

Collaborative Filtering

John Ashley Burgoyne

MUMT 611

22 March 2007

How do you find music?

- Chances are, you have found at least one piece of music in your collection as the result of collaborative filtering (CF).
- CF is a technique for making recommendations based on the preferences of other similar users.
- CF is very popular in commercial and open-source applications.

Outline

- Motivation for CF
- Overview of the CF algorithm
- Well-known applications
- Active research

Long (and heavy) tails

- coined by C.Anderson, *Wired*, Oct. 2004
- colloquialism for heavy tails of a PDF
- converse formulation of the ‘hit’ phenomenon
- potential for personal enjoyment and financial profit – if it can be accessed

Collaborative filtering

- typical example: user-rated products
- rating matrix is usually very sparse
- many singularities (i.e., dependencies)
- leverage dependencies to enrich matrix
- many techniques to accomplish the above, e.g., neural networks (Bengio et al.)

Varieties of CF

- user-based
 - identify users who behave similarly
 - combine ratings of this group of users
- item-based
 - cross-correlate items only
 - use purchasing history to recommend

amazon.com

- enormous purchasing history database
- item-based collaborative filtering
 - users who bought this item also bought...
 - users who viewed this item ultimately bought...
- Barnes & Noble uses a similar approach

TiVo and Netflix

- TiVo
 - U.S.-only DVR system with CF-based recommendations
- Netflix
 - U.S.-only DVD rental programme
 - recommends new movies to see based on viewing history

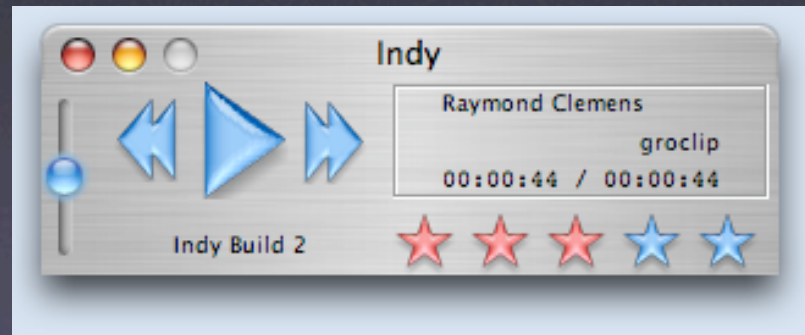
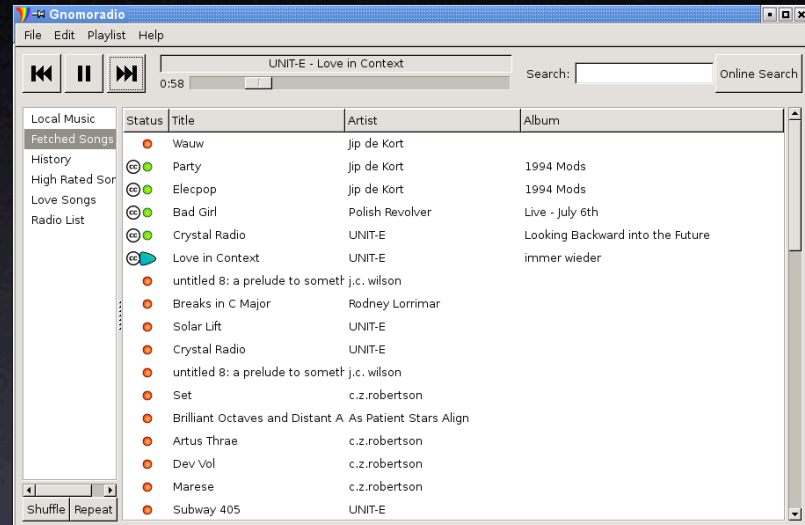


Musicmatch and last.fm

- Musicmatch jukebox has been purchased and integrated with Yahoo!
- Last.fm & Audioscrobbler
 - Audioscrobbler records the playing history of the music on users' computers
 - Last.fm uses CF to generate playlists from Audioscrobbler data and tags

Free music CF systems

- Gnomoradio
- Indy
- iRATE radio
- Musicmobs
- MyStrands
- Rate Your Music



Software packages

- Java: Taste, Cofi, CoFE, RACOFI
- C: SUGGEST
- PHP: Vogoo, Rating-Based Item-to-Item
- Python: consensus
- MATLAB: open-source toolkit

Active research

- Can CF be combined with content-based filtering (Yoshii, Goto, et al. 2006)?
- Is CF sufficient to generate playlists (Cunningham, Bainbridge, et al. 2006)?
- Where does one find data for CF, e.g., text-mining of blogs (Whitman 2004).

Summary

- *easy to use*
- considerable commercial interest
- considerable commercial success
- *easy to try*