AI that creates professional opportunities at scale

Deepak Agarwal
AI in practice, AAAI, San Francisco, USA
Feb 6th, 2017
We are seeking new professional opportunities all the time

To Advance our Careers
Mark Hull
Senior Director of Product Management at LinkedIn
1d - Edited

What's your walk-up song -- the music playing as you get on the stage (like a batter at a baseball game)? I'm giving a presentation in Singapore, and need to choose my walk-up song. Any recommendations? Goals: energize the audience, get me fired up, not get me fired, and be a dog whistle for music hipsters in the audience, with maybe a tip of the hat to my APAC audience. What’s yours? Or add your suggestions to my Spotify Playlist.

Mark's Walk-Up Music for APAC Presentation, a playlist by Mark Hull on Spotify
open.spotify.com

+12,000 Comments

+500,000 Views
Emily likes AI and deep learning. Got familiar with research on Dropout through an article shared by one of her connection. Followed through, implemented and got great recognition at work.
Impact of creating professional opportunities at scale

• Closing the skills gap by leveraging supply on a global basis, reducing income disparity, training the workforce to prepare for the future.

JOBS QUEUE in INDIA, over supply
Almost all user experience powered by AI

SEARCH, CONNECT, NURTURE
Search and connect, via recommendations, keep in touch

STAY INFORMED
through professional news and knowledge

GET HIRED
and build your career
All customer experience too

HIRE
Searching, researching, nurturing qualified candidates

MARKET
Targeting right audience via native ads, display ads

SELL
Searching and researching decision makers, following up on leads
How is AI used to drive value?

• Two broad classes of problems

  • **Recommendation:**
    ◆ Recommend items to a users to optimize one of more objectives of interest (e.g., connections, job applies, engagement/revenue)

  • **Search:**
    ◆ User asks a question through query, need to provide answer by searching through a large set of possibilities
High level Approach
Connecting long-term objectives to proxies that can be optimized by machines/algorithms

Long-term objectives (return visits to site, connections, quality job applies,...)

Short-term proxies (CTR, connection prob, apply prob, ...)

Large scale optimization via ML.

Experiment
Learn
Innovate
Recommendation Problem: Not merely supervised learning

User $i$ with user features $x_i$ (profile info, browse history, geo-location, search history, ...)

visits

Algorithm selects

item $j$ with item features $x_j$ (keywords, content categories, author, ...)

Interaction (Click, share, like, apply, connect,..)/no-interaction

(i,j) : response $y_{ij}$

Which item should we select?

- The one with highest predicted utility
- The one most useful for improving the utility prediction model

Exploit Explore
Data: Profile and Network Key to better AI
User Characteristics

Profile Information
Title, skills, work history, education, endorsements, presentations, follows,..

Behavioral
Activities, search,..
AI used to standardize member profile data

- Language detection
- Spell correction
- Segmentation
- Entity recognition
- Entity resolution

Profile IDs in taxonomy

Text mining

Machine learning

IDs in taxonomy
An AI-based holistic solution for taxonomy development

- Human-curation to create and structure taxonomies
- Slow, costly, not scalable

- Deep neural network for entity embedding
- Fast, cheap, scalable

Auto-generate:
- New entities
- Synonyms
- Entity hierarchical structure etc.
Proactively helping members with profile completion

Infer the skill reputation of LinkedIn member through user profiles (static), user behaviors (dynamic) and user connections (network).

Auto-generate profile summary through knowledge base.

Estimate the rewards of adding a new skill through peer analysis.

Standardization auto-generates summaries for members. Members with summaries receive more page views and InMails.

Standardization develops relevance models and improves skills quality to drive quality endorsements.
Better knowledge base powers better AI

Members (profile), Jobs, Companies etc. collectively form the LinkedIn professional knowledge base.

direct knowledge base approach based on logical inference rules

direct features

graph embedding

better representation

Search
Job recommendation
Profile edit
Ads targeting
Feed
Course learning
Insights
Endorsement
...

other training data like network, user behaviors
Network: Connections are fundamental to drive value

- A well connected member is much more likely to be able to use the LinkedIn eco-system for her professional development and growth
PYMK (growing connections)
Adding nodes: Viral Loops through recommendations

Contacts Upload

Abc
abc@myschool.edu

XYZ,
XYZ@myschool.edu

Landing Page
Adding nodes: Viral loops from Guest Invitation

• Members inviting others to join the network is crucial for growth of the network
• Anderson et. al. (WWW 2015)
• LinkedIn’s cascade trees are very viral

The growth of trees is almost linear in time
Item Features

Articles
author, sharer, keywords, named entities, topics, category, likes, comments, latent representation, etc.

Job Post
company, title, skills, keywords, geo, ...

Data Obsession and Risk Aversion
Published on March 7, 2015

Deepak Agarwal
VP Engineering, Head of Relevance at LinkedIn

For a statistician like me, the best way to make good decisions in the face of uncertainty is through data. That’s how I was trained during my academic life. I was interacting with like minded folks during the ten exciting years working as a research scientist at ATT and Yahoo! Labs. It got deeply ingrained into my thinking and personality. It has become a way of life for me, almost a religion. As a true devotee, I worship the data religion and the calculus of statistics. It keeps me cozy and comfortable when making decisions.

A recent event stirred up some doubts in this deep rooted belief. It started with a discussion at a meeting I was attending at work. There was a hypothesis floating around --- “we are sometimes too thoughtful and cautious in our approach, we tend to become obsessed with data and tend to be conservative in making quick decisions. We are not taking enough risks by trusting our gut and instinct”. As you can well imagine by now, I was quick to point out that data obsession has nothing to do with taking risks. It is merely a tool that helps in making better decisions. How can it influence a person’s
User Intent

• Why are you here?
  • Hire, get hired, stay informed, grow network, nurture connections, sell, market,..
  • Explicit (e.g., visiting jobs homepage, search query),
  • Implicit like job seeker (needs to be inferred, e.g., based on activities)
We know about users, their intent, about items

How to Recommend items to users algorithmically?

Framework, Infrastructure and Tools
Under the Hood of a Typical Recommender System at LinkedIn
Example Application: Job Recommendation

Jobs you may be interested in
Based on your job preferences: 4 locations, experience level, company size, 108 industries

- **Sr. Data Scientist**
  Comcast Silicon Valley Innovation C...
  San Francisco Bay Area
  - 1 alum works here
  - Posted 4 days ago

- **Staff Data Engineer**
  Credit Karma
  San Francisco, CA, US
  - 2 people hired from your company
  - Posted 2 days ago

- **Data Scientist – PhD / Master’s - Py...**
  FILD
  San Ramon, California
  - Posted 14 hours ago

- **Senior Tableau Architect - Healthc...**
  Cognizant
  US-California-Pleasanton - CA USA, CLT
  - 5 people hired from your company
  - Posted 18 days ago

- **Sr. Clinical Data Associate**
  Gilead Sciences
  US - California - Foster City
  - 4 people hired from your company
  - Posted 3 days ago

- **Sr. Software Engineer, Back End**
  Gliffy
  San Francisco, California
  - Posted 15 days ago

- **Sr Data Scientist/ Research Techno...**
  Comcast
  Sunnyvale, CA
  - 1 person hired from your company
  - Posted 4 days ago
Objective: Job Applications

Predict the probability that a user $i$ would apply for a job $j$ given ... 

- **User features**
  - Profile: Industry, skills, job positions, companies, education
  - Network: Connection patterns

- **Item (i.e., job) features**
  - Source: Poster, company, location
  - Content: Keywords, title, skills

- **Data about users’ past interactions with different types of items**
  - Items: Jobs, articles, updates, courses, comments
  - Interactions: Apply, click, like, share, connect, follow
Components

- Front End Service
- Ranking Service
- User DB
- Item DB
- User Feature Stores
- User Feature Pipelines
- Item Feature Pipelines
- Data Stream Processing
- Live Index Updater
- Model Training Pipelines
- Offline Data Pipelines
- ETL

Technologies:
- Apache Kafka
- Project Voldemort
- Apache Samza
- Photon-ML
- Apache Hadoop, Pig, Scalding, Spark

Data Streams

Experimentation Platform A/B testing

Offline

Online
Model Training

Feature Processing

Raw User Features

DAG of Transformers

Feature Vector of User $i$ $x_i$

Matching Feature Vector $m_{ij}$

Raw Item Features

DAG of Transformers

Feature Vector of Item $j$ $z_j$

Parameter Learning

$p(i \text{ applies for } j) = f( x_i, z_j, m_{ij} | \theta, \alpha_i, \beta_j )$

Global parameter vector

Parameter vector for each user $i$

Parameter vector for each item $j$
$p(i \text{ applies for } j) = f( x_i, z_j, m_{ij} | \theta, \alpha_i, \beta_j )$

Global parameter vector

Parameter vector for each user $i$

Parameter vector for each item $j$
Online A/B Experiments

**Experiment setting**
- Select a user segment
- Allocate traffic to different models

**Result reporting**
- Report experimental results
- Impact on a large number of metrics
How we do large scale Regression?
GAME: Generalized Additive Mixed-Effect Model
GAME as a framework, optimizer and library
GAME as a Framework

• Unifies and *mixes* different models into a principled *additive* model.

• Predicting the response of *user i* on *item j*:

\[ g(\mathbb{E}[y_{ij}]) = f_1 + f_2 + f_3 + \ldots \]

• \( f_i \): An effect (model)
• \( \mathbb{E}[y_{ij}] \): Expectation of response
• \( g(\cdot) \): Link function
GAME as a Library

- Basic models implemented as building block $f_i$.
  - Matrix factorization
  - Generalized Linear Model
  - ...
- New models can be directly composed by mixing existing building blocks.

$$g(\mathbb{E}[y_{i,j}]) = f_1 + f_2 + f_3 + \ldots$$
Ex: GLM + MF

- Predicting the response of *user i* on *item j*:

\[ g(\mathbb{E}[y_{ij}]) = f_1 + f_2 = x'_{ij}w + u'_{i}v_{j} \]

- \( f_1 \) : GLM
- \( f_2 \) : MF
- \( \mathbb{E}[y_{ij}] \) : Expectation of response
- \( g(\cdot) \) : Link function
GLMix¹: Fine-Grained GAME with Linear Components

¹: GLMix: Generalized Linear Mixed Models For Large-Scale Response Prediction
X. Zhang et al., KDD2016
• Jobs homepage
  • Ramped to serve **100%** traffic (400 million LinkedIn members)
  • **significant** lift in job application rate
• Jobs homepage
  • Ramped to serve 100% traffic (400 million LinkedIn members)
  • significant lift in job application rate

• Advertising, PYMK, Feed, Recruiter,…
  • significant lifts in performance
Generalized Linear Model (GLM)

- Predicting the response of user \( i \) on item \( j \):

\[
g(\mathbb{E}[y_{ij}]) = f_1 = x'_{ij}w
\]

- \( x_{ij} \): Feature vector (profile, items, intent, cross-prod, GBDT)
- \( w \): Coefficient vector
- \( \mathbb{E}[y_{ij}] \): Expectation of response
- \( g(\cdot) \): Link function
GLM for Job Recommendation

$$g(\mathbb{E}[y_{ij}]) = x'_{ij}w$$

- Alice and Annie are about the same age, similar majors in college… (similar member features $x_{ij}$)
- Alice likes to take more risks with start-ups
- Annie likes more stable career just like her parents
- GLM may return similar set of jobs to both
GLM for Job Recommendation

\[ g(\mathbb{E}[y_{ij}]) = x'_{ij} w \]

- Alice and Annie are about the same age, similar majors in college… (similar member features \( x_{ij} \))
- Alice likes to take more risks with start-ups
- Annie likes more stable career just like her parents
- GLM may return similar set of jobs to both
- Need more fine-grained modeling at different granularity to better personalize the model!
GLMix: Generalized Linear Mixed Model

• Predicting the response of *user i* on *item j*:

\[
g(\mathbb{E}[y_{ij}]) = f_1 + f_2 + f_3 = x'_{ij}w + x'_{j}\alpha_i + x'_{i}\beta_j
\]

• Model coefficients with different **granularities**:
  • *Per-user* random effect coefficients \(\alpha_i\)
  • *Per-item* random effect coefficients \(\beta_j\)

• GLMix = GLM + per-user model + per-item model
GLMix for Job Recommendation

\( g(\mathbb{E}[y_{ij}]) = x'_{ij} w + x'_j \alpha_i + x'_i \beta_j \)

- **Global fixed effect model**
  - Similarity between member profile and jobs profile, e.g. do the member skills and job skills look similar?
GLMix for Job Recommendation

\[ g(\mathbb{E}[y_{ij}]) = x'_{ij}w + x'_{j}\alpha_i + x'_{i}\beta_j \]

- **Global fixed effect model**
  - Similarity between member profile and jobs profile, e.g. do the member skills and job skills look similar?

- **Per-member random effect model**
  - E.g. If a member has applied to a job with title = “software engineer”, we will boost “software engineer” jobs more in her results.
GLMix for Job Recommendation

\[ g(\mathbb{E}[y_{ij}]) = x'_{ij} w + x'_j \alpha_i + x'_i \beta_j \]

- **Global fixed effect model**
  - Similarity between member profile and jobs profile, e.g. do the member skills and job skills look similar?

- **Per-member random effect model**
  - E.g. If a member has applied to a job with title = “software engineer”, we will boost “software engineer” jobs more in her results.

- **Per-job random effect model**
  - E.g. If a job gets an apply with a member titled “software engineer”, we will boost this job more for members with this title.
Alice and Annie’s problem revisited

\[ g(\mathbb{E}[y_{ij}]) = x'_{ij}w + x'_j\alpha_i + x'_i\beta_j \]

- Per-user random effect coefficients for Alice: \( \alpha_{\text{Alice}} \)
- Per-user random effect coefficients for Annie: \( \alpha_{\text{Annie}} \)
- Alice and Annie now may have different job recommendations given their per-user coefficients.
Takeaways

• GAME unifies and *mixes* different models into a principled *additive* model.
  • MF + GLM = RLFM/Matchbox
  • GLM + DNN = Wide & Deep Learning
  • ...
Takeaways

• GAME unifies and *mixes* different models into a principled additive model.
  • MF + GLM = RLFM
  • GLM + DNN = Wide & Deep Learning
  • ...

• GLMix is GAME with linear component that captures signal from *different granularity*
  • GLMix = GLM + Per-member model + per-item model + …
Takeaways

• GAME unifies and mixes different models into a principled additive model.
  • MF + GLM = RLFM
  • GLM + DNN = Wide & Deep Learning
  • ...

• GLMix is GAME with linear component that captures signal from different granularity
  • GLMix = GLM + Per-member model + per-item model + …

• GAME is part of an open-source library
  • Search for Photon-ML
  • https://github.com/linkedin/photon-ml
Some Challenges and Learnings
1. Cost of a Bad Recommendation

- Making ML robust when a few bad recommendations can hurt product brand
  - Maximize precision without hurting performance metrics significantly
    - Collect negative feedback from users, crowd; incorporate within algorithms
    - Create better product focus, filter unnecessary content from inventory
      - E.g., unprofessional content on Feed
- Better insights/explanations associated with recommendations help build trust
2. Data Tracking

• Proper data tracking and monitoring is not always easy!
  • Data literacy and understanding across organization (front-end, UI, SRE)
  • Proper tooling, continuous monitoring very important to scale the process

• Our philosophy: Loose coupling between FE and BE teams!
  • FE (client) emits limited events along with trackingid
  • BE emits more details and joins against trackingid

• Tracking events can have significant impact
  • View-port tracking (what the user actually saw) for more informative negatives
3. Content Inventory

• High quality and comprehensive content inventory as important as recommendation algorithms

• Examples: Jobs, Feed

• Supply and demand analysis, gap analysis, proactively producing more high quality content for creating appropriate inventory
4. A/B Testing with Network Interference

• Random treatment assignments (spillover effects, need to adjust)

• Treatment recommendations affect control group as well
  • A like/share in treatment may create a new item when ranking in control
Nicely written Summary