Week 1

Video 1.1.1 : Introduction to Recommender Systems

Learning Objectives:

- Understand what a recommender system is
- Some history and background

A Bit of History

1. Ants, Cavemen, and Early Recommender Systems
   - The emergence of critics
2. Information Retrieval and Filtering
3. Manual Collaborative Filtering
4. Automated Collaborative Filtering
5. The Commercial Era

2. Information Retrieval and Filtering

[Information filtering focuses on building profiles of long-term user interest while information retrieval focuses on building indexes of content.]

a) Information Retrieval [We have content, items, now based on user query-> retrieve]

a) Static content base

– Invest time in indexing content

b) Dynamic information need

– Queries presented in "real time"

c) Common approach: TFIDF

– Rank documents by term overlap

– Rank terms by frequency

b) Information Filtering [Information filtering evaluates new content items for match against user profiles.][We have userProfile now match it with new items]

a) Reverse assumptions from IR

– Static information need

– Dynamic content base

b) Invest effort in modeling user need

– Hand created “profile”

– Machine learned profile

– Feedback/updates

c) Pass new content through filters

3. Manual Collaborative Filtering

Collaborative Filtering

Premise

– Information needs more complex than keywords or topics: quality and taste

Small Community: Manual

– Tapestry– database of content & comments

– Active CF – easy mechanisms for forwarding content to relevant readers

4. Automated CF

• The GroupLens Project (CSCW ’94)

– ACF for Usenet News

• users rate items

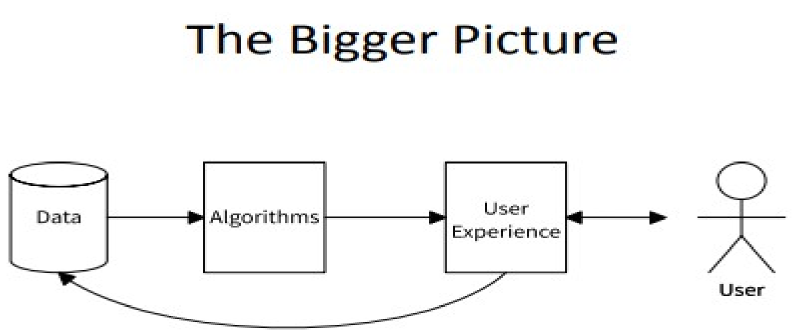
• users are correlated with other users

• personal predictions for unrated items

– Nearest Neighbor Approach

• find people with history of agreement

• assume stable tastes



Recommenders

• Tools to help identify worthwhile stuff

  – Filtering interfaces

    • E mail filters, clipping services

  – Recommendation interfaces

    • Suggestion lists, “top n”, offers and promotions

  – Prediction interfaces

    • Evaluate candidates, predicted ratings

Recommendation Approaches

• Non Personalized and Stereotyped

  – Popularity, Group Preference

• Product Association

  – People who liked/bought X, also like Y

• Content Based

  – Learn what I like (in terms of attributes)

• Collaborative

  – Learn what I like; use others’ experience to recommend

  (many different ways to implement)

Designing a Recommender

• Collecting Opinion and Experience Data

• Finding the Relevant Data for a Purpose

• Computing the Recommendations

• Presenting the Data in a Useful Way

Recommenders as Big Data

• Heavy Emphasis on Analysis and Evaluation

  – Exploring Data to Determine Best Recommendation Approaches

  – Algorithms Optimize Performance Against Metrics

  – Metrics Designed to Improve User Experience and Business Goals

• Continuing Adoption of New Machine Learning Techniques

Video 1.3.1 : Preferences and Ratings

Introduction

- To recommend, we need data (what users like, what goes together, etc.)
- Data comes from users, is collected somehow.
- This lecture's topic: what data we collect, how, and what it means

Learning Objectives

- Understand what data recommenders can use to learn what users like
- Identify types of data collected from users
- Understand when different data types are possible and appropriate
- Be able to identify types of preference data likely used in a system

Preference and Ratings

- We want to know: what do users like?

  –Or: what goes together?
- We can observe

  –What users tell us (ratings)

  –What users do (actions)
- These are noisy measurements of preference

Preference Model:

- Explicit
  - Rating
  - Review
  - Vote
- Implicit
  - Click
  - Purchase
  - Follow

Explicit Ratings

- Just ask the user what they think!

Ways to collect Explicit ratings:

- Start ratings
- Thumbs up and likes
- Pairwise preference
- Continuous scale
- Hybrid (e.g. 1-100 + never again)
- Temporary (e.g. Pandora 30-day suspend)

When are ratings provided?

- Consumption — during or immediately after experiencing the item
- Memory — some time after experience

  - Rating given after month/Year can be faulty as experience has been changed by memory

- Expectation — the item has not been experienced

  - Sometimes when we need ratings for Cars, Houses thennot many people have multiple houses. So without having experience we give them features and ask them what are their ratings. Like are the features complete your expectations.

Difficulties with Ratings

- Are ratings reliable and accurate?
- Do user preferences change?
- What does a rating mean?

Implicit Data

- Data collected from user actions
- Key difference: user action is for some other purpose, not expressing preference
- Their actions say a lot!

Reading Time

- Early implicit data: how long did user read?
- Listening and watching

  –IMMS

  –Video services

Binary actions

- Click on link (ad, result, cross-reference)
- Don't click on link
- Purchase
- Follow/Friend

Subtleties and Difficulties

- What does the action mean?
  - –Purchase: they might still hate it
  - –Don't click: expect bad, or didn't see
- How to scale/represent actions?
- Lots of opportunity to be creepy(use data in wrong ways)
  - –Education may help
  - –So can respecting privacy

Conclusion

- Recommenders mine what users say and what they do to learn preferences
- Ratings provide explicit expressions of preference
- Implicit data benefits from greater volume

Video 1.4.1 : Predictions and Recommendations

Learning Objectives

- To understand the ways in which recommender output can be used
- To understand the distinction between predictions and recommendations
- To understand the distinction between organic and explicit presentation
- To review examples and understand which presentation makes most sense in different applications.

Predictions

- Estimates of how much you'll like an item
  – Often scaled to match some rating scale
  – Often tied to search or browsing for specific products

Recommendations

- Recommendations are suggestions for items you might like (or might fit what you're doing)
  – Often presented in the form of "top-n lists"
  – Also sometimes just placed in front of you

Prediction and Recommendation

- Often, the two come together
- Predictions:
  – Pro: helps quantify item
  – Con: provides something falsifiable
- Recommendations
  – Pro: provides good choices as a default
  – Con: if perceived as top-n, can result in failure toexplore (if top few seem poor)

Another dimension to consider

• How explicit is the prediction or recommendation (vs. Organic)?

  - It means what words are used to reecommed. Are we saying "We are sure you will like this" and end up user disliking it and thus detoriating user experience or we are saying "Other users have also liked these items, we think you may also like" where user may or may not like any of the item but still we said "based on other users" and not "we know you very well". So organic means how well we present.

1. Historical note: we paid for it, we'll let you know
2. Today: balance between explicit prediction (falsifiable) and coarser granularity (you might like this!)
3. Today: balance between theses are the best (top-n) and softer presentation (here are some that might be interesting)

You should now understand

- Difference between prediction and recommendation
- Range of explicit to organic for both predictions and recommendations
- Advantages and disadvantages in both dimensions.

Video 1.5.1 : Taxonomy of Recommenders I

Learning Objectives

- To understand the different types of recommender systems
    - – A framework for analyzing recommender systems in general
    - – A specific overview of different recommendation algorithms
- To acquire a roadmap for the rest of the course, based on the algorithms studied

Analytical Framework

So here these 8 Dimensions of Analysis are the points that we must consider before going for making a recommendations system. These are very important points. We will cover each one of them in details.

- ***Dimensions of Analysis***

  – Domain

  – Purpose

  – Recommendation Context

  – Whose Opinions

  – Personalization Level

  – Privacy and Trustworthiness

  – Interfaces

  – Recommendation Algorithms

1. Domains of Recommendation

- Content to Commerce and Beyond
  - – News, information, “text”
  - – Products, vendors, bundles(if you are taking X item we can give offer on X+Y+Z)
  - – Matchmaking (other people)
  - – Sequences (e.g., music playlists)
- One particularly interesting property (When to recommend new and when old items)
  - – New items (e.g., movies, books, …)
  - – Re-recommend old ones (e.g., groceries, music)

2. Purposes of Recommendation

- The recommendations themselves
  - – Sales
  - – Information
- Education of user/customer (If you know A,B,C technology/command/concept then it will be benifitial for you to know X,Y,Z also)
- Build a community of users/customers around products or content

3. Recommendation Context

- What is the User doing at the time of recommendation?
  - Shopping (If shooping online than it could be good time to show some suggestions similar to the product he is using/searching)
  - Listening to Music (If I am listening to music I would never like a suggestion to pop up and break the music. Instead what I would like is that the song automatically plays)
  - Hanging out with other people (Maybe that time recommendation that the whole group like would be better)
- How does the context constrain the recommender?
  - Groups, automatic consumption (vs. suggestion), level of attention, level of interruption?

4. Whose Opinion?

- Experts
- Ordinary folks
- People like you

5. Personalization Level

- Generic / Non-Personalized
  - Everyone receives same recommendations
- Demographic
  - Matches a target group
- Ephemeral
  - Matches current activity [The way Amazons says: The users who buys this item also bought these items. Like based on your current activity recommending you. Not focusing on what you bought in past but focusing on what you are buying now]
- Persistent
  - Matches long-term interests

6. Privacy and Trustworthiness

- Who knows what about me?
  - – Personal information revealed
  - – Identity
  - – Deniability of preferences
- Is the recommendation honest?
  - – Biases built-in by operator
    - "business rules"
  - – Vulnerability to external manipulation
  - – Transparency of "recommenders"; Reputation

7. Interfaces

• Types of Output

– Predictions

– Recommendations

– Filtering

– Organic vs. explicit presentation

- Agent/Discussion Interface

• Types of Input

– Explicit

– Implicit

8. Recommendation Algorithms

a) Non-Personalized Summary Statistics

b) Content-Based Filtering

– Information Filtering

– Knowledge-Based

c) Collaborative Filtering

– User-User

– Item-Item

– Dimensionality Reduction

d) Others

– Critique / Interview Based Recommendations

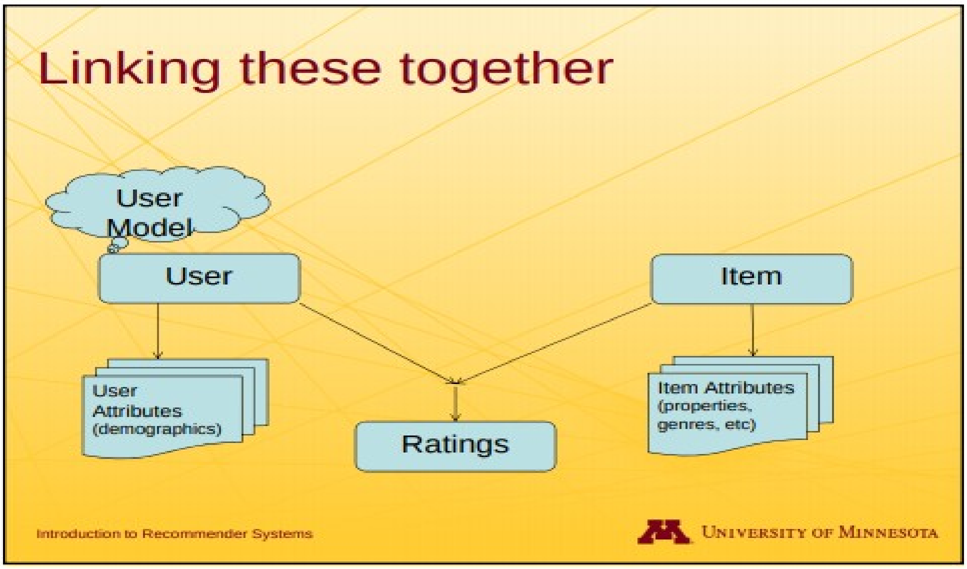
– Hybrid Techniques

Video 1.6.1 : Taxonomy of Recommenders II

Now we will look briefly at all the algorithms that we saw in the last page.

From the Abstract to the Specific

- Basic Model
  - Users
  - Items
  - Ratings
  - (Community)



The cloud representing the User Model is matrix of user and item-characteristics matrix. For example, a movie characteristics can be: Genre, director, actor etc. So for each user we will have a vector describing his different values for all characteristics. For example, based on all the movies that the user have watched we will calculate the action(genre) likeliness, like how much likeliness for action, or for a particular actor. Then based on this vector we will calculate a similarity with a movie and recommend. This is also called Content based recommendation. We will read about this later.

a) Non-Personalized Summary Stats

- External Community Data
  - Best-seller; Most popular; Trending Hot
- Summary of Community Ratings
  - Best-liked
- Examples
  - Zagat restaurant ratings
  - Billboard music rankings
  - TripAdvisor hotel ratings

b) Content-Based Filtering

- User Ratings x Item Attributes => Model [*This is what i mentioned in last page*]
- Model applied to new items via attributes [*Because of this we can also recommend an item that havent been bought or read by anone. We will do it based on attributes of that item*]
- Alternative: knowledge-based *[Basically what we do is provide our preferences, like types of news feeds that i want, or a particulat type of thing that I dislike. So it is more like highly personilized things.*]
  - Item attributes form model of item space
    - Users navigate/browse that space
- Examples
  - Personalized news feeds
  - Artist or Genre music feeds

c) Collaborative Filtering Techniques

- User-user
  - Select neighborhood of similar-taste people
- Variant: select people you know/trust
  - Use their opinions
- Item-item

    - Pre-compute similarity among items via ratings
    - Use own ratings to triangulate for recommendations
- Dimensionality reduction [The matrix factorization method]
    - Intuition: taste yields a lower-dimensionality matrix
    - Compress and use a taste representation

d) Other Approaches

- Interactive recommenders
    - Critique-based, dialog-based
- Hybrids of various techniques

Note on Evaluation

- To properly understand relative merits of each approach, we will spend significant time on evaluation
    - Accuracy of predictions
    - Usefulness of recommendations
- Correctness
- Non-obviousness
- Diversity
    - Computational performance

Video 1.8.1 : Recommender Systems: Past, Present, and Future

Before "Recommender Systems"

- Manual Personalization
- Cross   Sales and Early Product Associations
- Product Search

The Early Days

- 1992-1996
    - Research: GroupLens, RINGO, Video Recommender, and more
    - Industry Developments: Amazon and Beyond
    - Early Commercialization: Agents, Inc., Net Perceptions, and many more …
    - Mix of Business Excitement and Altruistic Dreams

The Tech Bubble and the Burst

- Recommendation as key technology …
- And then, Recommendation in Context

Wave Two: The Netflix Prize

- New Excitement in Recommender Systems
  - Netflix $1,000,000 prize
  - Recommendation as an application area for data mining, machine learning
  - Rapid Growth in the Field
- New Techniques
  - Algorithm Stacking
  - New Matrix Factorization Techniques

Mature Realizations

- Prediction and Basic Top N Algorithms Limited
  - Magic barriers
  - Value of Recommendation
  - Context, Content, much more …

State of the Field Today

- Algorithms Well Known
- Effective Recommendation Still a Craft
  - Exploring Data
  - Understanding Usage Cases and Value Proposition
- Still Largely Focused on Business Applications
  - Immense Creativity
  - Dream of Consumer Owned not Realized

Looking Forward

- Many Hard Problems Unsolved
  - Temporal Recommendation
  - Recommendations for Education
  - Low Frequency, High Stakes Recommendations
- Recognized Specialty that Brings Together
  - Machine Learning / Data Mining
  - Business / Marketing

- Human   Computer Interaction / Understanding
- Consumers

Promising Directions

- Context, Context, Context …
- Sequences of all types
  - Music
  - Education
- Lifetime Value
  - Includes exploit vs. explore