

Author(s): Rahul Sami, 2009

License: Unless otherwise noted, this material is made available under the terms of the **Creative Commons Attribution Noncommercial Share Alike 3.0 License:**

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

We have reviewed this material in accordance with U.S. Copyright Law **and have tried to maximize your ability to use, share, and adapt it.** The citation key on the following slide provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact open.michigan@umich.edu with any questions, corrections, or clarification regarding the use of content.

For more information about **how to cite** these materials visit <http://open.umich.edu/education/about/terms-of-use>.

Citation Key

for more information see: <http://open.umich.edu/wiki/CitationPolicy>

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }



Public Domain – Government: Works that are produced by the U.S. Government. (USC 17 § 105)



Public Domain – Expired: Works that are no longer protected due to an expired copyright term.



Public Domain – Self Dedicated: Works that a copyright holder has dedicated to the public domain.



Creative Commons – Zero Waiver



Creative Commons – Attribution License



Creative Commons – Attribution Share Alike License



Creative Commons – Attribution Noncommercial License



Creative Commons – Attribution Noncommercial Share Alike License



GNU – Free Documentation License

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }



Public Domain – Ineligible: Works that are ineligible for copyright protection in the U.S. (USC 17 § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }



Fair Use: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (USC 17 § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

SI 583
Recommender Systems

Rahul Sami
Winter 2009

Prerequisites

- Some exposure to basic statistics (e.g., from SI544) for the concepts of probability, expectation, variance.
- We will be covering and using linear algebra/matrix notation
- See me if you have any questions about whether you have sufficient background.

Course Goals

At the end of this course, you should be able to

- identify potential application domains for recommender systems
- generate recommender designs through an exploration of the design space
- critique a design to identify potential strengths and weaknesses, and compare design alternatives

What is a Recommender System?

- A working definition:

A system to guide users towards items/objects that they are likely to appreciate.

- The range of recommender systems is better grasped through examples

Example: Amazon Recommendations

Rahul's Amazon.com™ > **Recommended for You**

(If you're not Rahul, [click here.](#))

Recommendations by Category

Your Favorites [Edit](#)

[Books](#)

More Categories

[Apparel & Accessories](#)

[Baby](#)

[Beauty](#)

[Camera & Photo](#)

[Computer & Video Games](#)

[Computers & PC Hardware](#)

[DVD](#)

[Electronics](#)

[Gourmet Food](#)

[Health & Personal Care](#)

[Industrial & Scientific](#)

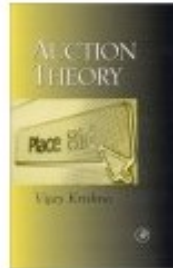
[Jewelry & Watches](#)

These recommendations are based on [items you own](#) and more.

view: **All** | [New Releases](#) | [Coming Soon](#)

[More results](#)

1.



[Auction Theory](#)

by Vijay Krishna

Average Customer Review: ★★★★★

In Stock

Publication Date: March 1, 2002

Our Price: \$52.46 [Used & new](#) from \$52.46

[Add to cart](#)

[Add to Wish List](#)

I Own It Not interested x|★★★★★ Rate It

Recommended because you purchased Putting Auction Theory to Work and more ([edit](#))

2.



[Canon Matte Photo Paper \(8.5x11, 50 Sheets\)](#)

by Canon

Average Customer Review: ★★★★★

Signed by Verisign, I

© FAIR USE

<http://www.amazon.com/>

Example: Slashdot comments

- a way of recommending which comments are worth reading..

↳ [Re:the Dual Music Player](#) by the_womble (Score:2) Monday January 07, @12:31PM

hmm. (Score:2)

by [apodyopsis \(1048476\)](#) on Monday January 07, @09:52AM ([#21941910](#))

The Jam trousers Q? now whose idea were those, bloody silly.

To be fair the only part of TFA that got to me were the iRing (jokes abound for the silly name) and Sony apple remote controls. They are very cool. Expect contactless recharging to be the norm in a few years – now how can I retrofit that into my antique desk?

The concept of jigsaw mini screens seems cool (build your own supersize screen by combining unlimited numbers of smaller screen. Of course then the pricing regime follow the square law – $2 \times \text{area} = 4 \times \text{cost}$. Damn it, I want my wall screen.

[[Reply to This](#)]

↳ [Re:hmm.](#) by ByOhTek (Score:2) Monday January 07, @10:04AM

Anyone watch Top Gear? (Score:3, Insightful)

by [ByOhTek \(1181381\)](#) on Monday January 07, @09:53AM ([#21941914](#)) [Journal](#)

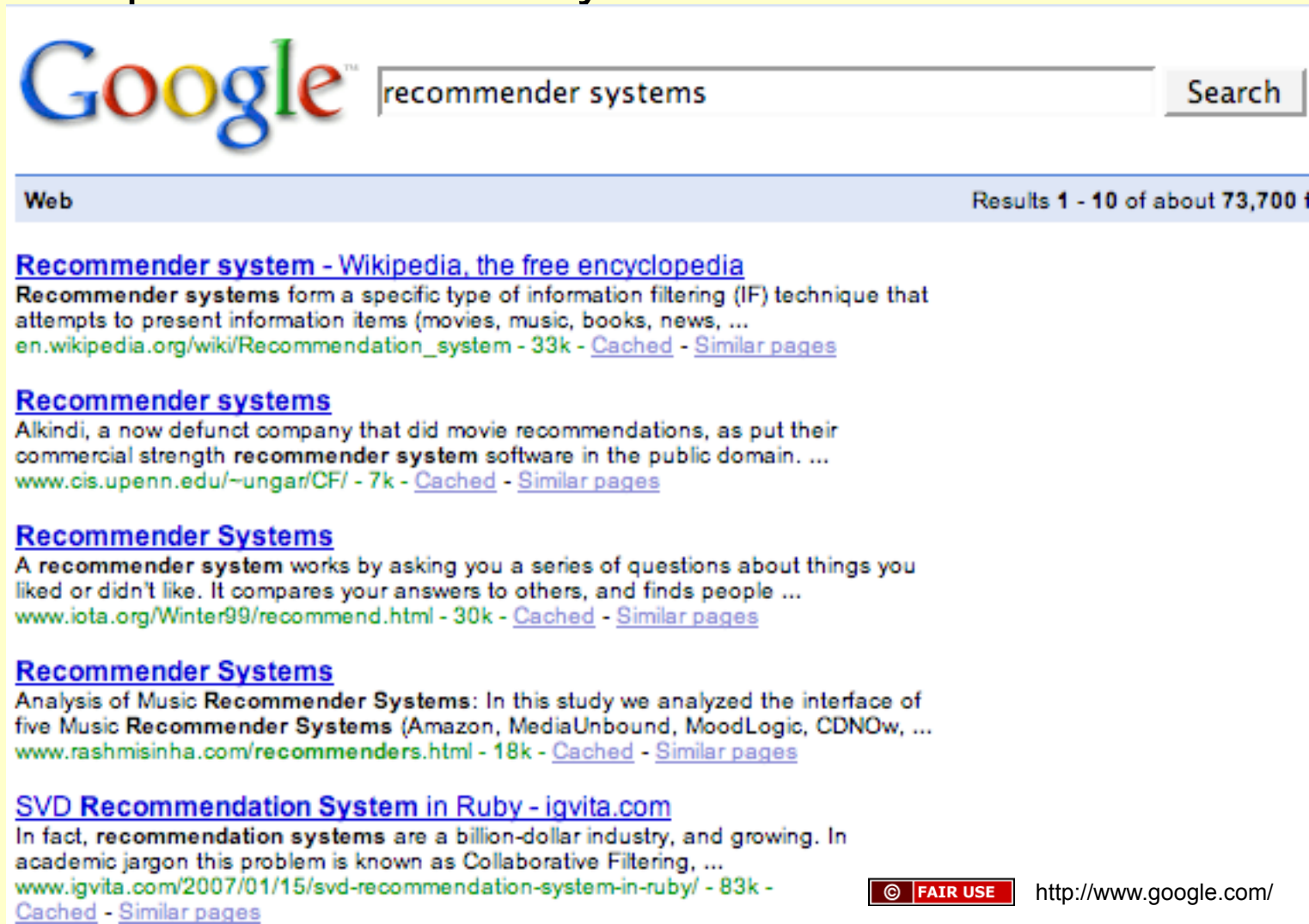
The amphibious car made me think of that. Aside from that there were certainly interesting (and what-where-they-thinking!?) ideas in there.

[[Reply to This](#)]

↳ [Re:Anyone watch Top Gear?](#) by Goffee71 (Score:3) Monday January 07, @10:01AM

Example: Search engines

- Recommends which web pages are worth reading for a particular set of keywords



The image shows a screenshot of a Google search results page. At the top left is the Google logo. To its right is a search input field containing the text "recommender systems" and a "Search" button. Below the search bar, the page indicates "Web" and "Results 1 - 10 of about 73,700". The first four search results are listed, each with a blue title link, a brief description, and a green URL link with "Cached" and "Similar pages" options.

Web Results 1 - 10 of about 73,700

[Recommender system - Wikipedia, the free encyclopedia](#)
Recommender systems form a specific type of information filtering (IF) technique that attempts to present information items (movies, music, books, news, ...
en.wikipedia.org/wiki/Recommendation_system - 33k - [Cached](#) - [Similar pages](#)

[Recommender systems](#)
Alkindi, a now defunct company that did movie recommendations, as put their commercial strength recommender system software in the public domain. ...
www.cis.upenn.edu/~ungar/CF/ - 7k - [Cached](#) - [Similar pages](#)

[Recommender Systems](#)
A recommender system works by asking you a series of questions about things you liked or didn't like. It compares your answers to others, and finds people ...
www.iota.org/Winter99/recommend.html - 30k - [Cached](#) - [Similar pages](#)

[Recommender Systems](#)
Analysis of Music Recommender Systems: In this study we analyzed the interface of five Music Recommender Systems (Amazon, MediaUnbound, MoodLogic, CDNOW, ...
www.rashmisinha.com/recommenders.html - 18k - [Cached](#) - [Similar pages](#)

[SVD Recommendation System in Ruby - igvita.com](#)
In fact, recommendation systems are a billion-dollar industry, and growing. In academic jargon this problem is known as Collaborative Filtering, ...
www.igvita.com/2007/01/15/svd-recommendation-system-in-ruby/ - 83k - [Cached](#) - [Similar pages](#)

© FAIR USE <http://www.google.com/>

Example: top lists

- Bestseller lists/charts for movies, music, websites (del.icio.us) are a way of guiding users to items that they are *likely* to like
 - because many people seem to like them

Other examples?

- online/offline recommender systems you've come across?

Reputation vs. Recommender systems

Bose AWRC-1G CD Bose Wave Radio w/ Remote

Item number: 120071696472

Bidder or seller of this item? [Sign in](#) for your status

[Watch this item](#) in My eBay | [Email to a friend](#)



[View larger picture](#)

Current bid: **US \$1.00** [Place Bid >](#)

[Reserve not met](#)

End time: **Jan-11-07 13:05:42 PST**
(6 days 22 hours)

Shipping costs: Check item description and payment instructions or contact seller for details

Ships to: United States, Canada

Item location: Los Angeles, CA, United States

History: [1 bid](#)

High bidder: [trik50](#) ([1699](#) ★)

You can also: [Watch this item](#)

Get alerts via [Text message](#), [IM](#) or [Cell phone](#)
[Sell one like this](#)

Meet the seller

Seller: [unpico](#) ([2629](#) ★)

Feedback: **97.9% Positive**

Member: since Dec-04-00 in United States

- [Read feedback comments](#)
- [Ask seller a question](#)
- [Add to Favorite Sellers](#)

[View seller's other items](#)

Buy safely

1. Check the seller's reputation

Score: 2629 | 97.9% Positive
[Read feedback comments](#)

2. Learn how you are protected

Shop without sharing your financial details [Learn more](#)

Reputation vs. Recommender Systems

- Similarities between recommendation and reputation systems:
- Both based on users' past reports
- Fundamental goal of both is to reduce a user's uncertainty about her satisfaction with a particular activity.

Reputation vs. Recommender

Differences between reputation systems and recommendation systems:

- Active agents vs. passive “items”
- Different emphasis: predicting future satisfaction vs. inducing appropriate actions
- Different typical mode of operation: summarizing information (about an agent) vs. selecting from a group (of items.)
- Edges are blurry, e.g., PageRank

Outline of course

- Today: understanding the design space
- Eliciting feedback/recommendation inputs
- Aggregation: Collaborative filtering algorithms (user-user, item-item, singular-value decomposition)
- Implementation and Architecture
- Interface alternatives and effects
- Methods of Evaluating Recommender Systems
- Anonymity and privacy issues
- Deliberate Manipulation

Coursework and evaluation

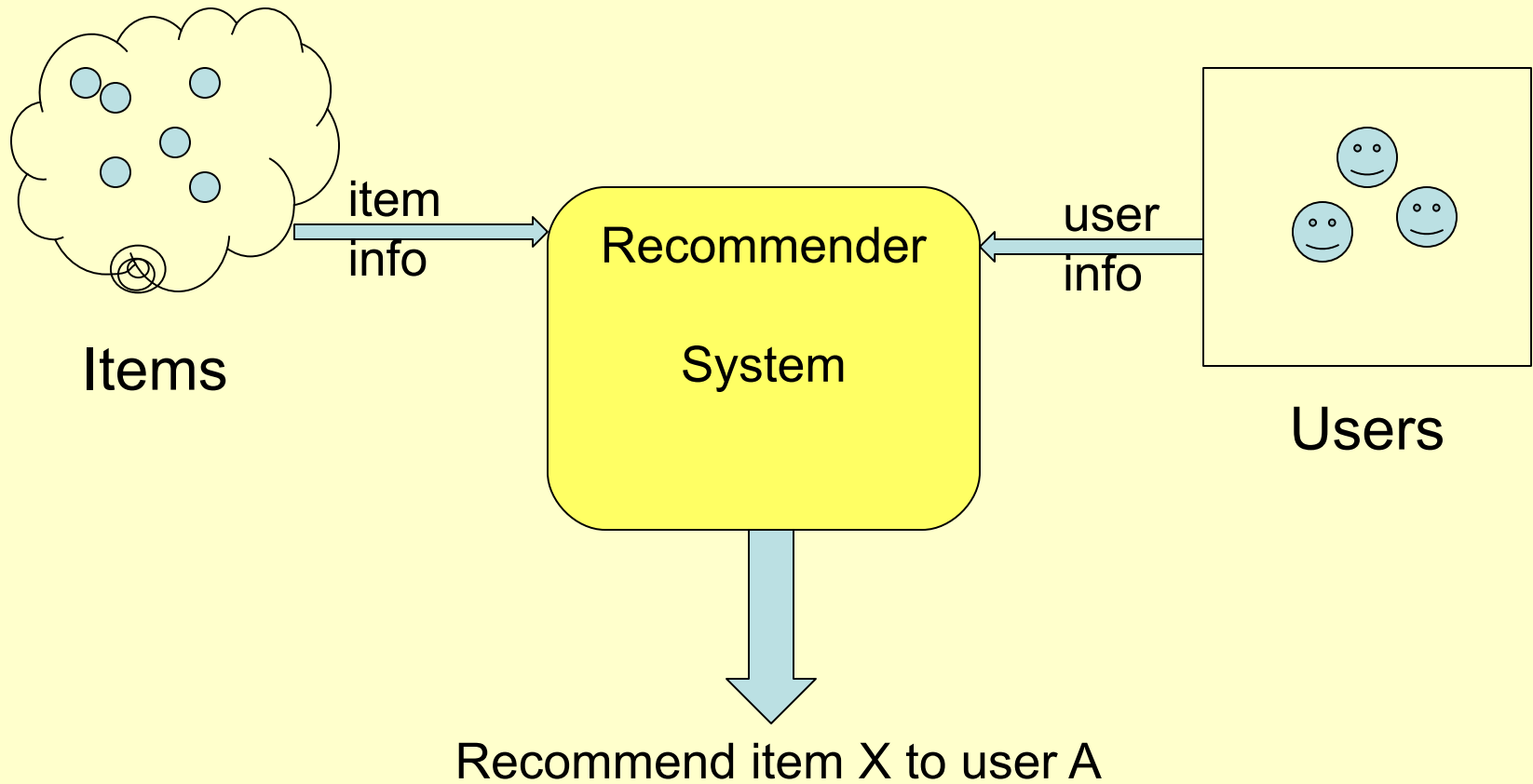
- Every class, read required readings before class.
- 4 Homework Assignments (30%)
- Class Participation (10%)
 - in class, and posting comments, relevant links, and articles to the Ctools discussion forum
 - Intended primarily for motivation, not evaluation
- Term paper (60%)

Term papers

- A short paper that is a mock “consultant’s report” which
 - identifies a potential application for a recommender system
 - explores the design space of a recommender system for that domain
 - suggests a design
 - points out strengths and weaknesses/pitfalls
- Due by Feb 20th (before winter break)

Waitlist

- Come see me at the end of class if you are on the waitlist
- If you are registered, and want to drop, please do so as soon as you are sure.



Sketching The Design Space

Major elements of the technical design space

- Domain (set of items)
- Identity management
- Information sources
- Aggregation: how is the information combined/processed?
- Presentation and interface

Online identity management

- Anonymous, pseudonymous, or attributed users
- Related: personalized vs. nonpersonalized recommendations

Sources of information

- Explicit ratings on a numeric/ 5-star/3-star etc. scale
- Explicit binary ratings (thumbs up/thumbs down)
- Implicit information, e.g.,
 - who bookmarked/linked to the item?
 - how many times was it viewed?
 - how many units were sold?
 - how long did users read the page?
- Item descriptions/features
- User profiles/preferences

Methods of Aggregating inputs

- Content-based filtering
 - recommendations based on item descriptions/features, and profile or past behavior *of the “target” user only*
- Collaborative filtering
 - recommendations based on past behavior of *other users as well as the target user*
- Hybrids

Content-filtering recommenders

- e.g., Pandora music recommender
- Overall operation: categorize items, or identify items with similar features; then recommend either
 - categories that match stated user profile
 - items similar to others the target user has liked/bought etc.



<http://www.pandora.com/>

About Pandora®

When was the last time you fell in love with a new artist or song?

At Pandora, we have a single mission: To play music you'll love - and nothing else.

To understand just how we do this, and why we think we do it really, really well, you need to know about the [Music Genome Pro](#)

Since we started back in 2000, we have been hard at work on the Music Genome Project. It's the most comprehensive analysis of music ever undertaken. Together our team of fifty musician-analysts has been listening to music, one song at a time, studying and collecting thousands of musical details on every song. It takes 20-30 minutes per song to capture all of the little details that give each recording its magic - melody, harmony, instrumentation, rhythm, vocals, lyrics ... and more - close to 400 attributes! We continue this work every day to keep up with the incredible flow of great new music coming from studios, stadiums and garages around the country.

With Pandora you can explore this vast trove of music to your heart's content. Just drop the name of one of your favorite songs on Pandora and let the Genome Project go. It will quickly scan its entire world of analyzed music, almost a century of popular recordings, both well known and completely obscure - to find songs with interesting musical similarities to your choice. Then sit back and enjoy as it plays a listening experience full of current and soon-to-be favorite songs for you.

Content-based filtering

- Example:
 - use number of common words as a similarity measure
 - Recommend “closest” item to liked items
- Content filtering similarity measures are domain-specific
- We will not cover them in this course

Collaborative filtering

- Main idea: users with similar tastes will tend to like similar items
- Use implicit/explicit ratings to:
 - find users similar to the current target user and recommend items they like
 - or, find item Y similar to item X for which most users who liked X like Y
 - or more complex approaches to learn a preference model from ratings
- Do not rely on domain-specific inputs-- basic algorithms can be applied to any CF setting

Hybrid methods

- Combine both content-based and collaborative filtering
- e.g., web search engines use keyword frequency metrics as well as link frequency to come up with a page list

Interface & Presentation

- Personalized/non-personalized recommendations
- Are recommendation levels/predictions used to:
 - filter out bad items
 - displayed next to items
 - sort items to show most recommended items first
- Add explanations?
 - e.g., “This book was recommended to you because you bought ABC”, “Previous customers who bought this also bought”.
- Other feedback to users about how much they have rated
- Quick ways to bootstrap new users

Business Models

- How is the recommendation site supported?

Business Models

- How is the recommendation site supported?
 - Value-addition attached to a purchase/circulation etc. service
 - Advertisements
 - Paid for by content owners
- Any others?

The Netflix challenge

- Netflix released 100 million anonymized ratings
- Kept a set of about 1 Million ratings as a secret “test”
- Challenge: come up with algorithms to accurately predict “test” ratings
- Goal: 10% improvement over Netflix’s own algorithm
- Current leader: Team “Pragmatic Theory”, 9.41 %
- Prize: \$1,000,000
- Any takers? (www.netflixprize.com)