

Data Mining

Versuch 3 Recommender Systeme mit Collaborative Filtering

von

Armin Schwarz (as219)

Florian Tatzel (ft020)

Marc Walter (mw136)

Das git Repository mit den Quellcodes findet sich unter <https://github.com/floriant/DataMining>
Alle Dateien zu diesem Versuch befinden sich im Unterordner Versuch_3.

2 Durchführung Teil 1: Fiktive Filmbewertung

2.2 Ähnlichkeitsbestimmung

1. Welche Bedeutung hat der Übergabeparameter normed in der Funktion sim_euclid?

Durch den Parameter normed kann festgelegt werden, dass die Distanzen normiert werden sollen. Dadurch fallen Unterschiede zwischen den Größenordnungen einzelner Werte nicht mehr so stark ins Gewicht wie es sonst der Fall ist.

Erreicht wird dies dadurch, dass die berechnete Distanz durch die Anzahl aller Werte größer Null geteilt wird.

3. Vergleichen Sie die beiden Ähnlichkeitsmaße. Welches Ähnlichkeitsmaß erscheint Ihnen für diesen Anwendungsfall sinnvoller und warum?

Ähnlichkeitswerte für Michael Philips

Distanz	Toby	Jack Mathews	Gene S.	Mick LaSalle	Claudia Puig	Lisa Rose
Euklid	0.558	0.653	0.674	0.717	0.776	0.782
Pearson	1.000	0.135	0.205	0.258	1.000	0.405

Mit der euklidischen Distanz scheinen Michael Philips und Toby am Ähnlichsten zu sein, allerdings haben sie nur zwei Filme beide gesehen, die anderen drei hat nur einer von ihnen gesehen. Diese zwei Filme wurden jedoch ähnlich bewertet.

Bei der Pearson Distanz scheinen Michael Philips und Jack Mathews am ähnlichsten zu sein, diese haben beide vier von sechs Filmen gesehen und relativ ähnlich bewertet. Nur ein Film hat Jack nicht gesehen, dieser ihm vorgeschlagen werden.

Darum scheint in diesem Anwendungsfall die Pearson Distanz besser zu sein.

2.3 User based collaborative filtering

Enthält keine Fragen. Quellcodes befinden sich In **Versuch_3/Program/recommtest1.py** und **Versuch_3/Program/recommendations.py**.

2.4 Item based collaborative filtering

Enthält keine Fragen. Quellcodes befinden sich In **Versuch_3/Program/recommtest1.py** und **Versuch_3/Program/recommendations.py**.

3 Durchführung Teil 2: last.fm Musikempfehlungen

Enthält keine Fragen. Quellcodes befinden sich In

Versuch_3/Program/pylastRecTest.py und **Versuch_3/Program/recommendations.py**