

Metódy získavania používateľských preferencií

Veronika Vaneková¹

¹ Ústav informatiky, Prírodovedecká fakulta, Univerzita Pavla Jozefa Šafárika v Košiciach.
Jesenná 9, 040 01 Košice, Slovakia
veronika.vanekova@upjs.sk

Abstrakt. Používateľské preferencie zohrávajú dôležitú úlohu v mnohých systémoch, napríklad v oblasti vyhľadávania alebo elektronického obchodu. Môžu byť založené na rôznych teoretických modeloch. Kritickým momentom je zistenie preferencií pre nových používateľov, o ktorých ešte systém nemá dostatok informácií. V tomto článku porovnávame niekoľko spôsobov explicitného a implicitného získavania informácií od používateľa. Popisujeme tiež systém založený na fuzzy množinách a agregáčnych operátoroch, ktorý spája obidva prístupy podľa množstva informácií aktuálne získaných od používateľa.

Kľúčové slová: model používateľa, fuzzy množiny, získavanie preferencií

1 Úvod

Stále narastajúce množstvo informácií dostupných na internete spôsobuje problémy pri ich vyhľadávaní a filtrovaní. Mnohé systémy riešia tento problém pomocou personalizácie, teda automatického prispôbenia svojho obsahu pre konkrétneho používateľa. Personalizácia je rozšírená predovšetkým v oblasti elektronického obchodu, ale môže byť súčasťou ľubovoľnej aplikácie, ktorá poskytuje veľké množstvo dát.

Personalizované systémy sa navzájom odlišujú tým, ako reprezentujú používateľské preferencie a ako ich od používateľa získavajú. V tomto článku popisujeme systém získavania preferencií, ktorý používa priame zadanie, skupiny podobných používateľov a akceptuje aj preferencie získané indukčnými metódami z ohodnotených objektov alebo implicitnými metódami zo správania používateľa na stránke. Všetky preferencie vytvárajú model používateľa založený na fuzzy množinách a agregáčnych operátoroch.

2 Modely preferencií

Najjednoduchším modelom preferencií je *vektor ohodnotení* jednotlivých objektov. Tento model využívajú *recommendery* [1] - systémy, ktoré v rámci internetových obchodov odporúčajú zákazníkom výrobky. Preferencie používateľa sú reprezentované ako vektor (p_1, \dots, p_n) , kde p_i je ohodnotenie i -teho výrobku. Keďže internetový obchod môže obsahovať niekoľko miliónov výrobkov, tento vektor je typicky riedky, teda väčšina hodnôt je neznámych. Úlohou recommendera je

odhadnúť ohodnotenie výrobku, ktorý používateľ ešte nevidel a odporúčať používateľovi výrobky s najvyšším odhadovaným ohodnotením.

Ohodnotenie je možné vypočítať metódou *kolaboratívneho filtrovania* na základe vzťahov medzi používateľmi alebo medzi objektmi. V prvom prípade systém nájde skupinu podobných používateľov podľa ich vektorov preferencií. Potom pre každý výrobok spočíta priemerné ohodnotenie v rámci skupiny a vyberie výrobky s najvyšším ohodnotením. V druhom prípade systém podľa danej metriky určí skupinu podobných výrobkov a na základe priemerných ohodnotení v tejto skupine ich odporúča používateľovi.

Nevýhodou recommenderov je to, že nedokážu určiť preferencie pre nového používateľa, ktorý ešte neohodnotil nijaké výrobky. Na odstránenie tohto problému boli navrhnuté rôzne metódy [5].

Ak sú hľadané objekty v textovom formáte alebo sú textovo anotované, je možné využiť vyhľadávanie v týchto textoch. Preferencie používateľa sú v tomto prípade vyjadrené ako *množina kľúčových slov*. Používateľ zadáva kľúčové slová ručne alebo ich systém zistí z ohodnotených objektov a zaradí do profilu. Následne odporúča používateľovi tie objekty, ktoré sa s profilom najviac zhodujú. Existujú rôzne metódy [2] na zistenie preferovaných kľúčových slov z ohodnotených dokumentov (data mining, štatistické učenie), aj na určovanie podobnosti (tf-idf model, clustering).

Pri budovaní používateľského profilu z ohodnotených objektov nastáva problém nového používateľa, rovnako ako pri kolaboratívnom filtrovaní. Navyše je ťažké vyjadriť pomocou kľúčových slov dostatočne presné preferencie.

Pri *multikriteriálnom rozhodovaní* systém eviduje rôzne vlastnosti (atribúty) objektov. Používateľ môže vybrať hodnoty vlastností, ktoré ho zaujímajú a systém vyberie objekty, ktoré tieto obmedzenia spĺňajú. Všetky získané objekty síce vyhovujú zadaným obmedzeniam, ale nie je možné ich podľa takýchto kritérií usporiadať. Ďalšou nevýhodou je to, že systém nezíska tie objekty, ktorým chýba jedna požadovaná vlastnosť, ale pritom spĺňajú ostatné vlastnosti. Riešením je použitie namiesto klasických množím preferovaných hodnôt fuzzy množiny.

Fuzzy množiny pripúšťajú čiastočné členstvo prvkov. Hodnota charakteristickej funkcie je zo spojitého intervalu $[0,1]$ a vyjadruje preferencie používateľa k nejakej vlastnosti objektu. Čím lepšia je hodnota vlastností pre daného používateľa, tým má táto hodnota vyššiu mieru členstva v danej fuzzy množine.

Ak chceme brať do úvahy viac vlastností toho istého objektu a vypočítať celkové ohodnotenie, použijeme agregáčnu funkciu. Jednoduchým príkladom agregáčnej funkcie je vážený aritmetický priemer. Váhy vyjadrujú dôležitosť príslušných vlastností. Inou alternatívou je monotónna klasifikácia, ktorá sa dá získať z dát a zapísať vo forme pravidiel [4].

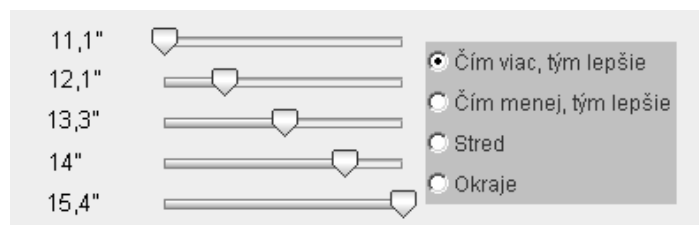
3 Metódy získavania dát o používateľoch

Vo všeobecnosti poznáme dve základné metódy získavania preferencií: explicitnú a implicitnú. Explicitná metóda vyžaduje priamy vstup od používateľa. Týmto vstupom môže byť ohodnotenie vzorky objektov, usporiadanie dvojíc objektov, výber preferovaných hodnôt vlastností. Čím zložitejší je model preferencií, tým ťažšie je

získať ho priamo od používateľa. Explicitné získavanie preferencií musí byť podporené jednoduchým a intuitívnym rozhraním. Nazbierané dáta je možné analyzovať, napríklad z ohodnotenej vzorky objektov odvodiť preferencie vzhľadom na atribúty.

Implicitná metóda je založená na myšlienke, že preferencie používateľa sú zjavné z jeho správania. Používatelia objednávajú výrobky, sťahujú dokumenty, čítajú stránky a podobne, pričom väčšinou sa jedná práve o tie objekty, ktoré ich zaujímajú. Systém môže zaznamenávať významné akcie používateľa, analyzovať ich a získať z nich preferencie. Tento spôsob vedie k úplnejším a objektívnejším dátam. Zbiera informácie o všetkých používateľoch, aj o tých, ktorí by svoje preferencie nešpecifikovali priamo.

4 UPreA



Obrázok 1. Preferencie k atribútu “uhlopriečka displeja” v doméne notebookov.

User Preference Acquirer (UPreA) používa fuzzy model preferencií popísaný v kapitole 2. Pre každý atribút umožňuje explicitne zadefinovať fuzzy množinu pomocou grafického rozhrania, ktoré znázorňuje obrázok 1. Pre každú hodnotu atribútu obsahuje toto rozhranie jeden posuvník, ktorým sa nastaví hodnota preferencie. Navyše je možné zvoliť jeden zo štyroch základných typov preferencie.

O každom používateľovi ukladáme aj osobné údaje ako veková skupina, rodinný stav, bydlisko a podobne, ktoré získame z vyplneného dotazníka pri registrácii v systéme. Používatelia s rovnakými údajmi môžu mať aj podobné preferencie, aspoň v niektorých atribútoch. Preto v prípade, že používateľ nezádá preferencie sám, UPreA ich doplní podľa prevládajúcich preferencií ostatných používateľov rovnakého typu. Preferencie využije top-k algoritmus [3] na vyhľadávanie najlepších ponúk a neskôr je možné ich spresniť.

	Osobné údaje	Explicitné fuzzy množiny	Implicitné obmedzenia atribútov	Ohodnotenie vzorky objektov
Fuzzy množiny	Kolaboratívne filtrovanie	Prepíše	Prispôsobí podľa obmedzení	Indukciou odvodí nové fuzzy množiny.
Agregačná funkcia	Nastaví rovnaké váhy	Nastaví rovnaké váhy	-	Indukciou odvodí pravidlá.

Tabuľka 1 Správa preferencií v systéme UPreA.

UPreA umožňuje zaradiť do modelu používateľa aj klasifikačné pravidlá získané z ohodnotených ponúk [4] a tiež implicitné preferencie získané zo záznamu používateľových akcií v systéme [6]. Podľa toho, koľko informácií o používateľovi máme, môžu nastať viaceré prípady. Tabuľka 1 opisuje to, ako aktualizuje UPreA fuzzy množiny alebo agregáčne funkcie pri rôznych vstupoch od používateľa. Typ vstupu je špecifikovaný v hlavičke každého stĺpca, napríklad „ohodnotenie vzorky objektov“.

5 Záver

Problém určenia preferencií pre nového používateľa je možné riešiť kombináciou rôznych prístupov: explicitného zadania cez grafické rozhranie, kolaboratívneho filtrovania, indukčných metód a analýzy správania používateľa v systéme. Program UPreA poskytuje jednotný model, ktorý spravuje a aktualizuje podľa aktuálneho stavu profilu a podľa vstupov od používateľa.

Referencie

1. Candillier, L., Meyer, F., Boullé, M. Comparing State-of-the-Art Collaborative Filtering Systems. *Machine Learning and Data Mining in Pattern Recognition*. Springer Berlin / Heidelberg 2007. ISBN 978-3-540-73498-7.
2. Grossman, D. A., Frieder, O.: Information Retrieval: Algorithms and Heuristics, Kluwer Academic Publishers, 2000. ISBN 1402030037.
3. Gurský, P.: Towards better semantics in the multifeature querying. *Proceedings of Dateso 2006*.
4. Horváth, T., Vojtáš, P.: Ordinal Classification with Monotonicity Constraints. *Proceedings of ICDM 2006*. Springer, 2006. ISBN 0-7695-2702-7.
5. Schein, A. I., Popescul, A., Ungar, L. H., Pennock, D. M. Methods and metrics for cold-start recommendations. *Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in information Retrieval*. ACM, New York 2002. ISBN 1-58113-561-0.
6. Tvarožek M., Bieliková M.: Adaptive faceted browser for navigation in open information spaces. *Proceedings of the 16th international conference on World Wide Web*. ACM Press, New York, USA. ISBN 978-1-59593-654-7.

Annotation:

User Preference Acquisition Methods

User preferences have an important role in many search or e-commerce systems. They can be based on different theoretic models. Critical phase is determining preferences for a new user. In this paper we compare several explicit and implicit user preference acquisition methods. We propose a system based on fuzzy sets and aggregation operators which combine both approaches depending on actual amount of information gained from user.