



Università degli Studi di Milano-Bicocca

Dipartimento di Informatica
Corso di Informatica - E3101Q
A.A. 2016-2019

Inferenza di alberi tumorali tramite Particle Swarm Optimization

Adrian David Castro Tenemaya, 816015

25 maggio 2019

Tutor: *Dott.* Ciccolella Simone

Relatore: *Prof.* Della Vedova Gianluca

Indice

0.1	Premessa e ringraziamenti	1
1	Introduzione	2
1.1	Descrizione	2
1.2	Storia	2
1.3	Stato dell'arte	3

Elenco delle figure

Elenco delle tabelle

0.1 Premessa e ringraziamenti

Il presente lavoro è frutto del lavoro svolto come tirocinio all'interno dell'Università di Milano-Bicocca, e viene anche utilizzato come tesi finale ai fini del conseguimento della laurea in Informatica. È però necessario chiarire che il progetto in questione non sarà abbandonato nè una volta terminata la stesura di questa relazione, nè dopo il conseguimento della laurea. È mia intenzione contribuire al meglio delle mie possibilità in quello che ritengo essere uno dei campi con il quale mi sento più legato, sia a livello di interesse professionale, che a livello strettamente personale: la ricerca sul cancro. Secondo il *National Cancer Institute*, nel 2012 sono stati riportati 14.1 *milioni* di nuovi casi, e di questi, 8.2 *milioni* hanno portato alla morte [1]. I dati mostrano anche quelli che può sembrare all'apparenza una realtà discordante: il numero totale di morti per cancro è in crescita, ma il rapporto delle morti per individuo sta calando [2]. Nel 1990, 161 persone su 100.000 nel mondo sono morte a causa del cancro. Nel 2016, questo numero è calato a 134 su 100.000. Questo miglioramento è dovuto indubbiamente ad un numero molto elevato di fattori, tra cui l'aumento della qualità di vita ed un migliore sistema sanitario, ma è anche grazie alla crescita incessante della ricerca sul cancro, ed ai campi sui quali essa si appoggia. Lo sviluppo di algoritmi sempre più efficienti e performanti, e l'utilizzo di calcolatori super-veloci, ha permesso a questo settore di ricerca di ottenere dei considerevoli risultati.

Con questo progetto spero, quindi, di aver dato un contributo in questo settore, anche se in una percentuale minuscola.

Vorrei ringraziare mia mamma **Laura**, mio padre **José**, mia sorella **Valeria**, i miei fantastici nonni, e tutte le bellissime e meravigliose persone che hanno contribuito, in maniera diretta ed indiretta, a farmi appassionare all'informatica e, in questo caso, alla bioinformatica.

Capitolo 1

Introduzione

1.1 Descrizione

Il presente lavoro è stato svolto sotto la guida ed il supporto di AlgoLab, laboratorio presso il dipartimento di informatica dell'Università di Milano-Bicocca, che ha lo scopo di progettare, studiare, analizzare ed implementare algoritmi efficienti per problemi computazionali. Il tirocinio è cominciato il 22 Marzo 2019. Il problema affrontato è l'*inferenza di progressioni tumorali* su dati single-cell, al fine di determinare l'ordine e la frequenza con cui le variazioni somatiche vengono acquisite durante una progressione tumorale. Spesso ciò è basato sulla "Infinite Sites Assumption", dove le mutazioni possono solo essere acquisite, e mai perse. Lo stage si colloca nella ricerca del superamento di tale assunzione, utilizzando il modello della *filogenesi persistente*, dove ogni mutazione può essere persa al massimo una volta nell'intero albero. Più precisamente, si è investigata la tecnica *Particle Swarm Optimization*. I dati single-cell sono caratterizzati da un elevato tasso di errore e di valori mancanti: ciò rende inutilizzabili gli approcci noti in letteratura per i dati di *bulk sequencing*. In particolare, sono state analizzate quali strutture dati utilizzare per rendere l'algoritmo efficiente ed efficace, e quali operazioni considerare per inferire predizioni accurate.

1.2 Storia

Era il 1869 quando venne isolato per la prima volta nella storia dell'umanità l'*Acido Desossiribonucleico*, anche conosciuto come *DNA*. Il pioniere di questa scoperta è Friedrich Miescher, medico e ricercatore nato in Svizzera nel 1844. Durante il processo di scoperta, Miescher aveva realizzato che nonostante avesse proprietà simili alle proteine, la nuova sostanza – il DNA – non lo era. Prima di isolare le cellule dal pus presente nelle bende chirurgiche dell'ospedale in cui lavorava, Miescher fu molto attento ad assicurarsi che il materiale che stava utilizzando fosse fresco e non contaminato. Fu solo più tardi, nel 1871, che il ricercatore iniziò a lavorare sullo sperma di salmone, una specie di pesce che affluiva numerosa durante il periodo autunnale nella città di Basel.

1.3 Stato dell'arte

Bibliografia

- [1] *Cancer Statistics*. URL: <https://www.cancer.gov/about-cancer/understanding/statistics>.
- [2] Max Roser e Hannah Ritchie. *Cancer*. Lug. 2015. URL: <https://ourworldindata.org/cancer>.