Learning Relevance from Heterogeneous Social Network and Its Application in Online Targeting

Chi Wang (UIUC), Rajat Raina (FB), David Fong (Stanford), Ding Zhou (FB), Jiawei Han (UIUC), Greg Badros (FB)
Facebook’s Social Graph
>750M active users
Jill Sparks commented on Frank Sparks’ Video.

Grandma’s Rose Garden.

0:45 Recorded 2 days ago

Mark Zuckerberg commented on Dave Morin’s Photo.

May the iforce be with you.

Jane Richie wrote on Jim Morgan’s wall.

Did you see the game tonight!?!
Opportunity

Can we use the rich, social information flowing through online social network to personalize user recommendations?
Relevance Matters

People You May Know

Alex Li
20 mutual friends
Add as friend

David Braginsky
31 mutual friends
Add as friend

Social Recommendations

“Sponsored Story”

Kang-Xing Jin and Janaka P Liyanage like Fuki Sushi.

Fuki Sushi

“Sponsored Story”

ICC CWC 2011 Online Game cricketpower.com
Will India continue their winning streak? Is it England’s year? Log in and lead your nation to glory – the official online game is here.

Office Master OMS
A new generation of ergonomic seating for the information based worker. Elegance, simplicity, and usability.

Like · 104 people like this.

Seeking VC/Angel Funding?
lawpivot.com
Before you do, check out the 10 biggest booby traps of term sheets at lawpivot.com.
Ads Targeting on Facebook

- Users:
  - Age, gender, location, pages liked, …
- Advertisers:
  - Targeting criteria
    - Demography
    - Keyword; no immediate intent (e.g., search query).
  - Bid per click.
Targeting: Important for Users
Targeting: Important for Advertisers

“Half the money I spend on advertising is wasted; the trouble is I don’t know which half.”

-- John Wanamaker, “the father of modern advertising”
Ads Ranking Overview

- Users: Age, gender, location, pages liked, ...
- Advertisers: Targeting criteria. Bid per click.
- Simplified Auction: Find top ad ranked by (bid * estimated CTR)

- **Machine learning goal:** Estimate CTR for a given user/ad pair.

\[ f ( \text{user}, \text{ad} ) = \text{CTR} \]
Modeling Relevance

Concept Extraction

Concepts

$50 off Your 1st Bonobos

cim.meebo.com

"Best Mens Pants" – NY Magazine. Get $50 off your first purchase of Bonobos stylish men’s apparel.
Modeling User Interest

- Many different sources of information.
- Entities are represented as text.
  - User writes a status update, friends give comments.
  - Pages have descriptions.
- Relations: Users like pages, click on ads.
Concept Extraction Model

- Design Decision: Use a hand-built taxonomy of labels.
  - Fixed labels – Easy to visualize, modify.
  - Hierarchical – Precise labels, but with high recall.
Text Classification Model

Input Text

$50 off Your 1st Bonobos
cim.meebo.com


Output Labels

Games

Shopping

Computers

Clothing

Sports

Food

\[ \text{Expected output:} \]

Shopping/Clothing \( \Rightarrow \) 0.8

Arts/Fashion/Men \( \Rightarrow \) 0.3

Dekel et al, 2004
Text Classification Model

- Open Directory Project: ~50K labels, 6 levels deep.
Modeling Relevance

Concept Vector from Source 1

Concept Vector from Source 2

Concept Vector from Source x

Concept Vector from Source x+1

Concept Aggregation

Concepts

$50 off Your 1st Bonobos
cim.meebo.com

Vector Space Model

\[
\cos(U, A) = \frac{(\sum_i U_i)^T (\sum_j A_j)}{\|\sum_i U_i\| \|\sum_j A_j\|}
\]

\(U_1\) (User concept from source 1)
- FIFA series: 0.2
- Society: 0.1

\(U_2\) (User concept from source 2)
- Society: 0.2
- Society / Relationship / Dating: 0.4

\(A_x\) (Ad concept from source \(x\))
- Soccer: 0.5

\(A_{x+1}\) (Ad concept from source \(x+1\))
- Beer: 0.6

\(\cos(\quad , \quad ) = 0?\)
Idea 1

- Different sources of information
Idea 2

- Different sources of information.
- Concept spaces are not identical.
  - Users who like Skiing might like ads about Hotels, Hiking, Beer, ...
  - General concepts: Society

Map between user and ad concepts
Better Model

- Train $p$, $q$, $W$ on past click data.
  - Optimizing one parameter = logistic regression.
  - Optimizing all with online gradient descent $\rightarrow$ local optimum
  - L1 regularization

\[
\text{Score}(U, A) = \frac{(\sum_i p_i U_i)^T W (\sum_j q_j A_j)}{||\sum_i p_i U_i|| \cdot ||\sum_j q_j A_j||}
\]

Sports/Golf $\Leftrightarrow$ Shopping/Sports/Golf/Apparel

Sports/Tennis $\Leftrightarrow$ Games/Video_Games/NintendoWii
Experiments

- 5M users, 1% of all
- Sample of 7 days of ad impressions
  - 4 for training
  - 3 for test
- 1.2M training samples / day
- On average 80 connected objects per user
- Offline Test: accuracy of estimated CTR VS. real CTR
- Online experiments: Two groups of users targeted with different relevance feature + a common set of other existent features
Offline Test

\[
Score(U, A) = \frac{(\Sigma_i p_i U_i)^T W (\Sigma_j q_j A_j)}{\|\Sigma_i p_i U_i\| \|\Sigma_j q_j A_j\|}
\]

Test Set Logloss Gain

- Uniform: 1.0
- \(p, q\) only: 3.8x
- All Learned: 14.3x
- L1 regularization: 13.7x

(80% sparser, 53x speed up for online score computation)

Learned user source weight \(p\)
A/B tests

\[
Score(U, A) = \frac{(\Sigma_i p_i U_i)^T W (\Sigma_j q_j A_j)}{||\Sigma_i p_i U_i|| ||\Sigma_j q_j A_j||}
\]

All Learned

\(p, q\) only
Interpret the matrix $W$

**ODP as user concepts**
- Ad concepts associated with most user concepts
  - Turn based video games
- Strong positive associations
  - Browser Based Casinos $\rightarrow$ Facebook Platform
  - Games $\rightarrow$ Point of Purchase Displays
- Negative associations
  - US Colleges $\rightarrow$ US Colleges

**Demography as user concepts**
- Ad concepts associated with most user concepts
  - Christianity Denominations
- Strong positive associations
  - Female $\rightarrow$ Dance
  - Male $\rightarrow$ Guitar
  - Married $\rightarrow$ Home & Garden
- Negative associations
  - US $\rightarrow$ Mah Jongg
Conclusion

- Interest modeling
  - Vector representation for hierarchical concepts
  - Heterogeneous sources and links
- Relevance matching
  - Related concepts and their association weight
- Ads ranking
  - Optimized parameters from historical data
  - Strong signal of relevance
  - Good interpretability