-processing-spacy-python/.

[104] Stanford typed Dependencies Manual. (n.d.). Retrieved February 8, 2022, from https://downloads.cs.stanford.edu/nlp/software/dependencies_manual.pdf

[105] Google. (n.d.). Fact check tools. Google. Retrieved February 8, 2022, from https://toolbox.google.com/factcheck/explorer.

[106] Using humans as sensors: An estimation-theoretic perspective. IEEE Xplore. (n.d.). Retrieved December 30, 2021, from https://ieeexplore.ieee.org/document/6846739.

[107] Dong Wang University of Illinois at Urbana Champaign, Wang, D., University of Illinois at Urbana Champaign, Lance Kaplan US Army Research Labs, Kaplan, L., Labs, U. S. A. R., Hieu Le University of Illinois at Urbana Champaign, Le, H., Tarek Abdelzaher University of Illinois at Urbana Champaign, Abdelzaher, T., Asia, M. R., University, J. H., Virginia, U. of, & Metrics, O. M. V. A. (2012, April 1). On truth discovery in social sensing: A maximum likelihood estimation approach. On truth discovery in social sensing | Proceedings of the 11th international conference on Information Processing in Sensor Networks. Retrieved December 30, 2021, from https://dl.acm.org/doi/10.1145/2185677.2185737.

[108] H. Shao et al ., "Truth Discovery With Multi-Modal Data in Social Sensing", in IEEE Transactions on Computers , vol. 70, nr. 9, pp. 1325-1337, 1 settembre 2021, doi: 10.1109/TC.2020.3008561.

[109] Truth Discovery With Multi-Modal Data in Social Sensing, IEEE Xplore, December 03, 2021, https://ieeexplore.ieee.org/document/9138694/authors#authors.

[110] H. Shao, S. Yao, Y. Zhao, C. Zhang, J. Han, L. Kaplan, L. Su, and T. Abdelzaher, "A constrained maximum likelihood estimator for unguided social sensing," in IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018, pp. 2429–2437.

[111] Shu, Kai and Mahudeswaran, Deepak and Wang, Suhang and Lee, Dongwon and Liu, Huan. (2018). FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media.

[112] F. Concone, R. Giaconia, G. Lo Re and M. Morana. A Smart Assistant for Visual Recognition of Painted Scenes. In 2021 Joint ACM Conference on Intelligent User Interfaces Workshops, ACMIUI-WS 2021 (Vol. 2903). CEUR-WS.

[113] Shu, Kai and Sliva, Amy and Wang, Suhang and Tang, Jiliang and Liu, Huan. (2017). Fake News Detection on Social Media: A Data Mining Perspective.

[114] Shu, Kai and Wang, Suhang and Liu, Huan. (2017). Exploiting Tri-Relationship for Fake News Detection.