Video recommendations Based on Visual Features Extracted with Deep Learning

By: Tord Kvifte

Supervisor: Assoc. Prof. Dr. Mehdi Elahi

Outline

1. Background	2. Methodology	3. Results	4. Conclusion
Research context, problem, and proposed solution	Feature extraction, recommendation technique, framework demo, and study design	Exploratory Analysis, Recommendation Quality, User Study	Conclusion of results and future plans

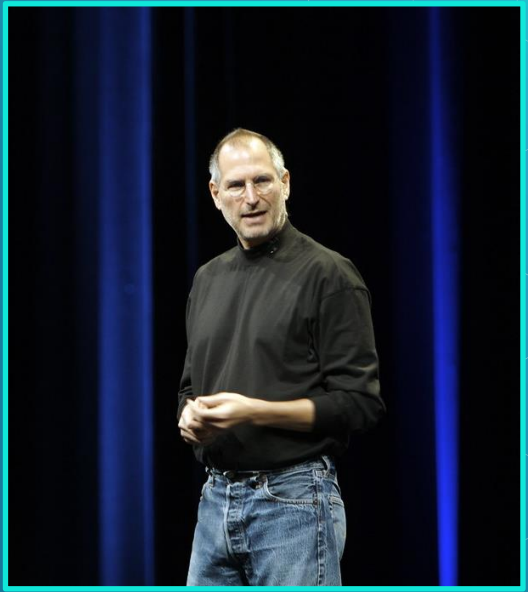
1

Background

Research context, problem, and proposed solution

Choice overload


SEAL OF THE PRESIDENT OF THE UNITED STATES

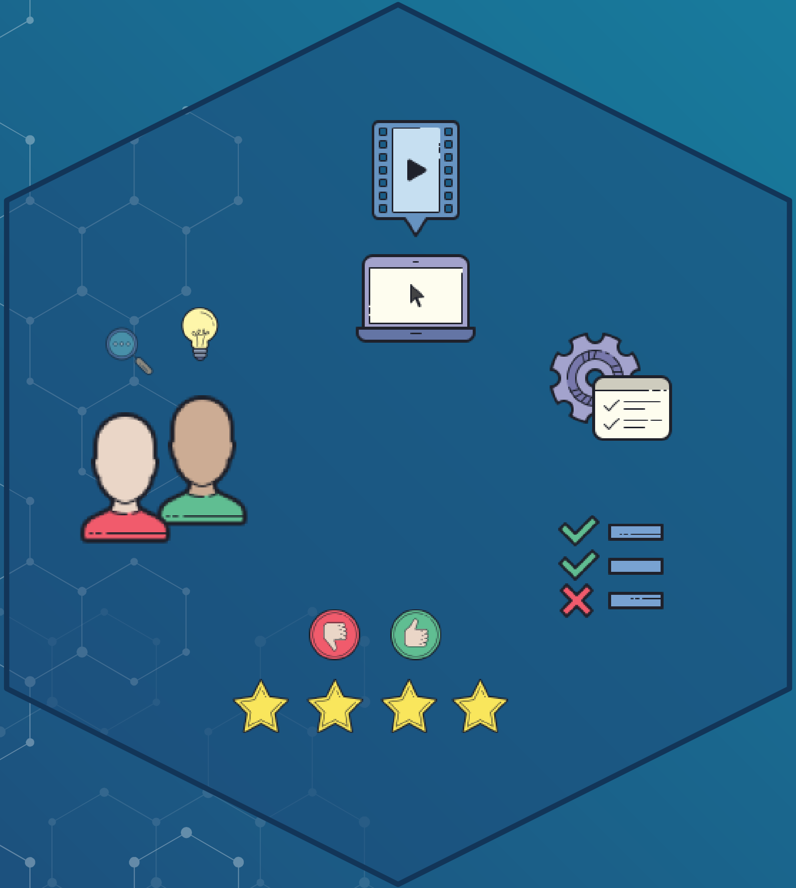


Hours of video uploaded to YouTube **every minute**:

500+

Number of titles in the Netflix library:

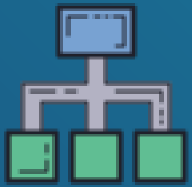
5,000+



Recommender Systems (RSs)

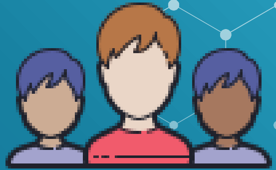
Tools that help users discover items they may like.

Problem



Recommendation algorithms

Dependency on manually created data

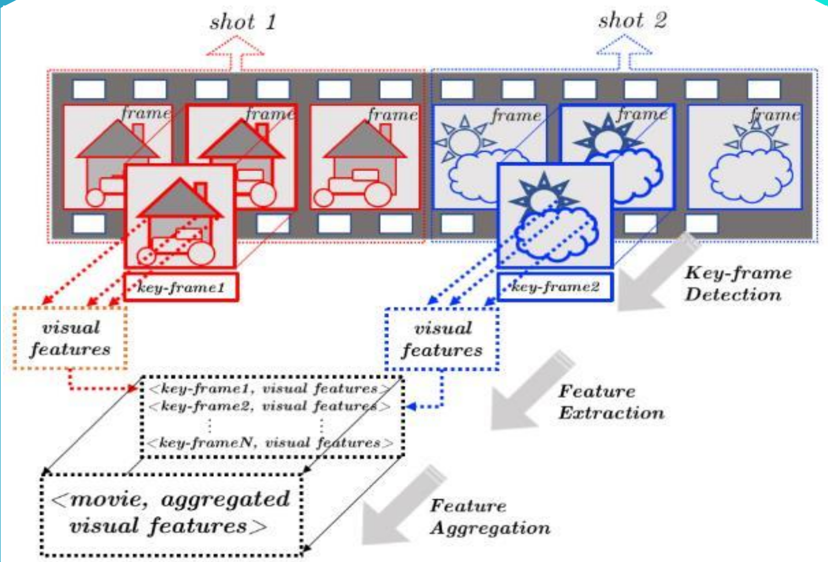


Evaluation of RSs:
High algorithmic performance ≠ good user-experience

Visual Features

- **Automatic** feature extraction

- Demonstrated capabilities.
  - Starke, Willemsen, and Trattner (2021)
  - Messina et. al. (2019)
  - Deldjoo et. al. (2016)



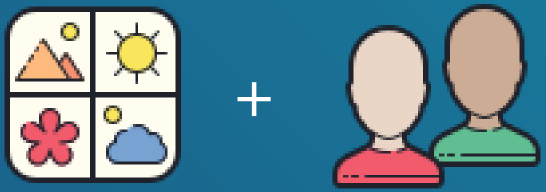
Levels of features

High-level (semantic)

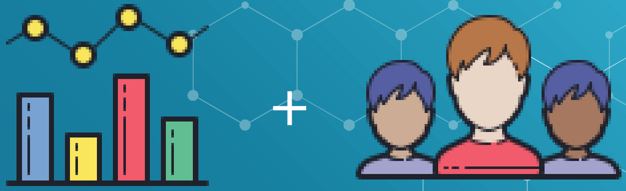
Mid-level (syntactic)

Low-level (stylistic)

Approach



- Deep Learning-based visual features
- Novel hybrid technique



- Offline evaluation
- User-centric evaluation (N=150)
- Evaluation framework
- Baselines:
  tag, genre (manual),
  subtitles (automatic).

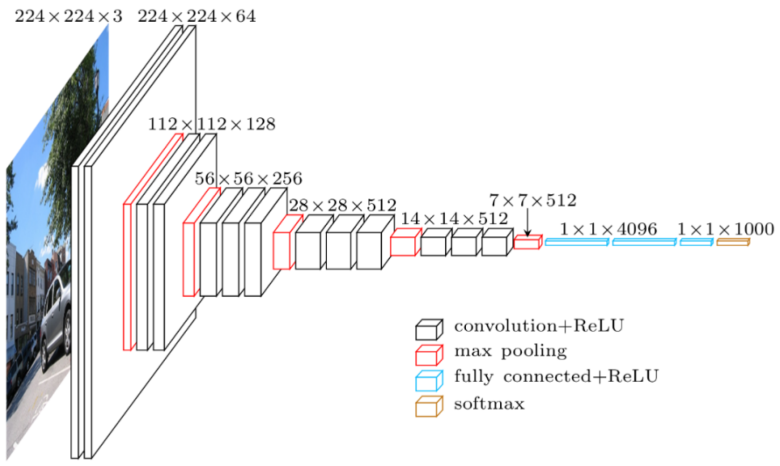
2

Methodology

Feature extraction, recommendation technique, framework demo, and study design

Feature Extraction

- Key frames
  - 12,875 movies
- VGG-19 CNN image classification
  - Trained on ImageNet
- Subtitles – CineSub
  - 3,405 movies



Predicted label: ‘liner’, Confidence: 0.56

Predicted label: ‘pay-phone’, Confidence: 0.56

Predicted label: ‘spotlight’, Confidence: 0.56

Datasets

Visual Features

DeepCineProp-f

TF-IDF

DeepCineProp-c

Confidence

Subtitle Features

CineSub

TF-IDF

Train/test

Interactions

MovieLens10M

80% | 20%

Manual Features

Genre

MovieLens10M

Tag

MovieLens10M

Recommendation model (Kula, 2015):

Latent representation of user *u* and item *i*:

$$q_u = \sum_{j \in f_u} e_j^U \qquad q_i = \sum_{j \in f_i} e_j^I$$

The scalar bias term of user *u* and item *i*:

$$b_u = \sum_{j \in f_u} b_j^U \qquad b_i = \sum_{j \in f_i} b_j^I$$

Predictions produced by:

$$\hat{r}_{u,i} = f(q_u \cdot p_i + b_u + b_i)$$

Where dot f· is given by:

$$f(x) = \frac{1}{1 + \exp(-x)}$$

Loss functions:

Warp loss function:

$$Err_{\mathrm{WARP}}(\mathbf{x}_i, y_i) = L[rank(f(y_i|\mathbf{x}_i))]$$

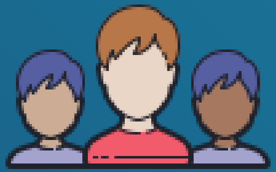
BPR loss function:

$$\min_{\Theta} \sum_{(u,i,j):(u,i)\succ(u,j)} f_{uij}(\Theta) + \mathcal{R}_{uij}(\Theta)$$

Logistic loss function:

$$\min_{U,M,C} \sum_{i}^{n} \sum_{j}^{m} [w_{ij}(p_{ij} - \langle U_{i*} M_{j*} \rangle)^2 + \frac{\lambda}{n}||U_{i*}||^2 + \frac{\lambda}{m}||M_{i*}||^2]$$

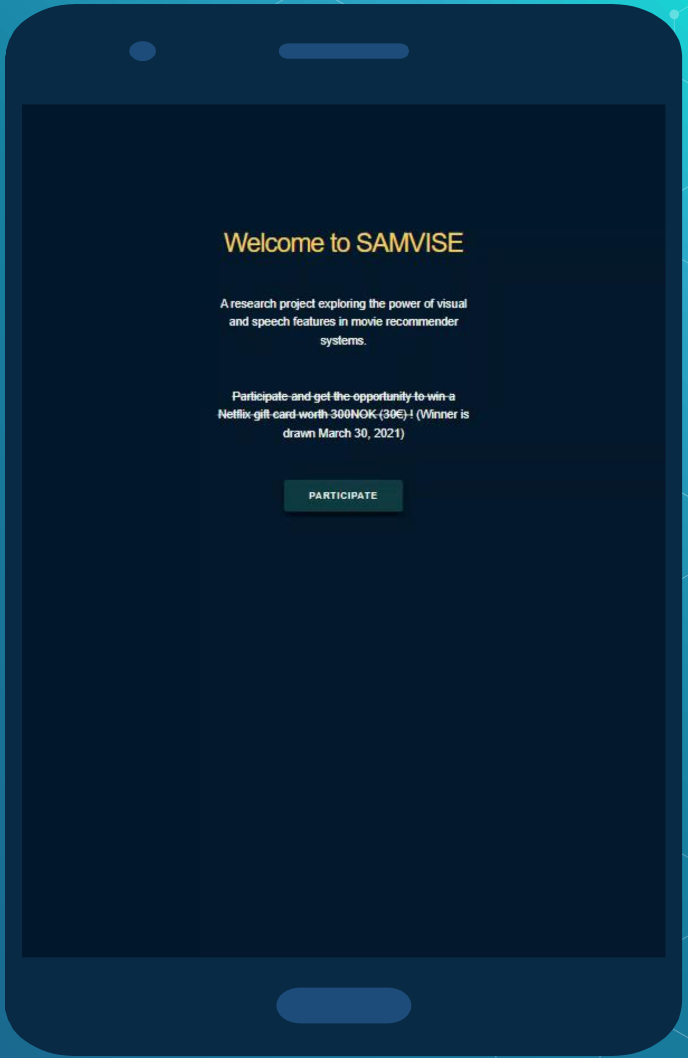
User Study



- 150 participants
- Voluntary + crowdsourcing
- 28 nationalities, 104 native English speakers

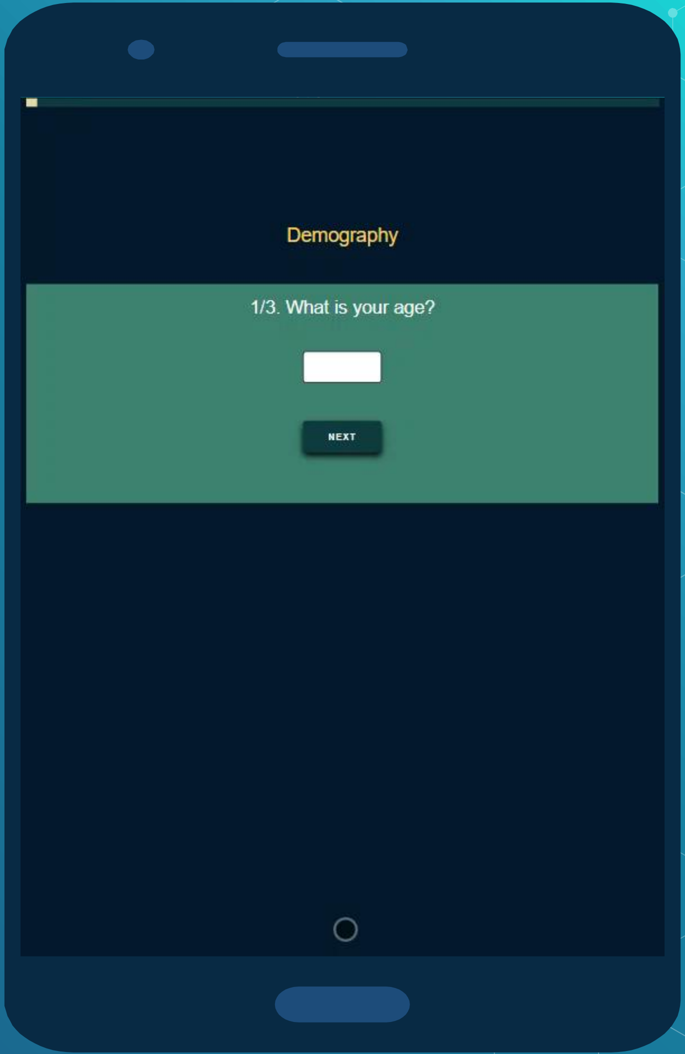
- Recommendation quality metrics:
  - Accuracy
  - Diversity
  - Personalization
  - Satisfaction
  - Novelty
- Usability evaluation
  - System Usability Scale (SUS)

Demo of SAMVISE Evaluation Framework



USER CHARACTERISTICS

The user answers demographic and personality questions.



PREFERENCE ELICITATION I

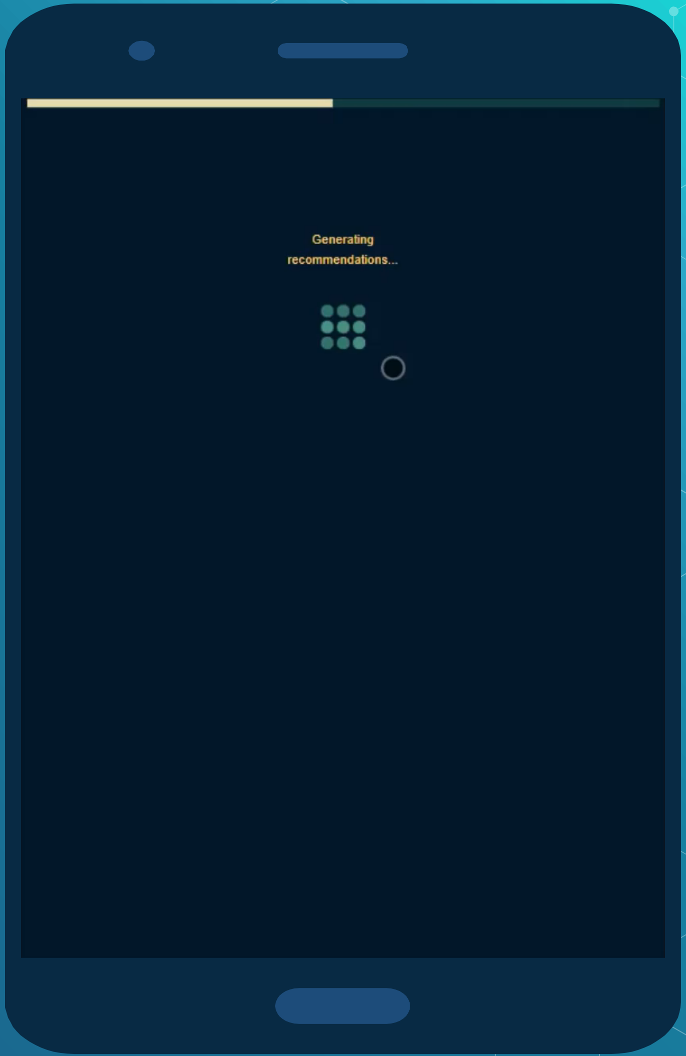
The user choose movies to rate. Filtering options include genre, decade, rating, and popularity.

PREFERENCE ELICITATION II

The user provides ratings for selected movies. Possibility to watch trailer and read info.

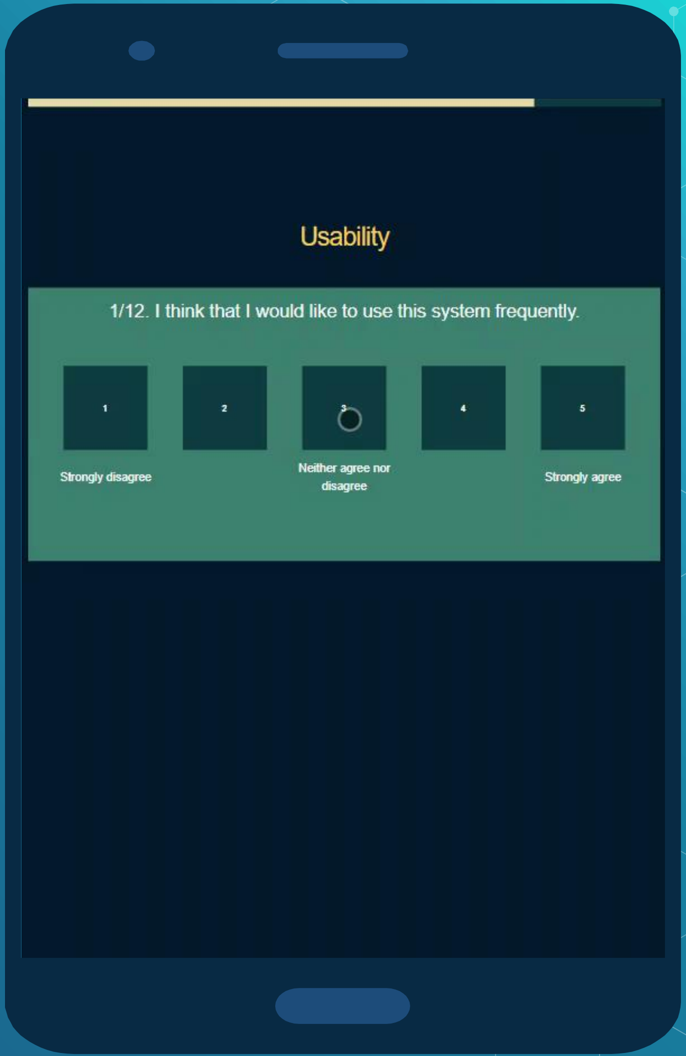
RECOMMENDATION EVALUATION

The user responds to questions by comparing the quality of 2 separate recommendation lists.



USABILITY EVALUATION

The user responds to the questions of the System Usability Scale.

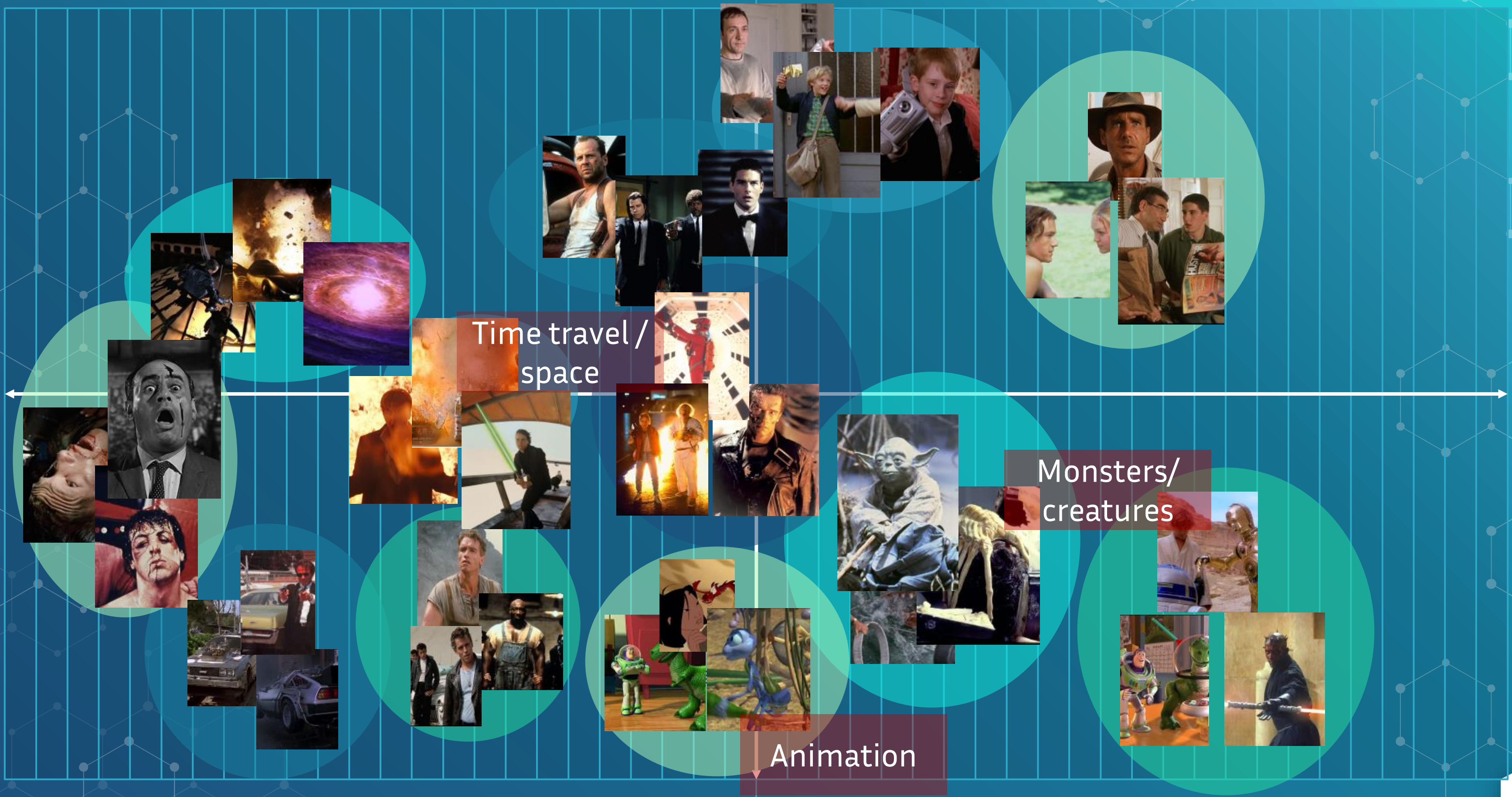


3

Results

Exploratory Analysis, Recommendation Quality, User Study

EXPLORATORY ANALYSIS – DeepCineProp

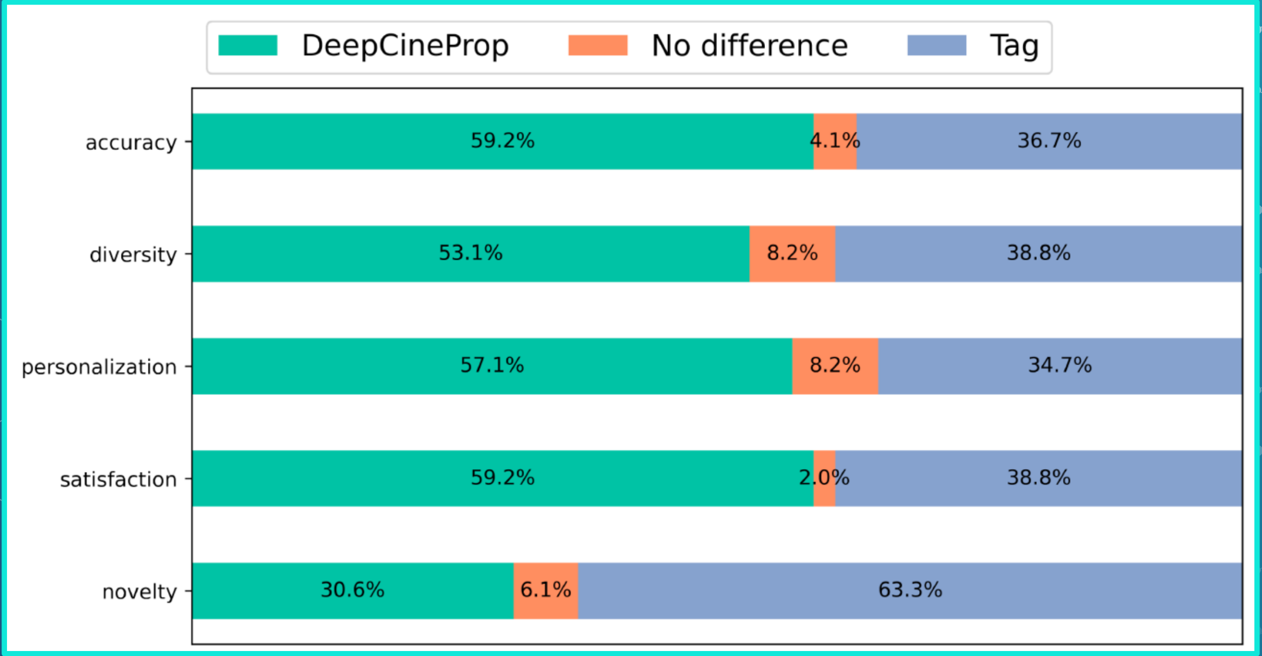


Recommendation Quality - Offline

Features	Type	P@K	R@K	AUC	Reciprocal Rank
Genre	*manual*	0.008	0.007	0.661	0.035
Tag	*manual*	0.053	0.068	0.721	0.147
DeepCineProp-c	*automatic*	0.116	0.123	0.885	0.270
DeepCineProp-f	*automatic*	0.122	0.123	0.890	0.282
CineSub	*automatic*	**0.177**	**0.172**	**0.962**	**0.381**

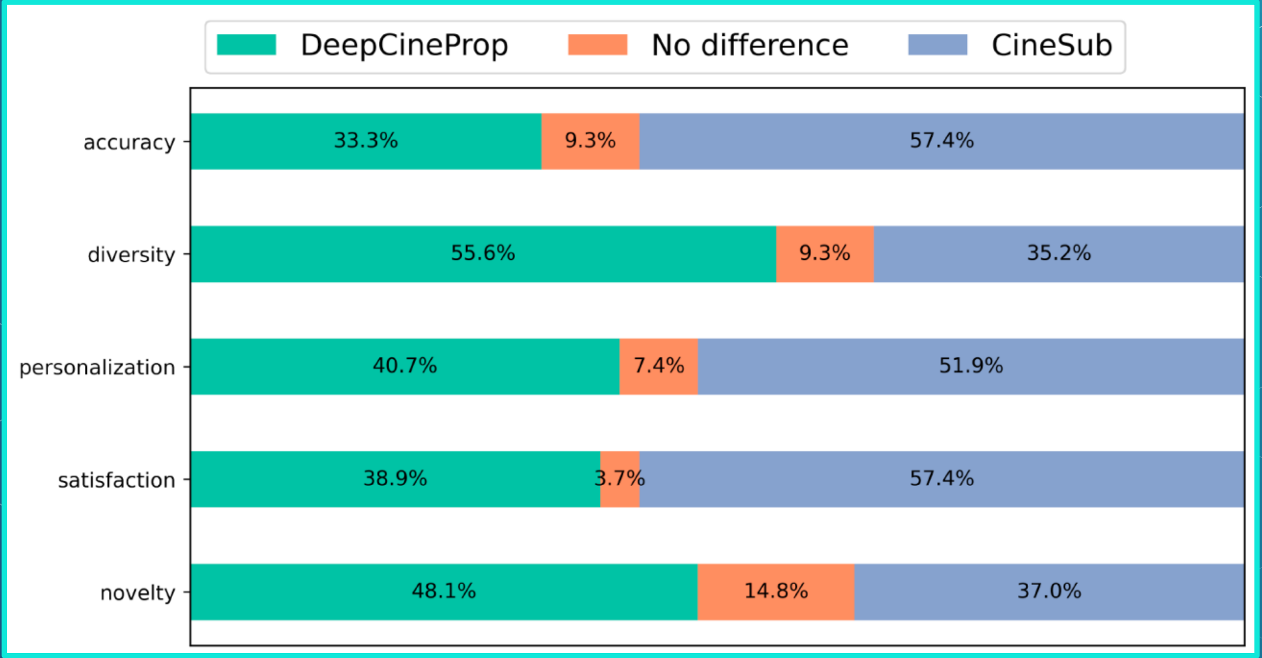
Recommendation Quality – User Study

DeepCineProp (visual) vs. Tag



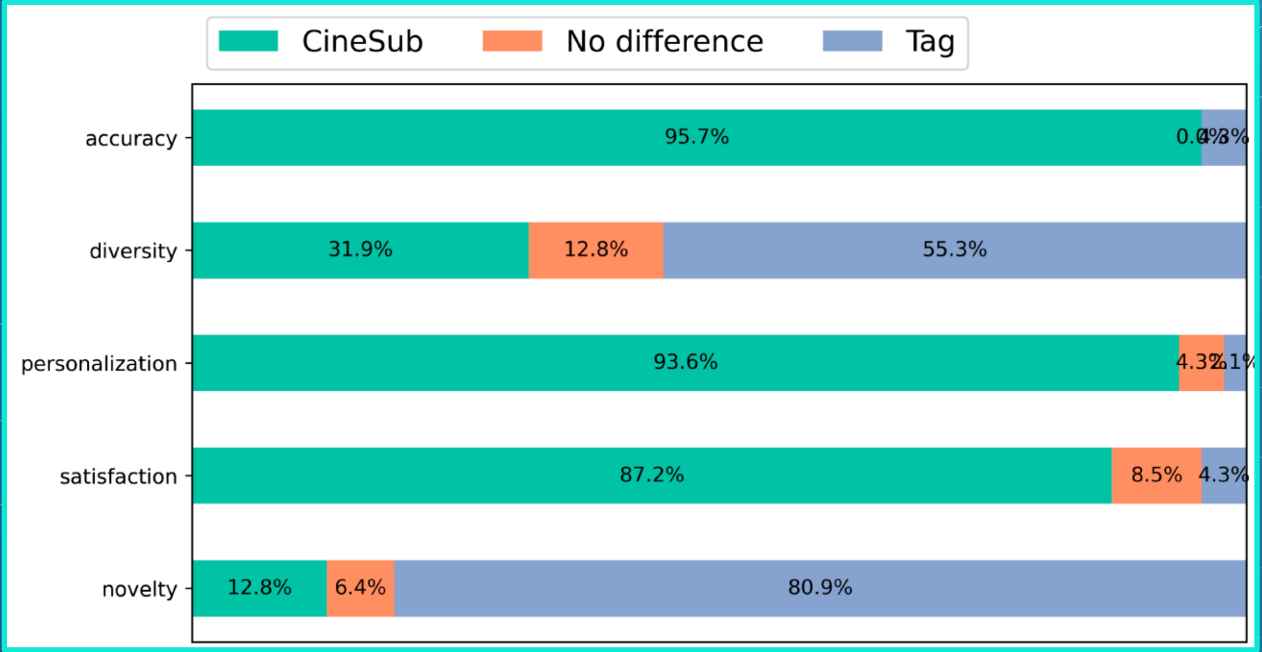
Recommendation Quality – User Study

DeepCineProp (visual) vs. CineSub (subtitles)

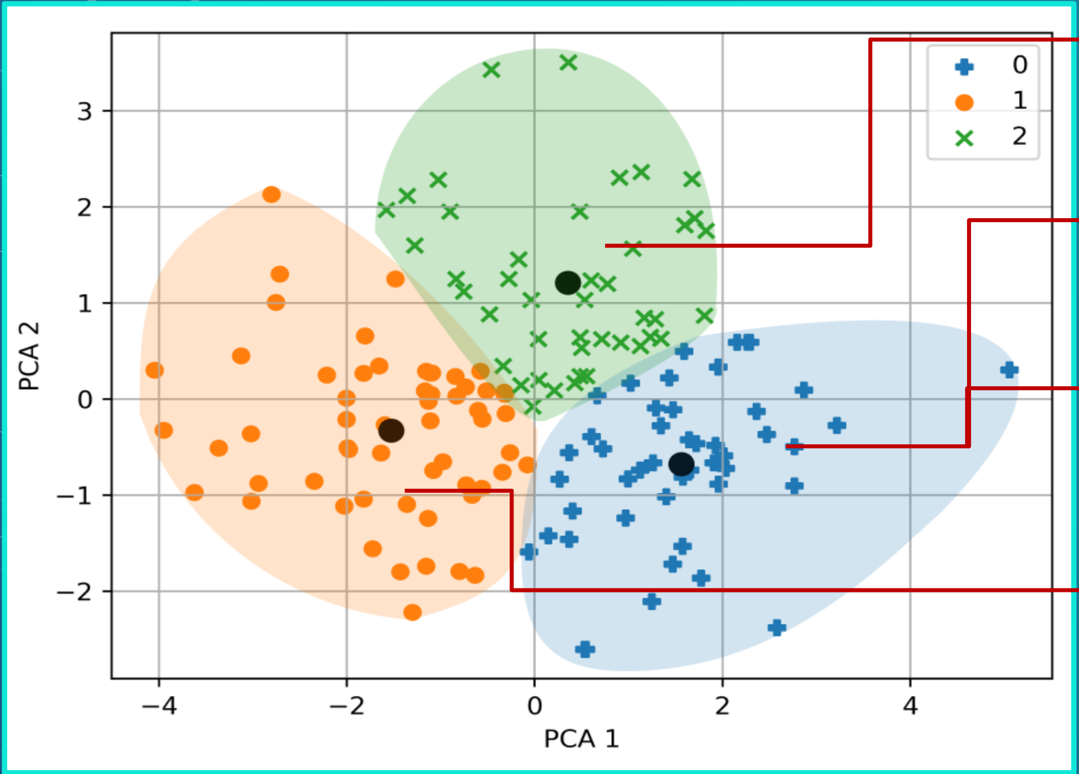


Recommendation Quality – User Study

CineSub (subtitles) vs. Tag



Recommendation Quality – User Study



Introverted, conscientious

Emotionally stable,
low conscientiousness

More likely to prefer DeepCineProp - Diversity

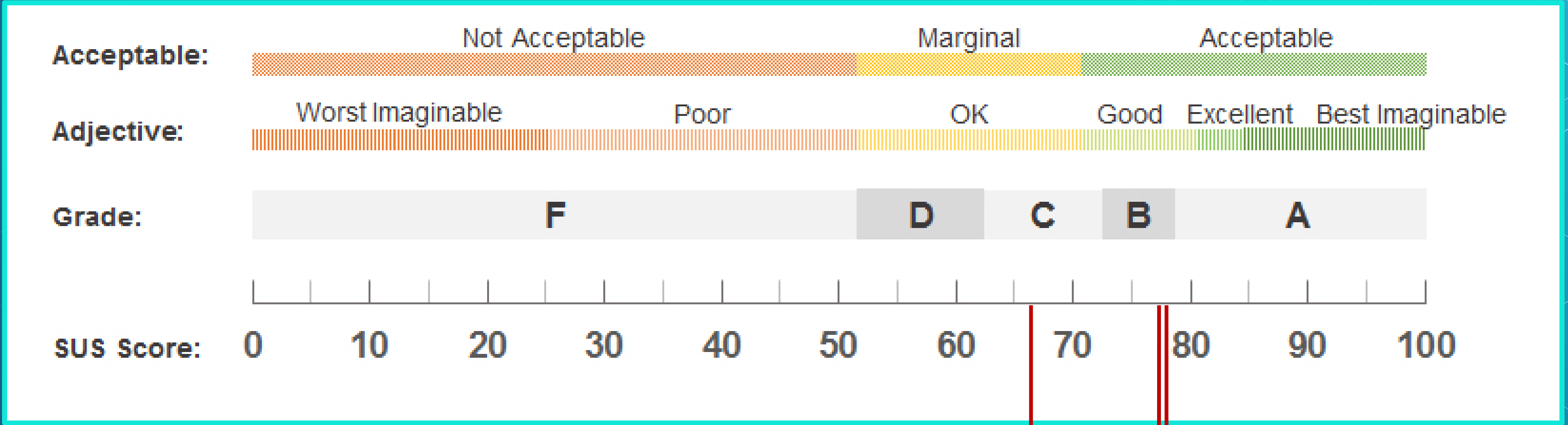
Neurotic, conscientious

Recommendation Quality – User Study

Main observations:

- Accuracy results resemble offline evaluation
- Automatic features outperform manual features
  - Exception: novelty
- DeepCineProp outperforms CineSub in diversity and novelty
- Diversity is an orthogonal factor

User Study – System Usability Scale

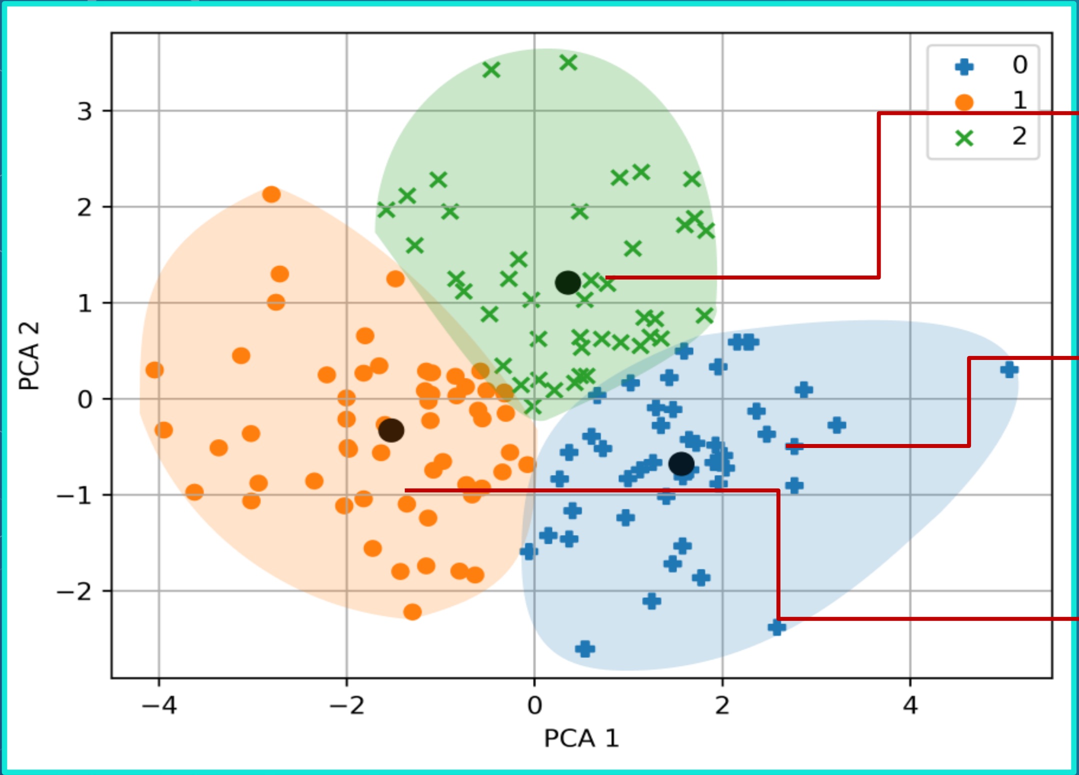


15 of the most popular mobile applications

Public-facing websites

Proposed framework

User Study – System Usability Scale



Introverted, conscientious
SUS score: 82.6

Emotionally stable,
low conscientiousness
SUS score: 74.3

Neurotic, conscientious
SUS score: 75.7

Conclusion

- Visual features based on deep learning
  - Algorithmic measures
  - User perception of performance
  - Beyond-accuracy metrics

- Proposed framework
  - Usability

Future plans

- **Write and submit to conference/journal.**
  - Expand user study?

- Further exploration of subtitles for movie recommendation

THANKS!

ANY QUESTIONS?