Week 2

Non-Personalized and Stereotype-Based Recommenders

Video 2.1.1: Introducing Nonpersonalized and Stereotyped Recommendation

Why Non-Personalized?

• New users – we know little about them (cold start)

• Simple but beneficial

• Online communities around common displays (e.g. Reddit, Slashdot)

• Applications & media where personalization is impossible

It can be remarkably effective.

Intro: Recommendation in Print

Long tradition of recommendation in print

- Book, movie, and music reviews
- ‘Goings On About Town’ in The New Yorker
- Michelin restaurant guides

Each of these is editorially selected

Some Example [1-4]:

1. Print Recommendation

- The Negro Motorist Green-Book:
- Listed accommodations friendly to AfricanAmerican travelers in segregation-era U.S.

2. Aggregate Opinion: Zagat Survey

Most print recommendation is editorial

Zagat Survey aggregated opinions about restaurants from individual reviewers

- Produced 30-point aggregate ratings on multiple dimensions
- Textual review compiled from individual reviewer reports

3. Aggregated Behavior: Billboard

Billboard Top 200

- Derived from sales and radio play
- Albums and songs rise based on last week's activity
- Computed over nationwide data

4. Many Other Examples

- E-commerce rating & review summaries
- Box office charts
- 'Popular Now' on any news site

Weak Personalization

Sometimes we know a little about a user

- Zip code or location
- Age, gender, nationality, ethnicity

This can be used for first-pass 'stereotyped' personalization

Product associations allow recommendations based on current page/item/context

In This Module

- Summary Statistics
    - Computing and displaying
    - Using for non-personalized recommendation
- Lightly Personalized Recommendation
    - Demographics and Stereotypes
- Identifying Related Items

Learning Objectives

- Understand the value and drawbacks of nonpersonalized recommendations
- Be able to compute nonpersonalized and weakly personalized recommendations using
    - Aggregated preferences

- Product associations
- User demographics

• Design basic user experiences around simple aggregation and recommendation algorithms

Video 2.2.1: Summary Statistics I

Computing The Scores

• What should they mean?

 – Popularity

 – Average Rating

 – Probability of You Liking

• How to compute?

 – Frequency

 – Average

 – More Complicated

Secrets Revealed!

• The “secret” formula for Zagat (Restaurant recommendation site)

 Rating = {0, 1, 2, 3}

 Score = round (MEAN(ratings) * 10)

  – OK, maybe not so secret – but effective!

Breaking it Down

• Popularity is an Important Metric

• Averages Can be Misleading

 – Can adjust by summing % who like

 – Can adjust by normalizing user ratings

• normalization addresses different rating scales

 – May want to consider credibility of individual raters (history of ratings)

• More data is better … to a point

 – Average, Count, Distribution

What’s missing here?

• Who you are:

- If I’m looking for popular new songs, I might not be looking for songs popular among 15 year-old girls

• Your context:

- If I’m ordering an ice-cream sundae and want a recommendation for a sauce, do I want to hear that ketchup is the most popular sauce?

Some take-away lessons

• Non-personalized popularity statistics or averages can be effective in the right application

  – Need to understand relationship between average and user need; correct average

• In many cases it can be best to show count, average, and distribution together

• For ranking, one alternative to average is the percentage who score above a threshold

  – Or below!

• Personalization would address many limitations!

Video 2.2.1: Summary Statistics II

Introduction

- Last 2 lectures:
  - how to collect data
  - what we present to users
- This lecture: how to do it
  - what predictions to show
  - how to rank

Learning Objectives

- Understand several ways of computing and displaying predictions
- Understand how to rank items with sparse, time-shifting data
- Understand several points in the design space for prediction and recommendation, and some of their tradeoffs

Overview

1. Example
2. Displaying Aggregate Preferences (predict)
3. Ranking Items (recommend)

1. Example

- Social news aggregator
- Non-personalized news recommender
- Users vote on items to determine top item

2. Displaying Aggregate Preferences (predict)

Simple Display Approaches

- Average rating / upvote proportion
- Net upvotes / # of likes
- % >= 4 stars (‘positive’)
- Full distribution

Goal of Display

- To help users decide to buy/read/view the item.

Simple Display Approaches(again)

- Average rating / upvote proportion
  - Of people who vote, do they like it?
  - Doesn't show popularity
- Net upvotes / # of likes
  - Shows popularity
  - No controversy
- % >= 4 stars (‘positive’)
- Full distribution
  - Complicated

3. Ranking Items (recommend)

Ranking

- What do you put at the top of Reddit?
- What is at the top of the e-Bay search list?
- You don't have to rank by prediction

Why not rank by score?

- Too little data (one 5-star rating)
- Score may be multivariate (histogram)
- Domain or business considerations
  - Item is old
  - Item is ‘unfavored’

Ranking Considerations

- Confidence
  - How confident are we that this item is good?
- Risk tolerance
  - High-risk, high-reward
  - Conservative recommendation
- Domain and business considerations
  - Age
  - System goals

Damped means

- Problem: low confidence w/ few ratings
- Solution: assume that, without evidence, everything is average
- Ratings are evidence of non-averageness k controls strength of evidence required

$$\frac{\sum_u r_{ui} + k\mu}{n + k}$$

Confidence Intervals

- From the reading: lower bound of statistical confidence interval (95%)
- Choice of bound affects risk/confidence
  - Lower bound is conservative: be sure it's good
  - Upper bound is risky: there's a chance of amazing
- Reddit uses Wilson interval (for binomial) to rank comments

Domain Consideration: Time

- Reddit: old stories aren't interesting
  - even if they have many upvotes!
- eBay: items have short lifetimes

Scoring news stories

Hacker News

$$\frac{(U - D - 1)^{\alpha}}{(t_{\text{now}} - t_{\text{post}})^{\gamma}} \times P$$

- Net upvotes, polynomially decayed by age
- Old items scored mostly by vote
- Multiplied by item penalty terms
    - incorporate community goals into score

Reddit algorithm (c. 2010)

$$log_{10}max(1, |U - D|) + \frac{sign(U - D)t_{post}}{45000}$$

- Log term applied to votes
    - decrease marginal value of later votes
- Time is seconds since Reddit epoch
- Buries items with negative votes
- Time vs. vote impact independent of age
- Scores news items, not comments

Ranking Wrap-Up

- There are some theoretically grounded approaches (confidence interval, damping)
- Many sites use ad-hoc methods
- Most formulas have constants, will be highly service-dependent
- Can manipulate for ‘good’ or ‘evil’
- Build based on domain properties, goals

Predict with sophisticated score?

- Theoretically a fine thing to do
- Be careful with transparency/scrutability
    - If you say ‘average rating’ for damped mean, and show ratings, users may be confused
    - Most important case (low ratings) also easiest to hand-verify

Conclusion

- Sparsity, inconsistency, temporal concerns make data messy
- Simple scoring doesn't necessarily match the domain or business
- There are good ways to deal with this (decay, time, penalties, damping)
- We'll see more normalizations later

Video 2.3.1: Demographics and Related Approaches

Demographics: What and Why?

- Motivation
  - – Popularity may not represent my tastes
  - – I may be part of an identifiable cohort with different tastes
    - Age
    - Gender
    - Race/Ethnicity
    - Socio Economic Status
    - Location
  - – Including non demographics that may be predictive

OK, But How?

- Start by identifying available demographics (and correlates)
  - – Many will require processing or bucketing
    - Age is often divided into groups
    - Postal codes can be transformed into socio economic status, urban/rural, dominant ethnicity, etc.
- Then explore where your data correlates with demographics
  - – Scatterplots, correlations, ….

If You Find Relevant Demographics

- Step 1: Break down summary statistics by demographic
  - E.g., most popular item for women, for men
  - Maybe even factorial (most popular item for men., age:45-60)
- Step 2: Consider a multiple regression model
  - Predict items based on demographic statistics
    - Linear regression for multi valued (e.g., rating) data
    - Logistic regression for 0/1 (e.g. purchase) data

Important!

- You need defaults for unknown demographics
  - May simply be overall preferences
  - May reflect expected demographics of newcomers
  - May be modeled separately
- If demographics are useful, getting data on user is key
  - Various sources of data, from advertising networks to loyalty club sign ups and surveys
  - In some cases, demographics can be “predicted” from data.

The Power and Limits of Demographics

- In many cases, demographics work because products or content is created to reach them
  - Television programs
  - Magazine articles and advertisements
  - Personal products
- Or products simply naturally appeal to different groups
- Of course, demographics fail miserably for people whose tastes don’t match their demographics!

Video 2.4.1: Product Association Recommenders

Ephemeral, Contextual Personalization

- Personalized to "what you are doing"
  - Your current navigation is a reflection of your current interest
  - Does not reflect any long-term knowledge of your preferences

How to Compute?

- Version I: Manual Cross-Sell Tables
  - Generated by marketers to reflect perceived cross-sales or up-sales
- Version II: Data Mining Associations
  - What are we looking for?
    - Mostly likely to be bought in this context?
    - Most "extra-likely" to be bought in this context?

Start Simple

• Start simple: percentage of X-buyers who also bought Y; divide the counts …

$$\frac{X \wedge Y}{X}$$

- Intuitively right, but is it useful? What if X is anchovy paste and Y is bananas??
- Challenge – doesn't compensate for overall

Bayes' Law

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Now, we look at how much more likely Y is than it was before:

$$\frac{P(Y|X)}{P(Y)}$$

Other solutions …

- Association rule mining brings us the lift metric:

P(X AND Y) / P(X) * P(Y)

    - This looks at non-directional assocation
    - More generally association rules look at baskets of products, not just individuals

Association Rules in Practice

1. The Beer and Diapers Story
2. Not Just Products … Link Associations
3. Sports Cars and Leather Driving Gloves
4. Are All Recommendations Worth Making
    1. Business Rules

Some take-away lessons

- Product association recommenders use present context to provide more relevant recommendations
    - Current products; path; links
- Can compute product associations from prior transaction history
    - Balance between high-probability and increased probability
- Such recommendations can be targeted for additional or replacement purchases