

Esame di Ingegneria della Conoscenza

Presentazione Caso di Studio

Università degli Studi di Bari «Aldo Moro»

Dipartimento di Informatica

A.A 2020/2021

Docente: Nicola Fanizzi

Gruppo di Lavoro:

Luca Lobascio, matricola 676614

Link utili

Repository Progetto:

https://github.com/Il2909/ICon_mental_diagnosis

Documentazione:

https://github.com/Il2909/ICon_mental_diagnosis/tree/main/doc/build/html (HTML)

https://github.com/Il2909/ICon_mental_diagnosis/blob/main/doc/doc_final.pdf (PDF)

Introduzione al caso di studio

Il progetto di questo caso di studio si focalizza sulla diagnosi dei disturbi mentali, tema piuttosto attuale nell'ultimo periodo, per via delle numerose restrizioni, tra cui il lockdown, a causa della pandemia del virus Sars-CoV-2, iniziata nel 2020 e tutt'ora in corso, seppur con misure di restrizione ridotte, allo stato attuale.

Il lockdown ha decisamente accentuato le problematiche legate ai disturbi mentali, poiché l'essere umano, per via del bisogno quasi naturale di socializzare, non è abituato a mantenere distanze sociali, o ad essere isolati per tanto tempo, così come altre forme di restrizione sociale.

Spesso, quando si parla di qualcuno che soffre di un disturbo mentale, il pensiero comune lo indirizza subito alla depressione, nonostante ci siano numerose diagnosi di diverse categorie, e in certi casi anche con sintomi e criteri apparentemente simili fra loro.

Il progetto lo si potrebbe definire come un *proof of concept*, con l'obiettivo di dimostrare che è possibile realizzare un sistema che sfrutta l'Intelligenza Artificiale per assistere gli utenti, che siano esperti o meno in materia, ad individuare i vari disturbi mentali e ad avere informazioni su di esso.

Descrizione generale del sistema

Per questo caso di studio è stato implementato, in linguaggio Python, un **sistema esperto** che, in base ai sintomi/criteri che l'utente inserisce tramite interfaccia a riga di comando (CLI), restituisce una lista delle possibili diagnosi (con i rispettivi codici identificativi standard ICD-10-CM, sezione F) e la probabilità correlate ad esse.

Successivamente l'utente potrà effettuare una ricerca delle diagnosi, inserendo in input il codice della diagnosi che si vuole cercare, e il sistema fornirà in output le informazioni della diagnosi, con eventuali termini di inclusione/esclusione e le varie sottodiagnosi.

Composizione del sistema

Il sistema è composto da quattro moduli, ognuno di loro responsabile di una specifica funzionalità del sistema, descritte nelle slides successive:

- BBN.py
- database.py
- xml_parser.py
- main.py

Il modulo BBN.py

Il modulo è responsabile del calcolo della probabilità che l'utente abbia una certa diagnosi, così come le probabilità della categoria di una diagnosi condizionata da quella dei sintomi, ottenute tramite la costruzione di una **Bayesian Belief Network**, sfruttando la libreria [*pomegranate*](#).

Inoltre sono presenti funzioni utili per il preprocessing dei dati, elencate di seguito:

- Calcolo della percentuale delle categorie dei sintomi fornite in input dagli utenti
- Calcolo della probabilità pesata
- Calcolo della probabilità totale.

Il modulo database.py

Questo modulo sfrutta la libreria [pysql](#) per il collegamento ad un database server, al fine di effettuare operazioni (query) all'interno di esso.

Il modulo implementa funzioni per:

- Creazione e utilizzo dei database
- Creazione tabelle e popolazione tramite lettura di file CSV
- Verifica dell'esistenza di un dato all'interno di una tabella
- Interrogazione (Select) di specifiche tabelle
- (Right) Join tra due tabelle

Il modulo xml_parser.py

Il modulo ha il compito di leggere un file XML e di effettuare il parsing dei dati all'interno del file.

All'interno vengono implementate due funzioni:

- La prima legge le diagnosi (con i rispettivi codici) all'interno del file XML e successivamente li scrive su un file CSV.
- La seconda è per la ricerca di tutte le informazioni inerenti una specifica diagnosi (termini di inclusione/esclusione e sottodiagnosi), quando l'utente fornisce in input il codice della diagnosi da ricercare.

Il modulo main.py

Il modulo principale contiene la funzione per l'inserimento dei dati e richiama le funzioni delle altre librerie per calcolare e stampare a video il risultato finale.

All'interno vengono implementate ulteriori funzioni di utilità, come il conteggio delle occorrenze, l'unione di più liste e la tabulazione dei dati.

Dataset usati

I dataset utilizzati all'interno del progetto sono quattro:

1. **Diagnosis.csv**: documento contenente gli ID (ICD) delle diagnosi, con i rispettivi nomi e categorie. Creato tramite parsing del file `icd-f-section.xml`.
2. **Criteria.csv**: documento contenente gli ID dei sintomi/criteri, con i rispettivi nomi e categorie. Il dataset è stato creato manualmente, tramite l'ausilio del DSM-5, manuale di riferimento per gli studi psichiatrici.
3. **CritToDiag.csv**: documento contenente gli ID dei sintomi, correlati agli ID delle diagnosi. Il dataset è stato creato manualmente, anche qui facendo riferimento al manuale precedentemente menzionato.
4. **icd-f-section.xml**: base di conoscenza contenente tutti le diagnosi, con le rispettive sottodiagnosi, termini di inclusione/esclusione. Utilizzato per ricavare il dataset **Diagnosis.csv** e per la funzionalità di ricerca. Il dataset è ottenibile dal sito dell'ente americano Centers for Disease Control and Prevention (<https://www.cdc.gov/nchs/icd/icd10cm.htm>).

Ulteriori note sui dataset utilizzati

Sia le diagnosi che i sintomi sono suddivise in dieci categorie, etichettate con le lettere dell'alfabeto. Ogni categoria ha un range di diagnosi correlate ad esso.

- A (F01-F09): Disturbi dovuti a condizioni psicologiche conosciute
- B (F10-F19): Disturbi e atteggiamenti dovuti all'uso di sostanze psicoattive
- C (F20-F29): Disturbo schizofrenico, schizotipico, delirante e altri disturbi psicotici non umorali
- D (F30-F39): Disturbi dell'umore (affettivi)
- E (F40-F48): Disturbo d'ansia, dissociativo, stress, somatoforme e altri disturbi non psicotici
- F (F50-F59): Sindromi comportamentali associati a disturbi psicologici e a fattori fisici
- G (F60-F69): Disturbi della personalità e del comportamento nell'età adulta
- H (F70-F79): Disabilità intellettuali
- I (F80-F89): Disturbi evolutivi specifici e pervasivi
- J (F90-F98): Disturbi emotivi e comportamentali ricorrenti durante l'infanzia ed adolescenza

Esempio di esecuzione del sistema

```
Database mentaldisorder exists.
Insert at least one symptom. Press Enter when you're done.
impulsivity
temper outburst
negative
worthlessness

code          name probability
0 f90  attention-deficit hyperactivity disorders  28.9039%
1 f32                depressive episode  43.4893%
2 f20                schizophrenia  27.6068%
```

Schermata 1: Input dei sintomi e output dei risultati

```
If you would like to have more information about a specific disorder, type the ICD code (Fxx),
otherwise press enter to terminate.
Input: f32

F32 : Depressive episode

Includes:
single episode of agitated depression
single episode of depressive reaction
single episode of major depression
single episode of psychogenic depression
single episode of reactive depression
single episode of vital depression
Excludes (1):
bipolar disorder (F31.-)
manic episode (F30.-)
recurrent depressive disorder (F33.-)
Excludes (2):
adjustment disorder (F43.2)

Subdiagnosis:

F32.0 : Major depressive disorder, single episode, mild
F32.1 : Major depressive disorder, single episode, moderate
F32.2 : Major depressive disorder, single episode, severe without psychotic features
F32.3 : Major depressive disorder, single episode, severe with psychotic features
Inclusion Term:
Single episode of major depression with mood-congruent psychotic symptoms
Single episode of major depression with mood-incongruent psychotic symptoms
Single episode of major depression with psychotic symptoms
Single episode of psychogenic depressive psychosis
```

Schermata 2: Ricerca delle informazioni su una specifica diagnosi

Grazie per l'attenzione

