

UNIVERSITÀ DEGLI STUDI DI PARMA

Corso di Laurea in Ingegneria dei Sistemi Informativi

Tesi di Laurea di primo livello

Realizzazione di un assistente vocale per il triage ospedaliero

Triage semi-automatizzato per la sicurezza e la tempestività



Relatori

prof.ssa Mordonini Monica
prof. Tomaiuolo Michele

Laureando

Simone Montali
matricola 288144

Probabilmente 2020

Dedica toccante.

Citatemi dicendo che sono stato citato male.
Groucho Marx

Ringraziamenti

Grazie

Sommario

Abstract della tesi

Indice

Elenco delle figure	VIII
1 Introduzione	1
2 Stato dell'arte	2
2.1 Machine Learning	2
2.1.1 Reti neurali	3
2.2 Assistenti vocali	4
2.2.1 Assistenti vocali in ambito medico	5
2.3 Triage	7
2.3.1 Come il COVID19 ha modificato le procedure di triage	7
3 Mycroft	9
3.1 Open source	9
3.1.1 Cosa si intende con open source?	9
3.1.2 Perché open source?	10
3.2 Introduzione a Mycroft	10
3.2.1 Stack di funzionamento	10
3.2.2 Interfaccia grafica	12
4 Realizzazione di skills	13
4.1 Struttura dei file della skill	13
4.2 Funzionamento della skill	14
4.2.1 Metodi relativi alle sintomatologie	15
4.2.2 Decorators utilizzati	16
4.2.3 Helper methods	18
4.2.4 Oggetto <code>med_record</code>	19
5 Possibilità di inference tramite reti neurali	21
5.1 Classificazione del testo	21
5.1.1 Dataset di trascrizioni	21
5.1.2 Inference sulle trascrizioni	22
5.1.3 Inference sulle keywords	26
A An appendix	28

Elenco delle figure

2.1	Modello lineare	3
2.2	Modello a 2 strati, ancora lineare	3
2.3	Rete neurale	4
2.4	Percentuali di utenti che hanno utilizzato/vorrebbero utilizzare assistenti vocali in ambito medico	5
2.5	Interesse dei consumatori negli assistenti vocali in ambito medico, per fasce di età	6
2.6	Schema del protocollo CESIRA	7
3.1	Schema di funzionamento di Precise	11

Capitolo 1

Introduzione

In una società sempre più digitalizzata e dinamica, spesso viene a crearsi un netto distacco tra i settori capaci di **evolvere assieme alla tecnologia**, e quelli che, per un motivo o per l'altro, rimangono ancorati a procedure e metodologie tradizionali. Il settore medico, in continuo rinnovamento sul lato scientifico, è, soprattutto in Italia, affidato ad **infrastrutture informatiche progettate tempo fa**. Questo, soprattutto per motivi di stabilità e affidabilità: gli errori, qui, non sono ammissibili.

Per questo motivo, spesso non si notano le evidenti possibilità di miglioramenti che la ricerca informatica potrebbe apportare agli ospedali, agli ambulatori, agli studi. L'obiettivo di questa tesi è mettere luce su una delle possibili modalità con cui l'informatica potrebbe, in un futuro prossimo, migliorare la praticità ma soprattutto la sicurezza degli ambienti ospedalieri.

La procedura di triage, ossia il processo di selezione dei pazienti richiedenti cure, è oggi affidata totalmente ad infermieri. Questa scelta è dovuta, oltre ad un evidente bisogno di poter osservare il paziente, alla necessità dell'**immediatezza di un contatto verbale** con il personale sanitario.

Perciò, la sfida nella realizzazione della tesi è soprattutto legata al rendere il **più umana possibile** l'interazione con un bot automatizzato. Il bot deve quindi accogliere il paziente, comprenderne le problematiche ed i sintomi, e farlo sentire protetto. Non si escluderà del tutto un apporto umano: gli infermieri sono addestrati per poter osservare e comprendere il richiedente cura, e l'apporto dell'osservazione diretta del paziente è ancora troppo importante per escluderla. È senz'altro possibile, però, affidare la prima parte di **profilazione dell'utente** a procedure automatizzate.

Capitolo 2

Stato dell'arte

Illustriamo ora lo stato dell'arte delle varie tecniche e tecnologie che coinvolgono il bot.

2.1 Machine Learning

Il machine learning (in seguito, ML) è una disciplina che si occupa di rispondere, sostanzialmente, a due domande: *Come si può costruire un sistema informatico che migliora con l'esperienza?* e *Quali sono le leggi teoriche fondamentali che governano ogni sistema di apprendimento, sia esso implementato nei computer, negli umani, nelle organizzazioni?*. Il ML copre una serie di compiti diversi, dalla classificazione di email spam, al riconoscimento facciale, al controllo di robot [4]. Ogni problema di ML può essere riassunto come il miglioramento di una determinata performance (data, ad esempio, dall'errore sulle predizioni) durante l'esecuzione di un'operazione. Seguendo l'esempio delle email spam, un algoritmo di ML potrebbe apprendere, tramite un *training set* di email già etichettate come spam/non-spam, a categorizzarne altre. Portando il problema in termini matematici, vogliamo, dato un grafico in cui i punti rossi sono email spam, e quelli blu no, trovare una funzione che ci permetta di evidenziare le email da eliminare. Ciò che faremo sarà quindi definire un insieme di **features**, consistenti, ad esempio, nel contenuto testuale della mail (vettorizzato in termini numerici), l'orario di arrivo, la lunghezza in parole... Oltre a ciò, definiremo una **label**, ossia la caratteristica che vogliamo predire: sarà, per esempio, 1 se l'email è spam, 0 altrimenti. Ora, il problema consiste in una semplice regressione lineare: date le features x_0, x_1, x_2 e la label y , vogliamo trovare dei pesi w_0, w_1, w_2 tali che

$$w_0x_0 + w_1x_1 + w_2x_2$$

approssimi al meglio y . Per fare ciò, procediamo per iterazioni: partiremo con dei pesi arbitrari, calcolando y per ogni esempio (già etichettato) fornito, e ne calcoleremo l'errore rispetto alla y reale. Potremo quindi calcolare il gradiente (ossia la somma delle derivate parziali sui rispettivi pesi) della funzione che lega l'errore ai pesi w_i , ottenendone un'informazione fondamentale: indicherà infatti la direzione nella quale la funzione errore è decrescente. Ci potremo quindi spostare, di una quantità pari al *learning rate* (uno degli iperparametri definiti arbitrariamente), nella direzione di diminuzione dell'errore. Scegliere un corretto *learning rate* è una parte fondamentale dello sviluppo dell'algoritmo di ML:

se è troppo piccolo, la soluzione richiederà troppo tempo, se troppo grande, la soluzione non sarà abbastanza precisa. Dopo un sufficiente numero di esecuzioni dell'algoritmo, il nostro grafico avrà ora questo aspetto: e saremo quindi in grado di determinare, con una determinata *accuracy*, l'appartenenza di un esempio ad una classe o all'altra.

2.1.1 Reti neurali

Per illustrare il concetto di rete neurale, produciamo anzitutto un grafico del modello lineare appena definito.

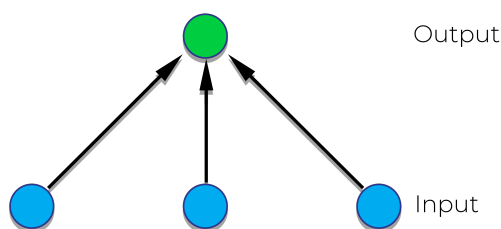


Figura 2.1: Modello lineare

Ogni cerchio blu rappresenta una feature, ed il cerchio verde rappresenta la somma pesata degli input. Vorremmo che questo modello fosse in grado di approssimare anche problemi non-lineari. Potremmo quindi decidere di inserire un nuovo strato di valori intermedi, ottenendo un modello così composto:

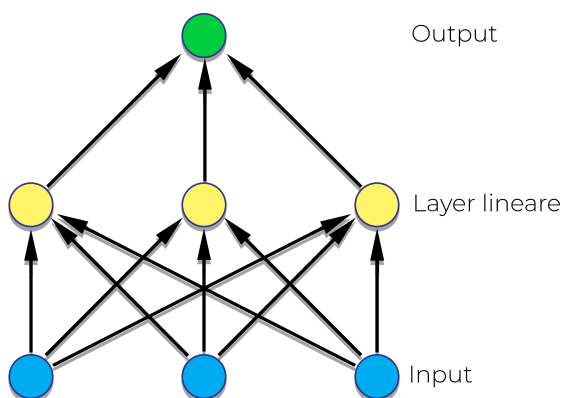


Figura 2.2: Modello a 2 strati, ancora lineare

c'è, però, un problema: essendo questo ancora un modello lineare, è possibile semplificarlo, ottenendo il primo grafico. È utile, allora, inserire una non-linearità all'interno del modello, che chiameremo *funzione di attivazione*: ora avremo, oltre ai layer lineari, una funzione tra di loro non lineare. Grazie a essa, l'aggiunta di nuovi strati è diventata significativa: non sarebbe più possibile semplificare il sistema ad uno lineare.

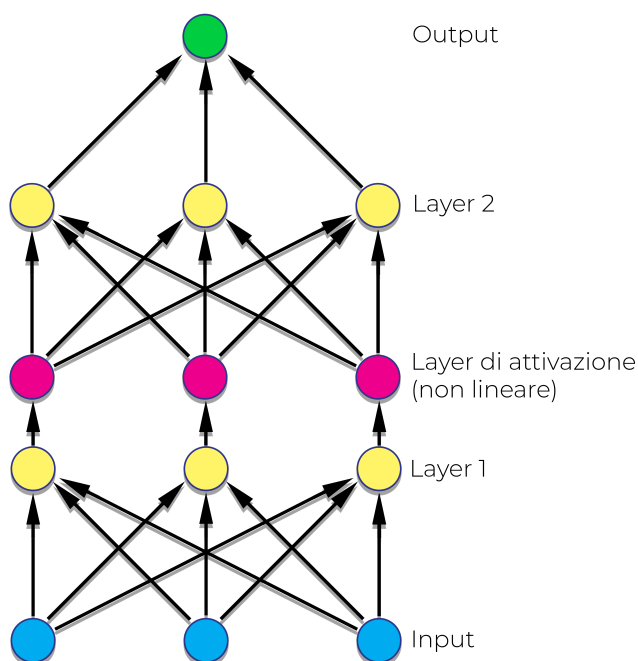


Figura 2.3: Rete neurale

Tra le funzioni di attivazione, citiamo la funzione sigmoidea $F(x) = \frac{1}{1+e^{-x}}$, o la *Rectified Linear Unit* (ReLU) $F(x) = \max(0, x)$. Non esiste una funzione di attivazione più corretta delle altre: ogni problema avrà una configurazione dei layer e delle loro attivazioni diversa. Trovare la configurazione corretta della rete neurale è la sfida più grande nel lavoro di *data analysis*.

Avremo ora quindi un modello capace di ottenere degli output date le input features. Ma come possiamo addestrarlo per ottenere dei risultati validi? Una soluzione ricorrente è quella della **retropropagazione dell'errore**, o *backpropagation*. Durante questo processo, la rete neurale calcola il gradiente dell'errore (ottenuto da una funzione di *loss*) sull'output, rispettivamente ai vari pesi degli strati. Questo permette alla rete neurale di individuare quali nodi hanno più effetto sulla generazione dell'errore, e di conseguenza, regolarne i pesi per ridurre l'errore. Fatto ciò sullo strato più vicino all'output, si tratta solo di ripetere il procedimento "*a salire*" per ogni layer della rete neurale.

2.2 Assistenti vocali

Gli assistenti vocali sono software in grado di interpretare dialoghi umani e rispondere attraverso voci sintetizzate. Alcuni esempi famosi sono Apple Siri, Amazon Alexa, Google Assistant. Gli utenti possono effettuare domande, controllare dispositivi domotici, e svolgere un numero di operazioni in continua espansione [3]. Il software è costantemente in attesa di una *wake word*; una volta riconosciuta questa parola, si mette in ascolto di comandi. La richiesta viene quindi tradotta in testo tramite un motore di *Speech To Text* ed interpretata dal software. Questa è senz'altro la parte più complessa: il compito di interpretazione di

linguaggio naturale è reso complesso dalle svariate sfaccettature che ogni diversa lingua presenta. Interpretata la richiesta, l'assistente procede con le operazioni collegate, e prepara una risposta testuale da dare all'utente. Questa risposta verrà poi sintetizzata in audio da un motore di *Text To Speech*. I moderni dispositivi di assistenza vocale come il Google Home o l'Amazon Echo si appoggiano a server in cloud per l'interpretazione delle richieste: questo permette di ridurne notevolmente le dimensioni ed i requisiti tecnici. Lo sviluppo, negli ultimi anni, di nuove tecniche di ML, unite ad un grande incremento delle potenze di calcolo e alla disponibilità di vastissimi *dataset linguistici*, hanno permesso grandi miglioramenti nella tecnologia, ma soprattutto l'applicazione ad ambiti, fino a qualche anno fa, impensabili.

2.2.1 Assistenti vocali in ambito medico

Negli anni sono emerse alcune possibili applicazioni delle tecnologie per assistenti vocali in ambito medico. La startup **Saykara** ha creato un software che permette di ascoltare le conversazioni medico-paziente, allo scopo di creare schede cliniche accurate senza dover impegnare il medico in lunghe procedure di *data entry*. Un report [5] di Voicebot, organizzazione per la diffusione degli assistenti vocali, afferma che il 52% degli intervistati sarebbe intenzionato ad utilizzare assistenti vocali in ambito medico nel futuro. Solo un 7.5% (1 su 13) l'ha già fatto.

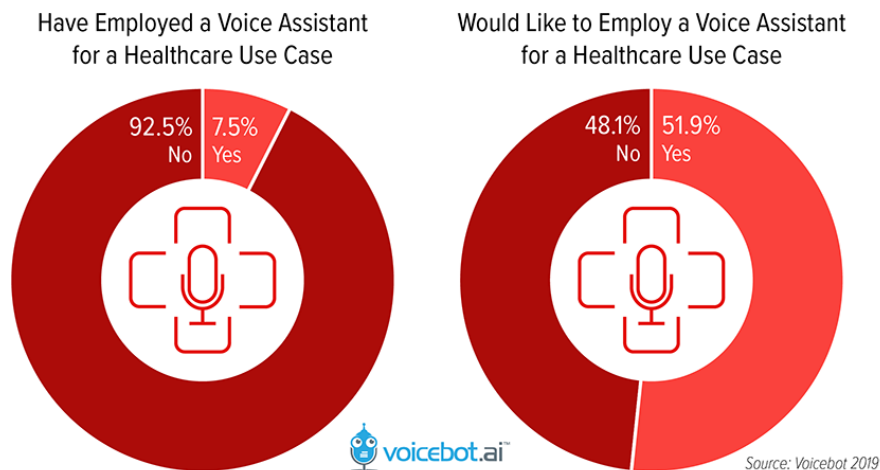


Figura 2.4: Percentuali di utenti che hanno utilizzato/vorrebbero utilizzare assistenti vocali in ambito medico

È interessante notare come anche le fasce di età più avanzate siano, seppur con percentuali più basse, interessate a provare simili applicazioni della tecnologia.

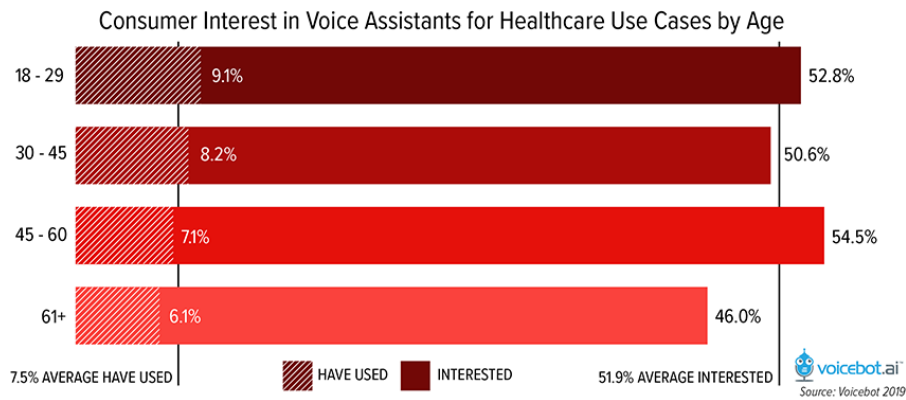


Figura 2.5: Interesse dei consumatori negli assistenti vocali in ambito medico, per fasce di età

Un altro esempio interessante di applicazione della tecnologia in ambito medico è *Dragon Medical Virtual Assistant* rilasciato da **Nuance**, una suite di strumenti vocali capaci di creare cartelle cliniche, eseguire diagnostica sul paziente, fornire risultati in diretta di ricerche mediche, supportare le operazioni di radiologia.

2.3 Triage

Il **triage** è, per definizione, il sistema utilizzato per selezionare i soggetti coinvolti in infortuni, gravi o leggeri che siano, secondo classi di emergenza, in base alla gravità delle lesioni riportate o del loro quadro clinico. Il termine nacque in ambiti di guerra (se ne cominciò a parlare durante le *guerre napoleoniche*), durante i quali si classificavano i soldati feriti in base alla probabilità di sopravvivenza. Il procedimento venne poi, negli anni, iniziato ad utilizzare in ambito ospedaliero, con l'applicazione di metodi più scientifici come il modello *START*, basato su un algoritmo che prevede la verifica di sintomi visibili, frequenza respiratoria, pulsazioni, risposta ai comandi. Sebbene inizialmente le procedure di triage fossero sfruttate solo in ambito di maxiemergenze, oggi tutti i pronto soccorso italiani ne fanno uso in qualche modo. Non esiste però uno standard di codici utilizzati in ogni ospedale: le normative cambiano tra un pronto soccorso e l'altro. Distinguiamo però, generalmente, 4 tipi di codice:

- Codice **rosso**: paziente critico con cedimento di una o più funzioni vitali. *Accesso immediato.*
- Codice **giallo**: potenziale rischio di vita o compromissione di una o più funzioni vitali. *Accesso entro 15 minuti.*
- Codice **verde**: valutazione medica differibile nel tempo senza alcun rischio evolutivo. *Accesso entro 60-90 minuti.*
- Codice **bianco**: paziente senza emergenza, con problematiche risolvibili a domicilio o dal medico di famiglia.

Il triage infermieristico produce quindi una **scheda di triage**, che ha valore giuridico, contenente anagrafica, data/ora di accettazione, motivo, codice colore. Durante le emergenze è comune utilizzare il protocollo **CESIRA** (Coscienza, Emorragia, Shock, Insufficienza respiratoria, Rotture, Altro).

2.3.1 Come il COVID19 ha modificato le procedure di triage

La pandemia mondiale di COVID19 ha modificato i processi di triage in ospedale: lo scopo è limitare al massimo le possibilità di contagio tra i pazienti, rendendo il pronto soccorso più sicuro tanto per i pazienti, quanto per il personale sanitario. Alcuni ospedali in Inghilterra hanno addirittura iniziato a praticare triage remoto: il paziente comunica le sintomatologie online

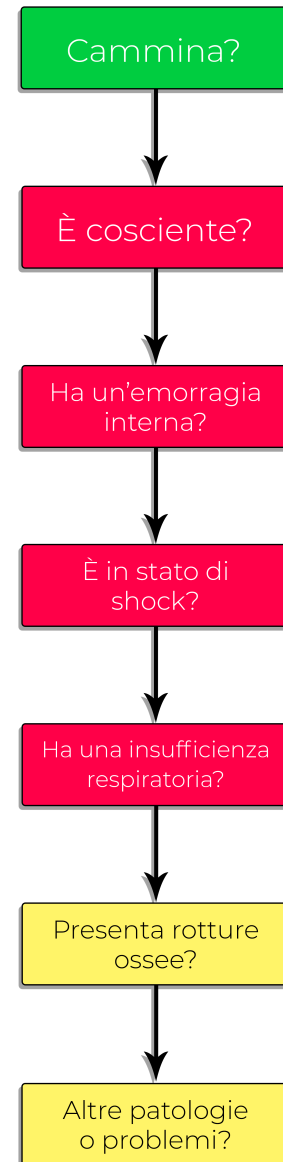


Figura 2.6: Schema del protocollo CESIRA

e, solo se necessario, può recarsi in ospedale per la visita medica [2]. In Italia, ogni ospedale ha provveduto a definire delle procedure di **pre-triage** atte ad evitare il contatto tra pazienti affetti da COVID19 e non. Generalmente, si tratta di percorsi obbligati in cui i pazienti con sintomi compatibili con COVID19 (febbre, tosse, insufficienza respiratoria) vengono separati in un'*area di decontaminazione* e visitati senza accedere agli ambienti del pronto soccorso comune. Nell'ospedale Maggiore di Parma sono state previste, durante i momenti di picco dell'epidemia, due tende di pre-triage esterne all'ospedale stesso.

Capitolo 3

Mycroft

There's an entire community of developers looking to access this technology, but so far, it's been the purview of a few large companies. The technology is walled-off, proprietary, and secretive

Joshua Montgomery
CEO di Mycroft AI, Inc.

Nel panorama degli assistenti vocali disponibili al pubblico, le grandi aziende catturano la maggior parte delle attenzioni dei consumatori. I nomi di Amazon Alexa, Google Home, Microsoft Cortana sono familiari a tanti. Meno familiare, è, però, lo scenario degli stessi strumenti in ambito di assistenti vocali. Piattaforme come *Jarvis*, *Linto*, *Open Assistant* e **Mycroft** sono sconosciute ai più. Quest'ultimo ha, negli ultimi anni, catturato l'attenzione di molti esperti di settore per la sua stabilità, espansibilità, popolarità.

3.1 Open source

3.1.1 Cosa si intende con open source?

Il termine **open source** si riferisce a qualcosa di modificabile e condivisibile dalle persone, avente quindi un **design pubblicamente accessibile**. Il termine nacque nell'ambito dello sviluppo software, ma oggi rappresenta piuttosto un'etica. Il software open source è software il cui codice sorgente è ispezionabile, modificabile, e migliorabile da chiunque voglia farlo. Con *codice sorgente* intendiamo il codice che definisce il comportamento del programma, ossia il prodotto del lavoro di un programmatore.

Molti dei software che oggi utilizziamo sono invece **closed source** (o *proprietario*), ossia software il cui codice sorgente non è accessibile agli utenti, e sul quale l'organizzazione che crea il software ha *pieni poteri*. Alcuni esempi di software proprietario possono essere Windows, Photoshop, Safari. Solitamente, durante l'installazione di uno di questi software,

l'utente accetta dei termini e condizioni che lo vincolano a non fare nulla di non autorizzato dall'azienda creatrice del software.

3.1.2 Perché open source?

Le motivazioni per preferire software open source rispetto a quelli proprietari sono svariate. Sebbene i software proprietari spesso abbiano funzionalità più avanzate e meglio sviluppate, hanno gravi mancanze in progetti come questo. Il software open source, a scapito della possibilità di minori funzionalità (nonostante ciò non sia sempre vero, basti pensare alla potenza del kernel Linux), dà ad un utente esperto tante possibilità:

- Possibilità di **controllo** sul software: la modifica e il miglioramento del software è permessa e ben vista. La comunità open source lavora con coesione per produrre software sempre migliore.
- Possibilità di conoscere e studiare i **meccanismi interni**: l'utilizzo di software per fini scientifici richiede la piena conoscenza di come un software funziona. Inoltre, la disponibilità del codice a tutti permette agli studenti e agli interessati di scoprire le logiche che ne permettono il corretto funzionamento.
- Grandi garanzie di **sicurezza**: se tutti possono accedere al codice, la presenza di bug di sicurezza e problematiche di privacy viene rilevata molto più velocemente di come accade nel software proprietario.
- Certezze sulla **privacy**: la società in cui viviamo è sempre più basata sulla vendita di dati personali acquisiti tramite software. L'open source, spesso e volentieri, si oppone a tale tendenza: essendo software della comunità e non di un'azienda che punta all'ottenimento di capitale, nessuno ha interesse nel *monetizzare gli utenti*.
- Supporto della **comunità**: lo sviluppo di software open source è molto supportato dalla comunità informatica. In caso di problematiche o necessità, ci sarà sempre un utente più esperto disposto ad aiutare.
- Correttezza **etica**: software come quello oggetto di questa tesi hanno l'obiettivo di salvare vite. È giusto che tutti possano accedervi, migliorarlo e conoscerlo. La ricerca scientifica dovrebbe essere il più possibile **libera**.

3.2 Introduzione a Mycroft

Mycroft nacque tramite crowdfunding nel 2015 e catturò subito molto interesse da parte della comunità informatica. L'obiettivo era chiaro: opporsi alla sola presenza di software proprietario in ambito open source. Ricevette presto l'appoggio di molte organizzazioni come la Canonical (promotrice di Ubuntu) e la Mozilla Foundation. La repository su Github di Mycroft conta oggi *4300 stars* in continua crescita.

3.2.1 Stack di funzionamento

Come spiegato nel capitolo 2.2, un assistente vocale si compone principalmente di:

Rilevamento della wake word

L'assistente ha bisogno di rilevare una parola per attivarsi. Gli esempi più famosi sono *Hey, Google* o *Alexa*. Mycroft permette di personalizzare la propria wake word, che all'installazione è semplicemente *Hey, Mycroft*. Il progetto utilizza **Precise**, un wake word listener basato su reti neurali addestrate su esempi sonori. Questo componente, totalmente open source, è basato su *pattern sonori*, non sul testo. Questo ne riduce la dipendenza da accenti e linguaggi diversi. Precise offre la possibilità di addestrare il listener su un proprio dataset. Il funzionamento è basato su una singola rete neurale ricorrente, più precisamente una **Gated Recurrent Unit** o GRU.

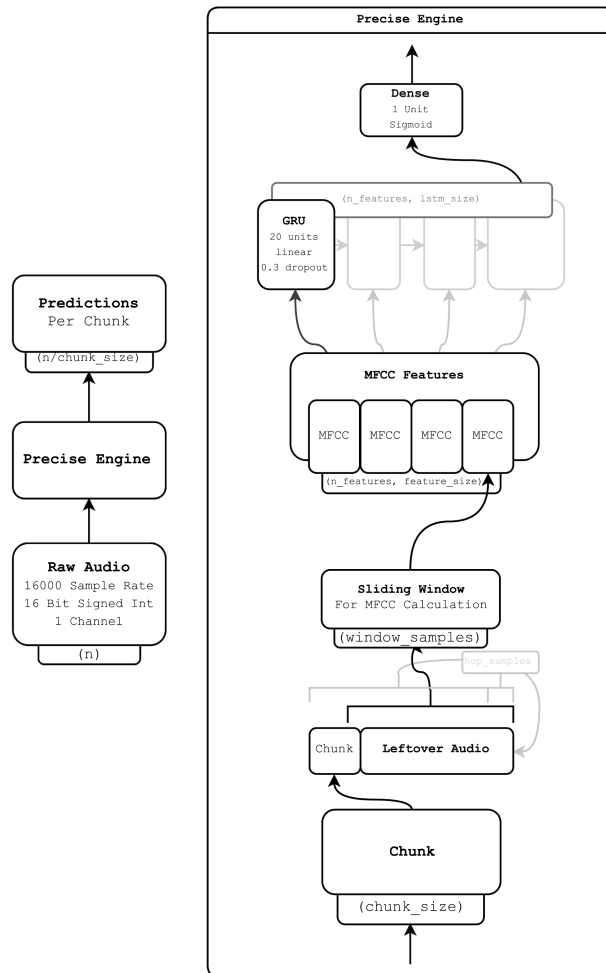


Figura 3.1: Schema di funzionamento di Precise

Speech To Text

Mycroft sfrutta il motore STT di Google per effettuare questo processo. Per aggiungere uno strato di privacy a queste richieste, esse vengono fatte passare attraverso i server Mycroft: Google non può ricollegare una richiesta all'utente che l'ha fatta. In questo modo, tutte le query effettuate a Google contengono solamente l'audio e hanno lo stesso mittente. In futuro sarà possibile sfruttare il dataset di **Mozilla Deepspeech**, che non è, ad oggi, ancora sufficientemente affidabile.

Interpretazione degli intent

Nell'ambito della speech recognition, un intent è l'operazione che l'utente *intende* svolgere. Un utente può richiedere un'operazione in modi molto diversi. Lo scopo dell'intent parser è proprio quello di riuscire a superare questo scoglio, estraendo dalla richiesta scritta gli elementi chiave. Per esempio, se avessimo una richiesta del tipo "Hey Mycroft, domani pioverà a Parma?", l'intent parser dovrebbe capire:

- L'utente desidera conoscere il **tempo atmosferico**
- Il tempo atmosferico deve essere cercato per **Parma**
- La data di interesse è **domani**

Mycroft rende disponibili due software per la rilevazione di intenti: Padatious e Adapt. Mentre il secondo è basato sul riconoscimento di parole chiave, ed è quindi molto suscettibile alle richieste, il primo è composto da una rete neurale addestrata su frasi intere. Nel corso del progetto sfrutteremo **Padatious**. Questo motore permette di avere creazione semplice degli intent, quantità di dati relativamente bassa, estrazione delle *entities* semplice, addestramento rapido.

Text To Speech

Il componente di Text To Speech ha il compito di generare una traccia audio partendo dal testo di risposta codificato nella skill. Mycroft rende disponibile il progetto **Mimic**, basato sul software FLITE dell'università Carnegie Mellon. Purtroppo quest'ultimo non è disponibile in italiano, quindi nel corso del progetto utilizzeremo il motore TTS di Google, ottimizzato per l'italiano. Come sempre, le richieste passano attraverso i server Mycroft per ottenere un buon livello di privacy. I testi di risposta sono codificati nei file vocabolario delle skills.

3.2.2 Interfaccia grafica

I dispositivi più avanzati di Mycroft come il *Mark II* forniscono la possibilità di aggiungere interazioni visive con gli utenti. L'interfaccia è gestita da Mycroft-GUI, un componente aggiuntivo del bot che sfrutta i KDE Plasmoids, oggetti dell'interfaccia grafica KDE Plasma, simili a widget. La tecnologia è basata sul linguaggio QML, che permette una totale libertà nella creazione di interfacce, ma anche dei template standard semplici e rapidi da implementare. QML fa parte dello stack di tecnologie di Qt, una libreria multiplatforma per lo sviluppo di programmi con interfaccia grafica.

Capitolo 4

Realizzazione di skills

Mycroft permette di espandere le abilità del bot in maniera semplice: è infatti costruito modularmente, ed è in ogni momento possibile espandere o ridurre i moduli attivi. Questi moduli sono detti **skills**. Ogni skill è responsabile di una funzionalità: avremo ad esempio la skill del meteo, la skill che racconta barzellette o la skill che si occupa della domotica. Alla ricezione di una richiesta, Mycroft verifica gli intent e le keywords di ogni skill, attivando quella giusta. La creazione di nuove skills è possibile grazie alla classe python `MycroftSkill`, tramite la cui espansione possiamo definire i comportamenti della skill. All'installazione di Mycroft verrà reso disponibile il **Mycroft Skills Kit**, una command line interface che permette di eseguire le operazioni fondamentali collegate alle skills Mycroft. Tramite il comando `mycroft-msk create` verrà avviata una CLI che, dopo aver richiesto al programmatore alcune informazioni, genera il *boilerplate code* necessario alla realizzazione della skill. Tra le informazioni richieste avremo nome, frasi d'esempio, dialoghi di risposta, descrizioni. . . Il template viene inoltre fornito di `git`, un sistema di versioning che permette di tenere traccia delle modifiche al codice e semplificare l'interazione tra programmatori.

4.1 Struttura dei file della skill

La struttura delle skills è ben organizzata, e abbiamo diverse tipologie di file. Oltre a licenza e README, abbiamo tre componenti fondamentali per il funzionamento.

Directory locale Questa directory contiene i file dipendenti dalla lingua. Le sottocartelle seguono i tag linguistici IETF, quindi l'italiano sarà rappresentato da `it-it` e l'inglese (americano) da `en-us`. All'interno di queste, possiamo incontrare quattro tipologie di file:

- **dialog**: questi file contengono i dialoghi di risposta del bot. Sono presenti più frasi simili, in modo da avere interazioni più "umane" con il bot, che sceglie una frase casuale tra queste. Ad esempio, `get_age.dialog` conterrà le frasi *"Mi potresti dire quanti anni hai?"*, *"Per favore, dimmi la tua età"*, *"Qual è la tua età?"*...

- **intent**: questi file contengono le frasi utilizzate per addestrare la rete neurale di Padatious. Ovviamente, più esempi vengono fatti, meglio la rete funzionerà. Per esempio, il file `symptoms.bleeding.intent` conterrà frasi come *"Ho un'emorragia", "Sto sanguinando molto", "Ho una grande perdita di sangue"...*
- **voc**: questi file contengono keywords utilizzate in Adapt e nelle domande a risposta chiusa. Siccome il progetto si basa fortemente su Padatious, troviamo solo i file vocabolario per le domande sì/no ed il riconoscimento di numeri. Ad esempio, il file `yes.voc` conterrà *"Sì", "Certo", "Esatto"...*
- **entity**: questi file contengono entità da estrarre negli intent. Ad esempio, l'intent *fracture* estrae l'arto coinvolto nella frattura. Il file `limb.entity` conterrà quindi termini come *"gamba", "pollice", "spalla"...*

settingsmeta.yaml Questo file contiene la definizione di impostazioni modificabili dall'utente sul sito di gestione di Mycroft. Ad esempio, l'utente potrebbe voler impostare i comportamenti di default, il nome del file esportato dal bot, una chiave API per l'integrazione con altri tool.

__init__.py Questo è il file principale della skill. Contiene il codice che definisce i comportamenti della skill. Il funzionamento è basato sull'espansione della classe `MycroftSkill`, che contiene i decorator ed i metodi necessari alla gestione della skill. Il prossimo capitolo si occuperà di definire la realizzazione vera e propria.

4.2 Funzionamento della skill

Tutta la realizzazione della skill è articolata intorno alla classe `MycroftSkill`.

```
"""Mycroft skill that does a pre-triage on hospital patients.

The skill tries to ask the patient its symptoms, its personal data,
and more. Then, it assigns a color code, stating a priority for
medical interventions.
"""

from mycroft import MycroftSkill, intent_file_handler
import json

class HospitalTriage(MycroftSkill):
    """Main skill class for the triage.

    This is the main skill class (extending MycroftSkill),
    which contains all the operations we need to perform the
    triage.

    Attributes:
        med_record: a dict containing all the patient data.
    """
```

```
def __init__(self):
    MycroftSkill.__init__(self)
    self.med_record = {}
```

4.2.1 Metodi relativi alle sintomatologie

La skill è stata strutturata come segue: ogni tipologia di sintomi equivale ad un intent, ed è quindi gestita da un metodo specifico. In base alla tipologia di sintomo, l'interazione viene poi gestita in modalità diverse: verrà assegnato al paziente un codice di triage, viene confermata al paziente la corretta comprensione del sintomo, viene mostrata sulla GUI l'informazione recepita. terminate le operazioni specifiche del sintomo, le restanti procedure vengono gestite da decorators: la richiesta dell'anagrafica, dei sintomi minori, della febbre vengono eseguite indipendentemente dal problema riportato dal paziente. Un esempio di metodo relativo alle sintomatologie può essere il seguente:

```
# BREATH
@intent_file_handler('symptoms.breath.intent')
@symptom_handler
@ covid_symptom
def handle_breathing(self, message):
    """This function handles a "breathing fatigue" symptom.

    Breathing fatigue is recognized as a red code, and is a
    COVID-compatible symptom.

    Args:
        message: the message object returned from Mycroft

    GUI: show open mouth emoji
    """
    self.med_record["main_symptom"] = "breathing"
    self.med_record["code"] = "red"
    self.speak_dialog('symptoms.breath')
```

Il metodo segna nell'oggetto *scheda clinica* il sintomo principale, ossia problemi di respirazione, ed il codice relativo, ossia il rosso (accesso immediato alle cure). Poi, tramite il metodo `speak_dialog` di `MycroftSkill`, pronuncia uno a caso tra i dialoghi contenuti nel file `symptoms.breath.dialog`, come

- *"Capisco, qualche problema di respirazione."*
- *"Ok, segno nella scheda problemi di respirazione."*
- *"Capito: problemi di respirazione."*

Da qui in poi, le operazioni sono affidate ai metodi chiamati dai decorators.

4.2.2 Decorators utilizzati

Che cos'è un decorator?

Anzitutto, occorre definire il comportamento di un decorator. Esso permette di agganciare responsabilità aggiuntive ad un oggetto, dinamicamente. Fornisce quindi un'alternativa flessibile al *subclassing*. Definire un decorator in Python significa definire dei comportamenti da eseguire prima e dopo l'esecuzione del metodo vero e proprio.

Applicazione del concetto al programma

Alcune operazioni sono, all'interno del progetto, ricorrenti. Ad esempio, la richiesta di anagrafica o di sintomi minori deve essere fatta in tutti e soli i metodi relativi alla gestione di sintomi. Alcuni di questi (quelli compatibili con il COVID19, come la tosse), avranno bisogno inoltre di approfondimenti per valutare la possibilità che il paziente sia affetto da COVID19, facendogli domande specifiche. Sono stati quindi definiti due decorators aggiuntivi:

symptom_handler Questo metodo, dopo aver gestito le operazioni specifiche di un sintomo, prosegue con le procedure generali, come la richiesta dell'anagrafica, di una valutazione del dolore, di altri sintomi, e, per ultima, l'esportazione della cartella clinica.

```
def symptom_handler(handler):  
    """Decorates a symptom with the needed operations.  
  
    This function is used as a decorator for symptoms, adding  
    operations like personal data asking, age, other symptoms...  
  
    Returns:  
        The decorator function  
    """  
    def ask_about_symptoms(*args, **kwargs):  
        returned = handler(*args, **kwargs)  
        args[0].med_record["symptom_declaration"] = args[1].data["utterance"]  
        # I'm using args[0] here instead of self, but it works the same  
        args[0].request_age()  
        args[0].request_other_symptoms()  
        args[0].evaluate_pain()  
        args[0].request_name()  
        args[0].export_med_record()  
        return returned  
    return ask_about_symptoms
```

covid_symptom La situazione di emergenza che il mondo sta vivendo al momento della stesura di questa tesi ha richiesto particolare attenzione verso le sintomatologie compatibili con COVID19: distanziare i pazienti che potrebbero essere portatori di virus al più presto è la sfida più grande del momento. Per questo, il decorator, aggiunto ai soli sintomi compatibili con la malattia, permette di interrogare il paziente riguardo alcune caratteristiche tipiche della malattia, come il fiato corto, la difficoltà a percepire i sapori, la febbre alta.


```
def covid_symptom(handler):  
    """Decorates a COVID-compatible symptom.  
  
    This function is used as a decorator in the COVID-compatible  
    symptoms. It proceeds to ask the COVID-related questions  
    to the patient.  
  
    Returns:  
        The decorator function  
    """  
    def check_if_covid(*args, **kwargs):  
        returned = handler(*args, **kwargs)  
        args[0].ask_covid_questions()  
        return returned  
    return check_if_covid
```

Il metodo `ask_covid_questions` procede ad intervistare il paziente con il modello di diagnosi più usato al momento. Ogni domanda posta al paziente ha un suo *score* caratteristico, che moltiplica un indice relativo al paziente in base alla gravità ed alla connessione con la malattia. Le caratteristiche indagate sono:

- **Febbre:** moltiplicatore 2
- **Mal di gola:** moltiplicatore 1.3
- **Raffreddore:** moltiplicatore 1.3
- **Fatica respiratoria:** moltiplicatore 1.6
- **Tosse:** moltiplicatore 1.6
- **Contatti con infetti:** moltiplicatore 2
- **Mancanza di gusto:** moltiplicatore 1.7

Ogni paziente sospetto ha inizialmente un `covid_score` di 1, che viene moltiplicato per i vari indici se il paziente risponde in modo affermativo. Se lo score supera una soglia predeterminata (al momento fissata a 15), è sospetto COVID19 e verrà destinato ad un'area di triage apposita.

```
def ask_covid_questions(self):  
    """Checks for COVID symptoms.  
  
    When triggered by a COVID-compatible symptom,  
    this function evaluates the patient symptoms to  
    try to guess if he/she has COVID19.  
  
    GUI: show face mask emoji  
    """  
    self.speak_dialog('gotta_check_covid')  
    covid_score = 1  
    # Let's check if the patient knows the temperature. Skip if he already declared it.
```

```
if not "fever" in self.med_record:
    self.check_fever()
if "fever" in self.med_record:
    if self.med_record["fever"] > 37.5:
        covid_score = covid_score * 2
# Let's define an array of tuples, each containing the yes/no question
# string and its COVID index multiplier
yesno_questions = [("has_sore_throat", 1.3), ("has_cold", 1.3),
                    ("has_breathing_difficulties", 1.6), ("has_cough", 1.6),
                    ("has_had_contacts", 2), ("misses_taste", 1.7)]
self.speak_dialog('will_ask_yesno')
# Check if he/she has COVID-compatible symptoms
for question in yesno_questions:
    self.med_record[question[0]] = self.ask_yesno(question[0])
    if self.med_record[question[0]] == 'yes':
        covid_score = covid_score * question[1]
    self.log.info(covid_score)

self.med_record["covid_score"] = covid_score
if covid_score > 15:
    self.speak_dialog('probably_has_covid')
else:
    self.speak_dialog('doesnt_have_covid')
self.log.info(self.med_record)
```

4.2.3 Helper methods

Sono, oltre ai metodi caratteristici della classe, presenti alcuni *helper methods* utili ad operazioni interne del software. Di seguito elenchiamo i più importanti:

extract_temperature Questo metodo, data una stringa in ingresso rappresentante la temperatura corporea che può essere in vari formati (ad esempio, il paziente potrebbe pronunciare *"Trentasette e mezzo"*, *"Trentasette virgola cinque"*, *"Trentasette punto cinque"*, *"Trentasette e cinque"* indicando sempre la stessa temperatura), la estrae in formato decimale.

```
def extract_temperature(utterance):
    """Extracts the patient temperature from the utterance.

    This is needed because of the various ways of Mycroft interpreting
    floating point numbers. Some examples:
    - 38 e 1
    - 38/1
    - 38.1
    - 38,1
    - 38 1

    Returns:
        The floating point value of the temperature, or
        None if it is impossible to extract.
```

```
"""
# Beware: the ' e ' has to be before the simple space!
possible_separators = ['/', '.', ',', ' e ', ' ']
try:
    for separator in possible_separators:
        if separator in utterance:
            temperature_strings = utterance.split(separator)
            if temperature_strings[1] == "mezzo":
                temperature_strings[1] = "5"
            temperature = int(
                temperature_strings[0])+float(temperature_strings[1])*0.1
            return temperature
    return None
except TypeError:
    return None
```

age_validator Questo metodo, fornito al metodo `get_response`, definisce se un'età è valida. In caso negativo, viene chiesta nuovamente.

number_validator Questo metodo verifica se l'input vocale è numerico, distinguendo anche il caso in cui un *sei* venga interpretato come *verbo essere* e non 6.

4.2.4 Oggetto `med_record`

Tutti i dati della procedura di triage confluiscono in un oggetto `med_record`, in cui vengono salvati per l'inserimento nel database ospedaliero. Questo oggetto verrà poi esportato in formato JSON (JavaScript Object Notation) con la possibilità di essere inviato ad un'API REST del sistema esistente.

```
def export_med_record(self):
    """Exports the data to JSON.

    This function is called at the end of the interaction
    to export the fetched data from the patient. It then
    assigns a desk to the patient based on his/her severeness.

    GUI: show hands emoji
    """
    with open("med_record.json", "w") as med_record_file:
        med_record_file.write(json.dumps(self.med_record))
    self.speak_dialog('thanks_and_bye', {"desk": self.med_record["code"]})
    self.med_record = {}
```

Questo formato standard permette di essere integrato semplicemente con infrastrutture preesistenti. Per esempio, un paziente affetto da COVID19 potrebbe avere la seguente scheda clinica:

```
{
    "can_talk": "yes",
    "main_symptom": "breathing",
```

```
"code": "red",  
"fever": 38.7,  
"has_sore_throat": "yes",  
"has_cold": "no",  
"has_breathing_difficulties": "yes",  
"has_cough": "yes",  
"has_had_contacts": "yes",  
"misses_taste": "yes",  
"covid_score": 22.6304,  
"symptom_declaration": "ho il fiatone",  
"age": 21,  
"other_symptoms": "ho un po' di mal di testa",  
"pain_index": "8",  
"full_name": "marco rossi"  
}
```

Capitolo 5

Possibilità di inference tramite reti neurali

Le moderne tecnologie di intelligenza artificiale espandono continuamente le possibilità di nuove applicazioni nell'interazione uomo-macchina. Questo capitolo si occupa di studiare le tecnologie necessarie ad attuare un processo di classificazione dei pazienti in base a conversazioni avute con un bot, ma anche con un medico. L'obiettivo è semplice: data in input la trascrizione di una conversazione, vogliamo ottenere in output il reparto al quale il paziente andrà assegnato.

5.1 Classificazione del testo

Esiste, nell'intelligenza artificiale, una branca che si occupa di **Natural Language Processing**, ossia processo del linguaggio naturale. Vorremmo quindi trovare un *sistema* che approssimi al meglio la corrispondenza tra un input testuale (che sarà necessario *tradurre* in termini matematici) ed un output che indichi una specifica classe. Il processo sarà quindi diviso in più parti: per primo, dovremo trovare un dataset testuale già categorizzato, o categorizzarne manualmente uno. Fatto ciò, dovremo trovare un modo di esprimere questo dataset in termini matematici, gli unici *comprensibili* da un computer. Dopodiché, ci occuperemo di trovare un algoritmo matematico che approssimi le corrispondenze.

5.1.1 Dataset di trascrizioni

È fondamentale trovare un dataset con cardinalità notevole e poco rumore. Online sono disponibili molti progetti di dataset medici, ma le stringenti leggi sulla privacy impongono grandi limitazioni sui suddetti. Sfruttando però siti di **sample di trascrizioni mediche**, come *MTSamples*, è possibile ottenere delle trascrizioni curate e verosimili, categorizzate in base al reparto di appartenenza. Trovando un modo di automatizzare l'operazione di download dei suddetti, ad esempio attraverso **Selenium**, riusciremo ad ottenere un dataset valido, verosimile, e con poco rumore, essendo utilizzato da professionisti medici per spunti e studi. Effettuando lo scraping da *MTSamples* è possibile ottenere un dataset contenente poco meno di 5000 records, cifra bassa ma sufficiente per sperimentare le tecniche di ML.

Fatto ciò, bisognerà però escludere dal set tutte quelle trascrizioni che non apportano valore alla nostra ricerca: ad esempio, non saremo interessati agli elettrocardiogrammi. Un modo rapido per filtrare i dati è quello di tenere le sole righe della tabella la cui trascrizione contiene il termine *HISTORY*, ossia le patologie preesistenti nel paziente. I record si riducono così a poco meno di 2000. Un record di esempio potrebbe essere così costruito:

description:	Consult for laparoscopic gastric bypass.
medical_specialty:	Bariatrics
transcription:	PAST MEDICAL HISTORY:, He has difficulty climbing stairs, difficulty with airline seats, tying shoes, used to public seating, and lifting objects off the floor. He exercises three times a week at home and does cardio. He has difficulty walking two blocks or five flights of stairs. Difficulty with snoring. He has muscle and joint pains includingknee pain, back pain, foot and ankle pain, and swelling. He has gastroesophageal reflux disease.,PAST SURGICAL HISTORY:, Includes reconstructive surgery on his right hand 13 years ago. ,SOCIAL HISTORY:, He is currently single. He has about ten drinks a year. He had smoked significantly up until several months ago. He now smokes less than three cigarettes a day.,FAMILY HISTORY:, Heart disease in both grandfathers, grandmother with stroke, and a grandmother with diabetes. Denies obesity and hypertension in other family members.,CURRENT MEDICATIONS:, None.,ALLERGIES:, He is allergic to Penicillin.,MISCELLANEOUS/EATING HISTORY:, He has been going to support groups for seven months with Lynn Holmberg in Greenwich and he is from Eastchester, New York and he feels that we are the appropriate program. He had a poor experience with the Greenwich program. Eating history, he is not an emotional eater. Does not like sweets. He likes big portions and carbohydrates. He likes chicken and not steak. He currently weighs 312 pounds. Ideal body weight would be 170 pounds. He is 142 pounds overweight. If ,he lost 60 of his excess body weight that would be 84 pounds and he should weigh about 228.,PHYSICAL EXAMINATION:, He is alert and oriented x 3. Cranial nerves II-XII are intact. Afebrile. Vital Signs are stable.
keywords:	bariatrics, laparoscopic gastric bypass, weight loss programs, gastric bypass, atkin's diet, weight watcher's, body weight, laparoscopic gastric, weight loss, pounds, months, weight, laparoscopic, band, loss, diets, overweight, lost

5.1.2 Inference sulle trascrizioni

Per prima cosa, potremmo tentare di addestrare la nostra rete neurale sulla trascrizione testuale del paziente, per prevedere il reparto. Queste trascrizioni sono lunghe ed eterogenee,

quindi non ci aspettiamo un'accuratezza troppo alta.

Preparazione del dataset

Dopo aver importato il dataset in formato CSV, lo inseriremo in un dataset **pandas** e terremo solo i record la cui trascrizione contiene *HISTORY*:

```
transcriptions = pd.read_csv("/content/drive/My Drive/Dataset/mtsamples.csv")
histories = transcriptions[transcriptions.transcription.str.contains("HISTORY", na=False)]
```

Dopodiché, sarà necessario ottenere delle label numeriche dalle classi testuali riportate nei samples. Sfrutteremo quindi la libreria **preprocessing** di *SciKit Learn*, più precisamente la classe **LabelEncoder**:

```
le = preprocessing.LabelEncoder()
histories["medical_specialty"] = histories.apply(lambda x : x["medical_specialty"].strip(), axis=0)
classes = le.fit(histories["medical_specialty"])
histories["class"] = histories.apply(lambda x : le.transform([x["medical_specialty"]]), axis='columns')
```

Stampando la dimensione di `le.classes_` contenuta nell'attributo **size**, scopriamo che il dataset ha 38 classi diverse. Alcune righe della tabella contengono dei NaN, valori nulli che possono causare errori. Per questo, vanno rimossi. Inoltre, è utile rimuovere le stopwords, ossia quelle parole che non apportano valore significativo al testo, come *che, io, lui, il, la...* Per fare ciò, sfruttiamo la libreria **gensim.parsing.preprocessing**.

```
histories = histories.dropna()
histories["transcription"] = histories.apply(lambda x : remove_stopwords(x["transcription"]), axis=0)
y = histories.apply(lambda x : x["class"][0], axis = 'columns')
y = y.values
```

Vettorizzazione del dataset

Abbiamo ora un dataset testuale, è necessario esprimerlo in termini numerici per poter addestrare la rete. Per questo, procediamo alla **vettorizzazione del testo**, ossia la trasposizione di un testo in forma vettoriale. Esistono varie soluzioni a questo problema, come la creazione di *sparse arrays*, ossia vettori contenenti un grande numero di 0, e 1 nel caso una parola sia presente. Se ad esempio avessimo un dataset di due frasi, *"mi piacciono le reti neurali"* e *"non mi piacciono i serpenti"*, la rappresentazione della prima frase sarebbe la seguente: Questo, però, porta ad un problema: ci sono parole molto più significative,

non	piacciono	mi	i	delfini	neurali	reti	le
0	1	1	0	0	1	1	1

come "reti" o "delfini", rispetto a "non" o "le". Un primo passo verso la risoluzione di questo problema è la rimozione di queste stopwords. Ma esiste un'idea interessante che ci permette di migliorare le performance: più una parola è presente nel dataset, meno sarà significativa. Vogliamo per questo dare più importanza a quelle parole che compaiono raramente nel testo. TF-IDF serve esattamente a questo. Definiamo la quantità $tf_{i,j}$ come il rapporto tra

il numero di occorrenze del termine i nel documento j ed il numero di termini totali del documento:

$$tf_{i,j} = \frac{n_{i,j}}{|d_j|}$$

Anche il fattore IDF è utile a determinare l'importanza di un termine: rappresenta infatti l'importanza generale del termine i nella collezione. Lo troviamo calcolando il logaritmo della divisione tra il numero di documenti j nella collezione ed il numero di documenti che contengono il termine i :

$$idf_i = \log \frac{|D|}{|\{d : i \in d\}|}$$

Infine,

$$(tf - idf)_{i,j} = tf_{i,j} \times idf_i$$

La libreria *SciKit Learn* fornisce un metodo che accorpa la vettorizzazione in sparse array (CountVectorizer) e l'applicazione al suddetto della funzione TF-IDF per regolarne i pesi:

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(histories["transcription"].to_numpy())
```

Ora, possiamo dividere il dataset in *training* e *test*, in modo da poter verificare il corretto addestramento della rete:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)
```

Costruzione della rete neurale

Per costruire la rete neurale, sfruttiamo la libreria **Keras** fornita da **TensorFlow**. Questa libreria permette di definire reti neurali con molta semplicità. Sappiamo che il vocabolario in input è composto da 15875 termini, e le classi in output sono 38. Inseriamo strati di dimensione decrescente, tutti con attivazione ReLU e dropout di 0.3. L'ultimo strato avrà attivazione softmax: quest'ultima è necessaria per la task di classificazione, permettendo un output di probabilità di appartenenza alle classi, che avrà somma 1.

```
from keras.models import Sequential
from keras.layers import Dense, Dropout
from keras.wrappers.scikit_learn import KerasClassifier
from keras.utils import np_utils
def build_model_keras():
    model = Sequential()
    model.add(Dense(256, input_dim=15875, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(200, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(200, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(150, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(100, activation='relu'))
```



```

model.add(Dropout(0.3))
model.add(Dense(80, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(60, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(38, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
model.summary()
return model

```

```

estimator = KerasClassifier(build_fn=build_model_keras, epochs=20, batch_size=128)
estimator.fit(X_train, y_train)

```

Ottenendo, quindi:

Layer (type)	Output Shape	Param #
dense_65 (Dense)	(None, 256)	4064256
dropout_56 (Dropout)	(None, 256)	0
dense_66 (Dense)	(None, 200)	51400
dropout_57 (Dropout)	(None, 200)	0
dense_67 (Dense)	(None, 200)	40200
dropout_58 (Dropout)	(None, 200)	0
dense_68 (Dense)	(None, 150)	30150
dropout_59 (Dropout)	(None, 150)	0
dense_69 (Dense)	(None, 100)	15100
dropout_60 (Dropout)	(None, 100)	0
dense_70 (Dense)	(None, 80)	8080
dropout_61 (Dropout)	(None, 80)	0
dense_71 (Dense)	(None, 60)	4860
dropout_62 (Dropout)	(None, 60)	0
dense_72 (Dense)	(None, 38)	2318

Total params: 4,216,364
 Trainable params: 4,216,364
 Non-trainable params: 0

Infine, verifichiamo l'accuracy sui dati di test.

```
from sklearn import metrics
print ("Predict on test data ... ")
y_pred = estimator.predict(X_test)
y_pred_converted = le.inverse_transform(y_test)
predictions = np.column_stack((y_pred, y_test))
print("Accuracy score: "+ str(metrics.accuracy_score(y_test, y_pred)))
```

L'accuracy sui dati di test è risultata intorno al 30%, non sufficiente per analisi affidabili. È risultato inoltre molto evidente come, non essendo questo dataset particolarmente grande, il rischio di overtraining sia decisamente verosimile. Con overtraining intendiamo un fenomeno per il quale l'addestramento sui dati di training sembra dare un'accuratezza alta, ma quando calcolata sui dati di test risulta pessima. Questo accade perché il modello si adatta troppo ai dati di training, risultando poco generalizzabile. Potremmo però pensare di utilizzare, al posto delle trascrizioni, le keywords fornite nel dataset.

5.1.3 Inference sulle keywords

Si potrebbe quindi pensare di effettuare l'inference non più sul testo della conversazione, ma solo su alcune parole chiave estrapolate dal suddetto. Questa pratica, in linea teorica, riduce nettamente il rumore sul dataset generato da tutti quei termini, nel testo, che non apportano informazioni sulla sintomatologia.

Preparazione dei dati e del modello

Procedendo con le stesse tecnologie di cui sopra, vettorizziamo il testo tramite `TfidfVectorizer`, e manteniamo una rete neurale costruita con Keras, dalla forma simile:

```
def build_model_keras():
    model = Sequential()
    model.add(Dense(256, input_dim=5814, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(200, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(200, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(150, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(100, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(80, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(60, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(39, activation='softmax'))
    model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    model.summary()
    return model
```

Ottenendo, quindi:

Layer (type)	Output Shape	Param #
dense_94 (Dense)	(None, 256)	1488640
dropout_82 (Dropout)	(None, 256)	0
dense_95 (Dense)	(None, 200)	51400
dropout_83 (Dropout)	(None, 200)	0
dense_96 (Dense)	(None, 200)	40200
dropout_84 (Dropout)	(None, 200)	0
dense_97 (Dense)	(None, 150)	30150
dropout_85 (Dropout)	(None, 150)	0
dense_98 (Dense)	(None, 100)	15100
dropout_86 (Dropout)	(None, 100)	0
dense_99 (Dense)	(None, 80)	8080
dropout_87 (Dropout)	(None, 80)	0
dense_100 (Dense)	(None, 60)	4860
dropout_88 (Dropout)	(None, 60)	0
dense_101 (Dense)	(None, 39)	2379
Total params: 1,640,809		
Trainable params: 1,640,809		
Non-trainable params: 0		

Questa volta, avremo in input 5814 features (equivalente alla cardinalità dell'insieme dei termini), ed in output 39 classi.

Risultati

Questa volta le prestazioni sono decisamente più alte: con la configurazione di cui sopra, l'accuratezza sul test set è di più del 75%. Questo risultato apre le porte a tantissime possibilità tecnologiche per i prossimi anni. Oltre alla creazione di assistenti vocali e chatbot, anche l'autodiagnosi di patologie sarebbe più semplice e gioverebbe al sistema sanitario, spesso in situazioni di sovraccarico. Per esempio, servizi di triage online per la comprensione dei sintomi e delle patologie sono già possibili. Con l'evoluzione della visione digitale, e l'aumento dei dispositivi wearable, con la possibilità di tenere sempre sotto controllo i parametri vitali, in futuro la coesione tra medicina ed informatica sarà sempre più forte.

Appendice A

An appendix

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text

will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliografia

- [1] Geoffrey I. Webb (Editor) Claude Sammut (Editor). *Encyclopedia of Machine Learning and Data Mining*. 2^a ed. Springer, 2017. ISBN: 148997685X, 978-1489976857.
- [2] Piers Ford. «Digital triage changes the game for UK GP practices in the age of COVID-19». In: (mag. 2020). URL: <https://www.mobihealthnews.com/news/europe/digital-triage-changes-game-uk-gp-practices-age-covid-19>.
- [3] Matthew Hoy. «Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants». In: *Medical Reference Services Quarterly* 37 (gen. 2018), pp. 81–88. DOI: [10.1080/02763869.2018.1404391](https://doi.org/10.1080/02763869.2018.1404391).
- [4] T.M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997, pp. 81–88. ISBN: 9780071154673. URL: <https://books.google.it/books?id=EoYBngEACAAJ>.
- [5] Voicebot - Orbita. «Voice Assistant Consumer Adoption in Healthcare». In: (ott. 2019). URL: https://voicebot.ai/wp-content/uploads/2019/10/voice_assistant_consumer_adoption_in_healthcare_report_voicebot.pdf.