HPS

Lorenz Glißmann

Real-Time data analysis in education

Table of Contents

Do you have used any of these apps?

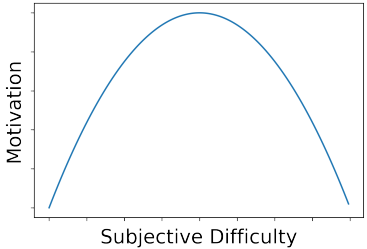
- Duolingo
- Drops
- Khan Academy
- Skillshare
- Coursera
- Udemy

Motivation for adaptive learning

Bored?

Overwhelmed?

Choose the right difficulty:

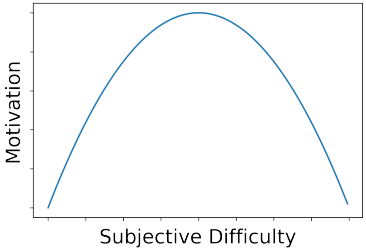


Motivation for adaptive learning

Bored?

Overwhelmed?

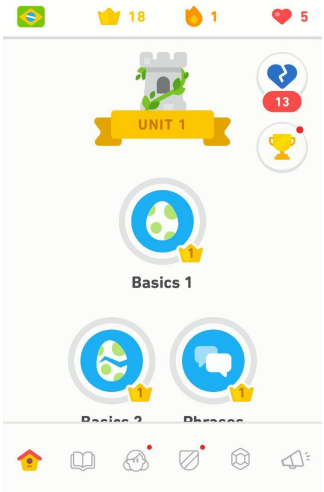
Choose the right difficulty:



We also learn the most in the "Zone of proximal development", where we can barely succeed, with help.

But for that we need a system that reacts and interacts with us!

Example: Duolingo

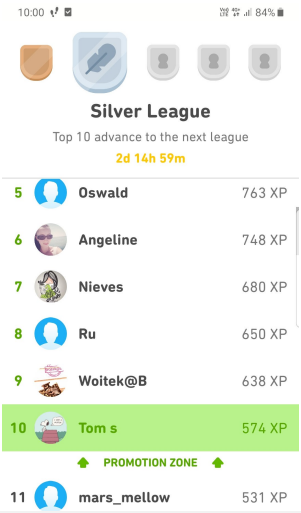


Example: Duolingo Leagues

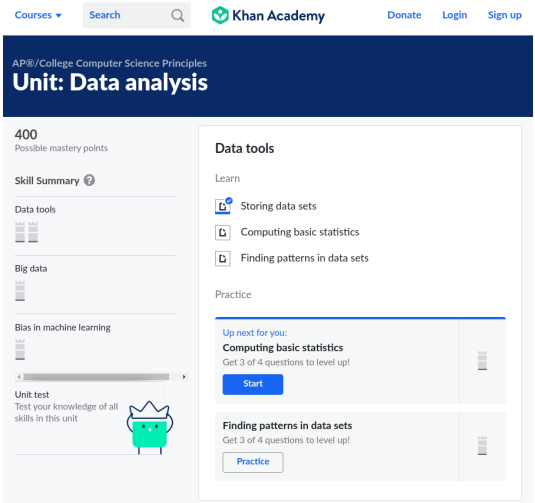
Leages are

- ... a competitive element to increase motivation.
- ... 50 learners compete for 1 week.
- ... of different rank: Bronze, Silver, Gold, ..., Diamond.

The top contenders advance to the next league.



Example: Khan Academy



- nonprofit Organization
- goal: "provide free education for everyone, everywhere"
- partially open source
- content under CC (Creative Commons) licenses

Khan Academy: mathematics course [Khan3]

Research Questions

Questions

1. How do we handle the data for so many students?
   - especially real-time data
2. How do we react to the learner in a high quality, intelligent way?
   - e.g. recommendations

Research Questions

Questions

1. How do we handle the data for so many students?
   - especially real-time data
2. How do we react to the learner in a high quality, intelligent way?
   - e.g. recommendations

We will look at representative examples from both research and practice.

Research Questions

Questions

1. How do we handle the data for so many students?
   - Typical application architectures (Khan Academy)
   - Backend design (Duolingo)
2. How do we react to the learner in a high quality, intelligent way?
   - Adaptively generate learning content (Duolingo)
   - Recommender Systems

Research Questions

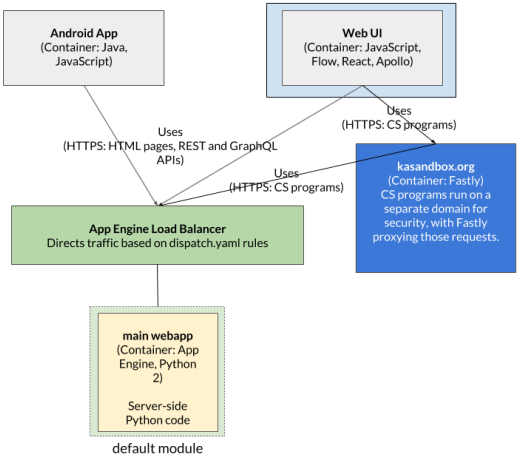
Questions

1. How do we handle the data for so many students?
   - Typical application architectures (Khan Academy)
   - Backend design (Duolingo)
2. How do we react to the learner in a high quality, intelligent way?
   - Adaptively generate learning content (Duolingo)
   - Recommender Systems

We will look at representative examples from both research and practice.

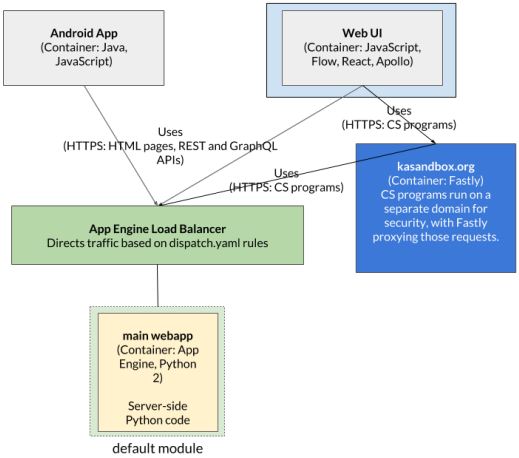
Table of Contents

Khan Academy: A typical architecture



Khan Academy Backend [Khan1]

Khan Academy: A typical architecture



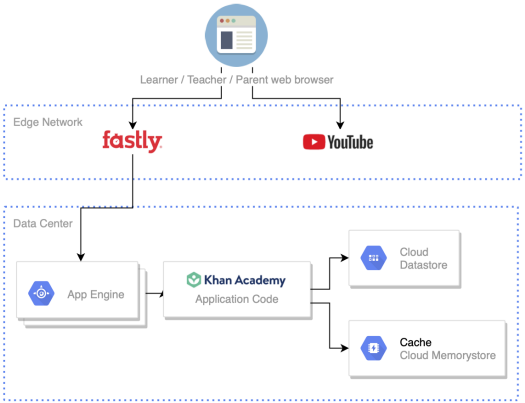
Khan Academy Backend [Khan1]

The Software Stack is typical for a modern (web-)application:

- clientside frontend
- backend
- load balancer
- (micro)services (kasandbox)
- HTTPS/GraphQL-APIs

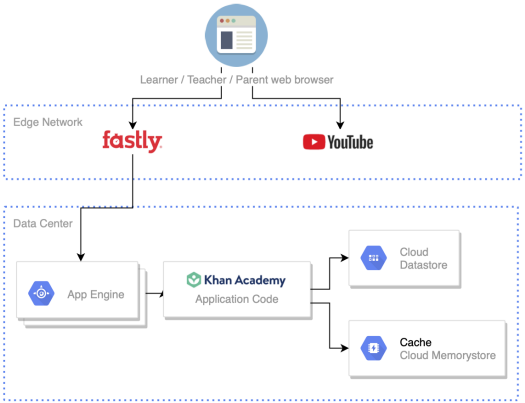
Moved backend from Python 2 to Go in April 2022.

So how do Khan Academy handle data?



Khan Academy Infrastructure [Khan2]

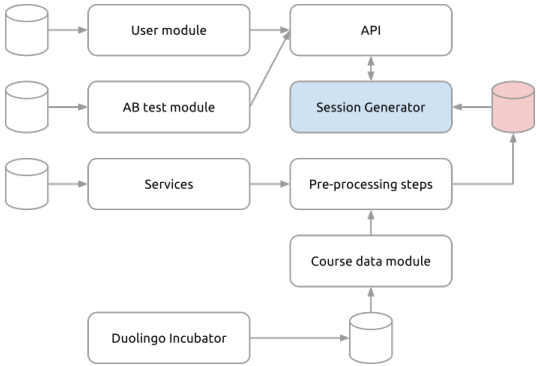
So how do Khan Academy handle data?



Khan Academy Infrastructure [Khan2]

- offloading videos to YouTube
- caching static content using fastly
- load balancing
- Plattform as a Service
- immutable data structures in the backend
- in-memory cache (memcache)

Duolingo: Backend Architecture



Duolingo backend schema, since 2017 [Duo6]

- Duolingo Incubator: Create/Edit Courses
- User module: all user specific information, e.g. last test runs
- Session generator: Generate individual test *adaptively*

Duolingo: Rewriting Session Generator

Challenges:

- individual and adaptive
- complex
- performance i.e. latency/speed, memory usage

Decisions:

- rewritten in Scala in 2017 (from Python)
- immutability and functional programming
- test extensively

Duolingo: Rewriting Session Generator

Challenges:

- individual and adaptive
- complex
- performance i.e. latency/speed, memory usage

Results:

- Latency reduction from 750ms to 14ms
- increased uptime
- smaller codebase

Decisions:

- rewritten in Scala in 2017 (from Python)
- immutability and functional programming
- test extensively

Duolingo: Rewriting Session Generator

Challenges:

- individual and adaptive
- complex
- performance i.e. latency/speed, memory usage

Results:

- Latency reduction from 750ms to 14ms
- increased uptime
- smaller codebase

Decisions:

- rewritten in Scala in 2017 (from Python)
- immutability and functional programming
- test extensively

But how does it work?

Table of Contents

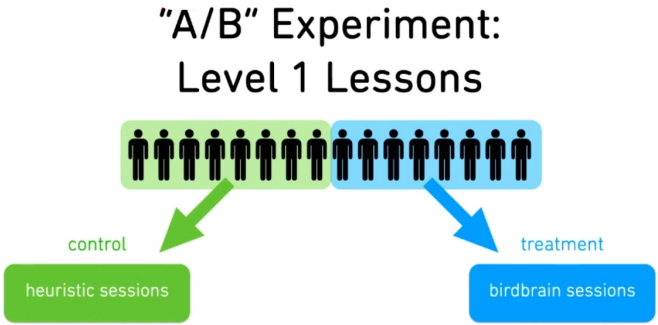
Duolingo: Birdbrain

"Birdbrain" is a new (2020) machine learning based learner model.

- Goal: Choose test items of optimal difficulty
- Approach: predict success probability for a test item
- Method: logistic regression model
- Data:
  - (individual) knowledge tracing
  - collectively experienced difficulty
  - probability of forgetting based on past learning experience

The session generator uses Birdbrain to generate a test of fitting difficulty.

Duolingo: Birdbrain results



Results of a randomized blinded controlled trial [Duo5]

Duolingo: Adaptive Features

- Difficulty
- Previous learning experience
- Forgetting
- Time, frequency and selection of push notification reminders
- Immediate feedback for grammatical errors, based on first order logic rules

Do you see problems with Duolingo's approach?

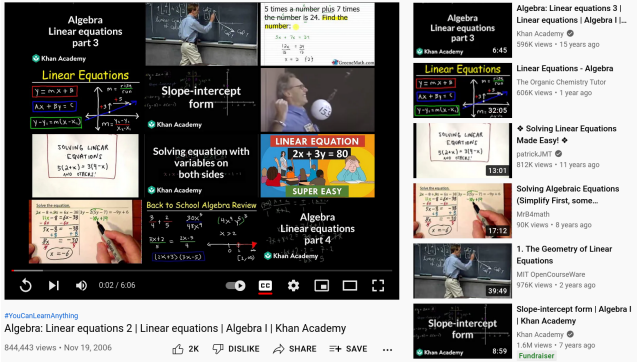
Duolingo: Adaptive Features

- Difficulty
- Previous learning experience
- Forgetting
- Time, frequency and selection of push notification reminders
- Immediate feedback for grammatical errors, based on first order logic rules

Do you see problems with Duolingo's approach?

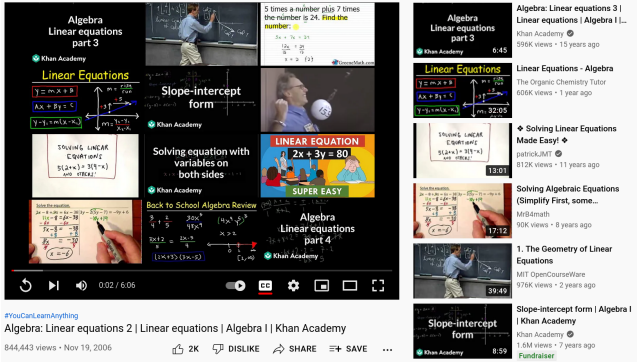
It takes away meaningful choices and *forces* the learner to do what the system thinks is best.

Youtube’s recommender system



An example showing YouTube’s recommender systems [You1]

Youtube’s recommender system



An example showing YouTube’s recommender systems [You1]

Why do we need recommender systems?

Recommender Systems

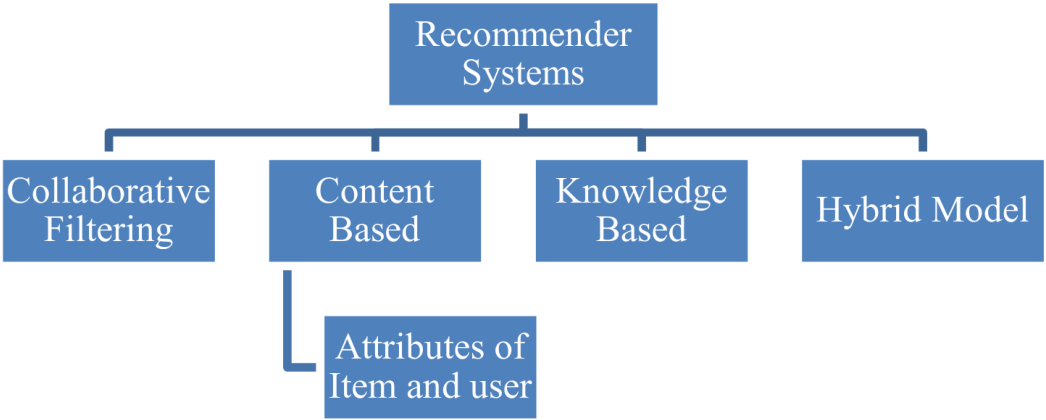
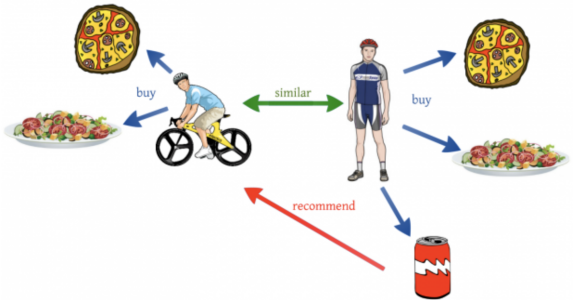


Figure: Overview over different recommender system methods [Rec1]

Recommender Systems: Collborative Filtering Algorithms

Collaborative Filtering

"collaborative filtering is a method of **making automatic predictions** (filtering) about the interests of a user by collecting preferences [**from other users**] (collaborating)" [Rec2]



User-based Collaborative Filtering [Rec4]

Collaborative Filtering

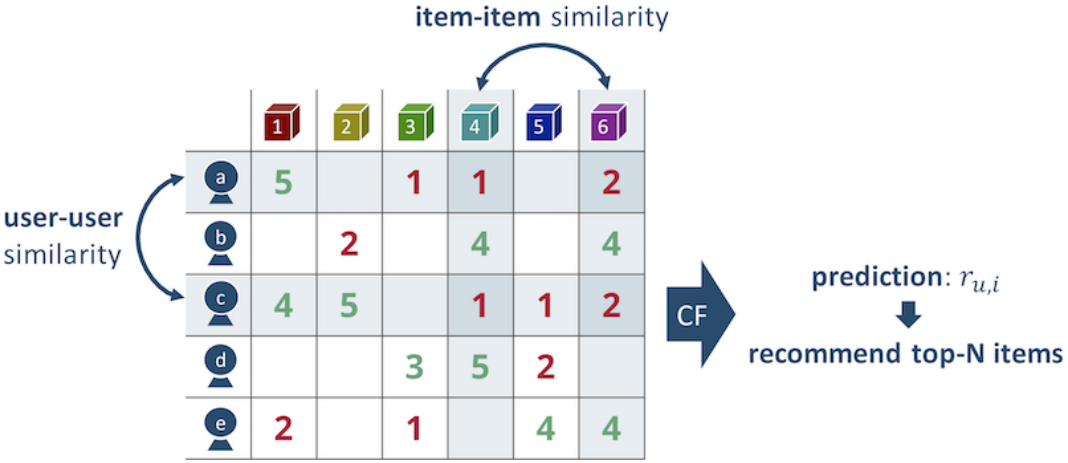


Figure: Item-based vs. User-based similarity collaborative filtering [Rec3]

Table of Contents

Summary

Answers

1. How do we handle the data for so many students?
   - *distributed* architecture
   - *caching* and *immutability*
   - pre-process as much as we can
   - make the dynamic parts of the backend *efficient*
   - Cloud / Infrastructure as a Service help
2. How do we react to the learner in a high quality, intelligent way?
   - model the learner
   - model the content
   - use experiments to evaluate viability
   - decide for the user OR make recommendations

Sources

Duo1 "Methods for Language Learning Assessment at Scale: Duolingo Case Study", Portnoff, Lucy et al., International Conference on Educational Data Mining (EDM), 2021

Duo2 "Learning how to help you learn: Introducing Birdbrain!" https://blog.duolingo.com/learning-how-to-help-you-learn-introducing-birdbrain/, 2020-10-07

Duo3 "How we learn how you learn", https://blog.duolingo.com/how-we-learn-how-you-learn/, 2016-12-14

Duo4 "Personalized feedback through smart tips", https://blog.duolingo.com/smart-tips-effective-feedback/, 2021-08-12

Duo5 "Improving Language Learning and Assessment with A.I.", https://youtu.be/HQyGblZ8tx8, 2022-05-19

Duo6 "Rewriting Duolingo's engine in Scala", https://blog.duolingo.com/rewriting-duolingos-engine-in-scala/, 2017-01-31

You1 "Algebra: Linear equations 2 | Linear equations | Algebra I | Khan Academy", https://youtu.be/DopnmxeMt-s, 2006-11-19

Rec1 "Recommender System for Big Data in Education", Surabhi Dwivedi, ELELTECH, 2017

Rec2 "Collaborative Filtering", https://en.wikipedia.org/w/index.php?title=Collaborative_filtering&oldid=1083148203, 2022-05-19

Rec3 "Collaborative Filtering" https://takuti.github.io/Recommendation.jl/latest/collaborative_filtering/, 2022-05-19

Rec4 https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0

Khan1 "Beating the odds: Khan Academy's successful monolith→services rewrite" https://blog.khanacademy.org/beating-the-odds-khan-academys-successful-monolith%E2%86%92services-rewrite/

Khan2 "How Khan Academy Successfully Handled 2.5x Traffic in a Week" https://blog.khanacademy.org/how-khan-academy-successfully-handled-2-5x-traffic-in-a-week/

Khan3 "Khan Academy: Unit: Data analysis", https://www.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101, 2022-05-18

Summary

Answers

1. How do we handle the data for so many students?
   - *distributed* architecture
   - *caching* and *immutability*
   - pre-process as much as we can
   - make the dynamic parts of the backend *efficient*
   - Cloud / Infrastructure as a Service help
2. How do we react to the learner in a high quality, intelligent way?
   - model the learner
   - model the content
   - use experiments to evaluate viability
   - decide for the user OR make recommendations

Discussion

- Why did I choose Khan Academy and Duolingo?
  - much used plattforms
  - share information about their internals
  - most scientific literature on the topic was of very low quality