



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. [1]

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“. [2] Podrobne sa pozrieme aj na filtračné algoritmy, ktoré sa delia na dva hlavné typy [3]. 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 prispôsobiť obsah každému jednotlivému používateľovi.

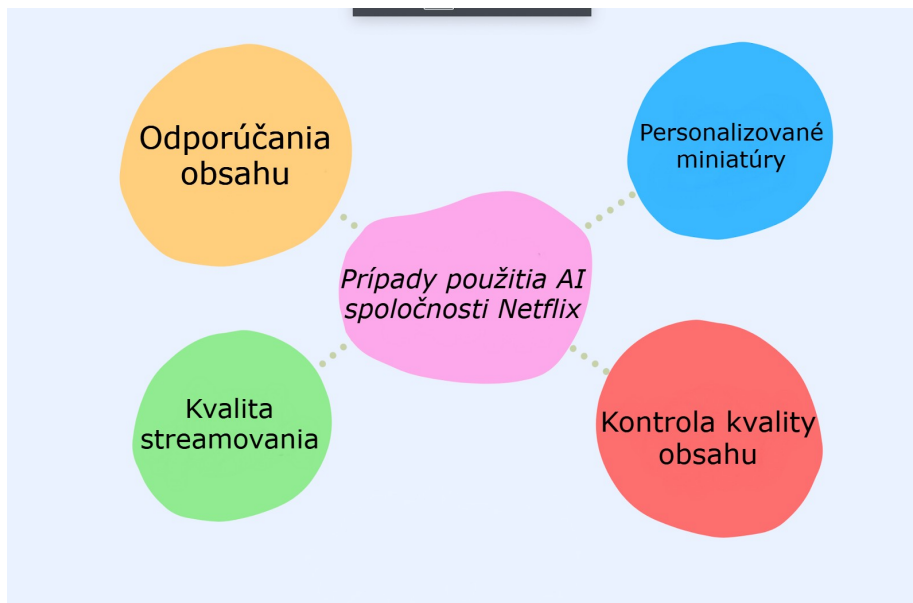
Netflix má obrovskú knižnicu obsahu a obmedzené rozhranie obrazovky, čo veľmi sťažuje odporúčanie relevantného obsahu, a každý používateľ má iný vkus, preferencie a zvyky pri sledovaní, takže personalizácia musí byť veľmi presná. Čo robí Netflix na dosiahnutie tohto cieľa, sa dozviete po tomto článku.

2 Technológie a Umelá inteligencia v Netflix

Na zlepšenie používateľského zážitku, Netflix využíva umelú inteligenciu a strojové učenie. Medzi kľúčové aplikácie týchto technológií patrí [4]

1. **Odporúčania obsahu:** na základe vašich zvykov a preferencií algoritmy spoločnosti Netflix môžu vám ponúknuť personalizované odporúčania. Analýza údajov obsahuje aj žáner a obľúbenosť obsahu. [5]
2. **Personalizované miniatúry:** Ak si zoberieme ako príklad film „Shadow and Bone“, miniatúry filmu sú prispôbené vašim záujmom a môžu používať obrázky zobraté s iných populárnych programov, skrátka čokoľvek, čo upúta vašu pozornosť. [6]
3. **Kvalita streamovania:** Na zabezpečenie plynulého prehrávania videa využíva Netflix údaje na predpovedanie preťaženia siete a ukladá obsah do medzipamäte na blízkych serveroch, takže si môžete vychutnať filmy bez oneskorenia. [6]
4. **Kontrola kvality obsahu:** Pred odporúčaním nového obsahu používateľom prechádza niekoľkými fázami testovania vrátane zvuku a obrazu, aby boli všetci používatelia spokojní. [6]

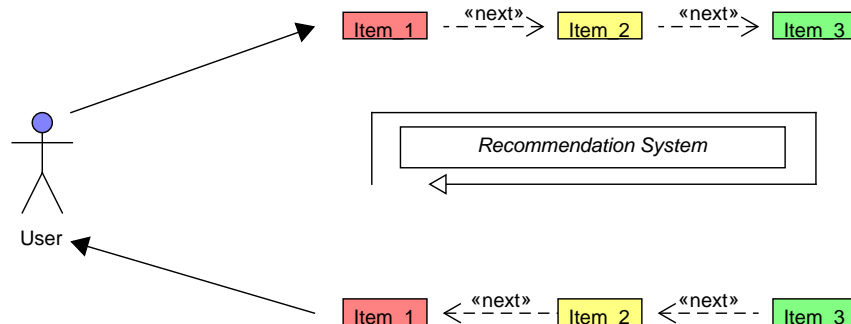
Táto technológia nielenže zvyšuje spokojnosť používateľov, ale podporuje aj dlhodobú angažovanosť.



Obr. 1: Systém umelej inteligencie

3 Typy Odporúčaní na Netflix

1. **Odporúčania obsahu:** na základe vašich zvykov a preferencií vám algoritmy spoločnosti Netflix môžu ponúknuť personalizované odporúčania. Analýza údajov zahŕňa aj žáner a obľúbenosť obsahu. [7]
2. **Kolaboratívne filtrovanie:** vytvorením skupín používateľov, ktorí majú podobný vkus, a na základe správania a preferencií používateľov táto metóda navrhuje obsah, ale má problém so studeným štartom [8].
3. **Filtrovanie na základe obsahu:** tento typ filtrovania zhromažďuje údaje, napríklad aký žáner filmu preferujete, akého režiséra alebo akých hercov uprednostňujete, a odporúča vám podobný obsah.
4. **Hybridné modely:** Tento typ kombinuje kolaboratívne filtrovanie a filtrovanie založené na obsahu, pretože len tak sa dá vyriešiť problém studeného štartu.
5. **Odporúčania založené na hlbokom učení:** Používajú sofistikované algoritmy umelej inteligencie, ktoré modelujú interakciu používateľa ako postupný rozhodovací proces.



Obr. 2: Systém odporúčaní

4 Filtračné algoritmy

Netflix používa rôzne typy filtrovania, ako je kolaboratívne filtrovanie, filtrovanie na základe obsahu [9], hybridné modely alebo odporúčania založené na hlbokom učení, ale v tomto článku sa budeme zaoberať len prvými dvoma, pretože sú najpoužívannejšie. Rozdiel medzi nimi spoznáte podľa názvu, ale poďme sa na ne pozrieť bližšie.

Kolaboratívne filtrovanie je algoritmus, ktorý sleduje, aké sú vaše preferencie pri výbere filmov, a na základe týchto údajov vás zoskupí s ostatnými ľuďmi, ktorí majú podobný vkus. Algoritmus využíva matematické modely na hodnotenie podobnosti a predikciu preferencií. Jedným z príkladov výpočtu skóre podobnosti môže byť vzorec:

$$\frac{\arcsin^2(x)}{2} - \frac{\sqrt{1-x^2} \cdot (x^2 + 2)}{3} + C$$

Tu x predstavuje normalizovanú hodnotu pre konkrétny atribút alebo interakciu používateľa, zatiaľ čo C je konštanta upravujúca celkové skóre. Takéto výpočty pomáhajú algoritmu rozhodnúť, ktoré filmy alebo seriály budú odporúčené používateľovi. Názornejšie, ako funguje, nájdete na Obr.3, kde je graficky znázornená metóda kolaboratívneho filtrovania. Hoci sú na obrázku len 2 ľudia, v skutočnosti ich môže byť 100 alebo viac. [10] Hlavným problémom tohto systému je problém studeného štartu, ktorý sa objavuje ako dôsledok nedostatku údajov o používateľovi alebo filmoch v systéme. Tento problém sa rieši použitím niekoľkých typov filtrovania, ale Netflix sa naďalej snaží tento problém vyriešiť. Metóda kolaboratívneho filtrovania sa delí na dve ďalšie metódy, prvou je **Kolaboratívne filtrovanie založené na používateľovi** (v tabuľke označene ako UBCF), zameriava sa na vyhľadávanie ľudí s podobným vkusom ako vy. [11] Ako presne? Najprv vypočíta podobnosť medzi všetkými používateľmi a v závislosti od výsledku začne odporúčať filmy, ktoré ste ešte nevideli, pričom sa páčia ľuďom s podobným vkusom. Druhou čiastkovou metódou je **Kolaboratívne filtrovanie založené na položkách** (IBCF), líši sa tým, že nehľadá podobnosti medzi ľuďmi, ale medzi položkami, s ktorými ste komunikovali, a

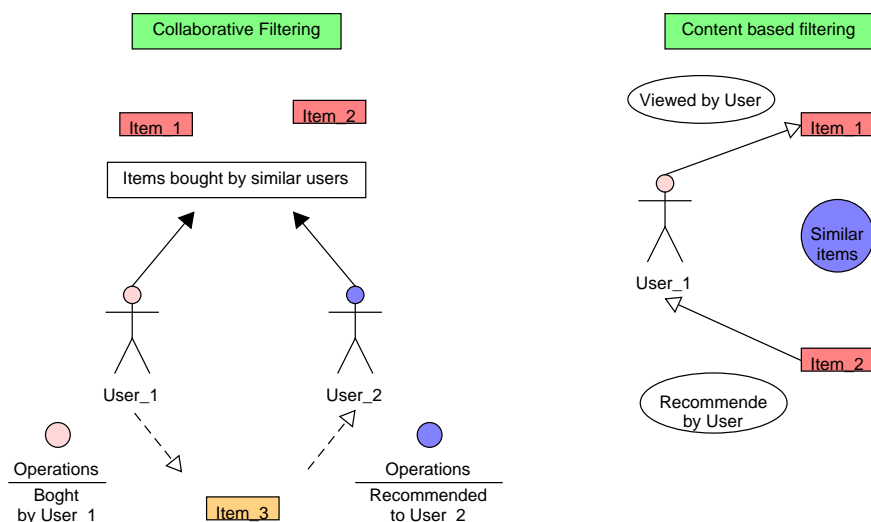
odporúča podobné. Ak sa vám napríklad páči film, v ktorom hral Robert Downey Jr, systém vám začne odporúčať filmy alebo seriály, v ktorých hral tento herec. Podrobnejšie, aký je medzi nimi rozdiel a aké aspekty zahŕňajú, si môžete pozrieť v nasledujúcej tabuľke.

Aspekt	Kolaboratívne filtrovanie na základe používateľov (UBCF)	Kolaboratívne filtrovanie na základe položiek (IBCF)
Zameranie	Odporúčania sú založené na podobnosti medzi používateľmi.	Odporúčania sú založené na podobnosti medzi položkami (produkty, filmy, atď.).
Vstupné dáta	Matica hodnotení používateľov (riadky reprezentujú používateľov, stĺpce reprezentujú položky).	Matica hodnotení používateľov (riadky reprezentujú používateľov, stĺpce reprezentujú položky).
Cieľ odporúčania	Na odporúčanie sa používajú používatelia, ktorí sú podobní cieľovému používateľovi.	Na odporúčanie sa používajú položky, ktoré sú podobné položkám hodnoteným cieľovým používateľom alebo s nimi
Logika algoritmu	1. Vypočíta sa podobnosť medzi cieľovým používateľom a všetkými ostatnými používateľmi. 2. Identifikujú sa používatelia podobní cieľovému. 3. Odporúčania sa generujú na základe položiek, ktoré tieto podobní používatelia hodnotili, ale cieľový používateľ ešte nie.	1. Vypočíta sa podobnosť medzi všetkými položkami. 2. Identifikujú sa položky podobné tým, ktoré cieľový používateľ hodnotil alebo s nimi interagoval. 3. Odporúčania sa generujú na základe podobných položiek.
Výpočet podobnosti	Typicky používa metriky ako kosínová podobnosť alebo Pearsonova korelačná podobnosť medzi hodnoteniami používateľov.	Typicky používa metriky ako kosínová podobnosť alebo Jaccardova podobnosť medzi charakteristikami položiek.
Riedkosť dát	UBCF môže byť menej účinný s riedkymi dátami, kde používatelia hodnotili len niekoľko položiek.	IBCF môže lepšie zvládať riedke dáta, pretože podobnosti medzi položkami sú zvyčajne hustejšie.
Škálovateľnosť	UBCF je menej výpočtovo náročný, ak sa pracuje s menším počtom používateľov.	IBCF je často škálovateľnejší, najmä ak sa pracuje s veľkým množstvom položiek.
Príklady použitia	1. Odporúčanie filmov na základe preferencií používateľov. 2. Odporúčania produktov na e-commerce platformách.	1. Odporúčanie podobných produktov k aktuálne prezeranému produktu. 2. Odporúčanie súvisiacich článkov v spravodajských aplikáciách.

Obr. 3: UBCF/IBCF

Filtrovanie na základe obsahu tento typ filtrovania je veľmi ľahko pochopiteľný, pretože podstatou algoritmu je odporučiť obsah, ktorý je podobný predchádzajúcim výberom používateľa, Obr.3. Algoritmus venuje pozornosť charakteristikám, ako sú „názov“, „krajina“, „režisér“, „obsadenie“, „uvedené“ a „opis“ filmu/seriálu, aby odporučil filmy, ktoré zodpovedajú vášmu vkusu. [12] Ak napríklad radi sledujete komédie, Netflix vám začne odporúčať filmy s Jackie Chanom alebo komédie s inými populárnymi hercami. Algoritmus tiež uvidí, že máte radi filmy s Jackie Chanom, a začne odporúčať filmy súvisiace s východnou kultúrou, a tak Netflix postupne zachytí vaše preferencie a bude sa stále

zlepšovať. Najväčším problémom tohto typu filtrovania je monotónnosť obsahu. V dôsledku toho používateľ stráca záujem o sledovanie filmov na tejto platforme. [13]



Obr. 4: Typy filtrovania

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.

1. **Personalizovaný rebríček videí (PVR)** je jedným z najmodernejších nástrojov [14] hodnotenia, ktoré Netflix používa. Pretože nielen chápe správanie používateľov v minulosti, ale dokáže predpovedať aj budúce potreby. Aby sme lepšie porozumeli tomuto algoritmu, pozrime sa, ako funguje, v prvom rade je založený na systéme hlbokého učenia a neustále sa prispôsobuje používateľovi. Základnou myšlienkou je použiť historické údaje (napríklad zvyky používateľov) na určenie vašich zvykov a uprednostnenie obsahu, ktorý vás zaujíma. Celkovo systém používa personalizované údaje používateľa na poskytnutie celkového relatívneho poradia položiek v rade.
2. **Systém Top-N Ranking** je založený na kombinácii kolaboratívneho a obsahového filtrovania. Ako už názov napovedá, poskytuje najlepšie hodnotené filmy, ktoré zodpovedajú vkusu používateľa. Je podobný ako PVR s tým rozdielom, že sa pozerá len na hodnotenie titulu. Netflix využíva Tento algoritmus aj na doplnenie riadkov, ako napríklad "Pre teba".

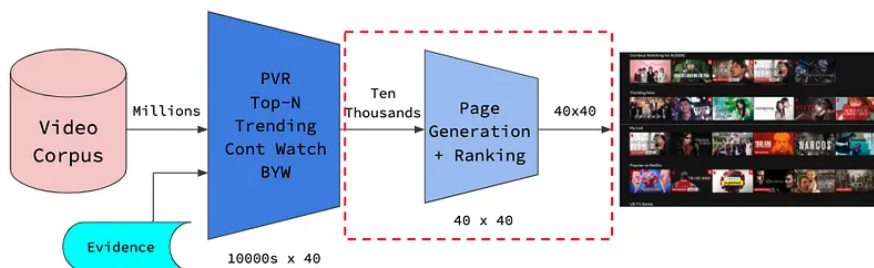
3. **Obľúbene medzi používateľmi**, ktorý predpovedá časové trendy na základe globálnych a regionálnych údajov o filmoch a televíznych seriáloch. Vo všeobecnosti existujú dva typy trendov: sezónne a špeciálne. Na základe názvu by ste mohli predpokladať, že čo je čo, ale poďme sa na ne pozrieť bližšie.

Do sekcie **sezónnych trendov** sa zvyčajne zaraďujú filmy ako *Sám doma* alebo *Polárny expres*, pretože získavajú popularitu v období Vianoc, alebo horory počas Halloween.

Sekcia **špeciálne** alebo **krátkodobé** udalosti môže patriť Pandémii 2021, ktorá zaznamenala nárast populárnosti dokumentárnych filmov a filmov o infekciách.

4. **Pozrite sa neskôr**, ide o pomerne zaujímavý algoritmus, ktorý je založený na dvoch kľúčových bodoch. Prvým je, že do tejto časti patria filmy, ktoré ste ešte nedopozerali, zvyčajne televízne seriály s niekoľkými sezónami, ale môžu to byť aj filmy alebo seriály, ktoré ešte neboli dokončené. A druhým bodom je, že na základe regionálnych údajov a pomocou kolaboratívneho filtrovania algoritmus zmieša vaše nepozreté filmy s podobnými filmami, ktoré sa páčia iným používateľom.

Po tom, ako algoritmy analyzujú vás a vaše správanie, začnú vytvárať vaše personalizované odporúčania, krok za krokom prostredníctvom všetkých parametrov, čo je graficky znázornené na obrázku 4. plynulo prechádzajú z backendu na frontend, takýto proces sa nazýva **generovanie stránok**.



Obr. 5: Hodnotenie

6 Typy Údajov v Netflix

A teraz nehovoríme len o tom, aké filmy uprednostňujete, ale o výsledkoch systému, ktorý má viac ako 1 300 skupín odporúčaní skenujúcich vás a vaše zvyky. Netflix vie, že máte radi nielen thrillery; on vie že zaoberáte sa psychologické thrillery so silnými ženskými hrdinkami, ktoré sledujete väčšinou cez víkend večer. Asi vás zaujíma ako. Odpoveď nájdete v tejto časti.

1. **Údaje o účtoch a predplatnom**, na prvý pohľad nie je jasné, prečo Netflix používa tieto údaje, ale je to na lepšiu správu vašich účtov a zaistenie bezpečnosti monitorovaním vašej aktivity na účte.

2. **Zariadenie a dáta**, z ktorého sledujete filmy a v akom čase, aby sú vedeli, čo a kedy vám máme odporučiť. [15] Vezmime si, že v práci počas prestávky dve hodiny pozeráte komediálne seriály a potom skončí a večer si na televízore zapnete Netflix, kde pozeráte horory, zhromažďovaním týchto informácií sa vaše odporúčania začnú meniť podľa dátumu a času.
3. Spoločnosť Netflix zhromažďuje údaje ob **históriu dopytov**, aby mohla sledovať trendy a prispôbovať ponuky obsahu s cieľom zdôrazniť to, čo používatelia hľadajú.
4. Samozrejme, že spracúvajú **hodnotenie obsahu a spätnú väzbu** používateľov, aby pochopili kvalitu a hodnotenie obsahu a rozhodli sa, či obsah spracovať alebo odstrániť.

7 Záver

Toto je koniec článku a stojí za to ho zhrnúť. Netflix je rýchlo rastúca platforma, ktorá stojí na vrchole odporúčacieho reťazca vďaka svojim algoritmom a pokročilej umelej inteligencii, ktorá je založená na hlbokom učení. Pomocou rôznych metód filtrovania, umelej inteligencie a systémov hodnotenia, ako sú PVR, Top-N Ranking atď. dokáže vytvoriť pre vás najrelevantnejší a najvhodnejší obsah.

Okrem toho Netflix využíva zhromaždené údaje, ako je zariadenie, čas sledovania a poloha, aby pochopila váš vkus a poskytla vám a ostatným používateľom najvhodnejší obsah. Netflix pracuje aj s rozhraním a poskytuje vám miniatúry, ktoré vytvárajú prvý dojem z filmu alebo seriálu a udržiavajú váš záujem. [16]

Netflix v súčasnosti skúma problém studeného štartu s cieľom vyriešiť ho. Týmto spôsobom sa spoločnosť Netflix snaží vytvoriť čo najpríjemnejší zážitok pre každého používateľa, aby dosiahla väčší úspech na trhu. Okrem toho pokračuje v štúdiu a zdokonaľovaní umelej inteligencie a techník filtrovania na presnejšiu personalizáciu obsahu.

Literatúra

- [1] Arkadiusz Krysiak. Netflix algorithm: How netflix uses ai to improve personalization. <https://stratoflow.com/how-netflix-recommendation-algorithm-work/>, jun. 2024.
- [2] Arkadiusz Krysiak. Netflix algorithm: Everything you need to know about the recommendation system of the most popular streaming portal. <https://recostream.com/blog/recommendation-system-netflix>, aug. 2021.
- [3] Ms Phalle and Prof Bhushan. Content based filtering and collaborative filtering: A comparative study. *Journal of Advanced Zoology*, 45:96–100, mar. 2024.
- [4] Bogdan Walek and Vladimir Fojtik. A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications*, 158:113452, nov. 2020.

- [5] Jinchao Huang, Zhipu Xie, Han Zhang, Bin Yang, Chong Di, and Runhe Huang. Enhancing knowledge-aware recommendation with dual-graph contrastive learning. *Information*, 15:534, sep. 2024.
- [6] Harald Steck, Linas Baltrunas, Ehtsham Elahi, Dawen Liang, Yves Raimond, and Justin Basilico. Deep learning for recommender systems: A netflix case study. *AI Magazine*, 42:7–18, nov. 2021.
- [7] Souptik Saha and S. Ramamoorthy. User centric and collaborative movie recommendation system under customized platforms. pages 699–704, may 2021.
- [8] Sreenivas Mekala, Ch Reddy, V Kumar, and D Reddy. To solve cold start problem in collaborative filtering recommender system. 12:458–463, jun. 2023.
- [9] Xiaoyuan Su and Taghi Khoshgoftaar. A survey of collaborative filtering techniques. *Adv. Artificial Intelligence*, 2009, okt. 2009.
- [10] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. Republic and Canton of Geneva, CHE, jul. 2017. International World Wide Web Conferences Steering Committee.
- [11] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. pages 263–272, dec. 2008.
- [12] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. *Content-based Recommender Systems: State of the Art and Trends*, pages 73–105. feb. 2011.
- [13] Michael Pazzani and Daniel Billsus. *Content-Based Recommendation Systems*, pages 325–341. feb. 2007.
- [14] Mehdi Elahi, Amin Beheshti, and Srinivasa Goluguri. *Recommender Systems: Challenges and Opportunities in the Age of Big Data and Artificial Intelligence*. jul. 2021.
- [15] Linas Baltrunas. Exploiting contextual information in recommender systems. pages 295–298, okt. 2008.
- [16] Pradeep Singh, Pijush Dutta Pramanik, Avick Dey, and Prasenjit Choudhury. Recommender systems: An overview, research trends, and future directions. *International Journal of Business and Systems Research*, 15:14–52, feb. 2021.
- [17] DeepL. <https://www.deepl.com/en/translator>, 2024. Translations of several parts of the text.