

### 1) Come funziona il text-mining?

È un metodo di estrazione della conoscenza che trasforma un testo non strutturato della letteratura scientifica in un testo strutturato. La letteratura scientifica non tratta di innanzitutto devo annotare i termini che mi interessano, associando i sinonimi a tutti i termini. Poi devo elencare i termini che si riferiscono ai termini che sto cercando, perché il tipo di relazione che voglio è sintattica e non semantica....

Quando usare pathway analysis? Crescita degli adipociti in vari ambienti. Le ha raccontato l'esempio. 15

### 2) Come funziona la pathway analysis? Si fa per capire i funzionamenti di un particolare processo biologico. Si parte da una rete di interazione proteica su cui si applica l'algoritmo di flusso dell'interazione. In questo modo otteniamo una classifica dei percorsi. Come si identifica il percorso più attivo? Come funziona l'algoritmo di flusso? Individua il percorso più probabile (non era molto convinto della sua risposta rispetto a come si individua il percorso più attivo)

Esempio di dove utilizzeresti biomarkers identification? Diagnosticare autismo o vedere come funziona un vaccino. Oppure posso identificare quelli che sono i soggetti su cui un farmaco potrebbe funzionare. 14

### 3) Come funziona biomarkers identification? Valutare l'espressione di geni con uno studio di trascrittomico, ranking dei geni più e meno espressi, una volta fatto ciò si ha una firma dell'individuo (geni maggiormente e minormente espressi). Dopodiché si va a valutare la distanza tra gli individui, in base alle differenze nel ranking di espressione dei geni. In questo modo, rappresentando poi gli individui sul grafico vado ad individuare dei cluster.

Abbiamo un paziente con una patologia complessa multifattoriale, quale terapia potrebbe essere adatta a questo individuo? Se la malattia è già conosciuta posso ricavare dei dati di trascrittomico dall'individuo. Posso individuare i markers dell'individuo, la sua firma. Se la firma risulta molto simile a quella di altri pazienti che hanno avuto la malattia, posso applicare sul paziente la stessa terapia. 15

### 4) In quali casi è ragionevole utilizzare dei metodi di modelling & simulation? Bisogna considerare se ci troviamo nella condizione adatta in quanto questa tecnica richiede molto tempo. Se ci troviamo di fronte allo studio di un paziente, non è funzionale eseguire modeling & simulation, in quanto quando ci troveremo di fronte alla soluzione il paziente sarà già morto. Lo dobbiamo analizzare in prospettiva nel futuro e per poter aiutare pazienti del futuro.

Quando vuoi integrare dati di tipo diverso in uno studio scientifico cosa si fa? Posso integrare le reti trovate su data base con il text-mining, per poi individuare quelle che sono le proteine più o meno espresse, per poi eventualmente proseguire con la pathway analysis.

### 5) Come funziona modeling & simulation? Devo eseguire un'astrazione di tutte le informazioni che non ci servono nel modello. Questo perché passiamo da una rappresentazione statica a una dinamica, per cui cresce esponenzialmente la quantità di informazioni che si trovano nel modello. Modelling significa rappresentare con formalismi matematici il mio modello reale, in modo poi da poter fare una simulation sul modello, ovvero lasciar andare la variabile tempo e vedere quello che succede. Si tratta di un processo iterativo, non sono sicuro che funzioni subito non appena ho prodotto il modello. Devo fare una serie di verifiche e accorgimenti per vedere se effettivamente il modello può essere utilizzato per simulare il mio modello reale e poter fare simulazioni.

Come velocizzeresti il tempo necessario allo sviluppo di un farmaco? Tecnica di riposizionamento dei farmaci: prendo il modulo del farmaco da riposizionare, poi rappresento il modulo del farmaco e vado a vedere se vi si possono riposizionare tutti i moduli delle patologie rare. Se è possibile che il farmaco agisca sulla patologia rara, la CF dice che il farmaco è stato prodotto per quella patologia. In quel modo il farmaco viene approvato in modo più veloce. **IL FARMACO DEVE FUNZIONARE PER LA PATOLOGIA RARA!**

Perché il metodo rank based funziona meglio di altri metodi. Misuro tutti i valori, li ordino, elimino i dati centrali, prendo solo gli estremi e questi sono le firme dei miei pazienti, a questo punto posso calcolare la distanza e rappresentarlo sul grafo. Ma perché è meglio di altri metodi che non sono rank based? Non lo so (14)

6) perché il rank based è meglio di altri metodi? perché si ha un collegamento migliore tra marcatore e patologia e ciò consente di vedere meglio quali sono i nodi di riferimento che classificano un determinato paziente. Ma perché? Potrei fare un ordinamento dei geni e usare un metodo diverso, ma poter comunque visualizzare i pazienti su di un grafo.???

Esempio su cui useresti il text-mining? Esempio del metabolismo degli sfingolipidi, si comprendono i rapporti tra le proteine, estrapolando la conoscenza dalla letteratura. Si ottiene così un qualcosa di strutturato e siamo in grado di realizzare dei nodi costituiti da nodi e da archi.

Esempio su cui useresti basi di dati? Riposizionamento dei farmaci. Perché abbiamo già i grafi delle malattie, i moduli delle malattie e i moduli dei farmaci sui DB. Si prende il modulo farmaco, si eliminano le malattie per cui il farmaco è stato approvato e si fa un'intersezione tra il modulo del farmaco e quello delle altre malattie. In questo modo si individuano le malattie su cui il farmaco potrebbe avere effetto. 14

7) perché la pathway analysis è un passo essenziale per poi costruire modelli di simulazione? Questo perché per costruire un modello è necessaria l'astrazione. Il modello costruito non può possedere tutti i dettagli. Ci aiuta la pathway analysis perché con questa e con l'algoritmo di flusso, individuiamo quelli che sono i pathway più attivi. Spiega come funziona la pathway analysis. Il grafo che ottengo ha un'informazione topografica, su cui aggiungo un'informazione sperimentale e poi un'informazione riguardo quelle che sono le molecole più interessanti per il fenomeno che sto studiando.

Dobbiamo applicare tutta la pipeline come l'abbiamo descritta? No. L'abbiamo descritta così perché percorrendola vanno ad incrementare sia il tempo che i costi. Tutto dipende dalle condizioni in cui mi trovo. Se devo agire immediatamente su un paziente non devo arrivare a modelling, mentre se ho più tempo posso arrivare a questo step della pipeline. 15.

8) Sto sviluppando un farmaco per una patologia, ma osservo che su alcuni soggetti il farmaco funziona e su altri no. Come faccio a capire perché? Faccio un'analisi di trascrittomico e proteomico sul paziente su cui voglio applicare il farmaco e posso vedere a quale gruppo appartiene, in base ai suoi markers. (sbaglia perché sta parlando dal punto di vista del paziente, non è quello che gli ha chiesto) Come posso individuare sottoclassi di una patologia?

Casa è una rete multilivello? È una rete in cui inserisco diversi dati di diverso tipo, come trascrittomico, lipidomico e proteomico. In questo modo non visualizzo i cambiamenti di un solo tipo di dati, ma ottengo una visione più completa. (non so quanto sia giusta, mi sa che questa la dobbiamo cercare)

Cosa sono i nodi e gli archi? Archi sono le interazioni tra le singole componenti. Cosa significa se A è collegato a B e a C e a D? non ha risposto bene. 14.

9) come si individuano i percorsi più attivi con l'analisi di pathway? Partiamo da una rete che rappresenta l'universo dello studio che stiamo facendo. Su questa etichettiamo quelle proteine che risultano avere un'espressione differente nelle condizioni che stiamo analizzando. Etichettiamo poi quelle che sono le molecole più importanti per il tipo di studio che stiamo facendo. A questo punto analizziamo quelli che sono i percorsi di collegamento tra questi nodi. L'algoritmo di flusso farà quindi fluire un'informazione sugli archi. Otteniamo così un Nas, uno score per ogni percorso. Individuiamo così quello che è il percorso più attivi. A questo punto si fa un'analisi di arricchimento, visualizzando l'intersezione tra i processi biologici canonici e il pathway che ho individuato.

Esempio pratico in cui questo tipo di analisi è fondamentale? Ricavare i processi più attivi di una patologia, che sono quelli più disregolati, per poi vedere se vi si può utilizzare un farmaco.

Se hai un paziente in corsia con un problema. Quale delle tecniche della pipeline sono ragionevolmente applicabili? Estrazione della conoscenza da database, con text-mining, con cartelle cliniche di altri pazienti. Posso vedere quindi le similitudini di questo paziente con gli altri. Non mi sogno neanche di fare pathway analysis e modelling and simulation. 15.

10) Cosa sono nodi e archi? I nodi sono gli elementi che sto prendendo in relazione in quel momento, quindi ad esempio proteine, mentre i nodi rappresentano le interazioni tra questi elementi. Supponiamo di avere una proteina A e degli archi che la collegano a B, a C e a D. Cosa vuol dire? Si tratta delle varie interazioni che la proteina A può avere

In quale modo si può introdurre conoscenza umana all'interno di un approccio computazionale? Intesa come conoscenza del medico, che deve poter essere condivisa con l'informatico che costruisce il modello. Come funziona l'analisi dei pathway? In che modo posso influenzare prima l'analisi con la mia conoscenza umana?

Ho un farmaco che ho iniziato a sperimentare su modelli animali. Il farmaco funziona, ma la sua tossicità è troppo elevata per eseguire una sperimentazione sull'uomo. Non so perché sia tossico però? Che tecniche usi? 14.

1) Quando hai la pipeline ideale, devi utilizzare tutti gli step?

Riposizionamento dei farmaci come funziona?

2) Cosa mi vuoi dire? Pathway analysis

Esempio in cui utilizzeresti biomarkers identification 15. (Margheritaaa)

3) Modeling & Simulation?

In cosa consiste fare un'integrazione dei dati? Rappresentare su una rete multilivello le varie tipologie di informazioni che stiamo considerando.

Quando voglio fare uno studio devo usare tutte le tecniche della pipeline ideale?

4) pathway analysis? Aggiungiamo una conoscenza a quella precedentemente acquisita e diamo un'interpretazione funzionale al sistema biologico che stiamo studiando.

Vantaggio dei metodi rank based rispetto ad altri metodi per l'identificazione di biomarcatori? perché si basa solo sul ranking e non su quelli che sono realmente i miei valori di espressione genica. In questo modo si calcola la distanza tra due individui dello studio solo sulla base di questo ranking, per cui anche se le misurazioni vengono effettuate in laboratori diversi, con magari possibilità che si ottengano risultati leggermente diversi da quelli reali, per quanto riguarda l'espressione genica, i valori misurati non intaccano lo studio. Ci basiamo infatti solo sull'ordine dei geni espressi e non sugli specifici valori di espressione per effettuare il clustering. 15.

5) text-mining?

Cosa è la data-integration?