

Je odporúčací systém Netflix-u problém ?*

Adam Glogovský

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

5. október 2024

Abstrakt

Cieľom tejto práce je analyzovať, ako spoločnosť Netflix spracúva užívateľské informácie a posúdiť, či je tento proces správny. Na úvod objasním problematiku všeobecných odporúčacích systémov, ktoré slúžia ako základ pre pochopenie kontextu. Následne sa zameriam na to, kam Netflix posiela užívateľské informácie, akým spôsobom ich spracúva a ako tieto dáta využíva na zvýšenie užívateľského komfortu na platforme. Zároveň preskúmam, či takéto zlepšenie užívateľského zážitku znamená zvýšené náklady pre platformu a či je tento proces efektívny a finančne prospešný pre Netflix.

Veľká pozornosť bude venovaná analýze spôsobu spracovania údajov. Poukážem na etické aspekty tohto procesu a preskúmam, či je správne tieto informácie spracúvať daným spôsobom. Kľúčovým prvkom, na ktorý sa zameriam, je algoritmus, ktorý spracúva užívateľské dáta, a infraštruktúra, v ktorej tento proces prebieha. [1]

Ďalej sa budem zaoberať otázkou, či algoritmy Netflixu neznevýhodňujú menej populárny alebo „nemainstreamový“ obsah tým, že ho menej zobrazujú užívateľom. Preskúmam, či je tento obsah iba menej favorizovaný a posúvaný nižšie v poradí vyhľadávania, alebo je pre určitých užívateľov úplne skrytý.

Na záver tejto práce získate prehľad o tom, ako Netflix spracúva užívateľské informácie, posúdenie etickej správnosti tohto procesu, jeho profitabilitu pre spoločnosť a jeho vplyv na užívateľov platformy. [8]

Úvod

Zaujímalo vás niekedy, ako Netflix určuje, čo by sa vám pravdepodobne páčilo keď prechádzate cez ich ponuku filmov a seriálov ? Viac o tejto tématike sa dozviete tu: 3. Alebo ako určuje kde a ako tieto filmy a seriály rozložiť na hlavnú stránku ? Viac o tomto tu:2. Možno vás niekedy zaujímalo či je takýto algoritmus profitabilný a prečo vlastne chce Netflix zlepšiť svojim používateľom používanie ich platofrmy ? Odpoveď nájdete tu:4

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2024/2025, vedenie: Mgr. Yevheniia Kataieva, PhD.

1 Krátka história

Už v roku 2006 mal Netflix záujem o odporúčacie systémy. [1] Preto, oznámil Cenu Netflixu ako súťaž strojového učenia o to, kto vytvorí aspoň o 10% lepší odporúčací systém ako bol ten ich, Cinematech. Nový odporúčací systém mal byť lepší v konkrétne $RMSE$ (*root mean squared error*)¹ pre predpokladané hodnotenie. A zrazu sa veľmi veľa rozdielných firiem snažilo zlepšiť svoje algoritmy.

2 Problem zlatej rybky

Keď majú ľudia príliš veľa vecí, na ktoré sa môžu pozrieť, nepozrú sa na nič. Štúdia spomenutá v [4] zistila, že v priemere typický používateľ stráca pozornosť po približne minúte prezerania si možností, čo si pozrie. No a teda sa Netflix veľmi usiluje aby počas tejto minúty, ukázal užívateľovi čo najlepší výsledok. Určite sa aj vám niekedy stalo, že ste otvorili Netflix hlavnú stránku, chvíľku popozerali a potom Netflix vyplí. To je pre Netflix najhoršia vec, čo môže nastať, pretože vy ste si nič nepozreli a aj napriek tomu, že im platíte mesačne, ak sa toto stane viac krát, určite to predplatné zrušíte. Nikto predsa nechce službu ktorá im neponúka to, čo chcú.

3 Algoritmus

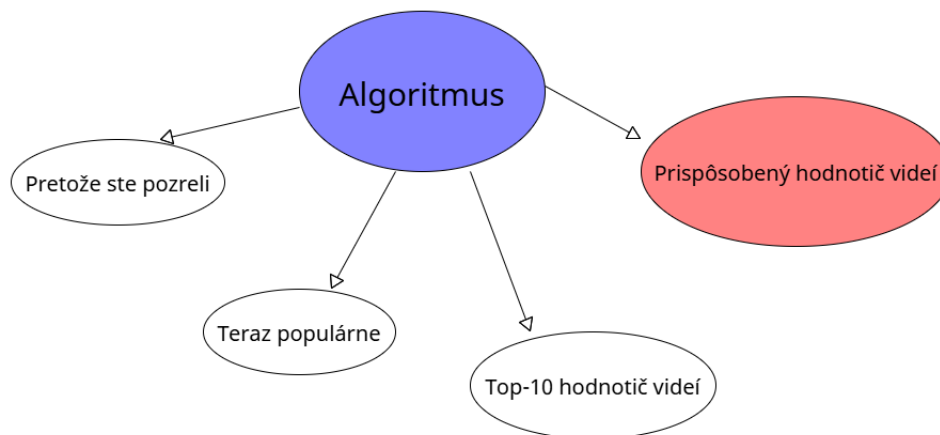
Ak človek nemá žiadne odporúčanie od kamaráta alebo známeho, je odkázaný prehľadávať veľké množstvo filmov a seriálov, o ktorých nevieme či sú dobré alebo nie. Jednoducho si o každom z filmov alebo seriálov dokážeme dohľadať hodnotenia a žánre, ale nikdy nevieme, či sa to bude páčiť práve nám. A preto prichádzajú "na scénu", veľmi potrebné algoritmy ktoré vďaka veľkému množstvu dát dokážu predpovedať ako veľmi sa nám bude daná šou páčiť, ukázať nám hodnotenia ľudí podobnej filmovej chute ako my a dokonca aj predpovedať aké hodnotenie jej dáme. "V Netflixu nieje jeden model ktorý poháňa všetky odporúčania ale skôr set techník ktoré sa navzájom prepájajú v rovnakom ciele - zvýšiť užívateľskú spokojnosť" [10]. No a teda aké všetky algoritmy vlastne používa Netflix na svojej platforme? Viac sa dozvieme v najbližších podsekcích. . .

3.1 Prispôbený hodnotič videí

Na hlavnej stránke môžete vidieť naozaj veľa filmov a seriálov. Sú ale rozdelené do riadkov, vždy podľa toho čo ich spája. Pre uvedenie príkladu, riadok komediálnych filmov sa pre nás vytvorí, ak sme si niekedy predtým pozreli nejakú komédiu. Samozrejme sa to odvíja od toho, aký člen domácnosti práve pozerá, tak isto od toho aké hodnotenie dal danej komédii a tak isto ako dlho ju pozeral a koľko takých komédií si pozrel. Na všetkom z tohto pracuje PVR (*Personalised Video Ranker*)² [4]

¹RMSE - root mean squared error - odmocnina priemerného rozdielu medzi reálnymi a očakávanými hodnotami [11]

²PVR - personal video ranker - Prispôbený hodnotič videí [4]



Obr. 1: Diagram popisujúci rozdielne časti algoritmu

3.2 Top-N hodnotič videí

Iný typ riadku ktorý je tiež na hlavnej stránke ale ukazuje trošku niečo iné. Jeho cieľ je taký istý - ukázať užívateľovi najlepší obsah. Pre zmenu, to teraz nefiltruje podľa iba určitého žánru ale z celého katalógu nám ukazuje to najlepšie čo bol schopný nájsť pre daného užívateľa a ukáže prvých 10, inokedy viac, odporúčaní, ktoré sú na začiatku "listu" ktoré vytvára tento algoritmus. [1]

3.3 Teraz populárne

Tento algoritmus je veľmi jednoduchý. Keďže nezohľadňuje takmer žiadnu personalizáciu, je najčastejšie používaný na začiatku resp. pre nových používateľov. Jeho hlavný zmysel je, ukázať obsah ktorý je práve populárny v danom ročnom období alebo časti roka. Tak isto môže zohľadňovať polohu a teda uprednostňovať obsah napr. zo Slovenska atď..

3.4 Pokračujte v pozeraní

Keďže je veľmi dôležité ukladať pokrok pozerania seriálov alebo filmov na pokračovanie, tento riadok je veľmi potrebný. Aj prvky v tomto riadku sú usporiadované pomocou vlastného algoritmu, tento algoritmus zohľadňuje čas ktorý ubehol od posledného pozretia daného obsahu, alebo aj či užívateľ pozrel podobný obsah.

4 Profitabilitnosť

Netflix má približne 235 miliónov platených užívateľov a úspech Netflixu vykvetol hlavne vďaka tomu, že začali používať vyššie spomenuté algoritmy skôr ako konkurencia. Keďže 80% sledovania na platforme tvorí sledovanie odporúčaného obsahu práve predtým spomenutými algoritmami, tak to znamená, že fungujú veľmi dobre a, že sa im vyplatia [9]. V podstate bez ohľadu na to, koľko by ich

to stálo, vyplatí sa im udržiavať a vyvíjať tieto algoritmy. Keďže za prehrávanie určitého obsahu, či už populárneho a dobre hodnoteného alebo nepopulárneho a menej dobre hodnoteného, platí Netflix prakticky tú istú sumu a teda im je to v podstate jedno, čo užívateľovi zobrazí, jediné na čom záleží je, že si to užívateľ pozrie, je spokojný a príde na platformu znova. [1]

Pre jednoduché porovnanie odporúčacieho systému Netflixu s napr. Spotify odporúčacím systémom. Môžeme vidieť veľmi veľký problém v tom, že Spotify uprednostňuje populárnu hudbu a obsah oproti tomu menej populárnemu, pretože z toho má väčší profit a teda je stavaný aj ich odporúčací systém okolo profitu [3]. Keďže každá firma chce mať profit, je to v pochopiteľné, ale problém je, že potom sú utláčaní nepopulárny umelci a majú problém vytvoriť zisk. Na rozdiel od Spotify, Netflix profituje za všetok obsah rovnako, je menej podstatné uprednostňovať populárny obsah. O tom, že Netflix je na tom finančne veľmi dobre a rastie aj naďalej oproti minulým rokom, hovorí aj ich výročná správa o profite [5].

Peňažná bilancia zisku Netflixu:

Dátum	Čistý zisk
31.12.2020	7,57 Mld \$
31.12.2021	12,69 Mld \$
31.12.2022	17,18 Mld \$
31.12.2023	22,59 Mld \$

[6]

Tento profit sa odvíja aj na hodnote ich akcie, tá bola v 01. novembra 2023 iba 420,19\$ ale tohto roka v ten istý dátum sa vyšplhala na neuveriteľných 756,10\$ za akciu. To je zvýšenie o takmer 80% za jeden rok. [7] A tak isto ich peňažná bilancia.

Enormný rozdiel hodnoty akcie Netflixu:

Dátum	Hodnota
01.11.2023	420,19 \$
01.11.2024	756,10 \$

[7]

5 Rozloženie hlavnej stránky

Keďže už sme si popísali ako fungujú algoritmy, teraz si môžeme popísať ako personalizovaný obsah rozkladá na stránku. Ako už predtým bolo spomenuté, užívateľ má krátku dobu pozornosti a Netflix preto rozdeľuje tento obsah do riadkov hneď na hlavnú stránku. Treba si ale uvedomiť že Netflix je zväčša používaný v domácnostiach a teda do napr. top 10 sa snaží dávať obsah ktorý si užije napr. otec, mama a ďalší členovia rodiny. Tak isto sa snaží nájsť položku pre celú rodinu. Ak je rodina jednočlenná, tak sa snaží nájsť výber z rozdielnych nálad a záľub daného užívateľa ktoré zisťujú dané algoritmy. [1]

Netflix chce aby sme vedeli, ako sa ich systém adaptuje našej chuti. Tak isto sa snaží vybudovať dôveru v tieto odporúčania a tiež sa snaží vysvetliť prečo by sme si to chceli pozrieť. Toto reprezentuje popisok k danému filmu/serialu, tak isto to ukazuje predikciu hodnotenia ake by sme dali danému filmu/seriálu.

6 Porovnanie HBO Max a Netflix

Kým Netflix sa snaží používať rozličné počítačové algoritmy a predikcie hodnotení ktoré získáva z našich akcií na platforme spomenutých v predchádzajúcich sekciách. Tak HBO Max sa snaží uchovať si iný prístup, a to tak, že implementuje aj ľudsky vybraný obsah. Tento prístup je podľa nich kľúčový k úspechu. A do budúcnosti, sa budú snažiť ešte viac zvýšiť tento ľudský dotyk pri výbere obsahu. A teda tento model odporúčania je bližšie k odporúčaciemu systému Spotify než Netflix. [2] Pre číselné porovnanie, Netflix má, ako už bolo spomenuté, 235 miliónov užívateľov, HBO Max má približne 73 miliónov. Veľkosť knižnice seriálov a filmov pre Netflix je 5831 a pre HBO Max to je 3235. [2] Keďže akcie HBO niesú verejne predajné, a samotný profit z HBO je zaštitený pod Warner Bros, nedostal som sa k dôveryhodným informáciám o profite z HBO Max. [12]

Ale ak porovnáme známe čísla používateľov a knižnice, dokážeme si vyvodiť porovnanie týchto konkurenčných firiem. Všimnúť sme si mohli, že Netflix má oveľa viac používateľov, pravdepodobne je to vďaka algoritmom ktoré očividne fungujú, ale tak isto, to môže byť tým, že Netflix, je na trhu oveľa dlhšie a teda, ak u ľudí vyvolal spokojnosť, ľudia nemajú potrebu skúšať nové platformy. To znamená, že Netflix, robí dobrú prácu s udržiavaním si užívateľov. Tak isto je ale dôležité podotknúť, že relatívne nová služba v porovnaní s Netflixom, dokázala za takú krátku dobu, získať tretinu používateľov ktorý používajú Netflix. Tak isto treba poznamenať, že veľké percento z týchto užívateľov pravdepodobne používa obe tieto platformy a teda sú v štatistike započítaní dva krát.

Záver

Touto prácou som prišiel na to, že odporúčací systém Netflixu, je veľmi komplexný, spolieha sa na niekoľko algoritmov, ktoré implementuje priamo do svojej stránky.

Porovnaním Netflixu s ďalšími platformami, som zistil, že ich systém, je veľmi efektívny, a v praxi funkčný. Iné platformy sa snažia vytvárať vlastné systémy a meniť ich originálne. Podľa čísel, je Netflix ale stále najpoužívanejší.

Taktiež analýzou samotného odporúčacieho systému Netflixu, som prišiel na to, že zbiera množstvo informácií od užívateľov, ale mal by ich používať len pre zlepšenie používateľovho zážitku na platforme. Samozrejme podľa výročných správ som prišiel na to, že rast samotnej firmy je kladný a teda môžeme usúdiť, že je to tak, aj vďaka odporúčaciemu systému.

Pokračovaním práce by mohlo byť spracovanie hlbšej analýzy rozličných algoritmov ako Netflixu, tak aj ďalších platforiem a ich porovnanie. Zaujímavé by bolo zistiť ako veľa údajov u užívateľov sa ukladá a ako veľa sa reálne použije nevinne.

Literatúra

- [1] X. Amatriain and J. Basilico. Recommender systems in industry: A netflix case study. In *Recommender systems handbook*, pages 385–419. Springer, 2015.

- [2] L. Arman and T. Ahmed. Identifying potential problem perceived by consumers within the recommendation system of streaming services, 2022.
- [3] T. Bachal, A. Kumar, A. Bachal, and V. S. Gutte. Tuning in to personalized music: A spotify api-based hybrid recommendation system integrating content-based and popularity-based approaches. In *2023 2nd International Conference on Futuristic Technologies (INCOFT)*, pages 1–6, 2023.
- [4] C. A. Gomez-Urbe and N. Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Trans. Manage. Inf. Syst.*, 6(4), Dec. 2016.
- [5] Netflix. 2023 annual report. https://s22.q4cdn.com/959853165/files/doc_financials/2023/ar/Netflix-10-K-01262024.pdf. Citované: 03.11.2024.
- [6] Netflix. Annual reports of netflix. <https://ir.netflix.net/financials/annual-reports-and-proxies/default.aspx>. Citované: 12.12.2024.
- [7] Netflix. Value of stock of netflix. <https://ir.netflix.net/stock-information/stock-chart/default.aspx>. Citované: 03.11.2024.
- [8] OpenAI. Chatgpt, 2024. Použitý v Abstracte pre formátovanie. Dostupné na: <https://chat.openai.com>.
- [9] D. Rastogi, T. S. Parihar, and H. Kumar. A parametric analysis of ava to optimise netflix performance. *International Journal of Information Technology*, 15(5):2687–2694, 2023. <https://link.springer.com/content/pdf/10.1007/s41870-023-01281-z.pdf>.
- [10] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, and J. Basilico. Deep learning for recommender systems: A netflix case study. *AI Magazine*, 42(3):7–18, Nov. 2021.
- [11] K. Stewart. Mean squared error. <https://www.britannica.com/science/mean-squared-error>. Citované: 03.11.2024.
- [12] WarnerBros. Annual reports of warner bros. <https://ir.corporate.discovery.com/financials/annual-reports-and-proxies/default.aspx>. Citované: 13.12.2024.