

# Identifying and Classifying Symptoms within a Patient's Statement

Nicola Sartorato

Università di Trento

23 settembre 2019



# Il macro-obiettivo

## Il macro-obiettivo

Sviluppare un agente conversazionale capace di condurre l'**anamnesi** sostituendosi al medico.

# Il macro-obiettivo

## Il macro-obiettivo

Sviluppare un agente conversazionale capace di condurre l'**anamnesi** sostituendosi al medico.

## L'anamnesi

È la raccolta dalla voce diretta del paziente di tutte le informazioni utili al medico per effettuare una diagnosi.

# Il problema e le motivazioni

I medici, infatti:

- svolgono il loro lavoro sotto stringenti **limitazioni di tempo**

# Il problema e le motivazioni

I medici, infatti:

- svolgono il loro lavoro sotto stringenti **limitazioni di tempo**
- sono occupati principalmente da **lavori di scrivania** (49.2% del loro tempo) e solo il 27% del loro tempo sono a contatto con i pazienti

# Il problema e le motivazioni

I medici, infatti:

- svolgono il loro lavoro sotto stringenti **limitazioni di tempo**
- sono occupati principalmente da **lavori di scrivania** (49.2% del loro tempo) e solo il 27% del loro tempo sono a contatto con i pazienti
- l'attività per loro più intensa è l'**anamnesi**, la quale è:
  - ❶ **fondamentale**

# Il problema e le motivazioni

I medici, infatti:

- svolgono il loro lavoro sotto stringenti **limitazioni di tempo**
- sono occupati principalmente da **lavori di scrivania** (49.2% del loro tempo) e solo il 27% del loro tempo sono a contatto con i pazienti
- l'attività per loro più intensa è l'**anamnesi**, la quale è:
  - 1 **fondamentale**
  - 2 ma a volte **incompleta**

# A piccoli passi

## Il primo ostacolo

**Identificare** e **classificare** i sintomi all'interno di una frase detta da un paziente.



# A piccoli passi

## Il primo ostacolo

**Identificare** e **classificare** i sintomi all'interno di una frase detta da un paziente.

## Identificare un sintomo

Individuare all'interno di una frase ogni gruppo di parole adiacenti che, insieme, vogliono dire un sintomo.

# A piccoli passi

## Il primo ostacolo

**Identificare** e **classificare** i sintomi all'interno di una frase detta da un paziente.

## Identificare un sintomo

Individuare all'interno di una frase ogni gruppo di parole adiacenti che, insieme, vogliono dire un sintomo.

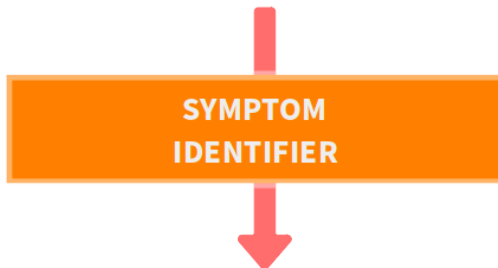
## Classificare un sintomo

Mappare i sintomi precedentemente identificati in sintomi specifici che fanno riferimento ad una classificazione.

# Metodi

## Metodi: identificare un sintomo

*"I started to have dizziness and sore throat 3 days ago."*



Group 1  
Group 2

*"I started to have dizziness and sore throat 3 days ago."*

Il **Symptom Identifier** riceve in input la frase del paziente e restituisce i sintomi identificati (i *tokens for predictions*).

Il **Symptom Identifier** è composto da tre componenti:

Il **Symptom Identifier** è composto da tre componenti:

- 1 dal **Body Part Finder** che evidenzia le parti del corpo all'interno della frase

Il **Symptom Identifier** è composto da tre componenti:

- 1 dal **Body Part Finder** che evidenzia le parti del corpo all'interno della frase
- 2 dal **Question Answering System** che si occupa di estrarre, usando Machine Reading Comprehension, il problema generico del paziente e quelli delle parti del corpo presenti nella frase. Lo fa sfruttando uno di questi due modelli:

Il **Symptom Identifier** è composto da tre componenti:

- 1 dal **Body Part Finder** che evidenzia le parti del corpo all'interno della frase
- 2 dal **Question Answering System** che si occupa di estrarre, usando Machine Reading Comprehension, il problema generico del paziente e quelli delle parti del corpo presenti nella frase. Lo fa sfruttando uno di questi due modelli:
  - BERT (Devlin, 2018)



Il **Symptom Identifier** è composto da tre componenti:

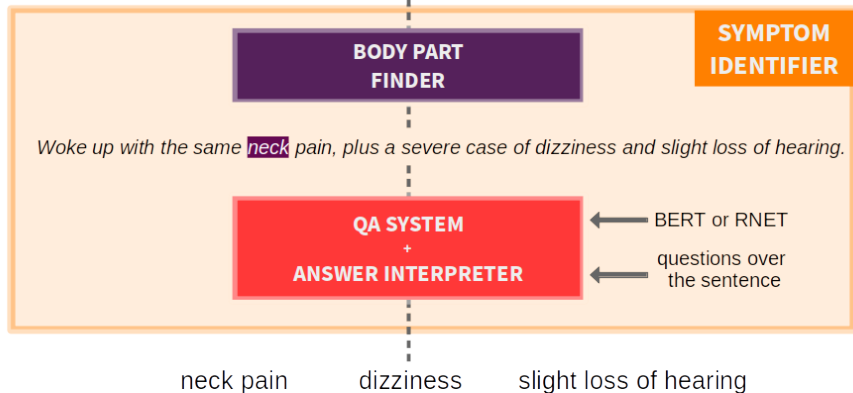
- 1 dal **Body Part Finder** che evidenzia le parti del corpo all'interno della frase
- 2 dal **Question Answering System** che si occupa di estrarre, usando Machine Reading Comprehension, il problema generico del paziente e quelli delle parti del corpo presenti nella frase. Lo fa sfruttando uno di questi due modelli:
  - BERT (Devlin, 2018)
  - R-NET (Wang, 2017)

Il **Symptom Identifier** è composto da tre componenti:

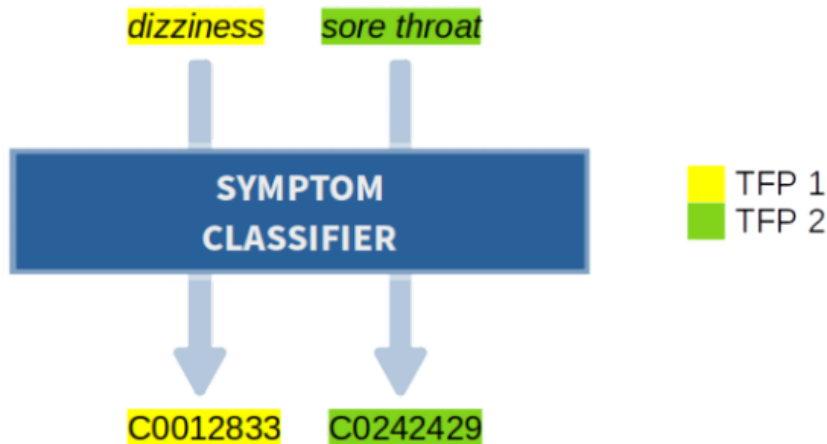
- ❶ dal **Body Part Finder** che evidenzia le parti del corpo all'interno della frase
- ❷ dal **Question Answering System** che si occupa di estrarre, usando Machine Reading Comprehension, il problema generico del paziente e quelli delle parti del corpo presenti nella frase. Lo fa sfruttando uno di questi due modelli:
  - BERT (Devlin, 2018)
  - R-NET (Wang, 2017)
- ❸ dall'**Answer Interpreter** che converte le risposte estratte dal QA System in *tokens for predictions*; essi saranno input del Symptom Classifier

# Metodi: il Symptom Identifier

*Woke up with the same neck pain, plus a severe case of dizziness and slight loss of hearing.*



## Metodi: classificare un sintomo



Il **Symptom Classifier** riceve in input i *tokens for predictions* e restituisce i codici univoci dei sintomi identificati (i CUIs).

# La classificazione di sintomi

La classificazione di sintomi utilizzata è MEDCIN:

# La classificazione di sintomi

La classificazione di sintomi utilizzata è MEDCIN:

- si trova all'interno dei dizionari supportati da UMLS (Unified Medical Language System), che fornisce un CUI (Concept Unique Identifier) per ogni suo termine

# La classificazione di sintomi

La classificazione di sintomi utilizzata è MEDCIN:

- si trova all'interno dei dizionari supportati da UMLS (Unified Medical Language System), che fornisce un CUI (Concept Unique Identifier) per ogni suo termine
- ha un'organizzazione gerarchica (i.e. ad albero)

# Metodi: il Symptom Classifier

Il Symptom Classifier è composto da due componenti:



# Metodi: il Symptom Classifier

Il Symptom Classifier è composto da due componenti:

- 1 il **Vectorifier**, che ha il compito di vettorizzare i tokens for predictions in due possibili tipi di embeddings:

# Metodi: il Symptom Classifier

Il Symptom Classifier è composto da due componenti:

- ① il **Vectorifier**, che ha il compito di vettorizzare i tokens for predictions in due possibili tipi di embeddings:
  - ① GloVe embeddings (Pennington, 2014)

# Metodi: il Symptom Classifier

Il Symptom Classifier è composto da due componenti:

- 1 il **Vectorifier**, che ha il compito di vettorizzare i tokens for predictions in due possibili tipi di embeddings:
  - 1 GloVe embeddings (Pennington, 2014)
  - 2 BERT embeddings (Devlin, 2018)

# Metodi: il Symptom Classifier

Il Symptom Classifier è composto da due componenti:

- 1 il **Vectorifier**, che ha il compito di vettorizzare i tokens for predictions in due possibili tipi di embeddings:
  - 1 GloVe embeddings (Pennington, 2014)
  - 2 BERT embeddings (Devlin, 2018)
- 2 il **Symptom Tree Mapper**, il cui scopo è ritornare, dato un *token for prediction*, il CUI del sintomo predetto. Ogni sintomo appartenente all'albero ha un suo rappresentante (la media dei vettori delle parole nel nome del sintomo).

# Metodi: il Symptom Classifier

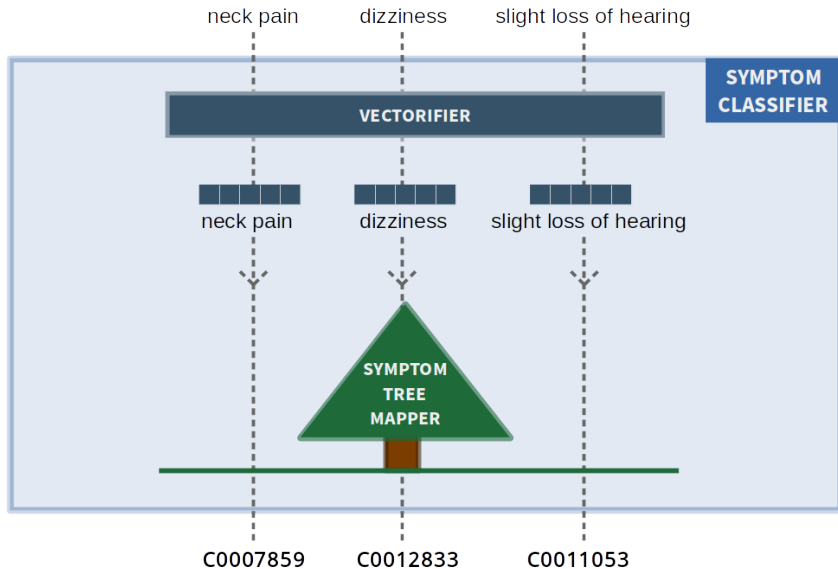
Il Symptom Classifier è composto da due componenti:

- 1 il **Vectorifier**, che ha il compito di vettorizzare i tokens for predictions in due possibili tipi di embeddings:
  - 1 GloVe embeddings (Pennington, 2014)
  - 2 BERT embeddings (Devlin, 2018)
- 2 il **Symptom Tree Mapper**, il cui scopo è ritornare, dato un *token for prediction*, il CUI del sintomo predetto. Ogni sintomo appartenente all'albero ha un suo rappresentante (la media dei vettori delle parole nel nome del sintomo).

## Criterio di predizione

Viene calcolata la cosine similarity tra ogni *token for prediction* vettorizzato e il rappresentante di ogni sintomo appartenente all'albero. Il sintomo con maggiore similarità verrà predetto.

# Metodi: il Symptom Classifier



Al momento, non ci sono dataset utilizzabili per la classificazione di sintomi. Si è rivelato necessario:

- estrarre 336 frasi dai post di un forum medico ([doctorslounge.com](http://doctorslounge.com))
- taggare i sintomi presenti nelle frasi

# Risultati



# Risultati: valutare il Symptom Identifier

Per valutare il Symptom Identifier i *tokens for predictions* sono stati classificati in:

# Risultati: valutare il Symptom Identifier

Per valutare il Symptom Identifier i *tokens for predictions* sono stati classificati in:

- 1 correct tokens

# Risultati: valutare il Symptom Identifier

Per valutare il Symptom Identifier i *tokens for predictions* sono stati classificati in:

- 1 correct tokens
- 2 redundant tokens

# Risultati: valutare il Symptom Identifier

Per valutare il Symptom Identifier i *tokens for predictions* sono stati classificati in:

- 1 correct tokens
- 2 redundant tokens
- 3 non-sense tokens

# Risultati: valutare il Symptom Identifier

Per valutare il Symptom Identifier i *tokens for predictions* sono stati classificati in:

- 1 correct tokens
- 2 redundant tokens
- 3 non-sense tokens
- 4 wrong tokens

# Risultati: valutare il Symptom Identifier

$$\text{score} = \# \text{correct} - \# \text{non sense} - \# \text{wrong} - \# \text{missing} \quad (1)$$

# of	BERT	R-NET
correct	423	427
redundant	275	195
non sense	122	247
wrong	8	14
missing	61	74
score	232	92

**Tabella:** Confronto dello score tra BERT e R-NET come calcolato in (1).

# Risultati: valutare il Symptom Classifier

Per valutare il Symptom Classifier le predizioni sono state divise in:

# Risultati: valutare il Symptom Classifier

Per valutare il Symptom Classifier le predizioni sono state divise in:

- `correct`, se la predizione si trova in uno dei sottoalberi di un sintomo presente nella frase; a loro volta divise in:



Per valutare il Symptom Classifier le predizioni sono state divise in:

- **correct**, se la predizione si trova in uno dei sottoalberi di un sintomo presente nella frase; a loro volta divise in:
  - **non-redundant**, se il sottoalbero in questione non è stato marcato come già associato

Per valutare il Symptom Classifier le predizioni sono state divise in:

- **correct**, se la predizione si trova in uno dei sottoalberi di un sintomo presente nella frase; a loro volta divise in:
  - **non-redundant**, se il sottoalbero in questione non è stato marcato come già associato
  - **redundant**, se invece è già stato marcato come associato

Per valutare il Symptom Classifier le predizioni sono state divise in:

- **correct**, se la predizione si trova in uno dei sottoalberi di un sintomo presente nella frase; a loro volta divise in:
  - **non-redundant**, se il sottoalbero in questione non è stato marcato come già associato
  - **redundant**, se invece è già stato marcato come associato
- **wrong**, se la predizione non si trova in nessun sottoalbero

# Risultati: valutare il Symptom Classifier

Assieme al Symptom Classifier sono state valutate le seguenti opzioni:

# Risultati: valutare il Symptom Classifier

Assieme al Symptom Classifier sono state valutate le seguenti opzioni:

- diversi tipi di embeddings

# Risultati: valutare il Symptom Classifier

Assieme al Symptom Classifier sono state valutate le seguenti opzioni:

- diversi tipi di embeddings
- se usare o meno il Body Part Finder

# Risultati: valutare il Symptom Classifier

Assieme al Symptom Classifier sono state valutate le seguenti opzioni:

- diversi tipi di embeddings
- se usare o meno il Body Part Finder
- se cercare o meno il sintomo solo nei sottoalberi relativi alle parti del corpo trovate nella frase (l'opzione “pruning”)

# Risultati: valutare il Symptom Classifier

Assieme al Symptom Classifier sono state valutate le seguenti opzioni:

- diversi tipi di embeddings
- se usare o meno il Body Part Finder
- se cercare o meno il sintomo solo nei sottoalberi relativi alle parti del corpo trovate nella frase (l'opzione “pruning”)
- diversi *minimum similarity thresholds*: il valore minimo di similarità affinché una predizione sia ritenuta valida

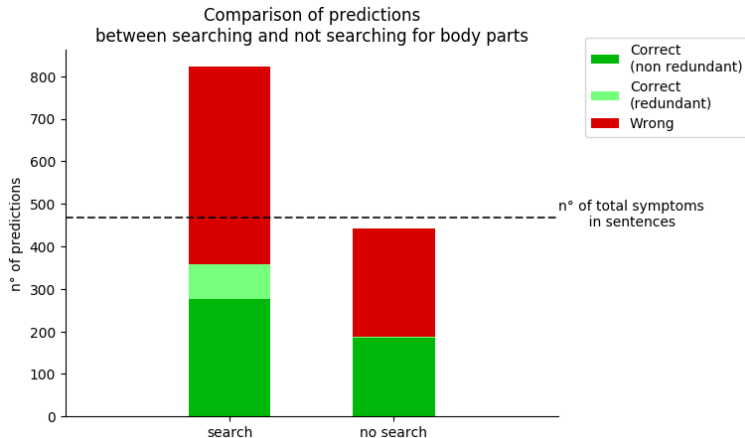


# Risultati: valutare il Symptom Classifier

embedding type	accuracy	correct predictions	harmonic mean	missed symptoms
GloVe 50	55.5 %	40.1 %	46.5 %	208
GloVe 100	55.2 %	39.1 %	45.8 %	209
GloVe 200	59.1 %	43.7 %	50.2 %	191
GloVe 300	59.3 %	42.2 %	49.3 %	190
BERT emb.	46.7 %	32.9 %	38.6 %	249

**Tabella:** Confronto tra diversi tipi di embeddings usando indici diversi. La media armonica tra accuracy e correct predictions indica che GloVe 200 e GloVe 300 hanno performance migliori degli altri tipi.

# Risultati: valutare il Symptom Classifier



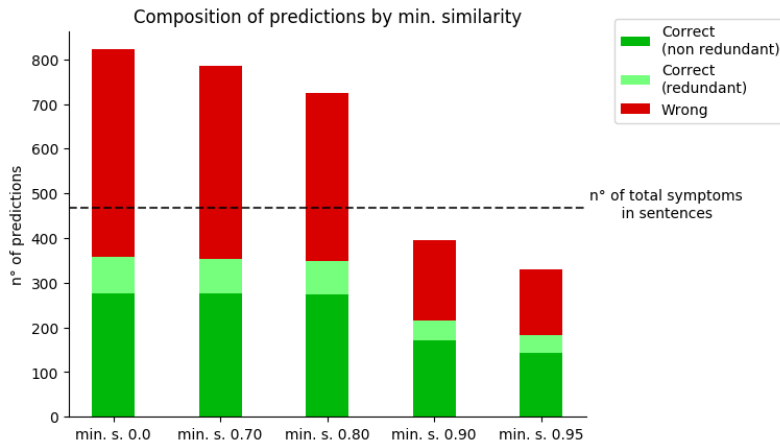
**Figura:** Confronto della composizione delle predizioni quando il Body Part Finder è attivo (search) e quando non lo è (no search).

# Risultati: valutare il Symptom Classifier

	accuracy	correct predictions	missed symptoms
pruning	59.1 %	43.7 %	191
no pruning	43.7 %	36.0 %	263

**Tabella:** Confronto fra indici quando l'opzione "pruning" è attiva (pruning) e quando non lo è (no pruning). L'introduzione di questa opzione provoca un miglioramento del 15.4% nell'accuracy e del 7.7% nelle predizioni corrette.

# Risultati: valutare il Symptom Classifier



**Figura:** Confronto della composizione delle predizioni usando diversi *minimum similarity thresholds* (0.00, 0.70, 0.80, 0.90, 0.95).

- I risultati sono complessivamente positivi perché, anche usando usando una rappresentazione semplificata, il modello raggiunge:
  - il 58.7% di accuracy
  - il 48.0% di predizioni corrette
  - facendo 1.5 medium attempts

- I risultati sono complessivamente positivi perché, anche usando una rappresentazione semplificata, il modello raggiunge:
  - il 58.7% di accuracy
  - il 48.0% di predizioni corrette
  - facendo 1.5 medium attempts
- Il fatto che i risultati dipendano molto dalle rappresentazioni dei sintomi indica che c'è molto margine di miglioramento.

- I risultati sono complessivamente positivi perché, anche usando una rappresentazione semplificata, il modello raggiunge:
  - il 58.7% di accuracy
  - il 48.0% di predizioni corrette
  - facendo 1.5 medium attempts
- Il fatto che i risultati dipendano molto dalle rappresentazioni dei sintomi indica che c'è molto margine di miglioramento.
- Il Symptom Classifier è la prima componente da migliorare: infatti, si passa dall'84.8% di *tokens for predictions* corretti al 48.0% di previsioni corrette.

## Conclusioni: future work

Questa componente può essere usata come base per una futura che riesca a focalizzare le domande su sintomi correlati a quelli già scoperti:



Questa componente può essere usata come base per una futura che riesca a focalizzare le domande su sintomi correlati a quelli già scoperti:

- usare `cui2vec` (Beam, 2018) come rappresentazioni per i sintomi e algoritmi di clustering per esplorare gruppi di sintomi correlati

*Grazie per l'attenzione*