

Data Mining in Action

Recommender Systems I

Terminology

- User
- Item
- Rating
 - Explicit / Non-explicit
 - Real- or integer-valued / binary / unary

- Content-based Recommenders
- Collaborative Filtering

Popular

Главная Набирающие популярность Подписки



19 марта 2016 Ростов-на-Дону. Упал самолёт в районе аэропорта, все 55 человек погибли.
Екатерина Киселева
11 часов назад • 178 553 просмотра
В Ростове-на-Дону разбился самолет с 55 россиянами на борту
...



This is Хорошо - Фляк-фляк-фляк! #506
This is Хорошо
21 час назад • 470 245 просмотров
Ссылка на наш магазин: <https://zileshop.com/>
...



Кукла Барби. Ярославла получает посылку на почте. Видео для девочек. Дом для Барби. Tiki Taki Cook
Tiki Taki Cook
1 день назад • 588 553 просмотра
Привет! Скорей к Экрану! Сегодня Ярослава с Куклой Барби отправится на почту, чтобы получить посылку. Ярослава еще не...



"А если здесь выйти со свастикой?"
Анатолий Шарий
1 день назад • 202 299 просмотров
В Германии - о запретных символах



Последние переговоры пилотов упавшего «Боинга» с диспетчерами аэропорта Ростова: аудио
DonDay Ростов
9 часов назад • 50 439 просмотров



10 ТАИНСТВЕННЫХ АРХЕОЛОГИЧЕСКИХ НАХОДОК
DaiFiveTop
22 часа назад • 264 297 просмотров
ПОДПИШИСЬ НА НОВЫЕ ВЫПУСКИ ► <http://bit.ly/daifiveTOP>
Канал Dameoz ► <http://bit.ly/DameozChan...>



по пивасику?
Главные Герои
1 день назад • 568 928 просмотров



Огромный желевый Мег Медведь с мармеладными червяками делаем сами World...
Miss Katy
20 часов назад • 673 284 просмотра
Катя с папой делают огромного 10 кг Медведя из фруктового желе с червячками и мармеладными фруктами...



БУДЬ КРАНОМ -||- MaxiMum Override
FROST
1 день назад • 611 269 просмотров
ГОРОД в ОПАСНОСТИ! Фрост\Сквидвард КРАНЕ! Классная ИГРА)
Я в ВК <http://vk.com/id192035364...>



В Ростове на Дону разбился пассажирский самолёт 19 марта 2016 года
Аварии Катастрофы
9 часов назад • 27 678 просмотров
В Ростове-на-Дону разбился пассажирский самолёт
Все пассажиры разбившегося в Ростове-на-Дону «Боинга» В Южно...



Лучшие Реплеи Недели с Кириллом Орешкиным #79 [World of Tanks]
World of Tanks. Официальный видеоканал
1 день назад • 298 854 просмотра
ЛРН снова готов радовать вас крутыми боями мира танков! Сегодня вы увидите, как елка дрифтует за медальку колобанова, объект 14...



Я БЕЗ ФОТОШОПА | ИГРОФЕСТ
EeOneGuy
1 день назад • 1 958 650 просмотров
Игрофест Google Play (10 Марта - 16 Апреля)
Скачивай Star Wars: Galaxy of Heroes сейчас на Google Play, и...

Content-based

- Systems implementing a *content-based recommendation* approach analyze a set of documents and/or descriptions of items previously rated by a user, and build a model or profile of user interests based on the features of the objects rated by that user

- Content Analyzer — how to represent items
- Profile Learner — how to learn user interests
- Filtering Component — how to filter items

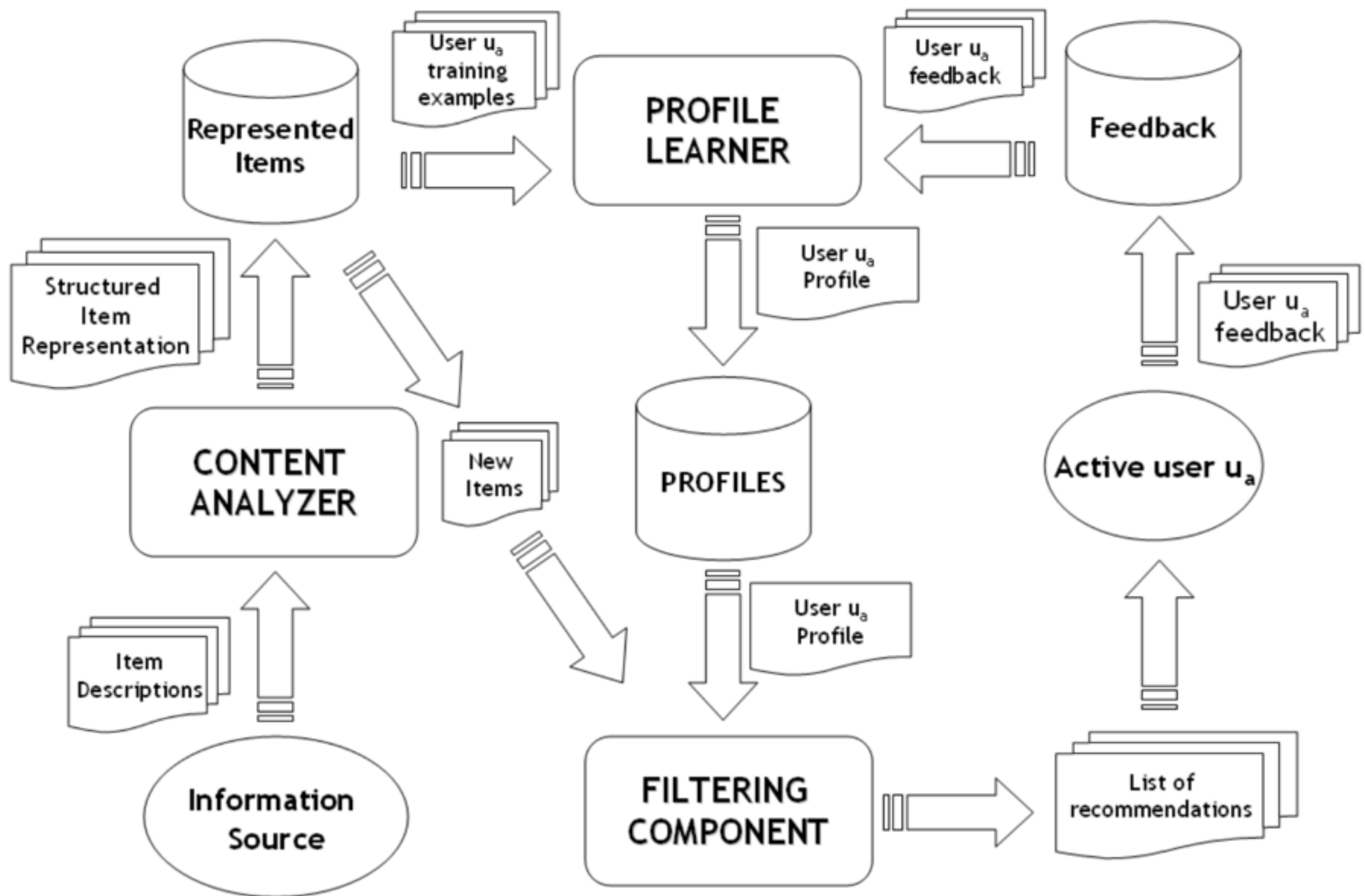


Fig. 3.1: High level architecture of a Content-based Recommender

Content Analyzer

- Most often items consist of text: web-pages, product descriptions, etc.
- Represent items as vectors: Bag-Of-Words, TF-IDF

TF-IDF

$D = \{d_1, d_2, \dots, d_N\}$ – corpus of documents

$T = \{t_1, t_2, \dots, t_n\}$ – dictionary

$d_j = (w_{1j}, w_{2j}, \dots, w_{nj})$, where w_{kj} is the weight for term t_k in document d_j

$$\text{TF-IDF}(t_k, d_j) = \underbrace{\text{TF}(t_k, d_j)}_{\text{TF}} \cdot \underbrace{\log \frac{N}{n_k}}_{\text{IDF}}$$

$$\text{TF}(t_k, d_j) = \frac{f_{k,j}}{\max_z f_{z,j}}$$

n_k – number of documents in the collection in which the term t_k occurs at least once

$f_{z,j}$ – frequency of term t_z in document d_j

$$\text{sim}(d_i, d_j) = \frac{\sum_k w_{ki} \cdot w_{kj}}{\sqrt{\sum_k w_{ki}^2} \cdot \sqrt{\sum_k w_{kj}^2}}$$

Profile Learner

- Neighbourhood based approach:

$$\text{User } u \text{ profile } \mathbf{x}_u = \sum_{i \in I_u} \alpha_i r_{ui} \mathbf{x}_i$$

I_u – a set of rated items by the user u

r_{ui} – rating of item i given by the user u

- Build supervised model on I_u training set:
 - Naive Bayes
 - Decision trees
 - etc.

Issues

- Serendipity problem
- Cold-start problem

What's next?

- Recommender Systems Handbook
- “Recommender Systems” on Coursera