

Lessons Learned from Building real-life Recsys



Xavier Amatriain (Quora)
Deepak Agarwal (LinkedIn)

4. Recommendations @Quora

Our Mission

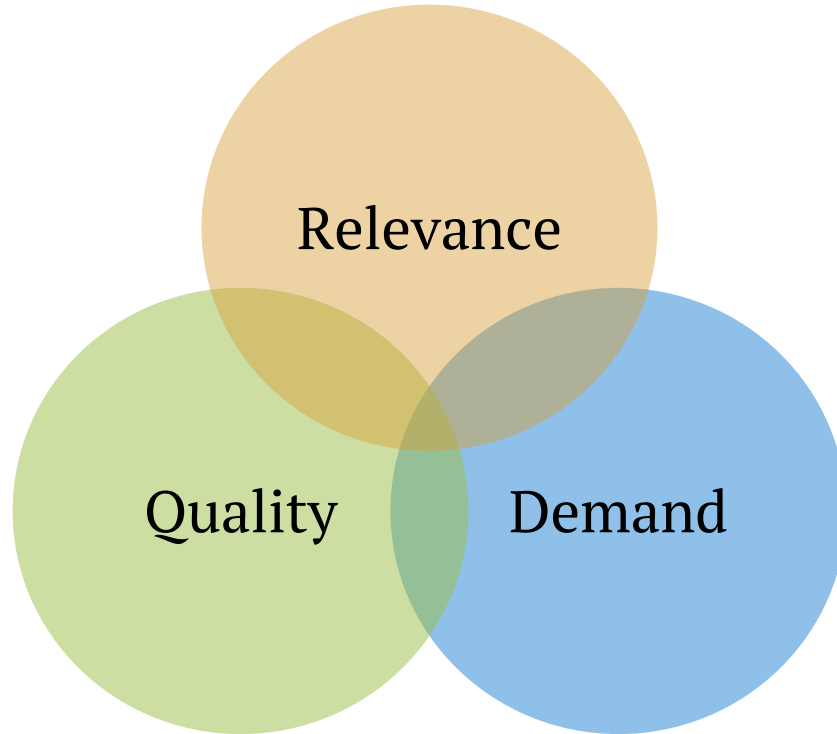
“To share and grow the world’s knowledge”

- Millions of questions & answers
- Millions of users
- Thousands of topics
- ...

The screenshot shows the Quora website interface. At the top, there's a navigation bar with the Quora logo, a search bar, and links for Home, Write (with a 10 notification badge), Notifications, and a user profile for Xavier. Below the navigation bar, the main content area is titled 'Machine Learning'. On the left, there's a sidebar with 'FEEDS' and 'EDIT' options. The 'FEEDS' section lists various topics like Data Science, Software and Applications, Recommendation Systems, Football (Soccer), Parenting, and Machine Learning. The 'Machine Learning' section further lists sub-topics like Barcelona, Spain; Spain; FC Barcelona; and MLconf 2015 Seattle. Below this, there's a 'TRENDING NOW' section with topics like Avengers: Age of Ultron, Release of Apple Watch, Game of Thrones Season 5 Episode 3, Silicon Valley Season 2 Episode 3, and Mayweather-Pacquiao Fight.

The main content area displays several questions and answers. The first question is 'What are the advantages of different classification algorithms?' by Naeem Siddiqi. It has 1 upvote and 12 downvotes. The second question is 'Is correlation a good indicator for choosing parameters in linear regression?' by Peter Flom, an independent statistical consultant. It has 12 upvotes and 1 downvote. The third question is 'What do employers think of Insight Data Science Bootcamp?' by Big Data. It has 2 upvotes and 2 downvotes. The fourth question is 'What is the data science scene like in Boston?' by Cortlandt Johnson, @FKAAAtlas, @TerribleLabs, @TicketZen. It has 2 upvotes and 2 downvotes. The fifth question is 'What are some examples of things neural networks can or can't do?' by Artificial Neural Networks. It has 2 upvotes and 2 downvotes.

What we care about



Data @ Quora

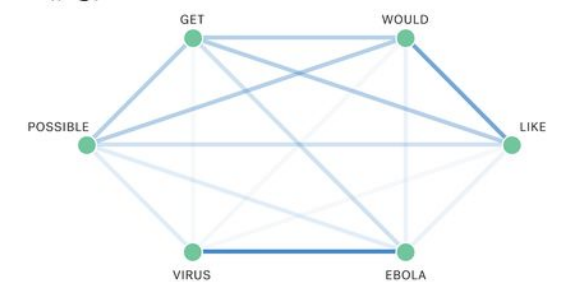
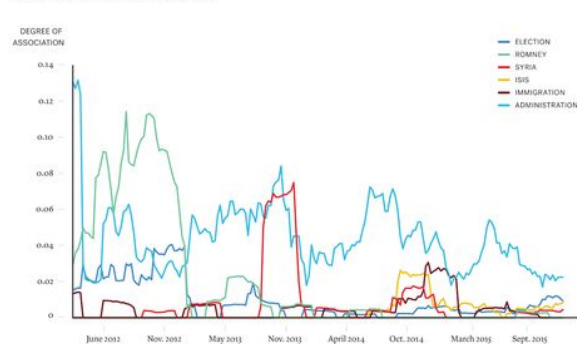
DATA @ QUORA

Mapping the Discussion on Quora Over Time
through Question Text

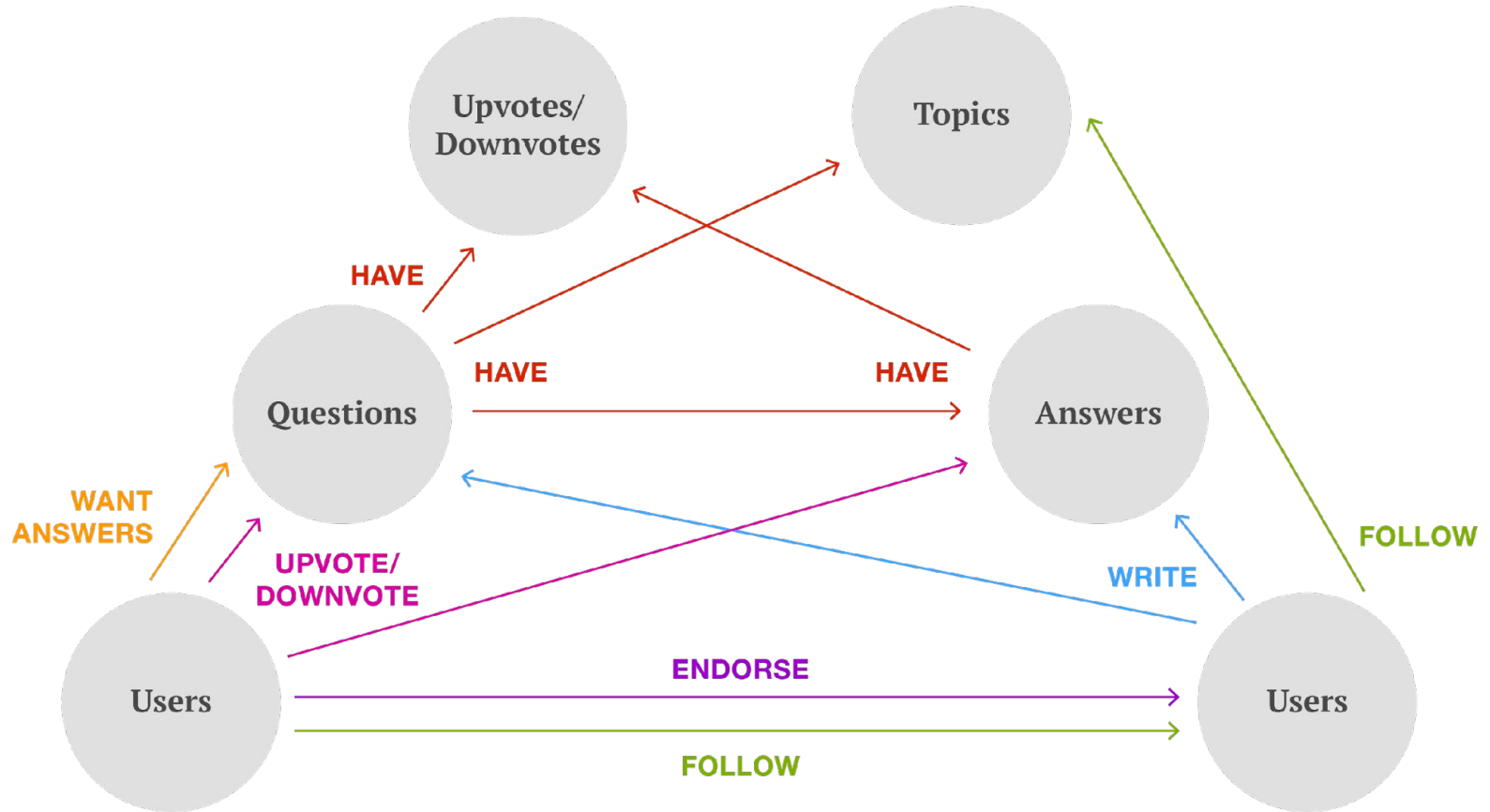
14, Q4 GET WENWEN TAO (陶雯雯)

14, Q4 GET WENWEN TAO (陶雯雯)

14, Q4 GET WENWEN TAO (陶雯雯)

[illegible]

Lots of data relations



Recommendations @Quora

Recommendations at Quora

- Homepage feed ranking
- Email digest
- Answer ranking
- Topic recommendation
- User recommendation
- Trending Topics
- Automated Topic Labelling
- Related Question
- ...

The screenshot shows a Quora interface with several annotations indicating recommendation points:

- click**: Points to the question title "What is the genesis of Instagram?".
- expand**: Points to the "(more)" link at the end of the answer text.
- upvote**: Points to the "Upvote" button.
- downvote**: Points to the "Downvote" button.
- share**: Points to the "Share" button.

The main content area shows a question "What is the genesis of Instagram?" by Kevin Systrom, CEO, co-founder. The answer text begins: "First off, we have to say that we never expected the overwhelming response that we've seen. We went from literally a handful of users to the #1 free photography app in a matter of hours. But as my cofounder ... (more)". Below the answer are buttons for "Upvote | 3.8k", "Downvote", "Comments | 32+", and "Share | 248".

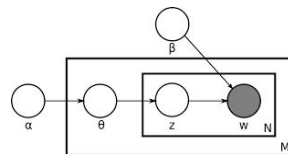
Below the main content, there are two sidebars:

- Discover new people**: Lists users like James Altucher, Feifei Wang, and Ellen Vrana with their follower counts.
- RELATED QUESTIONS**: Lists related questions such as "How do you decide to regularize between L1/L2 or best/greedy subset selection?" and "What's a good way to provide intuition as to why the lasso (L1 regularization) results in sparse weight vectors?".

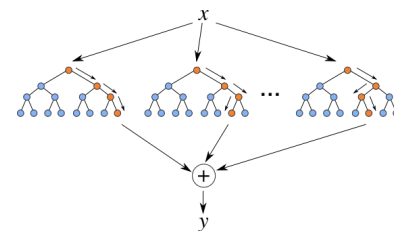
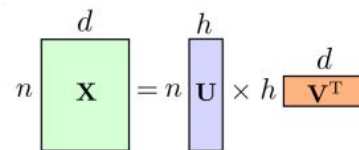
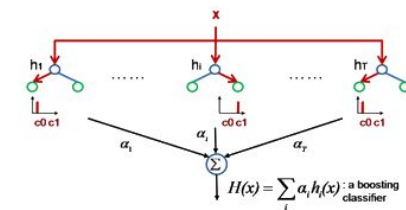
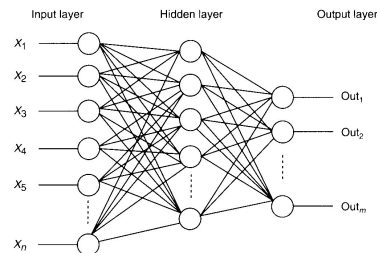
At the bottom, there is a section for "What US government agency would deal with an alien found on Earth?" with a list of answers from Loren Petrich, Michael Hessler, and James Card.

Models

- Deep Neural Networks
- Logistic Regression
- Elastic Nets
- Gradient Boosted Decision Trees
- Random Forests
- LambdaMART
- Matrix Factorization
- LDA
- ...
-



$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$



$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} (\|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1).$$

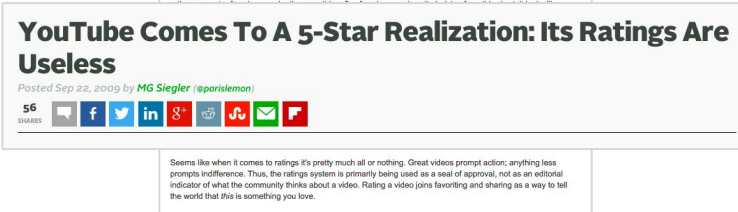
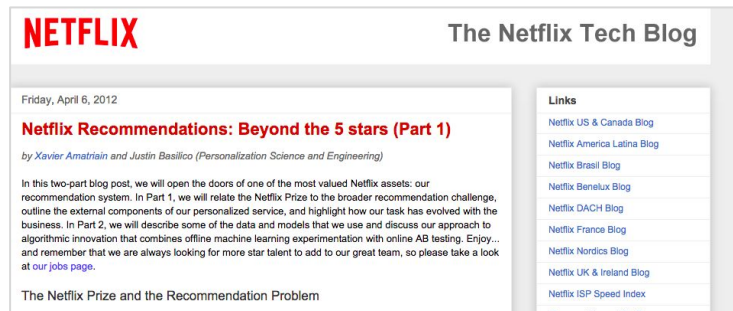
5. Lessons learned (Part II)

1. IMPLICIT SIGNALS BEAT
EXPLICIT ONES
(ALMOST ALWAYS)

Implicit vs. Explicit

Quora

- Many have acknowledged that implicit feedback is more useful
- Is implicit feedback really always more useful?
- If so, why?



Implicit vs. Explicit

- Implicit data is (usually):
 - More dense, and available for all users
 - Better representative of user behavior vs. user reflection
 - More related to final objective function
 - Better correlated with AB test results
- E.g. Rating vs watching

The image displays two screenshots from the IMDb website, illustrating the difference between explicit and implicit data in movie recommendations.

Top-US-Grossing Feature Films Released In 2014

1-50 of 9,031 titles.

Sort by: Popularity | A-Z | User Rating | Num Votes | US Box Office | Runtime | Year | US Release Date

Rank	Movie Title	Year	Rating	Box Office
1	American Sniper	2014	7.3	\$350M
2	The Hunger Games: Mockingjay - Part 1	2014	6.8	\$337M
3	Guardians of the Galaxy	2014	8.1	\$333M
4	Captain America: The Winter Soldier	2014	7.2	\$260M
5	The Lego Movie	2014	7.8	\$258M

Highest Rated Feature Films Released In 2014

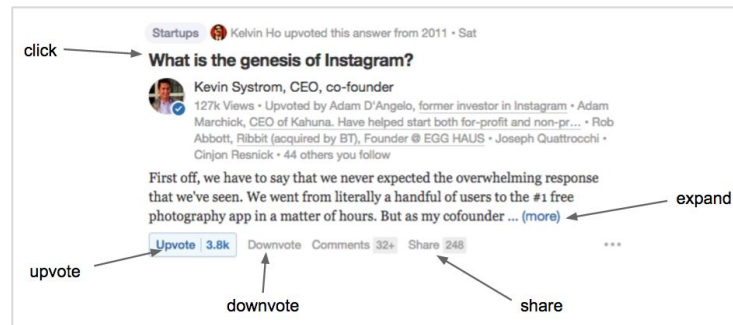
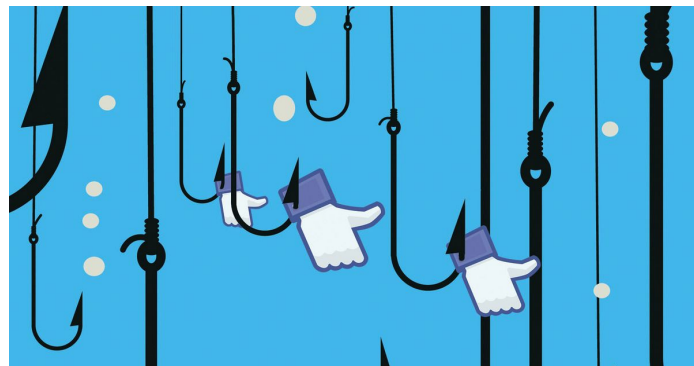
1-50 of 9,031 titles.

Sort by: Popularity | A-Z | User Rating | Num Votes | US Box Office | Runtime | Year | US Release Date

Rank	Movie Title	Year	Rating
1	Forgive and Forget	2014	9.5
2	Mahjong and the West	2014	8.4
3	National Theatre Live: Coriolanus	2014	9.1
4	Burning Dog	2014	8.9
5	The Rule of Law	2014	9.0

Implicit vs. Explicit

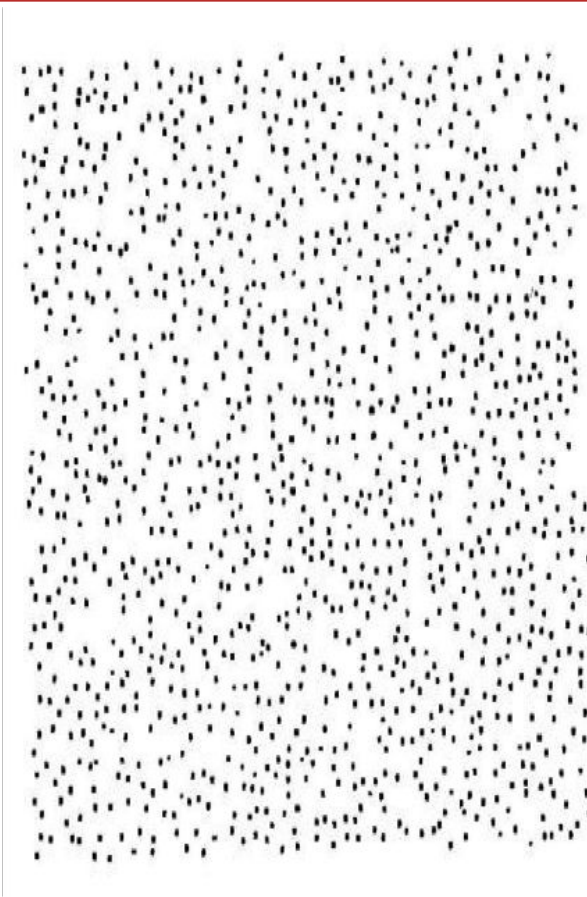
- However
 - It is not always the case that direct implicit feedback correlates well with long-term retention
 - E.g. clickbait
- Solution:
 - Combine different forms of implicit + explicit to better represent long-term goal



2. BE THOUGHTFUL ABOUT YOUR TRAINING DATA

Defining training/testing data

- Training a simple binary classifier for good/bad answer
 - Defining positive and negative labels -> Non-trivial task
 - *Is this a positive or a negative?*
 - funny uninformative answer with many upvotes
 - short uninformative answer by a well-known expert in the field
 - very long informative answer that nobody reads/upvotes
 - informative answer with grammar/spelling mistakes
 - ...



3. YOUR MODEL WILL LEARN
WHAT YOU TEACH IT TO LEARN

- Model will learn according to:
 - Training data (e.g. implicit and explicit)
 - Target function (e.g. probability of user reading an answer)
 - Metric (e.g. precision vs. recall)
- Example 1 (made up):
 - *Optimize probability of a user going to the cinema to watch a movie and rate it “highly” by using purchase history and previous ratings. Use NDCG of the ranking as final metric using only movies rated 4 or higher as positives.*

Example 2 - Quora's feed

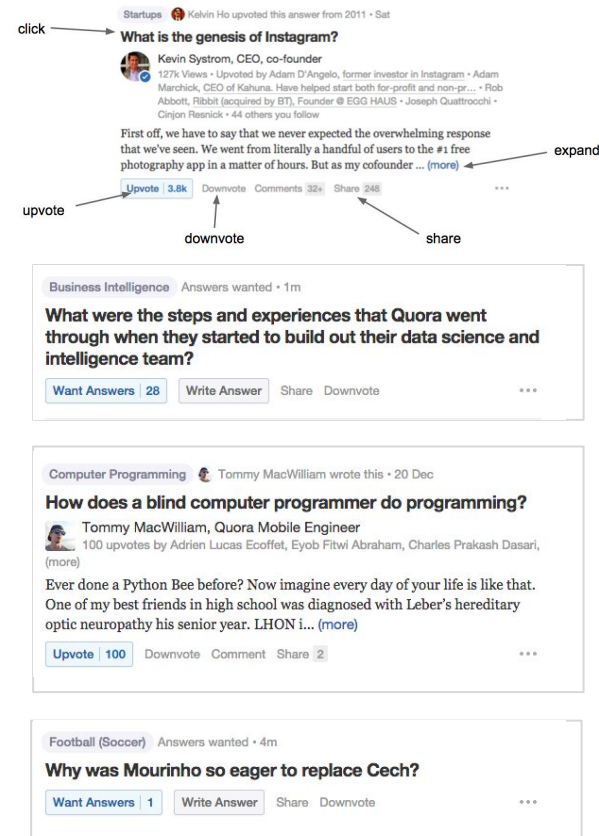
- Training data = implicit + explicit
- Target function: Value of showing a story to a

user \sim weighted sum of actions:

$$v = \sum_a v_a 1\{y_a = 1\}$$

- predict probabilities for each action, then compute expected value: $v_{\text{pred}} = E[V | x] = \sum_a v_a p(a | x)$

- Metric: any ranking metric



4. EXPLANATIONS MIGHT MATTER
MORE THAN THE PREDICTION

Explanation/Support for Recommendations

 Sarah Smith  Richard Henry and 3 more upvoted this • 7h

How can I complain about my roommate who is cheating on his Google phone interviews?



Ben Garrison, Software Engineer at Google

304.3k Views • Upvoted by Jeremy Miles, Quantitative analyst at Google, Mayeesha Tahsin, Sarah Smith, and 3 others you follow

First off, I really appreciate your trying to make sure the right thing happens. I think that's great. Cheating sucks. However, the answer is "don't worry about it". Phone screens here at Google ar... [\(more\)](#)

Upvote 968 Downvote Comments 23+ Share

 Discover new topics

 **Last.fm**
Last.fm builds detail...
Followed by Neal Lathia and 8 more

Follow 21.9k

 **Quantitative Finance**
Quantitative finance ...
Followed by Katie Hoban and 22 more

Follow 74.1k

 **California Stat**
California State
Followed by Rachelle Baratto

Follow 2.4k

PIRIT
abc family

RUSSELL CROWE THE NEXT THREE DAYS
FROM THE DIRECTOR OF CRASH

from PRADA
CAMPBELL SCOTT KERRY WASHINGTON MICHAEL FASSBENDER THE DRAKE

RACHEL ELSON TOM STURRIDGE WAITING FOR FOREVER
"THIS ONE GRABS YOUR HEART!"

The Next Three Days
2010 PG-13 133 minutes

When his wife is sent to jail on murder charges she fervently denies, a college professor hatches a meticulous plan for the ultimate prison escape.
[More Info](#)

Starring: Russell Crowe, Elizabeth Banks
Director: Paul Haggis

Based on your interest in: *Iron Man 2*, *John Q* and *X-Men Origins: Wolverine*

Our best guess for Xavier:
★★★★★

BRIDGES DAMON BRÖLIN

Recreation

5. IF YOU HAVE TO PICK ONE SINGLE APPROACH,
MATRIX FACTORIZATION IS YOUR BEST BET

Matrix Factorization

- MF can be interpreted as
 - Unsupervised:
 - Dimensionality Reduction a la PCA
 - Clustering (e.g. NMF)
 - Supervised:
 - Labeled targets \sim regression
- Very useful variations of MF
 - BPR, ALS, SVD++
 - Tensor Factorization, Factorization Machines
- However...

$$\begin{matrix} & d \\ n & \mathbf{X} \end{matrix} = \begin{matrix} & h \\ n & \mathbf{U} \end{matrix} \times \begin{matrix} & d \\ h & \mathbf{V}^T \end{matrix}$$

6. EVERYTHING IS AN ENSEMBLE

Ensembles

- Netflix Prize was won by an ensemble
 - Initially Bellkor was using GDBTs
 - BigChaos introduced ANN-based ensemble
- Most practical applications of ML run an ensemble
 - Why wouldn't you?
 - At least as good as the best of your methods
 - Can add completely different approaches (e.g. CF and content-based)
 - You can use many different models at the ensemble layer: LR, GDBTs, RFs, ANNs...

The BellKor Solution to the Netflix Grand Prize

Yehuda Koren
August 2009

The BigChaos Solution to the Netflix Grand Prize

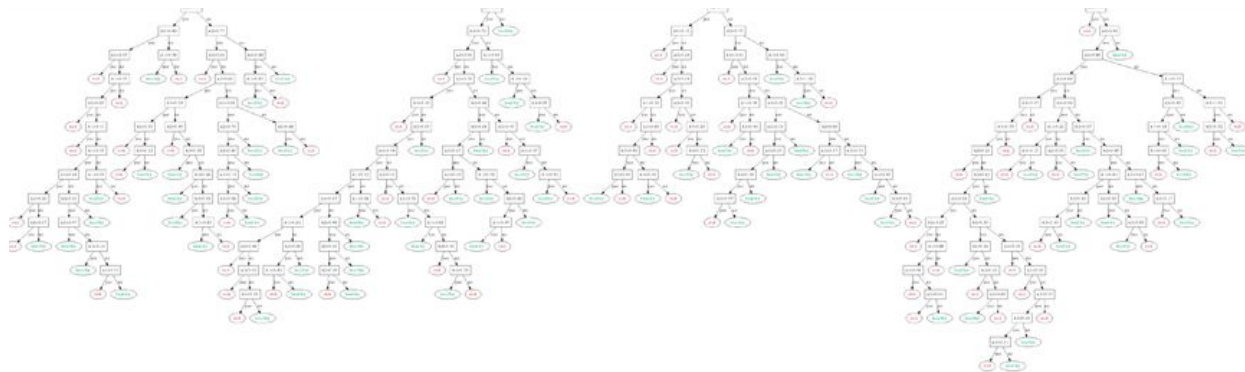
Andreas Töschel and Michael Jahrer
commendo research & consulting
Neuer Weg 23, A-8580 Köflach, Austria
{andreas.toeschel,michael.jahrer}@commendo.at

Robert M. Bell*
AT&T Labs - Research
Florham Park, NJ

September 5, 2009

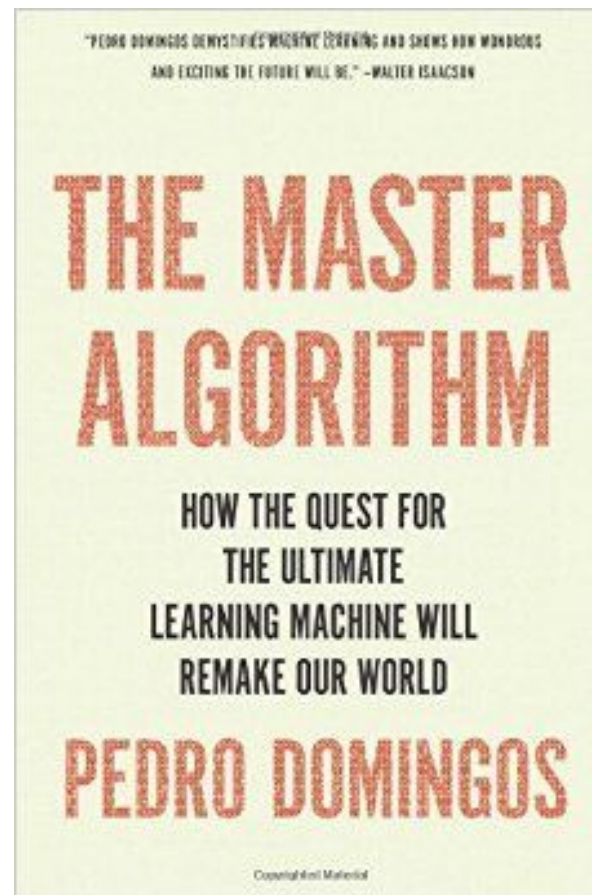
Ensembles & Feature Engineering

- Ensembles are the way to turn any model into a feature!
- E.g. Don't know if the way to go is to use Factorization Machines, Tensor Factorization, or RNNs?
 - Treat each model as a “feature”
 - Feed them into an ensemble



The Master Algorithm?

It definitely is an ensemble!



7. BUILDING RECOMMENDER SYSTEMS IS ALSO
ABOUT FEATURE ENGINEERING

Need for feature engineering

In many cases an understanding of the domain will lead to optimal results.

What is a good Quora answer?

- truthful
- reusable
- provides explanation
- well formatted
- ...

What music do data scientists usually listen to while working?



Paula Griffin, data scientist and biostatistics PhD ... (more)

13 upvotes by William Chen, Alexandr Wang (王誉舜), Sheila Christine Lee, (more)

I was figuring that this question was just fishing for someone to answer that Big Data is their favorite band. Unfortunately, the question log indicates this was asked about 6 months before their EP came out, so there goes that theory.

This is going to be a pretty odd list, but here's the list, in order of decreasing social acceptability:

- Electropop -- Banks and CHVRCHES are my favorites at the moment.
- Miscellaneous alt-rock -- this category basically includes anything I found out about from listening to Sirius XM in the car.
- Nerd rock -- What kind of geek would I be if Jonathan Coulton wasn't on this list?



Shankar Iyer, data scientist at Quora

10 upvotes by William Chen, Sheila Christine Lee, Don van der Drift, (more)

Based on the Pandora stations that I've been listening to, my recent work-time listening consists of:

1. **Acoustic folk music:** John Fahey, Leo Kottke, Six Organs of Admittance, etc.
2. **Post-Rock / Ambient Music:** Sigur Rós, Gregor Samsa, the Japanese Mono, Eluvium, El Ten Eleven, etc.
3. **Hindustani:** mostly Vishwa Mohan Bhatt
4. **Carnatic:** recently Rajeswari Pariti
5. **Classical Guitar:** recently Paul Galbraith, Konrad Ragossnig, etc.

How are those dimensions translated into features?

- Features that relate to the answer quality itself
- Interaction features (upvotes/downvotes, clicks, comments...)
- User features (e.g. expertise in topic)



Paula Griffin, data scientist and biostatistics PhD ... (more)

13 upvotes by William Chen, Alexandr Wang (王誉舜), Sheila Christine Lee, (more)

I was figuring that this question was just fishing for someone to answer that Big Data is their favorite band. Unfortunately, the question log indicates this was asked about 6 months before their EP came out, so there goes that theory.

This is going to be a pretty odd list, but here's the list, in order of decreasing social acceptability:

- Electropop -- Banks and CHVRCHES are my favorites at the moment.
- Miscellaneous alt-rock -- this category basically includes anything I found out about from listening to Sirius XM in the car.
- Nerd rock -- What kind of geek would I be if Jonathan Coulton wasn't on this list?
- Straight-up nostalgia -- I have an admittedly weird habit of listening to the same album (sometimes just one song) over and over for hours on end which was formed during all-nighters in high school. Motion City Soundtrack, Jimmy Eat World, and Weezer are my go-to's in this category.
- Soundtracks of all sorts -- *Chicago*, *Jurassic Park*, *Bastion*, *The Book of Mormon*, the Disney version of *Hercules*... again, basically anything that works on a repeat loop for ~3 hours.
- Pop -- don't make me list the artists. I've already told you I listen to Disney soundtracks; you can't possibly need more dirt on me. The general principle is that if you can dance to it, you can code to it.

Now, if you don't mind, I'm just going to sit at my desk and be super-embarrassed that my coworkers know what's in my headphones.

Written 4 Dec. 353 views. Asked to answer by William Chen.

Upvote

13

Downvote

Comment

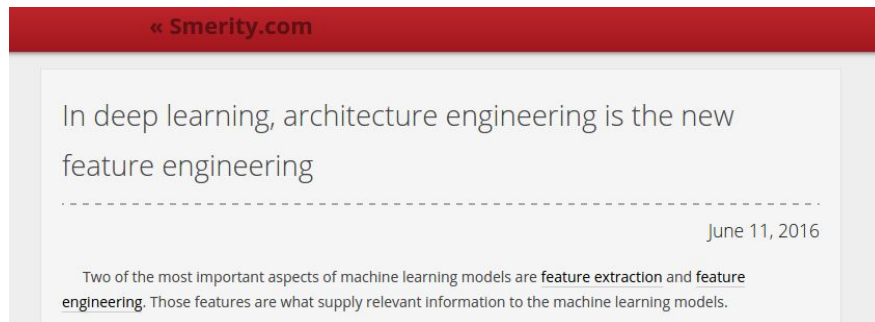
Share

...

● Properties of a well-behaved

ML feature:

- Reusable
- Transformable
- Interpretable
- Reliable



Deep Learning

NIPS'2015 Tutorial

Geoff Hinton, Yoshua Bengio & Yann LeCun



Deep Learning:
Automating
Feature Discovery

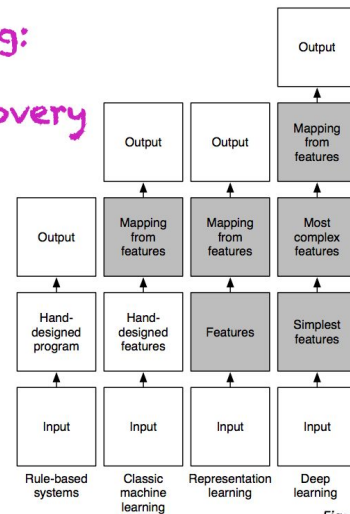
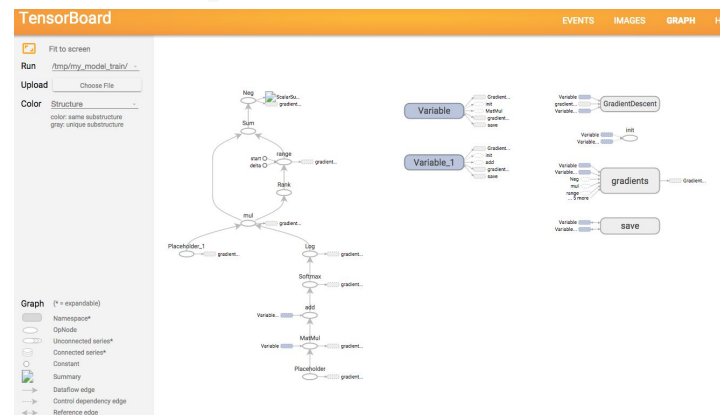
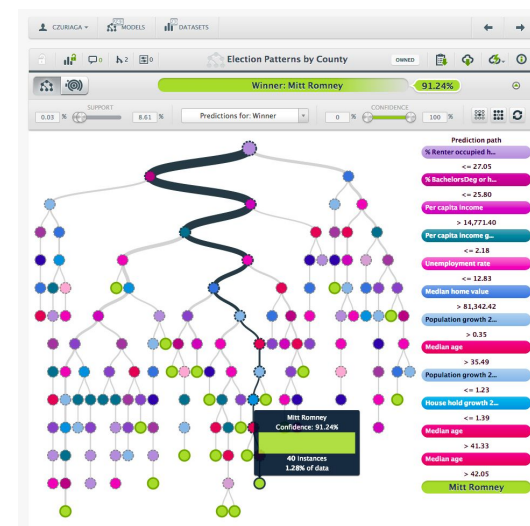


Fig. 1. Goodfellow

8. WHY YOU SHOULD CARE ABOUT
ANSWERING QUESTIONS
(ABOUT YOUR RECSYS)


Model debuggability

- Value of a model = value it brings to the product
- Product owners/stakeholders have expectations on the product
- It is important to answer questions to why did something fail
- Bridge gap between product design and ML algos
- Model debuggability is so important it can determine:
 - Particular model to use
 - Features to rely on
 - Implementation of tools



Model debuggability

- E.g. Why am I seeing or not seeing this on my homepage feed?

Feature Name		aid 14862324	aid 2546362
US	What is more dangerous, road or mountain biking?		
US			
OB	 Jack Rae, Gold medalist at British XC University Champs. President of UoBCC 2011-2012....		
US	Upvoted by Richard Henry · Vo Nghi Nguyen		
US	Encountering minor injuries : You'll get a lot more of that in cross country mountain biking . Brushing a tree, going over the bars and bruising your shoulder, scraping your legs... Maybe even fractu... (more)		
USER LONG HISTORY ACTION TYPE UPVOTE RATE BY STORY TYPE		0.0094589	0.0787334

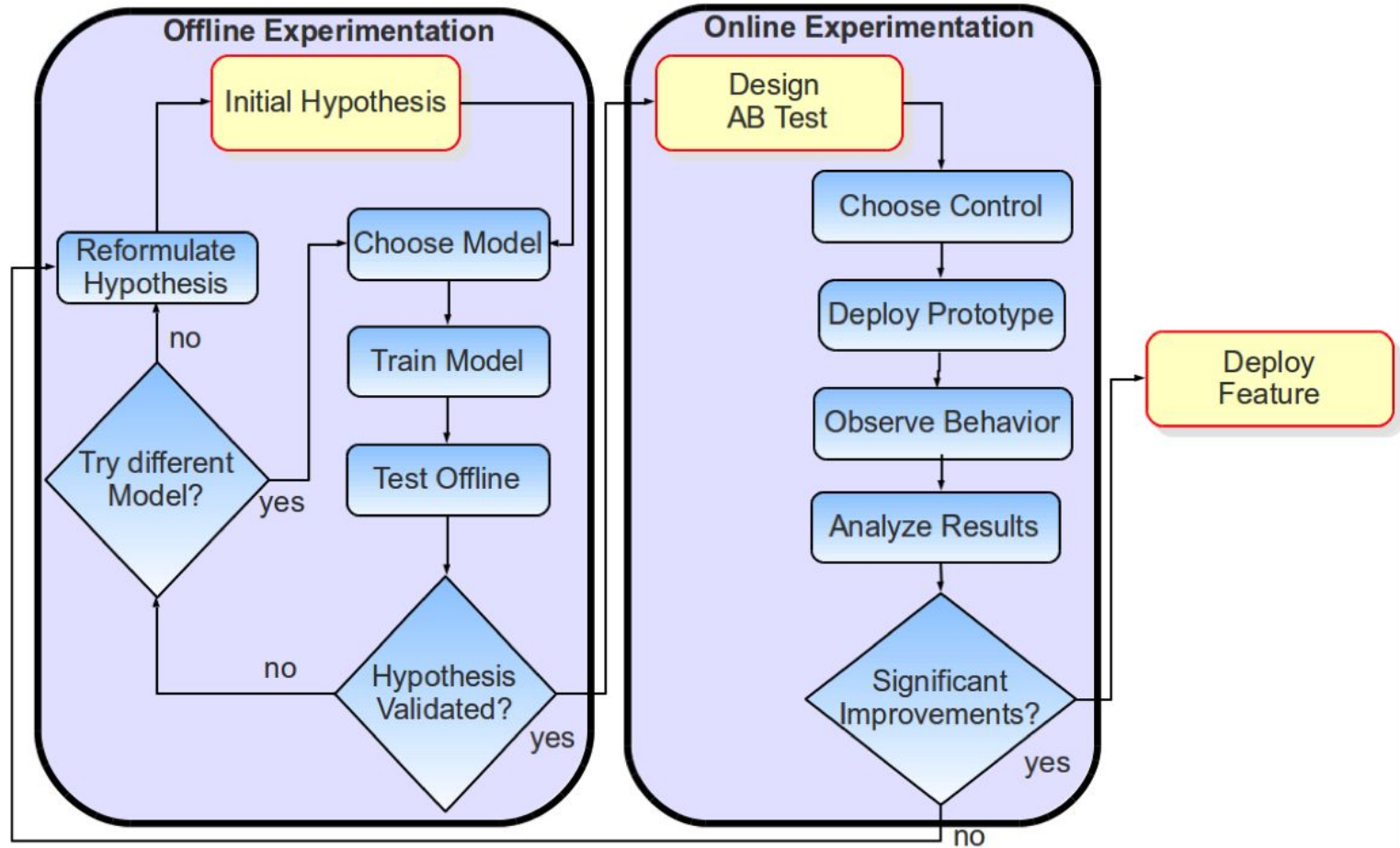
feed / feature analysis using score / feature analysis using model score

This table shows feature values for the debug story (using feedStory or debug_aid/qid above) and for the top 10 comparison stories from the same leaf node. For each comparison story, the color (and hover text) of a feature cell shows how the score of the debug story would change if feature values were swapped between the debug story the comparison story. Feature rows are sorted by the maximum absolute score gain among the comparison stories.

Feature Name		aid 14862324	aid 2546362	aid 2296
USER_I		0.0094589	0.2130526	0.213052
USER_I		0.0514545	0.2039045	0.203904
OBJEC'		8	None	7
OBJEC'		128263005100	70919435147759	7538566
USER_I		0.0648323	0.2112874	0.211287
USER_S		0	None	1
USER_I		0.0094589	0.0787334	0.078733
OBJEC'		0	0.3824919	0.245169
OBJEC'		0.1047419	None	None
NUM_R		1	None	None
USER S		0	None	1

9. DATA AND MODELS ARE GREAT. YOU KNOW
WHAT'S EVEN BETTER?
THE RIGHT EVALUATION APPROACH!

Offline/Online testing process

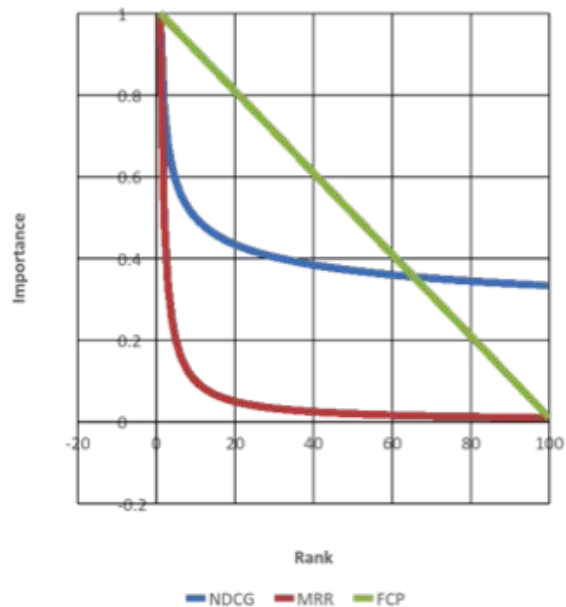


Executing A/B tests

- Measure *differences* in metrics across *statistically identical* populations that each experience a different algorithm.
- Decisions on the product always data-driven
- Overall Evaluation Criteria (OEC) = member retention
 - Use long-term metrics whenever possible
 - Short-term metrics can be informative and allow faster decisions
 - But, not always aligned with OEC

Offline testing

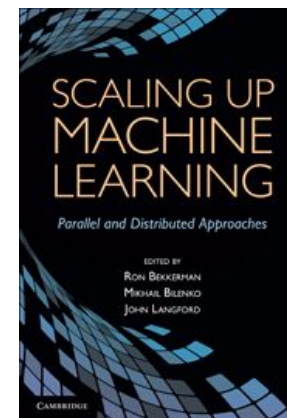
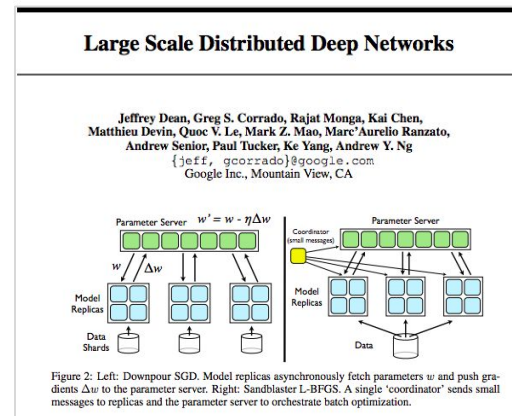
- Measure model performance, using (IR) **metrics**
- Offline performance = indication to make decisions on follow-up A/B tests
- A critical (and mostly unsolved) issue is how offline metrics correlate with A/B test results.



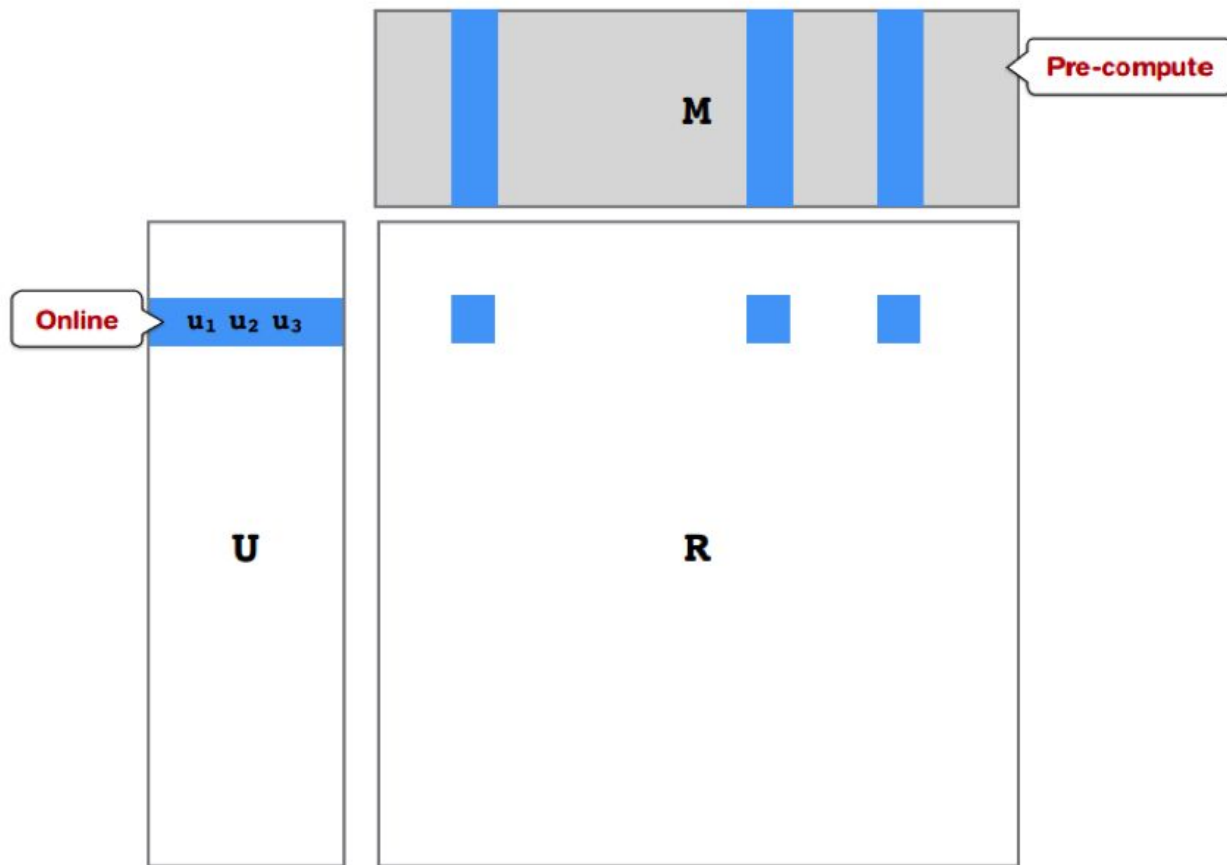
10. YOU DON'T NEED TO DISTRIBUTE YOUR
RECSYS

Distributing Recommender Systems

- Most of what people do in practice can fit into a multi-core machine
 - As long as you use:
 - Smart data sampling
 - Offline schemes
 - Efficient parallel code
- (... but not Deep ANNs)
- Do you care about costs? How about latencies or system complexity/debuggability?



Matrix Factorization Example



Conclusions

- Recommender Systems are about much more than just predicting a rating
- Designing a “real-life” recsys means paying attention to issues such as:
 - Feature engineering
 - Training dataset
 - Metrics
 - Experimentation and AB Testing
 - System scalability
 - ...
- Lots of room for improvement & research

Questions?



Xavier Amatriain (Quora)
xavier@amatriain.net



Deepak Agarwal (LinkedIn)
dagarwal@linkedin.com

Quora

Linked in