

Konzept für die Implementierung von MRMR

Beim Minimum redundancy - Maximum relevance (MRMR) handelt es sich um Verfahren zur Festlegung des Nutzens von einem Feature aus einem Feature Set.

(1) Hierbei kann davon ausgegangen werden, dass mengenweise Features zur Analyse angesammelt werden und die Betrachtung dieser zu kostspielig und redundant ist und um das zu vermeiden wird beim MRMR (1.1: Relevanz-Bestimmen) zuerst die Relevanz jeder einzelner Features iterativ durch die Menge hinweg bestimmt.

(1.2: Sortieren) Falls die Relevanz-Bestimmung fertig ist, sollen die zwei Mengen(, wobei eine der zwei Menge die Features als Inhalt hat und die andere Menge die dazu passenden Scores beinhalten,) nach Größe des Scores sortiert werden, wobei hier der größte Wert zuerst kommt und darauffolgende kleinere Werte folgen müssen.

(2: Redundanz prüfen) Die Auswahl, welches der Features aufgenommen wird, verläuft beim ersten Feature/Wert nur nach dem wählen des ersten Features des Feature-Sets, da die Sortierung dafür sorgen muss, dass die Entnahme des ersten Wertes das relevanteste Feature entspricht.

Bei den darauffolgenden müssen wir die Redundanz bedenken, weil jedes Feature gewisse Korrelationen zu schon aufgenommen Features haben kann und somit nicht mehr die Wertgewinnung bietet, die es laut der Relevanz hätte. Die Aufnahme von Features verläuft so lange, dass entweder keine Features in der zu betrachtenden Menge sind oder wir bei der Subtraktionen der Redundanz von der Relevanz auf null kommen oder negativ werden, was Indizien für eine nicht mehr an notwendigen Feature zunehmenden Feature-Set ist.

Die Bestimmung der Werte für die Relevanz und der Redundanz basieren auf den zugrunde liegenden Daten bezüglich dessen über welches Verfahren die Koeffizienten bestimmt werden (Die Bestimmung des Wertes unter den Bedingungen werden auf dem Blatt „“ dargestellt).

Umsetzung:

Score-Berechnung:

If (**Relevanz** — **Redundanz**) >= 0):
Annahme des Features

Relevanz:

Korrelationsbestimmung(Feature_subset,
Endpunkte
Methode)

X erhält das zu analysierende Feature
Y erhält Endpunkte wogegen die Analyse verläuft

Redundanz:

Korrelationsbestimmung(Feature_subset,
Feature_subset2,
Methode)

X und Y erhalten jeweils ein eigenes Feature, um die Abhängigkeit untereinander zu bestimmen

Korrelationsbestimmung:

Corr(X, Y, method)

X, Y = Menge

method = Vorgehensweise der Berechnung per
Default immer „pearson“