

Handbuch: 8.6. Korrelation

Wenn zwei unterschiedliche Tags in Beziehung zueinander stehen, nennen wir das Korrelation. Wenn wir von Korrelation sprechen, meinen wir damit typischerweise, dass die Beziehung linear ist. Das heißt, dass wenn eines dieser Tags variiert, sich auch der andere Tag proportional zum ersten verändert; mit anderen Worten: Die Beziehung wird durch die Konstante der Proportionalität zum Ausdruck gebracht.

Ein Beispiel für zwei korrelierte Tags ist die Temperatur und der Druck einer Substanz in einem Behälter. Solange der Behälter sein Volumen beibehält und keine Substanz dem Behälter zugeführt wird oder von ihm abfließt, stehen Druck und Temperatur in direkter Beziehung zueinander: Steigt das eine, steigt auch das andere. Weil diese beiden Tags somit streng verbunden sind, gelten sie in hohem Maße als korreliert.

Ein anderes Beispiel von zwei korrelierten Tags ist die Rotationsrate und die Schwingungsamplitude. Allerdings hängt diese Beziehung noch von einer Reihe anderer Tags ab. Während diese zwei Tags gemeinsam anwachsen oder abnehmen, kann es aber auch sein, dass sich ein Tag ohne das andere verändert, und zwar aufgrund von anderen Bedingungen. Diese Korrelation ist somit schwächer.

Die Stärke der Korrelation wird gemessen durch den Korrelationskoeffizienten. Das ist eine Zahl zwischen -1 und 1 . Ist der Koeffizient 1 , werden sich die zwei Elemente exakt gleich verändern. Ist der Koeffizient -1 , gibt es eine umgekehrte Korrelation, so dass, wenn ein Tag ansteigt, das andere sinkt. Ist der Koeffizient irgendwo dazwischen, ist die Beziehung zwischen beiden schwächer. Unter Praxisbedingungen werden zwei Tags allerdings nie perfekt korreliert sein, weil es bei jeder Messung stets eine natürliche willkürliche Variabilität gibt, zumindest eine Messunsicherheit.

Im Allgemeinen ist es nicht möglich, einen einzelnen Korrelationskoeffizienten in präziser Weise zu interpretieren, weil eine Korrelation stets vom Kontext abhängt. Was bedeutet es beispielsweise, wenn zwei Tags mit einem Koeffizienten von $0,758$ korrelieren? Für sich genommen, heißt das noch nicht viel. Erst wenn wir die Korrelationskoeffizienten miteinander vergleichen, gewinnen sie an Bedeutung. Wenn zwei Tags den Koeffizienten von $0,758$ haben und zwei andere den Koeffizienten von $0,123$, so können wir sagen, dass das erste Paar in einer sehr viel engeren Beziehung zueinander steht als das zweite. Welcher Korrelationskoeffizient hoch oder niedrig genannt wird, ist eine subjektive Entscheidung des Beobachters.

Bei industriellen Datensätzen beobachten wir manchmal Korrelationskoeffizienten von über $0,95$ zwischen verbundenen Tags, und häufig erhalten wir Koeffizienten von über $0,8$.

Wir können eine Korrelationsmatrix für eine ganze Kollektion von Tags berechnen, wobei wir die Korrelationskoeffizienten zwischen jedem Tagpaar berechnen. Diese Matrix kann dann analysiert werden, etwa um zu die Frage zu stellen: Welche Tags stehen zu einem bestimmten Tag in besonders enger Verbindung? Wir könnten solche Tags auswählen, die am engsten mit diesem einen uns besonders interessierenden Tag korreliert sind. Wir könnten uns zudem die Korrelationen zwischen den Tags anschauen und dann genau diejenigen eliminieren, bei denen wir eine hohe Korrelation erkennen, weil diese Tags nahezu dieselben Informationen enthalten.

Die Korrelationsmatrix kann somit die Grundlage dafür sein, Tags in Cluster zu gruppieren. Im Allgemeinen sind eng korrelierte Tags physisch miteinander verbunden und gehören deshalb zum selben Cluster.

© 2005 — 2024 algorithmica technologies Inc., alle Rechte vorbehalten