

An Alternative Approach to Ranking Videos and Measuring Dissimilarity Between Video Content and Titles

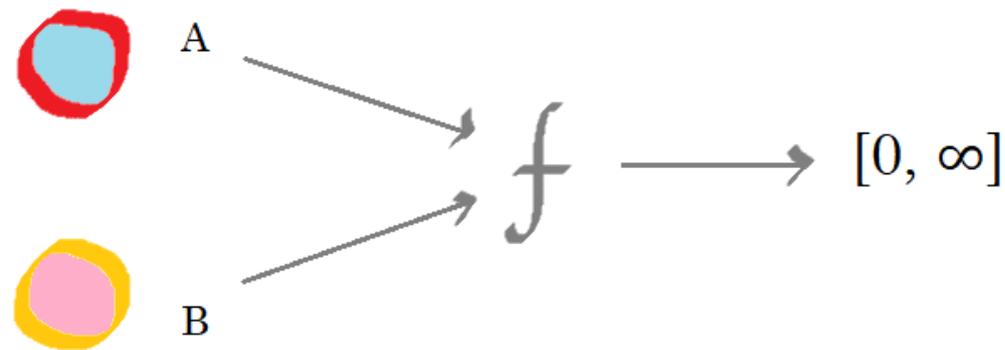
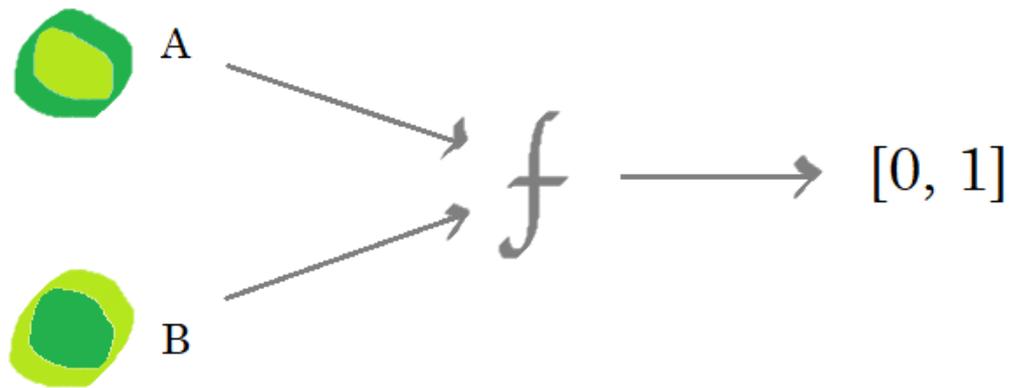
Kalp Abhinn Aghada
PES University

Presentation for
ROMCIR 2022: The 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval, held as part of ECIR 2022: the 44th European Conference on Information Retrieval, April 10-14, 2022, Stavanger, Norway

Overview

- Alternative dissimilarity measure to rank videos
- A way to detect clickbait by using dissimilarity measure

(Dis)similarity Measure



Clickbait Detection

- Current methods rely on either -
 - Meta-information
 - Individual modalities
- Dissimilarity between content and titles = clickbait
- Alternative dissimilarity measure modified to measure this dissimilarity

Related Work

- IR
 - Multitude of dissimilarity measures in IR
 - MIL-NCE (Multiple Instance Learning Noise Contrastive Estimation), HowTo100M^{1, 2}
- Clickbait Detection
 - Classifying textual clickbait content using similarity³
 - Challenge – multimodal aspect of data
 - Considering – audiovisual + title content

1. A. Miech, D. Zhukov, J.-B. Alayrac, M. Tapaswi, I. Laptev, J. Sivic, HowTo100M: Learning a Text-Video Embedding by Watching Hundred Million Narrated Video Clips, in: ICCV, 2019.

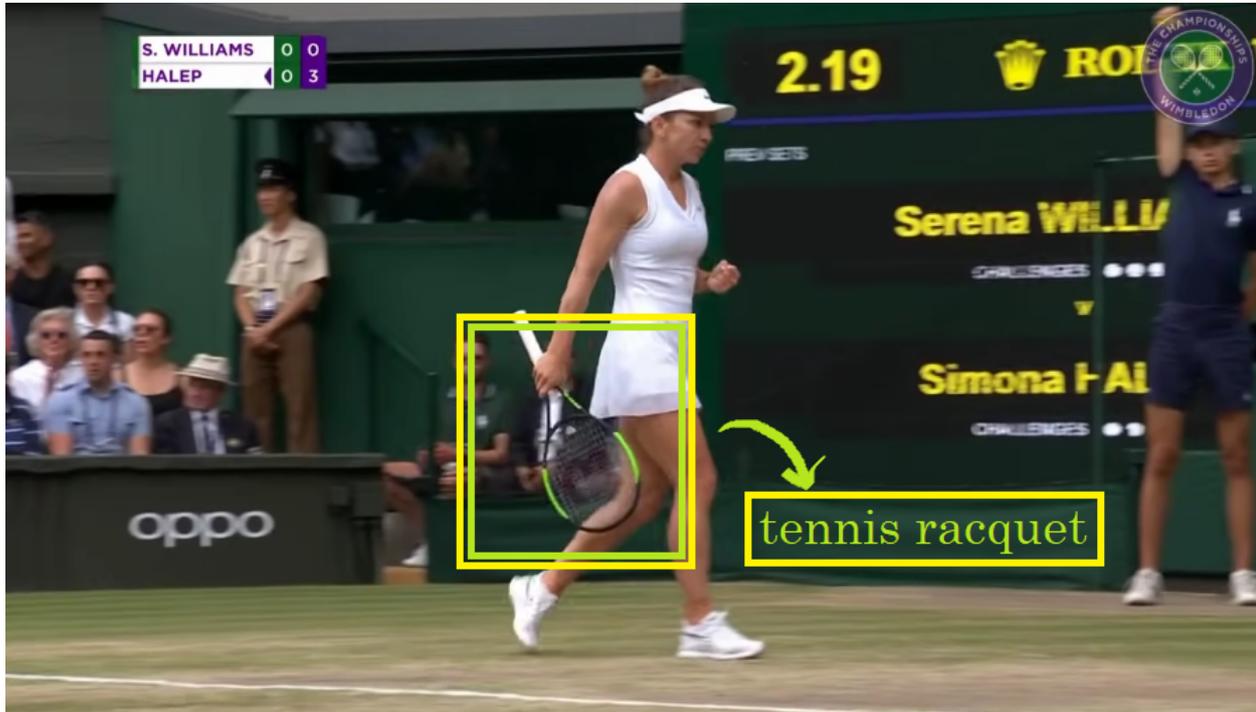
2. A. Miech, J.-B. Alayrac, L. Smaira, I. Laptev, J. Sivic, A. Zisserman, End-to-end learning of visual representations from uncurated instructional videos, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

3. M. Dong, L. Yao, X. Wang, B. Benatallah, C. Huang, Similarity-aware deep attentive model for clickbait detection, in: Q. Yang, Z.-H. Zhou, Z. Gong, M.-L. Zhang, S.-J. Huang (Eds.), Advances in Knowledge Discovery and Data Mining, Springer International Publishing, Cham, 2019, pp. 56–69.

Misinformation Mitigation

- Misinformation rife in clickbait content
- Video clickbait contributes to misinformation spread
- This work can potentially help attempt to mitigate this

Proposed Search Methodology: Object Detection



- YOLOv3 object detection model
- $\frac{\text{Occurrence of object}}{\text{Occurrence of all objects}}$
- Set of these ratios for k^{th} video
= C_k
- Set of objects for k^{th} video
= O_k

Proposed Search Methodology: Audio Recognition and Query Processing

- CMU Sphinx English model
- Set of tokens for k^{th} video = S_k
- Set of words in the title of the k^{th} video = T_k
- $tf(\text{word}, S_k), tf(\text{word}, T_k)$
- Input query tokenized, stop words removed = q



audio transcript

... one of the best defenders in the women's game and she gets so well on the run...

title

Simona Halep vs Serena Williams
Wimbledon 2019 final

Proposed Search Methodology: Dissimilarity Measure

$$f(x_1, x_2, x_3, \dots, x_n) = \frac{\sqrt{\frac{\sum(x_i - \frac{\sum x_l}{n})^2}{n}}}{(\frac{\sum x_l}{n})^{n-1}} + \frac{1}{\frac{\sum x_l}{n}}$$

Where,

- $x_i \in C_k \forall i \in q \cap O_k$ and $n = |q \cap O_k|$, or
- $x_i = tf(i, S_k) \forall i \in q \cap S_k$ and $n = |q \cap S_k|$, or
- $x_i = tf(i, T_k) \forall i \in q \cap T_k$ and $n = |q \cap T_k|$.

Proposed Search Methodology: Algorithm

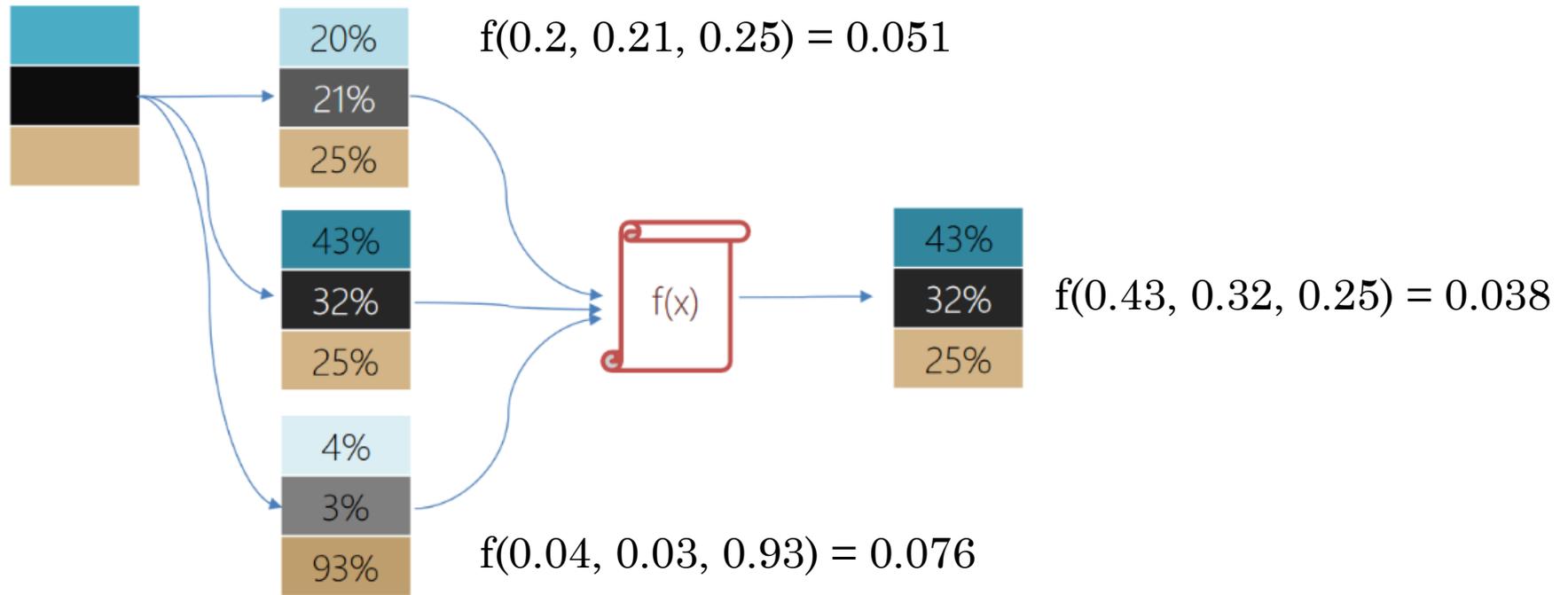
Algorithm 1 Algorithm blueprint for video retrieval using f

Input: $D = \{\text{set of all visual data or, set of all captions or, set of all titles}\}$, $q = \text{set of all words}$
in tokenized input query, $E_k = \{O_k \text{ or, } S_k \text{ or, } T_k\}$ where $k \in \text{set of all videos}$

Output: $R = \text{search output}$

- 1: for $entry$ in D do
 - 2: for $word$ in $q \cap E_{entry}$ do
 - 3: Calculate C_{entry} or $\text{tf}(\text{word}, \text{entry})$
 - 4: Calculate f .
 - 5: if Calculated f is lesser than previous f then
 - 6: Update R with video corresponding to $entry$
 - 7: if ($R \neq \text{empty}$) then
 - 8: Return R
-

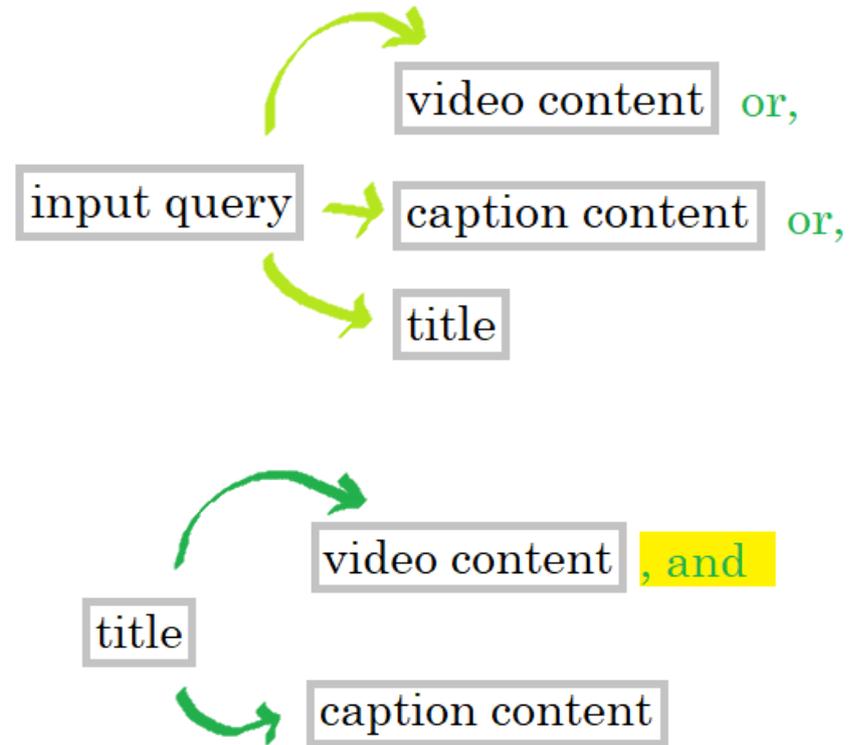
Search Methodology in Action



Measuring Dissimilarity between Video Content and Titles

$$f(x_1, x_2, x_3, \dots, x_n) = \frac{\sqrt{\frac{\sum(x_i - \frac{\sum x_l}{n})^2}{n}}}{(\frac{\sum x_l}{n})^{n-1}} + \frac{1}{\frac{\sum x_l}{n}}$$

$$x_i = \begin{cases} x & | x \in C_k \forall i \in T_k \cap O_k \\ tf(i, S_k) & \forall i \in T_k \cap S_k \end{cases}$$



Empirical Analysis: Dataset

Dataset Parameter	Min	Mean	Max
Video Length	0.86 minutes	5.05 minutes	14.99 minutes
Title Length	3 words	7 words	11 words
Caption Length	137 words	884 words	2796 words

Table 1
Dataset Statistics.

- 304 videos from HowTo100M dataset
- Selecting top 5 results for each detectable object from object detection model
- Repeated results, results other than in English, longer than 900 seconds removed
- 100 queries
- Sampling rate = 5fps

Empirical Analysis: Evaluation Metrics

- IR
 - Precision at k
 - Recall at k
 - Relevance assessment per query manually established
- Clickbait classification
 - Precision
 - Recall
 - Accuracy
 - F1-score
 - MCC (Matthews correlation coefficient)
 - $f = 100$, Ground truth manually established

Results

Table 2

Video Retrieval/Ranking Statistics for Precision at k.

Method	P@1	P@2	P@3	P@4	P@5
MIL-NCE [13, 4]	40.00%	60.00%	45.66%	57.50%	51.40%
Method described in this work (visual data)	75.00%	73.50%	76.67%	67.50%	70.00%
Method described in this work (caption data)	35.00%	59.50%	47.30%	57.52%	52.00%
Method described in this work (title data)	79.00%	75.00%	76.67%	77.50%	75.20%

Table 3

Video Retrieval/Ranking Statistics for Recall at k.

Method	R@1	R@2	R@3	R@4	R@5
MIL-NCE [13, 4]	5.71%	17.14%	19.57%	32.86%	36.71%
Method described in this work (visual data)	10.71%	21.00%	32.85%	38.57%	50.00%
Method described in this work (caption data)	5.00%	17.00%	20.27%	32.87%	37.14%
Method described in this work (title data)	11.28%	21.43%	32.86%	44.28%	53.71%

Table 4

Clickbait Classification Statistics.

Recall	Precision	Accuracy	F1-Score	MCC
0.77	0.85	0.97	0.81	0.79

Conclusion

- Objective
 - Formulate an alternative approach to ranking videos
 - Measure dissimilarity between audio-visual content and titles
- Results
 - Performs better than a current video retrieval system
 - Performs reasonably well in clickbait detection for small dataset

Future Work

- Dataset size can be increased
- Video retrieval method used with methods suited to large datasets
- The three modalities can be combined
- Clickbait classification testing with benchmark
- Threshold value can be treated as learnable parameter
- Integrating meta information into clickbait classification
- Large scale video library for clickbait classification

End of Presentation Questions and Comments Welcome!

Kalp Abhinn Aghada
PES University

Presentation for
ROMCIR 2022: The 2nd Workshop on Reducing Online Misinformation
through Credible Information Retrieval, held as part of ECIR 2022: the
44th European Conference on Information Retrieval, April 10-14, 2022,
Stavanger, Norway