

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-5220-64362

Bc. Ondrej Kaššák

SKUPINOVÉ ODPORÚČANIE PRE INTELIGENTÚ TELEVÍZIU

Diplomová práca

Študijný program: Softvérové inžinierstvo

Študijný odbor: 9.2.5 Softvérové inžinierstvo

Miesto vypracovania: Ústav informatiky a softvérového inžinierstva, FIIT STU

Vedúci práce: Ing. Michal Kompan, PhD.

máj 2014

Zadanie diplomovej práce

Meno študenta: **Bc. Ondrej Kaššák**

Študijný program: Softvérové inžinierstvo

Študijný odbor: Softvérové inžinierstvo

Názov práce: **Skupinové odporúčanie pre inteligentnú televíziu**

Samostatnou výskumnou a vývojovou činnosťou v rámci predmetov Diplomový projekt I, II, III vypracujte diplomovú prácu na tému, vyjadrenú vyššie uvedeným názvom tak, aby ste dosiahli tieto ciele:

Všeobecný cieľ:

Vypracovaním diplomovej práce preukážete, ako ste si osvojili metódy a postupy riešenia relatívne rozsiahlych projektov, schopnosť samostatne a tvorivo riešiť zložité úlohy aj výskumného charakteru v súlade so súčasnými metódami a postupmi študovaného odboru využívanými v príslušnej oblasti a schopnosť samostatne, tvorivo a kriticky pristupovať k analýze možných riešení a k tvorbe modelov.

Špecifický cieľ:

Vytvorte riešenie, zodpovedúce návrhu textu zadania, ktorý je prílohou tohto zadania. Návrh bližšie opisuje tému vyjadrenú názvom. Tento opis je záväzný, má však rámcový charakter, aby vznikol dostatočný priestor pre Vašu tvorivosť.

Riadte sa pokynmi Vášho vedúceho.

Pokiaľ v priebehu riešenia, opierajúc sa o hlbšie poznanie súčasného stavu v príslušnej oblasti alebo o priebežné výsledky Vášho riešenia alebo o iné závažné skutočnosti, dospejete spoločne s Vaším vedúcim k presvedčeniu, že niečo v texte zadania a/alebo v názve by sa malo zmeniť, navrhnete zmenu. Zmena je spravidla možná len pri dosiahnutí kontrolného bodu.

Miesto vypracovania: Ústav informatiky a softvérového inžinierstva FIIT STU v Bratislave

Vedúci práce: **Ing. Michal Kompan**

Termíny odovzdania:

podľa harmonogramu štúdia platného pre semester, v ktorom máte príslušný predmet (Diplomový projekt I, II, III) absolvovať podľa Vášho študijného plánu

Predmety odovzdania:

V každom predmete dokument podľa pokynov na www.fiit.stuba.sk v časti:
home > Informácie o > štúdiu > organizácia štúdia > diplomový projekt

V Bratislave dňa 18. 2. 2013



prof. Ing. Pavol Návrat, PhD.
riaditeľ Ústavu informatiky a softvérového
inžinierstva

Návrh zadania diplomovej práce

Finálna verzia do diplomovej práce ¹

Študent:

Meno, priezvisko, tituly: Ondrej Kaššák, Bc.
Študijný program: Softvérové inžinierstvo
Kontakt: xkassak@is.stuba.sk

Výskumník:

Meno, priezvisko, tituly: Michal Kompan, Ing.

Projekt:

Názov: Skupinové odporúčanie pre inteligentnú televíziu
Názov v angličtine: Group Recommendation for Smart Television
Miesto vypracovania: Ústav informatiky a softvérového inžinierstva, FIIT STU, Bratislava
Oblasť problematiky: Softvérové systémy pre personalizované odporúčanie

Text návrhu zadania ²

Personalizované odporúčanie predstavuje spôsob ako sa vysporiadať s množstvom informačného zahltenia, ktorému v súčasnosti čelíme. Môže mať veľký význam napríklad v doméne multimediálneho obsahu, kde položku nevieme dobre posúdiť kým si ju neprehráme (textové opisy sú často nepostačujúce), čo môže pri hľadaní zabráť veľa času. Viaceré činnosti pri ktorých je vhodné generovať odporúčania vykonávame skôr v skupine ako jednotlivo. Toto je zrejme najmä v doméne multimédií kedy sa ako skupina rozhodujeme aký film si pozrieme v TV, kine, prípadne aká hudba bude hrať a pod.

V dnešnej dobe pozorujeme tiež fenomén vzájomného prepájania zariadení a platforiem, čím vznikajú nové možnosti pre odporúčacie systémy. Prepájanie mobilných zariadení, počítačov či televízie často predstavuje možnosť ako o používateľoch získať väčšie množstvo informácií bez nutnosti interakcie, čím nám dáva možnosť vylepšiť generované odporúčania.

Analyzujte existujúce riešenia skupinového odporúčania a opíšte prínosy jednotlivých prístupov pre odporúčanie v doméne multimediálneho obsahu. Preskúmajte možnosť získania a/alebo odvodenia informácií o používateľoch a ich preferenciách z prostredia sociálneho webu a ich následnej aplikácie v procese odporúčania. Navrhňte prístup pre personalizované odporúčanie sa zohľadnením sociálneho kontextu pre jednotlivcov a skupiny používateľov, vhodný pre odporúčanie multimediálneho obsahu v obmedzenom čase. Experimentálne overte navrhnutý prístup na netriviálnej vzorke dát z oblasti multimediálneho obsahu.

¹ Vytlačiť obojstranne na jeden list papiera

² 150-200 slov (1200-1700 znakov), ktoré opisujú výskumný problém v kontexte súčasného stavu vrátane motivácie a smerov riešenia

Literatúra³

- Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. In: User Modeling and User-Adapted Interaction. Vol 14. Kluwer Academic Publishers. 2004, Feb, pp. 37-85.
- Senot, C., Kostadinov, D., Bouzid, M., et al.: Analysis of Strategies for Building Group Profiles. In: User Modeling, Adaptation, and Personalization. Vol. 6075. Springer Berlin. 2010, pp. 40-51.

Vyššie je uvedený návrh diplomového projektu, ktorý vypracoval(a) Bc. Ondrej Kaššák, konzultoval(a) a osvojil(a) si ho Ing. Michal Kompan a súhlasí, že bude takýto projekt viesť v prípade, že bude pridelený tomuto študentovi.

V Bratislave dňa 9.1.2013



Podpis študenta



Podpis výskumníka

Vyjadrenie garanta predmetov Diplomový projekt I, II, III

Návrh zadania schválený: áno / nie⁴

Dňa: 1.2.2013



Podpis garanta predmetov

³ Z vedecké zdroje, každý v samostatnej rubrike a s údajmi zodpovedajúcimi bibliografickým odkazom podľa normy STN ISO 690, ktoré sa viažu k téme zadania a preukazujú výskumnú povahu problému a jeho aktuálnosť (uveďte všetky potrebné údaje na identifikáciu zdroja, pričom uprednostnite vedecké príspevky v časopisoch a medzinárodných konferenciách)

⁴ Nehodiace sa prečiarknite

Anotácia

Slovenská technická univerzita v Bratislave

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ

Študijný program: SOFTVÉROVÉ INŽINIERSTVO

Autor: Bc. Ondrej Kaššák

Diplomová práca: Skupinové odporúčanie pre inteligentnú televíziu

Vedúci diplomovej práce: Ing. Michal Kompan, PhD.

máj 2014

Práca sa zaoberá problematikou personalizovaného odporúčania skupinám používateľov v doméne multimedialného obsahu. V rámci tejto domény sa sústreďuje na odporúčanie pre inteligentnú televíziu v zmysle odporúčania filmov z rozsiahlych databáz. Široká ponuka dostupného obsahu v takýchto televíziách a ich archívoch zapríčiňuje vznik problému informačného zahltenia. Personalizované odporúčanie predstavuje jedno z riešení tohto problému. Keďže sledovanie televízie je často skupinová aktivita, je žiaduce odporúčanie prispôbiť priamo pre potreby celých skupín namiesto odporúčania pre jednotlivcov.

V práci analyzujeme existujúce prístupy a nástroje venujúce sa problematike skupinového odporúčania multimedialného obsahu. Navrhujeme vektorový model používateľa uvažujúci rozličnú dôležitosť jednotlivých vektorov. Ich dôležitosť zakladáme na individuálnej preferencii typov metadát používateľom a na globálnych váhach vektorov. Ďalej navrhujeme vlastnú hybridnú metódu využívajúcu agregované kolaboratívne odporúčanie jednotlivým členom skupiny doplnené odporúčaním jednotlivcom na základe obsahu. Metóda berie do úvahy sociálny kontext členov skupiny v podobe vzájomných vzťahov členov reprezentovaných prepojeniami na sociálnych sieťach. Okrem odporúčania skupinám je využiteľná taktiež pre potreby personalizovaného odporúčania jednotlivcom.

Annotation

Slovak University of Technology Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Degree Course: SOFTWARE ENGINEERING

Author: Ondrej Kaššák

Master's Thesis: Group Recommendation for Smart TV

Supervisor: Dr. Michal Kompan

2014, May

This thesis deals with personalized recommendations to groups of users in the domain of multimedia content. Within this domain it specializes in recommendation for smart television in meaning of recommendation movies from big databases. Wide range of available content in these televisions and their archives gives rise to the information overload problem. Personalized recommendation represents one of possible solutions of this problem. Since watching television is often a group activity, it is appropriate to adapt recommendation to needs of groups rather than to individuals.

In this thesis we analyze existing approaches and systems dealing with the issue of group recommendation of multimedia content. We design a vector user model using different importance of individual vectors. The importance of them is based on user's individual preference of contained metadata types and on the global vector weights. We then propose our mixed hybrid method using aggregated collaborative recommendation for individual group members, complemented by aggregated individual content based recommendation. The proposed method considers social context, in meaning of relationships between group members represented by connections on social networks. Next to the group recommendation is method usable for purposes of personalized recommendation for individual users.

Obsah

1	Úvod	1
2	Personalizované odporúčanie.....	3
2.1	Model používateľa	3
2.2	Prístupy k personalizovanému odporúčaniam.....	6
2.2.1	Kolaboratívne odporúčanie	6
2.2.2	Odporúčanie založené na obsahu	8
2.2.3	Ďalšie typy odporúčaní	9
2.2.4	Hybridné odporúčacie prístupy	10
3	Charakteristiky skupín.....	17
3.1	Typy skupín.....	17
3.2	Predmet agregácie	18
3.3	Agregačné stratégie	19
3.4	Sociálne vzťahy v skupine	20
3.5	Metadáta opisujúce multimediálny obsah	21
4	Existujúce systémy	23
4.1	PolyLens.....	23
4.2	Yu's TV.....	24
4.3	Avatar	25
4.4	FIT-Family	25
4.5	Happy Movie.....	26
4.6	gRecs	26
4.7	Addaptive correlation-based RS.....	26
5	Ciele práce	27
6	Návrh metódy skupinového odporúčania	29
6.1	Reprezentácia multimediálneho obsahu	30
6.2	Model používateľa	32
6.3	Proces odporúčania	34
6.3.1	Kolaboratívne odporúčanie jednotlivcom	34
6.3.2	Odporúčanie jednotlivcom na základe obsahu.....	35
6.3.3	Hybridné odporúčanie skupine	35
6.4	Prezentácia odporúčania skupine	36
7	Realizácia metódy	37
7.1	Zdroje dát	38
7.2	Model používateľa	38
7.3	Odporúčanie	39
7.3.1	Kolaboratívne odporúčanie jednotlivcom	39
7.3.2	Odporúčanie jednotlivcom na základe obsahu.....	40
7.3.3	Hybridné odporúčanie skupine	41
8	Overenie metódy	43
8.1	Model používateľa	43
8.2	Hybridná odporúčacia metóda	46
8.3	Hybridná metóda skupinového odporúčania.....	48
8.3.1	Syntetický experiment.....	48

8.3.2	Reálny experiment.....	50
9	Zhodnotenie	55
	Literatúra.....	57
Príloha A.	Technická dokumentácia	61
A.1.	Diagram prípadov použitia.....	61
A.2.	Sekvenčný diagram	65
A.3.	Diagram aktivít.....	67
A.4.	Diagram tried.....	68
Príloha B.	Príspevok zaslaný na konfer. RECSYS 2014	73
Príloha C.	Príspevok publikovaný na konfer. IIT.SRC 2014	79
Príloha D.	Návrh článku do časopisu	87
Príloha E.	Obsah elektronického média	107

1 Úvod

Oblasť multimediálneho obsahu je rozsiahla, dynamicky sa zväčšujúca doména. Obsahuje množstvo zaujímavých položiek, ktoré majú potenciál človeka zabaviť, vzdeláť či rozšíriť jeho obzory atraktívnou formou. Konzumovanie multimediálneho obsahu je pre mnohých ľudí navyše každodenná činnosť, ktorej venujú nezanedbateľnú časť svojho voľného času. Vývoj riešení, ktoré pomôžu svojim používateľom vybrať pre nich zaujímavý obsah, preto pomôže používateľom tráviť svoj čas kvalitnejšie a plnohodnotnejšie. V našej práci sa zameriavame na subdoménu, s ktorou prichádzajú používatelia do styku najviac, teda na oblasť televízií, presnejšie na rozvíjajúce sa inteligentné televízie (angl. Smart TV), pretože práve v ich prípade je priestor pre personalizáciu obsahu. Prieskumy¹ ukazujú, že priemerný občan U.S.A. strávi denne sledovaním televízie 5,11 hodiny, za celý život potom sleduje televíziu 9 rokov. Priemerné školopovinné dieťa v U.S.A. pritom sleduje televíziu 1 200 hodín ročne, pričom pre porovnanie v škole strávi ročne 900 hodín. Z týchto údajov vidno, že televízny obsah tvorí významnú časť informácií, ktoré používateľ denne prijíma.

Sledovanie televízie predstavovalo v minulosti výber z obsahu, ktorý nadiktovali tvorcovia vysielania. Používatelia si často volili najmenšie zlo v podobe jediného z ich pohľadu sledovateľného programu. Čiastočné zlepšenie situácie nastalo vďaka zvýšeniu počtu dostupných staníc. Tu používatelia prvýkrát pocítili dilemu, ktorý program si zo zaujímavého výberu majú zvoliť, prípadne strávil zbytočný čas prepínaním medzi dostupnými kanálmi. Tretí stupeň vývoja nastal s príchodom inteligentných televízií. Tie totiž umožnili používateľom okrem bohatého výberu televíznych staníc taktiež možnosť výberu z rozsiahlych archívov obsahujúcich veľké množstvo položiek.

Výhoda, že si používatelia môžu pozrieť takmer čokoľvek sa tu však často mení na komplikáciu. Problém, ktorý nastáva, spočíva v tom, že si používateľ vyberá obsah z množiny tak rozsiahlej, že o existencii väčšiny jej položiek ani netuší. Týmto sa dostávame k problému informačného zahltenia, ktorý predstavuje odvrátenú stránku čoraz lepšej dostupnosti informácií spôsobenej technologickými možnosťami ľudstva. Práve preto sa naskytá výzva, ktorú máme za cieľ pomôcť zmierniť.

Personalizované odporúčanie predstavuje efektívne riešenie, ktoré môžeme využiť na redukciu spomenutého problému informačného zahltenia. Jeho využitie v doméne inteligentnej televízie môže spočívať v personalizovanom výbere obsahu z databázy archivovaných položiek ale taktiež z personalizácie samotného vysielania. Tu napríklad televízia divákovi vyberie z programov vysielaných v rovnakom čase taký, ktorý ho môže najviac zaujímať, prípadne televízia z večerných správ vysielaných na viacerých staniciach vyberie divákovi príspevky, o ktorých predpokladá, že ho budú zaujímať najviac.

Sledovanie televízie však predstavuje taktiež spoločenskú aktivitu (Cesar a kol., 2008). Podľa Masthoff, by sa preto adaptívne televízie mali prispôbovať skupinám používateľov, skôr ako jednotlivcom (Masthoff, 2004). To znamená, že by mali dokázať ponúknuť obsah, ktorý je čo najvhodnejší pre konkrétnu skupinu používateľov súčasne.

Fakt, že človek je spoločenský tvor je všeobecne známy. Liu a kol. dokonca tvrdia, že skupinové aktivity sú základnými ingredienciami spoločenského života ľudí (Liu a kol., 2012).

¹ BLS American Time Use Survey, A.C. Nielsen Co., dátum zverejnenia údajov 2. júl 2012

V našej práci zameriavame práve na personalizované odporúčanie skupinám používateľov. V tejto oblasti totiž vidíme veľký potenciál a zároveň dostatočný priestor pre výskum. Príkladom môže byť vytvorenie modelu používateľa zohľadňujúceho okrem záujmov tiež používateľov spôsob výberu položiek, prípadne využitie hybridných prístupov ku skupinovému odporúčaniam súčasne využívajúcich silné vlastnosti viacerých odporúčacích prístupov, prípadne zvyšujúcich spokojnosť používateľa.

V našej práci sa zaoberáme problematikou personalizovaného odporúčania skupinám v doméne multimediálneho obsahu so špecializáciou na subdoménu televízneho obsahu. Navrhujeme hybridnú metódu (angl. Mixed Hybrid Method) využívajúcu kolaboratívne odporúčanie multimediálneho obsahu doplnené odporúčaním na základe obsahu slúžiace na preusporiadanie kolaboratívnych výsledkov za účelom zvýšenia celkovej presnosti.

V práci v kapitole *Personalizované odporúčanie*, analyzujeme spôsoby reprezentácie preferencií používateľa a existujúce prístupy k personalizovanému odporúčaniam. V kapitole *Charakteristiky skupín* opisujeme možnosti odporúčania pre skupiny, typy skupín, spôsoby agregácie preferencií členov skupiny a sociálny vplyv členov. Kapitola *Existujúce systémy* približuje metódy odporúčania, agregácie a charakter skupín využívaných v existujúcich riešeniach skúmanej domény. Ciele, na ktoré sa v práci zameriavame sú zhrnuté v kapitole *Ciele práce*. Návrh modelu používateľa a hybridnej metódy skupinového odporúčania, spolu s jej komponentmi medzi ktoré zaradíme kolaboratívne a obsahové odporúčanie jednotlivcom a agregáciu odporúčaní sú uvedené v kapitole *Návrh metódy skupinového odporúčania*. Implementácia je opísaná v kapitole *Realizácia metódy*. Experimentálne overenie modelu používateľa, hybridnej metódy odporúčania a skupinovej hybridnej metódy odporúčania sú zhrnuté v kapitole *Overenie metódy*. Zhrnutie výsledkov práce uvádzame v kapitole *Zhodnotenie*.

2 Personalizované odporúčanie

Prvé systémy zamerané na personalizované odporúčanie obsahu vznikali už v 90-tych rokoch 20. storočia (Goldberg a kol., 1992). Patrili medzi ne systémy spoločností Xerox, Microsoft, MIT a Amazon. Ich účelom bolo zredukovať množstvo informačného zahltenia, ktorému boli používatelia nútení čeliť. Boli určené na odporúčanie obsahu jednotlivcom. Pomocou rozličných techník dokázali používateľom pomôcť vyhľadať relevantné položky, z rozsiahlych domén, ktoré by inak boli pre človeka, kvôli ich rozsahu len ťažko spracovateľné. Príkladom môžu byť digitálne knižnice, databázy obsahujúce množstvo multimediálneho obsahu, rozsiahle databázy s textovými dokumentmi a mnohé iné. Výskum v tejto oblasti naďalej aktívne prebieha, čo dokazuje neustávajúcu potrebu jednotlivé prístupy zdokonaľovať a prispôbovať novým požiadavkám, ktoré sú na ne kladené.

Samotné personalizované odporúčanie jednotlivcom však nedokáže pokryť všetky oblasti, v ktorých existuje záujem ho využívať. Ľudia sú totiž spoločenský jedinci a množstvo svojich aktivít vykonávajú spoločne s inými. Z tohto dôvodu nám personalizované odporúčanie zamerané len na jednotlivcov nepostačuje. Pri odporúčaní skupinám totiž platí, že skupina je viac ako súčet jej členov. Z tohto dôvodu sa na problematiku musíme pozrieť viac do hĺbky.

Riešením je napríklad skupinové odporúčanie. To predstavuje formu odporúčania, ktorá je určená pre skupinu ľudí. Skupina je množina osôb nachádzajúcich sa v rovnakom čase na rovnakom mieste, ktorá spoločne vykonáva určitú činnosť. Odporučený obsah je väčšinou jednotný pre všetkých členov skupiny a predpokladá sa, že skupina ho bude konzumovať naraz a spoločne. Takto vybraný obsah teda predstavuje určitú formu kompromisu medzi záujmami jednotlivých členov skupiny. To však vyvoláva otázku či je takéto odporúčanie možné pokladať za personalizované. Na prvý pohľad sa totiž môže zdať, že obsah vybraný pre viacero používateľov súčasne, by sa ako personalizovaný označovať nemal a teda, že sa jedná o samostatnú oblasť odporúčania. Pokiaľ si však uvedomíme, že odporúčame obsah, ktorý vyberáme pre každého člena zvlášť a až v rámci neho hľadáme taký, čo vyhovuje všetkým členom skupiny súčasne, môžeme skupinové odporúčanie označiť za podmnožinu personalizovaného odporúčania.

Nezávisle od zaradenia tohto typu odporúčania, o ňom v súčasnej dobe prebieha, rovnako ako pri odporúčaní jednotlivcom, aktívny výskum. Problematika aj vďaka komplexnosti svojho rozsahu poskytuje množstvo problémov, ktoré je potrebné riešiť. V tejto kapitole postupne opisujeme základné princípy, na ktorých je skupinové odporúčanie založené.

2.1 Model používateľa

Každý človek je jedinečný. Má svoju vlastnú osobnosť, záujmy, veciam pripisuje rôznu dôležitosť v porovnaní s ostatnými ľuďmi. Pri odporúčaní je preto v prvom rade nutné poznať preferencie používateľa. V prípade skupinového odporúčania je nutné poznať preferencie každého jej člena. Na tento účel slúži model používateľa (angl. User model). Pomocou neho reprezentujeme informácie o tom, čo používateľa zaujíma, aký obsah preferuje, či aká je osobnosť.

Podľa Senota a kol. je model používateľa v jednoduchom ponímaní tvorený množinou dvojíc [kategória, hodnota], pričom hodnotu predstavuje číslo z intervalu $<0,1>$, ktoré hovorí o záujme používateľa pre danú kategóriu položiek (napríklad komédií). Na základe času, ktorý používateľ venuje pozeraniu položiek z danej kategórie alebo priemernej výšky hodnotení položiek z danej kategórie vieme určiť jeho záujem a tým postupne modelujeme jeho profil (Senot a kol., 2010).

Títo autori model presnejšie definujú pomocou atribútov *hodnota príslušnosti*, *hodnota potreby* a *hodnota záujmu*. *Hodnota príslušnosti* (angl. Quantity of Affiliation - QoA) predstavuje

stupeň príslušnosti obsahu položky k určitému sémantickému konceptu (film „X-men: First class“, množina QoA {akčný = 0.9, sci-fi = 0.7}). *Hodnota spotreby* (angl. Quantity of Consumption - QoC) predstavuje stupeň spotreby položky používateľom, čiže napríklad akú časť filmu videl a nakoľko ho daná téma zaujala (film „X-men: First class“, QoC {132/132 min, 5/5 hodnotenie}). *Hodnota záujmu* (Quantity of Interest - QoI) stupeň záujmu používateľa daným sémantickým konceptom (komédia = 0.92). Vytvorenie modelu používateľa v tomto ponímaní pozostáva pre každého používateľa zo spojenia QoC pre množinu položiek hodnotených používateľom do množiny QoI daného používateľa, ktorá tvorí samotný model (Senot a kol., 2010).

Prehľad existujúcich prístupov k modelovaniu používateľa v doméne multimediálneho obsahu vo svojej práci opisujú O'Connor a kol. (O'Connor a kol., 2001). Títo autori spomínajú základné prístupy, ktorými boli preferencie používateľov uchovávané najčastejšie v minulosti, ako aj v súčasnosti využívané spôsoby. Prvý z opisovaných prístupov bol využívaný v najmä minulosti. Jeho princíp je založený na využití Bayesovských sietí, pričom model používateľa je reprezentovaný ako sieť, ktorá má v každom svojom vrchole rozhodovací strom a ktorej hrany reprezentujú konkrétne informácie o používateľovi (Breese a kol., 1998). V tomto období vznikali tiež iné prístupy. Ungar a kol. napríklad využívali princíp agregácie najpodobnejších používateľov do spoločných zhlukov, ktoré predstavovali množiny podobných používateľov využívaných pri kolaboratívnom filtrovaní (Ungar a kol., 1998). Aggarwal a kol., na modelovanie používateľa využívali grafy, ktorých vrcholy predstavovali používateľov a váhované hrany vyjadrovali mieru podobnosti medzi dvojicami takýchto vrcholov. Modely boli využívané v kolaboratívnom odporúčaní, kde boli do úvahy brané len najbližšie vrcholy (Aggarwal a kol., 1999).

Posledný prístup spomenutý v opisovanej práci je v súčasnosti najrozšírenejší. Jeho princíp spočíva v modelovaní používateľových preferencií ako množiny vektorov. Vektor je v tomto prípade tvorený vždy jedným typom metadát. V doméne multimediálneho obsahu sa môže jednať napríklad o kľúčové slová, hercov, názvy položiek, či roky výroby. Tento prístup sa často spája s využitím metód strojového učenia ako napríklad redukcia dimenzionality (angl. dimensionality reduction) alebo singularná dekompozícia hodnôt (angl. Singular value decomposition - SVD) (Goldberg a kol., 2000).

Existuje viacero systémov vychádzajúcich z opisovaného prístupu. Napríklad kolaboratívny odporúčací systém založený na vektorovom modeli používateľa zachytávajúcom žánre, hercov a kľúčové slová navrhli vo svojej práci Yu a kol. (Yu a kol., 2006). Model využívajúci totožný prístup, ktorý však uchovával preferencie len formou žánrov hodnotených položiek navrhli Ferman a kol. (Ferman a kol., 2003). Mukherjee a kol. použili rovnaký typ modelu používateľa uvažujúci žánre, kľúčové slová, herečky, hercov a režisérov (Mukherjee a kol., 2003). Wang a kol. rozšírili tento koncept dokonca do podoby množiny viacerých vektorových modelov používateľa, pričom každý z nich bol určený pre odlišnú náladu používateľa (Wang a kol., 2010). Vektorový model tvorený hercami, režisérmi, žánrami a kľúčovými slovami, kde má každý z vektorov svoju váhu využili Debnath a kol. na odporúčanie filmov založené na obsahu (Debnath a kol., 2011).

Množstvo existujúcich riešení založených na princípe vektorového modelovania preferencií používateľa dokazuje vhodnosť tohto konceptu jednak pre účely kolaboratívneho ale taktiež pre odporúčanie založené na obsahu.

Problém studeného štartu (angl. Cold Start Problem)

Preferencie nového používateľa však nepoznáme ihneď potom čo začne používať odporúčací systém. Pokiaľ nepoznáme dostatočné množstvo informácií o jeho aktivite, nevieme čo ho zaujíma, aký obsah chce alebo naopak nechce vidieť. Tento problém je známy ako problém studeného štartu. Existuje viacero jeho riešení. Príkladom môže byť situácia, kedy používateľ pred prvým odporúčaním explicitne ohodnotí určitý počet položiek, čím nepriamo vyplní svoj model

používateľa (van Meeteren, 2000). Inú možnosť predstavuje podobná situácia, kedy však používateľ pred prvým odporúčaním vyplní svoj model používateľa priamo, zadáním svojich preferencií.

Hodnotenie položiek prebieha tak, že používateľ z množiny všetkých položiek vyberá jemu známe položky a tieto potom hodnotí. Tým vyjadruje ako veľmi sa mu daný obsah páčil. Rovnako početnosť s akou vyberá zástupcov jednotlivých kategórií, napríklad žánrov, vypovedá o jeho záujmoch a preferenciách.

Informácie o používateľoch je možné získavať aj inak ako priamym hodnotením obsahu. Prvou možnosťou je nechať používateľa explicitne vyplniť svoj profil. Výhodou tohto prístupu je, že sa o ňom dozvieme informácie priamo a nemusíme ich odhadovať. Nevýhodou je, že používatelia neradi venujú svoj čas vyplňaniu informácií. Rovnako je otáznne ako objektívne dokážu v krátkom čase zhodnotiť svoje záujmy a preferencie (Wongchokprasitti a Brusilovsky, 2007).

Ďalšou možnosťou je preto informácie prebrať z existujúceho systému. Množstvo ľudí totiž informácie o sebe zverejňuje na rôznych portáloch či sociálnych sieťach. Tieto informácie poskytujú dlhodobu a je preto možné predpokladať, že do určitej miery vyjadrujú ich dlhodobé záujmy a sú teda relevantnejšie v porovnaní s informáciami zadanými do profilu pri jednom krátkom sedení, prípadne s narýchlo ohodnotenou množinou položiek. Tu je však nutný predpoklad, že používateľ poskytne k takémuto svojmu účtu prístup. Navyše aj po získaní informácií z účtu je nutné ich dokázať správne interpretovať a začleniť do vlastného modelu používateľa.

Spätná väzba

Okrem úvodného naplnenia používateľského profilu je nutné sledovať priebežnú aktivitu používateľa pomocou jeho explicitnej alebo implicitnej spätnej väzby prípadne pomocou sledovania zmien v jeho profile. Takto model používateľa postupne dopĺňame a skvalitňujeme.

Explicitná spätná väzba sa získava najmä formou hodnotenia položiek používateľom. Hodnotenie môže byť unárne (možnosť „Páči sa mi“ na sociálnej sieti Facebook²), binárne (palec hore / dole pri hodnotení na portáli Youtube³) alebo n-árne podľa použitej stupnice (6-bodová hodnotiacia stupnica filmov na portáli Csfid⁴ (0-5 hviezdíček, pričom 0 je zvýraznená slovom „odpad!“), 10-bodová hodnotiacia stupnica na portáli Imdb⁵ (10 hviezdíček s možnosťou zadať pol hviezdy)). Hodnotenie je zväčša reprezentované intuitívnymi grafickými prvkami (Obrázok 1).



Obrázok 1. Grafické prvky určené na zadávanie explicitnej spätnej väzby používané v reálnych systémoch: a) unárna b) n-árna, c) binárna, d) n-árna.

² <http://www.facebook.com/>

³ <http://www.youtube.com/>

⁴ <http://www.csfid.cz/>

⁵ <http://www.imdb.com/>

Implicitná spätná väzba je získavaná sledovaním ako často a ako dlho používateľ napr. pozerá videá s určitou tematikou. Prípadne v akom poradí číta novinky na spravodajskom portáli. Nevyžaduje žiadne dodatočné akcie od používateľa. Implicitná spätná väzba má menšiu presnosť ako explicitná, na jej získanie však nemusíme používateľa zaťažovať hodnotením položiek. Práve to je totiž najväčšou slabinou inak výhodnej explicitnej spätnej väzby.

2.2 Prístupy k personalizovanému odporúčaniu

Skupinové odporúčanie vychádza z personalizovaného odporúčania jednotlivcom. Až v momente, keď porozumieme možnostiam, ktoré máme pre používateľov jednotlivo, môžeme začať premýšľať o tom, ako pristupovať k celým skupinám naraz.

Poznáme dva základné prístupy k personalizovanému odporúčaniu: kolaboratívne odporúčanie (založené na sociálnych vzťahoch) a odporúčanie založené na obsahu. Oba tieto základné prístupy existujú v mnohých variáciách prispôbených konkrétnym doménam a situáciám, v ktorých sú používané. Okrem toho sa stretávame aj s hybridnými systémami kombinujúcimi oba základné prístupy, prípadne variácie jedného z nich s cieľom využiť silné stránky a zároveň sa vyhnúť slabším stránkam systémov.

2.2.1 Kolaboratívne odporúčanie

Tento spôsob personalizovaného odporúčania vychádza z predpokladu, že pokiaľ sa určitý obsah páčil používateľom s podobnými preferenciami ako má používateľ, ktorému aktuálne odporúčame, bude sa tento obsah páčiť aj jemu. Prístup sa prvýkrát objavil v roku 1992 v systéme Xerox PARC, ktorý vyhľadával dokumenty na základe podobných hodnotení rôznymi používateľmi (Goldberg a kol., 1992).

Atribútmi určujúcimi podobnosť používateľov môžu byť okrem príbuzných preferencií tiež podobné záujmy, aktivity či návyky. Rozhodujúce kritérium kolaboratívneho odporúčania predstavuje určitá zhoda v predchádzajúcom správaní používateľov, na základe ktorej sa odporúča vhodný obsah. Bežne sa odporúča obsah známy podobným používateľom a zároveň neznámy používateľovi, ktorému práve odporúčame. Tento prístup je široko využívaný, pretože pri odporúčaní dokáže dosahovať vysokú presnosť.

Pri využití tohto prístupu musíme pre odporúčanie poznať vybrané vlastnosti jednotlivých používateľov a tiež musíme vedieť určiť ich vzájomnú podobnosť. Z tohto dôvodu musíme pri nových používateľoch vyriešiť problémom studeného štartu, pretože inak by sme pre nich nevedeli odporučiť žiadny relevantný obsah.

Potrebné charakteristiky používateľov je možné odvodiť z informácií, ktoré o sebe v minulosti poskytli pomocou explicitnej alebo implicitnej spätnej väzby. V doméne inteligentnej televízie môže byť príkladom explicitnej spätnej väzby hodnotenie filmov po ich pozretí. Implicitnú spätnú väzbu reprezentuje napríklad čas, ktorý používateľ položku sledoval. V oboch prípadoch informácie transformujeme do podoby, ktorú dokážeme uchovávať v modeli používateľa. Tu postupne informácie zhromažďujeme, čím získavame čoraz presnejší model.

Vzájomnú podobnosť medzi používateľmi, na základe ktorej v kolaboratívnom odporúčaní vyberáme používateľov pre odporúčanie, vyhodnotíme zo sledovaných charakteristík modelu používateľa. Môže sa jednáť o históriu hodnotení, kde porovnávame nakoľko sa používatelia zhodujú v hodnotení filmov, prípadne či svoj čas trávia sledovaním podobných položiek. Pri tomto prístupe sa však stretávame s problémom, kedy používatelia hodnotili respektíve videli len malú časť z rozsiahlej množiny dostupných položiek (Tabuľka 1). Pokiaľ by sme model používateľa reprezentovali ako maticu s veľkosťou $M \times N$, kde M predstavuje počet používateľov systému, N počet všetkých existujúcich položiek systému a H_{ij} vyjadruje hodnotenie *Položky_j* od *Používateľa_i*,

výsledná matica by bola veľmi riedka a tým pádom nepoužiteľná pri určovaní podobnosti používateľov (Tabuľka 1). Primárny problém totiž spočíva v tom, že v prevažnej väčšine prípadov sa používateľove hodnotenia neprekrývajú s hodnoteniami iných používateľov na dostatočnej úrovni a preto mnoho používateľov s podobnými preferenciami týmto spôsobom odignorujeme len preto, že daní používatelia nehodnotili presne rovnaké položky. Druhý problém vyplývajúci z kolaboratívneho odporúčania spočíva v tom, že kým novému používateľovi dokážeme odporučiť, musíme najskôr poznať jeho preferencie. V prípade používateľa, ktorý hodnotil len málo položiek teda hrozí, že podobných používateľov identifikujeme nesprávne a odporučíme mu položky, ktoré sa mu nebudú páčiť.

Ďalšou nevýhodou prístupu je odporúčanie nových položiek. Položku je totiž možné odporučiť až potom, čo ju hodnotil podobný používateľ. Môže teda trvať pomerne dlhú dobu, kým položku komunita objaví. Následné šírenie v povedomí komunity už prebieha pomerne rýchlo.

Tabuľka 1. Matica hodnotení všetkých položiek používateľmi

Používatelia	Položky										
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_n
<i>Lolko</i>	5	-	-	-	-	-	-	4	-	4	...
<i>Alenka</i>	-	-	-	-	5	2	4	-	-	-	...
<i>Miško</i>	2	-	-	-	2	-	-	-	-	-	...
<i>Petka</i>	-	-	4	-	4	-	-	-	5	-	...
$Používateľ'_m$	...	...	...	...	...	...	...	...	...	...	H_{mn}

Riešením spomenutých problémov (riedkosť matice, studený štart) je napríklad v modeli používateľa ohodnotené (alebo videné) položky reprezentovať pomocou určitých všeobecnejších kategórií alebo pomocou metadát, ktoré ich opisujú. Príkladom v doméne inteligentnej televízie môžu byť žánre, do ktorých sú jednotlivé položky zaradené, prípadne kľúčové slová, ktoré položku charakterizujú. Tým totiž znížime pravdepodobnosť výskytu prázdnych miest v matici reprezentujúcej model používateľa, pretože každý prvok v matici bude napĺňaný viacerými položkami (na rozdiel od predchádzajúceho prípadu, kde každá položka predstavovala práve jeden prvok v matici).

Ďalším používaným riešením je dopočítavanie sledovaných akcií. Tento prístup je založený na predpoklade, že veľmi podobné položky by používateľ ohodnotil či preferoval podobne alebo by sa im venoval podobnú dobu. Na základe tohto predpokladu dokážeme určiť hodnoty sledovaných atribútov pre väčšiu množinu položiek než k akej používateľ fyzicky poskytol spätnú väzbu.

Kritériom podobnosti používateľov však nemusia byť len ich akcie, môžeme ho určovať aj na základe podobných záujmov, vzdelania, veku, či pohlavia používateľov. Pomocou vymenovaných a prípadne i ďalších atribútov následne dokážeme vyhodnotiť, podobnosť jednotlivých používateľov, vďaka čomu následne používateľovi dokážeme odporučiť obsah, ktorý on sám ešte nevidel, ale jemu najpodobnejším používateľom sa páčil.

Všeobecnou nevýhodou prístupu je, že nedokáže odporučiť obsah, ktorý ešte nehodnotili žiadni, prípadne len málo používateľov. Toto sa často týka novínok, ale rovnako sa môže jednať o obsah, ktorý používatelia jednoducho neodhalili. Aj v prípade, že obsah niekto ohodnotí môže trvať stále pomerne dlhú dobu kým sa rozšíri ku všetkým, ktorých môže zaujímať.

Výhody:

- Nie je potrebné analyzovať odporúčaný obsah (niekedy to dokonca nie je ani možné, pretože neexistujú metadáta a položky nie je možné priamo charakterizovať – napríklad videá alebo obrázky), rozhodujeme sa len podľa podobných znakov v konaní používateľov.
- Dokážeme odporučiť aj položky s inými vlastnosťami ako majú tie, ktoré používateľ obvykle pozerá a teda nemusí netušiť, že by sa mu mohli páčiť.

Nevýhody:

- Nevieme odporúčať novému používateľovi, pretože nevieme, ktorí používatelia sú mu podobní.
- Kým dokážeme nájsť dostatočné množstvo podobných používateľov, je potrebné aby používateľ ohodnotil pomerne veľa položiek.
- Môže trvať pomerne dlho kým sa nové položky rozšíria medzi používateľov. Pre odporúčanie položky je totiž nutné aby ju najprv objavili sami používatelia.

2.2.2 Odporúčanie založené na obsahu

Prístup je založený na predpoklade, že pokiaľ sa používateľovi páčil určitý obsah, budú sa mu pravdepodobne páčiť aj položky s podobnými vlastnosťami. Príkladom je situácia, kedy používateľ vysoko hodnotí alebo veľa pozerá komédie. Opisovaným prístupom mu odporučíme práve komédiu, ktorú ešte nevidel, pretože sa mu pravdepodobne bude páčiť viac ako napríklad horor. Ďalším príkladom môže byť odporúčanie podľa obľúbeného režiséra či herečky.

Preferencie používateľa je možné reprezentovať podobne ako pri kolaboratívnom prístupe. V doméne inteligentnej televízie to môže byť napríklad pomocou kategórií alebo kľúčových slov. V tomto prípade je spôsob, znázornený v Tabuľke 1 nepoužiteľný, keďže využíva model používateľa reprezentovaný samotnými položkami. Pri odporúčaní založenom na obsahu totiž potrebujeme vedieť rozhodnúť, ktoré položky spolu určitým spôsobom súvisia prípadne sú príbuzné. Používateľove preferencie pre jednotlivé položky nie sú preto postačujúce. Vhodnejšie je reprezentovať záujmy pomocou spoločných metadát, podľa ktorých vieme príbuznosť určiť (Weiß a kol., 2008). Výhodou multimediálneho obsahu, s ktorým pracujeme v doméne inteligentnej televízie, je jeho štruktúrovanosť. Jednotlivé položky totiž dokážeme zaradiť do určitých kategórií, podľa žánru, kľúčových slov a podobne. Tento prístup je preto v doméne multimediálneho obsahu podľa Melville dobre použiteľný (Melville a kol., 2002).

Ďalšou výhodou tohto prístupu, je nezávislosť od počtu podobných používateľov a tiež, že v porovnaní s kolaboratívnym odporúčaním ľahšie prekonáva úvodný problém studeného štartu, pretože mu na prvé odporúčanie stačí menšie množstvo informácií v modeli používateľa. Prístup rovnako odstraňuje problém nových položiek, ktoré ešte takmer nikto nevidel a tým pádom ani nehodnotil, ktorým trpí kolaboratívne odporúčanie. Z pohľadu tohto prístupu sa totiž jedná o plnohodnotnú položku, ktorú je možné ihneď po jej pridaní do množiny dostupných položiek odporučiť.

Naopak nevýhodou odporúčania založeného na obsahu je práve fakt, že odporúča jedine položky príbuzné tomu, čo už používateľ pozná. To znamená, že prístup používateľovi nikdy zásadne nerozšíri jemu známy obzor za hranicu domény, ktorú už pozná, čím ho v podstate uzatvára v úzko špecializovanej podmnožine.

Výhody:

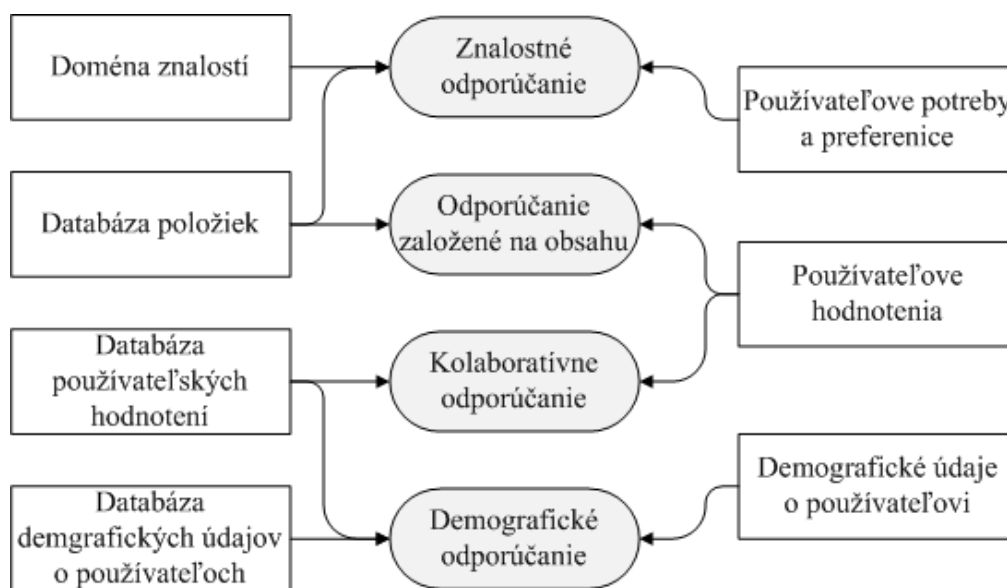
- Nie je potrebné poznať preferencie nikoho iného okrem používateľa, ktorému práve odporúčame.
- Novému používateľovi je možné odporúčať už po pár jeho hodnoteniach, pretože odporúčame obsah s podobnými parametrami.
- Používateľovi vieme opísať na základe čoho sme mu určitú položku odporučili (napr. hrá tam obľúbený herec, je to preferovaný žáner, a pod.).

Nevýhody:

- Je potrebné poznať a dokázať charakterizovať doménu, prípadne parametre obsahu, ktorý odporúčame.
- Analýza obsahu v dynamicky sa meniacej doméne vyžaduje značnú výpočtovú silu.
- Používateľa „uzatvárame v bubline“, pretože mu dokážeme odporúčať iba položky, ktoré sú podobné tomu, čo už pozná.

2.2.3 Ďalšie typy odporúčaní

Kolaboratívne odporúčanie a odporúčanie na základe obsahu sú síce najpoužívannejšie, no zďaleka nie jediné existujúce prístupy. V doméne odporúčania multimediálneho obsahu sa bežne využívajú tiež znalostné a demografické prístupy k odporúčaniam. Jednotlivé vymenované prístupy k odporúčaniam znázorňujeme na Obrázku 2.



Obrázok 2. Prístupy k personalizovanému odporúčaniam a ich zdroje dát (Burke, 2007): používané odporúčacie prístupy (v strede), dáta, ktoré prístupy pri odporúčaní využívajú (vľavo), vstupné dáta potrebné na inicializáciu odporúčania (vpravo)

Znalostné odporúčanie

V tomto type odporúčania sú položky používateľovi odporúčané na základe znalostí, postupne vyvodzovaných z jeho potrieb a preferencií. Tieto informácie sú štandardne získavané zbieraním explicitnej alebo implicitnej spätnej väzby z používateľovej aktivity. Druhou možnosťou však je, nadobúdanie znalostí ich explicitným definovaním doménovým expertom (Burke, 2000; Burke, 2007). Vďaka takto definovaným informáciám dokáže systém napríklad eliminovať problém studeného štartu.

Faktorom kvality tohto odporúčacieho prístupu je množina pravidiel, ktoré vytvára doménový expert. Vysoká náročnosť zostrojenia vhodných pravidiel prípadne mechanizmu, ktorý by ich automaticky vyvodzoval z aktivity používateľov je najväčšou nevýhodou brániacou masívnejšiemu rozšíreniu prístupu.

Demografické odporúčanie

Tento prístup používateľom odporúča na základe demografických údajov, medzi ktoré zaraďujeme napríklad lokalitu, vek pohlavie, vzdelanie či výšku zárobku (Burke, 2007; Pazzani 1999). Okrem týchto údajov, ktorých význam je pri odporúčaní pomerne zjavný, je možné pri demografickom odporúčaní sledovať tiež na prvý pohľad exotické parametre. Krulwich vo svojej práci sleduje napríklad či používateľ vlastní psa, aký druh whisky si kupuje, prípadne či hrá alebo sleduje golf (Krulwich 1997). Tieto údaje dokážeme využiť pokiaľ sa pri odporúčaní rozhodujeme, medzi viacerými položkami a chceme identifikovať, ktorá je pre používateľa vhodnejšia. Ak napríklad používateľ chová korytnačku, vieme mu odporučiť dokumentárny film o týchto zvieratách. Ak hrá hokej a v televízii súčasne vysielajú hokejový a futbalový zápas, odporučíme mu hokejový, pretože ho pravdepodobne bude preferovať viac.

Pri demografickom odporúčaní sledujeme aké sú spoločné črty používateľov, ktorí sa správajú podobne z hľadiska sledovanej charakteristiky. V doméne odporúčania multimediálneho obsahu si všímame, ktorí používatelia hodnotia alebo sledujú podobné položky. Pokiaľ následne do systému príde nový používateľ, nemusí u neho systém riešiť problém studeného štartu (potrebuje však o používateľovi poznať sledované charakteristiky) a môže mu ihneď odporučiť relevantný obsah. Výhodou v tomto prípade je, že pokiaľ sa používateľ do systému prihlási napríklad s účtom iného systému, kde už v minulosti svoje charakteristiky vyplnil a odporúčací systém dokáže tieto údaje využiť, dostáva ihneď odporúčania bez toho aby musel vyvinúť akúkoľvek námahu na to, aby ho systém spoznal.

2.2.4 Hybridné odporúčacie prístupy

Pojmom hybridný odporúčací systém označujeme ľubovoľný systém, ktorý kombinuje viacero odporúčacích prístupov. Je pritom jedno, či sa jedná o rovnaké alebo rozdielne prístupy. Dôvodom tohto kombinovania je snaha spojiť prípadne dokonca znásobiť silné vlastnosti jednotlivých nehybridných prístupov a súčasne eliminovať ich najväčšie slabiny. Spomenuté prístupy personalizovaného odporúčania totiž dokážu odporučiť výsledky, s ktorými sú používatelia často spokojní. Samostatne ale majú podstatné slabiny, vďaka ktorým je ich nevhodné alebo dokonca nemožné použiť v každej situácii.

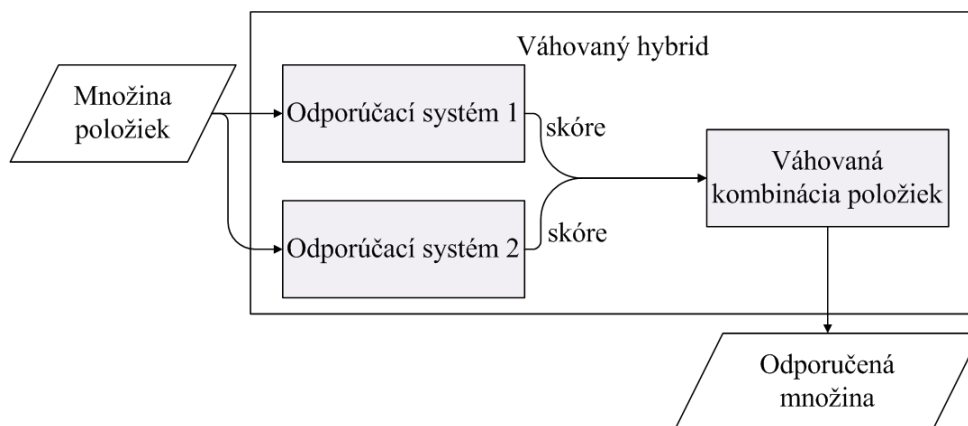
Z toho vyplýva, že pokiaľ vieme, že jeden prístup nedokáže odporučiť novú, ešte nehodnotenú položku a druhý pre zmenu neodporučí nič, čo prislúcha do žánrov, ktoré používateľ ešte nikdy nehodnotil, bolo by chybou napriek tomu odporučiť len pomocou metódy prislúchajúcej jedinému prístupu.

Logickým riešením je preto skombinovať viacero prístupov tak, aby sme dokázali využiť ich silné stránky a zároveň sme eliminovali tie slabé. Burke rozlišuje 7 základných typov hybridných systémov (Burke, 2002). Tie sa identifikujú na základe spôsobu akým sa výsledky jednotlivých prístupov podieľajú na výslednom odporúčaní systému, ktoré sa predloží používateľovi.

Váňovaný (angl. Weighted)

Jeden z najjednoduchších prípadov hybridných prístupov. Skladá sa z niekoľkých nezávislých odporúčacích metód, ktorých výsledky sa spájajú do jedného odporúčania pomocou pevne stanovenej lineárnej rovnice. Takto je čiastkovým výsledkom jednotlivých odporúčaní priradená váha, pomocou ktorej ich následne vieme usporiadať v spoločnom zozname.

Mobasher a kol. navrhujú určiť váhu, ktorú pridáme jednotlivým odporúčacím metódam, empiricky. Zistili totiž, že ich odporúčací systém kombinujúci znalostnú a kolaboratívnu metódu dosahuje najlepšie výsledky pri kombinácii v pomere 3:2 (Mobasher a kol., 2004). Príklad váhovaného hybridného prístupu je znázornený na Obrázku 3.

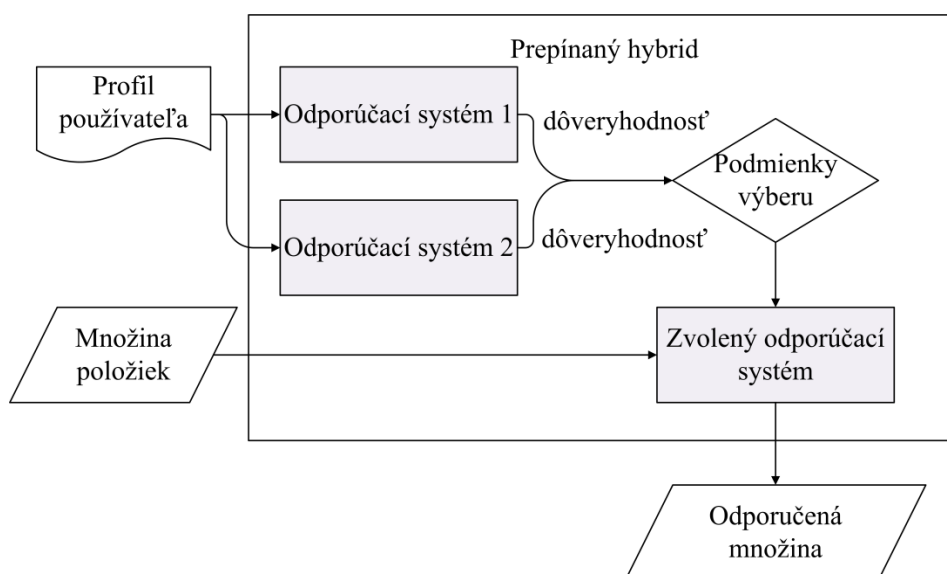


Obrázok 3. Váhovaný hybridný systém (Burke, 2007)

Prepínaný (angl. Switching)

Prepínaný hybridný prístup funguje na princípe jedinej aktívnej metódy. Samotný prístup síce obsahuje viacero metód, no výsledne zobrazuje vždy len výsledky jednej z nich. Jednotlivé metódy sú v systéme usporiadané v určitom poradí. Toto poradie je určené na základe vopred definovaných kritérií. Obslužný mechanizmus prístupu podľa tohto poradia zaktívuje prvú metódu, ktorá dovtedy nebola na rade a nechá ho vygenerovať odporúčanie. V prípade že dôveryhodnosť odporúčania prekoná vopred definovanú minimálnu hranicu, proces končí a používateľovi sú odporučené výsledky aktuálnej aktívnej metódy. Pokiaľ však výsledky tejto metódy potrebujú dôveryhodnosť nedosahujú, metóda je deaktivovaná a spustí sa proces pre nasledujúcu metódu v poradí.

Reálnym príkladom prepínaného systému je systém NewsDude určený na odporúčanie novinových článkov, ktorý prepína medzi tromi rôznymi prístupmi (Billsus a Pazzani, 2000; Pazzani, 1999). Princíp prepínaného hybridného prístupu je znázornený na Obrázku 4.

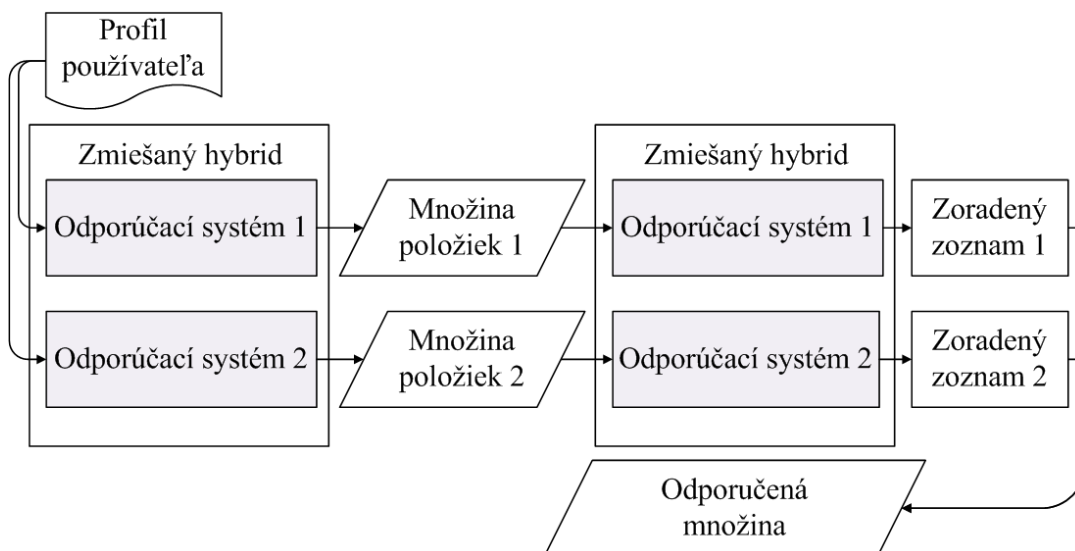


Obrázok 4. Prepínaný hybridný prístup (Burke, 2007)

Zmiešaný (angl. Mixed)

Tento hybridný prístup vychádza z predpokladu, že rozličné odporúčacie metódy dokážu na základe svojich silných stránok odporučiť množinu relevantných položiek, ktoré budú používateľa zaujímať. Každý prístup má však aj svoje slabiny, na základe ktorých časť vhodných výsledkov odhaliť nedokáže. Spomínali sme, že napríklad kolaboratívny prístup má slabinu v odporúčaní novínok, obsahovo orientovaný prístup pre zmenu odporúča len vo vymedzenej subdoméne, danej minulou aktivitou používateľa. Demografický prístup môže vynechať položky, ktoré nie sú prepojené s miestami, na ktorých sa používateľ bežne vyskytuje. Princípom zmiešaného hybridného prístupu je preto nechať odporučiť niekoľko rozličných metód a ich výsledky zobrazíť spoločne ako výsledky jedného odporúčania. Poradie jednotlivých položiek vo výslednom zozname je dané predpokladaným skóre, ktorým by mal dané položky používateľ ohodnotiť, prípadne mierou dôveryhodnosti jednotlivých odporúčacích metód. Tá môže byť určená empiricky.

Príkladom opisovaného prístupu je systém PTV, ktorý spája kolaboratívnu a obsahovo založenú metódu (Smyth a Cotter, 2000). Princíp zmiešaného hybridného prístupu zobrazuje Obrázok 5.

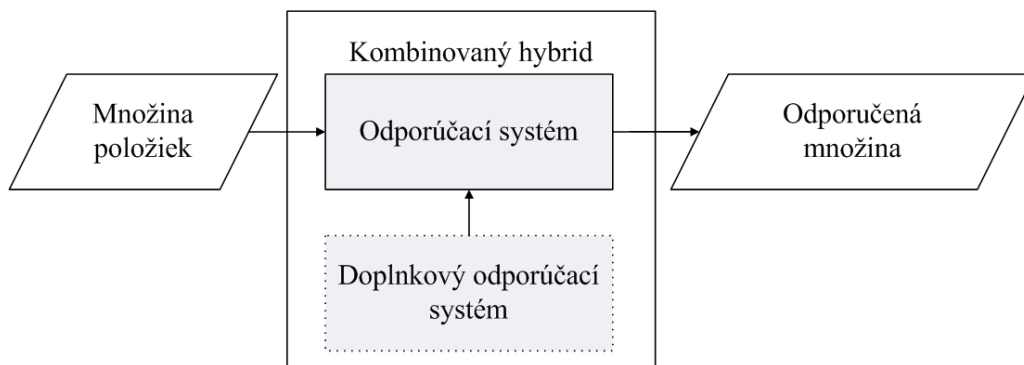


Obrázok 5. Zmiešaný hybridný prístup (Burke, 2007)

Kombinovaný (angl. Feature combination)

Takýto prístup pokladáme za hybridný z odlišných dôvodov ako predchádzajúce vymenované. Tu sa totiž nejdená o kombináciu výsledkov z viacerých metód využívaných v jednom systéme. Prístup obsahuje jedinou metódu, ktorej výsledky odporúčania zobrazí používateľovi bez zmeny. Hybridnosť tohto prístupu sa však prejavuje v obohatení odporúčacej metódy o informácie, ktoré z dostupných faktov dokázala vyvodit' doplnková metóda. Takto má odporúčacia metóda k dispozícii viacero informácií, na základe ktorých používateľovi dokáže odporučiť.

Príkladom môže byť práca Basu a kol. (Basu a kol., 1998). Tí vo svojej práci využili kombinovaný hybridný prístup na zlepšenie schopnosti učenia sa systému vytvárajúceho pravidlá určené na odporúčanie na základe obsahu. Svoj systém obohatili o poznatky získané pomocou kolaboratívneho odporúčania. Princíp kombinovaného hybridného prístupu je znázornený na Obrázku 6.

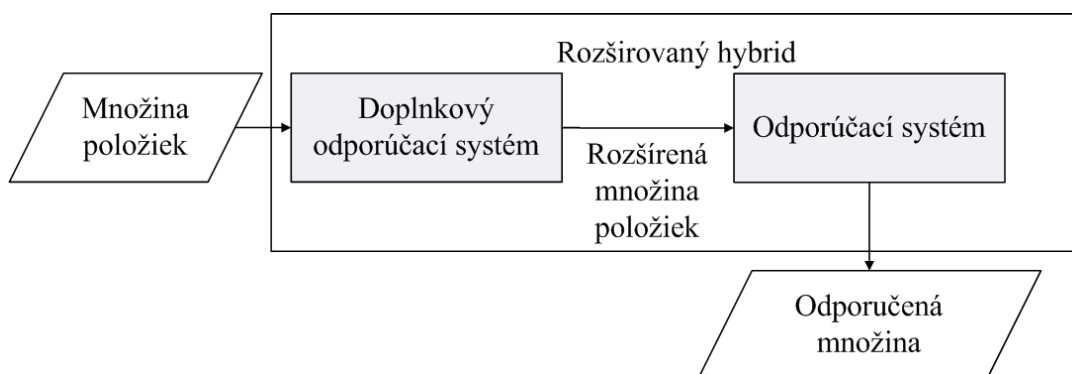


Obrázok 6. Kombinovaný hybridný prístup (Burke, 2007)

Rozširovaný (angl. Feature Augmentation)

Rozširovaný hybridný prístup sa podobá na kombinovaný. Je taktiež založený na jedinej odporúčacej metóde, ktorá je zodpovedaná za generovanie celkového výstupu, no môžeme nájsť aj jeden rozdiel. Ten spočíva v tom, že v prípade kombinovaného prístupu výsledky pomocnej metódy obohatili priamo hlavnú odporúčaciu metódu, ktorá tým pádom v podstate predstavovala kombináciu zúčastnených metód. V prípade rozširovaného systému však pracuje odporúčacia metóda úplne nezávisle a priamo nie je ničím ovplyvňovaná. Úlohou pomocných metód je totiž rozšíriť vstupnú množinu dát, z ktorých hlavná metóda potom vyberá.

Melville a kol. napríklad navrhli systém ktorý pomocou kolaboratívneho odporúčania predikuje hodnotenia pre položky, ktoré ešte používateľ nehodnotil a tým umožňujú hlavnej metóde, ktorá odporúča na základe obsahu odporúčať aj obsah mimo domén, ktoré používateľ skutočne hodnotil (Melville a kol., 2002). Princíp rozširovaného hybridného prístupu je znázornený na Obrázku 7.



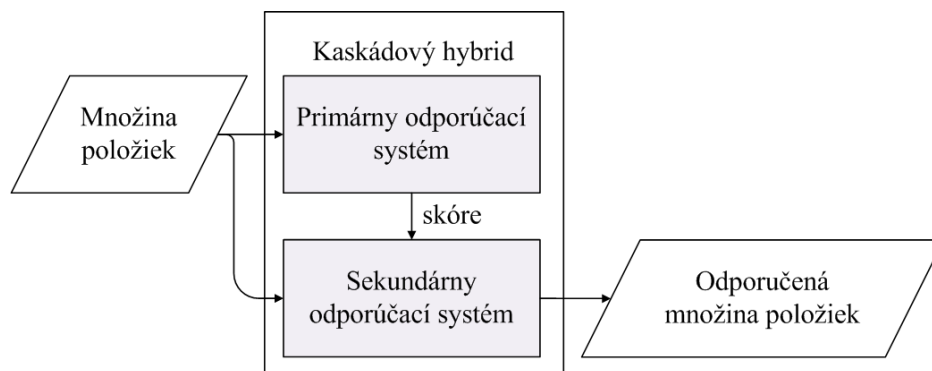
Obrázok 7. Rozširovaný hybridný prístup (Burke, 2007)

Kaskádový (angl. Cascade)

Kaskádový hybridný prístup je prípad prístupu, ktorý sa skladá z niekoľkých odporúčacích metód s presne stanoveným poradím. V tomto prípade môžeme hovoriť až o hierarchii. Systém vyžívajúci tento prístup v úvode vygeneruje odporúčanie pomocou metódy umiestnenej na vrchole hierarchie. Každému odporučenému má priradiť prioritu, určujúcu ako veľmi je tento prvok vhodný pre používateľa. Jednotlivé prvky následne zoradí podľa priority. V tomto poradí ich systém bude odporúčať používateľovi. Avšak v prípade, pokiaľ majú niektoré prvky priradenú totožnú prioritu, rozhodne o ich poradí nasledujúca odporúčacia metóda kaskádového odporúčacieho systému.

Pokiaľ sa vo výsledku opäť vyskytnú prvky s rovnakou prioritou, rozhodne o nich vždy nasledujúca metóda hierarchie s nižším postavením.

Príkladom môže byť Burkeho znalostný odporúčací systém reštaurácií, ktorý pomocou kaskádového prístupu vyriešil problém veľkého množstva rovnako hodnotených odporučených položiek (Burke, 2002). Princíp kaskádového hybridného prístupu je znázornený na Obrázku 8.

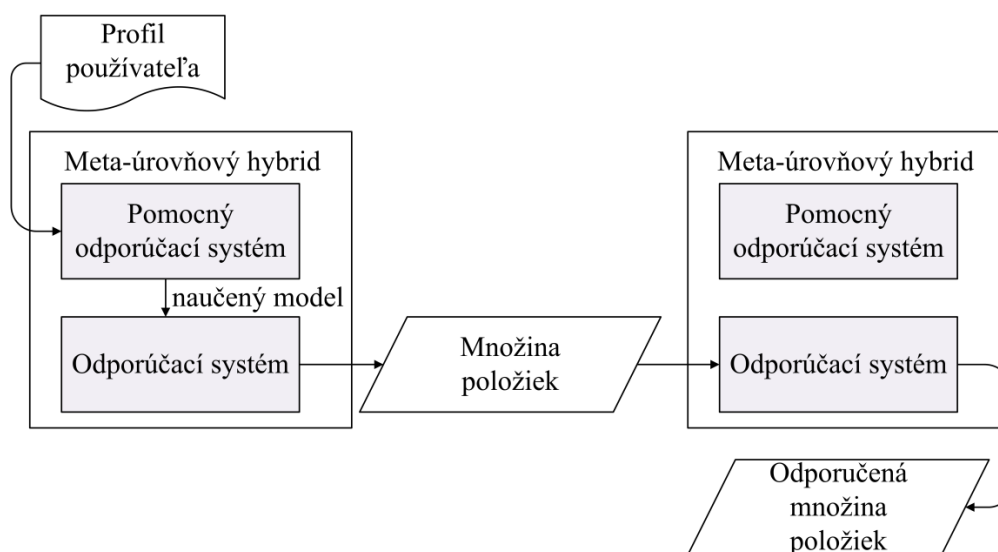


Obrázok 8. Kaskádový hybridný prístup (Burke, 2007)

Meta-úrovňový (angl. Meta-level)

Tento prístup k hybridným systémom je založený na spolupráci metód na úrovni metadát. V praxi to vyzerá tak, že pomocná metóda alebo metódy slúžia na naplnenie modelu používateľa. Na základe ich výsledkov sa napríklad vypočítajú predikované hodnotenia položiek, ktoré používateľ ešte nehodnotil a pod. Výsledky týchto metód môžu byť predpočítané mimo samotného odporúčania hlavnej metódy a to najmä z dôvodu aby nespomaľovali samotný výpočet hľadania odporúčaných položiek. Túto akciu zabezpečuje výhradne jediná, hlavná odporúčacia metóda meta-úrovňového hybridného prístupu.

Pazzani tento prístup využil v systéme na odporúčanie reštaurácií (Pazzani, 1999). Prístup nazval kolaboratívne odporúčanie pomocou odporúčania založeného na obsahu. Už samotný názov vypovedá o tom, že autor pred samotným kolaboratívnym odporúčaním vylepšil údaje v modeli používateľa pomocou techník odporúčania na základe obsahu. Princíp meta-úrovňového hybridného prístupu je znázornený na Obrázku 9.



Obrázok 9. Meta-úrovňový hybridný prístup (Burke, 2007)

V súčasnej dobe existuje viacero typov hybridných odporúčacích prístupov, v ktorých je navyše možné použiť rozličné základné odporúčacie prístupy, čo ešte rapídne zvyšuje počet známych hybridov. To na prvý pohľad vyvoláva pocit príliš rozsiahlej ponuky. Jednotlivé typy však majú často veľmi špecifické možnosti uplatnenia. Výberu hybridného odporúčacieho systému a jeho komponentov, by preto vždy mala predchádzať dôkladná analýza domény a úloh, na ktoré má byť systém použitý. Len tak môžeme naplno využiť výhody, ktoré nám jednotlivé prístupy poskytujú. Týmto spôsobom tiež rýchlo zistíme, že vymenované prístupy sú len špičkou ľadovca spomedzi známych a používaných prístupov.

3 Charakteristiky skupín

Potom čo dokážeme vhodne uchovávať preferencie používateľov pomocou modelov používateľa a poznáme jednotlivé prístupy k tvorbe personalizovaného odporúčania, je možné zamerať sa na samotné odporúčania skupinám.

Vieme, že televíziu sledujú ľudia často spoločne. V súčasnosti však neexistujú dostatočne rozsiahle štatistiky, ktoré by nám ešte pred samotným odporúčaním, napríklad na základe kontextu dokázali napovedať určité vlastnosti skupiny (Masthoff, 2004). Pre kvalitné skupinové odporúčanie je však kľúčové poznať veľkosť aktuálnej skupiny, jej presné zloženie, charakteristiku jednotlivých členov, ich vzájomné vzťahy či vplyv. Znalosť týchto informácií nám zásadným spôsobom pomáha pri výbere vhodného typu odporúčania pre aktuálnu situáciu. V nasledujúcej kapitole sa zameriavame na opis sledovaných charakteristík a spôsob akým ich je možné získať.

3.1 Typy skupín

Typ skupiny určujeme na základe spôsobu a účelu za akým vznikla. Pri skupinovom odporúčaní je tento aspekt pomerne podstatný, pretože na základe typu skupín, ktoré sa vyskytujú v doméne kde odporúčame, vyberáme rozličné spôsoby odporúčania. Príkladom môže byť situácia, kedy skupine kamarátov, ktorí sa dobre poznajú odporúčame inak ako skupine náhodných cestujúcich v autobuse. Amer rozdeľuje skupiny podľa typu nasledovne (Amer, 2009):

- **Stála skupina.** Skupina je tvorená členmi so stálymi spoločnými záujmami. Členovia sa rozhodli spojiť do skupiny. Napríklad skupina priateľov, ktorí pravidelne každý piatok sledujú filmy, ktoré im odporúčame.
- **Príležitostná skupina.** Skupina je tvorená ľuďmi, ktorí majú v danej chvíli spoločný cieľ. Všetci členovia skupiny sa vzájomne nemusia poznať. Napríklad skupina ľudí, ktorí idú naraz na poznávací zájazd a my im ideme odporučiť ciele výletov.
- **Náhodná skupina.** Skupina je tvorená ľuďmi, ktorí sa momentálne stretli na rovnakom mieste. Členovia skupiny sa vzájomne skôr nepoznajú a nemusia si ani uvedomovať, že sú členmi určitej skupiny. Napríklad sa jedná o návštevníkov rovnakej kaviarne, ktorým chceme pustiť čo najvhodnejšiu hudbu.
- **Automaticky vytvorená skupina.** Skupina je tvorená ľuďmi na základe ich podobnosti v určitom sledovanom aspekte. Skupina je vytváraná automaticky bez pričinenia jej členov. Napríklad sa jedná o študentov s podobným spôsobom učenia, ktorých adaptívny výučbový systém zaradí do rovnakej skupiny v rámci spoločnej prípravy študentov na test.

Okrem spôsobu vzniku je pre skupinové odporúčanie dôležité poznať či sa bude jednať o aktívnu alebo pasívnu skupinu. Túto vlastnosť skupín rozpoznávame podľa toho, či si skupina z odporúčeného obsahu aktívne vyberá (turisti vyberajúci si z odporučených výletov) alebo ho len pasívne prijíma (návštevníci kaviarne, v ktorej hrá odporúčená hudba) (Kompan a kol., 2014).

Ďalším dôležitým faktorom je homogenita skupiny. Tento atribút vyjadruje ako veľmi sú podobné preferencie jednotlivých členov v skupine. Príkladom homogénnej skupiny môže byť skupina fanúšikov fantastickej literatúry, príkladom heterogénnej zase skupina ľudí čakajúcich u zubára. Podľa Kompana platí, že čím podobnejšie sú preferencie členov skupiny, tým lepšie odporúčanie im dokážeme poskytnúť (Kompan a kol., 2014).

3.2 Predmet agregácie

V momente, kedy poznáme zloženie skupiny a máme k dispozícii modely používateľa jej členov, dokážeme začať tieto údaje agregovať za účelom vytvorenia jednotného odporúčania. Existujú dva základné prístupy určujúce predmet agregácie. Konkrétne sa jedná o agregáciu odporúčaní a agregáciu preferencií (Masthoff, 2011; Senot a kol, 2010).

Agregácia odporúčaní

Prístup je založený na princípe personalizovaného odporúčania jednotlivcom. Spočíva v tom, že každému členovi skupiny individuálne odporúčime obsah na základe jeho modelu používateľa. Predpokladá, že keďže boli personalizované metódy vyvinuté na odporúčanie jednotlivcom, poskytnú najlepšie výsledky práve pre nich.

Po vytvorení individuálnych odporúčaní opisovaný prístup agreguje takto vzniknutú množinu do jediného výsledku, ktorý predstavuje spoločný výsledok odporúčania pre danú skupinu. Ntoussi a kol. tvrdia, že tento prístup je flexibilnejší ako agregácia preferencií a skôr umožňuje optimalizovať proces tvorby skupinového odporúčania (Ntoussi a kol., 2012).

Agregácia preferencií

Prístup je založený na princípe vytvorenia agregovaného modelu „používateľa“ predstavujúceho preferencie celej skupiny. Takto vytvoríme akéhosi „typického zástupcu skupiny“, ktorému následne odporúčime obsah ako jednotlivcovi. Výsledok tohto odporúčania však predstavuje požadované skupinové odporúčanie (De Pessemier a kol., 2012; Quijano-Sanchez a kol., 2013).

Takto vytvorený spoločný model môže byť dokonca v prípade pravidelne sa stretávajúcich stálych skupín uchovaný a môže sa stať základom stáleho spoločného profilu skupiny.

Prístup agregácie preferencií je používannejší ako prístup agregácie odporúčaní, pretože odstraňuje jeho hlavný nedostatok, ktorý sa prejavuje pri výrazne nehomogénnych skupinách. Tento nedostatok spočíva v tom, že pri agregácii odporúčaní môže nastať prípad, kedy sú preferencie členov natoľko rozdielne, že položky, ktoré by boli akceptovateľné pre väčšinu, skončia pri individuálnom odporúčaní u jednotlivých používateľov pod hranicou vybraných prvkov. Pri výslednom spájaní odporúčaní tieto položky teda vôbec nefigurujú v množine položiek, z ktorých sa vyberá, napriek tomu, že by ako jediné mali podporu viacerých členov. Z tohto dôvodu výsledne odporúčime obsah, ktorý preferuje len malá časť skupiny, možno len jeden člen.

Pri použití agregácie preferencií, sa však spoločné preferencie agregujú ešte pred fázou odporúčania, čím zabráni vzniku vyššie opísanej situácie s výrazne nehomogénnou skupinou. V prípade, že sú spoločne dostatočne preferované, budú skupine odporúčené práve ony, čím zabezpečíme rovnomernejšie rozdelenie spokojnosti medzi jednotlivcami.

Podľa Guzziho sa však nedá povedať, že by bol niektorý z prístupov ideálny a vždy použiteľný (Guzzi, 2011). Výber predmetu agregácie je závislý od zvolenej domény a použitia a preto je nutné vždy individuálne zvážiť, ktorý sa v danom prípade hodí viac.

Jednotlivé spôsoby agregácie však nemusíme využívať len za účelom skupinového odporúčania. Je možné ich využiť tiež pri odporúčaní jednotlivcom, napríklad na zmiernenie problému studeného štartu (Masthoff, 2011). Pri využití viacerých charakteristík model používateľa je možné skvalitniť individuálne odporúčanie aj podľa Baltrunas-a. Ten vychádza z predpokladu, že skupinové odporúčanie je tým efektívnejšie, čím sú si členovia skupiny viac podobní. Model jedného používateľa zobrahaný z viacerých hľadísk je preto ukážkou vhodného adepta na skupinu (Baltrunas a kol., 2010).

3.3 Agregáčn  strateg ie

V predchádzajúcej  asti sme opisovali sp osoby agreg acie vyu ivan  v skupinovom odpor u an . Nevysvetlili sme v ak ako presne samotn  agregovanie prebieha. Senot a kol. rozpozn vajú tri z kladn  pr stupy (Senot a kol., 2010). Jedn  sa o strateg ie zalo en :

- na v    ine
- na konsenze
- hrani n  strateg ie

Masthoff vo svojej pr ci pribli uje 10 konkr tnych strateg  , ktor  v na ej pr ci priamo zara ujeme do troch spomenut ch z kladn ch pr stupov. Pri zara en  a opise strateg   vych dzame z autorkinej pr ce (Masthoff, 2004).

Strateg ie zalo en  na v    ine

Strateg ie patriace do tohto pr stupu, vyberaj  z mno iny dostupn ch polo iek tie, ktor  preferuje skupina ako celok. Jednotliv  strateg ie vyberaj  pod a r znich krit ri  celkovo najpreferovane  ie polo ky, pri om sa nezau imaj  o pr padn  negat vny postoj k nim zo strany  lenov skupiny, ktor  s  v men ine. Medzi z stupcov tohto pr stupu patria *Pluralitn  hlasovanie*, *Copelandove pravidlo* a *Schv len  hlasovanie*.

Pluralitn  hlasovanie (angl. Plurality Voting) umo ňuje ka  demu  lenovi hlasova  za v etky jeho najviac preferovan  alternat vy. Jednotliv  hlasy sa s  tavaj , polo ka, ktor  m  najviac hlasov vyhr va. V momente ke  s  vybran  v etky prvky ktor  dan   len preferoval najviac, hlasuje za v etky e te nezvolen  prvky s jeho najvy   ou preferenciou. T to strateg ia sa vyu  va napr klad pri vo b ach vo Ve kej Brit nii.

Copelandove pravidlo (angl. Copeland Rule) spo  va v tom,  e sa pre ka  deho hlasuj ceho  lena prid  k jednotliv m polo k m bod za ka  d  in  polo ku oproti ktorej aktu lnu preferuje  len viac a naopak sa bod odoberie za ka  d  polo ku, ktor  tento  len skupiny preferuje viac ako aktu lne hodnoten . Dosiahnut  body v etk ch  lenov skupiny sa v sledne spo  t j  a ur   sa poradie.

Schv len  hlasovanie (angl. Approval Voting) predstavuje strateg iu, v ktorej m  e ka  d   len skupiny hlasova  za ľubovoľn  po et polo iek. T to strateg ia je ur en  na podporu v etk ch alternat v, vo i ktor m nem   len skupiny negat vny postoj. T to strateg iu vyu  v j  pri hlasovan  odborn  spoločnosti ako napr klad IEEE.

Strateg ie zalo en  na konsenze

Cieľom strateg   zalo en ch na konsenze je dosiahnutie  o najvy   ej miery spokojnosti jednotliv ch  lenov skupiny. Tento cieľ m  vy   iu prioritu ako dosiahnutie  o najvy   ej celkovej spokojnosti skupiny. Strateg ie preto v  dy vezm  preferencie v etk ch  lenov, tie ur it m sp osobom agreguj  a polo ky vyber  na z klade tohto spoločn ho n zoru v etk ch  lenov.

Masthoff vo svojom v yskume zistila,  e pokiaľ je jednotlivec postaven  pred úlohu, aby s m vybral obsah pre skupinu, spr va sa ako *Priemern  strateg ia* pr padne ako *Strateg ia najmen ej nespokojnosti* (Masthoff, 2004). Obe tieto strateg ie zara ujeme pr ve do pr stupu zalo en ho na konsenze, z  oho vypl va,  e tento pr stup je najbli   i ľudsk mu sp osobu myslenia.

Okrem vymenovan ch strateg  , do tohto pr stupu zara ujeme tie  *Strateg iu u  ito nosti*, *Borda s   et*, *Strateg iu najvy   ej spokojnosti*, *Priemern  strateg iu bez nespokojnosti* a *Spravodliv  strateg iu*.

Priemern  strateg ia (angl. Average Strategy) ur   hodnotenie ka  dej polo ky ako priemer preferenci  v etk ch  lenov skupiny. N sledne vyberie polo ky s najvy    im hodnoten m.

Stratégia najmenšej nespokojnosti (angl. Least Misery Strategy) ohodnotí položky podľa najhorších preferencií, pre tieto položky v danej skupine. Následne vyberie položky s najvyšším hodnotením. Myšlienkou je, že skupina je len tak spokojná ako jej najnespokojnejší člen.

Stratégia užitočnosti (angl. Utilitarian Strategy) určuje užitočnosť každej položky na základe preferencií všetkých členov skupiny. Výsledné hodnotenie určí *aditívne* alebo *multiplikatívne*. Nevýhodou, je že pri väčších skupinách môže byť názor jednotlivca vždy prehlasovaný zvyšnými členmi.

Borda súčet (angl. Borda Count) prideluje položkám body podľa ich pozície v zozname preferencií jednotlivcov. Následne vyberie položky, ktoré získali v súčte najviac bodov.

Stratégia najvyššej spokojnosti (angl. Most Pleasure Strategy) ohodnotí položky podľa najlepších preferencií, pre tieto položky v danej skupine. Následne vyberie položky s najvyšším hodnotením.

Priemerná stratégia bez nespokojnosti (angl. Average Without Misery Strategy) určí hodnotenie položiek ako priemer preferencií všetkých členov skupiny. Neurčuje však hodnotenie položiek, ktoré aspoň jeden člen preferuje menej ako je určitá stanovená hranica (napríklad 4 na desaťbodovej hodnotiacej škále). Následne z položiek, ktoré boli hodnotené vyberie položky s najvyšším hodnotením.

Spravodlivá stratégia (angl. Fairness Strategy) vyberie najvyššie hodnotené položky od všetkých členov skupiny.

Hraničné stratégie

Stratégie tohto prístupu síce slúžia na agregáciu preferencií členov skupiny, no spôsob akým pracujú je na hranici zaradenia medzi skupinovú prístup. Typickým príkladom je *Diktátorská stratégia* nazývaná tiež *Stratégia najrešpektovanejšej osoby* (angl. Dictatorship Strategy, Most Respected Person Strategy). Táto stratégia vyberá položky na základe preferencií vybraného člena skupiny. Používa sa napríklad v situáciách kedy je niektorý člen skupiny aktuálne dôležitejší ako zvyšok skupiny (host, nadriadený, oslávenec).

Masthoff vo svojej práci vyhodnocuje experiment, v ktorom dospela k záveru, že skupiny sú v priemere najspokojnejšie, pokiaľ je im obsah vybraný pomocou *Bodra súčtu*, *Priemernej stratégie*, *Priemernej stratégie bez nespokojnosti* prípadne pomocou *Stratégie najvyššej spokojnosti*. Naopak v týchto experimentoch zlyhalo najmä *Copelandove pravidlo*, *Pluralitné hlasovanie* a *Stratégia najmenšej nespokojnosti* (Masthoff, 2011). Podľa nej sa pri vyhodnocovaní výsledkov agregácie používateľa starajú najmä o to, aby nebol niektorý člen skupiny príliš nespokojný a aby bola v skupine zachovaná spravodlivosť.

3.4 Sociálne vzťahy v skupine

Skupina je v prvom rade tvorená jej členmi. Tam však jej pôsobenie nekončí. Členovia skupiny totiž taktiež vytvárajú spleť vzájomných vzťahov a emócií. Liu a kol. vo svojej práci vyjadrujú presvedčenie, že pokiaľ chceme skupine odporúčať čo najrelevantnejší obsah musíme tieto prepojenia čo najlepšie pochopiť a dokázať ich zakomponovať do samotného procesu odporúčania (Liu a kol., 2012).

Prvým aspektom, na ktorý je vhodné pri skupinovom odporúčaní prihliadať je *emocionálne šírenie*. Tento jav spočíva v tom, že spokojnosť alebo nespokojnosť jedného člena skupiny pôsobí aj na ostatných. Sila tohto vplyvu závisí od sily osobnosti daného člena ale tiež od vzťahu medzi ním a osobou, ktoré ovplyvňuje. Quijano-Sanchez a kol. tvrdia, že ľudia sú najviac ovplyvnení tými, ktorých majú radi – partnermi, najlepšími priateľmi. Za nimi nasledujú rešpektované osoby –

nadriadení alebo rodičia, na tretej pozícii sa nachádzajú rovnocenné osoby – spolužiaci či kolegovia. Najmenej sa necháme ovplyvňovať osobami, ktoré nemáme radi alebo s nimi súperime (Quijano-Sanchez a kol., 2013).

Masthoff hovorí aj o prípade, kedy môžu členovia skupiny zmeniť názor alebo spokojnosť na základe presvedčenia zvyšku skupiny (Masthoff, 2011). Rozlišuje normatívny a informatívny vplyv. Normatívny vplyv predstavuje prípad, kedy jednotlivec upraví svoje stanovisko aby zapadol do skupiny, no naďalej si necháva svoj názor. Informatívny, naopak, predstavuje situáciu kedy jednotlivec upraví svoj názor pretože má pocit, že celá skupina sa nemôže myliť a teda, že sa musí myliť on.

Druhý dôležitý aspekt medzi sociálnymi vzťahmi v skupine je *dôležitosť jednotlivých členov*. V reálnom svete sa totiž nestretávame s prípadom kedy by boli názory všetkých členom skupiny brané s rovnakou váhou. Radi by sme však zdôraznili, že nemáme na mysli prípad formálneho hlasovania na porade manažmentu určitej korporácie, ale bežnú situáciu kedy sa napríklad skupina kamarátov neformálne rozhoduje ako strávi večer. Vplyvní používatelia totiž majú vyššiu pravdepodobnosť, že dokážu ovplyvniť rozhodovanie skupiny pri výbere z odporúčaných položiek.

De Pessemier a kol. klasifikujú dôležitosť člena v skupine na stupnici vysoká, neutrálna a nízka. Tento ukazovateľ dokáže napríklad vyjadriť fakt, že starší ľudia (rodičia) majú na skupinu väčší vplyv ako mladší členovia (deti) (De Pessemier a kol., 2012).

Liu a kol. vo svojej práci zaviedli pojem osobného vplyvu člena skupiny. Na základe histórie každého člena vyhodnotia ako veľký je jeho vplyv na výsledný výber skupiny. Autori v záznamoch z minulých výsledkov porovnávajú nakoľko sa individuálne preferencie daného člena zhodovali s výsledným výberom skupiny. Rovnako navrhujú zlepšenie kvality tohto parametra pomocou analyzovania aktivity používateľa na sociálnych sieťach (Liu a kol., 2012).

3.5 Metadáta opisujúce multimediálny obsah

Pri odporúčaní určitého obsahu je vhodné dokázať klasifikovať jednotlivé položky patriace do domény. Pod pojmom klasifikácia v tomto prípade rozumieme proces zatriedenia položiek do určitých kategórií podľa ich vlastností a atribútov. Vďaka tomu dokážeme napríklad nepriamo vyhodnotiť vzájomnú podobnosť položiek.

V prípade multimediálneho obsahu na jednej strane nedokážeme priamo zistiť o položke žiadne bližšie informácie okrem technických detailov ako sú veľkosť súboru, kvalita záznamu a podobne. V tomto smere je práca s multimediálnym obsahom ťažšia napríklad v porovnaní s textami, pre ktoré vieme priamo analýzou textu vyvodit' kľúčové slová, frekvenciu výskytu slov či pomenované entity a na základe týchto údajov zistiť bližšie informácie o danom texte. Na druhej strane je však výhodou multimediálneho obsahu dostatok metadát zameraných na samotný obsah. Údaje ako žánre, rok vydania, režisér, zoznam hercov či kľúčové slová vytvárajú dostatočnú informáciu o položke, na základe ktorej ju vieme zaradiť do klasifikačnej štruktúry v rámci domény. Položky vďaka tomu tiež dokážeme posudzovať podľa viacerých atribútov. Tento spôsob reprezentovania položiek bol pôvodne zavedený priamo pre ľudí, pretože podobne ako odporúčacie metódy potrebovali aj oni aspoň približne spoznať obsah bez toho, aby museli vidieť celú položku. Pre účely personalizovaného odporúčania sa však výborne hodí, vďaka čomu môžeme o multimediálnom obsahu prehlásiť, že patrí medzi dobre štruktúrované domény.

Nutnosť znalosti metadát o položkách sa prejavuje takmer v každom prístupe k personalizovanému odporúčaní. Na znalosti vlastností jednotlivých položiek a tiež ich vzájomnej podobnosti je postavený princíp personalizovaného odporúčania založeného na obsahu. Bez určenia podobnosti položiek, by sme totiž nedokázali vyhodnotiť, ktoré položky máme

používateľovi odporučiť ani v prípade, že by sme mali informáciu ktoré položky videl a ako sa mu páčili.

Rovnaký problém by nastal aj pri demografickom odporúčaní, pri ktorom potrebujeme poznať lokalitu ktorej sa položka týka. V prípade multimediálneho obsahu sú preto pre tento prístup k odporúčaniam opäť jedinou možnosťou informácie získané nepriamo z metadát.

Tretím prístupom, ktorý je do veľkej miery závislý na dodatočných informáciách je znalostné odporúčanie. Tu je totiž nutné jednotlivé znalosti vytvoriť a reprezentovať na základe záujmov a vlastností používateľa. Tie však dokážeme vyhodnotiť len ak poznáme charakteristické črty položiek, ktoré používateľ videl.

Jediný prístup, ktorý je teoreticky nezávislý od akýchkoľvek informácií o položkách, je kolaboratívne odporúčanie. V tomto prípade postačuje poznať používateľov, ktorí podobne hodnotili totožné položky a následne používateľovi odporučiť položky, ktoré nevidel ale podobní používatelia ich hodnotili pozitívne. V praxi sa však stretávame so situáciou kedy je doména natoľko rozsiahla, že ju hodnotenia používateľov nepokrývajú dostatočne. Na určenie podobnosti používateľov sa preto musíme namiesto porovnávania hodnotení konkrétnych položiek uchýliť k porovnávaniu hodnotení množín podobných položiek. Príkladom je vyhodnocovanie podobnosti používateľov na základe podobného hodnotenia žánrov alebo kľúčových slov. Z tohto dôvodu je preto znalosť dodatočných informácií opisujúcich položky potrebná aj pri tomto prístupe.

4 Existujúce systémy

V tejto kapitole opisujeme existujúce systémy určené na personalizované odporúčanie skupinám. Vychádzame z analýzy systémov určených na skupinové odporúčanie, ktorú vo svojej práci uvádza Kompan (Kompan a kol., 2014). V našej práci sa však špecializujeme na systémy primárne zamerané na skupinové odporúčanie multimediálneho obsahu (ostatné zo spomenutej analýzy preto vynechávame), pričom sa o týchto systémoch snažíme poskytovať podrobnejšie informácie (pridali sme informáciu o využívaných agregáčnych stratégiách a použitom odporúčačom prístupe) (Tabuľka 2).

Tabuľka 2. Zoznam najrelevantnejších skupinových odporúčačích systémov zameraných na odporúčanie multimediálneho obsahu (Kompan a kol., 2014).

Názov	Rok	Typ skupiny	Stálosť skupiny	Predmet agregácie	Spôsob agregácie	Odpor. prístup
PolyLens	2011	skutočná	stála	profil	s najmen. nespok.	kolab.
Yu's TV	2006	skutočná	dočasná	odporúčania	priemerná stratég.	kolab.
Avatar	2006	skutočná	stála	profil	nezistené	obsah. a kolab.
FIT-Family	2004	virtuálna	dočasná	profil	aditívna stratégia	obsah.
Happy Movie	2011	skutočná	stála	odporúčania	nezistené	obsah., hlasov.
gRecs	2012	virtuálna	dočasná	profil	s najmen. nespok., spravodlivá strat., s. najväčšou spok.	kolab.
AC -based RS	2011	skutočná	dočasná	profil	nezistené	kolab.

Rok predstavuje rok vzniku systému, typ skupiny hovorí o tom či systém pracuje so skupinami, pre ktoré je overené, že sú prítomní všetci členovia (*skutočná skupina*) alebo s odhadnutými skupinami (*virtuálna skupina*). Stálosť skupiny hovorí o tom, či sú skupiny uchovávané dlhodobo (*stála skupina*), alebo sa vždy vytvárajú nanovo (*dočasná skupina*). Predmet agregácie môže nadobúdať hodnotu *profil* alebo *odporúčania*. Spôsob agregácie uvádzame v prípade, že je známy (inak *nezistené*). Medzi používané odporúčačie prístupy patria kolaboratívny (*kolab.*) a založený na obsahu (*obsah.*). V prípade systému Happy Movie majú používatelia navyše možnosť hlasovania o odporúčanom obsahu (*hlasov.*).

4.1 PolyLens

PolyLens je systém slúžiaci na personalizované odporúčanie multimediálneho obsahu. Bol vyvinutý ako rozšírenie existujúceho voľne dostupného systému MovieLens slúžiaceho na odporúčanie jednotlivcom.

Používatelia majú v systéme možnosť hodnotiť položky na 5 hviezdíčkovej škále. Z týchto hodnotení sa im vypočítavajú individuálne modely používateľov, reprezentujúce ich záujmy.

Systém je určený pre stále skupiny, z čoho vyplýva, že používatelia si explicitne volia s kým vytvoria skupinu, ktorej systém následne odporúča. Nie je určený na odporúčanie náhodným alebo automaticky vytvoreným skupinám. Autori systému vo výsledkoch 9 mesačného experimentu z roku 2001 hodnotia (O'Connor a kol., 2002), že systém využívali prevažne menšie skupiny (Tabuľka 3).

Tabuľka 3. Veľkosti skupín v systéme PolyLens (O'Connor a kol., 2002)

Veľkosť skupiny	Počet skupín v systéme
2	257
3	53
4	16
viac	12
spolu	338

Systém využíva metódu kolaboratívneho odporúčania obsahu, ktorú autori prevzali zo systému MovieLens, kde sa táto metóda osvedčila a používatelia boli spokojní s jej odporúčaniami (O'Connor a kol., 2002). Preferencie členov skupiny sú pred samotným odporúčaním agregované pomocou stratégie najmenej nespokojnosti. Tá je však pozmenená o princíp, že položky, ktoré už sú ohodnotené niektorým používateľom priamo (čiže ich už videl), sú z výslednej množiny odstránené. Medzi ciele systému patrí snaha o dosiahnutie spokojnosti všetkých členov skupiny.

a)	TITLE	GENRE	GROUP	YOUR	
	King of Masks, The (Bian Lian) (1996)	Drama	★★★★★	★★★★★	
b)	TITLE	GENRE	REVIEWS	GROUP	YOUR
	Frequency (2000)	Drama, Thriller		★★★★★	★★★★★
	TITLE	GENRE	REVIEWS		oconnor@cs.umn.edu
	B. Monkey (1998)	Romance, Thriller		★★★★★	
c)	TITLE	GENRE	REVIEWS		lam@cs.umn.edu
	Last Night (1998)	Children's, Comedy		★★★★★	
	Get Bruce (1999)	Documentary		★★★★★	
	Last Night (1998)	Children's, Comedy		★★★★★	

Obrázok 10. a) Rozhranie systému PolyLens zobrazujúce iba preferencie skupiny ako celku, b) Kompozitné rozhranie dvojčlennej skupiny v systéme PolyLens, c) Rozhranie manuálneho odporúčania skupine v systéme PolyLens

Autori vo svojom experimente opisujú tri rozhrania, ktorými systém používateľovi prezentuje výsledky odporúčania. Na Obrázku 10 a) je zobrazená prvá možnosť kedy používateľ vidí svoju preferenciu a preferenciu skupiny. Takto je zachované súkromie jednotlivých členov skupiny a používateľ má prehľad ako veľmi sa stotožňuje so zvyškom skupiny. Obrázok 10 b) ukazuje prípad okrem vlastných preferencií a preferencií skupiny vidí tiež jednotlivo všetkých ostatných členov skupiny (na obrázku je príklad dvojčlennej skupiny). Obrázok 10 c) znázorňuje tretiu alternatívu, ktorou je individuálne zobrazenie odporúčených zoznamov pre členov skupiny jednotlivo. V tomto prípade je na skupine, ktorú položku si sama vyberie (O'Connor a kol., 2002).

Nevýhodou systému je neuvažovanie sociálnych vzťahov v skupine. Všetci používatelia majú v systéme rovnakú váhu, ktorou ovplyvňujú výber. Neexistuje tu vplyv používateľov, šírenie emócií či vzťahy medzi používateľmi v skupine.

4.2 Yu's TV

Yu's TV (Yu a kol., 2006) je systém určený na odporúčanie položiek z televízneho programu. Používatelia sú v systéme reprezentovaní na základe položiek hodnotených v minulosti. Jednotlivé atribúty ich modelov používateľa sú vytvárané a modifikované na základe metadát prislúchajúcim

k hodnoteným položkám. Príkladom metadát využívaných v systéme sú žánre, kľúčové slová či herci. Samotné modely používateľov sú tvorené vektormi, ktoré môžeme porovnávať.

Systém je založený na princípe kolaboratívneho odporúčania. Jeho stratégiou je nájsť používateľov, ktorých vektory sú si najpodobnejšie. Na to systém využíva metódu Pearsonovej korelácie. Tento koeficient slúži na vyhodnotenie podobnosti akcií medzi presne dvoma používateľmi. Jeho výhodou, je že určuje koreláciu v kladnom aj zápornom smere (v rozsahu $\langle -1; 1 \rangle$), čím dokáže okrem najpodobnejších používateľov (korelačný koeficient sa blíži k 1) určiť tiež tých, ktorí hodnotia opačne (korelačný koeficient sa blíži k -1). Pearsonov korelačný koeficient - PCC vypočítame pomocou Vzorca 1, kde x a y predstavujú množiny dát reprezentujúcich jednotlivých používateľov a N je celkový počet dát v jednej množine.

$$PCC(x, y) = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{N}\right) \left(\sum y^2 - \frac{(\sum y)^2}{N}\right)}} \quad (1)$$

Pri odporúčaní jednotlivcom systém pre každého z členov skupiny preto vyberie najpodobnejších používateľov a z nich vyberie najlepšie hodnotené položky. Vyberá pritom len také, ktoré členovia skupiny nevideli. Jednotlivé položky sú v systéme agregované do jednotného zoznamu odporúčaní. Pri agregácii položiek je využitá priemerná stratégia.

Nevýhodou systému, podobne ako pre systém PolyLens, je neuvažovanie sociálnych vzťahov v skupine. Všetci používatelia majú v systéme rovnakú váhu, ktorou ovplyvňujú výber. Neexistuje tu vzájomný vplyv používateľov, šírenie emócií či vzťahy medzi používateľmi v skupine.

4.3 Avatar

Systém je určený na odporúčanie multimediálneho obsahu. Informácie o obsahu v doméne sú reprezentované pomocou ontológií. Systém predstavuje hybridný koncept, v ktorom je využité obsahové a kolaboratívne odporúčanie. V prvom rade agreguje preferencie skupiny do spoločného profilu. Následne pomocou odporúčania na základe obsahu vyhodnotí položky, ktoré by pre skupiny mali byť zaujímavé. Na to využíva hierarchické a interferenčné vyhľadávanie. Tieto položky následne doplní o výsledky kolaboratívneho odporúčania. V rámci neho sa určuje podobnosť medzi používateľmi na základe Pearsonovej korelácie.

Využitie hybridnej metódy slúži tiež na odstránenie riedkosti preferencií používateľov pri položkách, u ktorých nie je známa preferencia priamo. V takýchto prípadoch použije preferenciu podobnej, známej položky (Fernandez a kol. 2006).

4.4 FIT-Family

Systém slúži na odporúčanie multimediálneho obsahu rodinám. Snaží sa minimalizovať interakciu používateľov so systémom. Z tohto dôvodu FIT-Family nevyžaduje pred spustením procesu odporúčania explicitné identifikovanie prítomných členov. Vychádza z toho, že v rámci rodiny sa pri sledovaní televízie obmieňa určitá malá komunita. Každý jej člen pri úvodnej registrácii zadá svoje pohlavie, vek a stanoví dobu, kedy zvyčajne sleduje televíziu. Následne už nemusí so systémom komunikovať, pretože ten na základe pravdepodobnosti výskytu jednotlivých členov rodiny vyhodnotí kto aktuálne pravdepodobne televíziu sleduje. Podľa toho prispôbuje obsah.

Proces personalizácie obsahu spočíva vo vynásobení preferencií používateľov s pravdepodobnosťou ich prítomnosti. Preferencie systém vypočíta ako súčet korelovaných hodnotení používateľov. Hodnotenia sú však len odhadované podľa času, v ktorom používatelia dané položky pravdepodobne sledovali. Z tohto dôvodu systém trpí nedostatkom overených dát získaných od používateľov (Goren-Bar a Glinansky, 2004).

4.5 Happy Movie

Systém je dostupný ako aplikácia na sociálnej sieti Facebook. Je špecializovaný len na odporúčanie filmov. Jednotliví členovia skupín sú v systéme identifikovaní podľa troch aspektov. Prvým sú informácie získané zo danej sociálnej siete. Tie slúžia na určenie sociálnej dôvery používateľa. Druhý aspekt predstavujú výsledky získané z dotazníka, ktorým sa stanoví osobnosť používateľa. Dotazník musí používateľ vyplniť pred prvým odporúčaním. Tretí aspekt reprezentujúci používateľa predstavujú jeho hodnotenia položiek. Vykonanie určitého počtu hodnotení je taktiež nutná vstupná podmienka pre prvé odporúčanie.

Samotné odporúčanie je založené na sociálnych vzťahoch medzi používateľmi a tiež na histórii výsledkov minulých odporúčaní skupine. Odporúčanie predstavuje formu filtrovania obsahu, pričom sú vybrané pravdepodobne zaujímavé filmy. Tie sa potom zobrazia používateľom, ktorí hlasovaním rozhodnú o tom čo si skutočne pozrú. (Quijano-Sanchez a kol., 2010).

4.6 gRecs

Odporúčací systém gRecs pracuje nad dátovou sadou položiek z datasetu MovieLens⁶. Odporúčanie sa tu skladá z dvoch krokov. Prvým je kolaboratívne odporúčanie jednotlivcom, v ktorom sú položky vyberané na základe vysokých hodnotení od podobných používateľov. Druhým krokom je agregácia individuálnych odporúčaní do zoznamu určeného pre skupinu.

Z dôvodu optimalizácie procesu odporúčania je proces výpočtu podobných používateľov, využívajúci hierarchické zhľukovanie, predpočítavaný (Ntoutsis a Stefanidis, 2012A, 2012B).

4.7 Addaptive correlation-based RS

Odporúčací systém Addaptive correlatin-based RS pracuje podobne ako systém gRecs nad dátovou sadou položiek z datasetu MovieLens. Myšlienkou systému je odhad preferovaných hodnotení skupiny na základe aplikácie korelácie medzi skupinou a jej členmi. Dôležitosť jednotlivých členov skupiny je len odhadovaná. Na základe nej je však vypočítaný výsledný model skupiny pričom sa do úvahy berú preferencie jednotlivých používateľov s prihliadnutím na dôležitosť ich autora.

Systém preukázal, že môže byť nápomocný pri rozhodovaní kedy má byť určitý multimediálny obsah odporúčený (Lin a kol., 2011).

Analýzou existujúcich riešení skupinového odporúčania v doméne inteligentných televízií sme zistili, že reálne využívaných systémov nie je veľa a časť opisovaných riešení dokonca existuje len vo forme algoritmov. Vidíme tu preto dostatočný priestor na vývoj nových riešení. Z analýzy je tiež viditeľné, že existujúce riešenia využívajú len základné odporúčacie algoritmy a žiadne z nich nepracuje s hybridným odporúčaním. Navyše vidíme možnosti v preskúmaní vplyvu sociálnych väzieb medzi členmi skupiny. Zaujímavým zistením pre nás boli výsledky výskumu (O'Connor a kol., 2002) uvedené na začiatku kapitoly v Tabuľke 3. Z nich totiž vyplýva, že v praxi v danej doméne existujú takmer výlučne skupiny zložené z veľmi malého počtu členov (do veľkosti 4). Túto informáciu využijeme pri overovaní správania metódy navrhutej v tejto práci. Z analýzy tiež vyplýva, že systémy prevažne využívajú skutočné skupiny používateľov, v ktorých agregujú profily používateľov do spoločného profilu skupiny, na základe ktorého odporúčajú. Na odporúčanie využívajú prevažne kolaboratívny prístup, existujú však aj zástupcovia využívajúci odporúčanie na základe obsahu. Použité spôsoby agregácie preferencií sa pomerne dosť líšia (Tabuľka 2).

⁶ <http://www.grouplens.org/node/12>

5 Ciele práce

V súčasnosti existuje množstvo možností ako sa môže používateľ dostať k multimediálnemu obsahu. Problém, „čo si má pozrieť“, ktorý poznáme z minulosti je aktuálny aj dnes, avšak dôvod jeho vzniku je radikálne iný. Zatiaľ čo v minulosti nemal pozerateľ čo, v súčasnosti nevie čo si z rozsiahlej ponuky vybrať. Ako príklad uvádzame inteligentné televízie, ktoré poskytujú okrem bohatej ponuky staníc taktiež výber z rozsiahlych archívov obsahujúcich veľké množstvo položiek. Z tohto dôvodu musíme čoraz aktívnejšie čeliť problému informačného zahltenia, ktorý pri množstve dát, akým je dnes bežne používateľ zavalený, vzniká. Personalizované odporúčanie predstavuje efektívne riešenie, ktoré môžeme využiť na redukciu tohto problému.

Sledovanie televízie je však spoločenská aktivita. Z tohto dôvodu sa v našej práci zameriavame na personalizované odporúčanie skôr skupinám ako jednotlivcom. Z analýzy existujúcich riešení sme zistili, že pri odporúčaní multimediálneho obsahu sa autori metód zameriavajú na vytvorenie kvalitného modelu používateľa takmer výlučne pre metódy určené jednotlivcom, pričom v prípade skupín často použijú základný jednoduchý model a následne sa zamerajú radšej na prepracovanie spôsobu agregácie, či metódy vyberajúcej odporúčaný obsah.

Naším prvým cieľom je vytvoriť model používateľa, ktorý nám pomôže pri procese odporúčania. Model bude založený na váhovaní jednotlivých jeho vektorov na základe vedomostí, ako veľmi si podľa nich používateľ vyberá obsah v porovnaní s priemerným používateľom.

Na samotné odporúčanie použijeme zmiešanú hybridnú metódu primárne používajúcu kolaboratívne odporúčanie. Podobní používatelia tu budú vyberaní váhovanou kosínusovou podobnosťou. Výsledky tohto prístupu budú následne preusporiadané na základe obsahového odporúčania, ktoré skupine pomôže nájsť potenciálne zaujímavé položky, ktoré však nevidel dostatočný počet skupine podobných používateľov. Parametre, podľa ktorých bude systém položky vyhľadávať, budú ovplyvnené váhami jednotlivých vektorov v modeloch používateľov. Odporúčania určené jednotlivým členom skupiny budú agregované do spoločného zoznamu, pričom budeme uvažovať rozličnú dôležitosť členov skupiny. Z výsledného zoznamu bude vybraných niekoľko najvhodnejších položiek, ktoré budú odporúčené skupine.

V práci sa zameriavame na nasledujúce ciele:

- Získanie preferencií používateľov pomocou explicitnej spätnej väzby z ich hodnotení
- Vytvorenie modelu používateľa, kde:
 - pomocou zložiek vektorov sledujeme záujem používateľa o jednotlivé žánre, skupiny kľúčových slov, režisérov, hercov a o položky samotné (konkrétny žánre, herec atď.)
 - sledujeme, mieru osobnej preferencie každej zo spomenutých zložiek ako celku (teda ako sú pre používateľa dôležití napr. režiséri ako celok)
- Kolaboratívne odporúčanie jednotlivcom podľa ich preferencií v modeli používateľa
- Odporúčania na základe obsahu uvažujúce rozličnú váhu zložiek modelu používateľa
- Získanie informácií o sociálnych väzbách medzi používateľmi, na základe ich priateľstiev na sociálnej sieti
- Hybridné odporúčanie skupinám používateľov vychádzajúce z agregovaných kolaboratívnych odporúčaní jednotlivým členom skupiny uvažujúcich sociálne vzťahy medzi týmito používateľmi, doplnené agregovanými obsahovými odporúčaniami členom skupiny
- Výber a odporúčenie podmnožiny N najvhodnejších vybraných prvkov skupine

6 Návrh metódy skupinového odporúčania

V práci navrhujeme metódu personalizovaného odporúčania skupinám zameranú na odporúčanie multimediálneho obsahu. Navrhujeme zmiešanú hybridnú metódu využívajúcu kolaboratívne odporúčanie doplnené odporúčaním na základe obsahu. Členov skupiny agregujeme do skupiny na úrovni konkrétnych odporúčaní vytvorených pre nich ako jednotlivcov (Obrázok 11).

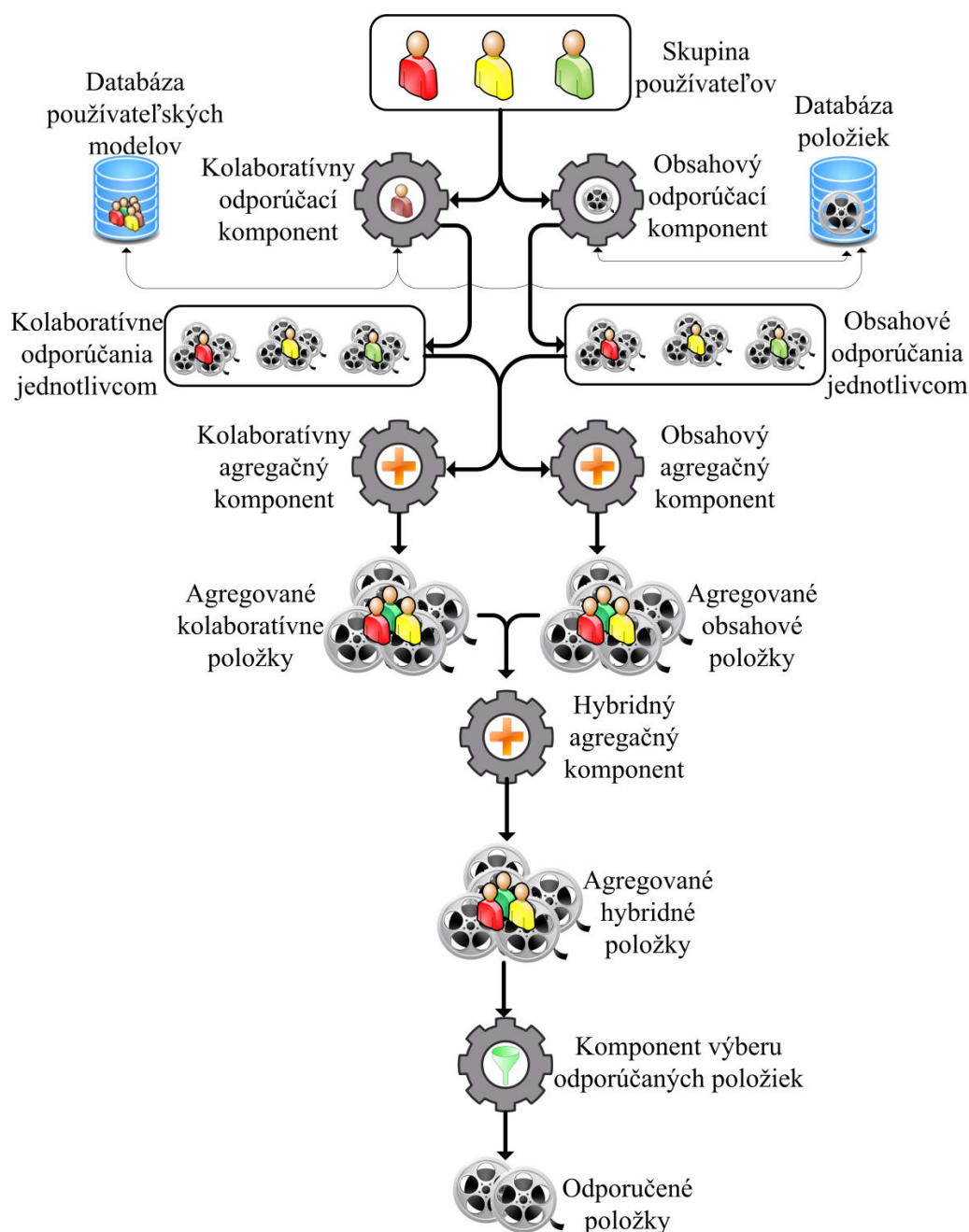
Preferencie jednotlivých používateľov získavame pomocou explicitnej spätnej väzby z ich hodnotení multimediálnych položiek. Tieto informácie uchováваме formou modelov používateľa, kde pomocou vektorových zložiek sledujeme záujem používateľov o jednotlivé žánre, kľúčové slová režisérov, hercov a o samotné položky. U jednotlivých používateľov sledujeme tiež, do akej miery preferujú každú zo spomenutých zložiek ako celok. Teda ako veľmi je pre nich vo všeobecnosti dôležité kto vo filme hrá, prípadne ho režíruje a podobne.

Pre každého používateľa evidujeme okrem jeho preferencií uložených v modeli používateľa tiež mieru jeho podobnosti voči iným používateľom, ktorú využívame pri výpočte kolaboratívneho odporúčania. Ďalej tiež výsledky kolaboratívneho a obsahového odporúčania určených modelovanému používateľovi ako jednotlivcovi.

Proces odporúčania skupine používateľov prebieha v niekoľkých krokoch. V prvom spojíme individuálne kolaboratívne odporúčania jednotlivcov do spoločného kolaboratívneho odporúčania, pričom berieme do úvahy tiež sociálne vzťahy medzi členmi skupiny.

V druhom kroku spájame výsledky odporúčaní na základe obsahu určených jednotlivým členom skupiny. Podobne ako v prvom kroku výsledky jednotlivcov spojíme do spoločného zoznamu, pričom opäť uvažujeme rozličnú dôležitosť členov skupiny.

Tretím krokom procesu je hybridný prístup, v ktorom kombinujeme agregované kolaboratívne a obsahové odporúčanie. Primárne tu vychádzame z výsledkov kolaboratívneho prístupu. Tieto výsledky však preusporiadavame, pričom na vyššie pozície posúvame položky, ktoré sa súčasne vyskytujú aj vo výsledkoch obsahového odporúčania. Následne z výsledného zoznamu vyberieme top N položiek, ktoré prezentujeme skupine používateľov.



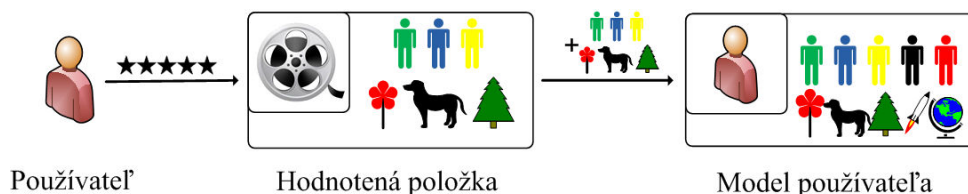
Obrázok 11. Tok dát v navrhovanej metóde

6.1 Reprezentácia multimediálneho obsahu

Pre potreby našej metódy potrebujeme vedieť o multimediálnom obsahu informácie, ktoré ho čo najlepšie charakterizujú. Za podstatné pokladáme pre každú položku poznať žánre, kľúčové slová a osoby, ktoré v nej pôsobia. Tu máme na mysli najmä režiséra a hercov v hlavných úlohách. Na základe týchto informácií o položkách dokážeme vyhodnotiť o aký obsah majú jednotliví používatelia záujem, čo využijeme pri kolaboratívnom odporúčaní a taktiež dokážeme z viacerých pohľadov charakterizovať položky, čo slúži pre účely odporúčania na základe obsahu.

Opísané druhy metadát (herci, žánre a pod.) reprezentujeme v modeli používateľa ako množinu nezávislých vektorov, spoločne charakterizujúcich jednotlivé položky (napr. filmy). Každý vektor je tvorený zoznamom prvkov, ktoré do neho patria. Pre prvky rozlišujeme názov (respektíve jedinečný identifikátor), priemerné hodnotenie a počet hodnotení. Hodnotenie

jednotlivých prvkov vektorov vzniká tým, že používateľ hodnotí konkrétne položky (napr. filmy). Tie sú definované prvkami (žánrami, kľúčovými slovami režisérmi, hercami a identifikátorom). Po ohodnotení položky je hodnotenie zarátané v modeli používateľa ku všetkým jej prvkom (Obrázok 12). Pri jednotlivých typoch prvkov je uvedená ich váha. Skutočná váha zarátavaná do modelu používateľa je však závislá aj od používateľovho explicitného hodnotenia celej položky. Vzorec vyjadrujúci princíp zarátavania je opísaný v aktuálnej kapitole v sekcii *Model používateľa* (nižšie).



Obrázok 12. Princíp vytvárania používateľského modelu hodnotením položiek

Žánre

Žánre sú pre jednotlivé filmy opísané ako rovnocenná skupina slov. Všetkým žánrom určitej položky preto pripisujeme rovnakú váhu, nezávisle od toho, koľko žánrov položku charakterizuje. Dôvodom rozličného počtu priradených žánrov je, že niektoré položky môžeme zaradiť súčasne do rôznych žánrov. Ako príklad môžeme uviesť položku, ktorej dej sa odohráva v minulosti, je vtipný, ale zároveň sa vyvíja akčne. Takúto položku teda označíme za historickú, komediálnu a akčnú. Nedokážeme pri nej však povedať, že je historická na x percent a vtipná na y , preto žánre explicitne neváhuje a predpokladáme u všetkých rovnakú váhu rovnú 1 (maximálna váha).

Kľúčové slová

Kľúčové slová predstavujú váhovaný zoznam prevažne jednoslovných charakteristík danej položky. Podľa pozície konkrétneho kľúčového slova v zozname slov charakterizujúcich určitú položku dokážeme určiť jeho relevantnosť pre konkrétnu položku. Kľúčové slová nám pomáhajú identifikovať obsah položky, pretože ten v prípade multimediálneho obsahu nedokážeme identifikovať priamo (ako je to napr. u textov, kde je možné urobiť analýzu textu) a sme odkázaní na nepriamy opis formou metadát.

Váhu kľúčových slov reprezentujeme reálnym číslom z intervalu $<0;1>$. V rámci 1 položky kľúčové slová rozmiestňujeme rovnomerne v tomto intervale, pričom najrelevantnejším priradíme váhu blízku sa maximálnej hodnote 1 a najmenej relevantným priradíme váhu blízku sa 0 .

Režiséri

V doméne multimediálneho obsah, konkrétne filmov, rozlišujeme dva typy osôb, ktoré sú zaujímavé z pohľadu charakteristiky položiek. Prvý typ predstavujú režiséri, ktorí často vnášajú do svojich diel charakteristické črty, na základe ktorých sú ich diela podobné. Používatelia si môžu k položkám vybudovať určitý vzťah len na základe ich sympatií s režisérovmi predchádzajúcimi dielami. Režiséri, podobne ako kľúčové slová, majú pre položku vždy maximálnu váhu rovnú 1 .

Herci

Druhý typ významných osôb predstavujú herci. Tu podobne ako pri režiséroch do určitej miery platí, že konkrétni herci dávajú položkám, v ktorých pôsobia určitý štýl a preto si používatelia často zvolia položku s hercami, ktorých poznajú a ktorých prejav sa im páči. Jedna osoba môže v rôznych dielach vystupovať raz ako režisér a inokedy ako herec. Z nášho pohľadu sa však jedná

o odlišné výskyty, ktoré zo sledovaného pohľadu nesúvisia. Hercom priradíme váhu podľa toho či pôsobia v hlavnej alebo vo vedľajšej roli.

Položky

Okrem opisných metadát, ktoré charakterizujú vlastnosti a črty položiek, uchováame v modeli používateľa tiež presnú informáciu o tom, ktoré konkrétne položky videl. Táto informácia síce nevypovedá o type položiek, ktoré používateľ preferuje, pomôže nám však presnejšie nájsť napríklad podobných používateľov. Zavedenie tohto typu informácií do modelu používateľa sa nám pri odporúčaní výrazne osvedčilo. Položka má vždy váhu 1.

6.2 Model používateľa

Model používateľa uchováva používateľove preferencie, jeho dôležitosť a priateľov. Model je vytváraný postupne, na základe používateľovej aktivity. Využívame ho na vytváranie jednotlivých odporúčaní (kolaboratívne jednotlivcovi, obsahové jednotlivcovi a hybridné skupine).

Preferencie každého používateľa uchováame pomocou 5 vektorov, reprezentujúcich jeho preferenciu jednotlivých žánrov, kľúčových slov, režisérov, hercov a samotných položiek. Vektory sú tvorené vždy množinou všetkých existujúcich prvkov v doméne (napr. všetky žánre, herci, a pod.). Každý prvok vektora má v modeli uvedený názov (respektíve identifikátor) hodnotu a váhu. Druhé dva údaje sú špecifické pre používateľa, ktorému model prislúcha. Hodnota je daná aritmetickým priemerom hodnotení tohto používateľa všetkých položiek obsahujúcich vo svojich metadátach daný prvok vektora, váha je daná počtom takýchto hodnotení. Tabuľka 4 ukazuje príklad vektorového modelu, pričom pre každý prvok uvádza jeho názov, hodnotu a váhu.

Tabuľka 4. Príklad vektorového modelu používateľa. Prvky vektora predstavujú trojice – názov, hodnota, váha prvku

Vektor	Prvky vektora
Žánre	[[komédia;4.0;3], [akčný;3.8;5], [dráma;1.5;2], [vojnový;2.6;4], ...]
Kľúčové slová	[[dobrodružstvo;3.4;6], [hobit;4.5;2], [kraj;3.1;1], [mordor;3.1;1], ...]
Režiséri	[[jackson;4.1;5], [tarantino;4.2;3], [conan;1.2;2], [spileberg;3.4;3], ...]
Herci	[[parsons;4.1;3], [streep;3.2;8], [cuoco;4.1;3], [cumberbatch;4.9;4], ...]
ID položiek	[[sherlock;5.0;1], [ironman;3.0;1], [watchmen;4.0;1], [hours;4.5;1], ...]

Po hodnotení položky (napr. filmu) sa v modeli používateľa hodnotenie ($hod_{položky\ nové}$) pridá ku každému prvku tejto položky (žánre, herci atď.). V prípade že používateľ položku hodnotil prvýkrát $hodnota\ prvku$ sa vypočíta podľa Vzorca 2, následne je potrebné zvýšiť $váhu\ prvku$.

$$hodnota_{prvku-priadnie} = \frac{hodnota_{prvku} * váha_{prvku} + hod_{položky\ nové}}{váha_{prvku} + 1} \quad (2)$$

V prípade, že používateľ znovu hodnotil položku, ktorú už ohodnotil v minulosti, započíta sa do $hodnoty\ prvku$ len zmena hodnotenia podľa Vzorca 3, $váha\ prvku$ sa v tomto prípade nemení.

$$hodnota_{prvku-zmena} = \frac{hodnota_{prvku} * váha_{prvku} - hod_{položky\ staré} + hod_{položky\ nové}}{váha_{prvku}} \quad (3)$$

Individuálne preferencie vektorov

Z prvkov jednotlivých vektorov a ich hodnôt je možné určiť oblasti záujmu používateľa. Predpokladáme však, že odlišní používatelia prikladajú zmieneným vektorom rozličnú váhu. Niektorí si napríklad film vyberá vyslovene podľa toho, kto ho režíroval, inému je táto informácia úplne ľahostajná. Z tohto dôvodu sledujeme, na ktoré prvky vektorov sa používateľ zameriava viac ako je bežné. Zistené odchýlky pokladáme za individuálne preferencie používateľa a podľa nich nastavíme váhu jeho jednotlivým vektorom. Na zistenie odchýlok porovnáваме model používateľa so štandardom, ktorý sme vytvorili na základe modelov veľkého množstva používateľov.

Definujeme pojem Priemerné preferencie, ktorý vyjadruje informáciu o záujme používateľov o jednotlivé typy vektorov. Hodnoty získame z veľkej množiny modelov používateľov (rádovo tisíce) ako aritmetický priemer preferencií jednotlivých používateľov.

Pre zistenie individuálnych preferencií používateľa, ktorému budeme odporúčať, vypočítame pre každý z vektorov sumu odchýlok početnosti voči Priemerným preferenciám delenú počtom uvažovaných položiek - Vzorec 4. Počítame teda priemernú odchýlku, ktorá predstavuje absolútnu hodnotu rozdielu početností u daného používateľa oproti priemernému používateľovi. Tento spôsob váhovania vektorov modelu používateľa v oblasti multimediálneho obsahu v minulosti použil Ferman (Ferman a kol., 2003). V jeho prístupe však počítali sumu odchýlok vypočítaných ako podiel preferencie hodnoty používateľa voči priemernej preferencii (my počítame rozdiel preferencií, pretože podľa nás reálnejšie zachytáva meranú odchýlku).

$$\text{Preferencia vektora} = \frac{\sum_{i=0}^N |\overline{\text{váha prvku}_i} - \text{váha prvku}_i|}{N} \quad (4)$$

N predstavuje počet prvkov vektora, pre ktorý počítame mieru *preferencie*. Váha určitého prvku je daná počtom položiek, ktoré používateľ hodnotil a ktoré zároveň obsahovali danú položku (napríklad hodnotené filmy, ktorých žánr bol komédia) krát váha prvku (napr. u žánrov rovná 1, u kľúčových slov v intervale $<0;1>$).

Po vypočítaní preferencií všetkých 5 vektorov, sú preferencie ešte znormované tak, aby sa ich súčet rovnal 1. Tento výpočet normovaných preferencií používateľa vypočítame podľa nasledujúceho postupu:

1. Podľa predchádzajúceho postupu vypočítame preferencie 5 vektorov modelu používateľa.
2. Vypočítame prevrátenú hodnotu zo sumy absolútnych hodnôt preferencií vypočítaných v kroku 1.
3. Pre každý vektor vynásobíme výsledok pre daný vektor z kroku 1 so spoločným výsledkom z kroku 2.

Takto dostaneme váhované preferencie, ktoré personalizujú dôležitosť jednotlivých vektorov modelu používateľa pre daného používateľa. Znормovaním dostaneme pre každý vektor hodnotu preferencie z intervalu $<0;1>$. Pokiaľ sčítame normované preferencie všetkých 4 vektorov v modeli používateľa, dostaneme vždy 1.

Sociálny koeficient používateľa založený na priateľstvách v skupine

Okrem vyššie opísaných vektorov a miery ich preferencie uchováваме v modeli používateľa aj zoznam jeho priateľov. Tieto údaje využívame na určenie sociálneho koeficientu používateľa v jeho jednotlivých skupinách. Koeficient určuje mieru vplyvu používateľa v rámci skupiny. Určuje sa individuálne pre každé pôsobenie modelovaného používateľa v konkrétnej skupine. Použitie takéhoto koeficientu vo svojej práci prezentovali napríklad Liu a kol. (Liu a kol., 2012) a rovnako tím okolo De Pessemier-a (De Pessemier a kol., 2012).

Podnetom k jeho zavedeniu bol fakt, že priatelia predstavujú množinu používateľov, ktorí sú modelovanému používateľovi blízki a ku ktorým má vzťah. Vďaka tomu je popri presadzovaní svojich preferencií skôr ochotný brať ohľad na priateľov a ich záujmy ako na iných členov skupiny a rovnako sú jeho preferencie realizované konaním priateľov. Pomocou sociálneho koeficientu teda modelujeme vzťahy v skupine, ktoré v našom ponímaní reprezentujú mieru emocionálneho šírenia. Význam emocionálneho šírenia v skupine vo svojej práci opisuje Qujano (Qujano, 2013).

V navrhovanej metóde pracujeme so šírením pozitívnych emócií. Predpokladáme, že používatelia sú spokojnejší v prípade ak vidia, že sú spokojné aj im blízke osoby. Pri určovaní priateľstiev využívame prepojenie odporúčacieho systému s vybranou sociálnou sieťou (napríklad sieť Facebook), odkiaľ vzťahy extrahujeme.

Sociálny koeficient každého používateľa je v skupine neutrálnych ľudí rovný 1 pre každého používateľa. Za každého priateľa v skupine sa však modelovanému členovi skupiny zvýši jeho koeficient o určitú pevne stanovenú hodnotu. Pri základe koeficientu rovnom 1, ktorý používame, je podľa Liu-ho a kol. vhodné použiť prídavok k dôležitosti na úrovni približne 0,1.

6.3 Proces odporúčania

Odporúčanie v našom systéme realizujeme na dvoch úrovniach. Primárne sa síce zameriavame na odporúčanie skupinám používateľov, no ako medzikrok tohto procesu vytvárame tiež odporúčania jednotlivcom. Na tejto úrovni odporúčame kolaboratívne a obsahovo.

Po tom čo viacero používateľov vytvorí skupinu, odporúčame im obsah pomocou hybridnej metódy. Takéto odporúčanie je určené jednotne pre celú skupinu. Proces opisujúci priebeh hybridného odporúčania je opísaný nižšie.

6.3.1 Kolaboratívne odporúčanie jednotlivcom

Kolaboratívne odporúčanie je založené na odporúčaní obsahu, ktorý sa páčil používateľom s podobnými preferenciami ako jednotlivec, ktorému aktuálne odporúčame.

Podobnosť preferencií vyhodnocujeme pomocou váhovanej kosínusovej podobnosti. Ako je možné vidieť vo Vzorcí 5, jedná sa o štandardnú kosínusovú podobnosť, v ktorej však pre každý vektor berieme do úvahy jeho váhu (danú preferenciou používateľa). Parameter M predstavuje počet vektorov, N_j predstavuje počet prvkov vektora j . $Pref_{xj}$ je súčin hodnoty individuálnej preferencie vektora j pre používateľa x so statickou preferenciou závislou od konkrétneho vektora (napríklad žánrom priradujeme staticky mierne vyššiu váhu ako kľúčovým slovám).

$$Podobnosť\ používateľov(x, y) = \frac{\sum_{j=1}^M pref_{xj} pref_{yj} \sum_{i=1}^{N_j} x_{ji} y_{ji}}{\sqrt{\sum_{j=1}^M pref_{xj}^2 \sum_{i=1}^{N_j} x_{ji}^2} * \sqrt{\sum_{j=1}^M pref_{yj}^2 \sum_{i=1}^{N_j} y_{ji}^2}} \quad (5)$$

Týmto spôsobom určíme podobnosť používateľa, ktorému odporúčame, voči ostatným používateľom. Podobných používateľov potom zoradíme podľa miery podobnosti, pričom ďalej pracujeme len s podmnožinou P používateľov najvyššou podobnosťou voči skupine.

Položky, ktoré títo vybraní používatelia videli, hodnotili a zároveň ich nevidel používateľ, ktorému odporúčame, vložíme do zoznamu položiek, ktoré by mali byť pre neho zaujímavé. Jednotlivým prvkom v zozname priradíme body, podľa ktorých ich vzájomne usporiadame. Body každej z položiek vypočítame podľa Vzorca 6. Kde p predstavuje počet používateľov z množiny P , ktorí danú položku hodnotili. K tomuto číslu ešte pripočítame aritmetický priemer hodnotení delený koeficientom k , ktorého úlohou je posunúť skóre hodnotenia o jeden rád doprava. Takto dosiahneme stav, kedy sú položky usporiadané podľa počtu používateľov, ktorí ich videli a v rámci rovnako videných pozícií sú položky usporiadané podľa výšky priemerného hodnotenia.

$$\text{Skóre položky}_{x-\text{kolaboratívne}} = p + \frac{\sum_{i=1}^p \text{hodnotenie}(i,x)}{k * p}; p \in P \quad (6)$$

Výsledkom navrhutej metódy kolaboratívneho odporúčania je usporiadaný a ohodnotený zoznam položiek, ktoré sú potenciálne zaujímavé pre používateľa, ktorému sme odporúčali. Za výstup metódy pokladáme top N položiek s najvyšším skóre.

6.3.2 Odporúčanie jednotlivcom na základe obsahu

V navrhovanej metóde využívame okrem kolaboratívneho odporúčania tiež odporúčanie na základe obsahu. Využívame ho najmä ako doplnok prvého menovaného prístupu. Kolaboratívne odporúčanie síce všeobecne dosahuje vyššiu presnosť ako odporúčanie založené na obsahu, avšak obsahové je menej náchylné na niektoré situácie, s ktorými má kolaboratívny prístup ťažkosti (napr. problém studeného štartu týkajúci sa nových používateľov, ktorí ešte nehodnotili dostatok položiek na to, aby sme im identifikovali podobných používateľov). Druhou výhodou obsahového prístupu je fakt, že ľahšie odporúča nové položky, ktoré ešte komunita podobných používateľov dostatočne neobjavila. Naším riešením známych obmedzení kolaboratívneho prístupu je úprava poradia jeho výsledkov na základe vhodnosti z pohľadu obsahového prístupu.

Pre každý zo 4 vektorov modelu používateľa opisujúcich položky z pohľadu ich metadát (teda okrem vektora obsahujúceho samotné videové položky), vyhľadáme v množine všetkých existujúcich položiek také, ktoré obsahujú niektorý z N prvkov s najvyššou preferenciou v rámci niektorého vektora v modeli používateľa. Inými slovami hľadáme filmy obsahujúce niektorého z top N používateľom najpreferovanejších režisérov, niektorého z top N najpreferovanejších hercov, žánrov alebo kľúčových slov. Jednotlivým položkám priradíme skóre za každé z hľadaných metadát, ktoré obsahujú. Priradené skóre sa líši v závislosti od typu metadát, pretože prikladáme napríklad rozličnú dôležitosť režisérom v porovnaní s kľúčovými slovami a podobne.

Skóre pre každú z položiek vypočítame podľa Vzorca 7 ako sumu M dopytov, kde M predstavuje počet používaných vektorov metadát, s ktorými pracujeme (v našom prípade 4 - herci, režiséri, žánre, kľúčové slová). Pre každý z týchto vektorov počítame súčin koeficientu dôležitosti tohto vektora k_M a veľkosti prieniku množiny N používateľom najpreferovanejších prvkov daného vektora (označené X) a množiny metadát opisujúcich hodnotenú položku (označené Y).

$$\text{Skóre položky}_{x-\text{obsahové}} = \sum_{i=1}^M k_M * |X \cap Y| \quad (7)$$

Výsledkom procesu priradovania skóre je pre každú existujúcu položku (film) informácia, či a nakoľko je vhodná na odporúčanie z pohľadu jej vlastností. Z tejto informácie vieme vytvoriť, podobne ako pri kolaboratívnom odporúčaní, usporiadaný a ohodnotený zoznam položiek, ktoré sú potenciálne zaujímavé pre používateľa, ktorému sme odporúčali. Do zoznamu zaradíme top N položiek s najvyšším dosiahnutým skóre, ktoré predstavujú výstup tejto metódy odporúčania.

6.3.3 Hybridné odporúčanie skupine

Primárnym cieľom našej práce je však odporúčanie skupinám používateľov. Na tento účel sme navrhli hybridnú metódu, vychádzajúcu z agregovaných kolaboratívnych odporúčaní jednotlivým členom skupiny a agregovaných odporúčaní jednotlivým členom skupiny založených na obsahu.

Pri agregácii kolaboratívnych odporúčaní metóda využíva sociálne koeficienty jednotlivých členov, ktoré vyjadrujú dôležitosť s akou sa jednotlivé odporúčania zarátajú do spoločného výsledku (Vzorec 8). Agregované skóre položky vypočítame ako sumu čiastkových skóre tejto

položky všetkých členov skupiny (S) vynásobených sociálnym koeficientom daného člena skupiny ($k_{soc.}$). Čiastkové skóre položiek však určujeme len na základe ich poradia vo výslednom zozname aktuálne agregovaného typu odporúčania, nie ako skóre, ktoré sme vypočítali v rámci daného typu odporúčania. Skóre, s ktorým pracujeme, teda vypočítame ako počet prístupom odporučených položiek (OP) mínus poradie hodnotenej položky v tomto zozname (P).

$$Skóre\ položky_{x\ kolab.\ agreg.} = \sum_{i=1}^{|S|} k_{soc. i} * skóre_{i\ čiast.}; skóre_{čiast.} = |OP| - P \quad (8)$$

Agregácia obsahových odporúčaní je jednoduchšia, pretože nás zaujíma len informácia, koľkokrát sa jednotlivé položky nachádzajú na zoznamoch jednotlivcov (Vzorec 9). S predstavuje počet členov skupiny, OP_i je množina odporučených položiek používateľovi i .

$$Skóre\ položky_{x\ obsah.\ agreg.} = \sum_{i=1}^{|S|} (1\ ak\ položka_x \in OP_i; inak\ 0) \quad (9)$$

Po vytvorení agregovaných zoznamov pre oba typy odporúčaní aplikujeme posledný krok navrhnutého hybridného procesu, ktorým je úprava skóre výsledkov kolaboratívneho odporúčania na základe výsledkov obsahového prístupu. Tu postupujeme podľa Vzorca 10, ktorý však aplikujeme len na položky nachádzajúce sa v oboch agregovaných množinách. Položkám, ktoré odporučil len kolaboratívny prístup skóre nemeňte.

$$Skóre\ položky_{x-hybrid.} = skóre\ položky_{x\ kolab.\ agreg.} * skóre\ položky_{x\ obsah.\ agreg.} \quad (10)$$

Výsledkom hybridného odporúčania je teda zoznam výsledkov kolaboratívneho prístupu usporiadaný podľa hybridného skóre. Z takto získanej množiny položiek vyberieme K položiek s najvyšším skóre, ktoré prezentujeme skupine používateľov.

6.4 Prezentácia odporúčania skupine

Po získaní finálneho zoznamu položiek, do ktorého sú zahrnuté výsledky oboch prístupov je potrebné určitý obsah vybrať a odporučiť ho skupine.

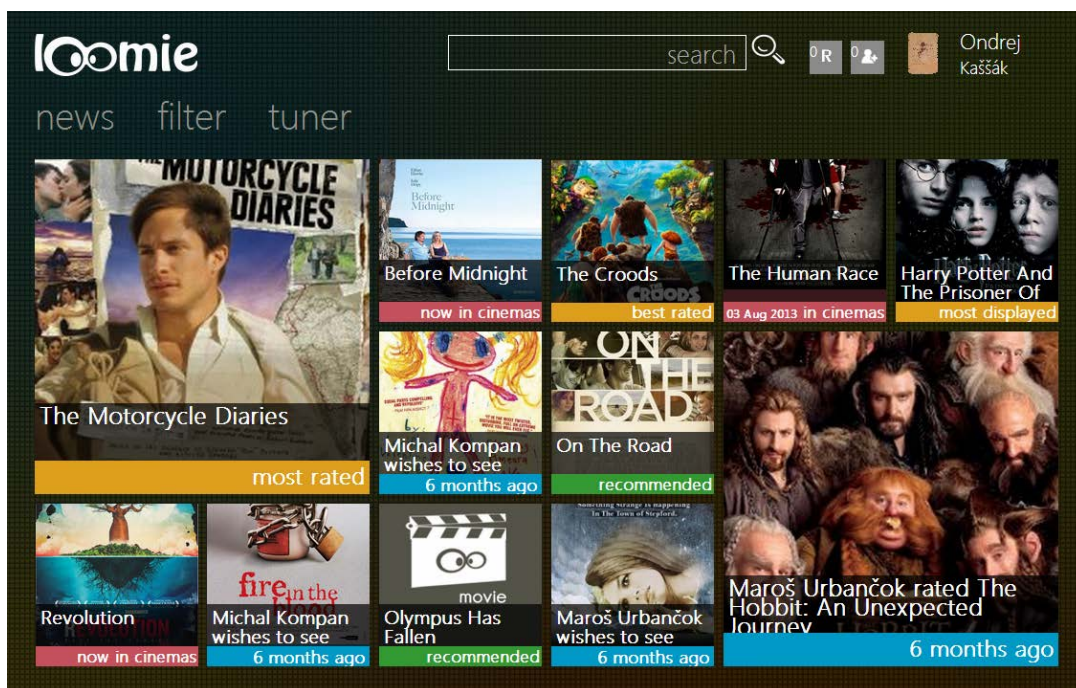
V oblasti, na ktorú sa zameriavame majú položky zvyčajne tak veľký rozsah, že si skupina v rámci jedného sedenia dokáže pozrieť jedinú položku. Domnievame sa však, že je vhodné skupine zobrazit' viac ako jeden výsledok a nechať jej členov aby si z odporučených položiek vybrali najvhodnejšiu podľa ich aktuálnej nálady, prípadne ďalších drobných aspektov, ktoré odporúčacia metóda neuvažuje.

Množina odporučených výsledkov, však na druhej strane nemôže byť príliš rozsiahla, pretože v tomto prípade by sa strácal zmysel odporúčania, keďže skupina by si v podstate musela sama zvoliť položku z predloženej množiny, ktorá by len bola mierne menšia ako bez odporúčania.

Pri určovaní rozdielu medzi tým, kedy je veľkosť zobrazenej množiny výsledkov akceptovateľná a kedy je už príliš veľká sme sa inšpirovali prácou Quijano-Sanchez a kol. (Quijano-Sanchez a kol., 2013). Tí vo svojej práci tvrdia, že používatelia chcú dostať odporučené najvyššie 5 položiek. Pri tomto množstve totiž dokážu rozhodnúť, ako veľmi chcú ktorú položku pozerať. Pri viacerých položkách totiž nedokážu určiť presné poradie preferencií a teda sa nedokážu dohodnúť kvalitne ako skupina. Z tohto dôvodu skupine odporúčame množinu 5 najvhodnejších položiek, ktoré prezentujeme ako zoradený zoznam.

7 Realizácia metódy

Navrhnutú metódu skupinového odporúčania multimediálneho obsahu sme implementovali v rámci portálu Loomie⁷, zameraného taktiež na poskytovanie informácií o multimediálnom obsahu. Tento portál bol vytvorený v rámci predmetu Tímový projekt na Fakulte informatiky a informačných technológií Slovenskej technickej univerzity v Bratislave. Náhľad úvodnej stránky portálu je uvedený na Obrázku 13.



Obrázok 13. Náhľad úvodnej obrazovky odporúčacieho portálu Loomie.

Portál Loomie o multimediálnom obsahu poskytuje komplexné informácie ako dátum vydania filmu, obsadenie, jazyk, trajler či komentáre používateľov. Registrovaným používateľom umožňuje taktiež položky hodnotiť k čomu ich dokonca motivuje zobrazovaním ich aktivity na nástenkách priateľov, či priamym prepojením s účtom na sociálnej sieti Facebook, kde majú používatelia možnosť zdieľať svoju vybranú aktivitu.

Hodnotenie položiek na portáli je realizované pomocou N -árnej bodovej stupnice, pričom N je aktuálne stanovené na úrovni 5 (Obrázok 14). Hodnotenia používateľov sú uchovávané a môžu byť využité na modelovanie používateľa, napríklad pre účely odporúčania.



Obrázok 14. Ukážka 5-hodnotovej bodovej stupnice slúžiacej na hodnotenie položiek.

⁷ <http://team04-12.ucebne.fiit.stuba.sk/>

7.1 Zdroje dát

Navrhovaná metóda skupinového odporúčania vyžaduje okrem zberu hodnotení položiek používateľmi aj dodatočné informácie o samotných položkách. Portál Loomie využíva dáta získané z Internetovej filmovej databázy - IMDb⁸. Takto získavané údaje o položkách obsahujú originálny názov diela, oficiálny preklad názvu v anglickom jazyku, žáner, kľúčové slová, zoznamy hercov, režisérov spolu s informáciami o nich, obsah diela, dátum jeho vydania, dĺžku a podobne. Samotný portál Loomie ako nadstavbu poskytuje napríklad doplnené kľúčové slová, nakoľko tieto informácie do IMDb zadávajú priamo používatelia a tak ich kvalita nie je garantovaná. Loomie pri ich doplnení a prípadnom preusporiadaní využíva textovú analýzu obsahu diela, založenú na využití algoritmu TF-IDF, čiže na frekvencii výskytu slov v dokumente a invertovanej početnosti slov vo všetkých dokumentoch (angl. term frequency – inverse document frequency).

V navrhovanej metóde pri individuálnom váhovaní vektorov modelu používateľa vyžadujeme na určenie priemerných preferencií množinu používateľov, pre ktorých potrebujeme poznať hodnotené položky, spolu s veľkosťou hodnotenia. Keďže pomocou navrhovaných experimentov by sme nevedeli zaistiť dostatočne veľkú vzorku používateľov, preto sme využili hodnotenia z datasetu MovieLens 10M⁹. Dataset obsahuje údaje o hodnotení 69 878 používateľov, z ktorých každý ohodnotil aspoň 20 položiek, celkovo obsahuje 10 000 054 hodnotení 10 681 filmov.

Údaje sú v datasete uvedené vo forme trojíc - identifikátor používateľa, identifikátor filmu, hodnotenie (5 bodová stupnica s možnosťou zadania pol bodu). Pre hodnotené položky je však v datasete uvedená aj dvojica - identifikátor filmu a reťazec obsahujúci názov v angličtine spolu s rokom vydania. Pre naše účely preto bolo potrebné mapovať hodnotené položky na položky v systéme Loomie, aby sme o nich získali informácie o žánroch, kľúčových slovách, režiséroch a hercoch.

Proces mapovania položiek sme realizovali formou skriptu, ktorý pre vyhodnotenie položiek ako zodpovedajúcich vyžadoval rovnaký rok vydania a názov zhodujúci sa aspoň v jednej z overovaných možností:

- Úplne zhodné názvy
- Názvy zhodné po odstránení nealfanumerických znakov
- Názvy zhodné po odstránení členov (a, an, the)

Po tomto mapovaní v datasete ostalo 8 568 390 hodnotení, pre ktoré sme vedeli jednoznačne identifikovať hodnotenú položku. Pri realizácii metódy pracujeme s takto upraveným datasetom.

7.2 Model používateľa

V navrhnutom modeli používateľa uchováваме preferencie používateľa jednak formou konkrétnych hodnotení a jednak formou spracovaných vektorov preferencií. Vďaka tomu, že si pamätáme samotné hodnotenia položiek, dokážeme následne pri ďalších používateľových hodnoteniach rozlišovať či hodnotil novú položku alebo modifikoval svoje rozhodnutie z minulosti. Vďaka uchovávaní preferencií aj vo forme spracovaných vektorov jednotlivých typov metadát dokážeme rýchlejšie vytvárať odporúčania.

V modeli používateľa si preferencie uchováваме ako trojice *identifikátor elementu x priemerné hodnotenie x počet hodnotení*. Kde *priemerné hodnotenie* predstavuje vážený aritmetický priemer hodnotení položiek obsahujúcich medzi svojimi metadátami daný element, *počet hodnotení* predstavuje počet takýchto hodnotených položiek. Každá z trojíc opisuje element

⁸ <http://www.imdb.com/>

⁹ <http://www.grouplens.org/node/12>

prislúchajúci do jedného z 5 modelovaných vektorov (žánre, herci a pod.). Okrem preferencií jednotlivých prvkov si v modeli používateľa pamätáme taktiež individuálne preferencie vektorov ako celkov. Tieto atribúty využívame na modifikáciu kosínusovej podobnosti pri určovaní podobnosti používateľov. Súčasťou modelu je tiež váha používateľa, ktorá je dôležitá pri vytváraní skupín. Závisí od osobnej dôležitosti a priateľov. Priatelia predstavujú množinu odkazov na iných používateľov.

Využívame vlastný model používateľa nezávislý od modelu využívaného v systéme Loomie, v ktorom sme našu metódu implementovali. Tento model používateľa totiž nezachytáva všetky parametre, ktoré vyžaduje naša metóda.

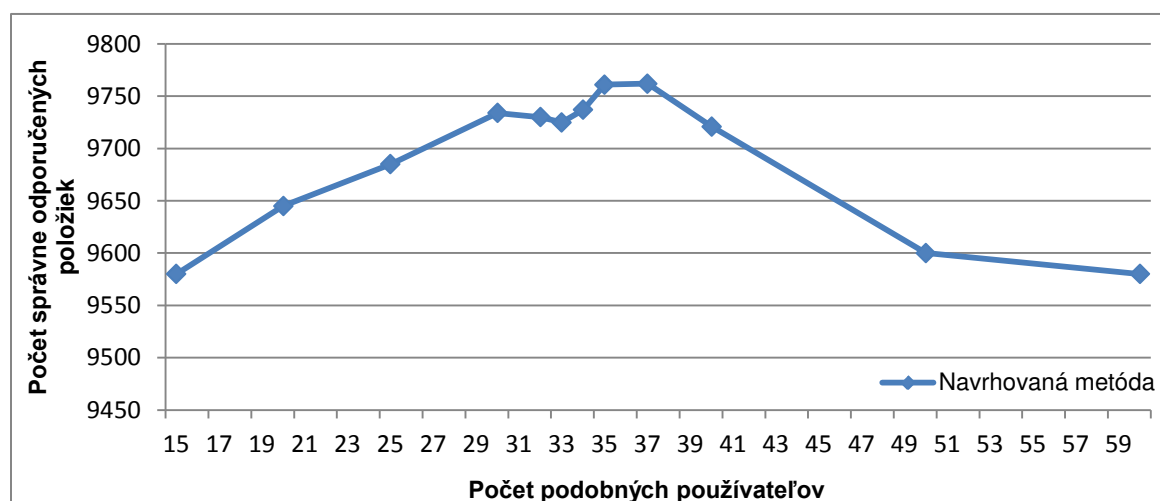
7.3 Odporúčanie

Navrhnuté hybridné odporúčanie je rozdelené do niekoľkých krokov. V úvode každému používateľovi individuálne vytvoríme kolaboratívne a obsahové odporúčanie. Tieto kroky sú predpočítavané, pretože nie sú závislé od aktuálneho zloženia skupiny a sú pomerne časovo náročné. Po vytvorení skupiny sú individuálne odporúčania agregované (oba typy nezávisle). Posledným krokom je spojenie oboch výsledkov do spoločného zoznamu, z ktorého niekoľko položiek, ktorým sme v predchádzajúcich krokoch priradili najvyššie skóre, odporučíme skupine.

7.3.1 Kolaboratívne odporúčanie jednotlivcom

Ako prvé sme implementovali kolaboratívne odporúčanie. V rámci neho určujeme vzájomnú podobnosť používateľa s ostatnými používateľmi systému. Na tento účel využívame tiež modely používateľov vytvorené na základe IMDb dát, vďaka čomu disponujeme pri porovnávaní s tisíckami používateľov.

Na určenie podobnosti používateľov využívame navrhnutú metódu váhovanej kosínusovej podobnosti. Následne pri kolaboratívnom odporúčaní jednotlivcovi pracujeme s množinou N najpodobnejších používateľov. Pri stanovení optimálnej hodnoty koeficientu sme vychádzali z výsledkov experimentu, ktorého výsledky vidíme na Obrázku 15.



Obrázok 15. Závislosť počtu správne odporučených položiek od počtu podobných používateľov braných do úvahy pri kolaboratívnom odporúčaní navrhovanou metódou.

V procese hľadania optimálneho počtu najpodobnejších používateľov, ktorých berieme do úvahy pri kolaboratívnom odporúčaní, sme spracovali hodnotenia 3 000 používateľov z MovieLens 10M datasetu. Každému používateľovi sme náhodne zakryli 20% jeho hodnotení a pomocou

navrhovanej metódy sme zo zvyšných hodnotení vytvorili jeho model používateľa. Následne sme vypočítali vzájomnú podobnosť medzi používateľmi a každému z používateľov sme odporučili množinu 15 položiek. Výsledky v grafe predstavujú absolútnu početnosť úspešne odporučených položiek. Tento experiment slúžil výhradne na stanovenie opísaného koeficientu, ktorého hodnotu sme stanovili na hodnotu 37.

Po určení najpodobnejších používateľov je potrebné rozhodnúť, ktoré z nimi videných položiek odporučíme a v akom poradí. Položky vyberáme z množiny všetkých, ktoré videli najpodobnejší používatelia skupiny a zároveň ich nevidel žiadny člen skupiny. V procese vývoja sme skúšali tieto položky zoradiť primárne podľa priemerného hodnotenia, ktoré im priradili najpodobnejší používatelia a v rámci zhodných výsledkov zoznam usporiadať na základe početnosti videnia, avšak tento postup sa ukázal ako nevhodný pretože dosahoval nižšie výsledky ako použitý spôsob opísaný v návrhu metódy. Ten spočíva v presne opačných prioritách, čiže zoradení položiek podľa početnosti videní a v rámci zhodne videných položiek podľa priemernej výšky hodnotení (v rámci množiny najpodobnejších používateľov).

Kolaboratívne odporúčanie jednotlivcom je v navrhutej metóde vopred predpočítavané. Určenie podobnosti voči takmer 70 000 MovieLens používateľom, ktorých modely máme spracované, totiž trvá netriviálny čas. Pri výpočte podobnosti dvoch používateľov trvajúcej rádovo milisekundy (4 x 2,3 GHz, 4GB RAM, v priemere 13,4 ms), trvá výpočet 70 000 podobností potrebných na úvodné spracovanie používateľa rádovo minúty (4 x 2,3 GHz, 4GB RAM, v priemere 15,6 min.). Skupiny navyše obsahujú viacero používateľov, takže výpočet podobností v reálnom čase neprichádza do úvahy. Spolu s podobnosťami predpočítavame aj výsledky kolaboratívneho prístupu. Tie by síce bolo možné počítať až v čase vzniku skupiny, no toto odporúčanie je od aktuálneho zloženia skupiny nezávislé, takže sme sa rozhodli tieto čiastkové výsledky zistiť v predstihu a tým ešte urýchliť následné skupinové odporúčanie.

7.3.2 Odporúčanie jednotlivcom na základe obsahu

Druhú časť odporúčania predstavuje odporúčanie na základe obsahu. Tu hľadáme vhodné položky z množiny všetkých existujúcich, ktoré však používateľ doposiaľ nehodnotil (a teda predpokladáme, že ich nevidel). Nad touto podmnožinou položiek postupne vykonáme vyhľadávaci dopyt pre každý vektor v modeli používateľa, ktorý opisuje určitú vlastnosť položiek, teda pre vektor žánrov, hercov, režisérov a vektor kľúčových slov. Nevyužívame pri ňom však piaty vektor obsahujúci samotné hodnotené položky, pretože z typu informácií, ktoré vektor opisuje, nevieme vyvodit' aké vlastnosti máme pri obsahovom odporúčaní hľadať.

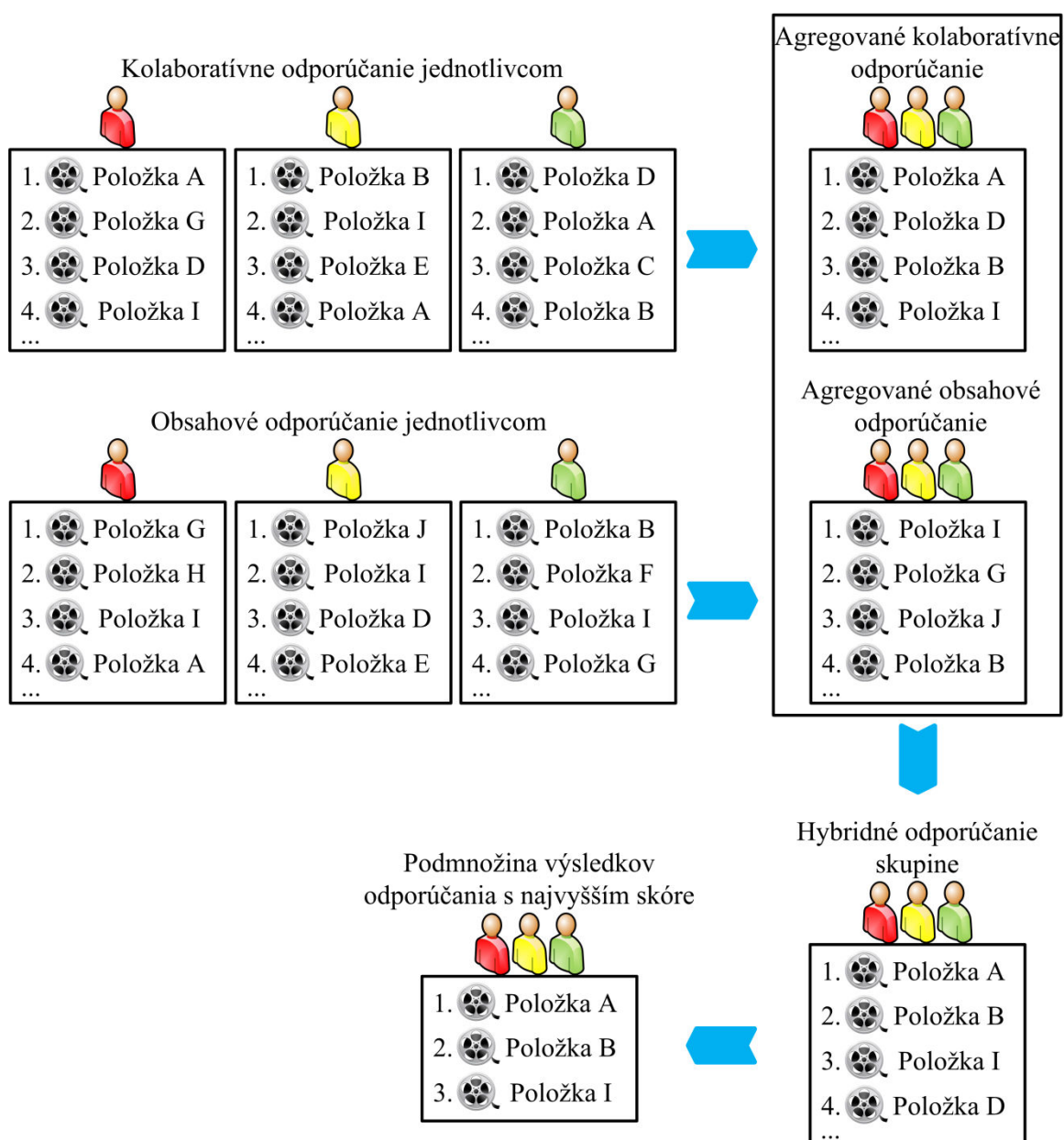
V rámci dopytov hľadáme položky obsahujúce niektorý z používateľom N najpreferovanejších elementov (teda napríklad N najobľúbenejších hercov, režisérov a pod.). Koeficient N bol experimentálne stanovený na hodnotu 10, kedy odporúčanie na vzorke 6 000 MovieLens používateľov dosahovalo najvyššiu presnosť. Položkám prideliujeme body, za každý z týchto top elementov, ktoré obsahujú vo svojich metadátach. Ak pre položku nájdeme vyhovujúci element vo viacerých dopytoch alebo nájdeme viacero elementov v jednom dopyte, body sčítavame. Za zhodného režiséra pripočítavame 25 bodov za herca 35, za žáner 12 a 6 bodov za kľúčové slovo. Tieto bodové ohodnotenia vektorov vznikli ako výsledok optimalizačného procesu, v ktorom sme sledovali presnosť obsahového odporúčania na vzorke 6 000 MovieLens používateľov v závislosti od meniacich sa váh vektorov.

Výsledky obsahového odporúčania jednotlivcom, podobne ako výsledky kolaboratívneho odporúčania jednotlivcom, predpočítavame. Samotný čas výpočtu síce nie je problémom a bolo by ho možné realizovať aj pri následnom skupinovom odporúčaní, no jeho priebeh nie je závislý od zloženia skupiny, keďže používateľom vytvárame tento typ odporúčania jednotlivco.

Predpočítanie tohto kroku navyiac urýchli proces, ktorý je nutné realizovať pri odporúčaní skupine v reálnom čase.

7.3.3 Hybridné odporúčanie skupine

Samotný proces skupinového odporúčania začína v momente, kedy sa sformuje skupina používateľov, ktorí majú záujem spoločne sledovať určitý obsah. Takáto skupina explicitne oznámi systému, ktorý bude odporúčanie realizovať, ktorí používatelia tvoria skupinu. Každý z nich musí mať v danom systéme svoj profil, ktorý obsahuje jeho preferencie. Samotné odporúčanie pozostáva z agregácie kolaboratívnych odporúčaní jednotlivým členom skupiny, agregácie obsahových odporúčaní jednotlivým členom skupiny, vytvorení hybridného odporúčania a prezentácii podmnožiny odporúčaní skupine (Obrázok 16).



Obrázok 16. Schematické znázornenie priebehu navrhutej metódy hybridného odporúčania skupine

Agregácia kolaboratívnych odporúčaní jednotlivým členom skupiny prebieha podľa spôsobu opísaného v návrhu metódy. Pre každého používateľa vytvoríme usporiadaný zoznam N jeho individuálnych kolaboratívnych odporúčaní, pričom každú položku obodujeme. Najvyššie odporúčenej položke pridáme N bodov, druhej $N-1$, a tak ďalej až po 1 bod pre N -tú položku. Koeficient N sme experimentálne stanovili na hodnotu 15, pričom sme hľadali najnižšiu možnú hodnotu, pri ktorej nie je ovplyvnená presnosť výsledného odporúčania. Samotná kolaboratívna agregácia odporúčaní jednotlivcov spočíva v spojení položiek do spoločného zoznamu, pričom body položiek, ktoré boli odporúčené viacerým používateľom vzájomne násobíme.

Agregácia obsahovo generovaných odporúčaní jednotlivým členom skupiny prebieha odlišne. Položky odporúčené jednotlivcom nebudujeme podľa poradia v zoznamoch individuálnych obsahových odporúčaní. V úvode procesu vytvoríme množinu s položkami, ktorá vznikla ako výsledok agregácie kolaboratívnych odporúčaní. Každá z týchto položiek dostane na úvod váhu 1. Následne k váham položiek pripočítavame 1 za každý výskyt danej položky v zoznamoch jednotlivcov. Zaujímá nás teda len početnosť výskytu, nie poradie.

Vytvorenie samotného hybridného odporúčania realizované podľa opísaného návrhu spočíva pre každú z položiek agregovaného kolaboratívneho odporúčania z pre násobenia jej kolaboratívnej váhy s váhou, ktorá jej bola priradená pri procese agregácie obsahovo orientovaných odporúčaní. V poslednom kroku odporúčania sú hybridne agregované položky usporiadané podľa výsledne prideleného skóre.

Skupine používateľov spomedzi výsledkov odporúčania skutočne zobrazíme len podmnožinu top 5 hybridne odporúčených položiek s najvyšším dosiahnutým skóre. Dôvody prečo neodporúčame viac položiek sú uvedené v návrhu metódy v kapitole Prezentácia odporúčania skupine. Výhodou toho, že disponujeme množinou viacerých položiek ako zobrazujeme, je že skupine vieme odporúčať viackrát bez nutnosti opätovného výpočtu odporúčania. Tieto prípadné neskoršie odporúčania však nebudú brať do úvahy názor používateľov na aktuálne odporúčené položky. Domnievame sa však, že ohodnotenie niekoľko málo položiek výrazne neovplyvní modely používateľov.

8 Overenie metódy

Overenie navrhutej metódy sme rozdelili do niekoľkých krokov. Pre každý z týchto krokov sme stanovili hypotézu, ktorú sme vyhodnotili pomocou samostatných experimentov. Spoločne majú za cieľ overiť model používateľa, hybridnú metódu pre odporúčanie jednotlivcovi a následne v hlavnej sérii experimentov overiť hybridnú metódu pre odporúčanie skupinám používateľov.

8.1 Model používateľa

Prvá z hypotéz je zameraná na model používateľa. Jej znenie sme formulovali nasledovne:

„Navrhnutý model používateľa uvažujúci vektory prvkov, ktoré opisujú hodnotené položky a uvažujúci individuálne preferencie týchto vektorov pomôže zlepšiť presnosť kolaboratívneho odporúčania v porovnaní s modelom používateľa, ktorý spomenuté vlastnosti neuvažuje.“

V rámci experimentu sme sa rozhodli porovnať výsledky kolaboratívneho odporúčania využívajúceho rozličné modely používateľa ako aj spôsob určovania podobnosti používateľov. Konkrétne porovnáваме:

- Navrhovaný váhovaný model používateľa využívajúci na určenie podobnosti používateľov kosínusovú podobnosť
- Aktuálny referenčný model používateľa v doméne multimediálneho obsahu využívajúci kosínusovú podobnosť
- Aktuálny referenčný model používateľa v doméne multimediálneho obsahu využívajúci Pearsonovu koreláciu

Dataset

Túto hypotézu sme kvantitatívne overili na vzorke používateľov získaných z MovieLens 10M datasetu. Uprednostnili sme syntetický experiment, na predpripravenej množine dát, na ktorých sme odporúčanie simulovali. Dôvodom pre tento krok bolo, že sme mohli vykonať pozorovanie rádovo na tisícoch používateľov, čo by sa nám pri živom experimente nepodarilo. V experimente sme spracovali hodnotenia 6 000 používateľov, pričom v pre dáta v použítom datasete je garantované, že každý používateľ vykonal minimálne 20 hodnotení. Spolu sme do modelov používateľov spracovali 662 517 hodnotení. Pre testovaných 6 000 používateľov sme v každej z overovaných metód vypočítali 18 000 000 vzájomných podobností medzi dvojicami používateľov.

Metodológia

Simulácia prebiehala spôsobom, že sme pre každého používateľa náhodne rozdelili jeho hodnotenia na 80% trénovaciu a 20% testovaciu množinu. Pomocou trénovacej množiny sme vytvorili modely používateľa (pre overovanú a referenčné metódy), na základe ktorých sme vyhodnotili vzájomné podobnosti používateľov a následne tiež kolaboratívne odporúčania. Odporúčané položky sme porovnali s testovacou množinou výsledkov, ktorá v tomto prípade nahradzovala skutočný výber používateľa.

Zvolený typ experimentu, nás však pripravil o situáciu, kedy by boli používatelia pri výbere položiek ovplyvnení informáciou, že sme im určité položky odporúčili. V našej situácii len retrospektívne pozorujeme, nakoľko sa naše odporúčanie zhoduje s používateľovým nezávislým

výberom. Môžeme sa tak pripraviť o situácie, kedy by si používateľ odporúčenú položku zvolil avšak vo svojej situácii o nej nevedel, prípadne si na ňu pri výbere nespomenul.

Tento hendikep, však v experimente mali všetky porovnávané metódy a preto môžeme jeho výsledky akceptovať. Cieľom experimentu totiž nebolo vyhodnotiť absolútnu úspešnosť niektoej z metód, ale vzájomne porovnať metódy v rovnakých podmienkach. Tento cieľ sme úspešne naplnili.

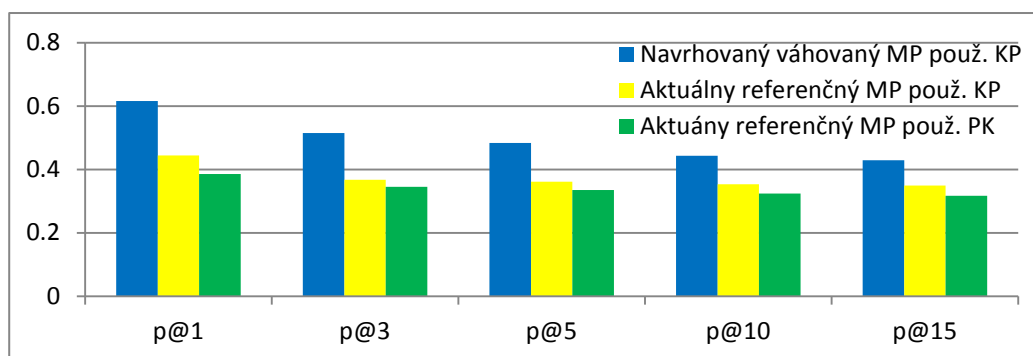
Na vyhodnotenie výsledkov porovnávaných metód sme zvolili štandardnú metriku *presnosť* (angl. precision), ktorá predstavuje podiel počtu používateľom vybraných položiek z množiny odporúčených a celkového počtu odporúčených položiek. Konkrétnejšie sme zisťovali presnosť z 1 odporúčenej položky ($p@1$), z 3 ($p@3$), z 5 ($p@5$), z 10 ($p@10$), z 15 ($p@15$).

Vyhodnotenie

Z výsledkov experimentu vyplýva, že kolaboratívne odporúčanie založené na navrhovanom modeli používateľa využívajúcom váhovanú kosínusovú podobnosť dosahuje výrazne lepšiu presnosť v porovnaní s totožnými odporúčaniami založenými na referenčnom modeli používateľa využívajúcom kosínusovú podobnosť respektíve Pearsonovu koreláciu (Tabuľka 5, Obrázok 17). Pri odporúčaní jednej položky (vyhodnotenie $p@1$) odporúčanie na základe nášho modelu dosahuje presnosť 61,61%, čo znamená zlepšenie 17,20% voči referenčnému modelu využívajúcomu kosínusovú podobnosť a 23,04% voči referenčnému modelu využívajúcomu Pearsonovu koreláciu. Pri odporúčaní skupiny 3 položiek ($p@3$) dosahuje odporúčanie na základe nášho modelu presnosť 51,53% a zlepšenie 14,79% a 17,05%, pre 15 položiek ($p@15$) dosahuje presnosť 42,94% a zlepšenie 7,98% a 11,26%.

Tabuľka 5. Výsledky experimentu porovnávajúceho presnosť kolaboratívneho odporúčania jednotlivcom na základe rozličných modelov používateľa (MP), KP – kosínusová podobnosť, PK – Pearsonova korelácia

MP využitý v odporúčaní	p@1	p@3	p@5	p@10	p@15
Navrhovaný váhovaný MP použ. KP	0,6161	0,5153	0,4838	0,4435	0,4294
Aktuálny referenčný MP použ. KP	0,4441	0,3674	0,3610	0,3530	0,3496
Aktuálny referenčný MP použ. PK	0,3857	0,3448	0,3351	0,3245	0,3168



Obrázok 17. Výsledky experimentu porovnávajúceho presnosť kolaboratívneho odporúčania jednotlivcom na základe rozličných modelov používateľa (MP), KP – kosínusová podobnosť, PK – Pearsonova korelácia

Na vyhodnotenie štatistickej signifikancie výsledkov odporúčania sme vykonali párový T-test. Dosiahnuté výsledky sú štandardnými kritériami považované ako vysoko štatisticky signifikantné

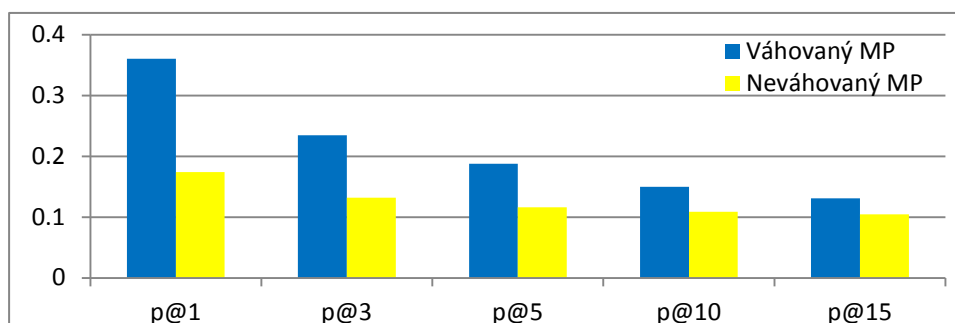
($p = 0.0047$, $\alpha = 0.05$ and $t = 7.6401$ pre referenčnú kosínusovú podobnosť a $p = 0.0058$, $\alpha = 0.05$, $t = 7.0895$ pre Pearsonovu koreláciu).

Na základe výsledkov môžeme konštatovať, že odporúčanie využívajúce nami navrhnutý model používateľa dosahuje najväčšie zlepšenie pri odporúčaní menšieho počtu položiek, čo predstavuje v oblasti odporúčania pre inteligentnú televíziu ideálnu situáciu. V tejto oblasti totiž trvá pozretie jednej položky pomerne dlhú dobu a používateľ si preto zvyčajne v rámci sedenia pozrie napríklad iba jeden film. Je teda zbytočné mu odporučiť príliš veľa položiek.

Okrem kolaboratívneho odporúčania sme model používateľa overili aj pre účely obsahového odporúčania. V tomto prípade sme však porovnávali len presnosť navrhnutého modelu používateľa bez a so zakomponovaním váhovania jednotlivých vektorov (Tabuľka 6 a Obrázok 18).

Tabuľka 6. Výsledky experimentu porovnávajúceho presnosť odporúčania na základe obsahu jednotlivcom využívajúceho váhovaný respektíve neváhovaný model používateľa (MP)

MP využitý v odporúčaní	p@1	p@3	p@5	p@10	p@15
Váhovaný MP	0,3608	0,2345	0,1879	0,1501	0,1307
Neváhovaný MP	0,1740	0,1322	0,1160	0,1088	0,1044



Obrázok 18. Výsledky experimentu porovnávajúceho presnosť odporúčania na základe obsahu jednotlivcom využívajúceho váhovaný respektíve neváhovaný model používateľa (MP)

Obsahové odporúčanie založené na váhovanom modeli používateľa dosiahlo vo všetkých meraniach vyššiu presnosť v porovnaní s odporúčaním založeným na neváhovanej variante navrhnutého modelu. Pre p@1 dokonca dosiahlo viac ako dvojnásobnú presnosť oproti neváhovanej variante (0,3608 oproti 0,1740).

Zistili sme taktiež, že v našich experimentoch dosiahlo kolaboratívne odporúčanie vždy vyššiu presnosť v porovnaní s výsledkami obsahového prístupu (Tabuľka 5 a Tabuľka 6). Z tohto dôvodu sme v navrhnutej hybridnej odporúčacej metóde primárne vychádzali z kolaboratívneho prístupu, ktorý sme len dopĺňali obsahovým.

8.2 Hybridná odporúčania metóda

V ďalšej časti overenia sme sa zamerali na zistenie miery použiteľnosti navrhnutej metódy pre účely personalizovaného odporúčania jednotlivcovi, keďže v ďalšej fáze sú tieto odporúčania využité pre odporúčanie skupine. Podobne ako v predchádzajúcej časti sme zvolili metriku presnosť, pretože sme chceli primárne zistiť akú časť z položiek odporučených metódou si používatelia vyberú. Samotný experiment slúžil na zistenie vlastností navrhnutej metódy bez toho aby bola ovplyvnená inými faktormi ako napríklad spôsobom agregácie, zavedením sociálneho koeficientu a podobne. Tie sú totiž súčasťou skupinového odporúčania overovaného v nasledujúcej podkapitole.

Dataset

Navrhnutú hybridnú metódu sme overili jednotlivo na používateľoch získaných z datasetu Movielens, pričom sme využili rovnakú časť datasetu a nastavenia ako v predchádzajúcom experimente, kde je dátová sada podrobnejšie opísaná.

Metodológia

V prvej časti experimentu sme pomocou metriky presnosť porovnali navrhnutú hybridnú metódu s najlepšou kolaboratívnou a najlepšou obsahovo založenou metódou, z experimentov opísaných v predchádzajúcej podkapitole venujúcej sa overeniu modelu používateľa, odkiaľ sme tiež prevzali spôsob tvorby modelov používateľa a výber testovacej a trénovacej množiny dát. Sledovali sme presnosť pri odporúčení 1 ($p@1$), 3 ($p@3$), 5 ($p@5$), 10 ($p@10$) a 15 ($p@15$) položiek.

Okrem celkovej presnosti, ktorú metóda dosahuje, nás v druhej časti experimentu zaujímalo akú presnosť odporúčania metóda dosahuje pre jednotlivé pozície výsledkov. Teda či si používatelia častejšie zvolia položku z vyšších miest v odporúčanom zozname alebo či je rozloženie rovnaké. Na vyhodnotenie tejto charakteristiky sme počítali presnosť odporúčania ($p@1$) nezávisle pre každú z prvých 15 pozícií zoznamu s výsledkami.

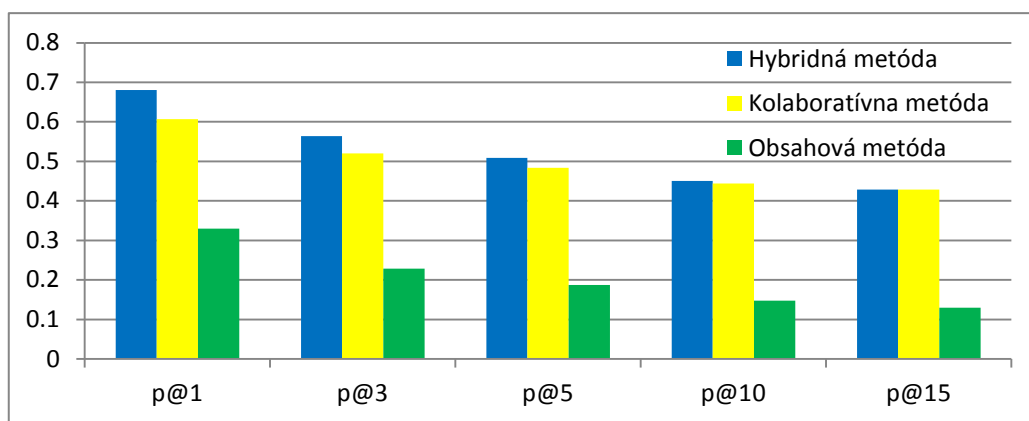
Vyhodnotenie

Z výsledkov na Obrázku 19 a v Tabuľke 7 vidíme, že vhodným hybridným spojením jednotlivých odporúčačích prístupov sme schopní zvýšiť presnosť odporúčania jednotlivcom práve v porovnaní s odporúčaniami založenými na kolaboratívnom respektíve obsahovom odporúčaní, ktoré naša hybridná metóda využíva. Najväčší rozdiel je pritom badateľný pri odporúčaní menšieho počtu položiek ($p@1$, $p@3$), pretože na týchto pozíciách sú odporúčané najvhodnejšie položky.

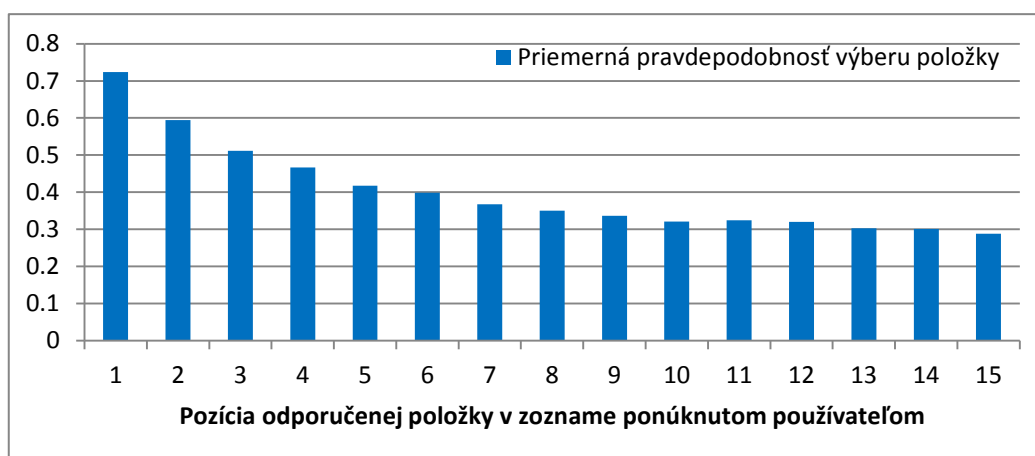
Daný stav je najmä výsledkom použitia navrhnutej hybridnej metódy. Dôkazom je fakt, že hybridná metóda dosahuje na prvých pozíciách výrazne vyššiu presnosť ako na nižších (Obrázok 19), čím sa odlišuje od referenčných metód.

Tabuľka 7. Výsledky experimentu porovnávajúceho presnosť hybridného odporúčania jednému používateľovi voči presnosti odporúčaní, z ktorých je zložené. Odporúčania boli realizované na rovnakej množine používateľov

Metóda/Presnosť	$p@1$	$p@3$	$p@5$	$p@10$	$p@15$
Hybridná	0,6803	0,5640	0,5092	0,4505	0,4287
Kolaboratívna	0,6072	0,5202	0,4835	0,4438	0,4287
Obsahová	0,3303	0,2284	0,1876	0,1478	0,1303



Obrázok 19. Výsledky experimentu porovnávajúceho presnosť hybridného odporúčania jednotlivcom voči presnosti odporúčaní, z ktorých je zložené. Odporúčania boli realizované na rovnakej množine používateľov



Obrázok 20. Priemerná pravdepodobnosť výberu odporúčenej položky používateľom v závislosti od jej pozície v ponúknutom zozname. (Presnosť odporúčania (p@1) pre jednotlivé pozície).

Výsledky uvedené na Obrázku 20 ukazujú, že presnosť odporúčania našou metódou rastie priamoúmerne s rastúcou pozíciou odporúčenej položky vo výslednom zozname (v zmysle, čím vyššia pozícia, tým skôr je položka odporúčená). To predstavuje ideálny stav, pretože vhodnejšie položky sú odporúčené skôr.

Využitie odporúčania založeného na obsahu sa totiž ukázalo ako vhodný spôsob preusporiadania kolaboratívnych výsledkov, pretože zvýšilo presnosť odporúčania najvyššie odporúčených položiek. To predstavuje jeden z najdôležitejších cieľov, ktoré je vhodné dosiahnuť pri odporúčaní dlhotrvajúcich položiek, u ktorých si používateľ dokáže pozrieť bežne len jednu položku za sedenie, ako napríklad v doméne multimediálneho obsahu.

8.3 Hybridná metóda skupinového odporúčania

Ďalším krokom v procese overenia navrhutej metódy bolo overenie jej vlastností pre skupiny používateľov. Tento krok sme overili dvomi rôznymi experimentmi. Najskôr sme vykonali syntetický experiment s využitím Movielens používateľov, ktorých sme umelo rozdelili do skupín. Následne sme vykonali riadený experiment s používateľmi v reálnom čase, kde sme sa porovnali s metódami založenými na princípoch referenčných metód v doméne. Výsledky tohto experimentu sme porovnali tiež s výsledkami referenčných metód publikovanými inými autormi.

V prípade tohto kroku overenia navrhutej metódy sme formulovali nasledujúcu hypotézu:

„Navrhnutá hybridná metóda skupinového odporúčania zvýši presnosť odporúčania v porovnaní s metódami používanými v referenčných odporúčacích systémoch – PolyLens, Yu’s TV a gRecs pomocou vhodnejšieho výberu položiek na odporúčenie a ich následného usporiadania.”

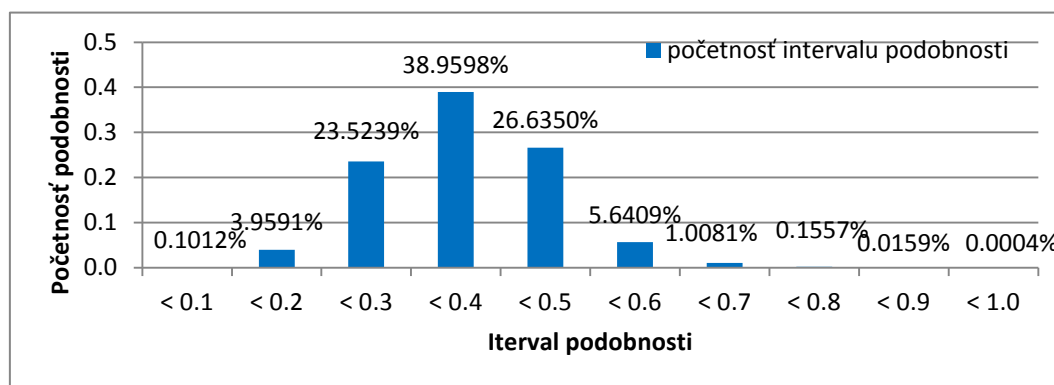
8.3.1 Syntetický experiment

Ako prvý z experimentov na overenie navrhnutého skupinového odporúčania sme vykonali kvantitatívny syntetický experiment. Tento experiment mal za úlohu primárne overiť metódu na veľkej vzorke dát. Takéto overenie má totiž dobrú výpovednú hodnotu, je možné ho zopakovať a nepodlieha rizikám vplyvu neželaných faktorov aké hrozí pri experimente s používateľmi v reálnom čase.

Dataset

Experiment bol realizovaný na rovnakom datasete, na ktorom sme overovali model používateľa (Movielens 10M dataset). Pracovali sme so vzorkou 20 000 používateľov. Pri vytváraní skupín sme pracovali s podobnosťou používateľov, ktorú sme určili ako váhovanú kosínusovú podobnosť ich modelov používateľa (opísané v kapitole 6.3.1 Kolaboratívne odporúčanie jednotlivcom).

V jeho úvode sme zisťovali mieru podobnosti medzi používateľmi v spracovanej podmnožine Movielens používateľov. Amer-Yahia vo svojej práci tvrdí, že skupiny vytvárajú zväčša používatelia, ktorých podobnosť je nad určitou prahovou hodnotou (Amer-Yahia, 2009). Na základe tohto tvrdenia sme identifikovali rozdelenie podobností (Obrázok 21). Identifikovali sme medián (0,359398), aritmetický priemer (0,363073), 1. a 3. kvartil (0,293812; 0,428561). V ďalšom priebehu experimentu sme prahovú hodnotu podľa vzoru Amer-Yahia určili ako medián podobností v použitom datasete.



Obrázok 21. Rozdelenie podobností medzi dvojicami používateľov do intervalov

Metodológia

Z používateľov sme vytvorili dvoj, troj a štvorčlenné skupiny, pre ktoré sme skúmali presnosť odporúčania. Zvolené veľkosti skupín vychádzajú z práce O'Connor a kol., ktorí v rok trvajúcom experimente zistili, že používatelia majú záujem práve o tieto veľkosti skupín. Väčšie skupiny používatelia nevytvárali (O'Connor a kol., 2002). Záujem používateľov o skupiny jednotlivých veľkostí zachytáva Tabuľka 3 v kapitole 4.1 PolyLens.

Používateľom sme podobne ako pri predchádzajúcich experimentoch stratifikáciou rozdelili ich videné položky (získané z datasetu) na trénovaciu a testovaciu množinu v pomere 80:20. Z trénovacích dát sme vytvorili modely používateľov, testovacie položky sme pokladali za položky, ktoré by si používatelia po ich odporúčení boli ochotní prezrieť.

Na to aby mohli byť používatelia v experimente spoločne zaradení do skupiny musela byť splnená podmienka, že je ich vzájomná podobnosť je nad stanovenou prahovou hodnotou. Pri vyhodnocovaní experimentu sme sledovali presnosť odporúčania jednotlivým skupinám, ktorým sme odporúčali podľa navrhutej hybridnej metódy. Na to aby sme prehlásili, že si skupina odporúčenú položku pozrie, musela byť táto položka v testovacej množine každého jej člena.

Aby sme zabezpečili nenulovú množinu položiek, ktoré by si vygenerovaná skupina bola ochotná po odporúčení pozrieť, sme pri vytváraní skupín stanovili okrem podmienky s nadprahovou podobnosťou jej členov tiež podmienku, že v skupine musí existovať minimálne N položiek, ktoré majú všetci členovia skupiny zaradené v testovacej množine. V experimente sme pre všetky 3 veľkosti skupín merali presnosť pre skupiny s N rovným 5 položiek a skupiny s N rovným 10 položiek. Pre každé nastavenie sme vygenerovali 1 000 skupín.

Vyhodnotenie

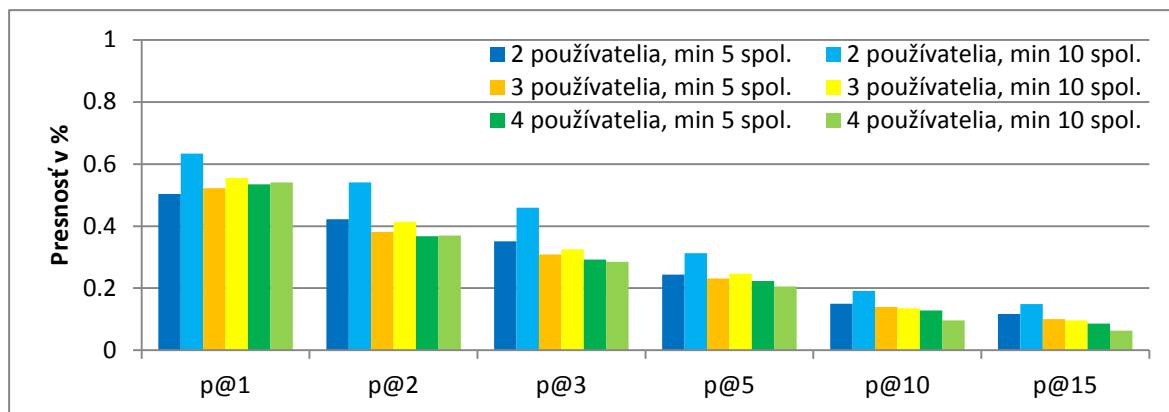
Výsledky zobrazuje Tabuľka 8 a Obrázok 22. Ako môžeme vidieť, presnosť odporúčania sa pri rastúcom počte členov skupiny taktiež mierne zvyšuje. Dôvodom je fakt, že pre vznik skupiny musela existovať množina minimálne 5 respektíve 10 spoločných položiek v testovacích množinách členov skupiny. V prípade väčšej skupiny sa teda s väčšou pravdepodobnosťou jedná o položky, ktoré aspoň jedno z nami využívaných odporúčaní identifikuje ako vysoko vhodné a vyberie ich na odporúčenie.

Presnosť odporúčania väčšieho množstva položiek klesá nezávisle od veľkosti skupiny. Je totiž náročnejšie nájsť napríklad 5 vhodných položiek na odporúčenie, ako 1 alebo 2. To je síce logické, no do istej miery tento jav zvyrazňuje tiež fakt, že sme využili syntetický experiment. V jeho prípade totiž spolu používatelia nemohli pri výbere z odporúčených položiek komunikovať a ovplyvňovať sa, ako v reálnej skupine. Pokiaľ sa v prípade tohto experimentu odporúčená položka nenachádzala v testovacej množine každého člena skupiny, odporúčanie sme vyhodnotili ako neúspešné. V skutočnej skupine by však mohla časť členov skupiny presvedčiť zvyšok, že odporúčená položka je vhodná voľba, ktorú daný zvyšok skupiny napríklad len nepoznal.

Odporúčanie bez tejto nevýhody zabezpečil nasledujúci experiment s reálnymi skupinami.

Tabuľka 8. Výsledky experimentu zobrazujúce presnosť hybridného odporúčania syntetickým skupinám pre rôzne veľkosti skupín a minimálny počet spoločne prijateľných položiek

Nastavenie/Presnosť	p@1	p@2	p@3	p@5	p@10
2 používatelia, min 5 spol.	0.5033	0.4225	0.3511	0.244	0.1504
2 používatelia, min 10 spol.	0.633	0.5405	0.4597	0.313	0.1917
3 používatelia, min 5 spol.	0.5225	0.3806	0.3091	0.2315	0.1398
3 používatelia, min 10 spol.	0.5556	0.4136	0.3251	0.2469	0.1356
4 používatelia, min 5 spol.	0.5347	0.3679	0.2926	0.2228	0.1287
4 používatelia, min 10 spol.	0.5412	0.3691	0.2849	0.2056	0.0962



Obrázok 22. Presnosť hybridného odporúčania syntetickým skupinám. Odporúčanie je testované pre rôzne veľkosti skupín a rôzny minimálny počet spoločne prijateľných položiek

8.3.2 Reálny experiment

Druhý z experimentov na overenie navrhnutého skupinového odporúčania sme vykonali s používateľmi v reálnom čase. Experiment sa skladal z 2 častí. V prvej sme prostredníctvom používateľov porovnávali našu metódu s referenčnými metódami založenými na princípoch využívaných v referenčných systémoch. V druhej časti sme získané výsledky porovnávali s publikovanými výsledkami iných autorov.

Dataset

V prípade tohto experimentu sme pracovali s množinou 31 používateľov. Každý z nich pred experimentom ohodnotil určité množstvo položiek. Toto množstvo mohlo byť ľubovoľné, minimálne sa však muselo jednať o 30 položiek. Položky boli hodnotené na 5 bodovej hodnotiacej škále (tak ako tomu bolo u používateľov z Movielens datasetu).

V experimente používatelia vytvárali skupiny s veľkosťou 2, 3 alebo 4. Vytvorili 63 2-členných skupín (32 z nich bolo odporúčané hybridnou metódou a metódou využívajúcou stratégiu najmenšej nespokojnosti, 31 bolo odporúčané hybridnou metódou a metódou využívajúcou agregáciu), 28 3-členných skupín (14 a 14) a 12 4-členných skupín (5 a 7).

Metodológia – experiment v reálnom čase

Druhý experiment mal za úlohu overiť vlastnosti navrhutej metódy v porovnaní s existujúcimi systémami. Rozhodli sme sa v ňom porovnať s princípmi využívanými v systémoch PolyLens, Yu's TV a gRecs, ktoré patria medzi najznámejšie systémy v doméne multimediálneho obsahu. Pre každý zo systémov je známy princíp odporúčania (kapitola 4 Existujúce systémy), avšak nie jeho konkrétna implementácia. Systémy však majú spoločný typ skupín, predmet agregácie a odporúčací prístup. V tomto ohľade sa zhodujú aj s nami navrhnutou hybridnou metódou.

Metódy sa však vzájomne líšia v spôsobe agregácie, kde systém PolyLens využíva stratégiu najmenšej nespokojnosti, Yu's TV využíva priemernú stratégiu, a gRecs v rôznych variáciách stratégiu najmenšej nespokojnosti, spravodlivú stratégiu a stratégiu najväčšej spokojnosti. Z vymenovaných spôsobov sme do experimentu vybrali stratégiu najmenšej nespokojnosti a spravodlivú stratégiu, ktoré spoločne pokrývajú všetky tri referenčné systémy. Tieto stratégie sme sa rozhodli porovnať voči našej hybridnej agregácii.

Keďže sme nemali k dispozícii konkrétne implementácie systémov aplikovali sme v experimente ich princíp na naše kolaboratívne odporúčanie, kde sme používateľov v skupine agregovali hybridne (na základe navrhnutého princípu vychádzajúceho z preusporiadania

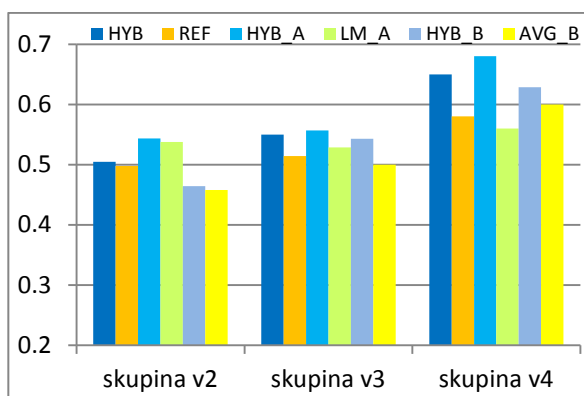
agregovaných kolaboratívnych výsledkov pomocou agregovaných obsahových výsledkov), pomocou stratégie najmenej nespokojnosti respektíve priemernej stratégie.

Skupinám sme vždy ponúkli dvojicu odporúčaní - nami navrhnuté a jedno z referenčných. Referenčné bolo zvolené náhodne a skupina nemala informáciu ktoré z odporúčaní je nami navrhované a ktoré referenčné. Poradie zobrazenia výsledkov metód sa menilo. Experimentovali sme so skupinami veľkosti 2, 3 a 4 používatelia.

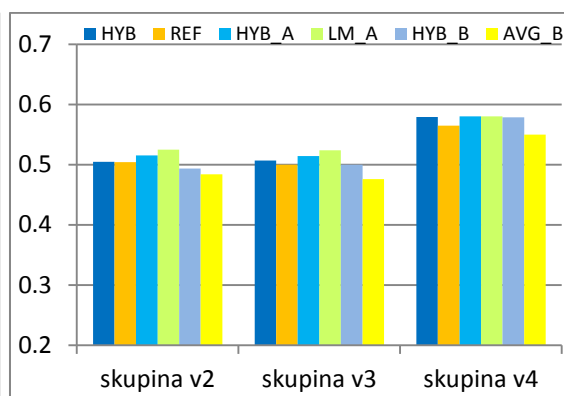
V rámci každého odporúčania sme skupine odporučili 5 položiek a požiadali jej členov aby pre každú odporúčenú položku vyjadrili ako veľmi by si ju v rámci danej skupiny spoločne pozreli. Hodnotenie skupina vyjadrovala najprv spoločne a následne každý jej člen individuálne. Hodnotenie prebiehalo priradením 0 až 5 bodov každej z položiek. 0 predstavovala situáciu, kedy by si skupina položku nevybrala, nenulový počet bodov vyjadroval vhodnosť odporúčania podľa názoru skupiny. Okrem jednotlivého hodnotenia položiek skupina (a následne členovia individuálne) hodnotila odporúčania tiež ako celky. Rovnakým bodovým systémom ako položky teda ohodnotila naše ponúknuté referenčné odporúčanie.

Vyhodnotenie – experiment v reálnom čase

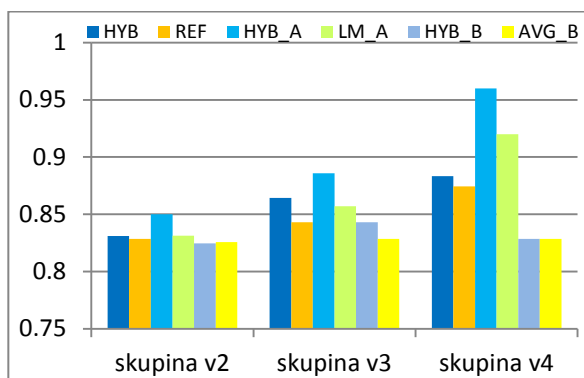
Výsledky experimentu sú znázornené na Obrázku 23. Skupiny vo výsledkoch označujeme ako A alebo B podľa toho ktorou referenčnou metódou sme im odporúčali. Skupiny so stratégiou najmenej nespokojnosti sú označené ako A (HYB_A a LM_A), skupiny s priemernou stratégiou sú označené ako B (HYB_B a AVG_B). HYB predstavuje priemer výsledkov HYB_A a HYB_B, REF priemer výsledkov LM_A a AVG_B.



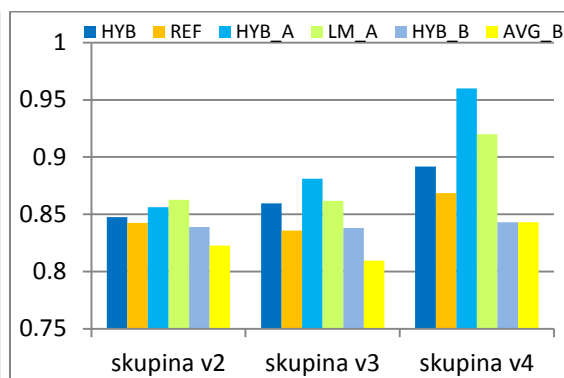
a) Hodnotenie odporúčaní ako celku z pohľadu skupín spoločne



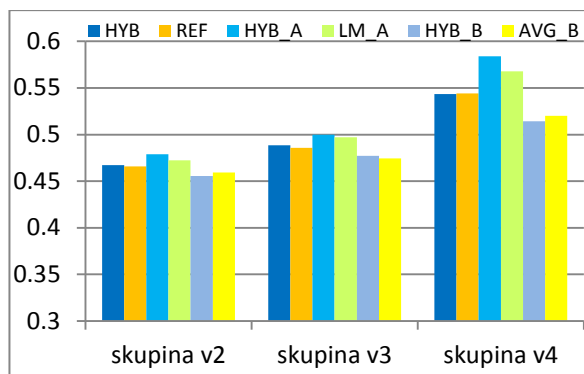
b) Hodnotenie odporúčaní ako celku z pohľadu členov skupín individuálne



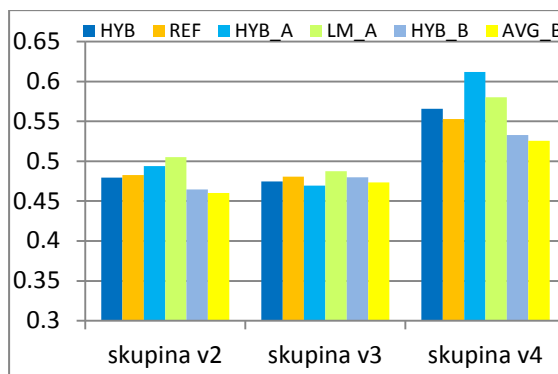
c) Presnosť odporúčania z pohľadu skupín spoločne



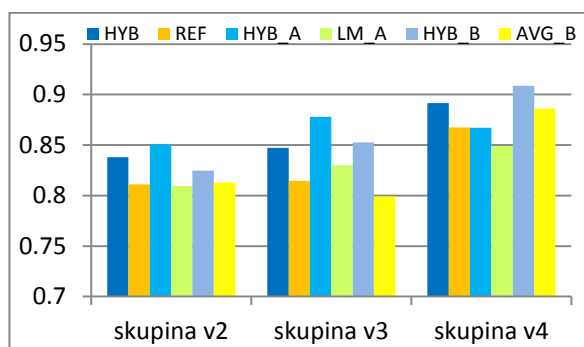
d) Presnosť odporúčania z pohľadu členov skupín individuálne



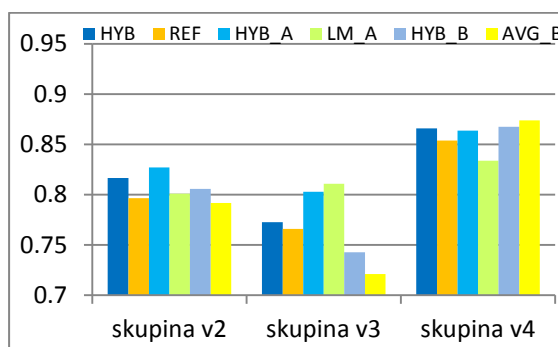
e) Priemerné hodnotenie odporúčenej položky z pohľadu skupín spoločne



f) Priemerné hodnotenie odporúčenej položky z pohľadu členov skupín individuálne



g) Výsledky metriky NDCG z pohľadu skupín spoločne



h) Výsledky metriky NDCG z pohľadu členov skupín individuálne

Obrázok 23. Výsledky experimentu porovnávajúceho navrhnutú metódu odporúčania s princípmi referenčných metód. HYB_A a HYB_B označujú navrhnutú metódu (ich priemer = HYB), LM_A označuje referenčnú metódu využívajúcu stratégiu najmenšej nespokojnosti, AVG_B označuje referenčnú metódu využívajúcu priemernú stratégiu (ich priemer = REF).

Pri experimente sme skupine zobrazovali vždy dve odporúčania, pričom ich poradie sme náhodne menili. Neuvedomili sme si však, že v prípadoch kedy sme ako referenčnú metódu zvolili metódu využívajúcu stratégiu najmenšej nespokojnosti, našu metódu sme zobrazili ako prvú, pred referenčnou, naopak v prípadoch kedy sme ako referenčnú metódu zvolili metódu s priemernou agregáčnou stratégiou, sme našu metódu zobrazovali ako druhú po referenčnej. Z výsledkov je jasne viditeľné, že skupina po ohodnotení prvej metódy, tomuto hodnoteniu prispôbila aj hodnotenie druhej metódy. Preto môžeme vidieť, že naša metóda dosiahla v prípade, že bola zobrazená ako prvá určité hodnotenie (HYB_A), avšak ak bola zobrazená až po referenčnej metóde dosiahla nižšie hodnotenie (HYB_B), podobné referenčnej aj napriek tomu, že sa jednalo o rovnakú metódu. Používatelia sa totiž zjavne snažili byť spravodliví a žiadnu z metód v hodnotení príliš nezvýhodňovali. Napriek tomu naša metóda vo všetkých prípadoch dosiahla vyššie alebo aspoň porovnateľné výsledky.

Pre lepšiu výpovednú hodnotu celého experimentu uvádzame aj priemer výsledkov našej metódy (HYB) a tiež priemer výsledkov referenčných metód (REF) aj napriek tomu, že sa jedná o dve nezávislé metódy.

Je potrebné si uvedomiť, že všetky testované prístupy v experimente majú totožný základ. Používajú rovnaký model používateľa a kolaboratívne odporúčanie. Líšia sa len spôsobom agregácie a v prípade navrhutej metódy ešte následným preusporiadaním odporučených výsledkov. Rozdiely vo výsledkoch teda logicky nie sú veľké.

Z výsledkov a) a b) na Obrázku 23 vidíme, že v hodnotení odporúčaní ako celkov dopadla navrhnutá metóda mierne lepšie, pričom rozdiel v spokojnosti je výraznejší pre väčšie skupiny, kde sa viac prejavovala rozdielnosť agregácií. Pri vyjadrení spoločného názoru skupiny je rozdiel medzi overovanými prístupmi viditeľnejší ako pri individuálnom vyjadrení názoru členov skupín.

Výsledky c) a d) na Obrázku 23 ukazujú, že pri určovaní presnosti odporúčania navrhovaná metóda taktiež dosahuje mierne vyššiu presnosť ako referenčné metódy. Prekvapením však je, že presnosť odporúčania rastie priamoúmerne zväčšujúcej sa skupine. Pri predchádzajúcich syntetických experimentoch, kde sme skupiny vytvárali umelo bol totiž trend presne opačný, pretože tu bolo pre zvolenie odporúčenej položky potrebné aby sa nachádzala v testovacej množine hodnotení každého z členov skupiny. Ako však vyplýva zo živého experimentu, pokiaľ majú členovia skupiny možnosť vzájomne interagovať, je pravdepodobnejšie, že odporúčenú položku niektorý z nich v skupine presadí a skupina je následne ochotná si ju pozrieť. Vzájomná interakcia členov skupiny skutočne presvedčila, pretože položky následne označovali ako vhodné aj individuálne. Tento jav nazývaný v angličtine conformity, vo svojej práci ako prví opísali Ricci a kol. (Ricci a kol., 2011).

Z výsledkov e) a f) na Obrázku 23 konštatujeme, že metóda skupinového odporúčania ani spôsob agregácie nemajú preukázateľný vplyv na výšku hodnotenia položky skupinou ani jej členmi. (Minimálne hodnotenie položky na osi y v grafoch e) a f) je rovné 0, maximálne možné hodnotenie je rovné 1. Jedná sa o normalizáciu 0-5 bodového systému hodnotenia z experimentu).

Výsledky g) a h) na Obrázku 23 ukazujú, že spôsob agregácie má vplyv na poradie usporiadania položiek v odporučenom zozname. Túto vlastnosť sme v experimente merali pomocou metriky NDCG (angl. Normalized Discounted Cumulative Gain), ktorá meria výkon systému podobne ako metrika Presnosť, avšak na rozdiel od nej berie do úvahy aj pozíciu jednotlivých výsledkov. Metrika vyššie hodnotí správne odporúčanie na vyššej pozícii. Z výsledkov experimentu môžeme vidieť, že navrhovaná metóda dosahuje vyššie hodnoty ako referenčné metódy.

Z výsledkov experimentu je možné vyvodiť záver, že navrhnutá metóda skupinového odporúčania dosahuje zlepšenie výsledkov v porovnaní s princípmi metód využívanými v referenčných systémoch. Zlepšenie sa prejavuje zvýšením presnosti odporúčania, zlepšením spokojnosti s odporúčaním ako s celkom a najmä v zlepšení usporiadania odporučených položiek (vyššie hodnotené položky sú odporúčané na vyšších pozíciách). Zlepšenie vo všetkých prípadoch nasalo z pohľadu skupiny ako celku a zároveň tiež z pohľadu členov skupín samostatne.

Metodológia – porovnanie s publikovanými výsledkami

V ďalšom experimente sme výsledky našej skupinovej metódy z druhého experimentu (HYB) porovnávali s publikovanými výsledkami iných autorov. Primárne sme hľadali výsledky zachytávajúce skupinové odporúčanie 5 položiek skupinám o veľkosti 2, 3 a 4 používateľov.

Referenčné metódy $VU + th + fix$, $VU + wt + fix$, $VU + th + var$ a $VU + wt + var$ pochádzajú z práce Kagita a kol. (Kagita a kol., 2013). Sú založené na princípe virtuálneho používateľa reprezentujúceho záujmy celej skupiny. Tomuto používateľovi jednotlivé metódy odporúčajú akoby sa jednalo o jednotlivca. Pri vytváraní virtuálneho používateľa z modelov používateľa jednotlivých členov skupiny sa tu využíva technika rozdielného váhovania členov skupiny ($VU + wt$) alebo maximálneho prahu – teda váhovania s obmedzením najvyššej možnej váhy ($VU + th$). Metódy navyše rozlišujú spôsob agregácie modelov používateľa (označené *fix* alebo *var*).

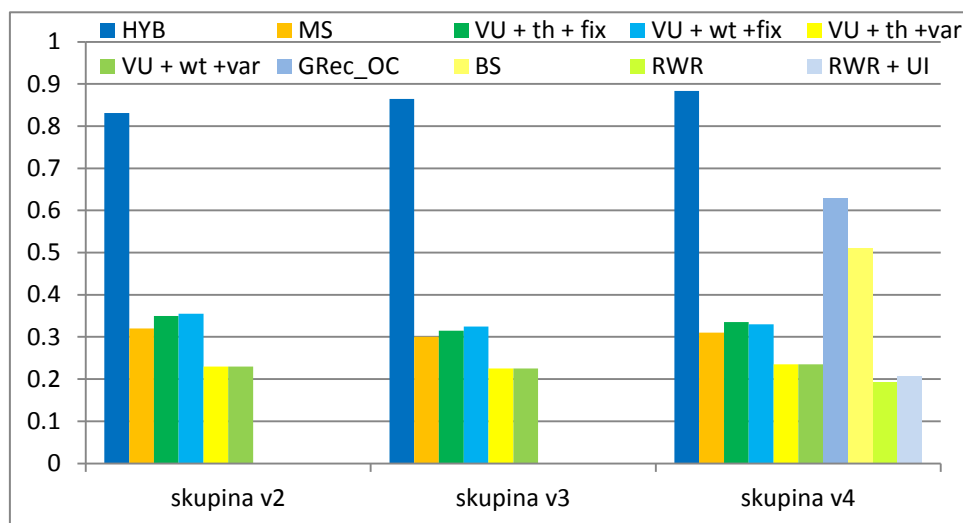
Referenčná metóda tejto práce je označená ako *MS*. Jej princíp spočíva v agregovaní preferencií členov skupiny, nie ich modelov ako tomu je v predchádzajúcich prípadoch.

Metódy *GRec_OC* a *BS* pochádzajú z práce Kim a kol. (Kim a kol., 2010). Využívajú kolaboratívne odporúčanie na základe kľúčových slov, pričom pre skupiny sa využíva agregácia odporúčaní jednotlivcom. Metóda *BS* (benchmark method) je prototypom metódy *GRec_OC*, ktorá k opísanému princípu navyše pridáva váhovanie členov skupiny na základe ich aktivity.

Metódy *RWR* a *RWR + UI* pochádzajú z práce Wang a kol. (Wang a kol., 2011). Využívajú na odporúčanie skupinám používateľov sociálny graf, v ktorom pomocou algoritmu *RWR* (angl. random walk with restarts) hľadajú používateľov podobných skupine. Na základe nimi vidенých položiek potom kolaboratívne odporúčajú obsah skupine. Metóda *RWR + UI* využíva taktiež sociálny graf, ktorý je však obohatený o dodatočné informácie o položkách, vďaka ktorým dokáže v grafe vyhľadať vhodnejších používateľov.

Vyhodnotenie – porovnanie s publikovanými výsledkami

Presnosť odporúčania 5 položiek ($p@5$) zistenú v druhom experimente overujúcom skupinovú metódu sme porovnali s výsledkami metód publikovaných inými autormi. V prípade, že výsledky experimentov referenčných metód boli pre dané veľkosti skupín známe, sme sledovali výsledky pre skupiny o veľkosti 2, 3 a 4 používateľov (Obrázok 24). Referenčné metódy odporúčali prevažne filmové položky, s výnimkou *GRec_OC* a *BS*, ktoré skupinám používateľov odporúčali knižné položky. Z výsledkov je viditeľné, že naša metóda dosahuje výrazne vyššiu presnosť odporúčania pre všetky 3 sledované veľkosti skupín v porovnaní s referenčnými systémami. Domnievame sa, že dosiahnutá presnosť je zapríčinená okrem samotnej hybridnej metódy odporúčania taktiež využitím navrhnutého modelu používateľa, pre ktorý sa nám s predchádzajúcich experimentoch podarilo taktiež preukázať vplyv na zvýšenie presnosti odporúčania pri jeho využití.



Obrázok 24. Porovnanie presnosti $p@5$ navrhovanej metódy s výsledkami publikovaných referenčných metód

9 Zhodnotenie

S rastúcou dostupnosťou informácií, prichádza v súčasnej dobe problém informačného zahltenia. V doméne multimediálneho obsahu sa oproti minulosti výrazne zvýšilo množstvo obsahu ktorý si používatelia môžu kedykoľvek pozrieť. Veľké množstvo televíznych staníc a prítomnosť rozsiahlych archívov často vytvára neprehľadnú spleť možností, z ktorých je náročné si vybrať. Pôžívateľia obľubujúci multimediálny obsah v súčasnej dobe trávajú čoraz viac času sledovaním televízie, či prezeraním rôzneho multimediálneho obsahu na internete. Tento čas zaberá netriviálnu časť ich voľného času a je preto veľkou výzvou ponúknuť im na sledovanie obsah, ktorý čo najlepšie naplní ich očakávania a podľa ich preferencií ich zabaví či vzdelá.

V prípade sledovania televízie je bežné, že viacerí používatelia sledujú obsah spoločne. V ich prípade teda nie je potrebné len vybrať im z množstva dostupných položiek obsah, ktorý sa im bude páčiť jednotlivo ale je tiež dôležité nájsť kompromis, ktorý ich zaujme súčasne. Medzi známe riešenia týchto požiadaviek patrí personalizované odporúčanie, respektíve odporúčanie skupinám. V týchto oblastiach v súčasnosti prebieha aktívny výskum.

V práci sme navrhli hybridnú metódu odporúčania multimediálneho obsahu určenú primárne skupinám používatel'ov, ktorú sme však overili aj pre účely personalizovaného odporúčania jednotlivcom. Pred samotným odporúčaním sme však v práci navrhli inovatívny model používatel'a. Jeho princíp je založený na bežne využívanej vektorovej paradigme zaznamenávajúcej používatel'ove preferencie z pohľadu viacerých typov metadát charakterizujúcich položky v doméne. Tento typ modelu obohacujeme o individuálne váhy jednotlivých vektorov, vyjadrujúce informáciu ako dôležité sú pre používatel'a typy informácií obsiahnuté v daných vektoroch. Váhy sú založené na statickej časti vyjadrujúcej priemerný záujem používatel'ov. Model však obohacujeme taktiež o individuálne časti vyjadrujúce mieru odlišnosti používatel'a od iných používatel'ov.

Príkladom v doméne multimédií je to napríklad informácia, či sa používatel' rozhoduje viac podľa režiséra, hercov či žánrov. Vďaka tomu vieme pri vyhľadávaní podobných používatel'ov využívaných pri kolaboratívnom odporúčaní v prvom rade nájsť používatel'ov ktorí sa rozhodujú podobným spôsobom (napríklad podľa režisérov) a až následne používatel'ov, ktorí preferujú rovnaké elementy (napríklad rovnakého konkrétneho režiséra alebo herečku).

Model je však možné využiť pre účely odporúčania na základe obsahu, kde hľadáme skôr položky obsahujúce preferovanejší typ informácií na úkor položiek s metadátami menej preferovaného typu. Model sme experimentálne overili pre účely kolaboratívneho odporúčania, kde sa nám podarilo pri odporúčaní založenom na tomto modeli dosiahnuť vyššiu presnosť v porovnaní s totožným odporúčaním založenom na štandardne používanom modeli.

Samotná navrhnutá odporúčacia metóda inovatívnym spôsobom spája výhody kolaboratívneho a obsahového odporúčania, ktoré sú v doméne inteligentných televízií na odporúčanie využívané v prevažnej väčšine prípadov. Hlavným prínosom našej metódy je hybridný spôsob spojenia oboch prístupov, kedy výsledky kolaboratívneho odporúčania preusporiadavame na základe ich vhodnosti z pohľadu obsahového odporúčania. Týmto spôsobom dokážeme na jednej strane znásobiť silné stránky oboch prístupov (nájdienie položiek, ktoré vybrala komunita, nájdienie položiek s vhodnými vlastnosťami) a na strane druhej dokážeme odstrániť známe slabiny (problém studeného startu pri kolaboratívnom prístupe, nižšia presnosť, či monotónnosť odporúčania pri odporúčaní na základe obsahu).

Týmto spôsobom dostávame v porovnaní s bežne využívanými prístupmi vyššiu presnosť odporúčania a to najmä u najskôr odporúčaných položkách. Dôvodom je, že na najvyššie pozície v odporúčaní umiestňujeme položky, ktoré z pohľadu vhodnosti vyhovujú obom tradične

používaným prístupom, a u ktorých je teda väčšia pravdepodobnosť, že zaujmú skupinu alebo používateľa, ktorým sú určené.

Tieto tvrdenia sme overili viacerými experimentmi. Či už synteticky na umelo vytvorených skupinách používateľov z Movielens datasetu, prípadne experimentom s reálnymi používateľmi. V experimentoch sme sa porovnávali s metódami založenými na známych riešeniach v doméne a s výsledkami publikovanými inými autormi. Môžeme konštatovať, že naša metóda zvyšuje presnosť odporúčania skupinám používateľov a taktiež usporiadanie odporúčených položiek podľa ich kvality. Zvýšenie presnosti odporúčania najmä na prvých pozíciách je veľmi pozitívnou správou najmä v doménach s rozsiahlymi položkami, ako je napríklad pri televízií, kde prezretie položky trvá pomerne dlhú dobu a používateľ si teda zvyčajne pozrie len málo položiek.

Odporúčacia metóda a rovnako ako model používateľa, sú navrhnuté ako doménovo nezávislé. K svojmu fungovaniu vyžadujú jedine vhodnú štruktúrovanosť obsahu oblasti, v ktorej budú použité a možnosť charakterizovať položky z viacerých pohľadov. Navrhnutú hybridnú odporúčaciu metódu je navyše možné použiť s akýmkoľvek modelom používateľa, ktorý súčasne umožňuje porovnávanie podobnosti medzi modelmi pre účely kolaboratívneho odporúčania a tiež odporúčanie na základe obsahu.

Literatúra

- Aggarwal, C., Wolf, J., Wu, K., Yu, P.S. (1999). Horting hatches an egg: A new graph- theoretic approach to collaborative filtering, in *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, CA, pp. 201-212, ACM.
- Amer-Yahia, S., Roy, S., Chawlat, A. & kol. (2009). Group recommendation: Semantics and efficiency. *Proceedings of the VLDB Endowment* (2), pp. 754–765, VLDB Endowment.
- Baltrunas, L., Makcinskas, T., & Ricci, F. (2010). Group recommendations with rank aggregation and collaborative filtering. *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*, pp. 119–126, ACM.
- Basu, C., Hirsh, H. & Cohen W. (1998). Recommendation as Classification: Using Social and Content-Based Information in Recommendation. *ProC. of the 15th Natl. Conf. on AI, Madison, WI*, pp. 714-720, ACM.
- Billsus, D. & Pazzani, M. (2000). User Modeling for Adaptive News Access. *UMUAI 10*(2-3), pp. 147-180, ACM.
- Breese, J.S., Heckerman, D., Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering, in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, pp. 43-52, ACM
- Burke, R. (2000). Knowledge-based recommender systems. *Encyclopedia of Library and Information Systems*, pp. 1–23.
- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *UMUAI 12* (4), pp. 331- 370, ACM.
- Burke, R. (2007). Hybrid web recommender systems. *The adaptive web*, pp. 377–408, Springer Berlin Heidelberg.
- Debnath, S., Ganguly, N., Mitra, P. (2011). Feature weighting in content based recommendation system using social network analysis. In *Proceedings of the 17th international conference on World Wide Web (WWW '08)*, NY, USA, ACM.
- De Pessemier, T., Dooms, S. & Martens, L. (2012). Design and evaluation of a group recommender system. *Proceedings of the sixth ACM conference on Recommender systems - RecSys '12*. pp. 225-228, ACM.
- Cesar, P., Chorianopoulos, K. & Jensen, J.F. (2008). Social television and user interaction. *Comput. Entertain.* 6, 1, Article 4, ACM.
- Ferman, A.M., van Beek, P., Errico, J.H. & Sezan, M.I. (2003). Multimedia content recommendation engine with automatic inference of user preferences. *Image Processing*,

2003. *ICIP 2003. Proceedings. 2003 International Conference on* , vol.3, no., pp. III, 49-52 vol.2, IEEE.
- Fernandez, Y.B., Pazos Arias, J.J., Nores, M.L. & kol. (2006). AVATAR: an improved solution for personalized TV based on semantic inference. *Consumer Electronics, IEEE Transactions*, pp. 223-231, IEEE.
- Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1995). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Goldberg, D., Nichols, D., Oki, B. M. & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35 (12), pp. 61–70, ACM.
- Goldberg, K., Roeder, T., Gupta, D., Perkins, C. (2000). Eigentaste: A Constant Time Collaborative Filtering Algorithm, *UCB ERL Technical Report M00/41*. ACM
- Goren-Bar, D. & Glinansky O. (2004). FIT-recommending TV programs to family members, *Computers & Graphics*, '28, pp. 149-156.
- Guzzi, F., Ricci, F. & Burke, R. (2011). Interactive multi-party critiquing for group recommendation. *Proceedings of the fifth ACM conference on Recommender systems RecSys '11*, pp. 265-268, ACM.
- Kagita, V.R., Pujari, A.K. & Padmanabhan, V. (2013). Group Recommender Systems: A Virtual User Approach Based on Precedence Mining.. In *S. Cranefield & A. Nayak (eds.), Australasian Conference on Artificial Intelligence*, pp. 434-440, Springer.
- Kim, J.K., Kim, H.K., Oh, H.Y. & Ryu, Y.U. (2010). A group recommendation system for online communities. *International Journal of Information Management*, vol. 30, pp. 212-219, Elsevier Science Publishers.
- Kompan, M. & Bieliková, M. (2014). Group Recommendations: Survey and Perspectives, *In Computing and Informatics*, ISSN 1335-9150, Vol. 33, No. 2.
- Krulwich, B. (1997). Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine* 18 (2), pp. 37-45, AAAI.
- Lin, K.H., Chiu Y.S. & Chen, J.S. (2011). An adaptive correlation-based group recommendation system, *Intelligent Signal Processing and Communications Systems (ISPACS)*, pp.1-5, IEEE.
- Liu, X., Tian, Y., Ye, M. & Lee, W. (2012). Exploring personal impact for group recommendation. *Proceedings of the 21st ACM international conference on Information and knowledge management CIKM '12*, pp.674–683, ACM.
- Masthoff, J. (2004). Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Modeling and User-Adapted Interaction*, 14 (1), pp. 37–85, Kluwer Academic Publishers.

- Masthoff, J. (2011). Group Recommender Systems: Combining Individual Models. *Recommender Systems Handbook*. pp. 677-702, Springer US.
- Melville, P., Mooney, R. J. & Nagarajan, R. (2002). Content-Boosted Collaborative Filtering for Improved Recommendations. *Proc. of the 18th Natl. Conf. on AI (AAAI-2002)*, pp. 187-192, ACM.
- Mobasher, B., Jin, X. & Zhou, Y. (2004). Semantically Enhanced Collaborative Filtering on the Web'. *Web Mining: From Web to Semantic Web. LNAI (3209)*, pp. 57-76, Springer.
- Mukherjee, R., Sajja, N., Sen, S. (2003). A Movie Recommendation System—An Application of Voting Theory in User Modeling. *User Modeling and User-Adapted Interaction*13, 1-2 (February 2003), pp. 5-33. ACM.
- Ntoutsi I. & Stefanidis.K. (2012A). gRecs: A group recommendation system based on user clustering. *Database Systems for Advanced Applications*, pp. 299-303, Springer Berlin Heidelberg.
- Ntoutsi, I., Stefanidis, K., Nørnvåg, K. & Kriegel H. (2012B). Fast Group Recommendations by Applying User Clustering. *Lecture Notes in Computer Science*. (vol. 7532) pp. 126–140. Springer Berlin Heidelberg.
- O'Connor, M., Cosley, D., Konstan, J., & Riedl, J. (2002). PolyLens: A Recommender System for Groups of Users, *Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work ECSCW'01*, pp. 199-218, Springer Netherlands.
- Pazzani, M. J. (1999). A Framework for Collaborative, Content-Based and Demographic Filtering. *AI Review 13 (5/6)*, pp. 393-408, ACM.
- Quijano-Sanchez, L., Recio-Garcia, J., & Diaz-Agudo B. (2010). Personality and Social Trust. in Group Recommendations, *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, pp. 121-126, IEEE.
- Quijano-Sanchez, L., Recio-Garcia, J., Diaz-Agudo, B. & Jimenez-Diaz, G. (2013). Social factors in group recommender systems. *ACM Transactions on Intelligent Systems and Technology*, 4(1), pp. 1–30, ACM.
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B., editors (2011). *Recommender Systems Handbook*. Springer.
- Senot, C., Kostadinov, D. & Bouzid, M. (2010). Analysis of strategies for building group profiles. *User Modeling, Adaptation, and Personalization*, pp. 40–51, Springer Berlin Heidelberg.
- Smyth, B. & Cotter, P. (2000). A Personalized TV Listings Service for the Digital TV Age. *Knowledge-Based Systems 13*, (pp.53-59), Elsevier.
- Ungar, L.H., Foster, D.P. (1998). Clustering Methods for Collaborative Filtering, in *AAAI Workshop on Recommendation Systems*, Menlo Park, CA, AAAI.

- van Meeteren, R., & van Someren, M. (2000). Using Content-based Filtering for Recommendation. In *Potamias, G., Moustakis, V., and van Someren, M., editors, ECML/MLNET Workshop on Machine Learning and the New Information Age*, pp. 47–56, ECML.
- Wang, L., Meng, X., Zhang, Y. & Shi, Y. (2010). New approaches to moodbased hybrid collaborative filtering. In *Proceedings of the Workshop on Context- Aware Movie Recommendation - CAMRa '10*, ACM.
- Wang, Z., Zhang, M., Tan, Y. & kol. (2011). Recommendation Algorithm Based on Graph-Model Considering User Background Information, In *'11 Proceedings of the 2011 9th International Conference on Creating, Connecting and Collaborating through Computing*, pp.32-39, IEEE.
- Weiß, D., Scheuerer, J., Wenleder, M. & kol. (2008). A user profile-based personalization system for digital multimedia content. *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*, pp. 281, ACM.
- Wongchokprasitti, C. & Brusilovsky, P. (2007). NewsMe: A Case Study for Adaptive News Systems with Open User Model. *Third International Conference on Autonomic and Autonomous Systems (ICAS07)*, pp. 69, IEEE.
- Yu, Z., Zhou, X., Hao, Y. & Gu, J. (2006). TV Program Recommendation for Multiple Viewers Based on user Profile Merging. *User Modeling and User-Adapted Interaction '16*, pp. 63-82, Springer Netherlands.