

Jeśli w przypadku dostosowywania klasyfikatora wykorzystywalibyśmy uczenie maszynowe, wagę każdej cechy oraz próg określilibyśmy na podstawie danych. Zgromadzilibyśmy je, prosząc dzieci o przygotowanie jakichś dań i odnotowując, które najbardziej im smakowały. Następnie wykorzystalibyśmy standardowy wzór statystyczny do określenia wag konkretnych cech na podstawie zebranych danych. Wzór ten jest czytelnikom zapewne znany ze szkoły średniej (choć został natychmiast wymazany z pamięci), z lekcji, na której uczono, w jaki sposób na kartce papieru poprowadzić linię przez zbiór punktów (x, y) . Tutaj wykorzystujemy ten sam wzór, z tą tylko różnicą, że każdej współrzędnej y przypisana jest więcej niż jedna współrzędna x .

Kiedy już na podstawie garstki przepisów – dajmy na to, że 100 z 15 tysięcy – dopasujemy wagi, możemy za pomocą komputera uruchomić klasyfikator, który zajmie się 14 900 przepisami i ustali, jaki z nich się nadaje, a jaki nie. Możemy wybrać 200 najlepszych przepisów ze zbioru 15 000 na podstawie tego, co ustalił klasyfikator, wypróbować je, by upewnić się, że się nadają, i zachować w publikacji tylko te najlepsze, a wtedy wszystko będzie gotowe.

Dysponując już umiejętnością dostosowywania klasyfikatora, wróćmy teraz do Wielkiej Nagrody Netflixsa i sprawdźmy, jak za jego pomocą przebiega proces rekomendacji filmów.

Cele konkursu

Jakimi kryteriami powinien kierować się Netflix, rekomendując filmy klientom? Jakie cele powinny mu przyświecać? Takie pytania padły w artykule opublikowanym przez Clive'a Thompsona w „New York Timesie” w roku 2008, a więc wtedy, gdy konkurs właśnie trwał¹². Czy Netflix – pytał autor – świadcząc usługę rekomendacji filmów, powinien ograniczyć się do „bezpiecznych wyborów”, filmów, które z pewnością polubimy, choć nie sprawiają, że opuścimy swoją strefę komfortu? A może powinien odegrać rolę ekscentrycznego sprzedawcy ze sklepu z filmami, polecającego filmy, które być może pokochamy miłością bezgraniczną, ale też takie, które uznamy za niewypały?¹³

W tradycyjnych wypożyczalniach filmowych tamtego czasu większość wypożyczanych tytułów to były filmy nowe i popularne; wypożyczalnie te mogły bazować na tej wąskiej selekcji filmów, by rekomendowanie filmów było łatwiejsze. Z Netflixem sprawa przedstawiała się inaczej: 70% wypożyczonych filmów stanowiły tytuły niezależne i starsze. Biorąc pod uwagę szeroki katalog filmów oraz duży odstęp czasu dzielący jedno wypożyczenie od drugiego, Netflix polegał na Cinematch, własnym systemie rekomendacji filmów, za pomocą którego polecał pozycje użytkownikom. Udoskonalanie Cinematch miało zasadniczy

wpływ na wynik finansowy firmy, jako że groziła jej utrata tych klientów, którzy obejrzeni zbyt mało filmów Netflixsa lub też nie spodobał im się film, na który czekali kilka dni: widzowie ci mogli w każdej chwili zrezygnować z subskrypcji¹⁴.

Tak więc programiści Netflixsa wciąż zajmowali się udoskonalaniem algorytmu Cinematch. Nie mogąc już posunąć się w pracach do przodu, podjęli decyzję o ogłoszeniu Wielkiej Nagrody Netflixsa, w ramach której zespół zdolny pobić ich własny algorytm o 10 procent miał otrzymać nagrodę w wysokości miliona dolarów. Jak oznajmił dyrektor Netflixsa Reed Hastings, wypłacenie nagrody nie stanowiło problemu dla firmy: korzyści finansowe płynące z ulepszenia rekomendacji filmów miały potencjał szybkiego zrekompensovania kosztów poniesionych na jej ufundowanie¹⁵. Nawet niewielkie poprawki wprowadzone do systemu rekomendacji mogły przynieść duże zyski, ponieważ mnożyły się przez setki milionów rekomendacji udzielanych każdego dnia¹⁶. Na wypadek, gdyby żadnemu zespołowi nie udało się osiągnąć celu 10 procent, Netflix przygotował Nagrodę Postępu: zgodnie z jej zasadami, jeśli uczestnicy dokonają wystarczających postępów w ciągu roku, najlepszy zespół otrzyma 50 tysięcy dolarów. Netflix obwarował otrzymanie tych nagród tylko jednym warunkiem: zwycięzca musiał opublikować szczegółowy opis stworzonego algorytmu rekomendacji.

Netflix ułatwił zadanie uczestnikom przez ustanowienie jasnego, obiektywnego celu ich dążeń. Uczestnicy mieli za zadanie dokonać prognozy, ile gwiazdek zostało przyznanych przez określone grono klientów określonym filmom w określonym czasie. Netflix oceniał zespoły, obliczając średnią kwadratową różnicę oceny prognozowanej i oceny rzeczywistej wystawionej przez klientów i zawartej w tajnej bazie danych, której uczestnicy mieli nigdy nie poznać¹⁷.

Za każdym razem, gdy jakiś zespół przedstawił Netflixsowi swoje ustalenia, firma oceniała pracę zespołu na podstawie tajnej bazy danych, po czym publikowała jego nowy wynik w rankingu cieszącym się zainteresowaniem zarówno innych zespołów, jak i dziennikarzy¹⁸. Technicznie rzecz biorąc, dany zespół mógł „podpatrzeć” oceny klientów przez przedstawianie Netflixsowi wielu wyników, które miały zostać poddane analizie na podstawie tajnej bazy danych. Netflix jednak był na tyle przebiegły, że trzymał w zanadrzu inną tajną bazę danych, której uczestnicy nigdy nie mieli zobaczyć. Ta podwójna tajna baza danych miała zostać użyta jedynie na zakończenie konkursu, aby ocenić wyniki najlepszych kandydatów do nagrody.

Potężna macierz ocen

Zważywszy że konkurs ogłoszony przez Netflixsa koncentrował się wyłącznie wokół ocen filmów wystawionych przez klientów, przydatne będzie rozważenie

nagrody w perspektywie gigantycznej macierzy ocen. Ilustracja 5.2 przedstawia niewielki wycinek tej macierzy (ze zmyślonymi liczbami).

	Użytkownik 1	Użytkownik 2	Użytkownik 3	Użytkownik 4	...	Użytkownik 480 185	Użytkownik 480 186	Użytkownik 480 187	Użytkownik 480 188	Użytkownik 480 189
Terminator 2	5	5	4	...	2	5				5
Skrawki	1	1	2	?	...		3	2		?
Słodkie zmartwienia		4		?	...	2		4		
Napoleon Wybuchowiec	4		2		...		5	5		
Labirynt Fauna	4				...				5	5
...
Tajemnica spalonego domu	3				4	...	?	?		
X-Men	?			4		...	2	4		5
Edward Nożycoręki	5			5		...		5		
Krótkie spięcie	4	4				...	1			
Toy Story		?	4		5	...		4		

Ilustracja 5.2. Przykłady ocen, które klienci Netfliksa mogli wystawić wybranym filmom. Netflix dostarczył badaczom część ocen w postaci macierzy (na ilustracji wyrażone są liczbowo). Zawodnicy musieli dokonać prognoz brakujących rekomendacji (na ilustracji oznaczone są pytajnikami)

Macierz była ogromna: zawierała oceny 177 700 różnych filmów wystawione przez 480 189 różnych osób¹⁹. Netflix przedstawił uczestnikom niektóre z tych ocen i poprosił ich, by dokonali prognoz dotyczących ocen brakujących (znaki zapytania w kratkach). Mimo swych rozmiarów tylko 1 procent macierzy zawierał jakiegokolwiek liczby: rzecz jasna, większość klientów Netfliksa nie wystawiła ocen większości obejrzanych filmów.

Od czego więc powinni byli zacząć uczestnicy?

W początkowej fazie zawodów większość zespołów z czołówki przyjęła podobne podejścia w analizie ocen. Członkowie zespołu BellKor – naukowcy z AT&T oraz Yahoo! – zwracali uwagę na wartość płynącą z rozpoczęcia analizy od prostego modelu bazowego tłumaczącego podstawowe trendy widoczne w macierzy ocen. Model bazowy zaproponowany przez BellKor rozpoczynał pracę od dwóch elementów. Pierwszy odnosił się jedynie do filmów; możemy nazwać go efektem *E.T.*²⁰. Efekt *E.T.* mierzył popularność danego filmu bez względu na to, kto poddał go ocenie. W bazie danych dostarczonej przez