ToViS
Topic Based Video Suggestion System

Tushaar Gangarapu (15IT117)
Himadri Pal (15IT119)
Pratyush Prakash (15IT130)
Suraj Hegde (15IT146)

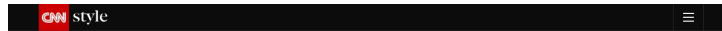# Agenda

- ❏ **Introduction**
- ❏ **Literature Survey**
- ❏ **Outcome of Literature Survey**
- ❏ **Motivation**
- ❏ **Problem Statement**
- ❏ **Objectives**
- ❏ **Proposed Work**
- ❏ **Conclusion**
- ❏ **Timeline of Project**
- ❏ **Individual Contribution**
- ❏ **References**

Introduction

CNN

Hulu

YouTube

Literature Review

Authors	Methodology	Advantages	Limitations
J. Davidson et al.	Collaborative filtering	Recommendations are more accurate	Presentation bias, only co-viewed videos
Shumeet Baluja et al.	Co-view graph	Same as above	Same as above
Bo Yang et al.	Hybrid model	Considers both video info and co-views (multimodal)	Small scale

Outcome of Literature Review

- Current state-of-the-art video suggestion systems are based on the *collaborative filtering analysis*– less applicable to fresh videos or tail videos with few views, since they have very sparse and noisy co-view data.

- Lack of *content based* video retrieval systems

- *Topic based indexing* retrieval has not be done on a large scale before

Issues and Challenges

- *Hand-crafted* **playlists**
  - Does not scale for large video collections

- *Metadata/Topic* **based playlists**
  - More of the same/non-diverse
  - Poor metadata
  - User feedback is not considered

- *Co-view* **counts**
  - Works well for popular videos with many views
  - Fresh/tail content will have very sparse and noisy co-view data

Motivation

- Abundance of content supply and demand: video retrieval, recommendation and discovery

- Least number of people are willing to spend more time on website

- Related video suggestion: ubiquitous on the web

- CNN, Netflix, Hulu, YouTube: related videos– improving these suggestions

- Improve "*user*" engagement and browsing experience (click-through analytics)

Problem Statement

- Model videos using *weighted topic vectors*

- Use an information retrieval approach to find related videos

  - **Inverted index**– topic → video index
  - **Query**– watch video
  - **Documents**– ranked videos
  - **Topic weights**– tf-idf / learn from user feedback

Research Objectives

- *Video Representation*
  - Topic Assignment
  - Topic Indexing
- *Retrieval* with weighted topics

Experimental Setup

1. **Twenty people chosen,not random.**

2. **Selected a random movie trailer and navigated through 10 other movies.**

3. **Co view graph constructed.**

4. **Users defined tags for movies watched.**

Proposed Model

Watch Video (V_w)

Collaborative Filtering Analysis

Topical Retrieval

Topic indexing → Retrieval using tf-idf scoring → Topic Transitions

Re-ranking model

Related Videos (V_R)
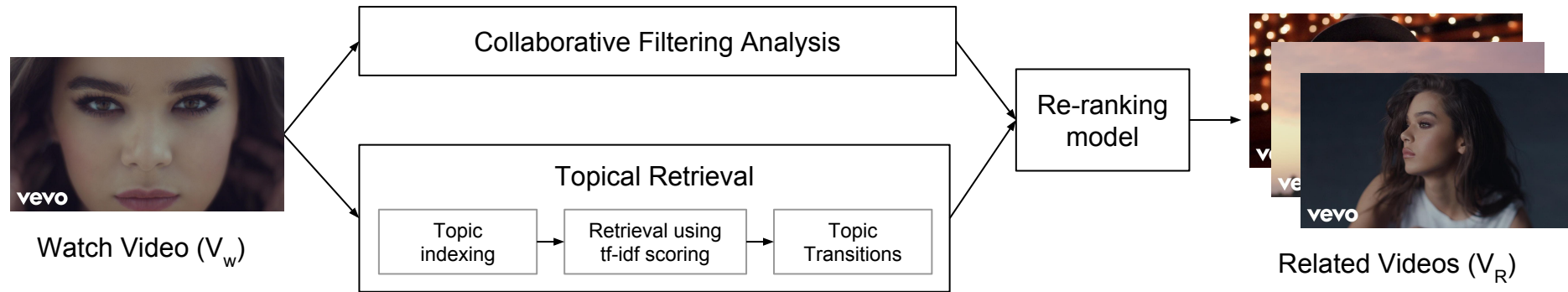
Methodology: Video Representation
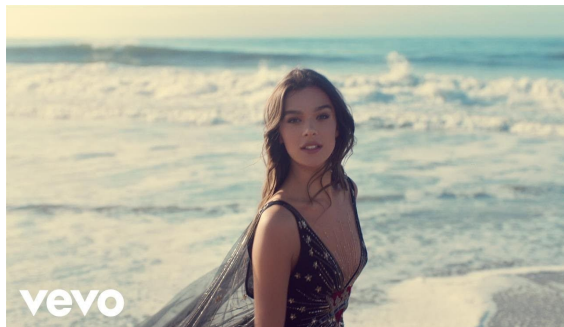
- *Semantic representation* of video content
- Annotate video 'V' using *textual annotations*, *query logs*, *anchor text information*

Soundtrack (0.335)
Fifty_Shades_Freed (0.894)
Hailee_Steinfeld (0.995)
Video_Song (0.112)

metadata
uploader keywords
common search queries
playlist names
Freebase entities
Wikipedia articles

...

Methodology: IR Weights

Quality of video: likes, dislikes, age etc.

Topic count: topicality of 'V' w.r.t. 'τ'

$$sc(V_W, V_R) = q(V_R) \sum_{\tau \in V_W \cap V_R} \mathcal{I}_s(\tau) \frac{c(\tau, V_W)}{log(1 + df(\tau))} c(\tau, V_R)$$

idf component: dampened

Indicator function: "stopword" removal

$$\mathcal{I}_s = \begin{cases} 1 & df(\tau) < df_{max} \\ 0 & else \end{cases}$$

some large constant

* Note that there is **no need** *for any document length normalization* since all the videos have roughly the same number of topics associated with them

Work Done

- **Evaluation Methodologies**
  - *Historical data, user studies, online evaluation*
  - *Choice of method–* population of users

- **Limitations to Large Scale Recommendations**
  - Biased Sample
  - Subjective w.r.t. users
  - Gap between *information need* and *information query*

Future Work

- **Query Optimization**

- **Dynamic Model**

- **Hybrid Model**

Results and Analysis

Table 4.1: Summary of Likert Evaluation (20 users)

Question Asked	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Did the recommender system recommend relevant videos?	3	8	6	1	2
Did your session time increase as result of suggestions?	1	3	10	4	2
Did you have Overall Satisfaction?	4	6	7	2	1

Results and Analysis
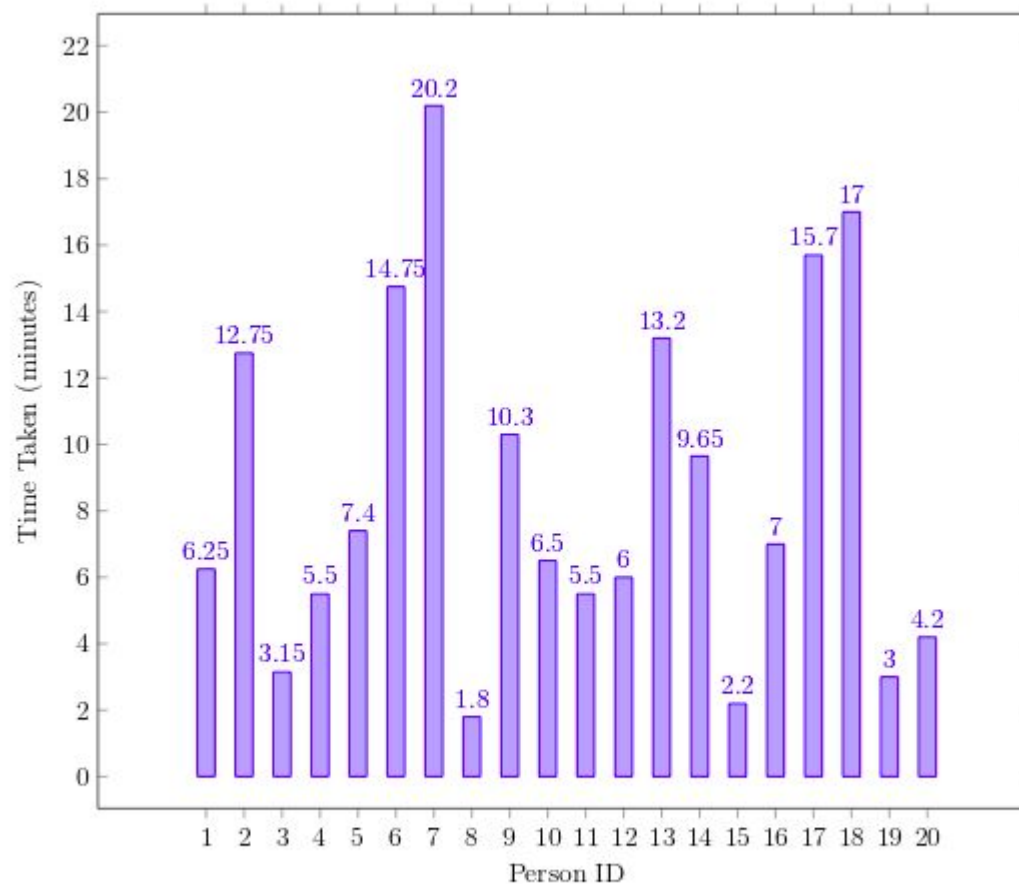

Figure 4.1: Watch Time for 20 Users (Based on Experimental Setup)

Results and Analysis
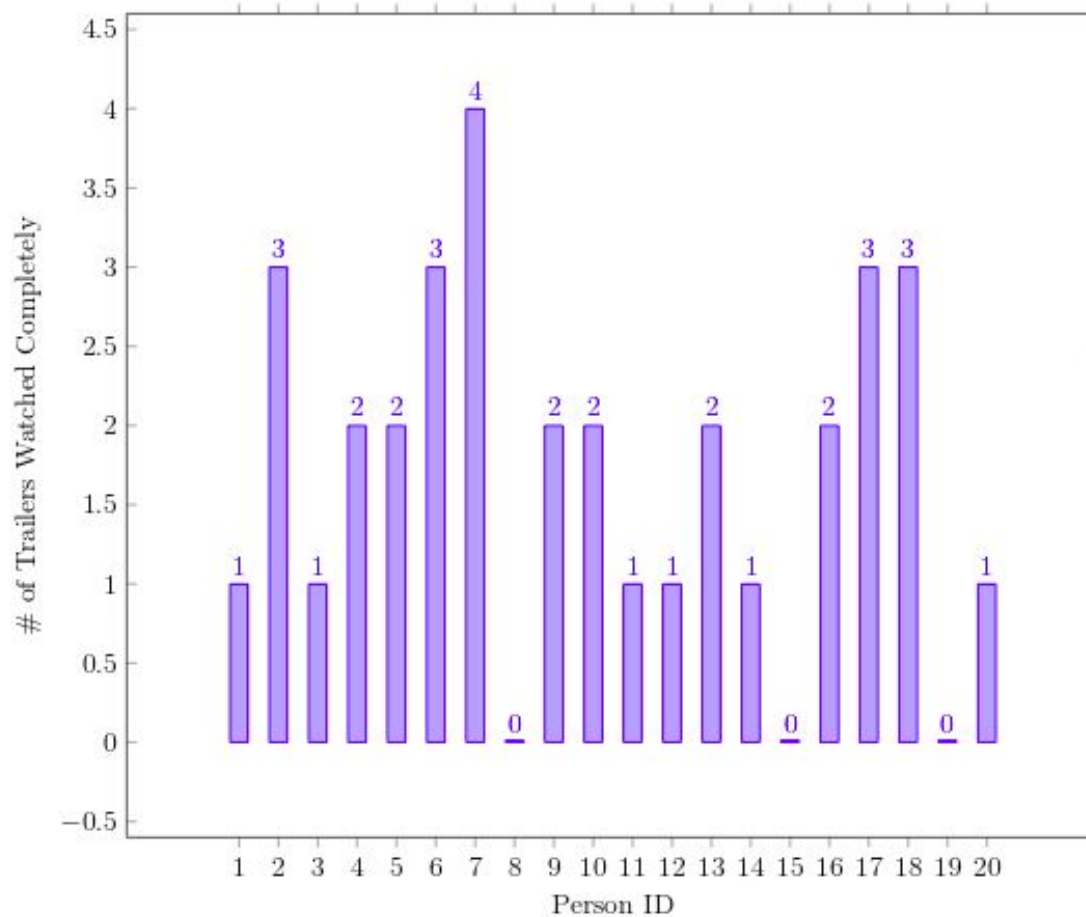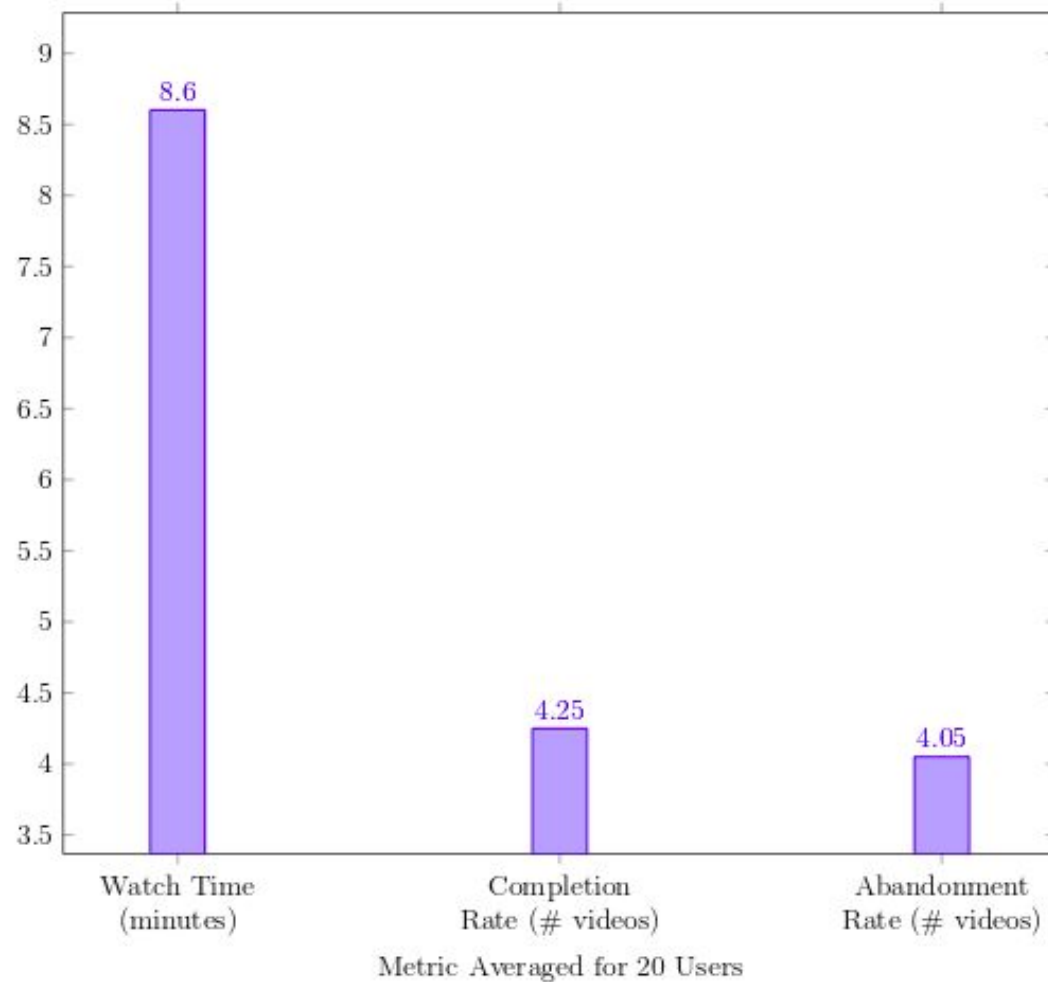

Figure 4.2: Completion Rate for 20 Users (Based on Experimental Setup)

Results and Analysis

Results and Analysis

Dunston Checks In (1996)

Watch Next:

- It Takes Two (1995)
- Toy Story (1995)
- Dracula: Dead and Loving It (1995)
- Big Green, The (1995)
- Mr. Wrong (1996)
- Don't Be a Menace to South Central While Drinking Your Juice in the Hood (1996)
- Four Rooms (1995)
- Muppet Treasure Island (1996)
- Happy Gilmore (1996)
- Black Sheep (1996)

Ace Ventura: When Nature Calls (1995)

Watch Next:

- Dracula: Dead and Loving It (1995)
- Mr. Wrong (1996)
- Don't Be a Menace to South Central While Drinking Your Juice in the Hood (1996)
- Four Rooms (1995)
- Happy Gilmore (1996)
- Black Sheep (1996)
- Friday (1995)
- Bio-Dome (1996)
- Two if by Sea (1996)
- It Takes Two (1995)

Timeline of Project

Milestones	Feb 2018	Mid-Mar 2018	End-Mar 2018	April 2018
Literature Survey	✔	✔	✔	✔
Topic Indexing		✔	✔	✔
Weighted Topic Retrieval (IRTopics)			✔	✔
User Simulated metric evaluation (Testing)				✔

Individual Contribution
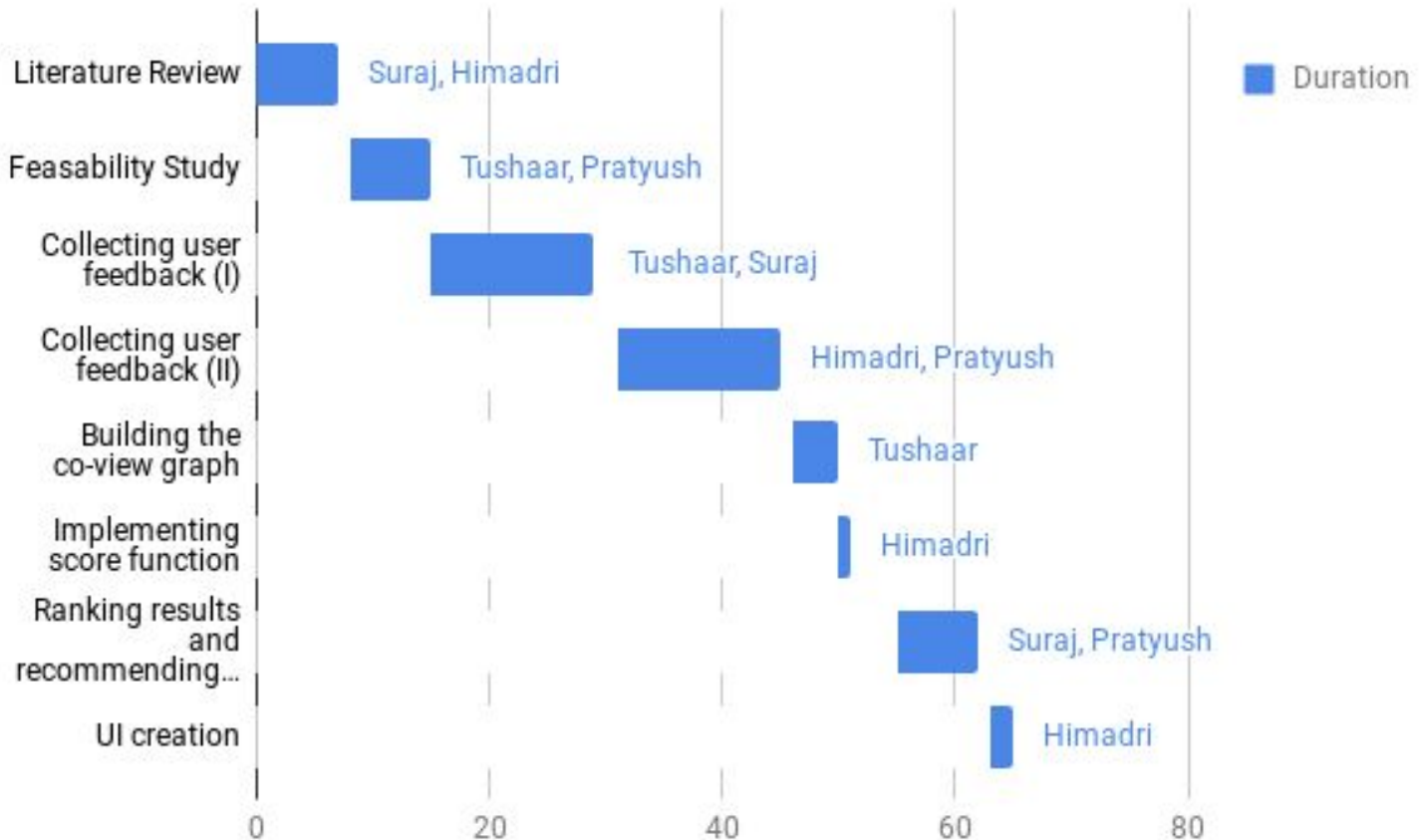
References

[1]   Bendersky, M., Harmsen, J.J., Josifovski, V., Lepikhin, D., & Pueyo, L.G. (2014). Up next: retrieval methods for large scale related video suggestion. KDD.

[2]   Broder, A. Z., Carmel, D., Herscovici, M., Soffer, A., & Zien, J. (2003, November). Efficient query evaluation using a two-level retrieval process. In Proceedings of the twelfth international conference on Information and knowledge management (pp. 426-434). ACM.

[3]   Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010, September). The YouTube video recommendation system. In Proceedings of the fourth ACM conference on Recommender systems (pp. 293-296). ACM.