

Indice delle figure

FIGURA 1: SCHEMA DELLA SINTESI PROTEICA; SONO RAPPRESENTATE LE DUE FASI DI TRASCRIZIONE E TRADUZIONE	7
FIGURA 2: ESEMPIO DI ANNOTAZIONE FUNZIONALE TRA UN GENE (K45) E LA CARATTERISTICA "È COINVOLTO NEL CANCRO AL FEGATO" ..	10
FIGURA 3: CRESCITA DELLE BANCHE DATI DI DATI BIOMOLECOLARI DISPONIBILI PUBBLICAMENTE ONLINE	12
FIGURA 4: (A) UN ALBERO; (B) UN DAG DI UNA ONTOLOGIA, RISPETTO ALL'ALBERO SI PUÒ NOTARE COME IL NODO X ABBAIA DUE PADRI E SIA CONTEMPORANEAMENTE AL SECONDO E TERZO LIVELLO DI PROFONDITÀ; (C) UN ESEMPIO DI DAG DI GO, ONTOLOGIA BIOLOGICAL PROCESS	13
FIGURA 5: (A) DAG DELLE ANNOTAZIONI DIRETTE; (B) DAG DELLE ANNOTAZIONI "UNFOLDATE" LUNGO LE DIREZIONI IN ROSSO	20
FIGURA 6: IL VETTORE v È UN AUTOVETTORE E QUANDO APPLICATO ALLA TRASFORMAZIONE LINEARE A VIENE SEMPLICEMENTE SCALATO; IL VETTORE x NON È AUTOVETTORE E QUANDO APPLICATO AD A CAMBIA IN DIREZIONE	22
FIGURA 7: INTERPRETAZIONE DELL'APPLICAZIONE DELLA TRASFORMAZIONE LINEARE A ; q_1 E q_2 SONO GLI AUTOVETTORI DELLA MATRICE A	23
FIGURA 8: IPOTETICO PROFILO DI ANNOTAZIONI DI UN GENE; (A) VALORI PREDETTI DALL'ALGORITMO; (B) ANNOTAZIONI PREDETTE DOPO LA FASE DI SOGLIATURA AL LIVELLO 0.7, I NODI A-B-D VIOLANO LA TRUE PATH RULE; (C) CORREZIONE DI ANOMALIE FROM LEAVES TO ROOT, CON SOGLIATURA A 0.7 IL TERMINE D SAREBBE ANCORA PREDETTO; (D) CORREZIONE DI ANOMALIE FROM ROOT TO LEAVES, CON SOGLIATURA 0.7 VERREBBE PREDETTO SOLAMENTE L'ANNOTAZIONE AD A.....	30
FIGURA 9: ARCHITETTURA GENERALE DEL SISTEMA	31
FIGURA 10: SCHEMA DELLE NUOVE COMPONENTI ARCHITETTURALI	33
FIGURA 11: PROCESSO DI VALIDAZIONE, ESEMPIO NEL CASO DI MODELLO LSI	34
FIGURA 12: WORKFLOW DEL PROCESSO DI PREDIZIONE, NELL'ESEMPIO SI USA IL METODO PLSA	36
FIGURA 13: RAPPRESENTAZIONE GRAFICA DELLO ASPECT MODEL.....	38
FIGURA 14: PARAGONE TRA LE DUE VERSIONI DEL MODELLO PREDITTIVO; (A) FORMULAZIONE ASIMMETRICA, (B) FORMULAZIONE SIMMETRICA.	39
FIGURA 15: ESEMPIO DI APPLICAZIONE DELL'ALGORITMO EM; NELL'ESEMPIO IN FIGURA L'ASSEGNAZIONE È AD UNA SOLA VARIABILE. PARTENDO DA UN ASSEGNAAMENTO CASUALE, AD OGNI ITERAZIONE SI TROVA UNA NUOVA SOLUZIONE MIGLIORE MUOVENDOSI NELLA DIREZIONE DEL GRADIENTE DELLA FUNZIONE OBIETTIVO. DAI DUE ESEMPI SI PUÒ VEDERE COME PARTENDO DA DUE INIZIALIZZAZIONI DIFFERENTI SI TROVINO DUE SOLUZIONI FINALI CON QUALITÀ DIVERSA.	45
FIGURA 16: ESEMPIO DI CALCOLO DEI VALORI DI G_FREQ ; (A) PROFILO DI ANNOTAZIONI DI PARTENZA, (B) VALORI CALCOLATI DI G_FREQ	50
FIGURA 17: CASO CON DUE FUNZIONALITÀ: IN AZZURRO I VETTORI-TOPIC, IN ROSSO LO SPAZIO DELLE PREDIZIONI; (A) NUMERO DI TOPICS MINORE DEL NUMERO DI FUNZIONALITÀ; (B) NUMERO DI TOPICS PARI AL NUMERO DI FUNZIONALITÀ; (C) NUMERO DI TOPICS MAGGIORE DEL NUMERO DI FUNZIONALITÀ.....	63
FIGURA 18: IN ALTO A SINISTRA LE CURVE ROC OTTENUTE, IN ALTO A DESTRA L'ANDAMENTO DELL'AREA DELLE CURVE ROC IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A SINISTRA L'ANDAMENTO DEL TEMPO MEDIO NECESSARIO ALL'ADDESTRAMENTO DEL MODELLO	

IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A DESTRA L'ANDAMENTO MEDIO DEL TEMPO NECESSARIO A PRODURRE LE PREDIZIONI SU UN GENE. I TEMPI SONO RIPORTATI IN MILLISECONDI. 65

FIGURA 19: IN ALTO A SINISTRA LE CURVE ROC OTTENUTE, IN ALTO A DESTRA L'ANDAMENTO DELL'AREA DELLE CURVE ROC IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A SINISTRA L'ANDAMENTO DEL TEMPO MEDIO NECESSARIO ALL'ADDESTRAMENTO DEL MODELLO IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A DESTRA L'ANDAMENTO MEDIO DEL TEMPO NECESSARIO A PRODURRE LE PREDIZIONI SU UN GENE. I TEMPI SONO RIPORTATI IN MILLISECONDI. 66

FIGURA 20: IN ALTO A SINISTRA LE CURVE ROC OTTENUTE, IN ALTO A DESTRA L'ANDAMENTO DELL'AREA DELLE CURVE ROC IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A SINISTRA L'ANDAMENTO DEL TEMPO MEDIO NECESSARIO ALL'ADDESTRAMENTO DEL MODELLO IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A DESTRA L'ANDAMENTO MEDIO DEL TEMPO NECESSARIO A PRODURRE LE PREDIZIONI SU UN GENE. I TEMPI SONO RIPORTATI IN MILLISECONDI. 67

FIGURA 21: IN ALTO A SINISTRA LE CURVE ROC OTTENUTE, IN ALTO A DESTRA L'ANDAMENTO DELL'AREA DELLE CURVE ROC IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A SINISTRA L'ANDAMENTO DEL TEMPO MEDIO NECESSARIO ALL'ADDESTRAMENTO DEL MODELLO IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A DESTRA L'ANDAMENTO MEDIO DEL TEMPO NECESSARIO A PRODURRE LE PREDIZIONI SU UN GENE. I TEMPI SONO RIPORTATI IN MILLISECONDI. 68

FIGURA 22: IN ALTO A SINISTRA LE CURVE ROC OTTENUTE, IN ALTO A DESTRA L'ANDAMENTO DELL'AREA DELLE CURVE ROC IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A SINISTRA L'ANDAMENTO DEL TEMPO MEDIO NECESSARIO ALL'ADDESTRAMENTO DEL MODELLO IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A DESTRA L'ANDAMENTO MEDIO DEL TEMPO NECESSARIO A PRODURRE LE PREDIZIONI SU UN GENE. I TEMPI SONO RIPORTATI IN MILLISECONDI. 69

FIGURA 23: IN ALTO A SINISTRA LE CURVE ROC OTTENUTE, IN ALTO A DESTRA L'ANDAMENTO DELL'AREA DELLE CURVE ROC IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A SINISTRA L'ANDAMENTO DEL TEMPO MEDIO NECESSARIO ALL'ADDESTRAMENTO DEL MODELLO IN FUNZIONE DEL NUMERO DI TOPICS, IN BASSO A DESTRA L'ANDAMENTO MEDIO DEL TEMPO NECESSARIO A PRODURRE LE PREDIZIONI SU UN GENE. I TEMPI SONO RIPORTATI IN MILLISECONDI. 70

FIGURA 24: CURVE ROC DATASET GALLUSGALLUS - CELLULAR COMPONENT 75

FIGURA 25: CURVE ROC DATASET GALLUSGALLUS - MOLECULAR FUNCTION 75

FIGURA 26: CURVE ROC DATASET GALLUSGALLUS - BIOLOGICAL PROCESS 76

FIGURA 27: CURVE ROC DATASET BOSTAURUS - CELLULAR COMPONENT 76

FIGURA 28: CURVE ROC BOSTAURUS - CELLULAR COMPONENT 76

FIGURA 29: CURVE ROC DATASET BOSTAURUS - MOLECULAR FUNCTION 77

FIGURA 30: CURVE ROC SCHEMI DI PESO, DATASET BOSTAURUS - CELLULAR COMPONENT, METODO PLSA 80

FIGURA 31: CURVE ROC SCHEMI DI PESO, DATASET BOSTAURUS - MOLECULAR FUNCTION, METODO PLSA 82