ASSESSING REVIEW RECOMMENDATION TECHNIQUES UNDER A RANKING PERSPECTIVE

Luciana B. Maroun, Mirella M. Moro, Jussara M. Almeida, Ana Paula C. Silva

{lubm,mirella,jussara,ana.coutosilva}@dcc.ufmg.br

27th ACM Conference on Hypertext and Social Media
WHY RECOMMEND REVIEWS?
88% of users have read online reviews

39% read reviews regularly

ON THE ONE HAND

Top of the Rock Observation Deck

- 47,375 Reviews
- #4 of 1,069 things to do in New York City

Customer Reviews

- 85,365
- 4.2 out of 5 stars

The Hunger Games (The Hunger Games #1)

by Suzanne Collins

- 4.37
- Winning will make you famous.
- Losing means certain death.

146,182 Reviews
ON THE OTHER HAND

67% of users read up to 6 reviews

REVIEW WRITING

- Author
- Rating
- Text
- Product
HELPFULNESS VOTING

helpfulness vote

reader

voter
Only 10% reviews have at least 10 votes in Amazon (all reviews from 1995-2013)

average helpfulness

rich-get-richer effect

COMMON SOLUTION
ANOTHER COMMON SOLUTION

global helpfulness

text
author
35 of 70 people found this helpful

**Great boat has a fatal flaw**

By P. G. Farrow on 29 July 2011

Having bought this boat to go Kayaking in France we found a fatal flaw in the design that meant the boat sat there useless for the majority of our holiday.

The boston valves used to inflate the boat detach very easily and get lost. When you unscrew the valve inner to release the pressure, to pack the boat away the valves are not secured to the boat.
MODELING AS A RECOMMENDATION PROBLEM

diagram:

- personalized helpfulness
- user
- content
- review-reader
- author
- text

diagram:

- user
- content
- review-reader
- author
- text
EXISTING SOLUTIONS

Two specialized algorithms: BETF and CAP

- Not compared against each other
- No significance report
- Only regression perspective
OUR EXPERIMENTS

Comparison of a rich set of personalized techniques
(with public source code\(^1\))

Analysis of best parameter values

Experiments with statistical tests

Evaluation as a ranking task

\(^1\)github.com/lucianamaroun/review_recommendation
2 PROBLEM STATEMENT
THE REVIEW RECOMMENDATION PROBLEM

product

author

review

reader

1

2

3

relative vote
3 OVERVIEW OF TECHNIQUES
OVERVIEW OF TECHNIQUES

- Mean-based Predictors (MBP)
- Regressors (REG)
- Learning to Rank Predictors (LTR)
- General-purpose Recommender Systems (GRS)
- Review Recommender Systems (RRS)
MEAN-BASED PREDICTORS (MBP)

> Overall Mean Helpfulness (OM)
> Author’s Mean Helpfulness (AM)
> Voter’s Mean Helpfulness (VM)
REGRESSORS (REG)

- Linear Regression (LR)
- Support Vector Machine Regression (SVR)
- Gradient Boosting Regression Trees (GBRT)

\[ \gamma_1 \cdot Y_1 + \gamma_2 \cdot Y_2 + \gamma_3 \cdot Y_3 \]
LEARNING TO RANK PREDICTORS (LTR)

> RankSVM (RSVM)

> LambdaMART (LMART)
GENERAL-PURPOSE RECOMMENDER SYSTEMS (GPRS)

- Matrix Factorization (MF)
  \[ W^T X \]

- Regression-based Latent Factor Models (RLFM)
  \[ V^T Y \]
REVIEW RECOMMENDER SYSTEMS (RRS)

- UnBiased Extended Tensor Factorization (BETF)
- Context-Aware Review Rating Prediction (CAP)

Jointly: MF for ratings

$$\text{review} = W^TX$$

$$\text{reader} = V^TY$$

$$\text{author} = U^TY$$
4 UNIFIED EXPERIMENTAL DESIGN
EXPERIMENTAL DESIGN

Split

Sort

457,679

1 2 3 4 5

training
validation

test

time
EXPERIMENTAL DESIGN

- RMSE, nDCG@p
- Paired t-tests
- Holm-Bonferroni Correction
- 95% of confidence
PARAMETER TUNING

- **nDCG@5**
  - Simple: ↑ regularization, ↓ # dimensions
  - Stochastic: ↓ iterations, ↓ # samples

- **RMSE**
  - Opposite trend
5 EXPERIMENTAL EVALUATION
## COMPARISON WITHIN CLASSES

<table>
<thead>
<tr>
<th>Technique</th>
<th>nDCG@5</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MBP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OM</td>
<td>0.8808 ± 0.0265</td>
<td>0.4148 ± 0.0489</td>
</tr>
<tr>
<td>AM</td>
<td>0.8944 ± 0.0278</td>
<td>0.4394 ± 0.0624</td>
</tr>
<tr>
<td>VM</td>
<td>0.8808 ± 0.0265</td>
<td>0.4323 ± 0.0681</td>
</tr>
<tr>
<td><strong>REG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.9269 ± 0.0178</td>
<td>0.3916 ± 0.0463*</td>
</tr>
<tr>
<td>SVR</td>
<td>*<em>0.9362 ± 0.0132</em></td>
<td>*<em>0.3997 ± 0.0524</em></td>
</tr>
<tr>
<td>GBRT</td>
<td>0.9233 ± 0.0185</td>
<td>0.3621 ± 0.0354*</td>
</tr>
<tr>
<td><strong>LTR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSVM</td>
<td>0.8790 ± 0.0279</td>
<td>—</td>
</tr>
<tr>
<td>LMART</td>
<td><strong>0.8929 ± 0.0277</strong></td>
<td>—</td>
</tr>
<tr>
<td><strong>GRS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>0.8886 ± 0.0286</td>
<td>0.4370 ± 0.0621</td>
</tr>
<tr>
<td>RLFM</td>
<td><strong>0.9229 ± 0.0197</strong></td>
<td><strong>0.4167 ± 0.0235</strong></td>
</tr>
<tr>
<td><strong>RRS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BETF</td>
<td>0.8869 ± 0.0281</td>
<td>0.4161 ± 0.0509</td>
</tr>
<tr>
<td>CAP</td>
<td><strong>0.9213 ± 0.0196</strong></td>
<td>0.4354 ± 0.0462</td>
</tr>
</tbody>
</table>
COMPARISON AMONG BEST OF CLASSES

nDCG@p

p

SVR
LR
RLFM
CAP
AM
LMART
Random split
SIMPLE VS. COMPLEX MODEL

Simple models perform better

- Parameters
- Linear > Non-linear

Less prone to overfitting
OBSERVED VS. LATENT FEATURES

Review scenario is sparse

- RLFM $\succ$ MF
- CAP $\succ$ BETF
- SVR $\succ$ Rest

Traditional solutions, such as MF, do not work well
RANKING VS. REGRESSION GOAL

Fitting ranking goal: LTR

- Only 11.5% non-tied reviews
- Few interaction features
- Regression: a good workaround

Evaluating with a ranking metric

- Regression ≠ Ranking
- Ranking: related to final purpose
OPTIMAL SVR: GENERAL CASE
OPTIMAL SVR: INTEGER RESPONSES
CONCLUSION
Contributions:
1) Rich set
2) Parameter
3) Significance
4) Ranking

Solution:
1) Average
2) Global
3) Personalized
OVERALL GOOD PRACTICES

- Simplicity
- Observed features
- Chronological split
- Ranking perspective
FUTURE WORK

- New interaction features
- Ranking optimization goal
- Other datasets
- Different user types
ASSESSING REVIEW RECOMMENDATION TECHNIQUES UNDER A RANKING PERSPECTIVE

Luciana B. Maroun, Mirella M. Moro, Jussara M. Almeida, Ana Paula C. Silva

{lubm,mirella,jussara,ana.coutosilva}@dcc.ufmg.br

27th ACM Conference on Hypertext and Social Media
ADDITIONAL SLIDES
Previous Experiments

BETF
OM
VM
LR (NP)
SVR (NP)
MF

CAP
OM
AM
VM
LR (NP)
MF

LR (P)
SVR (P)
GBRT
RSVM
LMART
RLFM
CAP & BETF

- Only for regression
- No significance report
## STATISTICS OF THE DATASET

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td># Reviews</td>
<td>58,826</td>
</tr>
<tr>
<td># Users</td>
<td>8,511</td>
</tr>
<tr>
<td># Authors</td>
<td>8,065</td>
</tr>
<tr>
<td># Voters</td>
<td>1,737</td>
</tr>
<tr>
<td># Authors &amp; Voters</td>
<td>1,291</td>
</tr>
<tr>
<td># Products</td>
<td>1,908</td>
</tr>
<tr>
<td># Votes</td>
<td>457,697</td>
</tr>
<tr>
<td>Mean # Votes by Review</td>
<td>8.05</td>
</tr>
<tr>
<td>Reviews’ Time Span</td>
<td>05.31.00 - 09.25.11</td>
</tr>
</tbody>
</table>
EVALUATION METRICS

- Root Mean Squared Error (RMSE)
- Normalized Discounted Cumulative Gain at position $p$ (nDCG@$p$)
PARTICIPATORY BEHAVIOR
DISTRIBUTIONS OF EVALUATIONS

![Bar chart showing frequency of product ratings](chart1)

![Bar chart showing frequency of helpfulness votes](chart2)
CONFORMITY

> Agreement on helpfulness for an entity
> ≈ How much entity explains helpfulness
> Measured by CV

CV → Conformity
CONFORMITY
FEATURES

- Review
- Author
- Reader
- Author-Reader Similarity
- Author-Reader Connection
RMSE for all (but LTR)
### Types of users

<table>
<thead>
<tr>
<th></th>
<th>Warm-Start</th>
<th>Cold-Start</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>$0.936262 \pm 0.013216$</td>
<td>$0.927939 \pm 0.015284$</td>
</tr>
<tr>
<td>LR</td>
<td>$0.928612 \pm 0.019309$</td>
<td>$0.907157 \pm 0.029005$</td>
</tr>
<tr>
<td>RLFM</td>
<td>$0.922199 \pm 0.019730$</td>
<td>$0.912220 \pm 0.029250$</td>
</tr>
</tbody>
</table>
## IMPACT OF ENTITIES

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>nDCG@5</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.9362 ± 0.0132</td>
<td>—</td>
</tr>
<tr>
<td>All \ Review</td>
<td>0.9078 ± 0.0256</td>
<td>0.0040</td>
</tr>
<tr>
<td>All \ Author</td>
<td>0.9356 ± 0.0130</td>
<td>0.0253</td>
</tr>
<tr>
<td>All \ Voter</td>
<td>0.9117 ± 0.0433</td>
<td>0.1300</td>
</tr>
<tr>
<td>All \ Similarity</td>
<td>0.9364 ± 0.0130</td>
<td>0.0973</td>
</tr>
<tr>
<td>All \ Connection</td>
<td>0.9356 ± 0.0130</td>
<td>0.0056</td>
</tr>
<tr>
<td>All \ Personalized</td>
<td>0.9319 ± 0.0053</td>
<td>0.0271</td>
</tr>
</tbody>
</table>