Recommender systems

Alaa BAKHTI - OCTO Technology

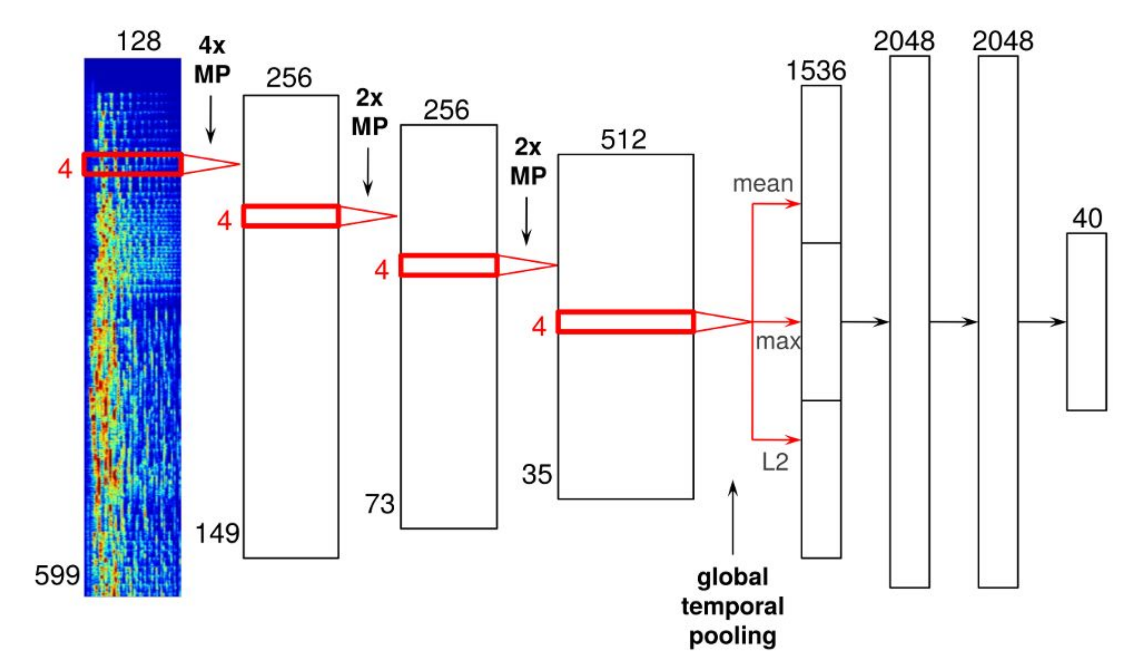
Deep neural networks in recommender systems

Motivation

- Nonlinear transformations (relu, tanh, etc)
- Representation learning
- Sequence modeling (next item to purchase)
- Flexibility with DL frameworks (changing architecture, loss function, etc)
- Speed of iteration

Deep content-based music recommendation

- Predicting listening preferences from audio signals by training a regression model to predict the latent representations of songs that were obtained from a collaborative filtering model.
- Maps the content of a song (audio) to its embedding (40 dimensions) in the collaborative filtering space.
- Tackles the song cold start problem : recommend a song to the right audience without having to rely on historical usage data.



[Blog post : Recommending music on Spotify with deep learning- Sander Dieleman]

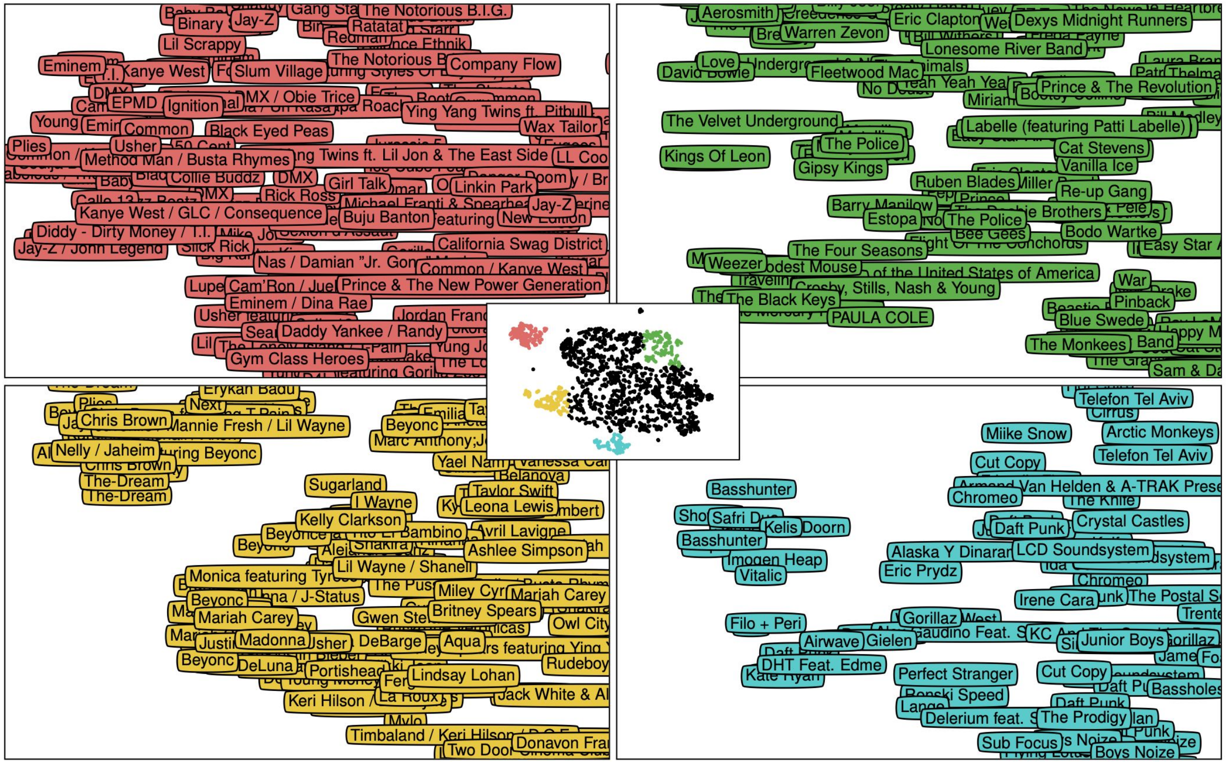


Figure 1: t-SNE visualization of the distribution of predicted usage patterns, using latent factors predicted from audio. A few close-ups show artists whose songs are projected in specific areas. We can discern hip-hop (red), rock (green), pop (yellow) and electronic music (blue). This figure is best viewed in color.

Source : Deep content-based music recommendation - Aaron van den Oord et al.

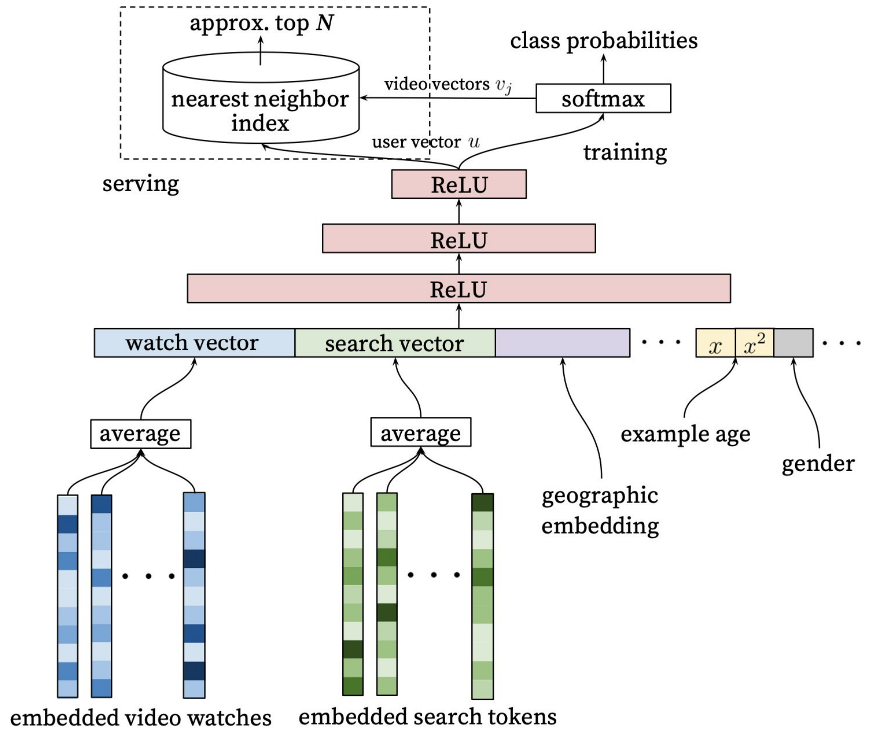


Figure 3: Deep candidate generation model architecture showing embedded sparse features concatenated with dense features. Embeddings are averaged before concatenation to transform variable sized bags of sparse IDs into fixed-width vectors suitable for input to the hidden layers. All hidden layers are fully connected. In training, a cross-entropy loss is minimized with gradient descent on the output of the sampled softmax. At serving, an approximate nearest neighbor lookup is performed to generate hundreds of candidate video recommendations.

[Deep Neural Networks for YouTube Recommendations - Paul Covington et al - 2016]()

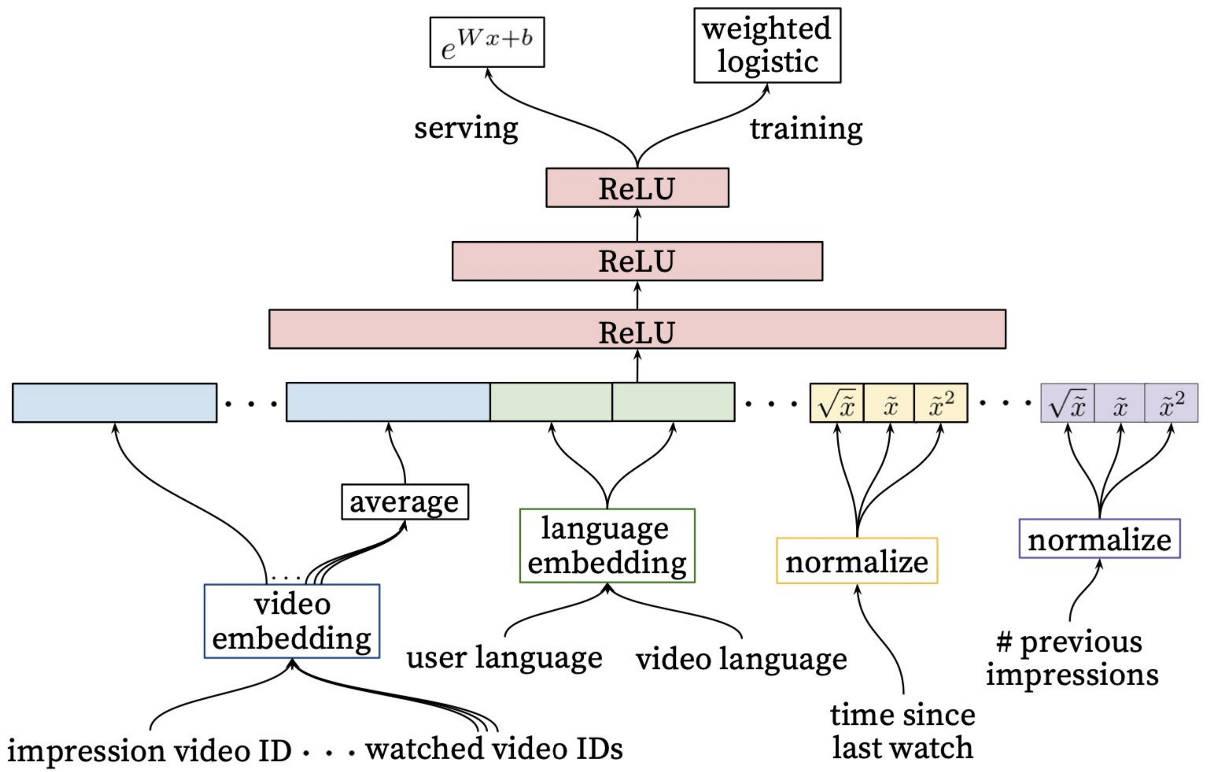
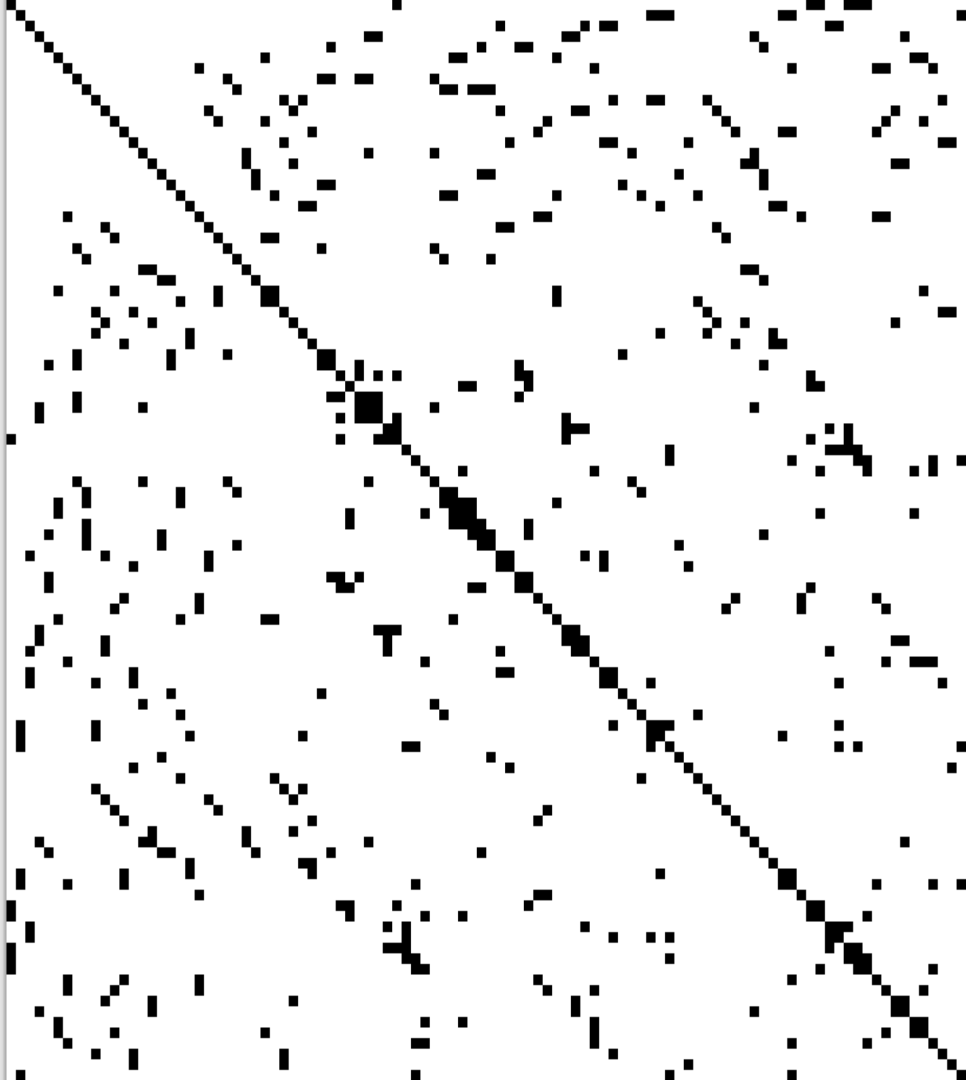


Figure 7: Deep ranking network architecture depicting embedded categorical features (both univalent and multivalent) with shared embeddings and powers of normalized continuous features. All layers are fully connected. In practice, hundreds of features are fed into the network.

[Deep Neural Networks for YouTube Recommendations - Paul Covington et al - 2016]

Limitations of Deep Learning

- ➔ Extreme sparsity
- ➔ Data-hungry
- ➔ Interpretability



How to evaluate a recommender system?

Motivation

Do the recommendations work? Do they increase sales? How much?

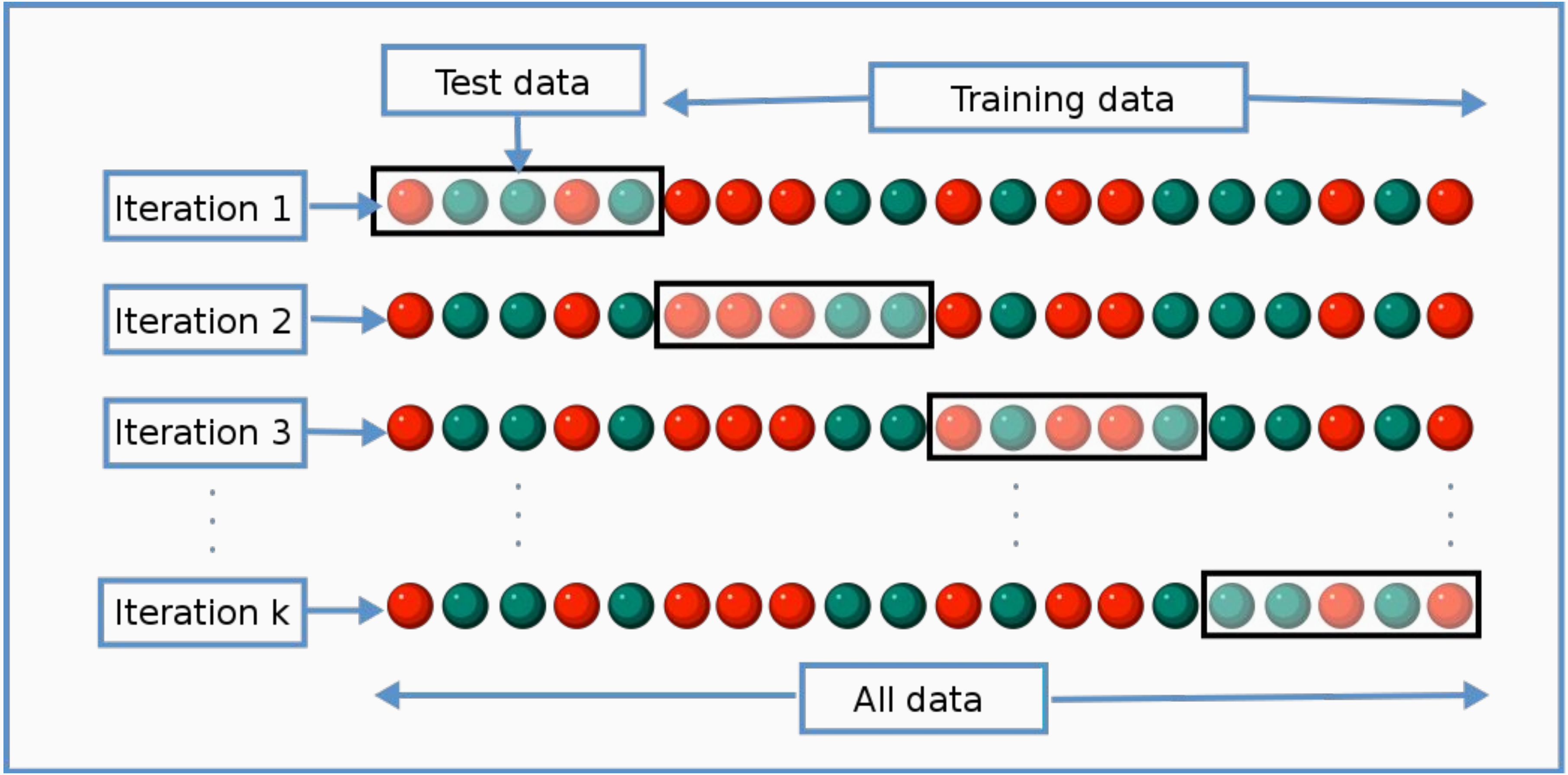
Which algorithm to choose for our application?

What are the best model parameters?

[Evaluating Recommender Systems: Choosing the best one for your business]

Offline evaluation

- Evaluate the recommender system before deploying it.
- Assumes that there will be no change in the user behavior between the data collection and the recommender system deployment.



K-fold cross validation - Author: Gufosowa - License: [CC BY-SA 4.0]

Offline evaluation metrics

Rating metrics

Are the recommendations relevant?

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)

- Penalizes you more when the predicted rating is way off penalizes you less when you are reasonably close.

Ranking metrics

Are the recommendations in the correct order?

- Precision @ k
  # of recommended items @k that are relevant / # of recommended items @k
- Recall @ k
  # of recommended items @k that are relevant / total # of relevant items
- Mean Average Precision @ k (MAP@k)
- Normalized Discounted Cumulative Gain @ k (nDCG@k)

You can find a benchmark of different algorithms on different datasets (MovieLens, Netflix, Million Song Dataset, etc) along with the papers code here.

Drawbacks

- Poor estimation of online performance.
- Does not correlate with the business metric.

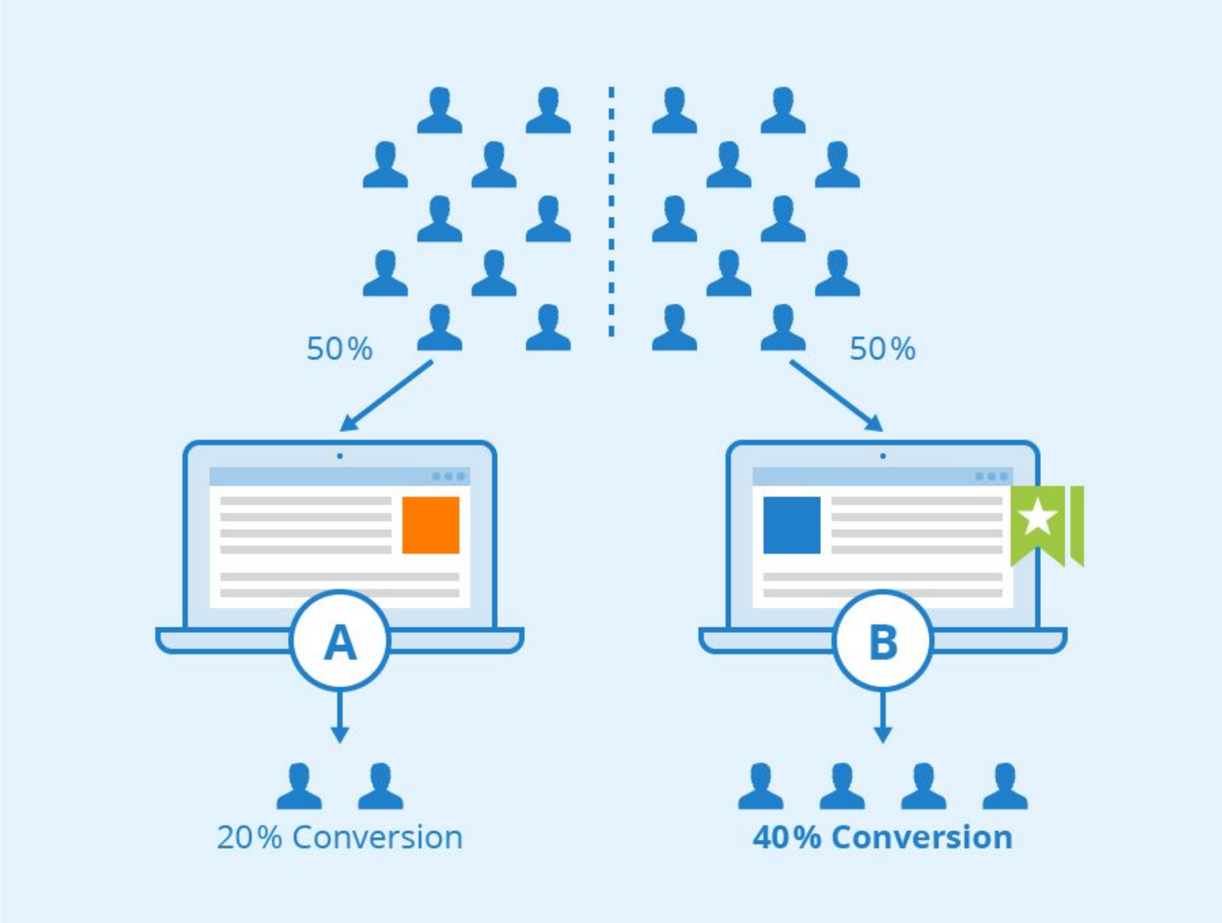
Online evaluation

Data-driven approach to determine if the new system variant improves the user experience.

Measures the acceptance rates of recommendations in real-world recommender systems using high-level interaction and business metrics.

KPIs for ROI measurement

- Click-Through Rate (CTR)
  - Did the user clicked on the recommended product?
- Conversion Rate (CR)
  - Did the user buy the recommended product?
  - Proportion of orders with recommendation.
  - Search for an unavailable product > recommend similar products > did the user purchased it?
- Customer lifetime value (CLV) : ultimate yet most difficult measure.
  - Customer retention.
  - Customer engagement.



AB testing - Author: Seobility - License: [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

AB testing

Definition

- Statistical method for comparing 2 or more variants to determine which one performances better.
- *"A user experience research methodology for testing a subject's response to variant A against variant B, and determining which of the two variants is more effective" - Wikipedia*
- Measures how the user behavior changes when he is interacting with different system variants.
- Can be interpreted as :
  - Implicit measure for user satisfaction.
  - Explicit measure of effectiveness of the new system.

Method

1. Split users in two or more groups : control (old system) and variation (new system).
2. Monitor yout target KPIs.

"Integrate few different recommender systems, divide your users into groups and put the recommender systems into fight." - Tomáš Řehořek

AB testing

- Used by web giants (Amazon, Facebook, Google, etc) to select the best recommendation engine.

Youtube & AB testing (source)

- Changes are validated only after doing an AB experiment.
- Each experiment can take from 1 week to months.
- Hundreds of different variables are measured with confidence intervals.

Challenges and downsides

- Identifying the right KPIs.
- Takes a lot of time to run an AB test experiment (weeks, months).
- Requires a lot coordination between multiple teams.

Scoring recommendation relevance is more an art than an exact science.

Annex

Suggested readings

Algorithms

- Learning the parts of objects by non-negative matrix factorization - Daniel D. Lee & H. Sebastian Seung
- Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. Computer, 42(8):30–37, 2009
- Deep Neural networks for YouTube Recommendations - Google
- Amazon.com Recommendations: Item-to-Item Collaborative Filtering - Greg Linden, Brent Smith, and Jeremy York - Amazon
- Recommending music on Spotify with deep learning - Sander Dieleman
- An Industrial-Strength Audio Search Algorithm - Avery Li-Chun Wang - co-founder of Shazam
- How does Shazam work? Music Recognition Algorithms, Fingerprinting, and Processing
- Spotify's Discover Weekly: How machine learning finds your new music

Scoring and ranking

- How Reddit ranking algorithms work - medium
- How Hacker News ranking really works: scoring, controversy, and penalties
- How Not To Sort By Average Rating - Evan Miller

AB testing

- A/B testing metrics — Evaluating the best metrics for your search - Algolia
- Using A/B testing to measure the efficacy of recommendations generated by Amazon Personalize | Amazon Web Services

Podcasts

- Gustav Soderstrom - Chief R&D officer @ Spotify - Lex Fridman Podcast
- Cristos Goodrow - Vice President of engineering @ YouTube - Lex Fridman Podcast

Other

- How Spotify Uses ML to Create the Future of Personalization
- Powered by AI: Instagram's Explore recommender system - Facebook AI
- YouTube Now: Why We Focus on Watch Time - Google Inc
- Netflix challenge