UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI INFORMATICA

TESI DI LAUREA

IN

INFORMATICA

METODI E ALGORITMI PER IL RICONOSCIMENTO DI FAKE NEWS

Relatore:

Prof.ssa Dajana Conte

Candidata:

Coccaro Benedetta

Matr. 05121/05357

Correlatore:

Prof. Francesco Colace

ANNO ACCADEMICO 2019/2020

ABSTRACT

L'utilizzo delle fake news è, attualmente, uno tra gli strumenti più efficaci di manipolazione e plagio dell'opinione pubblica.

Le notizie false, fin dai tempi più antichi, sono sempre state elemento caratterizzante di ogni società, ma, al giorno d'oggi, con la nascita di nuovi e potenti mezzi di comunicazione come Internet e i social network, le fake news sono diventate ancora più deleterie e pericolose. Esse, infatti, sono in grado di raggiungere e raggirare un numero sempre più elevato di utenti, per via della grande rapidità di diffusione su ampia scala che caratterizza questi nuovi mezzi di divulgazione.

Sempre più persone, ormai, fanno di Internet la loro unica fonte nella raccolta di conoscenze rischiando così ogni volta di incappare nella rete delle notizie false.

In questa tesi, verranno descritte e analizzate le principali soluzioni fino ad ora adoperate nell'ambito della lotta contro le fake news. In particolar modo verranno trattate metodologie Knowledge-based, Style-based, Network-based e Source-based (capitolo 1). Verrà descritto l'utilizzo e l'importanza delle reti bayesiane (capitolo 2), verrà mostrato e analizzato un metodo non supervisionato per l'individuazione delle fake news sui social (capitolo 3) ed infine verrà presentato un primo esempio di approccio per l'individuazione di fake news basato su reti bayesiane (capitolo 4).

Sommario

1. *Descrizione dei metodi esistenti per l'individuazione delle fake news sui social network*

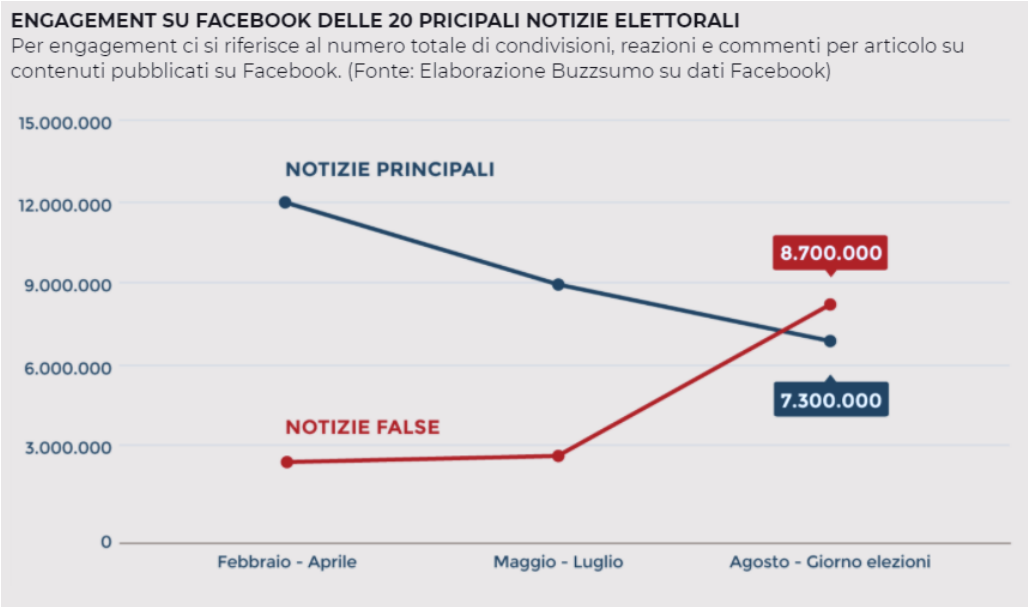
1.1. Introduzione

Fenomeno esistente fin dalla nascita delle prime civiltà, la manipolazione della percezione della realtà è uno degli strumenti più pericolosi ed efficaci in grado di condizionare la concezione di un essere umano.

Negli ultimi anni la società è stata soggetta ad un grande cambiamento consistente nella migrazione della divulgazione delle informazioni da vecchi mezzi di comunicazione di massa, come ad esempio le radio, i televisori, i giornali, ad un grande mezzo di comunicazione per le masse, ovvero Internet e in particolar modo i social network.

Questi ultimi, a differenza dei primi, sono privi di un mediatore culturale, ciò significa che chiunque può pubblicare una notizia e la può diffondere. Abbiamo tra le nostre mani un grande strumento di democrazia che purtroppo sta venendo contaminato dalla divulgazione massiva di fake news. A testimonianza di ciò, si riporta l'esempio delle elezioni presidenziali statunitensi del 2016, dove prevalse la diffusione delle notizie false su quelle vere.

Fig.1 [Il Sole 24 Ore]



1.2. Tecnologie utilizzate

La distorsione della realtà, tramite la diffusione delle fake news, è uno dei problemi che caratterizzano di più la nostra epoca. Molti sono gli studi che sono stati compiuti incentrati sullo sviluppo di nuove tecnologie e metodologie per la loro individuazione e denuncia.

Le tecniche fino ad ora implementate e maggiormente adoperate possono essere classificate in 4 categorie [1]:

- Knowledge-based methods
- Style-based methods
- Network-based detection [5]
- Source-based methods

Di seguito è riportata una breve descrizione di ogni metodologia, a cui verrà associata l'analisi di ogni sua caratteristica più importante.

1.2.1. Knowledge-based methods

Quando parliamo di individuazione di fake news tramite la tecnica del knowledge-based, uno degli strumenti più adoperati è il fact-checking.

Inizialmente sviluppato e utilizzato in ambito giornalistico, la logica adoperata nel fact-checking mira a valutare l'autenticità della notizia confrontando le informazioni presenti al suo interno con contenuti di altre notizie già verificate. Questa tecnica, più comunemente conosciuta come manual fact-checking, può essere generalizzata in expert-based fact-checking e crowd-sourced fact-checking [1].

La prima metodologia affida l'analisi delle notizie ad un piccolo gruppo di esperti nella tecnica del fact-checking i quali, grazie alle loro conoscenze, ne verificano la veridicità.

Adoperare l'expert-based fact-checking presenta numerosi vantaggi come, ad esempio, la facile gestione del gruppo di fact-checkers data la dimensione ridotta e la grande accuratezza dei risultati forniti. Essa presenta però una costosa e scarsa scalabilità all'aumentare della quantità di informazioni presenti nella notizia da dover verificare.

La tecnica del crowd-sourced fact-checking affida le sue ricerche ad un ampio gruppo di fact-checkers i quali, a differenza dell'expert-based fact-checking, vengono scelti tra normali gruppi di utenti. Questa metodologia però, rispetto alla precedente, è difficilmente gestibile data la quantità ingente di fact-checkers

adoperati ed inoltre i risultati ottenuti sono caratterizzati da scarsa credibilità e accuratezza a causa, ad esempio, della propensione politica e dei pareri contrastanti degli utenti.

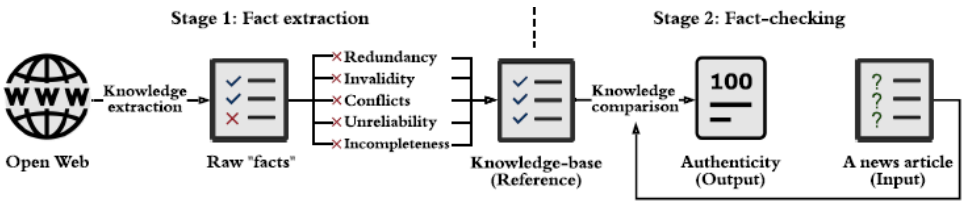
La tecnica del manual fact-checking, nonostante sia funzionale, presenta delle limitazioni. Al giorno d'oggi i social network ed Internet sono il mezzo più utilizzato dalla popolazione per la diffusione e la raccolta di informazioni e sono anche il luogo in cui avviene senza sosta il passaggio di un ingente flusso di notizie. Il manual fact-checking non è in grado di fronteggiare questo massiccio traffico di informazioni, dunque si è ricorso alla rapidità e alla scalabilità degli algoritmi, sviluppando così l'automatic fact-checking.

Esso analizza una determinata informazione passatagli secondo una specifica rappresentazione che consiste nella riorganizzazione delle proposizioni secondo la tripla (Soggetto, Predicato, Complemento) che ben ne rappresenta il significato.

Consideriamo la notizia "Joe Biden è il presidente degli Stati Uniti", essa può essere ristrutturata in (JoeBiden, Professione, Presidente).

La metodologia dell'automatic fact-checking viene organizzata in due specifiche fasi come possiamo vedere in Fig.2.

Fig.2



La fase del fact extraction si dedica all'assunzione e all'elaborazione delle notizie, organizzando le informazioni contenute in esse secondo la tripla precedentemente descritta; essa verrà presa in input dalla fase successiva di fact-checking.

Consideriamo un insieme di triple (s_i, p_i, o_i), con i = 1, 2, …, n le quali rappresentano rispettivamente il soggetto, il predicato e l'oggetto di ciascun predicato presente all'interno delle informazioni contenute nella news.

Sia G_{KB} un grafo contenente triple di informazioni verificate denotate (s_j, p_j, o_j), con j = 1, 2, …, m; per KB si intente "Knowledge Base", ovvero un set di triple verificate.

La procedura di fact-checking consiste nell'identificare una funzione f che assegna un valore di autenticità $A_i \in [0, 1]$, con $A_i = 1$ se la tripla è vera e $A_i = 0$ se è falsa, confrontando le informazioni di cui non si conosce il valore di veridicità, con ogni tripla del grafo G_{KB} che sappiamo essere informazioni sicuramente vere.

Al termine di questa fase, l'attendibilità della notizia analizzata è ottenuta analizzando tutti i valori A_i acquisiti.

$$\text{f} : (s_i, p_i, o_i) \xrightarrow{G_{KB}} A_i \in [0, 1],$$

$$\text{A} = \text{I}\,(A_1, A_2, \dots, A_n),$$

dove I è la funzione di aggregazione [1], che permette di raggruppare i valori di validità ottenuti, riuscendo così ad ottenere il livello di veridicità complessivo della notizia.

1.2.2. Style-based methods

È una metodologia che si rispecchia per molti aspetti nella logica knowledge-based, anch'essa incentrata sull'analisi approfondita del contenuto delle notizie.

Essa, a differenza della tecnica precedentemente descritta, non analizza l'autenticità delle informazioni ma bensì ne valuta le "intenzioni". Ciò avviene studiando in maniera approfondita il linguaggio e lo stile lessicale utilizzato nella realizzazione di fake news, la cui logica di creazione è improntata appositamente nell'attirare utenti, come ad esempio col fenomeno del clickbaiting, e a influenzarli negativamente.

Molti sono gli algoritmi fino ad ora realizzati con tecnologie di machine learning che usufruiscono della logica del style-based [1]. Primo passo da dover compiere è l'elaborazione dello stile lessicale che caratterizza la notizia secondo un set di features in grado di rappresentare e tradurne molti aspetti, come ad esempio la sintassi, la semantica e il lessico utilizzato.

Consideriamo un set di k features, il quale verrà rappresentato sotto forma di vettore f $\in R^k$. Per verificare la notizia, si individua una funzione S, tale che

$$S : f \overset{TD}{\rightarrow} \hat{y} \in [0,1]$$

dove $\hat{y}$ è il risultato della funzione che con 0 indica la notizia vera e con 1 la notizia falsa e TD = $\{(f_l, y_l) \mid f_l \in R^k,\ y_l \in \{0,1\}, l = 1 \dots n\}$ è il training dataset. Il training dataset consiste in un set di n informazioni rappresentate con il medesimo set di features f_l con le y_l già note.

Grande utilizzo della logica style-based viene fatto anche nell'ambito dell'intelligenza artificiale [2], ne è un esempio la metodologia sviluppata da Ozabary nell'articolo [3].

Primo passo da compiere è la cancellazione di numeri e segni di punteggiatura all'interno del testo analizzato. Ne segue la rimozione di articoli, congiunzioni e parole formate da un numero di caratteri inferiori ad un determinato valore N.

Il passo successivo consiste nell'associare ad ogni parola presente nel testo un valore che ne rappresenta l'importanza. Esso viene ottenuto, utilizzando il metodo Term Frequency (TF), dal rapporto tra il numero di volte che si presenta la parola e la somma di tutte le parole del testo. Al termine di ciò i documenti verranno convertiti in vettori di valori di importanza dai quali verrà elaborata una matrice m x n dove le righe rappresentano i documenti e le colonne le parole al loro interno.

Sempre nell'ambito dell'intelligenza artificiale, troviamo un'altra metodologia che si basa sempre sulla medesima logica style-based. Essa consiste nell'utilizzo di n-grammi per l'individuazione di fake news [4].

Un n-gramma è una sequenza di n elementi, come ad esempio parole, numeri, caratteri, fonemi, la cui lunghezza può variare. Un n-gramma di dimensione 1 viene chiamato unigramma, un n-gramma di dimensione 2 viene chiamato bigramma, un n-gramma di dimensione 3 viene chiamato trigramma e così andando avanti.

Questo metodo è caratterizzato da tre fasi. Nella prima fase avviene l'eliminazione dei segni di punteggiatura e di tutte le parole, come ad esempio articoli e congiunzioni, di scarsa importanza. Successivamente, dopo aver ridotto così facendo la

dimensione del testo, inizia la seconda fase nella quale le features per la percezione del contenuto vengono individuate. Nell'ultima fase, in seguito alla loro individuazione, viene applicata la tecnica TF precedentemente descritta, in modo tale da calcolare l'importanza di ogni feature.

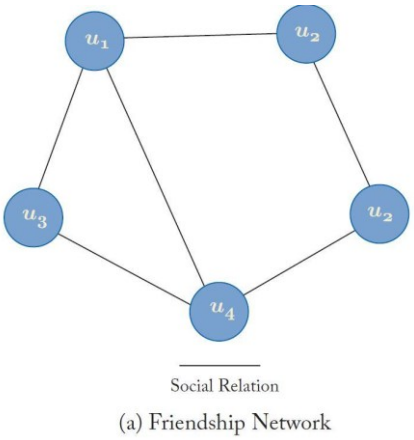
1.2.3. Network-based detection

Sono molte le reti che vengono comunemente utilizzate per facilitare l'individuazione delle fake news. Tra le più utilizzate troviamo la friendship network, la diffusion network e la propagation network [5].

La friendship network, Fig.3, viene rappresentata come un grafo $G_F = (U, E_F)$, dove U ed E_F sono rispettivamente l'insieme dei nodi e l'insieme degli archi del grafo. I nodi rappresentano gli utenti e gli archi, se esistono, vanno a rappresentare l'avvenuto scambio di informazioni tra quest'ultimi.

Come accade nella vita reale, anche sui social l'utente tende a condividere informazioni con persone con le quali ha interessi e concezioni comuni, perciò lo studio della rete che si viene a formare permette di acquisire importanti informazioni sulle community. L'analisi delle friendship networks è di grande importanza perché, oltre che a far comprendere il tipo di relazioni che intercorrono tra gli utenti, esse sono tra le reti attraverso cui passano il maggior numero di informazioni, quindi anche fake news [5].

Fig.3



(a) Friendship Network

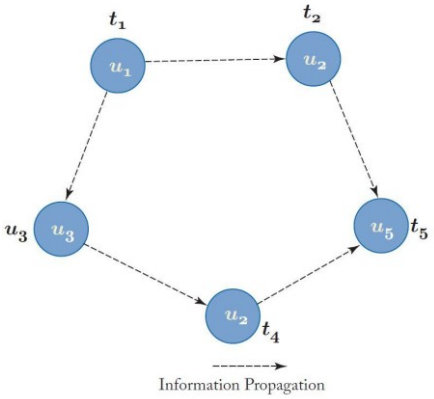
La diffusion network, Fig.4, viene rappresentata come un grafo diretto $G_F = (U, E_F, p, t)$, dove U ed E_F sono rispettivamente l'insieme dei nodi e l'insieme degli archi del grafo. Il nodo $u \in U$ rappresenta un utente che può pubblicare, ricevere e condividere informazioni al tempo $t_i \in t$. L'arco $(u_1 \rightarrow u_2) \in E_F$ tra i nodi $u_1, u_2 \in U$, rappresenta il verso di propagazione dell'informazione al quale è associata una probabilità $p = (u_1 \rightarrow u_2) \in [0,1]$.

La diffusione di contenuti all'interno di questa rete viene rappresentata attraverso una tupla $x_i = (\eta, \Delta t, u_i, c_i)$, dove η rappresenta il numero di utenti attraverso il quale l'informazione c_i è stata diffusa in un determinato tempo t e Δt rappresenta l'intervallo di tempo che intercorre tra la diffusione della notizia c_i da parte dell'utente u_i e la propagazione della medesima notizia da parte dell'utente u_{i+1}.

E' proprio attraverso lo studio dei modelli temporali ottenuti che, grazie all'utilizzo delle reti neurali come ad esempio la RNN (rete neurale ricorrente), avviene l'individuazione delle fake news sulla rete.

Primo passo da dover compiere, è l'estrazione delle features da c_i attraverso, ad esempio, la logica dell'n-gramma precedentemente illustrata. Successivamente, siccome il vettore di features così ottenuto non può essere preso in input dalla RNN, viene sottoposto ad un processo di standardizzazione. Al termine di ciò la rete neurale potrà elaborare le informazioni passategli, scandendo ed analizzando gli intervalli che intercorrono [6].

Fig.4

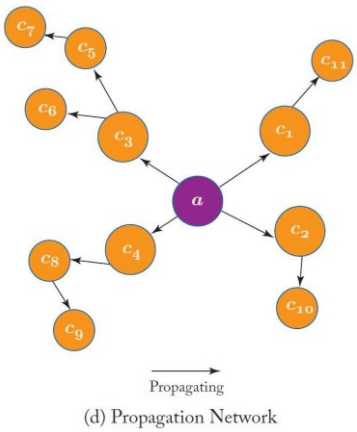


(b) Diffusion Network

La propagation network $G_P = (C, a)$, Fig.5, è formata da frammenti di notizie a e dai corrispondenti post C attraverso i quali è avvenuta la loro propagazione. Essa è strutturata in due livelli, il macro-level e il micro-level. Nel macro-level sono presenti i nodi che rappresentano le news, i tweet e i retweet, mentre nel micro-level troviamo alberi che ne rappresentano la diffusione tramite nodi di condivisione.

La propagation network è una rete che fornisce numerose informazioni sia dal punto di vista temporale, linguistico che strutturale, ed è proprio grazie a ciò che essa viene considerata tra le reti più utili per l'individuazione di fake news. La tecnologia che più ne sfrutta le potenzialità è il geometric deep learning [7, 8].

Fig.5



(d) Propagation Network

1.2.4. Source-based methods

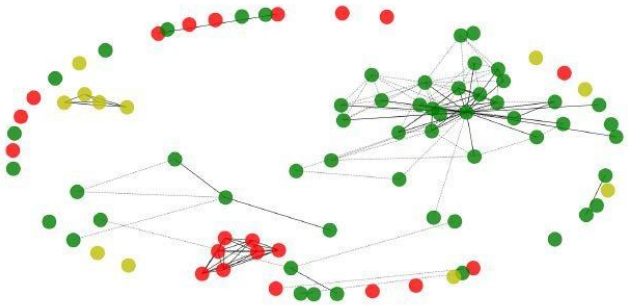
La logica source-based valuta la veridicità di una notizia partendo dall'analisi della fonte [1]. Coloro che creano e diffondono consapevolmente le fake news, a differenza di come si possa pensare, non raggiungono alti livelli di numerosità. Il più delle volte, infatti, le fake news sono frutto del medesimo utente.

Questa metodologia si basa sullo studio di un grafo $G = (U, E)$, i cui nodi rappresentano gli utenti che pubblicano e condividono informazioni e gli archi le relazioni che intercorrono tra di loro. E' così possibile effettuare una valutazione della reputazione di un utente, semplicemente analizzando i nodi con i quali è in relazione.

Come possiamo vedere in Fig.6, i nodi rossi rappresentano gli utenti che hanno diffuso fake news e, il più delle volte, anche i nodi con cui sono in relazione sono del medesimo colore. Abbiamo anche nodi gialli, questi rappresentano gli utenti che hanno condiviso per lo più notizie vere ma che delle volte, anche

inconsapevolmente, sono stati causa della diffusione di notizie false. Abbiamo infine i nodi verdi, i quali rappresentano gli utenti che non hanno mai contribuito alla divulgazione di fake news [1].

Fig.6



2. Le Reti Bayesiane

Prendere decisioni è una delle azioni più ricorrenti che caratterizzano le nostre giornate, dal privato, con scelte personali, fino al pubblico, come ad esempio in ambito lavorativo.

Fattore fondamentale e imprescindibile in ogni decisione è la gestione delle incertezze, condizionate dalle informazioni di cui disponiamo e soprattutto dalle relazioni, sia probabilistiche che deterministiche, che intercorrono tra di loro.

Le reti bayesiane rappresentano uno degli strumenti logici più potenti in grado di gestire problemi decisionali anche in presenza di un alto numero di variabili. Esse fanno affidamento su uno dei teoremi più importanti in campo probabilistico, il teorema di Bayes.

2.1. Teoria della probabilità - cenni

Prima di descrivere il funzionamento delle reti bayesiane, è importante fare dei richiami ad alcuni concetti della teoria della probabilità.

Definizione di probabilità 2.1. La probabilità è una funzione di insieme che associa ad ogni evento A appartenente allo spazio campionario Ω, un numero reale nell'intervallo [0, 1], dove 0 indica la probabilità dell'evento impossibile ovvero che non accadrà mai ed 1, al contrario, la probabilità dell'evento certo.

Essa viene calcolata dal rapporto tra il numero dei casi favorevoli all'evento e il numero di casi possibili purché essi siano tutti ugualmente possibili.

$$P(A) = \frac{n.casi\ favorevoli}{n.casi\ possibili}$$

Definizione di probabilità condizionata 2.2. Siano A e B due eventi tali che $P(B) \neq 0$. La probabilità condizionata di A dato B, indicata come $P(A|B)$, è data da

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

Essa permette di calcolare la probabilità che si verifichi l'evento A nell'ipotesi in cui si sia verificato l'evento B.

La probabilità condizionata rappresenta uno dei tasselli più importanti nello studio delle reti bayesiane.

Introduciamo ora il concetto di dipendenza, anch'esso fondamentale nello studio di quest'ultime, perché indispensabile nella definizione delle relazioni tra le variabili del nostro problema decisionale.

Definizione di dipendenza 2.3. Due eventi A e B si dicono dipendenti se il verificarsi dell'uno influisce sul calcolo della probabilità dell'altro.

Definizione di indipendenza 2.4. Viceversa, due eventi A e B si dicono indipendenti se il verificarsi dell'uno non influisce sul calcolo della probabilità del verificarsi dell'altro. Quando due eventi sono indipendenti, la loro intersezione ha probabilità $P(A \cap B) = P(A) \times P(B)$. Per questo motivo, considerando la formula di probabilità condizionata precedentemente descritta (1), la probabilità condizionata di due eventi indipendenti sarà, in conseguenza alla semplificazione di P(B),

$$P(A|B) = P(A)$$

L'indipendenza è simbolicamente indicata con A ⊥ B ed è simmetrica, quindi, A ⊥ B ≡ B ⊥ A.

Il teorema di Bayes, che come abbiamo già detto è fondamentale per lo studio delle relazioni all'interno di una rete bayesiana, è lo strumento che ci permette di calcolare le probabilità condizionate.

Teorema di Bayes 2.1. Dato un insieme di k eventi $B_1, B_2, \dots, B_k$ a due a due incompatibili ed un evento A, si ha che

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A)} = \frac{P(A \mid B_i)P(B_i)}{\sum_{i=1}^{k} P(B_i)P(A \mid B_i)}$$

Le probabilità $P(B_i)$ dei singoli eventi B_i per i = 1, 2, …, k sono dette probabilità a priori;

le probabilità condizionate $P(A \mid B_i)$ per i = 1, 2, …, k sono chiamate verosimiglianze;

le probabilità condizionate $P(B_i \mid A)$ per i = 1, 2, …, k che si riferiscono agli eventi B_i dopo che si è verificato l'evento A vengono chiamate probabilità a posteriori.

2.2. *Introduzione alle reti Bayesiane*

Una rete Bayesiana è un modello grafico probabilistico utilizzato per descrivere e studiare problemi decisionali.

Definizione rete Bayesiana 2.5.[9] Si considerino n variabili casuali $X_1, X_2, \dots, X_n$, un grafo diretto aciclico con n nodi numerati e si supponga che il nodo i del grafo sia associato alla variabile X_i. Il grafo è una rete Bayesiana che rappresenta le variabili $X_1, X_2, \dots, X_n$ se:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i \middle| Pa(X_i)\right)$$

dove $Pa(X_i)$ denota l'insieme dei genitori del nodo X_i, ovvero tutte le variabili per le quali esiste un arco diretto da ciascuna di esse al nodo X_i.

Le reti Bayesiane vengono rappresentate con la notazione $BN = (G, P)$ dove G rappresenta la struttura della rete e P la distribuzione di probabilità congiunta.

Definizione di probabilità congiunta 2.6. Per una rete Bayesiana definita sull'insieme delle variabili $X = (X_1, X_2, \dots, X_n)$, la distribuzione di probabilità congiunta della rete è definita come il prodotto delle probabilità condizionate e marginali di tutti i nodi [10].

$$P(X) = P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i \middle| Pa(X_i)\right)$$

La costruzione della rete Bayesiana, come un grafo diretto aciclico, permette di garantire il rispetto di alcuni vincoli:

- la fattorizzazione della probabilità congiunta come prodotto di probabilità condizionate
- l'esistenza di almeno una struttura aciclica adatta a rappresentare l'oggetto, qualunque sia il numero e la natura delle dipendenze tra le variabili [9]
- l'impossibilità che un nodo possa essere il suo stesso ascendente o discendente

Con l'utilizzo delle reti Bayesiane, le relazioni che intercorrono tra le variabili del problema vengono quantificate specificando, per ogni nodo, una distribuzione di probabilità condizionata.

Queste vengono rappresentate attraverso delle tabelle di probabilità condizionata (CPT).

Definizione di CPT 2.7. Per ogni variabile X_i, con n nodi genitori $(Y_1, Y_2, \dots, Y_n)$, la CPT è indicata con $P(X_i|Y_1, Y_2, \dots, Y_n)$ e contiene la probabilità associata ad ogni possibile combinazione tra gli stati di X_i e di tutti i suoi genitori [10].

Nel caso dei nodi radice, non avendo genitori, la CPT contiene la probabilità a priori associata ad ogni stato del nodo considerato.

Nella realizzazione di una rete Beyesiana, le relazioni che intercorrono tra le variabili del problema, possono essere rappresentate attraverso 3 diverse tipologie di connesioni: connessioni seriali, connessioni divergenti e connessioni convergenti [25].

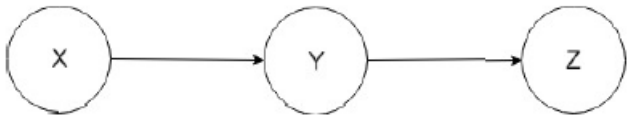
In fig. 7 è rappresentato un esempio di connessione seriale.

La rete è formata da tre nodi X, Y e Z; Z è dipendente da Y che a sua volta è dipendente da X.

$$P(Z|X \wedge \mathrm{Y}) = P(Z|Y) \equiv \mathrm{X} \perp \mathrm{Z}|\mathrm{Y}$$

Quest'uguaglianza indica che la probabilità di Z, sapendo che X e Y si sono verificati, è uguale a quella che si otterrebbe sapendo che solo Y si è verificato. La variabile X non ha alcuna influenza sulla probabilità di Z.

Fig.7

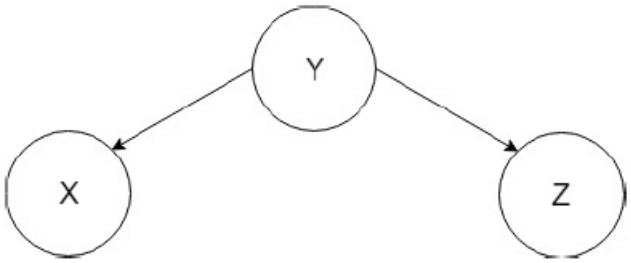


In fig. 8 è mostrato un esempio di connessione divergente.

La rete è formata da tre nodi X, Y e Z; X e Z sono rispettivamente figli del nodo Y con il quale instaurano una relazione di dipendenza. La probabilità dei singoli stati di un nodo figlio non è influenzata in nessun modo dalle probabilità riscontrate negli altri figli.

$$P(Z|X \wedge \mathrm{Y}) = P(Z|Y) \equiv \mathrm{X} \perp \mathrm{Z}|\mathrm{Y}$$

Fig.8

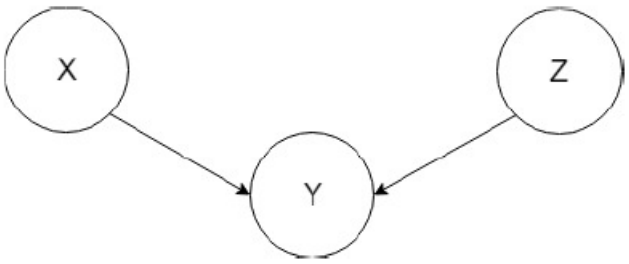


In fig. 9 è rappresentato un esempio di connessione convergente.

La rete è formata da tre nodi X, Y e Z; X e Z sono rispettivamente genitori del nodo Y, dunque, la probabilità di quest'ultimo verrà condizionata da entrambi i nodi X e Z.

Le probabilità dei nodi X e Z sono tra loro indipendenti.

Fig.9

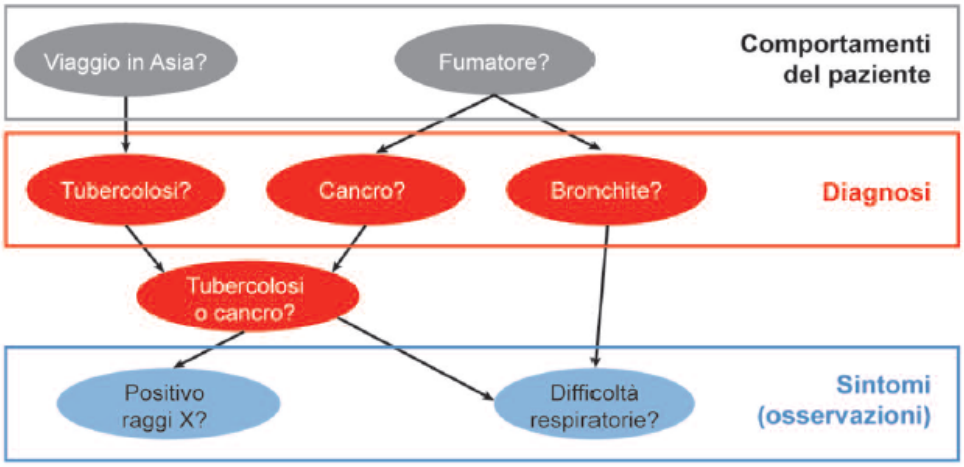


2.3. Esempio di rete Bayesiana

Uno dei primi esempi accademici di rete bayesiana, è la rete Asia rappresentata in fig. 10 [11].

Si tratta della versione semplificata di una rete per la diagnosi di malattie ai polmoni, in particolare tubercolosi, cancro e bronchite. Ogni nodo della rete corrisponde a un comportamento del paziente o al risultato di un esame medico, mentre la direzione degli archi descrive le relazioni tra le variabili.

Fig.10



Ad esempio, essere fumatore aumenta la probabilità di avere un tumore ai polmoni o di soffrire di bronchite, mentre non ha alcuna relazione con la tubercolosi.

Nel momento in cui vengono inseriti le informazioni relative ad un paziente, queste, propagandosi all'interno della rete, aggiornano in maniera quantitativa le probabilità associate ai nodi in essa presenti.

La rete Asia può essere facilmente ampliata includendo risultati di ulteriori test utili alla diagnosi, semplicemente aggiungendo nuovi nodi alla struttura già esistente, come d'altronde accade per ogni rete Bayesiana.

2.4. Vantaggi nell'utilizzo di reti Bayesiane

Le reti bayesiane sono strumenti decisionali estremamente flessibili nella gestione di problemi di incertezza e sono numerosi i vantaggi che offrono [13].

Esse permettono di integrare informazioni provenienti da fonti diverse, dall'utilizzo di dati sperimentali, allo studio di conoscenze teoriche e di poterle inserire in qualsiasi momento all'interno della rete. Infatti, essa è in grado di effettuare un aggiornamento globale delle probabilità, nel momento in cui si acquisiscono nuove conoscenze sul sistema e sul processo che si sta rappresentando.

Altro vantaggio nell'utilizzo delle reti bayesiane si riscontra dal punto di vista computazionale. Esse sono in grado di raggiungere una maggior efficienza in questo ambito poiché richiedono di inserire le probabilità delle sole variabili legate tra loro da un arco.

La facile lettura e comprensione di queste reti è un altro aspetto da non sottovalutare. Esse danno la possibilità di raggiungere facilmente un'accurata conoscenza delle variabili del problema e delle loro relazioni e, di conseguenza, anche dei dati ottenuti in output.

Un grande limite che per molto tempo ha caratterizzato le reti bayesiane, era l'elevato carico computazionale che era richiesto in alcune elaborazioni. Oggi tale ostacolo è stato in buona parte superato grazie allo sviluppo di appositi software di calcolo [12] e di algoritmi di elaborazione sempre più efficienti.

3. Metodo non supervisionato per l'individuazione delle fake news sui social

La maggior parte dei metodi fino ad ora utilizzati per l'individuazione delle fake news, sono caratterizzati da un sistema di apprendimento supervisionato. Ciò vuol dire che i dataset interrogati devono fornire, oltre le features che caratterizzano gli item, anche un valore di verità che classifichi gli elementi. Questo valore è fondamentale per allenare l'algoritmo e ottenere così in futuro output corretti [14].

L'utilizzo di questa tipologia di dataset, presenta però una limitazione. E' necessario infatti un grande dispendio di tempo e manodopera per la loro realizzazione, necessari per la corretta individuazione dei valori di verità che verranno poi forniti.

Nell'esempio di seguito proposto, la logica che viene utilizzata è differente. Infatti, questa nuova metodologia non supervisionata, non baserà il calcolo degli output su valori di verità già conosciuti.

Vengono analizzati e sfruttati due aspetti fondamentali per l'individuazione di fake news.

Il primo consiste nell'analizzare il coinvolgimento e l'interesse mostrato dall'utente nei confronti di una notizia, mentre il secondo aspetto, fondamentale per questo tipo di approccio, è lo studio della reputazione e quindi della credibilità degli utenti che vengono considerati [14].

Le tecnologie che vengono adoperate sono le seguenti:

- reti bayesiane, per la cattura delle opinioni degli utenti e lo studio della loro credibilità come divulgatori
- campionamento di Gibbs, per dedurre la verità delle notizie tramite il calcolo della probabilità condizionata

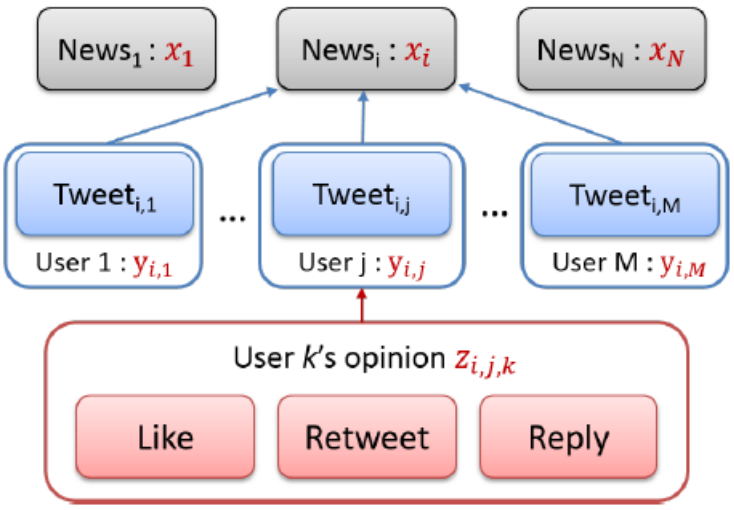
3.1. Analisi della metodologia

In Fig.11 è rappresentata graficamente la logica alla base dello sviluppo di questo metodo.

È necessario prima di tutto fare una precisazione.

Gli utenti vengono divisi in due diverse tipologie, gli utenti verificati e gli utenti non verificati. Gli utenti verificati, avendo maggiore popolarità sui social, avranno, da parte degli altri utenti, una maggiore attenzione e di conseguenza anche ciò che pubblicano e condividono [14].

Fig.11



In fig.11 è rappresentato ciò che abbiamo appena detto; le notizie vengono divulgate sui social da utenti verificati con le quali interagiscono gli utenti non verificati.

Vengono considerati in partenza solamente i tweet degli utenti verificati per due ragioni. In primo luogo, per semplificare il modello analizzato, infatti l'analisi di tweet di utenti non verificati potrebbe indurre molta confusione e imprecisione nei dati da analizzare senza fornirci informazioni utili per l'individuazione di fake news. Altra ragione riguarda il presupposto che gli utenti verificati, per via della loro popolarità ed influenza, divulghino informazioni con un livello di veridicità maggiore.

3.2. Modello del problema

Supponiamo di indicare con N un set di notizie, con M l'insieme degli utenti verificati e con K quello degli utenti non verificati.

Per ogni news $i \in N$, individuiamo solo gli utenti verificati che hanno interagito con questa notizia. Sia quindi $M_i \subseteq M$ l'insieme degli utenti verificati che hanno condiviso la notizia i. Sia invece $K_{i,j} \subseteq K$ l'insieme degli utenti non verificati che hanno interagito con il post dell'utente verificato j in merito alla news i.

Ad ogni news i, assoceremo un valore di verità $x_i \in \{0, 1\}$ che sarà uguale a 0 in caso di fake news e uguale ad 1 in caso di notizia vera.

Per ogni $j \in M_i$, indichiamo con $y_{i,j} \in \{0, 1\}$ l'opinione che l'utente verificato j ha della notizia i che sarà pari ad 1 se l'utente crede che la notizia i sia vera e 0 il contrario.

Allo stesso modo indichiamo, per ogni $k \in K_{i,j}$, con $z_{i,j,k} \in \{0,1\}$ l'opinione che l'utente non verificato k ha della notizia i che è stata condivisa dall'utente verificato j. $z_{i,j,k}$ sarà pari ad 1 se l'utente k è d'accordo con l'opinione espressa nel tweet e 0 altrimenti. Nel caso in cui $z_{i,j,k}$ sia uguale a 0, la sua opinione verrà successivamente estratta ed analizzata attraverso tecniche di sentiment analysis [15] o tecniche di conflicting opinion mining [16, 17].

E' ora necessario trovare il modo per calcolare questi valori appena presentati [14].
Per ogni news i, il valore di verità x_i viene ottenuto tramite la distribuzione di Bernoulli sul parametro Θ_i.

$$x_i \sim Bernoulli(\Theta_i)$$

Quest'ultimo, a sua volta, viene generato dalla distribuzione Beta su due parametri, γ_1 e γ_0, i quali rappresentano rispettivamente il numero di notizie vere e di notizie false.

$$\Theta_i \sim Beta(\gamma_1, \gamma_0)$$

La stima della credibilità dell'utente verificato viene invece calcolata grazie all'ausilio di altre due variabili, Φ_j^1 e Φ_j^0

$$\Phi_j^1 \coloneqq p\big(y_{i,j} = 1 \,\big|\, x_i = 1\big)$$

$$\Phi_j^0 \coloneqq p\big(y_{i,j} = 1 \,\big|\, x_i = 0\big)$$

Esse rappresentano rispettivamente la probabilità che l'utente j consideri vera la notizia i, la quale è a sua volta vera e la probabilità che l'utente j consideri ancora vera la notizia i che però in questo caso si rivela essere falsa.

Queste due probabilità possono essere ottenute anche attraverso la distribuzione Beta

$$\Phi_j^1 \sim Beta(\alpha_1^1, \alpha_0^1)$$

$$\Phi_j^0 \sim Beta(\alpha_1^0, \alpha_0^0)$$

dove i parametri α_1^1, α_0^1, α_1^0, α_0^0 rappresentato rispettivamente gli utenti che hanno considerato vera una notizia che effettivamente lo era, gli utenti che hanno giudicato vera una notizia che in realtà era falsa, gli utenti che hanno valutato vera una notizia che era però falsa e, in fine, gli utenti che hanno considerato falsa una notizia che effettivamente lo era.

Una volta definiti i parametri Φ_j^1 e Φ_j^0, si è ora in grado di generare il valore associato all'utente verificato $y_{i,j}$ attraverso la distribuzione di Bernoulli

$$y_{i,j} \sim Bernoulli(\Phi_j^{x_i})$$

La stima di credibilità dell'utente non verificato viene anch'essa calcolata grazie all'ausilio di variabili $\Psi_k^{u,v}$, cha rappresentano la probabilità che l'utente non verificato abbia un determinato giudizio k della news, la cui vera condizione è rappresentata da u tenendo conto anche del giudizio v che l'utente verificato ha della medesima notizia.

Abbiamo dunque 4 possibili casi da rappresentare

$$\Psi_k^{0,0} \coloneqq p\big(z_{i,j,k} = 1 \,\big|\, x_i = 0,\, y_{i,j} = 0\big)$$

$$\Psi_k^{0,1} \coloneqq p\big(z_{i,j,k} = 1 \,\big|\, x_i = 0,\, y_{i,j} = 1\big)$$

$$\Psi_k^{1,0} \coloneqq p\left(z_{i,j,k} = 1 \,\middle|\, x_i = 1,\, y_{i,j} = 0\right)$$

$$\Psi_k^{1,1} \coloneqq p\left(z_{i,j,k} = 1 \,\middle|\, x_i = 1,\, y_{i,j} = 1\right)$$

Anche queste quattro probabilità possono essere ottenute attraverso la distribuzione Beta

$$\Psi_k^{u,v} \sim Beta(\beta_1^{u,v}, \beta_0^{u,v})$$

dove $\beta_1^{u,v}$ rappresenta la condizione in cui l'utente non verificato abbia un giudizio positivo della notizia, mentre $\beta_0^{u,v}$ rappresenta la condizione in cui l'utente non verificato abbia un giudizio negativo della notizia. u e v sono valori fissati che possono assumere valori pari a 0 o 1 e rappresentano rispettivamente il valore di verità della news e la considerazione che l'utente verificato ha della notizia.

Una volta ottenuti dunque il valore di x_i e $y_{i,j}$, è possibile rappresentare anche la stima di credibilità dell'utente non verificato tramite la distribuzione di Bernoulli.

$$z_{i,j,k} \sim Bernoulli(\Psi_k^{x_i, y_i, j})$$

Una volta appreso il modo per calcolare questi valori, abbiamo la possibilità di trovare l'istanza della variabile di verità che restituisce l'elemento del dominio che massimizza la funzione di distribuzione congiunta

$$\hat{x}_{MAP} = \arg max \iiint p(x, y, z, \theta, \Phi, \Psi)\, d\theta\, d\Phi\, d\Psi$$

dove, per semplicità, Φ rappresenta $\{\Phi^0, \Phi^1\}$ e Ψ rappresenta $\{\Psi^{0,0}, \Psi^{0,1}, \Psi^{1,0}, \Psi^{1,1}\}$.

Tuttavia, la formula appena descritta è molto complessa da utilizzare, perciò, è stato formulato un nuovo algoritmo che, sfruttando il campionamento di Gibbs, riesce a calcolare il valore di verità della notizia.

3.3. Algoritmo di individuazione di fake news

Per la realizzazione di questo algoritmo, è stato utilizzato il campionamento di Gibbs, il quale è un algoritmo MCMC largamente utilizzato per approssimare una funzione multivariata quando il campionamento diretto è inutilizzabile [18].

Definizione MCMC. I metodi Monte Carlo, basati su Catena di Markov, sono una classe di algoritmi per il campionamento da distribuzioni di probabilità basata sulla costruzione di una catena di Markov avente come distribuzione di equilibrio la distribuzione desiderata.

La logica di questo algoritmo si basa sul campionamento iterativo del valore di verità di ogni notizia con la seguente formula di distribuzione condizionata [14].

$$p(x_i = s \mid x_{-i}, y, z)$$

x_i rappresenta la notizia che può assumere valore 0 in caso di notizia falsa ed 1 in caso di notizia vera, $s \in \{0, 1\}$;

il vettore x_{-i} rappresenta il valore di verità assunto da tutte le notizie eccetto la notizia i;

il vettore y rappresenta i giudizi degli utenti verificati;

il vettore z rappresenta le considerazioni degli utenti non verificati sempre in relazione con i giudizi degli utenti verificati.

La formula appena descritta, al termine di una serie di passaggi, viene ricondotta alla seguente distribuzione di probabilità [14]

$$p(x_i = s|\boldsymbol{x}_{-i}, \boldsymbol{y}, \boldsymbol{z}) \propto \gamma_s \prod_{j \in \mathcal{M}_i} \left(\frac{\alpha^s_{y_{i,j}} + m^s_{j,-i,y_{i,j}}}{\alpha^s_1 + m^s_{j,-i,1} + \alpha^s_0 + m^s_{j,-i,0}} \times \prod_{k \in \mathcal{K}_{i,j}} \frac{\beta^{s,y_{i,j}}_{z_{i,j,k}} + n^{s,y_{i,j}}_{k,-i,z_{i,j,k}}}{\beta^{s,y_{i,j}}_1 + n^{s,y_{i,j}}_{k,-i,1} + \beta^{s,y_{i,j}}_0 + n^{s,y_{i,j}}_{k,-i,0}} \right) \quad (8)$$

che viene utilizzata nel seguente algoritmo

Randomly initialize $x_i^{(0)}$ with 0 or 1, $\forall i \in \mathcal{N}$;
Initialize counts m for $\forall j \in \mathcal{M}$ and n for $\forall k \in \mathcal{K}$;
Sample record $R \leftarrow \emptyset$;
for $t = 1 \rightarrow iter\_num$ **do**
    foreach *news* $i \in \mathcal{N}$ **do**
        Sample $x_i^{(t)}$ using Equation (8);
        Update counts;
    if $t > burn\text{-}in$ & t % $thinning = 0$ **then**
        $R \leftarrow R \cup \{\boldsymbol{x}^{(t)}\}$;
return $\frac{1}{|R|} \sum_{\boldsymbol{x}^{(t)} \in R} \boldsymbol{x}^{(t)}$;

Primo passaggio da compiere è inizializzare in modo casuale il valore di verità, con 0 o 1 e il valore di ogni notizia, calcolando successivamente i valori associati a ciascun utente verificato e non. Conduciamo ora il processo di campionamento per un numero di iterazioni pari al valore di burn-in, ovvero il periodo di tempo necessario per l'allenamento del campionamento. Durante le iterazioni viene campionato il valore di verità di ciascuna notizia utilizzando la formula precedentemente enunciata (8). Aggiorniamo poi i valori associati ad ogni utente ad intervalli definiti dalla variabile thinning, ogni volta che il resto derivante dal rapporto tra questa e il numero di iterazioni t è uguale a 0. L'algoritmo restituirà come output la media tra tutti i valori x ottenuti [14].

3.4. Sperimentazione del metodo

La metodologia appena descritta è stata sperimentata avendo come metriche di valutazione accuracy, precision, recall e F1-score.

3.4.1. Metriche di valutazione – cenni

Per valutare le prestazioni degli algoritmi di individuazione di fake news, vengono utilizzate varie metriche di valutazione. Le metriche però più utilizzate per il rilevamento di notizie false sono accuracy, precision, recall e F1-score.

Esse vengono ottenute tramite le seguenti formule [19]

$$
\begin{aligned}
Precision &= \frac{|TP|}{|TP| + |FP|} \\
Recall &= \frac{|TP|}{|TP| + |FN|} \\
F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \\
Accuracy &= \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}
\end{aligned}
$$

dove i valori TP, TN, FN, FP rappresentano rispettivamente i veri positivi, i veri negativi, i falsi negativi e i falsi positivi.

L'utilizzo di queste specifiche metriche da l'opportunità di valutare una metodologia da prospettive differenti.
L'accuracy misura la somiglianza tra le notizie false che sono state previste e le notizie che lo sono effettivamente, la precision valuta l'abilità di non etichettare una notizia positiva quando in realtà è negativa, la recall valuta la capacità di trovare tra tutte le notizie considerate positive quelle che effettivamente lo sono ed in fine la F1-score rappresenta la media tra le metriche precision e recall.

Fattore da dover specificare è che maggiori sono i valori di queste metriche, migliori sono le prestazioni dell'algoritmo [19].

3.4.2. Logica di sperimentazione

La sperimentazione viene basata sul confronto dell'algoritmo realizzato con altri 4 algoritmi destinati anch'essi all'individuazione di fake news. La valutazione avviene comparando i valori di accuracy, precision, recall e F1-score ottenuti da ogni singolo algoritmo.

I dataset che sono stati consultati per la fase di sperimentazione sono il dataset LIAR e il dataset BuzzFeed. Mentre gli algoritmi che sono stati selezionati sono l'algoritmo Majority Voting, il quale per ogni notizia fornisce come risultato della stima l'opinione più frequente di un utente verificato; l'algoritmo TruthFinder [20], che calcola in modo iterativo il valore di verità di ogni notizia in base alle relazioni conflittuali tra giudizi di utenti verificati; l'algoritmo LTM [21] e l'algoritmo CRH [22].

Fig.12

Performance comparison on LIAR dataset

Methods	Accuracy	True			Fake		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Majority Voting	0.586	0.624	0.628	0.626	0.539	0.534	0.537
TruthFinder	0.634	0.650	0.679	0.664	0.615	0.583	0.599
LTM	0.641	0.654	0.691	0.672	0.624	0.583	0.603
CRH	0.639	0.653	0.687	0.669	0.621	0.583	0.601
UFD	**0.759**	**0.766**	**0.783**	**0.774**	**0.750**	**0.732**	**0.741**

Performance comparison on BuzzFeed dataset

Methods	Accuracy	True			Fake		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Majority Voting	0.556	0.532	0.373	0.439	0.567	0.714	0.632
TruthFinder	0.554	0.523	0.359	0.426	0.568	**0.720**	0.635
LTM	0.465	0.443	0.582	0.503	0.500	0.364	0.421
CRH	0.562	0.542	0.388	0.452	0.573	0.714	0.636
UFD	**0.679**	**0.667**	**0.714**	**0.690**	**0.692**	0.643	**0.668**

In Fig.12 sono riportati i risultati ottenuti in seguito all'analisi delle metriche per ogni singolo algoritmo. Come possiamo notare, l'algoritmo descritto in questo capitolo, presenta riscontri migliori su gli altri quattro analizzati [14].

4. Esempio di approccio per l'individuazione di fake news basato su reti bayesiane

Utilizzando la metodologia delle reti bayesiane, è di seguito riportata una soluzione per l'individuazione di fake news.

Le tecnologie che sono state utilizzate nell'implementazione sono le seguenti:

- linguaggio R nella versione 4.0.3, è un linguaggio di programmazione per l'elaborazione statistica e grafica. Le librerie che sono state utilizzate sono bnlearn, Rgraphviz, graph e BioGenerics
- RStudio nella versione 1.3.1093, è un ambiente di sviluppo integrato per il linguaggio R
- dataset LIAR, esso contiene 12836 news, provenienti da contesti diversi, ottenute attraverso l'API del sito PolitiFact.com. Il loro valore di verità è stato identificato in seguito a stime compiute dall'uomo che comprendono 6 diverse tipologie, "pants-fire", "false", "barely-true", "half-true", "mostly true" e "true".

4.1. Implementazione del sistema

Primo passo che è stato compiuto nella fase implementativa del sistema, è stato lo studio e l'analisi del dataset LIAR. Esso comprende molteplici news per ciascuna delle quali fornisce numerose informazioni, l'id della notizia, il valore di verità, il titolo, il contesto, la fonte, il ruolo che ha la fonte nella società, il

suo stato di provenienza, la preferenza politica, il mezzo di divulgazione utilizzato e i valori numerici che giustificano l'attribuzione del valore di verità alla news.

Ebbene, tutte queste informazioni che fornisce il dataset LIAR, sono state utilizzate solo in parte. Le nozioni di cui si è fatto uso sono il valore di verità, il titolo, il contesto e il ruolo che ha la fonte nella società. Inoltre, non tutte le news presenti nel dataset sono state utilizzate, solamente coloro che presentavano un valore di verità pari a "false" o a "true" ed un contesto che rientrasse nell'ambito politico, "politics", o sociale, "world".

Si è deciso di compiere questa selezione dal momento che, la rete bayesiana proposta, rappresenta solo un primo prototipo dell'utilizzo di questa tecnologia per l'individuazione delle fake news.

4.1.1. Costruzione della rete bayesiana

Una volta aver selezionato le informazioni necessarie dal dataset di partenza LIAR e averle raccolte in un data frame, i dati sono stati discretizzati così come segue.

I titoli di ciascuna notizia sono stati suddivisi in due diverse tipologie, i titoli clickbait e i titoli non clickbait [23], invece le informazioni relative al ruolo che ha la fonte all'interno della società, sono state utilizzate per classificare, secondo tre livelli di affidabilità A, B e C, la fonte stessa. Per quanto riguarda il valore di validità e la categoria esse si presentano già discretizzate.

Una volta selezionati i dati, il passo successivo prevede la fase dello structural learning.

Il problema dello structural learning ha come obbiettivo l'apprendimento della struttura della rete bayesiana utilizzando algoritmi di apprendimento automatico a partire da un insieme di dati iniziali. Questi algoritmi, in base alla logica da loro utilizzata, possono essere classificati in algoritmi score-based e algoritmi constraint-based [24].

L'algoritmo che si è deciso di utilizzare per l'implementazione è l'algoritmo Hill Climbing di tipo score-based. Gli algoritmi di tipo score-based classificano ogni modello in base ad un punteggio che ne descrive la qualità. Questo punteggio viene influenzato dalla presenza di archi, quindi di dipendenze tra le variabili del problema [24].

In Fig.13, è rappresentato lo pseudo codice dell'algoritmo Hill Climbing.

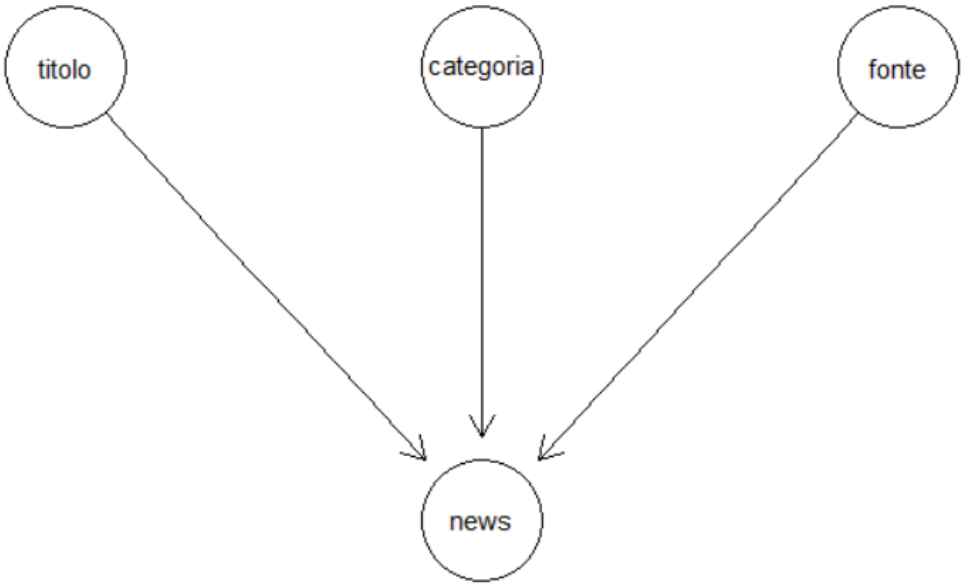
Fig.13

```
A = ∅;
G = (U, A);
score = −∞;
maxscore = score;
Per ogni possibile aggiunta o rimozione di archi all'interno di G
        se G* è aciclico
                calcola score*
                se score* > score
                        G = G*
                        maxscore = score*
return G
```

L’algoritmo considera inizialmente un grafo G privo di archi, la variabile score inizializzata a $-\infty$ e la variabile maxscore alla quale viene associato il valore iniziale di score. Per ogni possibile cambiamento che può essere effettuato all’interno della struttura della rete, quindi aggiungendo e rimuovendo archi, viene calcolato, ad ogni iterazione, il valore score* del nuovo grafo G*. Nel caso in cui il nuovo grafo G* non presenti cicli ed il nuovo valore score* sia maggiore del precedente valore score, la nuova struttura della rete sarà il grafo G* e la variabile maxscore verrà posta uguale a score*. In questo modo l’algoritmo restituirà in output la struttura migliore per la creazione della rete bayesiana [24].

La struttura ottenuta in seguito all’utilizzo dell’algoritmo Hill Climbing è riportata in Fig. 14.

Fig.14



Come possiamo vedere dalla rete, Fig.14, le uniche relazioni di dipendenza intercorrono tra il nodo titolo e il nodo news, tra il nodo categoria e il nodo news e tra il nodo fonte e il nodo news. Ciò significa che ciascuna delle variabili del problema influenza il valore di veridicità della notizia in analisi.

4.2. CPT ottenute

In seguito alla costruzione della rete bayesiana, utilizzando come dati iniziali le news selezionate dal dataset LIAR, le tavole di probabilità condizionate (CPT) ottenute, ci permettono di valutare, in maniera ancor più approfondita, le dipendenze tra le variabili del problema.

Qui di seguito sono riportate alcune tavole di probabilità condizionate restituite in output dalla rete bayesiana individuata.

Fig.15

```
, , categoria = world, fonte = A

       titolo
news           CL       NCL
  false 0.3333333 0.2000000
  true  0.6666667 0.8000000
```

Fig.16

```
, , categoria = world, fonte = B

       titolo
news           CL       NCL
  false 0.6000000 0.3333333
  true  0.4000000 0.6666667
```

Fig.17

```
, , categoria = world, fonte = C

       titolo
news           CL       NCL
  false 0.8750000 0.8000000
  true  0.1250000 0.2000000
```

Analizzando le CPT mostrate in Fig.15, 16 e 17, siamo in grado di fare un'importante osservazione. La variabile del problema riferita all'attendibilità della fonte, rappresenta un attributo di grande valenza per il calcolo del valore di verità della news.

Basti prendere in esempio la tavola di probabilità condizionate in Fig.17. Il valore associato all'attributo fonte è pari a C, dunque la fonte è poco attendibile, mentre, per quando riguarda il titolo, vengono mostrate entrambe le tipologie, clickbait e non clickbait.

Le fake news, il più delle volte, sono caratterizzate da titoli clickbait, finalizzati ad attirare quanti più utenti possibili [23].

Dalla CPT in figura ci rendiamo conto che la valenza della fonte è maggiore dell'influenza che può esercitare il titolo. Una news divulgata da una fonte non attendibile ha un'alta probabilità di essere falsa. Infatti, nonostante tra le news analizzate siano presenti anche notizie con titoli non clickbait, esse sono state valutate dalla rete bayesiana fake.

BIBLIOGRAFIA

[1] Zhou, Xinyi, and Reza Zafarani. "A survey of fake news: Fundamental theories, detection methods, and opportunities." ACM Computing Surveys (CSUR) 53.5 (2020).

[2] Giordano, Giuseppe, Serena Mottola, and Beatrice Paternoster. "Some mathematical aspects to detect fake news: a short review." 2020 International Conference on Mathematics and Computers in Science and Engineering (MACISE). IEEE, 2020.

[3] Ozbay, F.A., Alatas, B., Fake news detection within online social media using supervised arti_cial intelligence algorithms, Physica A: Statistical Mechanics and its Applications, 540, 2020.

[4] Wynne H.E., Wint Z.Z., Content Based Fake News Detection Using N N-Gram Models, Association for Computing Machinery, 669{673, 2019.

[5] Shu, Kai, and Huan Liu. "Detecting fake news on social media." Synthesis Lectures on Data Mining and Knowledge Discovery 11.3 (2019): 1-129.

[6] Natali Ruchansky, Sungyong Seo, and Yan Liu. CSI: A hybrid deep model for fake news. ArXiv Preprint ArXiv:1703.06959, 2017.

[7] Nikos Deligiannis, Tien Huu Do, Duc Minh Nguyen, and Xiao Luo. Deep learning for geolocating social media users and detecting fake news.

[8] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. Fake news detection on

social media using geometric deep learning. ArXiv Preprint ArXiv:1902.06673, 2019.

[9] O. Pourret, P. Naim, and B. Marcot. Bayesian Networks: A Practical Guide to Applications. John Wiley & Sons, Ltd, 2008.

[10] M. L. Krieg. A Tutorial on Bayesian Belief Networks. DSTO Electronics and Surveillance Research Laboratory, 2001.

[11] Lauritzen,S.L. e Spiegelhalter, D.J. Local computations with probabilities on graphical structures and their application to expert systems, 1988, Journal of the Royal Society, Vol. Series B, p. 157-224.

[12] Sparacino, Flavia. "The Museum Wearable: Real-Time Sensor-Driven Understanding of Visitors' Interests for Personalized Visually-Augmented Museum Experiences." (2002).

[13] S. Cenatiempo, G. D'Agostini, and A. Vanelli. Reti bayesiane: da modelli di conoscenza a strumenti inferenziali e decisionali. Notiziario tecnico Telecom Italia, (3):16–25, 2010.

[14] Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. Unsupervised fake news detection on social media: A generative approach. In Proceedings of the AAAI Conference on Artificial Intelligence, 2019

[15] Gilbert, C. H. E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media (ICWSM), 2014

[16] Trabelsi, A., and Zaiane, O. R. 2014. Mining contentious documents using an unsupervised topic model based approach.

In 2014 IEEE International Conference on Data Mining (ICDM), 550–559. IEEE.

[17] Dave, K.; Lawrence, S.; and Pennock, D. M. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web, 519–528. ACM.

[18] Robert, Christian, and George Casella. Monte Carlo statistical methods. Springer Science & Business Media, 2013.

[19] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36, 2017

[20] Yin, X.; Han, J.; and Philip, S. Y. Truth discovery with multiple conflicting information providers on the web. IEEE Transactions on Knowledge and Data Engineering 20(6):796–808, 2008

[21] Zhao, B.; Rubinstein, B. I.; Gemmell, J.; and Han, J. A bayesian approach to discovering truth from conflicting sources for data integration. Proceedings of the VLDB Endowment, 2012

[22] Li, Q.; Li, Y.; Gao, J.; Zhao, B.; Fan, W.; and Han, J. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data, 1187–1198. ACM, 2014

[23] Potthast, M., Köpsel, S., Stein, B., & Hagen, M. Clickbait detection. In European Conference on Information Retrieval (pp. 810-817). Springer, Cham, 2016

[24] Scutari, Marco, and Jean-Baptiste Denis. Bayesian networks: with examples in R. CRC press, 2014.

[25] Krieg, Mark L. A tutorial on Bayesian belief networks. No. DSTO-TN-0403. DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION SALISBURY (AUSTRALIA) ELECTRONICS AND SURVEILLANCE RESEARCH, 2001.