LEARNING TO RANK SOCIAL UPDATE STREAMS

Liangjie Hong*, Lehigh University
Ron Bekkerman, LinkedIn Corp.
Joseph Adler, LinkedIn Corp.
Brian D. Davison, Lehigh University

* Part of this work was done when the author was on an internship at LinkedIn Corp.
OVERVIEW

- Social Update Streams
- Overview of LinkedIn
- Social Stream Ranking & Dataset
- Methods
- Experiments
- Conclusion
Task
  • Improve user engagement by re-ranking social updates

Main results
  • We demonstrate that recommender systems + preference-based learning can be used to re-rank social updates.
  • A linear model can achieve 60% of the performance of latent factor models, on average.
  • A tensor factorization model with regression on explicit features works the best.
  • The cold-start problem makes it impossible to model some kinds of interactions.
SOCIAL UPDATE STREAMS
SOCIAL UPDATE STREAMS

Problems?
SOCIAL UPDATE STREAMS

Information overload
SOCIAL UPDATE STREAMS

Information shortage
SOCIAL UPDATE STREAMS
OVERVIEW OF LINKEDIN
OVERVIEW OF LINKEDIN

- Founded in Dec. 2002, launched in May 2003
- 160M¹ users in 200 countries and territories
- Biggest social network for professionals

¹ As of March 2012
OVERVIEW OF LINKEDIN
LinkedIn Homepage
LinkedIn Homepage

- has an updated profile (Interests, Associations, Honors)
- has an updated current title: Entrepreneur Platform Builder for Social & Mobile Experience Design Strategy at Thoroughbranding
- Craig Kessler Question 1: Do you have a budget for Valentine’s Day this year? - PLZ RT #bp
OVERVIEW OF LINKEDIN
OVERVIEW OF LINKEDIN
OVERVIEW OF LINKEDIN

LinkedIn Today - Mar 17

Hiring Managers Take Their Time Filling Jobs

Online HR

Recruiters say they are having trouble finding candidates for many skilled positions, and once candidates are found, hiring managers are taking longer to pull the trigger.

LinkedIn Groups

User Communications

Connect Your Brand To Over 1 Million Sales Professionals

LinkedIn Ad

Ad

LinkedIn

Social

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn

LinkedIn
PROBLEM DEFINITION

For a given recipient and updates from his/her social connections (senders), we want to re-rank these updates to optimize user engagement.
## Dataset

<table>
<thead>
<tr>
<th>Data Summary</th>
<th>April, 2011</th>
<th>September, 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impressions</td>
<td>3M-4M</td>
<td>10M-20M</td>
</tr>
<tr>
<td>Updates</td>
<td>30M-40M</td>
<td>100M-200M</td>
</tr>
<tr>
<td>Clicked Updates</td>
<td>3M-4M</td>
<td>10M-20M</td>
</tr>
<tr>
<td>Non-clicked Updates</td>
<td>27M-36M</td>
<td>90-180M</td>
</tr>
<tr>
<td>Distinct Updates</td>
<td>10M-20M</td>
<td>20M-30M</td>
</tr>
<tr>
<td>Recipients</td>
<td>1M-2M</td>
<td>4M-5M</td>
</tr>
<tr>
<td>Producers</td>
<td>4M-5M</td>
<td>6M-7M</td>
</tr>
</tbody>
</table>

The numbers are obfuscated for commercial reason.
EVALUATION METRIC

- **Precision@k**
  - $\frac{\# \text{ of clicks in top } k \text{ positions}}{k}$

- **Average Precision (AP) for ranked list $i$**
  - $\frac{\sum_{k=1}^{m} \text{Precision@ } k \times l_k}{\# \text{ of clicks for ranked list of ranked list } i}$
    - $l_k$: position $k$ is clicked.
    - $m$: total number of positions evaluated.

- **Mean Average Precision (MAP)**
  - average AP across all ranked lists
METHODS

- Linear Models
  - Feature Model
  - Bias Model
  - Hybrid Model

- Latent Factor Models
  - Matrix Factorization
  - Tensor Factorization
  - Regression-based Tensor Factorization
Methods

- Linear Models
  - Feature Model
  - Bias Model
  - Hybrid Model

- Latent Factor Models
  - Matrix Factorization
  - Tensor Factorization
  - Regression-based Tensor Factorization

From the simplest to the most complex
METHODS

Linear Models: Feature Model

\[ f_i^{(1)} = \beta_{r(i)}^T \phi_{r(i)} + \alpha_{r(i)}^T \phi_i \]

- utilize explicit features.
- \( f_i \) represents the estimation of user’s click on update \( i \).
- \( r(i) \) is the recipient of update \( i \).
- \( \phi \) is a feature vector.
- \( \beta \) and \( \alpha \) are coefficients.
METHODS

Linear Models: Latent Bias Model

\[ f_i^{(2)} = \mu + b_i + b_{t(i)} + b_{r(i)} + b_{c(i)} + b_{s(i)} \]

- utilize categorical features.
- \( t(i) \) is the type of update \( i \).
- \( c(i) \) is the type of sender of update \( i \).
- \( s(i) \) is the sender of update \( i \).
METHODS

Linear Models: Latent Bias Model

\[ f_i^{(2)} = \mu + b_i + b_{t(i)} + b_{r(i)} + b_{c(i)} + b_{s(i)} \]

- utilize categorical features.
- \( t(i) \) is the type of update \( i \).
- \( c(i) \) is the type of sender of update \( i \).
- \( s(i) \) is the sender of update \( i \).
METHODS

Combining Feature and Bias

\[ f_i^{(3)} = f_i^{(1)} + f_i^{(2)} \]

Incorporating Temporal Effects

\[ f_i^4 = f_i^{(*)} + \zeta \times t_{\text{recency}} \]
METHODS

Learning through $L_2$-regularized logistic regression

$$l_1(y_i, f_i^{(*)}) = \log \left[ 1 + \exp(-y_i f_i^*) \right]$$
METHODS

Linear Model Summary

- Simple
- Fast
- Intuitive
METHODS

Linear Model Summary

- Simple
- Fast
- Intuitive

Does not exploit user-user, user-item interactions at all
METHODS

Latent Factor Model: Matrix Factorization

How to utilize pair-wise interactions?
METHODS

Latent Factor Model: Matrix Factorization

- user-item interaction?

<table>
<thead>
<tr>
<th></th>
<th>User 1</th>
<th>User 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>File 1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>File 2</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>File 3</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>File 4</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>File 5</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
# METHODS

Latent Factor Model: Matrix Factorization

- user-user interaction?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="User 1" /></td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td><img src="image2" alt="User 2" /></td>
<td>?</td>
<td>2</td>
</tr>
<tr>
<td><img src="image3" alt="User 3" /></td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td><img src="image4" alt="User 4" /></td>
<td>?</td>
<td>4</td>
</tr>
</tbody>
</table>
METHODS

Latent Factor Model: Matrix Factorization

- user-user interaction?
- user-item interaction?
METHODS

Latent Factor Model: Matrix Factorization

\[ f_i = \mu + b_i + b_{t(i)} + b_{r(i)} + b_{c(i)} + b_{s(i)} + \eta_{r(i)}^{T} \eta_{s(i)} \]
METHODS

Latent Factor Model: Matrix Factorization

\[ f_i = \mu + b_i + b_{t(i)} + b_{r(i)} + b_{c(i)} + b_{s(i)} + \eta_{r(i)}^T \eta_{s(i)} \]

- Latent Bias Model
- Matrix Factorization

- very similar to basic MF model used in SVD++

[Koren 2010]
METHODS

Higher-order interactions?
METHODS

Latent Factor Model: Tensor Factorization

- Recipient-Type-Sender relationships
- CP decomposition
METHODS

Latent Factor Model: Tensor Factorization

\[ f_i = \mu + b_i + b_{t(i)} + b_{r(i)} + b_{c(i)} + b_{s(i)} + \sum_{k} \eta_{r(i),k} + \eta_{s(i),k} + \eta_{t(i),k} \]

- Latent Bias Model
- Tensor Factorization
METHODS

How about other explicit features?
METHODS

How about other explicit features?

- Regression-based latent factor models
  - another layer of regression
  - replacing zero-mean with regression-based mean

\[
\eta_{x(\ast)} = M_x \phi_{x(\ast)} + \epsilon_x \quad x \in \{R, S, T\} \\
\]

\[
b_{x(\ast)} = \pi^T_x \phi_{x(\ast)} + \epsilon_{b_x} \]
METHODS
METHODS
METHODS

Learning procedure

- Maximum A Posterior (MAP)
- Stochastic Gradient Descent
METHODS

Going beyond pointwise learning

- Optimizing Bayesian Personalized Ranked Ranking (BPR)

\[ \sum_{m \in O_{i,+}} \sum_{n \in O_{i,-}} \sigma(f_m - f_n) \]

[Rendle et al. 2009]
EXPERIMENTS

Models

<table>
<thead>
<tr>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (BL)</td>
</tr>
<tr>
<td>Feature Model (FM)</td>
</tr>
<tr>
<td>Latent Bias Model (LFM)</td>
</tr>
<tr>
<td>Feature Bias Model (FBM)</td>
</tr>
<tr>
<td>Matrix Factorization (MF)</td>
</tr>
<tr>
<td>Tensor Factorization (TF)</td>
</tr>
<tr>
<td>Matrix Factorization with Features (MF2)</td>
</tr>
<tr>
<td>Tensor Factorization with Features (TF2)</td>
</tr>
</tbody>
</table>
# Experiments

## Models

<table>
<thead>
<tr>
<th>Features</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seniority</td>
<td>the seniority level of a user</td>
</tr>
<tr>
<td>Visiting</td>
<td>how frequently a user visits LinkedIn</td>
</tr>
<tr>
<td>PageRank</td>
<td>discretized PageRank scores</td>
</tr>
<tr>
<td>Connectedness</td>
<td>how well a user is connected to others</td>
</tr>
<tr>
<td>Social strength</td>
<td>social strength between recipient and sender</td>
</tr>
<tr>
<td>Professionalism</td>
<td>how professional an update’s language is</td>
</tr>
<tr>
<td>Recency</td>
<td>the freshness of an update</td>
</tr>
</tbody>
</table>
EXPERIMENTS: PAIRWISE LEARNING

Relative improvement over baseline

MAP

0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35

4_01 4_08 4_15 9_01 9_10

TF2
TF
MF2
MF
LBM
FBM
FM
EXPERIMENTS: PAIRWISE LEARNING

0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35
Relative improvement over baseline

MAP

TF2
TF
MF2
MF
LBM
FBM
FM

4_01 4_08 4_15 9_01 9_10
Experiments: Pairwise Learning

Relative improvement over baseline MAP

- TF2
- TF
- MF2
- MF
- LBM
- FBM
- FM
EXPERIMENTS: PAIRWISE LEARNING

Relative improvement over baseline

MAP

TF2
TF
MF2
MF
LBM
FBM
FM

4.01 4.08 4.15 9.01 9.10
EXPERIMENTS: PARAMETER SENSITIVITY

The weight of recency

MAP

BL
LBM
MF
TF
MF2
TF2

The weight of recency

0.1 1 5 10 50 100 150 200 250 300 400

0.7

0.6

0.5

0.4

0.3

0.2

0.1

0.1

1 5 10 50 100 150 200 250 300 400

0
## Experiments

Example of highly ranked types of updates

<table>
<thead>
<tr>
<th>Type Description</th>
<th>Bias $b_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Seeker Product Update</td>
<td>0.5765</td>
</tr>
<tr>
<td>Joining Sub-Group</td>
<td>0.5407</td>
</tr>
<tr>
<td>Company News</td>
<td>0.4592</td>
</tr>
<tr>
<td>Joining Group</td>
<td>0.2625</td>
</tr>
<tr>
<td>Profile Picture Update</td>
<td>0.2516</td>
</tr>
<tr>
<td>Initiating Direct Ads Campaign</td>
<td>0.2253</td>
</tr>
<tr>
<td>Profile Update</td>
<td>0.1394</td>
</tr>
</tbody>
</table>
CONCLUSIONS

- We demonstrate that recommender systems + preference-based learning can be used to re-rank social updates.
- A linear model can achieve 60% of the performance of latent factor models, on average.
- A tensor factorization model with regression on explicit features works the best.
- The cold-start problem makes it impossible to model some kinds of interactions.
Thank you.

Liangjie Hong
PhD candidate
WUME Lab
Lehigh University
lih307@cse.lehigh.edu
EXPERIMENTS: COMPARISON

<table>
<thead>
<tr>
<th>Training/Testing</th>
<th>BL</th>
<th>FM</th>
<th>LBM</th>
<th>FBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>4_01(Tr.)/4_08(Te.)</td>
<td>0.5278</td>
<td>0.5317</td>
<td><strong>0.5943</strong></td>
<td>0.5520</td>
</tr>
<tr>
<td>4_08(Tr.)/4_15(Te.)</td>
<td>0.5435</td>
<td>0.5509</td>
<td><strong>0.6040</strong></td>
<td>0.5574</td>
</tr>
<tr>
<td>4_15(Tr.)/4_22(Te.)</td>
<td>0.5218</td>
<td>0.5246</td>
<td><strong>0.5823</strong></td>
<td>0.5235</td>
</tr>
<tr>
<td>9_01(Tr.)/9_10(Te.)</td>
<td>0.4829</td>
<td>0.4911</td>
<td><strong>0.5457</strong></td>
<td>0.4984</td>
</tr>
<tr>
<td>9_10(Tr.)/9_18(Te.)</td>
<td>0.4779</td>
<td>0.4798</td>
<td><strong>0.5432</strong></td>
<td>0.4915</td>
</tr>
<tr>
<td>9_18(Tr.)/9_25(Te.)</td>
<td>0.4768</td>
<td>0.4803</td>
<td><strong>0.5329</strong></td>
<td>0.4886</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training/Testing</th>
<th>MF</th>
<th>TF</th>
<th>MF2</th>
<th>TF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4_01(Tr.)/4_08(Te.)</td>
<td>0.5955</td>
<td>0.6258</td>
<td>0.5951</td>
<td><strong>0.6336</strong></td>
</tr>
<tr>
<td>4_08(Tr.)/4_15(Te.)</td>
<td>0.6079</td>
<td>0.6228</td>
<td>0.6088</td>
<td><strong>0.6535</strong></td>
</tr>
<tr>
<td>4_15(Tr.)/4_22(Te.)</td>
<td>0.5962</td>
<td>0.6014</td>
<td>0.5991</td>
<td><strong>0.6312</strong></td>
</tr>
<tr>
<td>9_01(Tr.)/9_10(Te.)</td>
<td>0.5511</td>
<td>0.5766</td>
<td>0.5523</td>
<td><strong>0.6003</strong></td>
</tr>
<tr>
<td>9_10(Tr.)/9_18(Te.)</td>
<td>0.5412</td>
<td>0.5833</td>
<td>0.5449</td>
<td><strong>0.6109</strong></td>
</tr>
<tr>
<td>9_18(Tr.)/9_25(Te.)</td>
<td>0.5359</td>
<td>0.5799</td>
<td>0.5362</td>
<td><strong>0.5992</strong></td>
</tr>
</tbody>
</table>
## Experiments

The effects of pairwise learning

<table>
<thead>
<tr>
<th>Training/Testing</th>
<th>LBM</th>
<th>MF</th>
<th>MF2</th>
<th>TF</th>
<th>TF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4_01(Tr.)/4_08(Te.)</td>
<td>0.6169</td>
<td>0.6033</td>
<td>0.6151</td>
<td>0.6358</td>
<td>0.6532</td>
</tr>
<tr>
<td>4_08(Tr.)/4_15(Te.)</td>
<td>0.6188</td>
<td>0.6168</td>
<td>0.6188</td>
<td>0.6528</td>
<td>0.6641</td>
</tr>
<tr>
<td>4_15(Tr.)/4_22(Te.)</td>
<td>0.5897</td>
<td>0.6104</td>
<td>0.6191</td>
<td>0.6014</td>
<td>0.6402</td>
</tr>
<tr>
<td>9_01(Tr.)/9_10(Te.)</td>
<td>0.5644</td>
<td>0.5716</td>
<td>0.5723</td>
<td>0.5966</td>
<td>0.6207</td>
</tr>
<tr>
<td>9_10(Tr.)/9_18(Te.)</td>
<td>0.5593</td>
<td>0.5621</td>
<td>0.5607</td>
<td>0.5999</td>
<td>0.6183</td>
</tr>
</tbody>
</table>
Experiments
EXPERIMENTS

Parameter Sensitivity

MAP vs. # of Dimensions

- MF
- TF
- MF2
- TF3