

Prístupy k vyhodnoteniu aplikácií modelujúcich používateľa

Michal Barla

02.05.2007

Aký je problém?

- Veľa z nás sa venuje niečomu, čo má v názve alebo v motivácií prispôsobovanie
- Ako to vyhodnotiť?
 - Vieme zistiť, či sme vyvinuli to, čo sme navrhli
 - Ako zistiť, či sme to navrhli správne, užitočne?
- Ako sa navzájom porovnať?
 - Doménovo špecifické časti
 - Neexistujúci benchmark

Malé pripomenutie

- Adaptívny systém pracuje v dvoch etapách
 - Tvorba modelu používateľa
 - Prispôsobovanie na báze modelu

Základná otázka

Vyhodnocovať systém ako celok alebo po etapách?

Dve možnosti

- Informačný priestor je dostatočne malý ~ poznáme všetky jeho koncepty
- Informačný priestor je rozsiahly ~ nevieme čo všetko obsahuje

Malý informačný priestor

Podmienka

Používateľ systému explicitne povie aké sú jeho záujmy a pod.

- Systém vytvorí a použije model
- Vyhodnocujeme, či systém odporučil, to čo by mal
 - Precision, recall, F1

Rozsiahly informačný priestor

Podmienka

Máme k dispozícii verziu bez prispôsobovania a s prispôsobovaním

- Máme dve skupiny používateľov, ktorí používajú dve verzie systému (T-test)
- Sledujeme rozdiely
 - Počet navštívených uzlov
 - Čas potrebný na vykonanie úlohy
 - Prípadné ďalšie dopady vzhľadom na cieľ (úspech v teste)

Vyhodnocovanie po etapách

- Keď pri vyhodnocovaní systému v celku niečo zlyhá
 - Je problém v modelovaní používateľa alebo
 - Je problém v prispôbovaní?
- *Layered Evaluation of Adaptive Applications and Services*
 - Odhalil systém úspešne charakteristiky používateľa a uložil ich do modelu?
 - Bolo zvolené adekvátne, zmysluplné prispôsobenie?

Príklad

- Model používateľa¹
 - *not ready to be learned*
 - *ready and recommended*
 - *nothing new*
- Veľmi málo krokov išlo cez anotované odkazy
- Korelácia času čítania a údajov o stránke v modeli
 - Prvá fáza je dobrá
 - Problém je v anotácií odkazov

¹Brusilovsky et al.: The Benefits of Layered Evaluation of Adaptive Applications and Services

Vplyv reprezentácie modelu

- Strojové učenie produkuje implicitné modely, ktoré sťažuje oddelené vyhodnotenie
 - Vieme aspoň posúdiť rozlišovaciu schopnosť modelu
- Explicitné, znalostné reprezentácie sú z pohľadu oddeleného vyhodnotenia lepšie

Porovnávanie AHS

Zdroj

Sadat, H., Ghorbani, A: On The Evaluation of Adaptive Web Systems

- Porovnávanie na základe črt systému
 - Poskytovaná funkcionálnosť
 - Použité technológie
 - Softvérové inžinierstvo
- Jednotlivé črty sú ováňované
- Na konci máme pre každý systém jedno číslo
 - Čím vyššie tým lepšie