



Relazione scientifica sull'attività svolta nell'ambito del programma di mobilità breve CNR 2011

Fruitore: *Mario Rosario Guarracino*

Istituto di afferenza del Fruitore: *Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche*

Descrizione dettagliata dell'Istituzione ospitante: *Center for Applied Optimization, Department of Industrial and System Engineering, University of Florida*

Titolo del programma: *Metodi di Data Mining robusti e loro applicazione alla bioinformatica*

Obiettivi

Obiettivo del programma è la progettazione e implementazione di nuovi algoritmi robusti di data mining e la loro validazione tramite l'analisi di dati provenienti da strumentazioni biomedicali. In particolare, l'attenzione è stata rivolta agli algoritmi di classificazione supervisionata basati sugli autovalori generalizzati. Tali tecniche hanno mostrato di essere efficaci nell'analisi di grandi quantità di dati e pertanto possono essere estesi al caso di dati affetti da errori.

Attività svolte

Nel corso del programma è stata effettuata una approfondita ricerca bibliografica e si sono studiate le soluzioni esistenti per la classificazione supervisionata di dati affetti da errori, con particolare attenzione per i metodi di tipo Support Vector Machines. Sono stati individuati dei casi di studio provenienti da applicazioni reali della bioinformatica e biomedicina da utilizzare per la validazione. Sono stati identificati i limiti delle soluzioni esistenti, come ad esempio il tipo di errori che tali metodi sono in grado di gestire e stati proposti nuovi algoritmi che permettono di superarli. Le nuove soluzioni sono state implementate in maniera prototipale, mediante l'ambiente di risoluzione di problemi Matlab. La valutazione delle prestazioni, in termini di accuratezza della classificazione e complessità di spazio e di tempo, è stata effettuata sui casi di studio preventivamente individuati. I risultati verranno raccolti in una o più pubblicazioni scientifiche.

Risultati

Si è dimostrato che è possibile ottenere un metodo robusto di classificazione supervisionata mediante l'introduzione di un nuovo termine di regolarizzazione. I risultati preliminari mostrano che nel caso di perturbazioni puntuali ed ellittiche, la nuova formulazione dell'algoritmo è in grado di gestire gli errori presenti nei dati. Si è completata la sperimentazione della versione lineare dell'algoritmo.

I risultati saranno a breve sottomessi per pubblicazione su rivista JCR.

Firma del Fruitore

.....

.....