

TELEPOLIS

Was macht den perfekten Film aus?

Roman Urbaner 29.06.2009

Ein Team hat nun zumindest einen besseren Algorithmus zur Vorhersage individueller Filmvorlieben entwickelt und hofft auf den Gewinn des Netflix-Preises

Ganz ausgestorben ist die alchemistische Suche nach der Weltenformel offenbar noch nicht, auch wenn sie inzwischen etwas zeitgemäßer – mit Datenreihen und Computeralgorithmen – daherkommt. So scheint es den Tüftlern in jüngster Zeit besonders das Mysterium des Kinos angetan zu haben: Schon seit Jahren jagen Forscher der Formel für den perfekten Film hinterher – mit mäßigem Erfolg. Jetzt scheint zumindest ein Algorithmus zur Vorhersage individueller Filmvorlieben gefunden zu sein, dessen Treffsicherheit alles Bisherige in den Schatten stellt. Eine Million Dollar lässt der amerikanische DVD-Online-Verleiher **Netflix (1) dafür springen.**

Für den Ausbau seines Empfehlungssystems hat es das US-Unternehmen Netflix ganz auf die Berechnung des persönlichen Kundengeschmacks abgesehen, was man sich auch einiges kosten lässt: **Eine Million Dollar** (2) warten seit Herbst 2006 auf denjenigen, der einen Algorithmus entwickelt, mit dem sich die Filmbewertungen der einzelnen Kunden um 10% exakter vorhersagen lassen als mit der firmeneigenen Software Cinematch. Dabei bemisst sich der prozentuelle Fortschritt nach der Fehlerquote der prognostizierten Punktebewertungen der Kunden. So liegen die Vorhersagen bei einem Fünf-Sterne-Raster derzeit im Durchschnitt deutlich weniger als einen Stern daneben (konkret: nur 0,8558 Sterne).



Eine harte Nuss also, an der sich zehntausende Hobby- und Profiprogrammierer jahrelang die Zähne ausbissen. Jetzt aber wurde sie tatsächlich geknackt. Dem Team "BellKor's Pragmatic Chaos" (das sich auch über die Namensgebung lange den **Kopf zerbrochen** (3) hat) ist es am Freitagabend erstmals gelungen, die Zehn-Prozent-Schallmauer zu **durchbrechen** (4). Doch bevor die Forscher dieser multinationalen Allianz, zu der sich vier der führenden Teams zusammengeschlossen haben, um dem "Netflix Prize" endlich näher zu rücken und das Preisgeld entgegennehmen können, haben die anderen Teilnehmer des Wettbewerbs noch exakt 30 Tage Zeit, das Resultat zu übertreffen. Erst dann gibt es die Million.

Dem amerikanisch-kanadisch-israelisch-österreichischen Gruppenteam von "BellKor's Pragmatic Chaos" stehen also bange Tage bevor. Mit dabei sind mit Andreas Töscher and Michael Jahrer (Team **Commendo** (5)) auch zwei Telematikstudenten aus Österreich, die ihren neuen Teamkollegen aus den Forschungsabteilungen von AT&T und Yahoo Research zuvor ein monatelanges Duell um die Spitze geliefert hatten.

Datenberge als Experimentierfeld

Empfehlungsdienste, wie sie Internet-Giganten wie Netflix oder Amazon anbieten, beruhen dabei auf dem simplen Prinzip, dass Kunden, die das gleiche Produkt mögen, mit einiger Wahrscheinlichkeit auch bei anderen Produkten ähnliche Vorlieben zeigen. Um diese halbwegs präzise vorhersagen zu können, bedarf es allerdings äußerst komplexer

Rechenoperationen, die der Fachwelt auf dem Gebiet der Künstlichen Intelligenz einiges abverlangen.

Als nun das US-Unternehmen im Oktober 2006 den Wettbewerb ausschrieb und zu diesem Zweck ein anonymisiertes Datenpaket gigantischen Ausmaßes freigab, ging ein erfreutes Raunen rund um den Globus. Für die Experten des "Maschinellen Lernens", die es gewohnt sind, mit viel kleineren Datenmengen Vorlieb zu nehmen, stellt der Wettbewerb nämlich einen seltenen Glücksfall dar: 100 Millionen Bewertungen, die 480.000 Netflix-Kunden zwischen 1998 und 2005 zu 18.000 Filmtiteln abgegeben hatten, stehen ihrem Forscherdrang zur Verfügung, ein Experimentierfeld, wie es die Programmiererszene in dieser Fülle wohl noch nie vorgefunden hat. Allerdings auch eines, bei dem Datenschützer den warnenden Zeigefinger hoben – und recht behielten: Denn schon nach einigen Wochen kam es zu einem peinlichen **Zwischenfall** (6), als es Hackern gelang, die anonymisierten Netflix-Daten mit den nicht anonymisierten Filmbewertungen der IMDb abzugleichen und sie so wieder konkreten Personen zuzuordnen.

"In unserem Forschungsbereich", erläutert Töschler den Stellenwert des Preises, "ist das zweifellos der größte Wettbewerb, den es je gegeben hat." 49.333 Teilnehmer in 40.540 Teams aus 184 Ländern machten sich an die Arbeit, warfen ihre Rechner an und wühlten sich unermüdlich auf der Suche nach der perfekten Formel durch die Datenmassen. Auf die Frage, welchem Erfolgsrezept sie ihren Durchbruch verdanken, wollen die beiden Österreicher, die den Wettbewerb lange ganz im Alleingang bestritten haben, nur verraten, dass sie, wie alle Spitzenteams, "ein ganzes Ensemble von Algorithmen" verwenden: "Wenn jemand das Ziel wirklich erreichen sollte", erklärten sie schon vor einem Jahr, "dann bestimmt nicht mit einem Einzelalgorithmus."

Das Bündeln einzelner Strategien mündete schließlich in die Kooperation mehrerer Spitzenteams "Die Teilnehmer haben erkannt, wie schwierig diese Aufgabe ist, und haben sich daher Leute gesucht, die ihre eigenen Fähigkeiten ergänzen", kommentierte Netflix-Sprecher Steve Swasey gegenüber der **New York Times** (7) die Formierung neuer Gruppenteams in der jüngsten Wettbewerbsphase.

Der Schweiß der Masse

Netflix selbst war mit ihrem Latein offenbar am Ende, wie sich die Treffsicherheit der Kundenempfehlungen noch verbessern ließe. Also lagerte man die Problemlösung einfach an die weltweite Forschergemeinschaft aus: ein schönes Beispiel für den ökonomischen Nutzen von **Crowdsourcing** (8) und **Prize economics** (9). Für das findige US-Unternehmen ist der Wettbewerb nämlich kein schlechtes Geschäft: Denn statt um teures Geld eine einzelne Softwarefirma anzuheuern, tüfteln und schufteten nun gleich Zehntausende Forscherteams, um ein ausgeklügeltes Empfehlungssystem zu entwickeln, dem die Konkurrenz sicherlich nicht so bald das Wasser reichen kann.

Während diese Experten jahrelang ihre Arbeit tun, ohne einen einzigen Cent Lohn zu verlangen, konnte man sich bei Netflix entspannt zurücklehnen und sich die Hände reiben über all den kostenlosen Werberummel, den das plötzlich entflammte Medieninteresse am Netflix Prize in Bewegung setzte. Zwischendurch spendierte das Unternehmen zwar einen eher bescheidenen "Fortschrittspreis". Ansonsten hieß es aber nur noch zuwarten, bis eines der Teams das Planziel erreicht und seine Algorithmen abgeliefert hat. Das kassiert dann zwar einen schönen Batzen Geld – alle anderen Mitstreiter aber gehen leer aus.

10 Prozent Handlung

Geht es Netflix darum, möglichst punktgenau persönliche Filmvorlieben zu prognostizieren, jagen andere der Weltenformel hinterher, um zur Bauanleitung für künftige Blockbuster zu gelangen. Immer wieder finden Meldungen von angeblichen Sensationen den Weg in die Medien, die aufgeregt davon berichten, dass schon wieder irgendjemand die Zauberformel für den perfekten Film entdeckt habe.

So behauptete die britische Filmautorin und Dozentin **Sue Clayton** (10) schon 2003, das Geheimnis des Kinos gelüftet zu haben. Alle Genres folgten demnach ähnlichen Gesetzen, wobei es allein auf die richtige Balance ankomme. "Ich war überrascht zu sehen, wie fein die unterschiedlichen Komponenten eines Films ausbalanciert sein müssen, um Perfektion zu erreichen", **erklärte** (11) Clayton damals und glaubte gleich auch die ideale Rezeptur anbieten zu können: "31 Prozent Action, 17 Prozent Comedy, 13 Prozent Gut gegen Böse, 10 Prozent Special Effects, 10 Prozent Handlung und 8 Prozent Musik." Von allen untersuchten Filmen komme die Disney-Produktion "Toy Story 2" diesem Schlüssel am nächsten (und spielte allein in Großbritannien mehr als 61 Millionen Euro ein). Produzenten und Medien waren jedenfalls kurz ganz aus dem Häuschen. Doch dann wurde es wieder still um Claytons Kinoformel.

Mr. Pink & Mr. Brown

Zuletzt war es die britische Firma **Epagogix** (12), die vor rund drei Jahren von sich reden machte: Mittels eines

raffinierten Computerverfahrens, das permanent mit Daten aus Filmanalysen und Einspielstatistiken gefüttert wird, will sie die Blockbuster-Formel gefunden haben, die ganz Hollywood auf den Kopf zu stellen verspricht, auch wenn man sich von Hollywoods sturen Studiobossen noch immer etwas unverstanden fühlt. Allein anhand des Drehbuchs lasse sich jetzt der spätere Kassenerfolg mit "verblüffender Trefferquote" vorhersagen:

Wir haben ein neuronales Netzwerk mit charakteristischen Details unzähliger Filme gefüttert, von denen wir genau wissen, wie viel Geld sie an den US-Kinokassen eingespielt haben. Wenn uns ein Filmstudio ein neues Drehbuch schickt, zerlegen wir das Skript in seine Bausteine und stecken diese in Form von Zahlen in unser Programm. Die Software kommt dann zu dem Schluss: Diese spezielle Kombination charakteristischer Elemente wird vermutlich so und so abschneiden. Am Ende spuckt das Programm eine Zahl aus, die den zu erwartenden Umsatz an der Kinokasse verrät.
Epagogix-Manager **Nick Meaney** (13) 2007 in einem Radiointerview

Nicht ganz traut **Thorsten Hennig-Thurau** (14), Marketingprofessor in Weimar und London, der seit Jahren ähnliche Computer-Kalkulationen für die Marktchancen neuer Filmproduktionen erstellt, den Behauptungen der Briten. Dass die Experten von Epagogix lediglich mit dem Filmscript auszukommen behaupten, macht ihn stutzig: "Bei der Entscheidung, einen Film am Startwochenende anzusehen, spielt das Drehbuch noch gar keine Rolle – das kennt zu diesem Zeitpunkt ja noch niemand", **meint** (15) der Professor und zieht Parallelen zwischen dem Anlaufen eines neuen Films und den Marktchancen einer neuen Joghurtsorte: Nicht das Drehbuch, sondern vor allem Marke und Marketing bestimmen demnach den Erfolg. An die 60 Eigenschaften, die für den Erfolg neuer Filme eine Rolle spielen, hat der deutsche Forscher ermittelt; statistisch wirklich signifikant sei jedoch nur ein gutes Dutzend.

Während man in Weimar das Modell, dessen Fehlerquote im Schnitt zwischen 15 und 20 Prozent liege, pflichtbewusst in **Fachpublikationen** (16) dokumentiert, blieb Epagogix den Beweis für die Treffsicherheit seiner Berechnungen offenbar bis heute schuldig: Selbst über die Identität der Formelfinder ließ die Firmenleitung nur lustige Pseudonyme à la Tarantino (**Mr. Pink** (17) & **Mr. Brown** (18)) nach außen dringen. Auch darüber, welche Faktoren durch den Rechner gejagt werden, bis dieser seine Prognose ausspuckt, hüllt sich die Firma in Schweigen. Man ließ nur verlauten, dass der Stellenwert von kostspieligen Stars grotesk überbewertet werde; und dass es vielmehr gerade belächelte Kleinigkeiten seien, auf die es ankomme – etwa ob der Protagonist einen Hut trägt oder ob er in Jeans herumspaziert.

Der perfekte Schrecken

Weniger geheimniskrämerisch gab sich da ein Mathematikerteam des **Londoner King's College** (19), das die Klassiker des Horrorfilms, von "The Shining" bis zum "Kettensägenmassaker", so lange zerpfückte, bis es 2004 zur Formel des maximalen Schreckens gelangte:

$$\text{Horror} = (\text{es} + \text{u} + \text{cs} + \text{t})2 + \text{s} + (\text{tl} + \text{f})/2 + (\text{a} + \text{dr} + \text{fs})/\text{n} + \sin x - 1$$

Das erste Glied der Gleichung soll die Spannung beschreiben, wobei es für die Musik (escalating music), u für das Unbekannte (the unknown), cs für Verfolgungsjagden (chase scenes) und t für Eingesperrtsein (sense of being trapped) steht, ergänzt um den Faktor s (shock) und die richtige Dosis Realismus (tl: true life) und Fantasie (f). Die Horrortauglichkeit der Umgebung nimmt die britische Horrorformel im Anschluss in den Blick: Die Variable a bildet dabei das Alleinsein des Protagonisten (character is alone) ab, und dr und fs stehen für Dunkelheit (dark room) und die passende Szenerie (film setting). Und weil Ausgesetztheit und Wehrlosigkeit der Figuren mit der Gruppengröße schwinden, teilt man das Ganze noch schnell durch die Zahl der beteiligten Personen(n).

Nach dieser Parade mehr oder weniger subtiler Ingredienzien geht es zu guter Letzt dann doch auch ans Eingemachte: Mit $\sin x$ (blood and guts) erfasst die Gleichung schließlich das unvermeidliche Gemetzel. Und weil allzu viel Schablonenhaftigkeit dem Gruselvergnügen im Weg steht, subtrahiert man am Ende noch 1 (für stereotypes). Der Weisheit letzter Schluss scheint jedoch auch das noch nicht gewesen zu sein – der wundersame Aufschwung des Horrorgenres steht zumindest noch aus.

Den Stein der Weisen haben also auch sie nicht gefunden. Das unfreundliche Verdikt, mit dem der Drehbuchautor **William Goldman** (20) einst Hollywoods übereifrige Marktstrategen bedachte, dürfte noch länger seine Gültigkeit behalten: "Nobody knows anything: Jedes Mal wird bloß drauflos geraten." Bei Netflix eben neuerdings ein wenig exakter.

Links

(1) <http://www.netflixprize.com/>

(2) <http://www.heise.de/tr/artikel/79639>

- (3) <http://pragmatictheory.blogspot.com/2009/06/whats-in-name.html>
- (4) <http://www.netflixprize.com//leaderboard>
- (5) http://www.commendo.at/index.php?lang=1&_0=2&_1=1
- (6) <http://arxiv.org/abs/cs/0610105>
- (7) <http://bits.blogs.nytimes.com/2009/06/26/and-the-winner-of-the-1-million-netflix-prize-probably-is/>
- (8) http://www.wired.com/wired/archive/14.06/crowds.html?pg=1&topic=crowds&topic_set=
- (9) http://www.wired.com/culture/geekipedia/magazine/geekipedia/x_prize
- (10) <http://www.sueclaytonfilm.com/>
- (11) http://www.welt.de/print-welt/article693926/Professorin_entdeckt_die_Formel_fuer_den_perfekten_Film.html
- (12) <http://www.epagogix.com/>
- (13) <http://www.dradio.de/dlf/sendungen/forschak/987379/>
- (14) <http://www.moviesuccess.org/>
- (15) <http://tages-anzeiger.ch/dyn/news/print/wirtschaft/767637.html>
- (16) <http://www.uni-weimar.de/medien/marketing/moviesuccess/research.php?tab=2>
- (17) http://www.tarantino.info/wiki/index.php/Mr._Pink
- (18) http://www.tarantino.info/wiki/index.php/Mr._Brown
- (19) <http://news.bbc.co.uk/2/hi/entertainment/3537938.stm>
- (20) <http://movies.ign.com/articles/386/386212p1.html>

Telepolis Artikel-URL: <http://www.heise.de/tp/r4/artikel/30/30622/1.html>

Copyright © Heise Zeitschriften Verlag