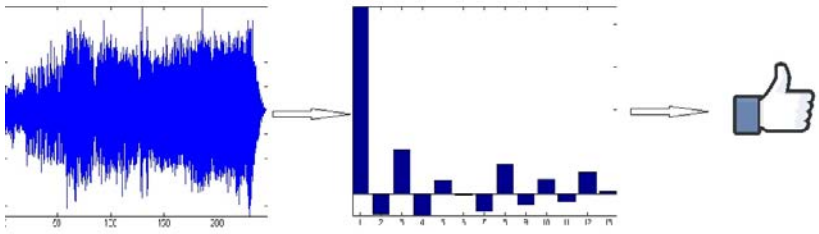
Proactive Caching of Music Videos based on Audio Features, Mood, and Genre

ACM MMSys 2017, June 20-23, Taipei, Taiwan

TECHNISCHE UNIVERSITÄT DARMSTADT

Source: https://www.digitaltrends.com/music/universal-music-mobile-streaming-app/



Feature Action

Christian Koch, M.Sc.
Ganna Krupii, M.Sc.
Prof. Dr. David Hausheer

Prof. Dr.-Ing. Ralf Steinmetz
KOM - Multimedia Communications Lab

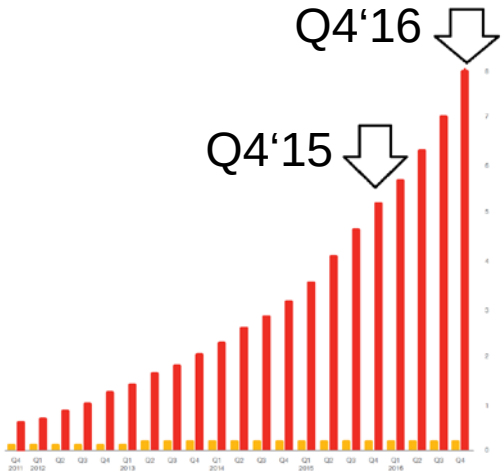
Music Videos Cause a Large Amount of Traffic

TECHNISCHE UNIVERSITÄT DARMSTADT

Mobile Traffic [1]

- Grew 55% between 2015 and 2016 → exponential trend
- Video is predicted to constitute 70% of mobile data in 20
- Higher qualities, e.g. 4K, 3D and more services, e.g., Sp YouTube, Netflix, Twitch, Twitter, etc. support this trend

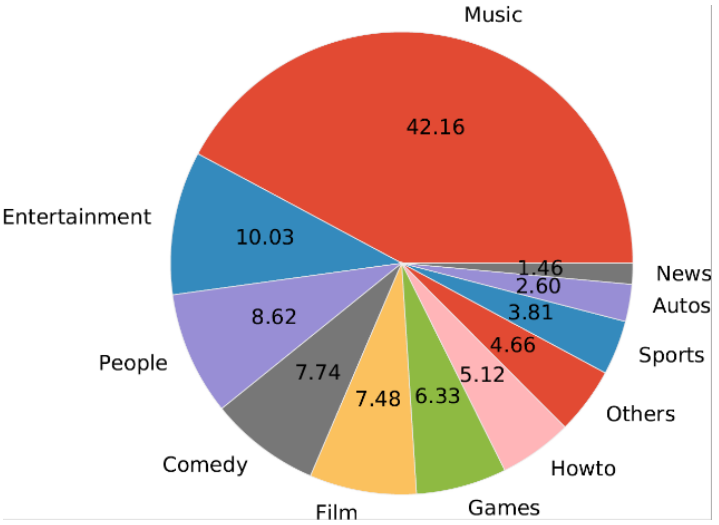
→ Even though 4G is becoming increasingly available, a gap between the user demand and the available bandwidth is predicted [22]



The Case of YouTube

- Largest single source of traffic in North America
- ~82% of the users watch music videos
- ~42% of YT requests in mobile networks are caused by music videos (2014)

→ Music videos are worth investigating to alleviate networks



Existing Solutions for Network Load Reduction

TECHNISCHE UNIVERSITÄT DARMSTADT

Focusing on Video Content

Approach	Content Type	User Satisfaction	Suitable for Music Videos?
ISP-internal Multicast	live IPTV	+	-
Prefetching	static, user-specific	+	o
CDNs	live and on-demand	+	+

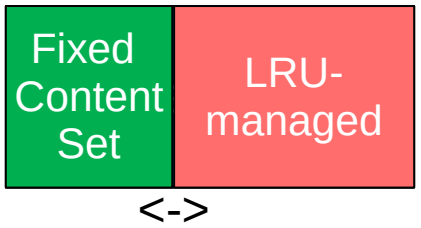
→ CDNs operating outside and potentially inside ISPs are most promising.
Yet, there is not much work on optimization for music video content having
a) different popularity live cycles
b) Requested repeatedly

Proactive Caching using Music Features

TECHNISCHE UNIVERSITÄT DARMSTADT

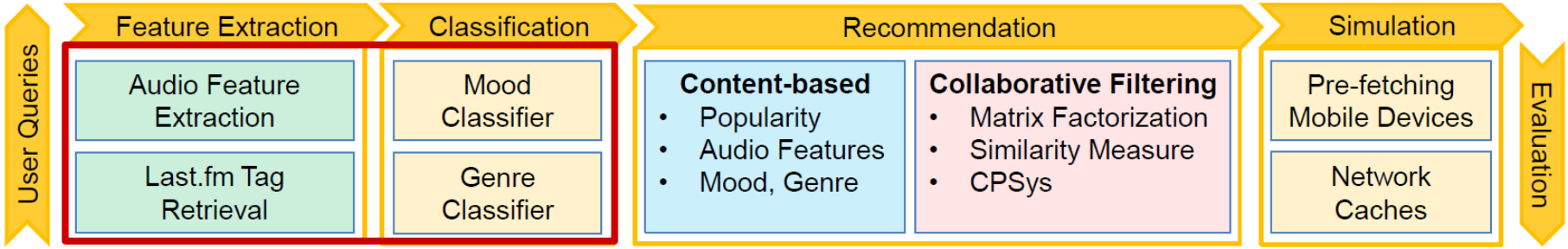
Proactive Caching

- A certain share of the cache is managed reactively, e.g., by LRU
- The remaining share is filled proactively and kept static



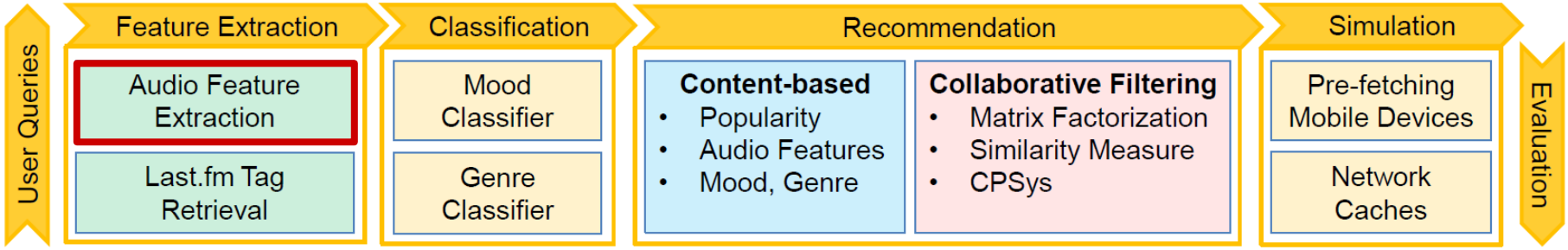
Challenges

1. Collecting information
   - User-related information might be large or even illegal to trace
   - Content has many dimensions, suitable categories required
2. Determining the content which is most popular
   - within the next time, e.g., couple of hours till a day



For $\frac{13{,}553}{44{,}704} \sim = 30\%$ tracks the mood and genre could be determined using last.fm

Audio Feature Extraction



For >40k music videos, features are extracted

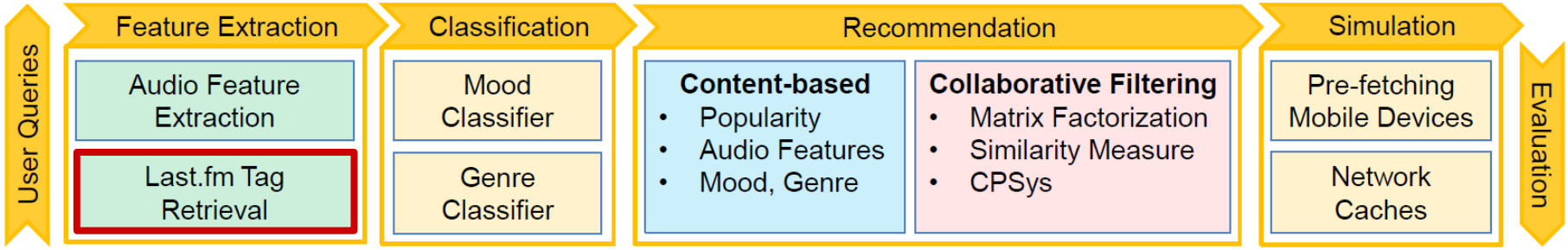
- If the music track has a duration >60s: take seconds 30-60 as a sample
- Else take 0-60s as a sample

Low Level Audio Features (aka. Descriptors)

- Zero-crossing rate,
  RMS energy,
  Spectral flux, etc.
- Further statistics like: mean, median, std., etc. of each feature

Framework Name	#Descriptors	Last update
MPEG-7 descriptors [11]	17	2004
Marsyas [25]	30	2015
jAudio [18]	40	2009
MIRtoolbox [14]	55	2014

Last.fm Tag Retrieval



Last.fm allows users to assign tags to the music tracks offered

For >13k music videos, last.fm tags could be used

1. YouTube video title cleaning: Remove strings like “official clip”
2. If a title contained a “-” the preceding part is considered as artist, the latter the track name
3. The resulting tags are only used if at least 50 people assigned this tag to a track

FROM THE ALBUM

Music

296,546 listeners

pop · dance · electronic · madonna · female vocalists

Music Label Extraction

TECHNISCHE UNIVERSITÄT DARMSTADT

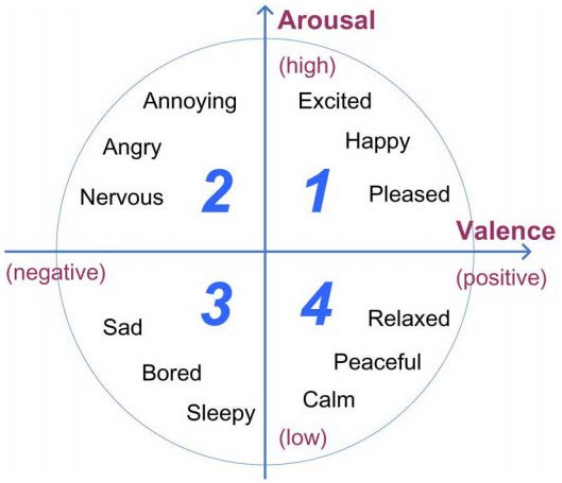
Tags can be ambiguous and noisy

- Manual association of tags and mood & genre categories

10 Genre Categories used

4 Mood Classes used

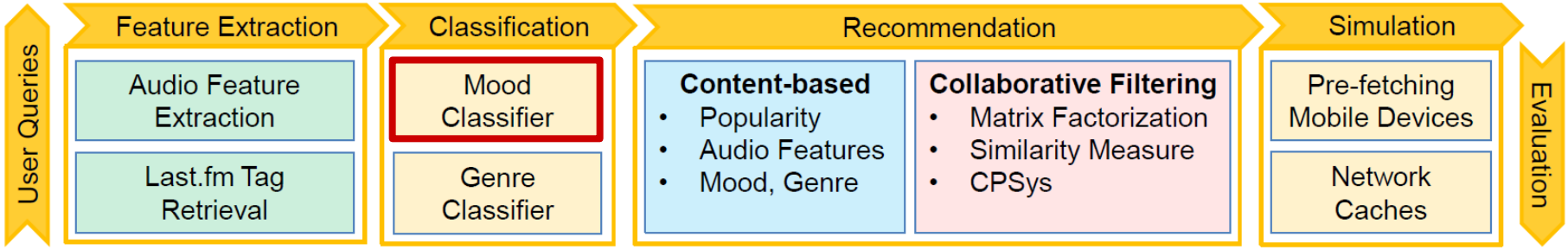
Happy	Sad	Angry	Relaxed
happy	sad	angry	relaxed
energetic	nostalgia	aggressive	calm
positive	depressive	banger	downtempo
fun	bittersweet	passion	chillout
cheerful	sentimantal	quirky	dreamy
humorous	melancholic	annoying	longing
feel good	dramatic	gangsta rap	spiritual



Thayer‘s mood model [27]

→ How can Mood be classified?

Music Mood Recognition



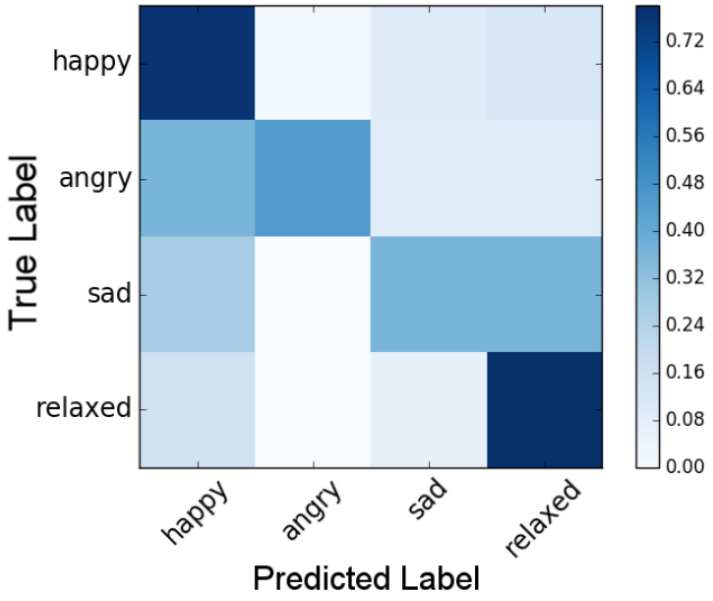
Related Work	Mood representation	Utilized features	Classification/ regression algorithm	Accuracy
[15]	Angry, sad, happy and relaxed category	Dynamics, rhythm, spectral, harmony, etc.	Support Vector Machine	Polynomial SVM 54% accuracy
[17]	Arousal/valence emotional plane	Spectral centroid, loudness, sharpness, timbral width, volume, spectral dissonance, tonal dissonance, pure tonal, etc.	Multiple linear regression (MLR), support vector regression (SVR) and AdaBoost	SVR 58.3 % for arousal and 28.1 % for valence
[16]	Emotion state transition model	Beat, tempo, intensity, spectral centroid, flatness, spread, flux, MFCC, etc.	SVM and one-class SVM Classification	Polynomial one-class SVM 87.78% accuracy
[19]	***Happy, sad, angry and relaxed category***	***MFCCs, loudness, spectral flatness, flux, roll off, centroid, kurtosis, etc.***	***SVM, trees, random forest, KNN, GMM***	***Polynomial SVM 90.44% accuracy***

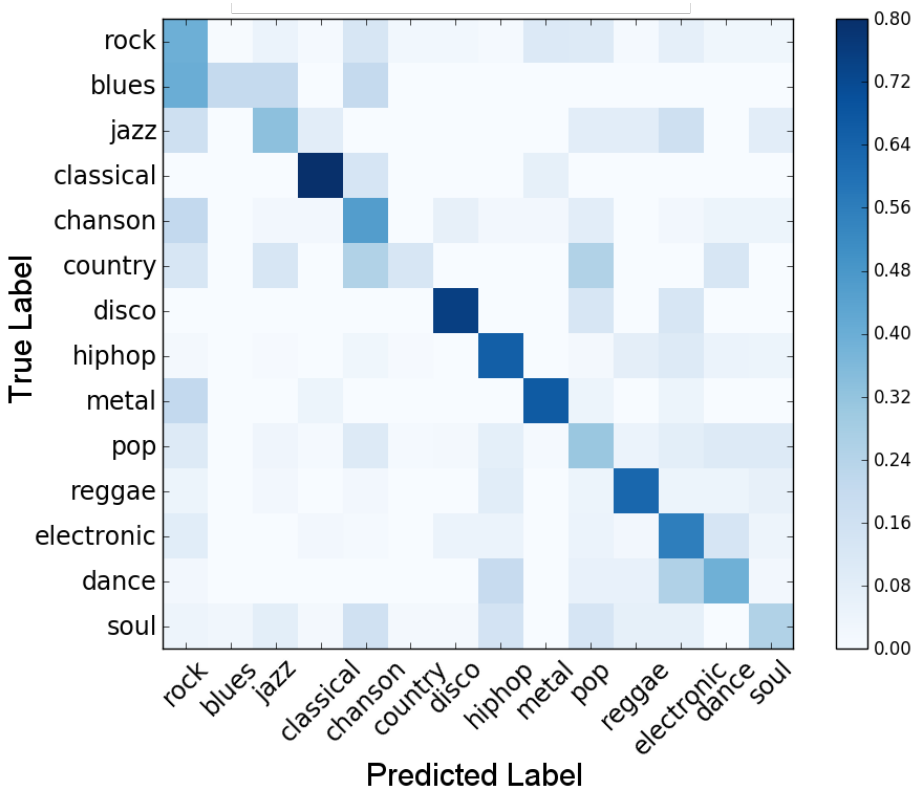
Mood and Genre Classification

TECHNISCHE UNIVERSITÄT DARMSTADT

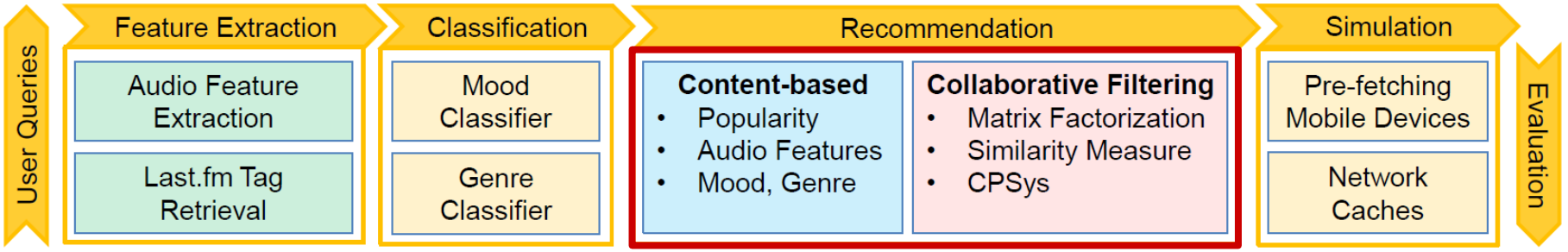
Tested several hundred SVM parameter settings

- 10-fold CV, grid search for hyper-parameter determination
- **Best**: kernel: rbf, accuracy: 0.64(mood), 0.5(genre) → **used to determine miss. labels**
  - Mood: 64% accuracy (random: 25%)
  - Genre: 50% accuracy (random 7%)





Recommendation for Proactive Caching



Two types of Recommender Systems

1. Content-based

- Uses only content-related information
- Privacy preserving
- Policy Classes
  - *Popularity*
  - *Time-aware caching*
  - *Audio features*

2. Collaborative Filtering

- Relies on user-item matrix
- User activities need to be stored
- Policy Classes
  - *Matrix factorization*
  - *User similarity*
  - *Audio feature similarity*
  - *CPSys[5]*

Content-based Recommendation

TECHNISCHE UNIVERSITÄT DARMSTADT

Popularity (NW)

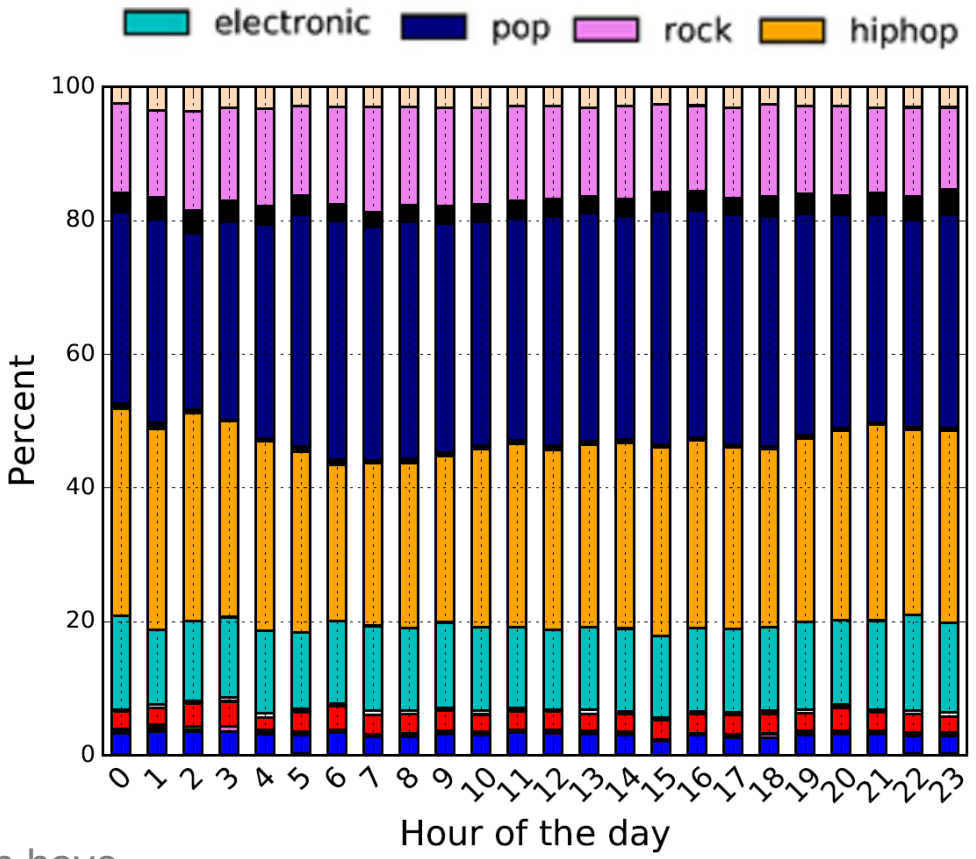
- Backwards-oriented: Recommend what was popular in the past
- Does not consider new videos
- Easy to implement
- Policies: *Popularity*

Time-aware Caching (NW)

- Using time-feature correlations, e.g., pop music in the morning, jazz in the evening
- Policies: *Mood*, *Genre*

Audio Features (UT)

- Music videos similar to the ones which have been watched more than once are selected
- Policies: *Feature*, *Feature Vector*



Collaborative Filtering (User Behavior-based Caching)

TECHNISCHE UNIVERSITÄT DARMSTADT

Matrix Factorization

- Uses Apache Spark (Alternating Least Squares)
- /w and w/o explicit rating: % of a video watched, i.e., 2 x 50% = 1

User Similarity

- *Similarity Measure*: Videos from similar users are weighted by the user-similarity
- *Aggregated Similarity*: The union of all users' *Similarity Measure* is used

Audio Feature Similarity

- Content from other users with similar low level feature mean & variance
- Policy: *Feature Range*

CPSys

- A mobile video prefetching system designed for non-music videos
- Candidates are selected by majority of similar users

Evaluation Scenarios

TECHNISCHE UNIVERSITÄT DARMSTADT

1. In-network Caching

- All ~700k users considered for trace-driven simulation
- Users randomly assigned to
  a. 1 cache (intra-ISP scenario)
  b. 5 caches (CDN scenario)

2. Client-side Caching

- Cache size depends on avg. #requests per day but is limited
- **Users with constantly high demand are selected for proactive caching**
  - requested at least 2 videos in 7 consecutive days within the 2 weeks

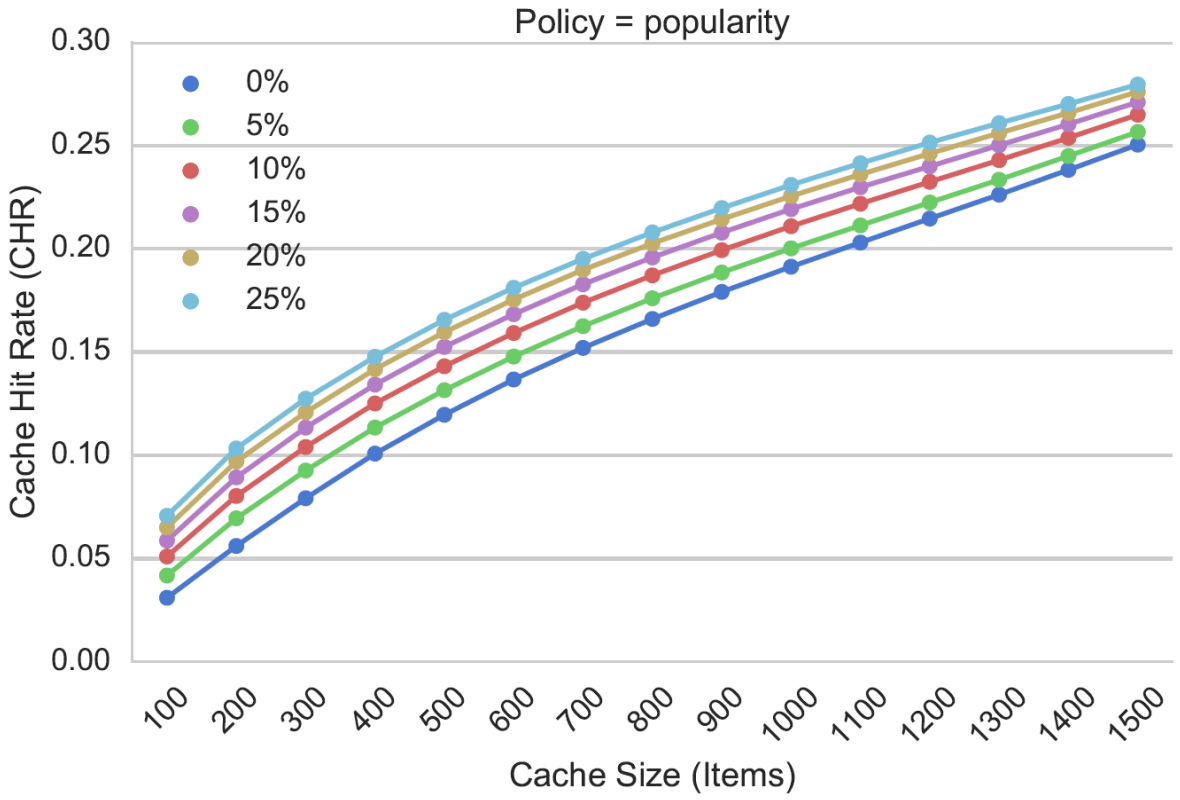
➔ 5,351 users constitute 1.64% of all users and 15.6% of total requests

Evaluation: a. Network Scenario (CDN)

TECHNISCHE UNIVERSITÄT DARMSTADT

1 Cache (intra-ISP Scenario)

- All user requests are send to one cache
- An increasing proactive cache share increases CHR, flattens at 25%

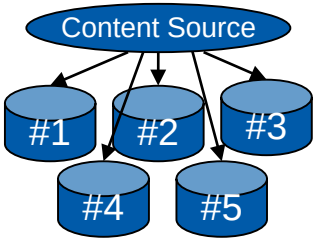


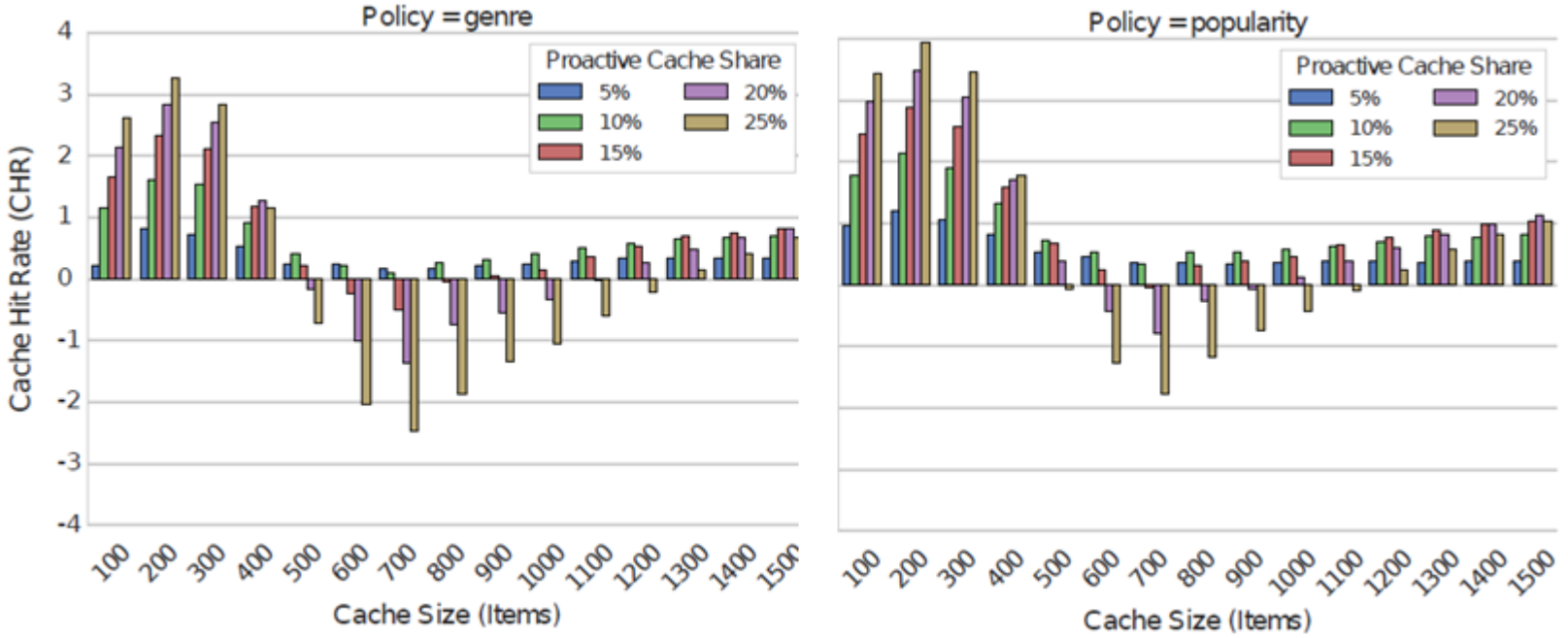
Evaluation: b. Network Scenario (CDN)

TECHNISCHE UNIVERSITÄT DARMSTADT

5 Caches (intra-ISP Scenario)

- Users are randomly assigned to one of the caches
- 5% proactive cache always increases CHR
- Popularity achieves the highest CHR at cache size 200: 55.1%



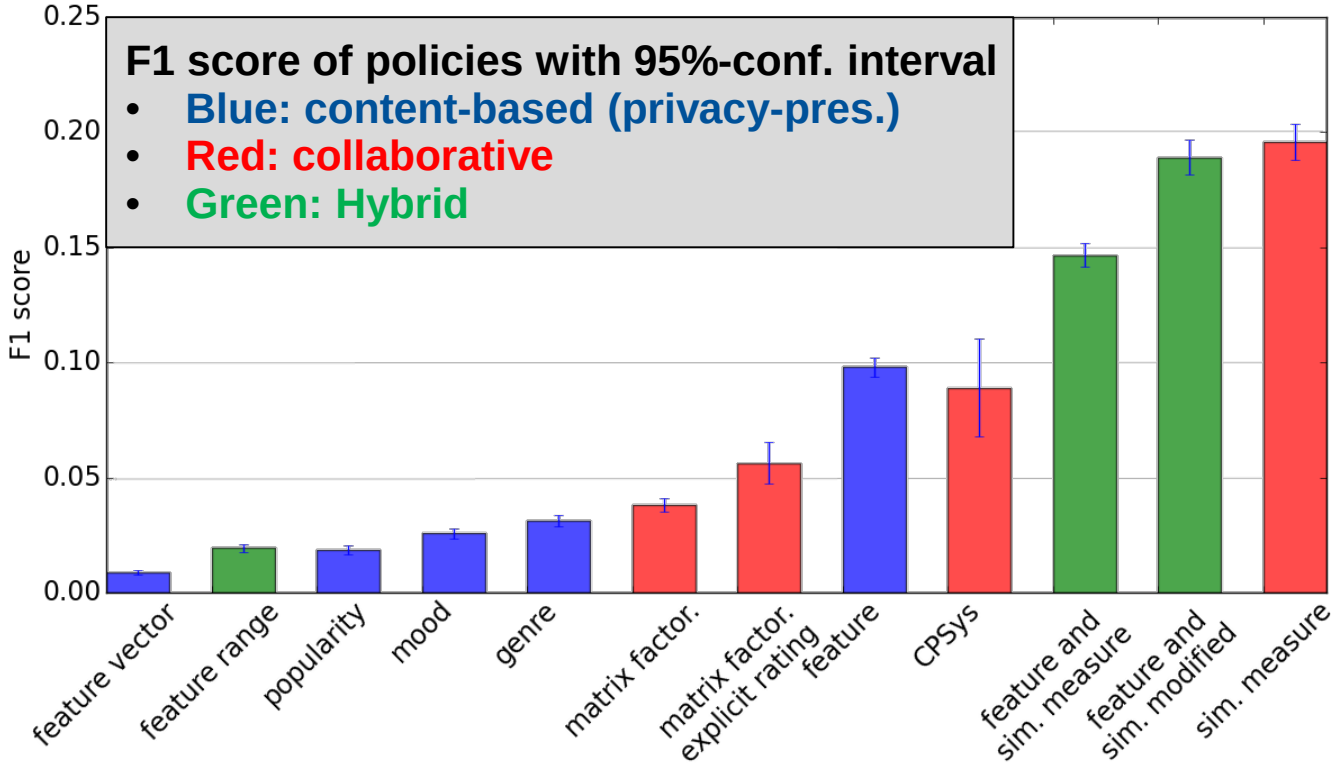


Differences between LRU and Proactive Caching

Evaluation: Client Device Prefetching

TECHNISCHE UNIVERSITÄT DARMSTADT

- Overall quite low F1 scores: Sufficient if downloaded via Wi-Fi
- Feature, the content-based policy is superior to matrix factorization, however pure similarity measure performs best
- F1 score up to 35% possible by choosing frequency and data in a clever way



Summary, Conclusion and Future Work

TECHNISCHE UNIVERSITÄT DARMSTADT

Summary

- Feature extraction and classification for mood and a genre labeling using last.fm
- Proactive caching policies using music features and user-item-matrix have been evaluated
- Evaluation on a two weeks dataset considering all of their ~10 million requests

In-Network Caching

→ For small & large caches, proactive caching increases cache hit rate up to 4%

Client-side Caching

→ Feature policy (privacy-preserving) achieves >2x F1 score than Matrix Factorization. Moreover, similarity measure is ~4x as efficient

Future Work

- Considering Video-Segments, policies using deep learning, cache hierarchies

Questions & Contact

TECHNISCHE UNIVERSITÄT DARMSTADT

Department of Electrical Engineering
and Information Technology
Multimedia Communications Lab - KOM

TECHNISCHE UNIVERSITÄT DARMSTADT

Christian Koch, M.Sc.

Christian.Koch@KOM.tu-darmstadt.de
Rundeturmstr. 10
64283 Darmstadt/Germany
www.kom.tu-darmstadt.de

Phone +49 6151 16-20894
Fax +49 6151 16-29109