

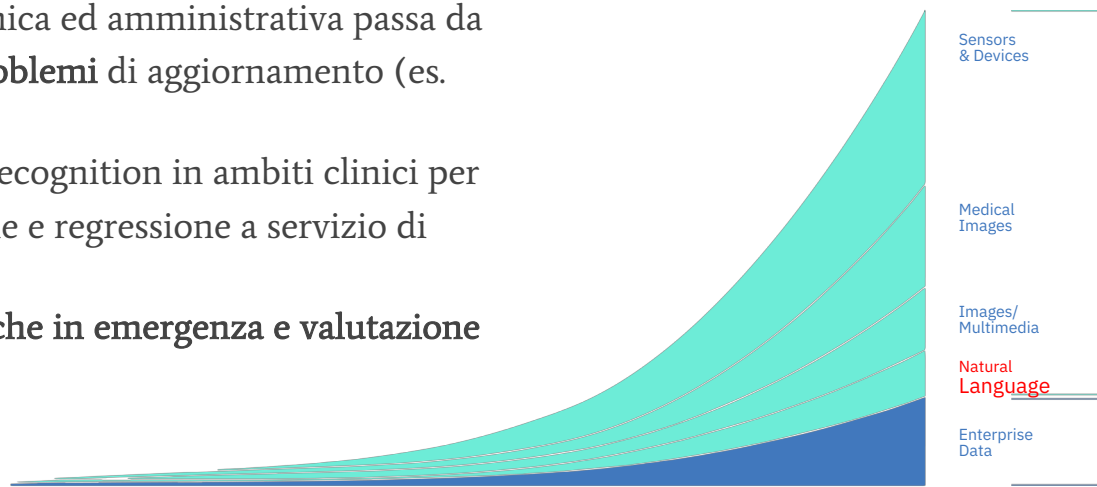
Natural Language Processing (NLP) nella ricerca e nell'assistenza medica: importanza, esperienze e visione per il futuro

Robert Alexander MD – IBM Italia

Idea progettuale

- Circa l'**80%** del volume di **dati** generati dall'assistenza sanitaria è di tipo **non strutturato** in larga misura costituito da **immagini e testo**.
- Il **testo libero** si ritrova in **prescrizioni, referti** (es radiologia, anatomia patologica), **storia clinica, esame obiettivo, cartella medica ed infermieristica, lettere di dimissione**. A differenza dalle immagini il testo ha **rilevanza solo nazionale**.
- Attualmente l'analisi scientifica, clinica ed amministrativa passa da **processi di codifica con i relativi problemi** di aggiornamento (es. ICD9), irreversibilità etc
- Costruire modelli di named entity recognition in ambiti clinici per supportare processi di classificazione e regressione a servizio di CDSS.
- **Supporto per i referti di TAC craniche in emergenza e valutazione rischio di morte.**

44 zettabytes



Tecniche di IA Utilizzate



- **Watson Knowledge Studio:** robust collaborative production ready PaaS offering
- Ambiente di gestione progettuale di sviluppo di **NER** pattern based e machine learning (RNN -> LSTM).
- Gestione preannotazione con modelli e/o dizionari (sviluppo ad hoc)
- Gestione team di annotazione di entità e relazioni (k score)
- Valutazione recursiva di statistiche di coverage ed accuratezza del modello (eg F1, ground truth coverage etc.)
- Export del modello in Watson Natural Language Understanding
- Sviluppo del modello di classificazione (SVM e random forests) per la previsione del rischio di esito infausto

Criticità per il settore

- Grave carenza di risorse per la lingua italiana in campo clinico
 - Es. No SNOMED-CT ITA Es. FMA
- Grande difficoltà accesso a corpora
 - GDPR, comitati etici, consenso informato, silos etc.
- Qualità delle pipeline OS (eg Spacy, NLTK etc)
 - Good for English, schifida per l'Italiano
- Necessità team multidisciplinari clinico/NLP



Visione per il futuro



- Accesso a pipeline stato dell'arte per l'Italiano
- Sviluppo collaborativo di corpora e risorse lessicali in campo clinic
- Coinvolgimento delle società scientifiche per lo sviluppo e la manutenzione di risorse e modelli.
- Definizione protocolli di validazione dei modelli "in vivo"
- Definizione dei modelli di erogazione dei modelli in produzione
- Individuazione delle proposizioni a maggior valore clinico-gestionale