

Validazione clinica in P.S.

1 Validazione accuratezza del sistema

La validazione dell'accuratezza del sistema ha lo scopo di produrre **valori di accuratezza**, secondo misure statistiche standard, per le codifiche ICD-9-CM proposte dal sistema. Queste misure ci permettono di valutare inequivocabilmente l'accuratezza del sistema e confrontare l'accuratezza del nostro sistema con quella di altri sistemi. Come?

Al fine di calcolare i valori di accuratezza bisognerà disporre, per ogni diagnosi considerata, del corretto insieme di codici ICD-9-CM, da confrontare con l'insieme dei codici proposti dal sistema.

2 Validazione funzionalità di apprendimento continuo

Questo processo ha lo scopo di validare la funzionalità "learn to rank" del sistema, ovvero la sua capacità di imparare con l'utilizzo. Si misura l'accuratezza del sistema prima di qualsiasi utilizzo e dopo un certo numero di utilizzi.

Il processo di validazione è composto da due fasi:

1. Validazione accuratezza del sistema allo stato attuale.
2. Validazione accuratezza del sistema dopo un numero di utilizzi.

La validazione fornirà una percentuale di miglioramento per le misure statistiche considerate. Questa validazione potrà essere portata avanti virtualmente (e istantaneamente) utilizzando la collezione descritta nella prossima sezione.

3 Costruzione della collezione di train/test

Al fine di valutare l'**accuratezza e l'apprendimento** del sistema è necessaria la costruzione di una collezione classificata: **una serie di diagnosi con i relativi codici ICD associati**.

Queste diagnosi codificate rappresenteranno per noi la **verità assoluta** con cui istruire e successivamente validare il nostro sistema.

3.1 Dimensione della collezione

Quante diagnosi bisogna considerare per la validazione?

Il numero di diagnosi dipende da

- Il numero totale di diagnosi a disposizione.
- Il numero di codici ICD che si vogliono considerare.

Per misurare le capacità di learning del sistema oltre che l'attuale accuratezza, la collezione sarà suddivisa in due insiemi distinti: uno per il learning e uno per la validazione. Per questo motivo ogni codice nella collezione dovrà essere presente almeno una volta in entrambi gli insiemi, ovvero **almeno due volte nella collezione**.

Da un'analisi quantitativa su una collezione di diagnosi, relativa ad un periodo di **4 mesi** nell'U.O. di Medicina d'Urgenza-Pronto Soccorso dell'Ospedale di Forlì, risultano:

- **13554 diagnosi totali** nella collezione
- 1463 i codici presenti nella collezione
- **852 i codici presenti almeno due volte**
- 611 i codici presenti solo una volta

Con una collezione di 4 mesi è pertanto possibile validare per il 6% dei codici ICD-9-CM.

Per ottenere una collezione che copra almeno il 15% dei codici ICD-9-CM (2100 codici su 14000), si stimano necessarie circa **40 mila diagnosi codificate**, ovvero il periodo di un anno di accessi in P.S. . Inoltre, scegliendo un periodo di un intero anno solare si eviterà di incorrere in bias indotti da malattie e incidenti con frequenza stagionale (es. eritema solare in agosto).

Un **numero maggiore** di diagnosi porterà ad **una validazione più estesa**: si potranno considerare metodi “cross-validation” k-fold. Ad esempio, se ogni codice considerato è ripetuto almeno 5 volte, potrà essere eseguita una cross-validation 5-fold.

Un **numero minore** di diagnosi porterà a **ridurre il numero di codici ICD-9-CM** considerati nella validazione. Mediante un campionamento stratificato sarà tuttavia possibile considerare una frazione della collezione, pur mantenendo la varietà di un intero anno solare.

3.2 Confronti

Si vuole confrontare l'accuratezza del sistema **con la codifica in reparto e con la verità assoluta**, pertanto si considerano 3 diversi livelli di codifica:

1. **Codifica in P.S.:** la collezione scelta dovrà già essere stata codificata durante i processi di Pronto Soccorso. Questa codifica ci permette di confrontare l'accuratezza del sistema con la codifica che avviene attualmente in ospedale.
2. **Codifica del sistema:** il sistema sarà lanciato sulla collezione ottenendo automaticamente la codifica per ogni diagnosi nella collezione.
3. **Codifica esperta:** la collezione dovrà essere codificata oggettivamente da 3 soggetti diversi. Questa codifica rappresenterà la verità assoluta.

Per calcolare l'accuratezza assoluta della codifica in P.S. e della codifica del sistema, queste saranno confrontate con la codifica esperta. In questo modo, oltre ad ottenere l'accuratezza del sistema in termini assoluti, si potrà confrontare l'accuratezza del sistema con l'accuratezza dell'attuale codifica in P.S.

3.3 Codifica esperta di una diagnosi

La codifica di una diagnosi deve rispettare le seguenti proprietà:

- **Codifica ICD oggettiva:** i codici ICD associati devono essere indipendenti dal giudizio soggettivo. Seguendo le modalità osservate in esperimenti simili, ogni diagnosi deve essere codificata **da 3 esperti diversi** per permettere un “**voto di maggioranza**”. Un codice viene associato a una diagnosi solo se proposto da 2 o più esperti.
- **Massima specificità:** il codice ICD scelto per una diagnosi deve essere il più specifico possibile. Bisogna sempre evitare codici di 3 cifre e codici generici qualora esistesse un codice più specifico.

4 Validazione miglioramento introdotto dall'utilizzo del sistema

Questo processo ha lo scopo di validare il miglioramento introdotto dal nostro sistema nella qualità di codifica di un soggetto (persona/reparto/ospedale)

Il processo si divide in due fasi:

1. Validazione accuratezza del soggetto in un periodo di n **utilizzi senza il nostro sistema.**
2. Validazione accuratezza del soggetto in un periodo di n utilizzi **con l'ausilio del nostro sistema.**

La validazione al punto 1. è rappresentata dalla validazione dell'accuratezza della Codifica in P.S. (confronto tra codifica in P.S. e codifica esperta)

Per la validazione al punto 2. invece è necessaria una ulteriore codifica esperta delle codifiche prodotte durante un periodo di utilizzo con il nostro sistema.

La validazione produrrà una percentuale di miglioramento dell'accuratezza del soggetto.