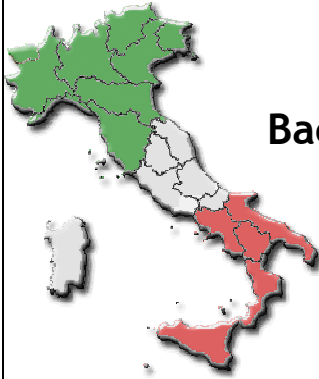


Philipps



Universität
Marburg



Bachelorarbeit in Bioinformatik mit Forschungsaufenthalt auf Sardinien

Prof. Dr. Dominik Heider
dominik.heider@uni-marburg.de

Thema: Implementierung von Phi-Delta-Diagrammen in R

Feature Selection (FS) Methoden spielen in der diagnostischen Medizin eine immer größere Rolle, um aus großen Patientendatensätzen relevante Parameter für ein Vorhersagemodell ausfindig zu machen. Es gibt bereits eine Vielzahl von FS Methoden, z.B. korrelationsbasierte FS, Support Vector Machines (SVMs) oder Variable Importance Measures von Random Forest Algorithmen.

Je nach Datensatzumfang und Variablentypen reagieren die verschiedenen FS Methoden mit verschiedenen Fehleranfälligkeiten. Um diese zu korrigieren wird vermehrt auf Ensembles von FS Methoden gesetzt. Unsere Arbeitsgruppe hat eine solche Ensemble Feature Methode (EFS) entwickelt, welche bereits als R-Paket für Benutzer zur Verfügung steht.

Eine weitere FS Methode für binäre Klassifikationen wurde von der AG Armano an der Universität von Cagliari, Sardinien entwickelt - die Phi-Delta-Diagramme. Die Diagramme stützen sich auf die Ergebnisse der Konfusionsmatrix und die sich daraus ergebenden Sensitivität und Spezifität, weshalb sie keine hohe Fehleranfälligkeit besitzt.

In diesem Projekt soll ein R-Paket erstellt werden und der Algorithmus der Phi-Delta-Diagramme implementiert werden. Das Paket soll anschließend mit dem EFS R-Paket zusammengefügt werden.

Die Arbeit ist in das Projekt ID:57265705, gefördert vom DAAD aus Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), eingebunden und beinhaltet einen 20-tägigen Forschungsaufenthalt an der Universität von Cagliari, Sardinien, Italien.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Literatur

[1] Saeys Y, Iñza I, Larrañaga P: A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007, 23(19): 2507-2517.

[2] Armano G: A direct measure of discriminant and characteristic capability for classifier building and assessment. *Information Sciences* 2015, 325:466-483.

Bitte melden Sie sich bei Interesse per Email bei:

03.11.2016

Prof. Dr. Dominik Heider
dominik.heider@uni-marburg.de