



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Informatica

APPLICAZIONI DELL'ALGORITMICA ALLA
BIOLOGIA: EVOLUTIONARY TREES E
CLUSTERING

APPLICATIONS OF ALGORITHMICS TO
BIOLOGY: EVOLUTIONARY TREES AND
CLUSTERING

MATTEO TORTOLI

MARIA CECILIA VERRI

Anno Accademico 2018-2019

CONTENTS

1	Capitolo 1: La bioinformatica	7
1.1	Che cosa è la bioinformatica?	7
1.2	Aree di ricerca	9
1.2.1	Analisi dei genomi	9
1.2.2	Analisi di sequenze	10
1.2.3	Filogenetica	10
1.2.4	Bioinformatica strutturale	10
1.2.5	Espressione genica	10
1.2.6	Genetica delle popolazioni	10
1.2.7	Biologia dei sistemi	10
1.2.8	Data mining	10
1.2.9	Database biologici	10
1.2.10	Bioimmagini	10
1.3	Storia	10

LIST OF TABLES

LIST OF FIGURES

CAPITOLO 1: LA BIOINFORMATICA

Per molti anni l'informatica è stata una scienza a sé stante, tuttavia negli ultimi decenni, grazie al progresso scientifico e tecnologico, sono nate nuove discipline chiamate genericamente **X-Informatics**. Queste sono il risultato dell'incontro tra l'informatica ed altre scienze di base (quali la biologia, la chimica, l'astronomia, la geologia etc) e tra queste citiamo la bioinformatica, la chemioinformatica, l'astroinformatica, la geoinformatica e così via. Anche se queste discipline sono diverse tra loro, ad esempio i dati raccolti in campo astronomico saranno di natura diversa rispetto quelli raccolti in campo biologico, condividono gli stessi obiettivi, come riportato nella pubblicazione [2] *X-Informatics: Practical Semantic Science*:

- Processamento ed estrazione delle informazioni
- Utilizzo trasparente ed efficiente dei dati in base al contesto scientifico, dalla raccolta, all'analisi fino alla catalogazione
- Integrazione di dati ottenuti tra sorgenti eterogenee
- Interazione con la raccolta dati adattata e personalizzata per l'utente
- Fornire supporto decisionale per l'utente, riducendo così i possibili errori e facilitando l'analisi dei risultati

Tra tutte queste discipline, risulta di particolare importanza la bioinformatica.

1.1 CHE COSA È LA BIOINFORMATICA?

Non esiste un unico modo con il quale definire la bioinformatica, infatti è possibile trovare definizioni diverse tra loro in quanto i professionisti non sempre concordano sulla portata del suo uso sia nel campo della biologia che dell'informatica. Tuttavia una possibile definizione è la seguente:

La bioinformatica è un campo multidisciplinare della scienza che coinvolge la genetica, la biologia molecolare, l'informatica, la matematica e la statistica, rivolta a studiare sistemi biologici utilizzando metodi e modelli informatici e computazionali.

Tra i vari obiettivi precedentemente elencati nell'introduzione al capitolo, va aggiunto quello che risulta l'obiettivo principale di questa disciplina, ovvero quello di aumentare la conoscenza di tutti quei processi di natura biologica.

A prescindere dalla natura del problema da affrontare, è possibile individuare un approccio standard, suddiviso in cinque steps:

- Studio ed analisi del problema da affrontare
- Collezionamento ed analisi di dati statistici a fronte di dati biologici in input
- Creazione di modelli ed uso di strumenti matematici che possano essere applicati al problema in esame, al fine di sviluppare un algoritmo
- Creazione, valutazione e test dell'algoritmo risolutivo del problema

Una parte fondamentale della bioinformatica consiste in esperimenti che generano dati ad alto throughput (high-throughput data), tra cui la misurazione dei modelli di espressione genica oppure la determinazione della sequenza genomica. Per **high-throughput data** si intendono quei dati biologici ottenuti tramite tecniche automatizzate e quindi non ottenibili attraverso metodi convenzionali. Il mining di questi dati può portare a nuove scoperte scientifiche non solo in campo biologico, ma anche medico, sia nel breve che nel lungo periodo.

Nel breve periodo, ad esempio grazie al *progetto genoma umano*¹, si possono scoprire nuovi geni legati alle malattie e nuovi bersagli molecolari, ovvero quei processi biologici, intesi come proteine, recettori, pathway biochimici etc su cui si può intervenire per modificare il decorso di una malattia.

Nel lungo periodo sarà possibile scoprire eventuali reazioni avverse ai farmaci da individuo ad individuo in base a dei tests, al punto tale che,

¹ Il progetto genoma umano (Human Genome Project) è stato uno dei più grandi progetti scientifici degli ultimi anni. L'obiettivo era quello di ottenere la sequenza del genoma umano (e quindi il suo intero DNA) e identificare in esso i geni contenuti. Il progetto è cominciato nel 1990, per poi essere completato nel 2003 ed ulteriori ricerche sono ancora in corso.

grazie all'informazione genetica ottenuta attraverso strumenti informatici, sarà possibile personalizzare l'uso di farmaci, portando ad una migliore efficacia alla terapia individuale, riducendo o addirittura eliminando possibili effetti collaterali.

1.2 AREE DI RICERCA

Data la natura eterogenea dei dati biologici, la bioinformatica comprende un vasto numero di aree di ricerca in continua crescita, di seguito presentate.

1.2.1 *Analisi dei genomi*

Uno dei principali focus della bioinformatica riguarda l'analisi dei genomi degli organismi il cui sequenziamento² è già stato completato, dal moscerino della frutta fino a quello dell'essere umano. L'analisi dei genomi è un'area di ricerca relativa alla genomica, ovvero quella disciplina che studia la struttura, il contenuto, la funzione e l'evoluzione del genoma

² Il sequenziamento è un processo mediante il quale viene determinata la struttura primaria delle macromolecole (composizione atomica e legami), quali il DNA, RNA e proteine.

1.2.2 *Analisi di sequenze*

1.2.3 *Filogenetica*

1.2.4 *Bioinformatica strutturale*

1.2.5 *Espressione genica*

1.2.6 *Genetica delle popolazioni*

1.2.7 *Biologia dei sistemi*

1.2.8 *Data mining*

1.2.9 *Database biologici*

1.2.10 *Bioimmagini*

1.3 STORIA

BIBLIOGRAPHY

- [1] Bayat Ardeshir (2002). *Science, medicine, and the future: Bioinformatics*. BMJ, 324(7344), 1018–1022. DOI: 10.1136/bmj.324.7344.1018
- [2] Borne, K. D. *X-Informatics: Practical Semantic Science*. <http://adsabs.harvard.edu/abs/2009AGUFMIN43E..01B>. George Mason University Fairfax VA, American Geophysical Union Fall Meeting, 12/2009. (Cited on page 7.)
- [3] Can T. (2013). *Introduction to Bioinformatics*. miRNomics: MicroRNA Biology and Computational Analysis. DOI: 10.1007/978-1-62703-748-8_4
- [4] ICAR CNR: Istituto Di Calcolo E Reti Ad Alte Prestazioni. *Bioinformatica*. <https://www.icar.cnr.it/bio-informatica/>
- [5] Jones C. Neil and Pevzner A. Pavel. *An introduction to bioinformatics algorithms*. Massachusetts, Massachusetts Institute of Technology, 2004.
- [6] Tramontano Anna (2003). *La grande scienza. Bioinformatica*. http://www.treccani.it/enciclopedia/la-grande-scienza-bioinformatica_%28Storia-della-Scienza%29/