



Algoritmy odporúčania filmov v Netflix*

Yehor Bohuslavskiyi

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xbohuslavskiyi@stuba.sk

17. október 2024

Abstrakt

Túto tému som si vybral, pretože je veľmi zaujímavé, ako technológie a umelá inteligencia (ktorá je v súčasnosti jednou z najrozvíjajúcejších sa tém) ovplyvňujú naše každodenné rozhodovanie o tom, čo sledujeme. Ja som zvedavý, že ako po zhliadnutí jedného tureckého filmu, ktorý som ani neohodnotil, sa mi v odporúčaní začali objavovať podobné filmy.

V systéme sa nachádza viac, ako 15 000 filmov a zobrazuje vám len tie, ktoré sa vám budú páčiť (neskôr dozvieme sa, ako to funguje). Fakt: ani jeden používateľ Netflix nebol by schopný si samostatne nájsť film alebo seriál, ktorý by sa mu páčil bez požívania algoritmu odporúčania filmov [5].

V tomto článku sa podrobne pozrieme na algoritmy spoločnosti Netflix. V súčasnosti je systém Netflix postavený na algoritme, ktorý využíva umelú inteligenciu a strojové učenie. Pozrieme sa aj na umiestnenie odporúčaní na obrazovke, kde je niekoľko typov návrhov. Prvým je „Odporúčané pre vás“, po ňom idú „Trendy“, na stránke sa nachádza aj záložka „New Releases“ [4].

Podrobne sa pozrieme aj na filtračné algoritmy, ktoré sa delia na dva hlavné typy. Akú rolu odohrávajú hodnotenia a správanie používateľov v rámci algoritmu collaborative filtering a „filtrovanie na základe obsahu“.

Zaujímavým aspektom sú aj systémy hodnotenia. Patrí medzi ne „Personalizované hodnotenie videa (PVR)“, „Top-N Ranking“, „Rebríček obľúbených filmov“ a „Zoznam zaujímavého obsahu na neskoršie sledovanie“. Každé z týchto hodnotení je personalizované pre jednotlivých používateľov, čo je veľmi zaujímavé.

V rámci článku sa budeme venovať aj typom údajov, ktoré Netflix používa. Tie sa zhruba delia na dva typy: základné typy zhromažďovaných údajov a dodatočné informácie. Medzi základné patria „interakcie používateľa so stránkou“, ktoré pozostávajú z histórie sledovania a hodnotení. Druhým typom sú „korelačné údaje“, „informácie o obsahu knižnice Netflix“, ako napríklad žánre. Ďalšie uvidíme ako ovplyvňujú špecifickejšie typy informácií, ako sú napríklad denný čas, používané zariadenie alebo priemerná dĺžka sledovania.

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2024/25, vedenie: Ing. Richard Marko, PhD

1 Úvod

Cieľom spoločnosti Netflix je používať odporúčacie algoritmy na maximalizáciu spokojnosti používateľov. Hlavným cieľom je prispôbiť obsah každému jednotlivému používateľovi.

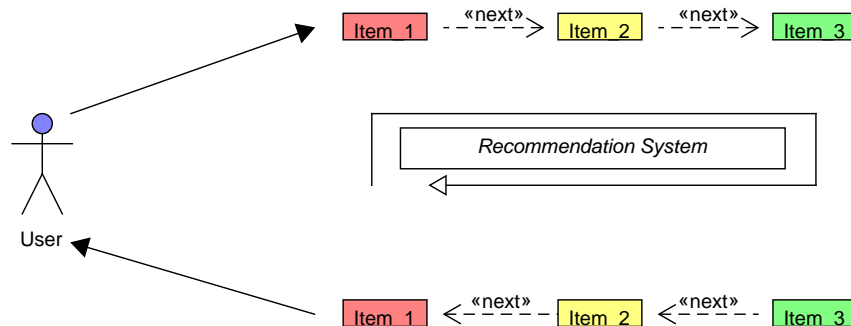
Netflix má obrovskú knižnicu obsahu a obmedzené rozhranie obrazovky, čo veľmi sťažuje odporúčanie relevantného obsahu, a každý používateľ má iný vkus, preferencie a zvyky pri sledovaní, takže personalizácia musí byť veľmi presná. [1] Čo robí Netflix na dosiahnutie tohto cieľa, sa dozvieme po tomto článku.

2 Technológie a Umelá inteligencia v Netflix

V súčasnosti veľké spoločnosti v IT priemysle využívajú umelú inteligencia, čo nezostalo bez povšimnutia spoločnosti Netflix, ktorá kombinuje niekoľko modelov a prístupov (tzv. ensemble learning) na vytvorenie najlepších odporúčaní. [1]

3 Typy Odporúčaní na Netflix

Netflix používa dvoj úrovňový systém hodnotenia založený na riadkoch, čo to znamená? Obsah je hodnotený dvakrát, v stĺpcoch a v každom riadku, ako sa dozvieme v článku ďalej. Na základe kombinácie týchto prístupov sa vytvárajú personalizované odporúčania pre konkrétného používateľa. [7]



Obr. 1: Recommendation System

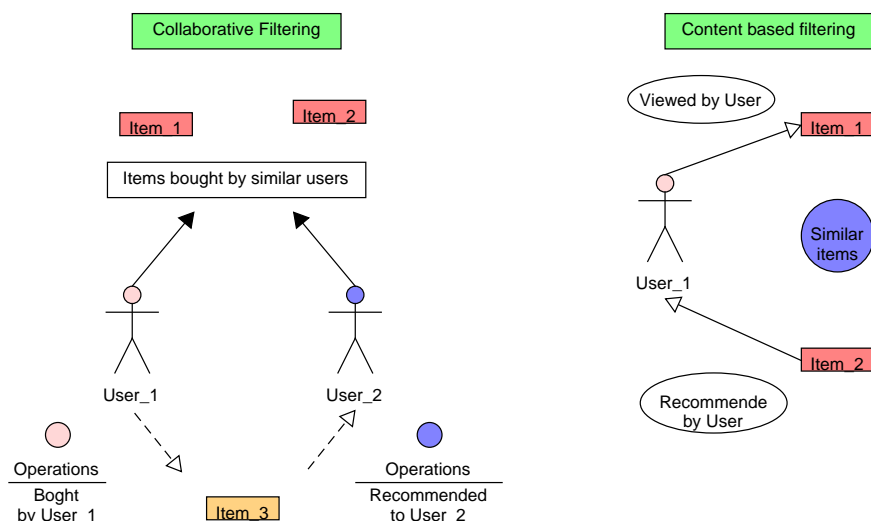
4 Filtračné algoritmy

Netflix používa dva typy filtrovania: kolaboratívne filtrovanie a filtrovanie podľa obsahu. Rozdiel môžete rozoznať podľa názvu, ale poďme sa naň pozrieť bližšie.

Kolaboratívne filtrovanie je algoritmus, ktorý pomáha odhadnúť záujmy nového používateľa na základe akcií predchádzajúcich členov cieľového publika, rozdeľuje používateľov podľa zadaných charakteristík. Najprv identifikuje podobné záujmy a potom hľadá podobné skóre. Každý používateľ dostane takz-

vanú váhu. Keď sa váha niekoľkých návštevníkov stane maximálne podobnou, sú spojení do skupiny „susedov“ a zobrazí sa im rovnaký obsah. [6]

Filtrovanie na základe obsahu zohľadňuje pri odporúčaní vlastnosti položky a preferencie používateľa. Tento prístup zahŕňa analýzu samotného obsahu, napríklad filmových žánrov, hercov alebo kľúčových slov, a jeho porovnanie s predchádzajúcimi preferenciami používateľov. [8]



Obr. 2: Type of Filtering

5 Systémy Hodnotenia

Kľúčovými zložkami odporúčacieho systému spoločnosti Netflix je niekoľko personalizovaných systémov hodnotenia, z ktorých každý je navrhnutý tak, aby zohľadňoval rôzne aspekty zvykov používateľa pri sledovaní. Poďme sa do týchto systémov ponoriť a preskúmať, ako fungujú.

Personalizovaný rebríček videí (PVR) je jedným z najsofistikovanejších nástrojov spoločnosti Netflix a neustále sa prispôbuje na základe spätnej väzby používateľov. Tento model založený na hlbokom učení využíva historické údaje o zvykoch používateľa pri sledovaní na uprednostňovanie obsahu, ktorý ho pravdepodobne najviac zaujíma. Sila PVR spočíva v jeho schopnosti nielen pochopiť minulé správanie, ale aj predpovedať budúce interakcie. [9] [10]

Systém Top-N Ranking - ako už názov napovedá, tento model generuje zoznam N najlepších odporúčaní obsahu pre každého používateľa. Robí to kombináciou kolaboratívneho a obsahového filtrovania. Obsah je zoradený na základe toho, ako dobre zodpovedá preferenciám používateľa, čím sa zabezpečí, aby sa najrelevantnejšie tituly zobrazovali na vrchole obrazovky. [3]

Poradie obľúbených filmov je ďalší model hodnotenia, ktorý využíva trendový obsah. Netflix používa tento systém na zaraďovanie filmov na základe ich

celosvetovej aj regionálnej popularity. Ukážkovým príkladom bolo obdobie pandémie COVID-19, keď prudko stúpol počet divákov filmu *Nákaza*. [3]

Zaraďovanie zaujímavého obsahu na neskoršie pozretie Netflix používa aj systém hodnotenia, ktorý sa zameriava na zaujímavý obsah, ktorý si používatelia môžu chcieť pozrieť neskôr. Sledovaním tohto správania Netflix zabezpečuje, aby platforma zostala relevantná pre široký vkus používateľov a aby sa mohli vždy vrátiť k predtým sledovanému obsahu. [3] [2]

Netflix sa neustále zlepšuje prostredníctvom A/B testovania. Netflix pravidelne testuje varianty svojich odporúčacích algoritmov na rôznych skupinách používateľov a analyzuje, ktoré modely vedú k vyššej angažovanosti a spokojnosti. [2]

6 Typy Údajov v Netflix

Cesta spoločnosti Netflix k personalizácii sa začína zberom údajov. A nejde len o to, aké seriály sa vám páčia. Netflix analyzuje všetko od žánrov, ktoré preferujete, až po režisérov, ktorých sledujete, dĺžku sledovania a dennú dobu. Nie je to len súbor údajov - je to základom schopnosti spoločnosti Netflix prispôbiť váš zážitok zo sledovania. Vďaka viac ako 1 300 klastrom odporúčaní Netflix vie nielen to, že máte radi thrillery; vie, že máte radi psychologické thrillery so silnými ženskými hrdinkami, ktoré sledujete väčšinou cez víkend večer. [1] [10]

7 Záver

Odporúčací systém Netflix je zložitý mechanizmus na spracovanie veľkého množstva údajov a preferencií používateľov. Je to kľúčový nástroj na udržanie záujmu používateľov a na to, aby sa nestratili v obrovskej knižnici obsahu.

Literatúra

- [1] James Bennett and Stan Lanning. An introduction to recommender systems at netflix.
- [2] David Chong. Deep dive into netflix's recommender system.
- [3] Arkadiusz Krysik. Inside the netflix algorithm: Ai's role in personalizing user experience.
- [4] Arkadiusz Krysik. Netflix algorithm: Everything you need to know about the recommendation system of the most popular streaming portal.
- [5] Arkadiusz Krysik. Netflix algorithm: How netflix uses ai to improve personalization.
- [6] Prem Kumar. Netflix movie recommendation — using collaborative filtering.
- [7] Yakov Livshits. How netflix's recommendation algorithm works: A look behind the scenes.

- [8] Sachin Sharma. The magic of recommendation systems: How netflix knows what you like.
- [9] Argoid Team. How does netflix's recommendation engine work?
- [10] The Product Folks Team. Unlocking viewer engagement: How netflix uses data to win over 230+ million users worldwide.