

Machine Learning per la Medicina: Classificazione Automatica di Referti Radiologici

Alfonso Emilio Gerevini¹, Alberto Lavelli², Alessandro Maffi¹, Roberto Maroldi^{1,3}, Luca Putelli¹,
Ivan Serina¹

¹Università degli Studi di Brescia, ²Fondazione Bruno Kessler, ³Spedali Civili di Brescia
{alfonso.gerevini, a.maffi007, roberto.maroldi, l.putelli002, ivan.serina}@unibs.it, lavelli@fbk.eu

Abstract

Centinaia di migliaia di referti medici vengono prodotti ogni giorno e contengono informazioni molto importanti riguardanti la salute dei pazienti e l'andamento dei trattamenti medici. Tuttavia, essendo espressi in testo non vincolato, non sono utilizzabili in modo efficace. Perciò l'utilizzo di tecniche automatiche per estrarre da queste fonti testuali conoscenza in forma strutturata può essere molto utile. In questo lavoro descriviamo un recente sistema per la classificazione automatica di referti radiologici, che è stato sperimentato sui referti di TAC al torace prodotti in uno dei reparti di radiologia degli Spedali Civili di Brescia. Il sistema utilizza tecniche di machine learning e uno schema di classificazione gerarchico proposto dai radiologi.

1 Introduzione

Negli ultimi anni i sistemi informatici hanno permesso di raccogliere una grande quantità di dati sotto forma di testo. Questo tipo di informazione presenta diversi problemi come l'assenza di struttura, la rappresentazione delle parole e le ambiguità tipiche del linguaggio. Le tecniche di Natural Language Processing e di Machine Learning possono essere di grande aiuto per estrarre conoscenza da queste fonti di informazioni.

Nel campo medico, in particolare, vengono prodotte centinaia di migliaia di documenti tra cui referti, cartelle cliniche o valutazioni di farmaci, da cui è possibile ricavare informazioni essenziali per scopi sia di ricerca, sia organizzativi e di valutazione della politica sanitaria in generale.

Presso l'Università degli Studi di Brescia è in corso da alcuni anni un progetto di ricerca multi disciplinare inerente la classificazione automatica dei referti prodotti dagli esami radiologici, che in una prima fase si è concentrato su un sistema per gli esami TAC al torace prodotti dagli Spedali Civili di Brescia. Nel contesto del progetto, i radiologi coinvolti hanno progettato uno schema gerarchico di classificazione, in cui particolare rilievo viene dato alla possibile presenza di neoplasie, ovvero di masse potenzialmente riconducibili a tumori maligni. Il sistema realizzato permette di processare un referto ottenendone la sua classificazione in tempo reale, durante la refertazione, che il medico medico può immediatamente validare o, eventualmente, modificare.

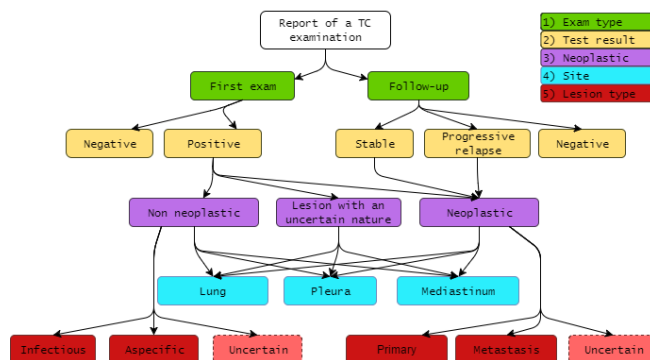


Figura 1: Gerarchia di classificazione

Questo primo sistema sviluppato nell'ambito del progetto è basato sull'annotazione delle parti rilevanti del referto. È stato realizzato un corpus di referti di cui i medici coinvolti hanno annotato le parti salienti. Tale corpus è stato utilizzato per creare, attraverso tecniche di NLP e machine learning, alcuni modelli per l'annotazione e la classificazione automatica del referto. Eventuali classificazioni multiple di differenti parti annotate nel testo sono combinate attraverso alcune regole elicitate da conoscenza medica fornita dai radiologi al fine di fornire una classificazione globale del referto.

2 Schema di classificazione

La Figura 1 mostra la gerarchia di classificazione che è stata progettata con i radiologi. E' composta da cinque livelli:

- tipo di esame: primo esame o follow-up;
- risultato dell'esame: negativo, positivo, oppure negativo, stabile o progressione recidiva;
- natura della lesione: non neoplastica, dubbia o neoplastica;
- sito della lesione: polmone, pleura o mediastino;
- tipo della lesione: infettiva, aspecifica o incerta, oppure primaria, metastasi o incerta.

All'interno dello schema, le frecce mostrano le relazioni gerarchiche della classificazione. Ad esempio, i referti che hanno come risultato dell'esame "stabile" o "progressione recidiva", possono avere come natura della lesione esclusivamente "neoplastica".

3 Ruolo delle annotazioni

Tipicamente in un referto i radiologi descrivono quanto è visibile nelle immagini ottenute dall'esame, identificando quali sono i rilievi riscontrati (ad es. la presenza di noduli, lesioni, adenopatie) e le loro caratteristiche. Queste porzioni di testo sono fondamentali perché forniscono evidenza empirica che giustifica la classificazione del referto (e fanno anche in modo che un altro medico possa capire le condizioni del paziente). Ad esempio, la presenza di "nodulo a margini irregolari e spiculati" nel referto è un forte indizio della presenza di una neoplasia. Analogamente, identificare queste "espressioni chiave" può aiutare nel processo di classificazione automatica. Per questo motivo, quando è stata costruito il training set per l'addestramento del classificatore, insieme alla classificazione in base allo schema descritto in precedenza, si è provveduto anche all'annotazione di queste porzioni di testo all'interno dei referti da usare per l'addestramento.

4 Classificazione automatica dei referti

I documenti vengono innanzitutto analizzati con TextPro (uno strumento per l'analisi testuale automatica) per effettuare: la separazione delle frasi, quella delle singole parole (token), l'analisi morfologica, l'estrazione di prefissi e suffissi e quindi del lemma, il riconoscimento delle parti del discorso e delle negazioni, l'eliminazione di numeri e caratteri speciali.

4.1 Annotazione automatica

Il riconoscimento delle porzioni di testo interessanti in un nuovo referto viene fatta attraverso un procedimento di annotazione automatica. Questo viene svolto con un sistema di apprendimento supervisionato in cui il training set è il corpus di documenti annotati manualmente dai radiologi coinvolti. Per questo scopo si è utilizzato l'algoritmo Conditional Random Fields, creando un modello di annotazione automatica per ogni livello della classificazione radiologica. Ognuno di questi modelli utilizza le feature estratte dalla fase di pre-processing (token, lemmi, parti del discorso) per riconoscere le porzioni di testo più simili a quelle annotate manualmente, isolando quindi le espressioni e le frasi più importanti relative a un particolare livello. Ad esempio, nella frase "lesione nodulare solida di 7mm nel segmento basale laterale del lobo inferiore destro", l'espressione "lesione nodulare solida" sarà attribuita al terzo livello (natura della lesione), "lobo inferiore destro" al quarto (sito della lesione).

4.2 Processo di classificazione

Per ogni livello della gerarchia di classificazione, le relative annotazioni automatiche vengono rappresentate come un vettore binario, tramite il modello bag-of-words. Questo vettore di feature binario viene quindi processato da un algoritmo di classificazione, addestrato utilizzando le annotazioni manuali dei radiologi, per predire il valore della classe di quel livello. Ad esempio, l'espressione "lesione nodulare solida" (livello 3), sarà classificata con il valore "neoplasia", mentre "lobo inferiore destro" (livello 4) sarà classificata come "polmone".

Alcune regole di aggregazione, derivate dalla metodologia utilizzata dai radiologi, definiscono il passaggio tra la classificazione delle singole annotazioni a quella della frase completa a cui appartengono. Per il terzo livello, ad esempio,

	RF		SVM		NN	
	Acc	FM	Acc	FM	Acc	FM
Exam type	95.6	94.9	95.6	94.9	95.6	94.9
Result	82.4	75.5	79.4	70.1	82.4	75.5
Result f-up	67.7	75.5	67.7	70.1	64.5	75.5
Neoplastic	75.0	57.0	77.1	66.0	70.8	56.3
Site lung	75.0	55.0	72.9	53.5	70.8	52.1
Site pleura	89.6	47.3	87.5	59.1	89.6	47.3
Site med.	72.9	67.9	72.9	67.9	66.7	63.6
L. t. neopl.	63.6	60.8	63.6	60.8	63.6	60.8
L.t.nonneopl.	60.0	57.8	60.0	57.8	60.0	57.8

Tabella 1: Valutazione del metodo proposto con tre algoritmi di classificazione: Random Forest (RF), Support Vector Machines (SVM), Reti Neurali (NN).

se una frase contiene un'annotazione di tipo "natura dubbia" e una "neoplasia", una regola impone la prevalenza del caso peggiore, per cui la frase viene classificata con il valore "neoplasia".

Ulteriori regole sono state definite anche per ottenere, a partire dalla classificazione delle singole frasi, i valori delle classi per l'intero referto. Ad esempio, se nel documento una frase è classificata per il secondo livello come "stabile" e un'altra come "progressione recidiva", il referto sarà classificato come "progressione recidiva".

5 Risultati

Nella Tabella 1 sono mostrati i risultati, per ogni livello della gerarchia, della classificazione dei referti con il metodo proposto e utilizzando tre diversi algoritmi di classificazione delle annotazioni: Random Forest (RF), Support Vector Machines (SVM) e Reti Neurali (NN).

Le performance sono state valutate attraverso due metriche: l'accuracy e la macro-averaged F-Measure, definita come la media armonica delle F-Measure delle singole classi.

I risultati complessivi migliori sono stati ottenuti utilizzando Random Forest, tuttavia anche SVM e le reti neurali producono risultati competitivi.

6 Discussione e sviluppi futuri

Abbiamo sviluppato un sistema per la classificazione automatica di referti di esami TAC al torace che attualmente è in fase di sperimentazione presso gli Spedali Civili di Brescia. La metodologia e le tecniche sviluppate sono estendibili per referti di altri esami di radiologia e di diversi ambiti biomedici. Un limite attuale del sistema è quello di basarsi sull'annotazione manuale dei referti per il training, un processo lungo, costoso e che necessita l'impiego di esperti. Per ovviare a questo problema, e quindi ridurre o eliminare il ruolo delle annotazioni, sono in studio e in implementazione tecniche basate su deep learning, sia per la rappresentazione delle parole nel referto e del loro significato, sia per l'utilizzo di reti neurali più complesse e sofisticate nella classificazione.

Riferimenti bibliografici

[Gerevini *et al.*, 2018] Alfonso E. Gerevini, A. Lavelli, A. Maffi, R. Maroldi, A.L. Minard, I. Serina, e G. Squassina. Automatic classification of radiological reports for clinical care. *Artificial intelligence in medicine*, 91:72–81, September 2018.