Recommender Systems: Overview and Package rectools

Norm Matloff
Dept. of Computer Science
University of California at Davis

December 13, 2016

Recommender Systems: Overview and Package rectools
Norm Matloff Dept. of Computer Science University of California at Davis

Norm Matloff
Dept. of Computer Science
University of California at Davis

What Are Recommender Systems?

Norm Matloff
Dept. of Computer Science
University of California at Davis

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

Norm Matloff
Dept. of Computer Science
University of California at Davis

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

- Input data format:

```
userID    filmID    rating
```

Norm Matloff
Dept. of Computer Science
University of California at Davis

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

- Input data format:

```
userID    filmID    rating
```

- Implies matrix $Y_{ij} =$, rating user i gives to film j.

Norm Matloff
Dept. of Computer Science
University of California at Davis

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

- Input data format:

```
userID   filmID   rating
```

- Implies matrix $Y_{ij} =$, rating user i gives to film j.
- Most of the Y_{ij} are **NAs**; most users haven't rated most films.

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

- Input data format:

```
userID    filmID    rating
```

- Implies matrix $Y_{ij} =$, rating user i gives to film j.
- Most of the Y_{ij} are **NAs**; most users haven't rated most films.
- Want to predict the unknown Y_{ij}.

Norm Matloff
Dept. of Computer Science
University of California at Davis

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

- Input data format:

```
userID    filmID    rating
```

- Implies matrix $Y_{ij} =$, rating user i gives to film j.
- Most of the Y_{ij} are **NAs**; most users haven't rated most films.
- Want to predict the unknown Y_{ij}.
- Also called the *matrix completion problem*, but we also want to predict for new users as they come in.

What Are Recommender Systems?

Various forms, but here is a common one, say for data on movie ratings:

- Input data format:

```
userID    filmID    rating
```

- Implies matrix $Y_{ij} =$, rating user i gives to film j.
- Most of the Y_{ij} are **NAs**; most users haven't rated most films.
- Want to predict the unknown Y_{ij}.
- Also called the *matrix completion problem*, but we also want to predict for new users as they come in.
- Above referred to as *collaborative filtering*. Various other approaches are popular as well.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Applications

Norm Matloff
Dept. of Computer Science
University of California at Davis

Applications

Why the interest?

Applications

Why the interest?

- Amazon book RS; no bricks and mortar stores to browse in these days.

Applications

Why the interest?

- Amazon book RS; no bricks and mortar stores to browse in these days.
- University of Minnesota class finder: A student considering taking Class Stat 101 would have estimates of (a) how well she would like the class and (b) what grade she would get.

Applications

Why the interest?

- Amazon book RS; no bricks and mortar stores to browse in these days.
- University of Minnesota class finder: A student considering taking Class Stat 101 would have estimates of (a) how well she would like the class and (b) what grade she would get.
- Prediction of adverse reactions to prescription drugs.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Applications

Why the interest?

- Amazon book RS; no bricks and mortar stores to browse in these days.
- University of Minnesota class finder: A student considering taking Class Stat 101 would have estimates of (a) how well she would like the class and (b) what grade she would get.
- Prediction of adverse reactions to prescription drugs.
- Etc.

Statistical Issues (Part I)

Norm Matloff
Dept. of Computer Science
University of California at Davis

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.
- Big gap between ML and stat people in general.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.
- Big gap between ML and stat people in general.
  - Stat: Sample from a population. Properties of estimators wrt that population. Derive MSE, asymptotics.

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.
- Big gap between ML and stat people in general.
  - Stat: Sample from a population. Properties of estimators wrt that population. Derive MSE, asymptotics.
  - ML: There is only the given data, no population. Derive inequalities related to various data quantities.

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.
- Big gap between ML and stat people in general.
    - Stat: Sample from a population. Properties of estimators wrt that population. Derive MSE, asymptotics.
    - ML: There is only the given data, no population. Derive inequalities related to various data quantities.
    - Discussed in (Breiman, 2001) and (Matloff, 2014).

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.
- Big gap between ML and stat people in general.
    - Stat: Sample from a population. Properties of estimators wrt that population. Derive MSE, asymptotics.
    - ML: There is only the given data, no population. Derive inequalities related to various data quantities.
    - Discussed in (Breiman, 2001) and (Matloff, 2014).
- ML people in denial.

Statistical Issues (Part I)

- RS research done almost entirely in machine learning (ML) community.
- Big gap between ML and stat people in general.
  - Stat: Sample from a population. Properties of estimators wrt that population. Derive MSE, asymptotics.
  - ML: There is only the given data, no population. Derive inequalities related to various data quantities.
  - Discussed in (Breiman, 2001) and (Matloff, 2014).
- ML people in denial. They talk about predicting new cases, overfitting etc. — all implicitly assuming sampling from a population.

Norm Matloff
Dept. of Computer Science
University of California at Davis

3 Broad Categories of Methods

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:**

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.*

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.* To predict how I might rate Movie X, average the ratings of the k people in the dataset who are most similar to me and who have rated Movie X.

Norm Matloff
Dept. of Computer Science
University of California at Davis

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.* To predict how I might rate Movie X, average the ratings of the k people in the dataset who are most similar to me and who have rated Movie X.
- **Matrix Factorization (MF) Methods:**

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.* To predict how I might rate Movie X, average the ratings of the k people in the dataset who are most similar to me and who have rated Movie X.
- **Matrix Factorization (MF) Methods:** *Currently the most popular method.*

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.* To predict how I might rate Movie X, average the ratings of the k people in the dataset who are most similar to me and who have rated Movie X.
- **Matrix Factorization (MF) Methods:** *Currently the most popular method.* Find known factors U, V such that

$$M = (Y_{ij}) \approx UV$$

Norm Matloff
Dept. of Computer Science
University of California at Davis

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.* To predict how I might rate Movie X, average the ratings of the k people in the dataset who are most similar to me and who have rated Movie X.
- **Matrix Factorization (MF) Methods:** *Currently the most popular method.* Find known factors U, V such that

$$M = (Y_{ij}) \approx UV$$

- **Random effects (RE) models:** *I believe these' are little known in the ML RS community.* The basic model is

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

with α_i and β_j being random and latent effects, mean 0.

Norm Matloff
Dept. of Computer Science
University of California at Davis

3 Broad Categories of Methods

- **k-Nearest Neighbor (kNN) Methods:** *One of the earliest RS methods.* To predict how I might rate Movie X, average the ratings of the k people in the dataset who are most similar to me and who have rated Movie X.
- **Matrix Factorization (MF) Methods:** *Currently the most popular method.* Find known factors U, V such that

$$M = (Y_{ij}) \approx UV$$

- **Random effects (RE) models:** *I believe these' are little known in the ML RS community.* The basic model is

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

with α_i and β_j being random and latent effects, mean 0. Then apply MLE (e.g. **lme4** package in R) or, better, Method of Moments (MM).

Norm Matloff
Dept. of Computer Science
University of California at Davis

Predicted Values

Predicted Values

- **kNN:** $\widehat{Y}$ = average of the neighbors' ratings of item j.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Predicted Values

- **kNN:** $\widehat{Y}$ = average of the neighbors' ratings of item j.
- **MF:** $\widehat{Y}_{ij} = (UV)_{ij}$

Norm Matloff
Dept. of Computer Science
University of California at Davis

Predicted Values

- **kNN:** $\widehat{Y}$ = average of the neighbors' ratings of item j.
- **MF:** $\widehat{Y}_{ij} = (UV)_{ij}$
- **RE:** $\widehat{Y}_{ij} = \widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j$

Norm Matloff
Dept. of Computer Science
University of California at Davis

Some Proposals for Novel Variations

Norm Matloff
Dept. of Computer Science
University of California at Davis

Some Proposals for Novel Variations

Hybrid kNN-MF Model

- Find U, V for $M \approx UV$.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Some Proposals for Novel Variations

Hybrid kNN-MF Model

- Find U, V for $M \approx UV$.
- The rows of V are (or should be) interpreted as forming a kind of basis, consisting of synthetic reprentative users.

Some Proposals for Novel Variations

Hybrid kNN-MF Model

- Find U, V for $M \approx UV$.
- The rows of V are (or should be) interpreted as forming a kind of basis, consisting of synthetic reprentative users.
- For each new case C to predict for item j, find the closest row of V to (the known elements of) C, denoted by R.

Some Proposals for Novel Variations

Hybrid kNN-MF Model

- Find U, V for $M \approx UV$.
- The rows of V are (or should be) interpreted as forming a kind of basis, consisting of synthetic reprentative users.
- For each new case C to predict for item j, find the closest row of V to (the known elements of) C, denoted by R. Predict the rating to be element j of R_j.
- Should be much, much faster, and maybe more accurate.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model for Binary R

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model for Binary R

- The case Y = 0,1.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model for Binary R

- The case Y = 0,1.
- Use of kNN straightforward.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model for Binary R

- The case Y = 0,1.
- Use of kNN straightforward.
- MF not designed for binary Y, but still could use them; they are only heuristics anyway.

Multiplicative Model for Binary R

- The case Y = 0,1.
- Use of kNN straightforward.
- MF not designed for binary Y, but still could use them; they are only heuristics anyway.
- For RE, **lme4** package offers GLM analysis for binary Y, but MLE is too slow, especially with large data.

Multiplicative Model for Binary R

- The case Y = 0,1.
- Use of kNN straightforward.
- MF not designed for binary Y, but still could use them; they are only heuristics anyway.
- For RE, **lme4** package offers GLM analysis for binary Y, but MLE is too slow, especially with large data.
- Desirable: Alternative using MM.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model (cont'd.)

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model (cont'd.)

- Instead of independent *additive* model, $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, have an independent *multiplicative* model,

$$P(Y_{ij} = 1) = \nu \alpha_i \beta_j$$

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model (cont'd.)

- Instead of independent *additive* model, $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, have an independent *multiplicative* model,

$$P(Y_{ij} = 1) = \nu \alpha_i \beta_j$$

- For identifiability, set $E\alpha_i = E\beta_j = 1$.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Multiplicative Model (cont'd.)

- Instead of independent *additive* model, $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, have an independent *multiplicative* model,

$$P(Y_{ij} = 1) = \nu \alpha_i \beta_j$$

- For identifiability, set $E\alpha_i = E\beta_j = 1$.
- Apply MM, get very simple closed-form expressions.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Our rectools Package

Our rectools Package

- At *https://github.com/Pooja-Rajkumar/rectools*, soon to be submitted to CRAN.

Our rectools Package

- At *https://github.com/Pooja-Rajkumar/rectools*, soon to be submitted to CRAN.
- kNN, MF, RE (MLE/MM)

Our rectools Package

- At *https://github.com/Pooja-Rajkumar/rectools*, soon to be submitted to CRAN.
- kNN, MF, RE (MLE/MM)
- Implements (or will implement) the above proposed methods and others.

Our rectools Package

- At *https://github.com/Pooja-Rajkumar/rectools*, soon to be submitted to CRAN.
- kNN, MF, RE (MLE/MM)
- Implements (or will implement) the above proposed methods and others.
- Some parallel computational capability, using Software Alchemy (Matloff, 2016).

Statistical Issues (Part II)

Norm Matloff
Dept. of Computer Science
University of California at Davis

Statistical Issues (Part II)

- After fitting a model, a statistician's natural inclination is to make plots, to assess the fit.

Statistical Issues (Part II)

- After fitting a model, a statistician's natural inclination is to make plots, to assess the fit.
- But to my knowledge, no RS researcher/package has done this.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Statistical Issues (Part II)

- After fitting a model, a statistician's natural inclination is to make plots, to assess the fit.
- But to my knowledge, no RS researcher/package has done this.
- So, **rectools** does various plots.

Covariate Data

Norm Matloff
Dept. of Computer Science
University of California at Davis

Covariate Data

- A few months ago, out of the blue, Netflix made an announcement about their RS:

Norm Matloff
Dept. of Computer Science
University of California at Davis

Covariate Data

- A few months ago, out of the blue, Netflix made an announcement about their RS:
  - They found that demographics (age, gender, income etc.) do NOT enhance predictive ability. (Already incorporated in the ratings.)

Covariate Data

- A few months ago, out of the blue, Netflix made an announcement about their RS:
  - They found that demographics (age, gender, income etc.) do NOT enhance predictive ability. (Already incorporated in the ratings.)
  - They stated that they found other data that DOES help — but said they won't tell us what kind of data!

Norm Matloff
Dept. of Computer Science
University of California at Davis

Covariate Data

- A few months ago, out of the blue, Netflix made an announcement about their RS:
  - They found that demographics (age, gender, income etc.) do NOT enhance predictive ability. (Already incorporated in the ratings.)
  - They stated that they found other data that DOES help — but said they won't tell us what kind of data! (So why did they bother with their announcement?)

Norm Matloff
Dept. of Computer Science
University of California at Davis

Covariate Data

- A few months ago, out of the blue, Netflix made an announcement about their RS:
  - They found that demographics (age, gender, income etc.) do NOT enhance predictive ability. (Already incorporated in the ratings.)
  - They stated that they found other data that DOES help — but said they won't tell us what kind of data! (So why did they bother with their announcement?)
- But covariate data should be helpful in some settings, e.g. kNN with a new case that has covariate data but not much of a ratings history.

Covariate Data

- A few months ago, out of the blue, Netflix made an announcement about their RS:
  - They found that demographics (age, gender, income etc.) do NOT enhance predictive ability. (Already incorporated in the ratings.)
  - They stated that they found other data that DOES help — but said they won't tell us what kind of data! (So why did they bother with their announcement?)
- But covariate data should be helpful in some settings, e.g. kNN with a new case that has covariate data but not much of a ratings history.
- The **rectools** package does allow covariates in most of its functions.

Norm Matloff
Dept. of Computer Science
University of California at Davis

Example: Czech Dating Site

Norm Matloff
Dept. of Computer Science
University of California at Davis

Example: Czech Dating Site

Ratings 1:10.

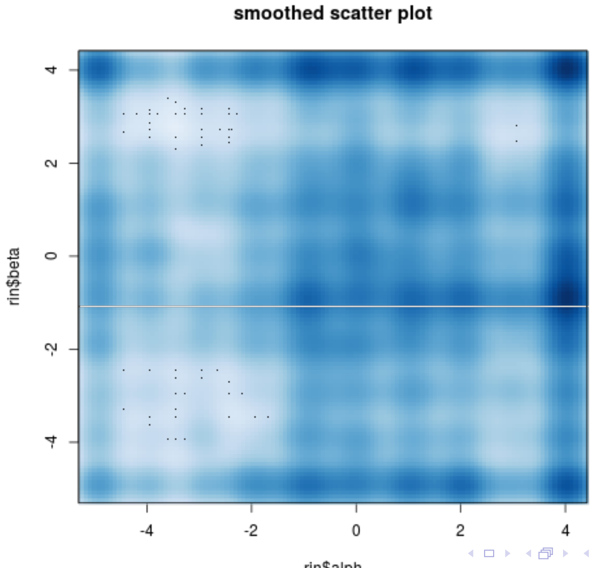
```
> dat <- read.csv('ratings.dat',header=FALSE)
> idxs <- sample(1:nrow(dat),100000)
> dat100k <- dat[idxs,]
> mmout <- trainMM(dat100k)
> dat100k[1,]
              V1     V2 V3
14857272 115894 14800  4
# how would user 115894 like member 2?
> z <- dat100k[1,]
> z$V2 <- 2
> predict(mmout,z)
[1] 8.45566
```

Norm Matloff
Dept. of Computer Science
University of California at Davis

Plotting Example

Plotting Example

MLE model assumes the α, β independent, normal.



Norm Matloff
Dept. of Computer Science
University of California at Davis

Plot Example, cont'd.

Plot Example, cont'd.