

48 Die Mathematik hinter den Empfehlungen von Netflix und Amazon Prime

Ein Algorithmus schlägt Ihnen auf subtile Weise vor, welches Buch Sie als nächstes lesen oder welchen Film Sie sich ansehen sollen. Dahinter steckt lineare Algebra.

Internetunternehmen, wie z.B. Google, Amazon, Netflix, wissen am besten, wie man zu gesuchten Informationen, wie etwa ein Buch oder ein Film, kommt. Dazu verwenden sie Methoden der linearen Algebra, also die Mathematik der Vektoren und Matrizen, genau genommen die Matrixvervollständigung und die *Singulärwertzerlegung*²⁹.

Wenn man einen Film bei einem Streamingdienst gesehen hat, hat man im Anschluß die Möglichkeit ihn mit einem bis fünf Sternen zu bewerten. Auf Basis dieser persönlichen Bewertungen und ihrer Auswahl von Filmen wird bestimmt, welche weiteren Filme Ihnen gefallen könnten. Anschließend bekommt man entsprechende Empfehlungen zugeschickt.

Ein einfacher Weg, um herauszufinden, ob Ihnen z.B. „Das Boot“ gefallen würde, wäre der Durchschnitt aller Bewertungen, die der Film von anderen Nutzern erhalten hat. Dies ist jedoch kein sehr guter Ansatz, da er ignoriert, ob Sie generell U-Boot Filme mögen oder nicht.

Aber wie kann man entscheiden, welche Benutzer wie Sie sind? An dieser Stelle kommt die lineare Algebra ins Spiel. Stellen Sie sich eine riesige Tabellenkalkulation mit Spalten vor, die nach jedem der Zehntausenden von Filmen benannt sind, die auf Netflix verfügbar sind. Jeder der mehreren Millionen Abonnenten entspricht einer Zeile in dieser Tabelle. Wenn ein Benutzer einen Film bewertet, wird ein Eintrag ausgefüllt. Wenn ich zum Beispiel „Avatar – Aufbruch nach Pandora“ drei Sterne gebe, wird die Spalte, die diesem Film entspricht, in der Zeile mit meinem Namen aktualisiert.

Nur wenige bewerten mehr als hundert Filme, so dass nur ein winziger Bruchteil dieser riesigen Tabelle tatsächlich ausgefüllt wird. Die Aufgabe des Empfehlungsalgorithmus ist es, die gesamte Tabelle auf der Grundlage dieser sehr spärlichen Informationen auszufüllen. Wie kann das effizient gemacht werden?

Der Netflix-Algorithmus

Vor etwa drei Jahren bot Netflix eine Million Dollar für jeden an, der die Genauigkeit der Empfehlungen um 10% verbessert³⁰. Teams von Software-Ingenieuren und Enthusiasten aus der ganzen Welt haben an dieser Herausforderung teilgenommen. Die erfolgreichsten Ansätze basieren auf linearer Algebra: Während die Tabelle der Benutzerpräferenzen riesig ist, gibt es vielleicht ein paar Dutzend stereotypische Bewer-

²⁹T. Hastie, R. Mazumder, J.D. Lee, R. Zadeh, *Matrix completion and low-rank SVD via fast alternating least squares*, The Journal of Machine Learning Research, 16(1) (2015), 3367-3402.

³⁰<https://www.netflixprize.com/>

tungsprofile. Der Geschmack eines jeden Menschen wird als eine Mischung aus diesen Profilen beschrieben. Zum Beispiel mögen Sie vielleicht Krimis und gelegentlich eine Dokumentation. Lineare Algebra kann diese stereotypen Profile identifizieren und die magische Mischung liefern, die Ihre Vorlieben beschreibt.

Verbesserungen von 8 bis 9 % gegenüber dem bestehenden Netflix-Algorithmus wurden schnell nach Bekanntgabe des Preises erreicht. Das verlangte 10 %-Ziel erwies sich jedoch als schwer zu erreichen. Nach drei Jahren hat es ein Team von Programmierern aus aller Welt im Juli 2009 endlich geknackt. Sie erreichten dies, indem sie berücksichtigten, dass sich unsere Vorlieben im Laufe der Zeit stark verändern.

Yehuda Koren von Yahoo! Research Israel war einer der ersten Teilnehmer, die diese 10 %-Marke überschreiten konnten, indem er den Bewertungen eine zeitliche Komponente gab³¹. Dazu hatte er ein Modell entwickelt, das die zeitliche Dynamiken in den kollaborativen Filtern oder Empfehlungssystemen sensibel genug beschrieb³². Obwohl aktuelle Daten mehr über die aktuellen Präferenzen eines Benutzers verraten, gehen durch die Untergewichtung älterer Bewertungen zu viele wertvolle Informationen verloren. Der Trick von Koren bestand darin, alles beizubehalten, was das langfristige Verhalten des Nutzers vorhersagt, und gleichzeitig das temporäre Rauschen, die "Launen" herauszufiltern. Zu diesem Zweck hat Korens Modell eine Komponente für die Aufnahme des dauerhaften Signals und eine andere für die Erkennung von Anzeichen von vorübergehendem Rauschen, wie z.B. bestimmte abrupte Änderungen in den Bewertungen.

Eine Unterhaltung mit Maschinen

Mächtige Algorithmen formen, was Sie in Ihrem Webbrowser sehen. Sie können Ihnen auch auf subtile Weise vorschlagen, welches Buch Sie als nächstes lesen und welchen Film Sie sich ansehen sollten. Wenn Sie auf diese Vorschläge mit Ihrem Feedback antworten, beeinflussen Sie nicht nur Ihre zukünftigen Empfehlungen, sondern auch die aller anderen Nutzer desselben Dienstes. Sie nehmen an einem Gespräch zwischen Maschinen und Menschen teil.

³¹<https://www.wired.com/2009/09/bellkors-pragmatic-chaos-wins-1-million-netflix-prize/>

³²Y. Koren, *The BellKor Solution to the Netflix Grand Prize*, (2009). https://www.asc.ohio-state.edu/statistics/dmsl/GrandPrize2009_BPC_BellKor.pdf