LINKE

Lab for Intelligent Networking
and Knowledge Engineering

High-Quality Activity-Level Video Advertising

Mu Yuan[1], Lan Zhang[1], Zhengtao Wu[1,2], Daren Zheng[1]

[1]University of Science and Technology of China

[2]Tencent Marketing Solution

中国科学技术大学
University of Science and Technology of China

Tencent 腾讯

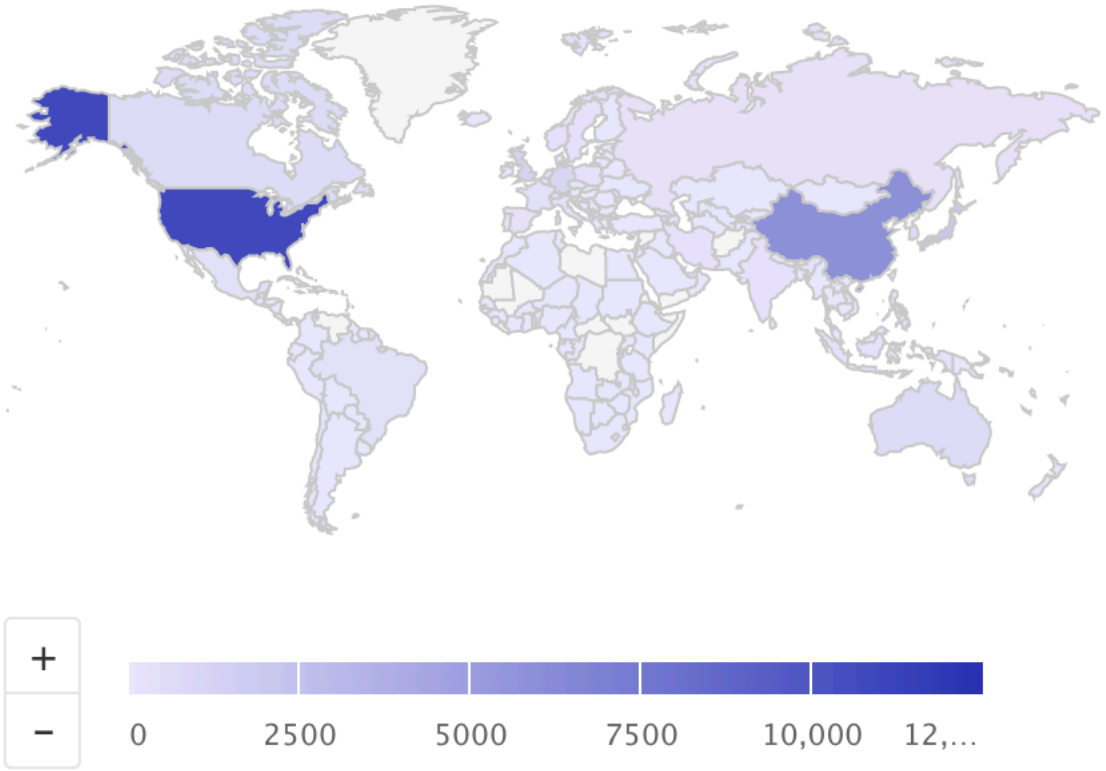
Outline

- **Introduction**
- Multimodal Content Embedding
- Activity-Level Video Advertising
- Evaluation
- Conclusion

Introduction

Billion-Dollar Business

2020 Video Advertising Market Volume

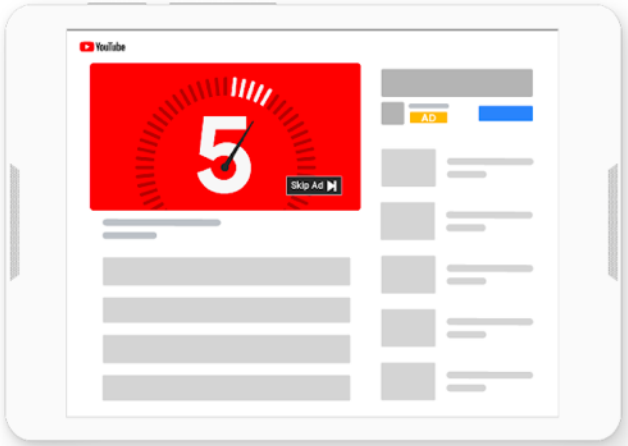


Top 5	
1. United States	**US$10,857m**
2. China	US$5,897m
3. Japan	US$2,153m
4. United Kingdom	US$1,222m
5. Germany	US$1,035m

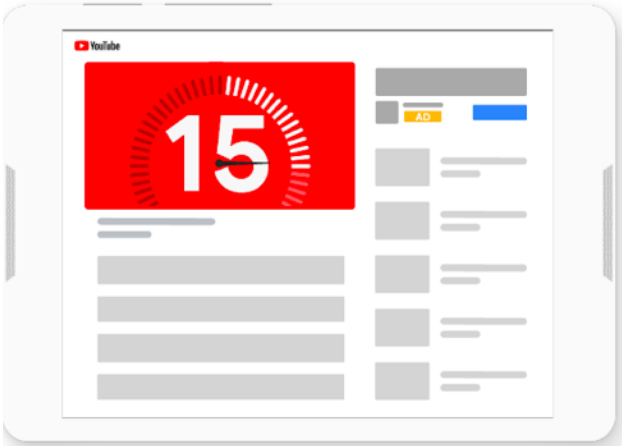
https://www.statista.com/outlook/218/109/video-advertising/united-states#market-globalRevenue

Introduction

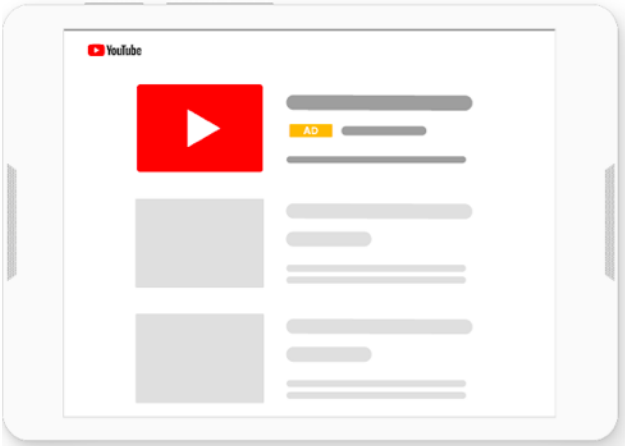
Video Advertising Formats



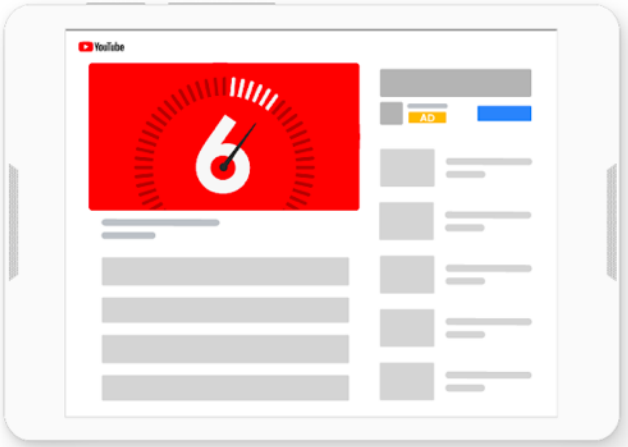
Skippable In-Stream Ads



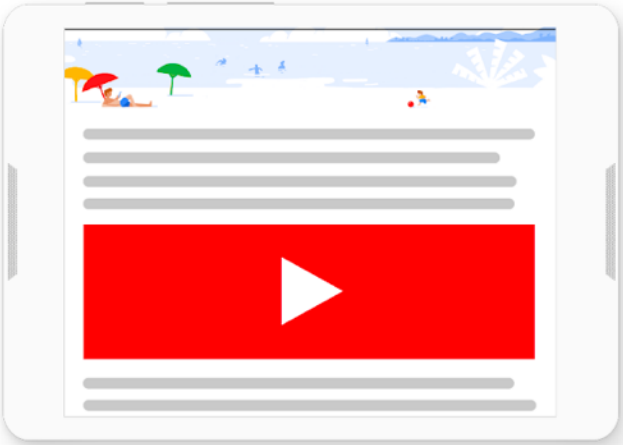
Non-Skippable In-Stream Ads



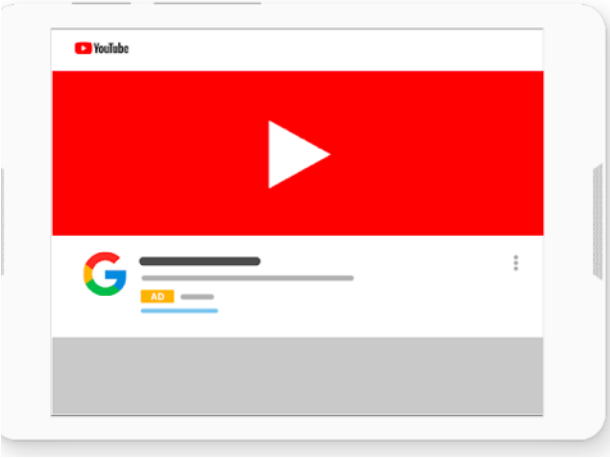
Video Discovery Ads



Bumper Ads



Out-Stream Ads

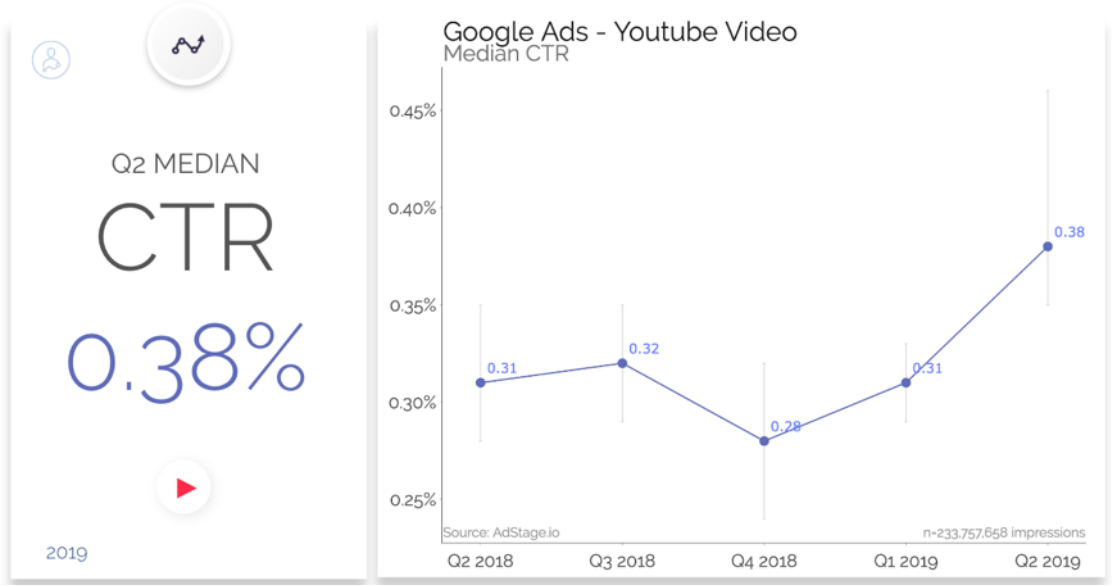


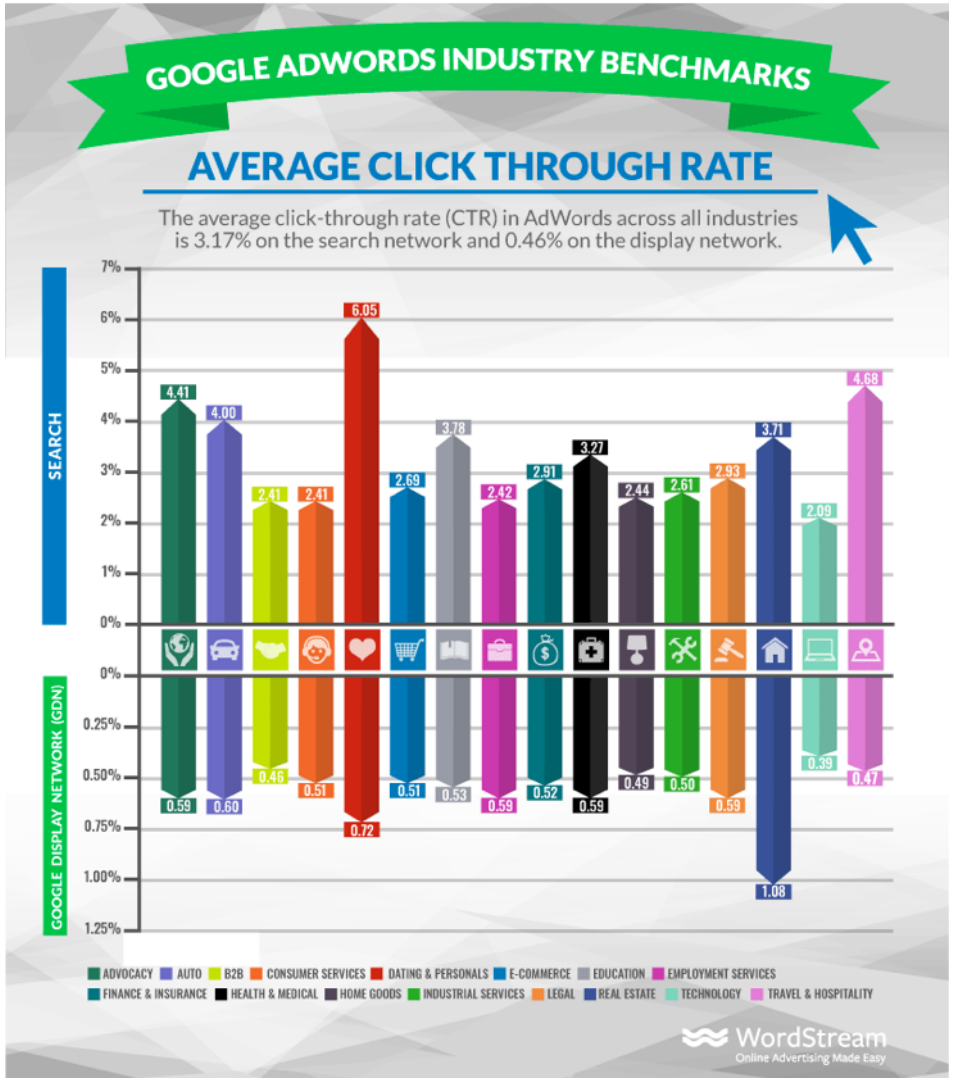
Masthead Ads

https://support.google.com/google-ads/answer/2375464?hl=en

Introduction

Low CTR Compared with Traditional Ads





Youtube 0.38% << AdWords 3.17%

https://blog.adstage.io/youtube-benchmarks-cpc-cpm-and-ctr
https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks

Introduction

Content-Related Video Advertising

Will you marry me?

Diamond Rings
White Sapphire 58...
HK$3,928.00
Glamira HK

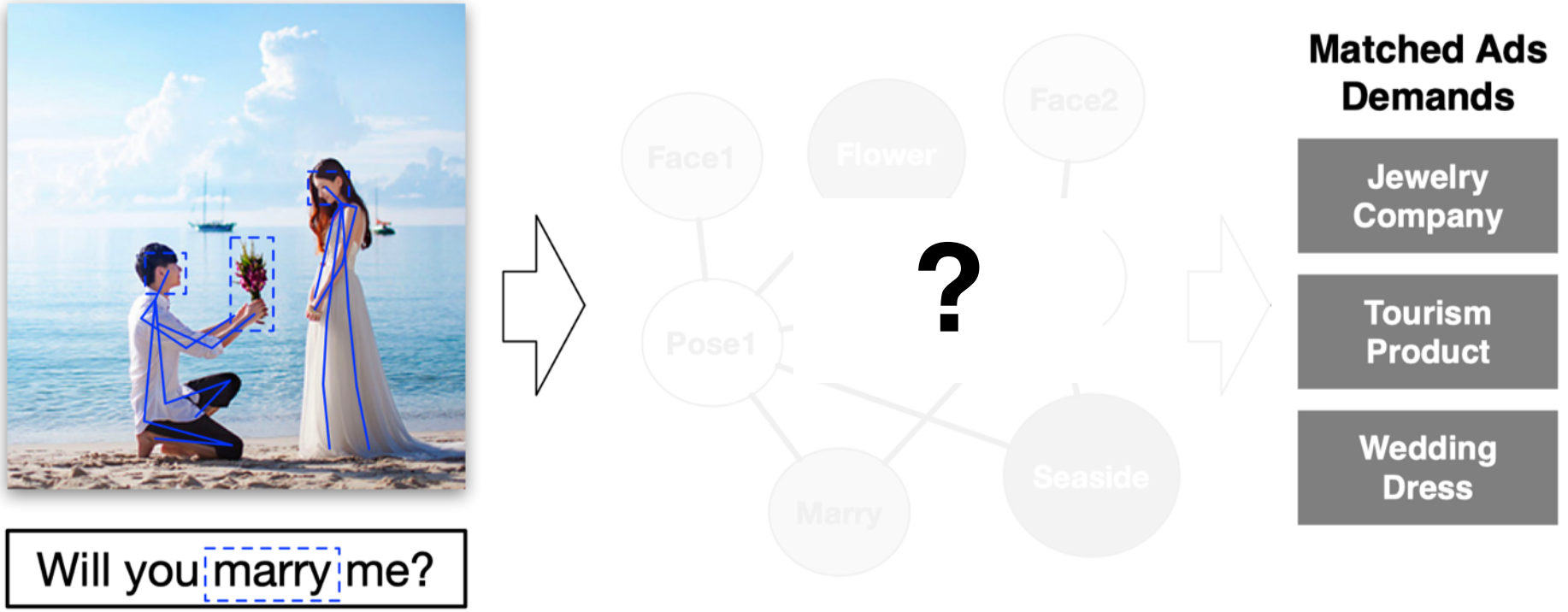
LoveShackFancy
Irene Gown in Whi...
HK$9,262.70
revolve.com

Matched Ads Demands

- Jewelry Company
- Tourism Product
- Wedding Dress

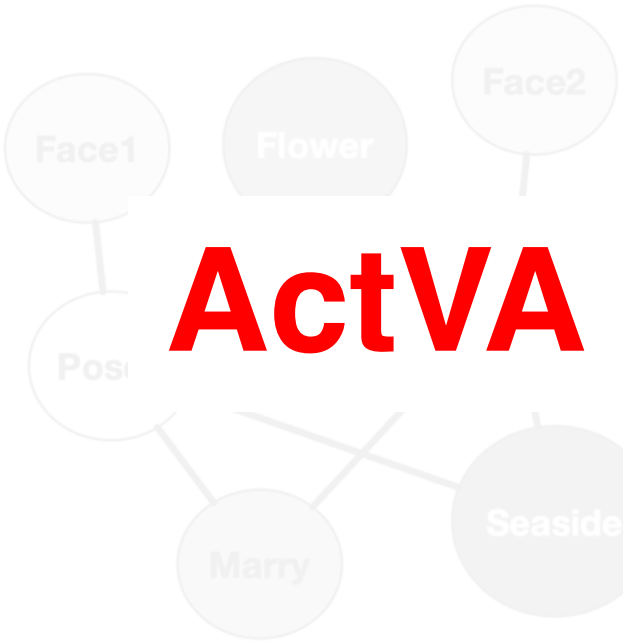
Introduction

Content-Related Video Advertising



Introduction

Activity-Level Video Advertising



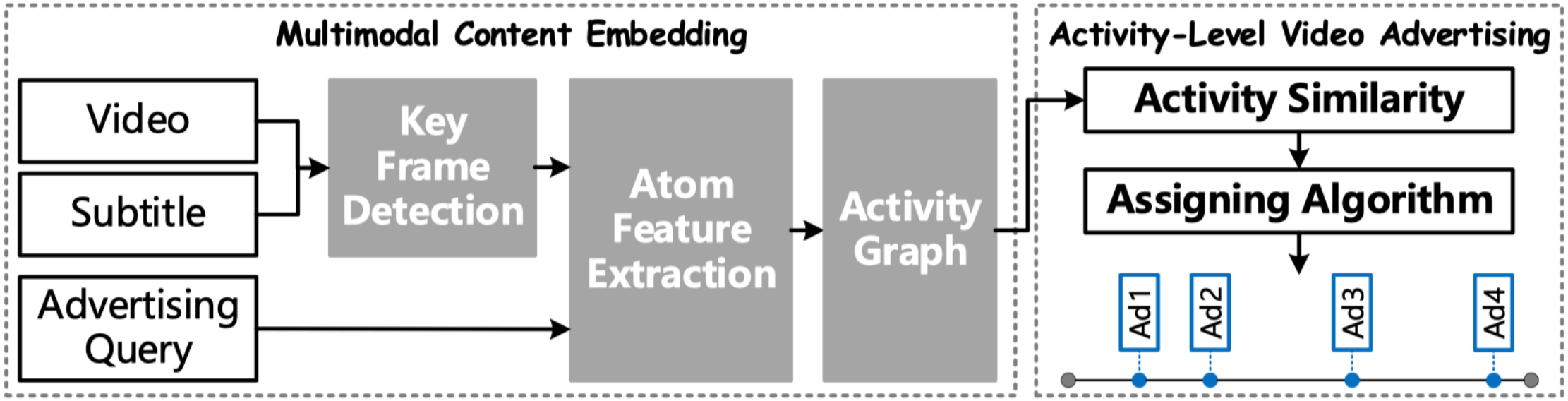
Matched Ads Demands

- Jewelry Company
- Tourism Product
- Wedding Dress

Introduction

Activity-Level Video Advertising

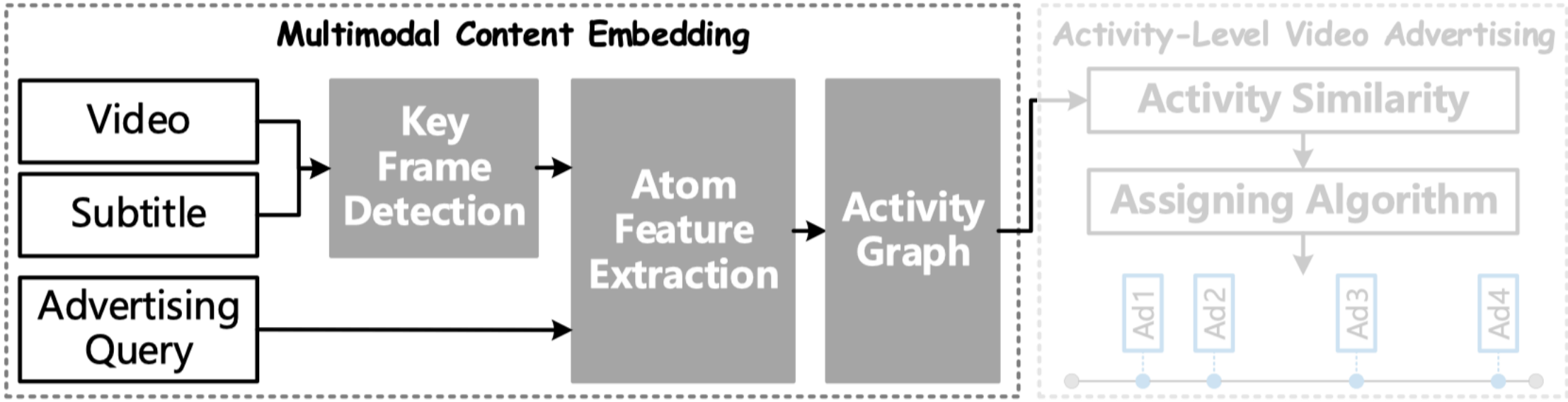
Main Contributions



1. The first non-predefined activity-level video advertising system;
2. Effective algorithm for optimizing advertising service over content relevance, revenue and intrusiveness perception.

Activity-Level Video Advertising

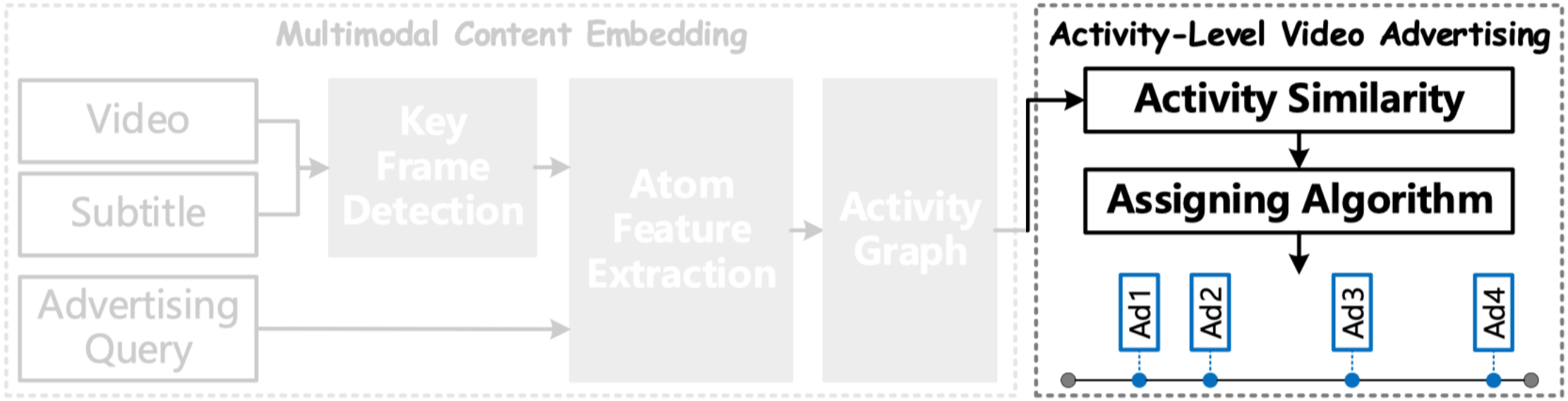
System Framework



Step#1: Find possible ads positions.

Activity-Level Video Advertising

System Framework



Step#2: Assign ads properly.

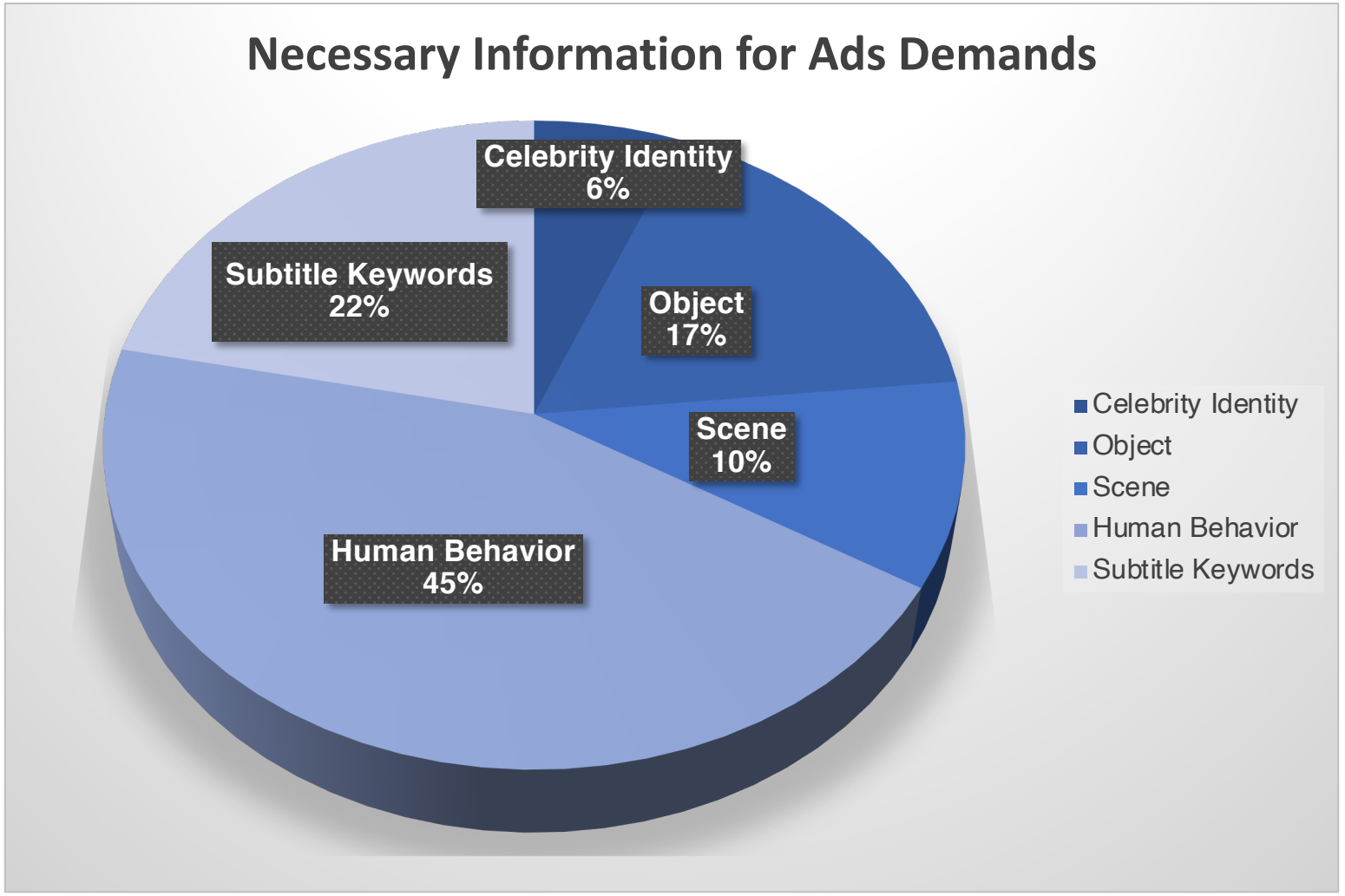
Outline

- Introduction
- **Multimodal Content Embedding**
- Activity-Level Video Advertising
- Evaluation
- Conclusion

Multimodal Content Embedding

Market Research @ Tencent

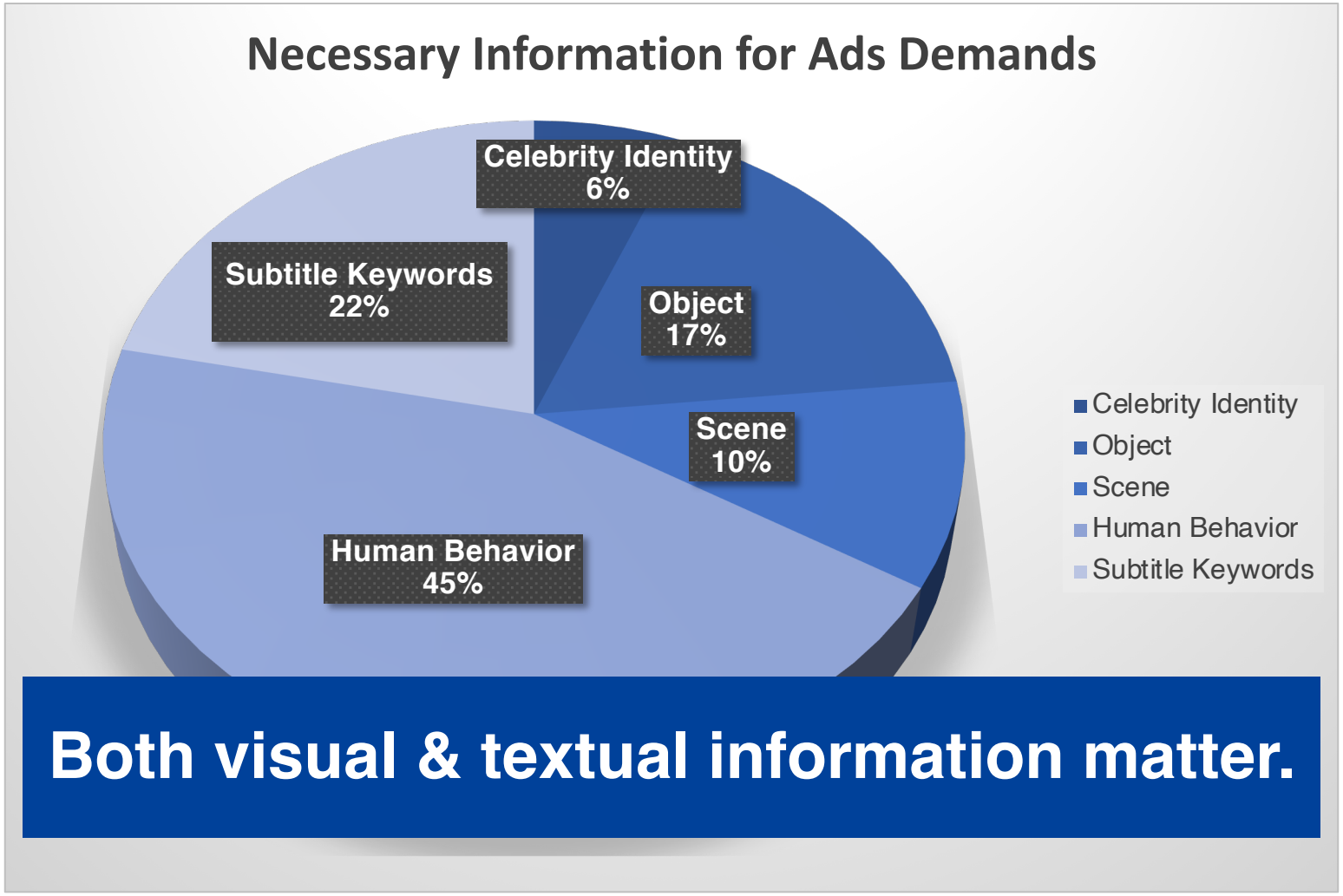
150 Content-Targeted Advertising Needs



Multimodal Content Embedding

Market Research @ Tencent

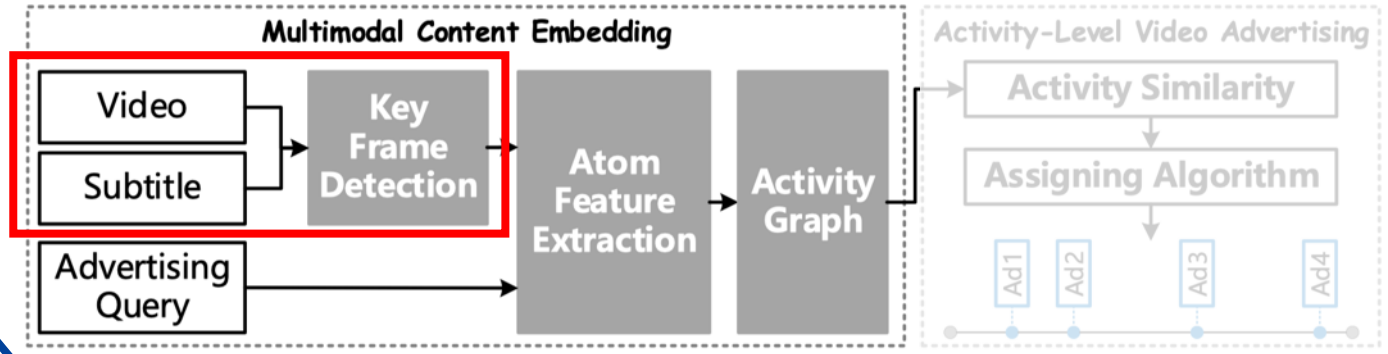
150 Content-Targeted Advertising Demands



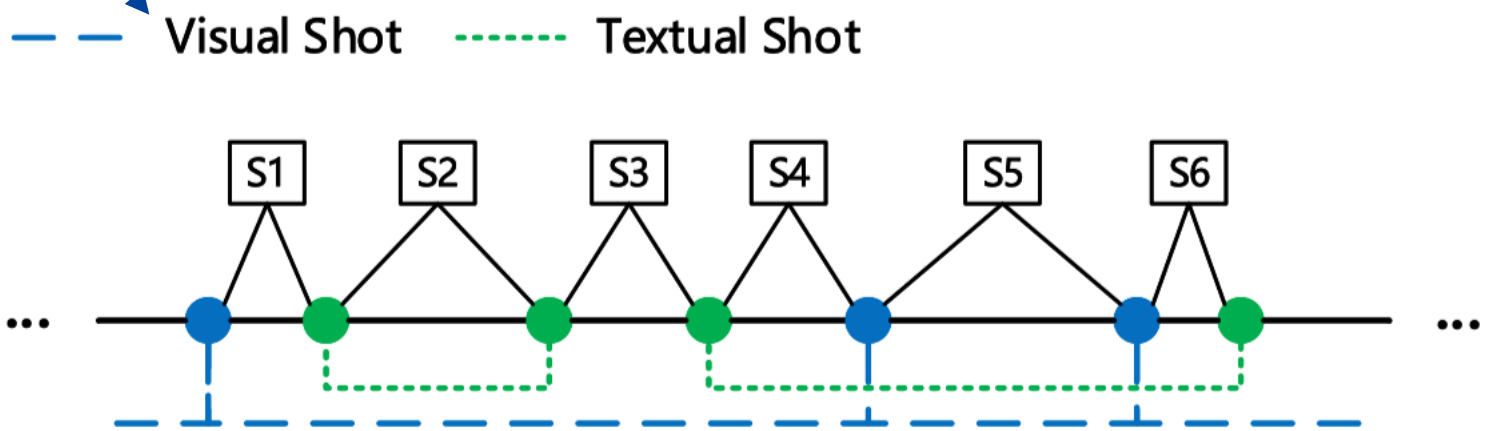
Both visual & textual information matter.

Multimodal Content Embedding

Key Frame Detection

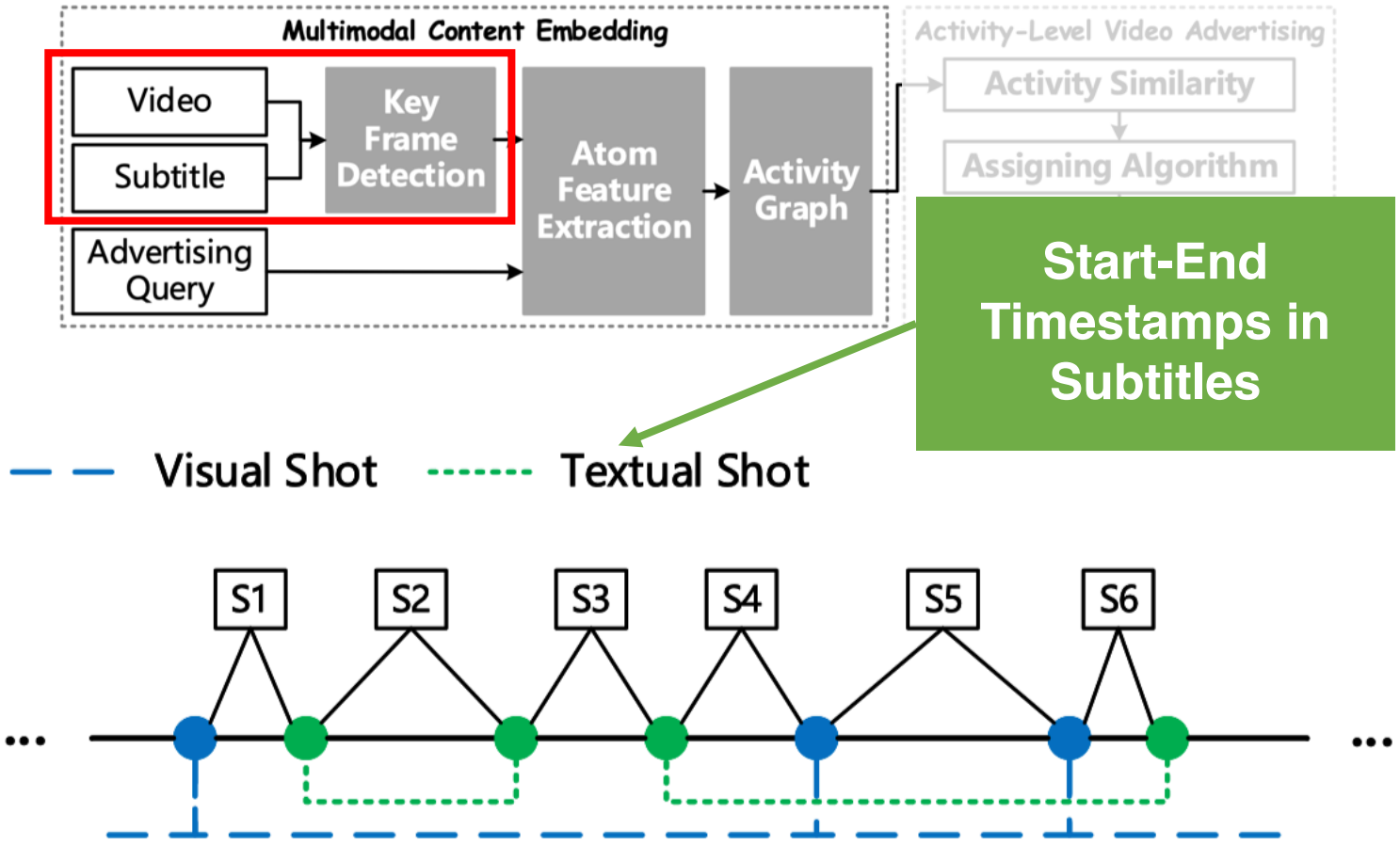


Color-Histogram based Shot Detection



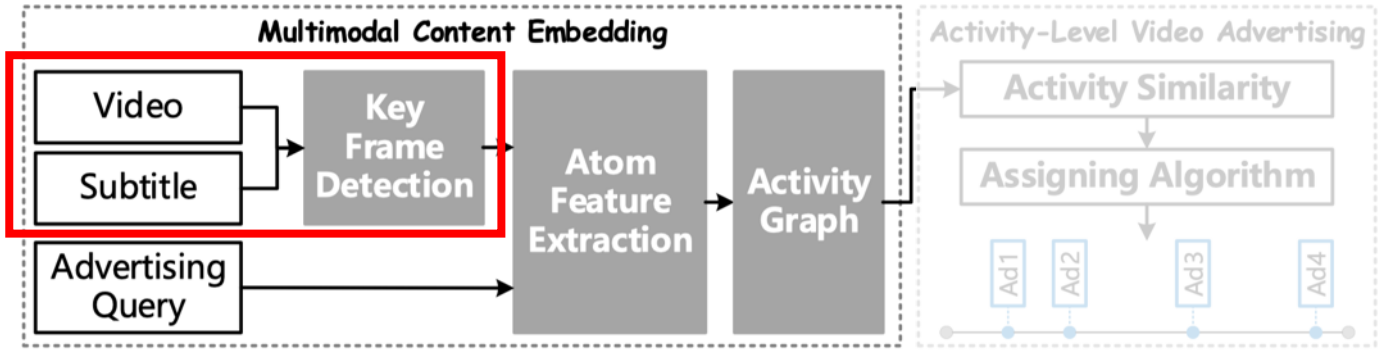
Multimodal Content Embedding

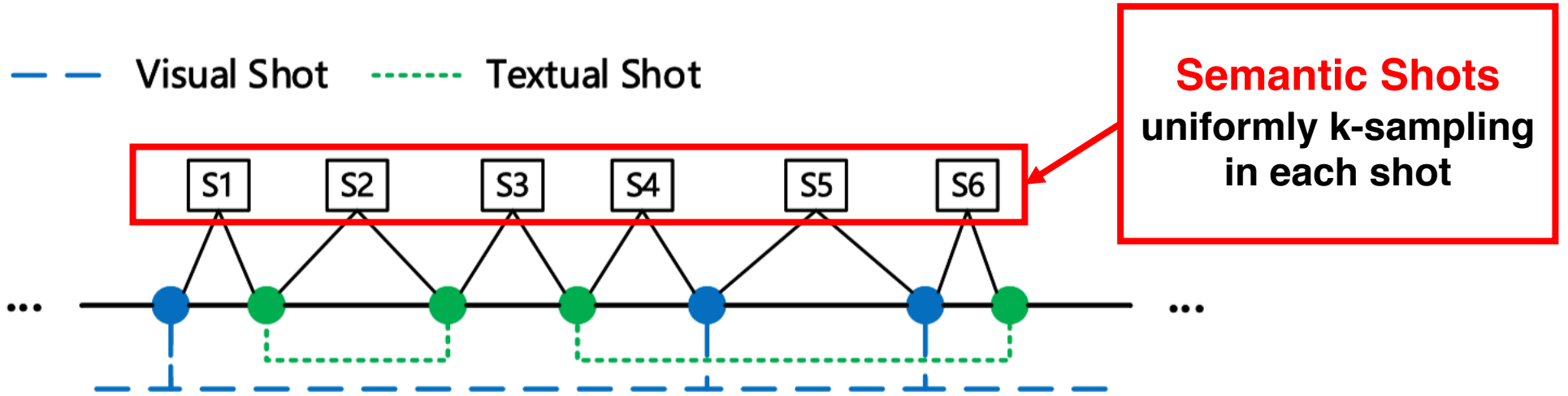
Key Frame Detection



Multimodal Content Embedding

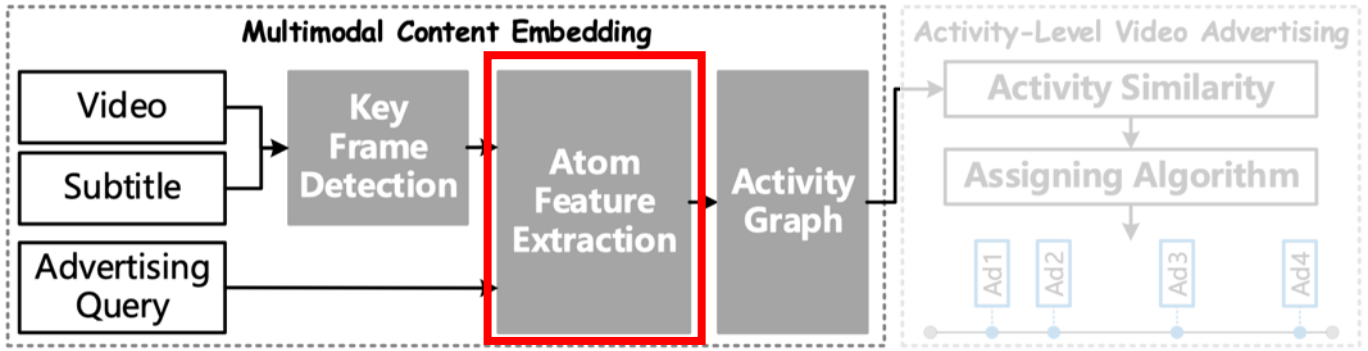
Key Frame Detection

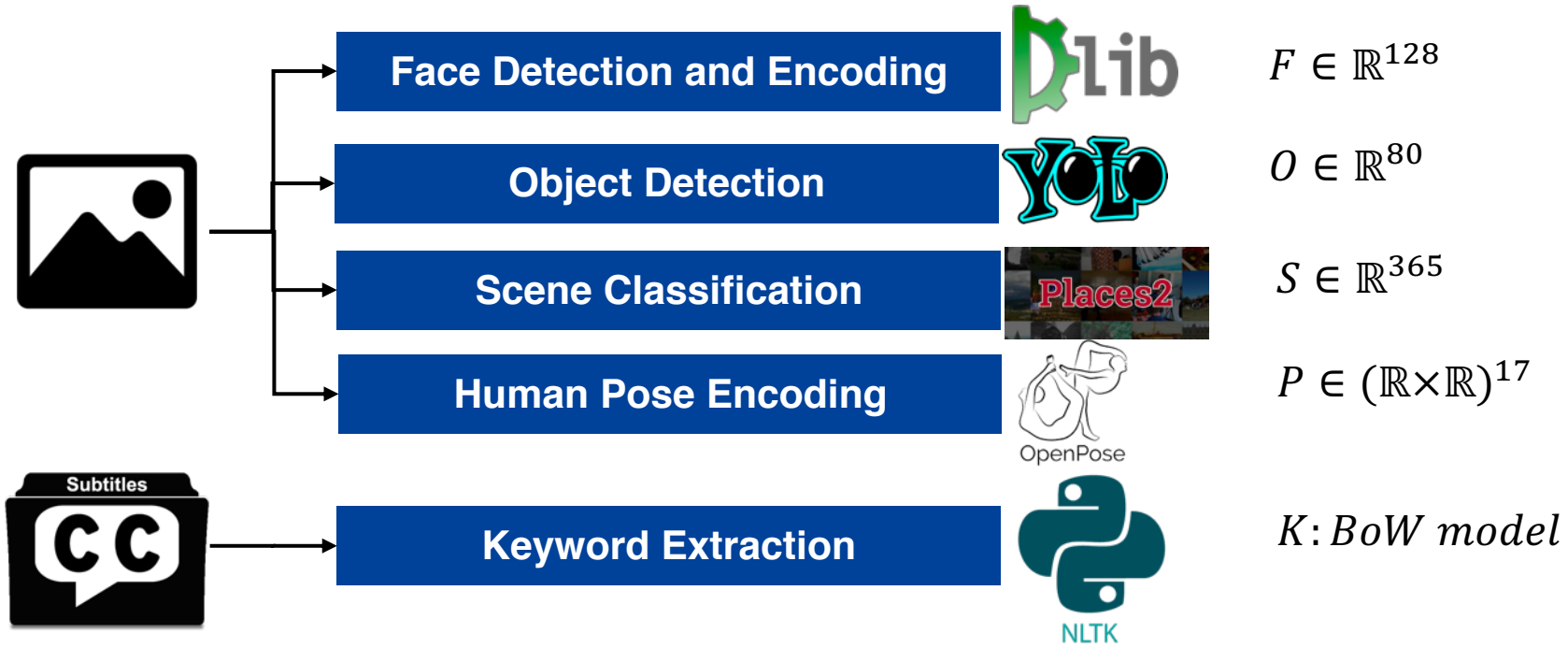




Multimodal Content Embedding

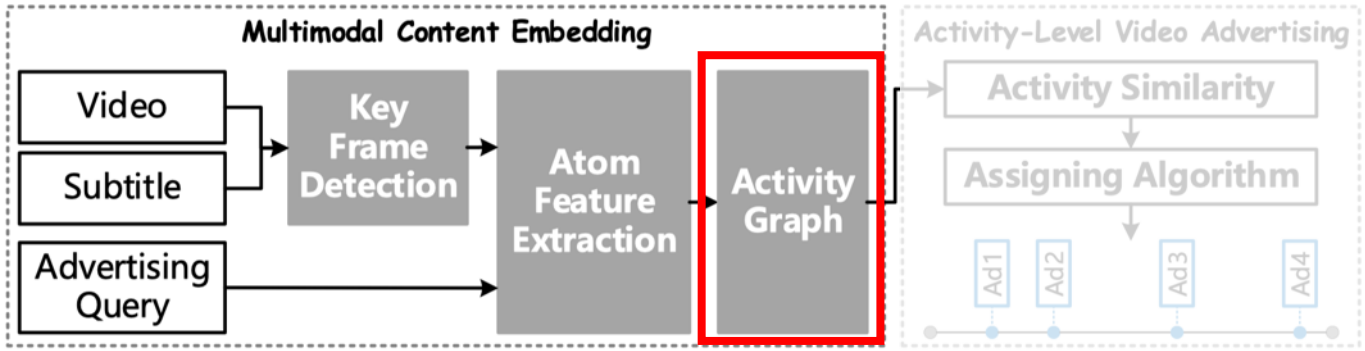
Atom Feature Extraction



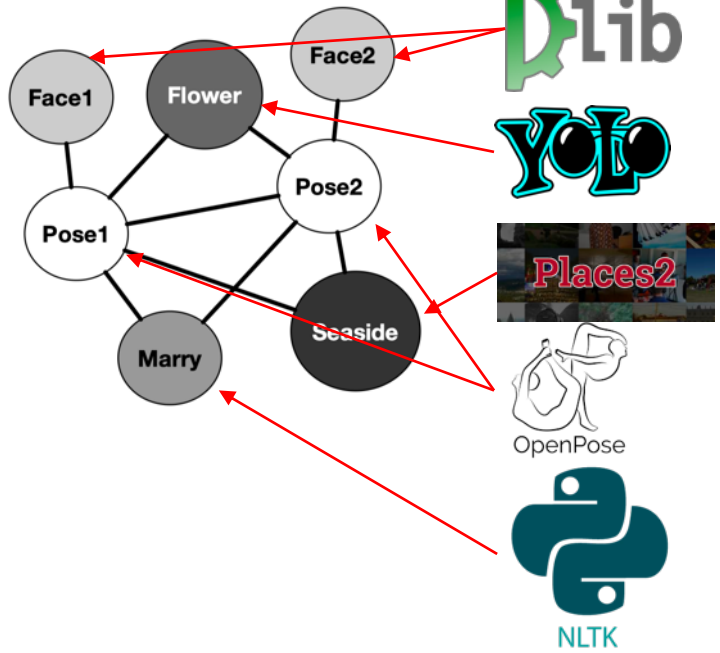


Multimodal Content Embedding

Activity Graph Representation



Vertices



$F \in \mathbb{R}^{128}$

$O \in \mathbb{R}^{80}$

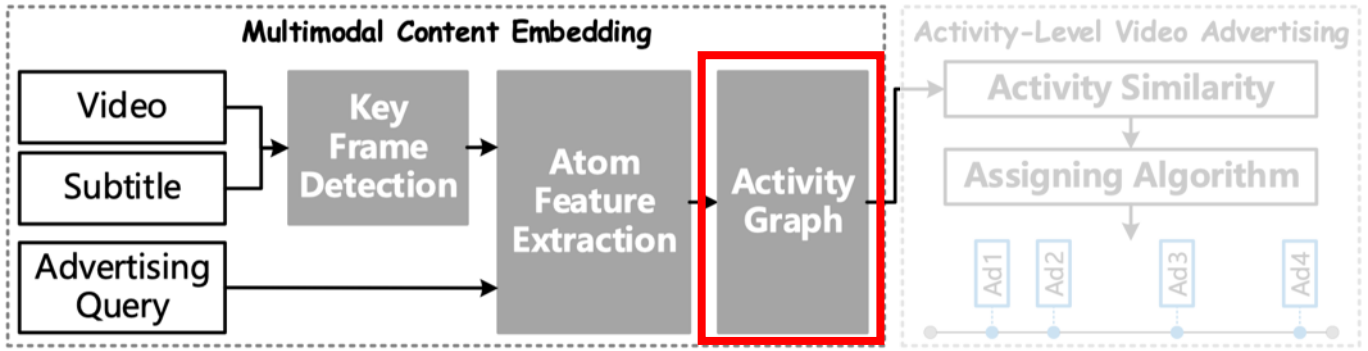
$S \in \mathbb{R}^{365}$

$P \in (\mathbb{R}\times\mathbb{R})^{17}$

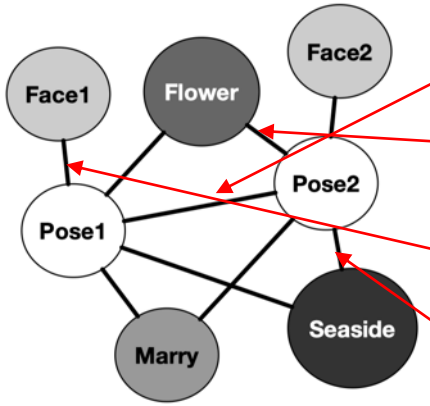
$K: BoW\ model$

Multimodal Content Embedding

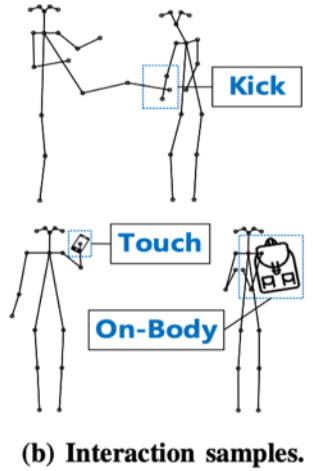
Activity Graph Representation



Edges



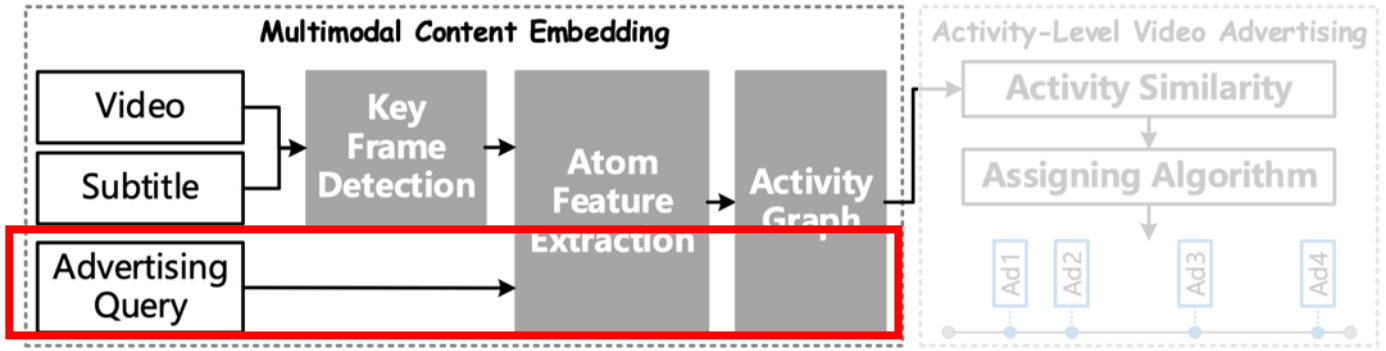
- **Pose-Pose Interaction** $R_{pp}: P \times P \rightarrow \mathbb{N}$
- **Pose-Object Interaction** $R_{po}: P \times O \rightarrow \mathbb{N}$
- **Pose-Face Matching** $R_{pf}: P \times F \rightarrow \mathbb{N}$
- **Pose-Scene Correlation** $R_{ps}: P \times S \rightarrow \mathbb{R}$

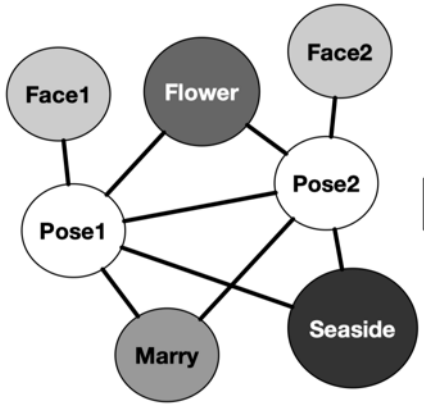


(b) Interaction samples.

Multimodal Content Embedding

Activity Graph Representation





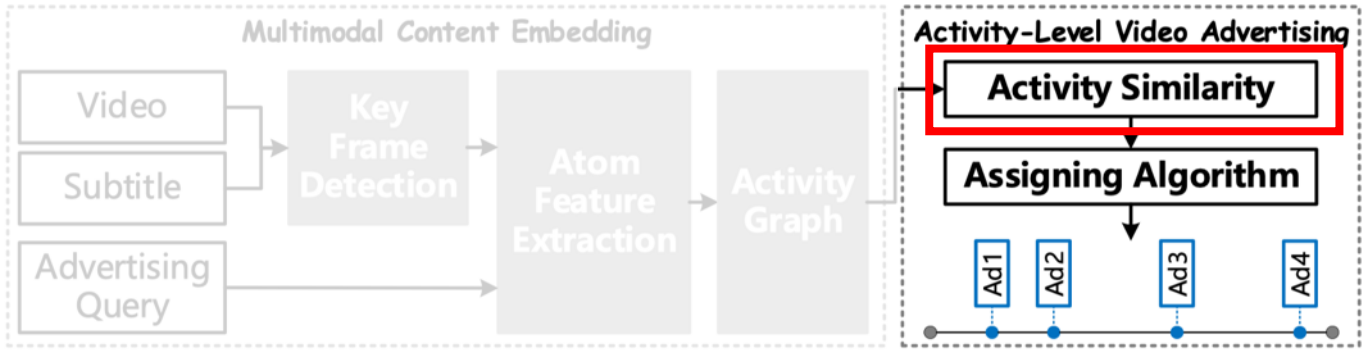
marry, travel, jewelry, ...

Outline

- Introduction
- Multimodal Content Embedding
- **Activity-Level Video Advertising**
- Evaluation
- Conclusion

Activity-Level Video Advertising

Activity Similarity



Face / Human Pose

$x = (encoding, confidence)$

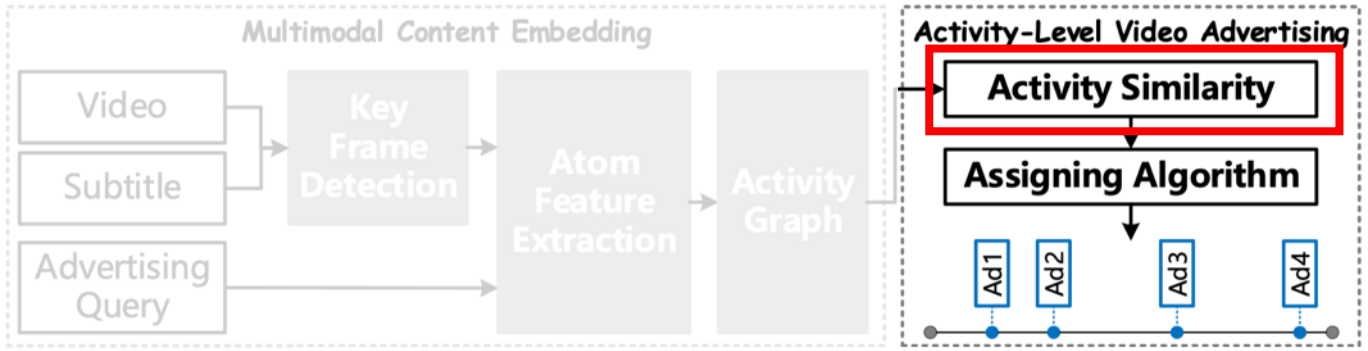
Atom-Feature / Vertex Similarity

$$s_1(x_1, x_2) = 1 - \frac{1}{|enc_1|} \sum_{\substack{e_i \in enc_1 \\ c_i \in conf_1}} \min_{\substack{e_j \in enc_2 \\ c_j \in conf_2}} \frac{\theta_1 ||e_i - e_j||_2}{(1 + \min(c_i, c_j))}$$

Interaction / Edge Similarity

Activity-Level Video Advertising

Activity Similarity



Atom-Feature / Vertex Similarity

Face / Human Pose

$x = (encoding, confidence)$

Object / Scene / Keyword

$x = (category, confidence)$

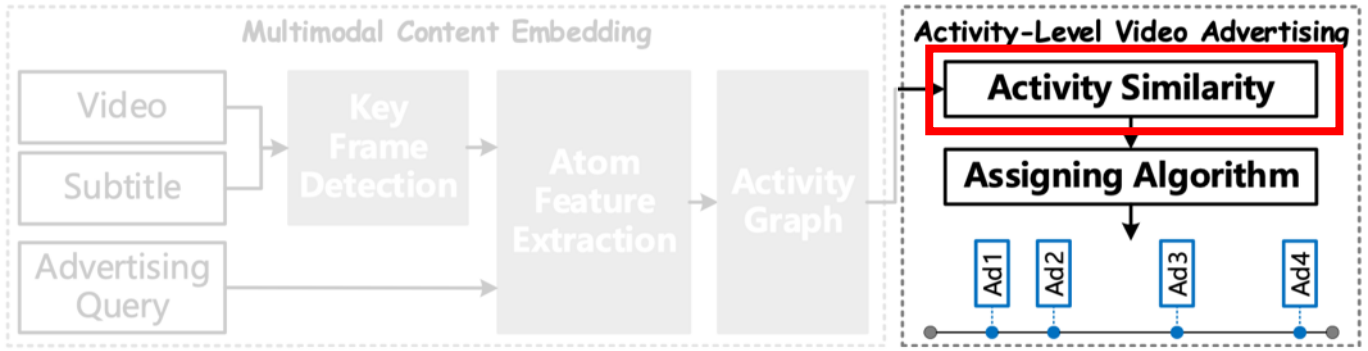
$$s_1(x_1, x_2) = 1 - \frac{1}{|enc_1|} \sum_{\substack{e_i \in enc_1 \\ c_i \in conf_1}} \min_{\substack{e_j \in enc_2 \\ c_j \in conf_2}} \frac{\theta_1 ||e_i - e_j||_2}{(1 + \min(c_i, c_j))}$$

$$s_2(x_1, x_2) = \frac{1}{|cls_1|} \sum_{\substack{e_i \in cls_1 \\ c_i \in conf_1}} \max_{\substack{e_j \in cls_2 \\ c_j \in conf_2}} (\theta_2 \min(c_i, c_j) \rho(e_i, e_j))$$

Interaction / Edge Similarity

Activity-Level Video Advertising

Activity Similarity



Atom-Feature / Vertex Similarity

Face / Human Pose
$x = (encoding, confidence)$

Object / Scene / Keyword
$x = (category, confidence)$

$$s_1(x_1, x_2) = 1 - \frac{1}{|enc_1|} \sum_{\substack{e_i \in enc_1 \\ c_i \in conf_1}} \min_{\substack{e_j \in enc_2 \\ c_j \in conf_2}} \frac{\theta_1 ||e_i - e_j||_2}{(1 + \min(c_i, c_j))}$$

$$s_2(x_1, x_2) = \frac{1}{|cls_1|} \sum_{\substack{e_i \in cls_1 \\ c_i \in conf_1}} \max_{\substack{e_j \in cls_2 \\ c_j \in conf_2}} (\theta_2 \min(c_i, c_j) \rho(e_i, e_j))$$

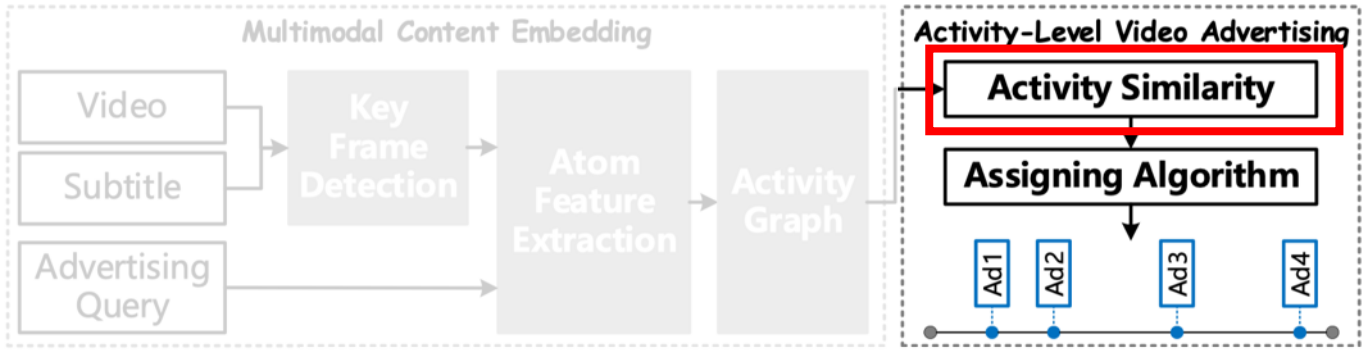
Interaction / Edge Similarity

Pose-Scene
numerical

$$s_3(e_i, e_j) = \begin{cases} 1 - ||e_i - e_j||_2, & \text{numerical} \\ \rho(e_i, e_j), & \text{categorical} \end{cases}$$

Activity-Level Video Advertising

Activity Similarity



Atom-Feature / Vertex Similarity

- **Face / Human Pose**
  $x = (encoding, confidence)$
- **Object / Scene / Keyword**
  $x = (category, confidence)$

$$s_1(x_1, x_2) = 1 - \frac{1}{|enc_1|} \sum_{\substack{e_i \in enc_1 \\ c_i \in conf_1}} \min_{\substack{e_j \in enc_2 \\ c_j \in conf_2}} \frac{\theta_1 ||e_i - e_j||_2}{(1 + \min(c_i, c_j))}$$

$$s_2(x_1, x_2) = \frac{1}{|cls_1|} \sum_{\substack{e_i \in cls_1 \\ c_i \in conf_1}} \max_{\substack{e_j \in cls_2 \\ c_j \in conf_2}} (\theta_2 \min(c_i, c_j) \rho(e_i, e_j))$$

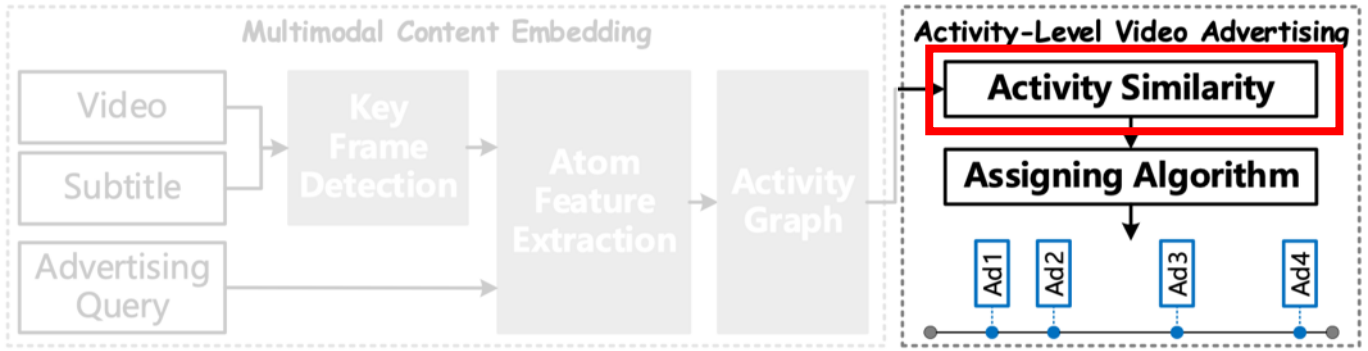
Interaction / Edge Similarity

- **Pose-Scene**
  numerical
- **Pose-Pose / Pose-Object / Pose-Face**
  categorical

$$s_3(e_i, e_j) = \begin{cases} 1 - ||e_i - e_j||_2, & \text{numerical} \\ \rho(e_i, e_j), & \text{categorical} \end{cases}$$

Activity-Level Video Advertising

Activity Similarity



Atom-Feature / Vertex Similarity

Face / Human Pose
$x = (encoding, confidence)$

$$s_1(x_1, x_2) = 1 - \frac{1}{|enc_1|} \sum_{\substack{e_i \in enc_1 \\ c_i \in conf_1}} \min_{\substack{e_j \in enc_2 \\ c_j \in conf_2}} \frac{\theta_1 ||e_i - e_j||_2}{(1 + \min(c_i, c_j))}$$

Object / Scene / Keyword

$$s_2(x_1, x_2) = \frac{1}{|cls_1|} \sum_{\substack{e_i \in cls_1 \\ c_i \in conf_1}} \max_{\substack{e_j \in cls_2 \\ c_j \in conf_2}} (\theta_2 \min(c_i, c_j) \rho(e_i, e_j))$$

$$S(g_1, g_2) = \sum w_i s_i(g_1, g_2)$$

Interaction / Edge Similarity

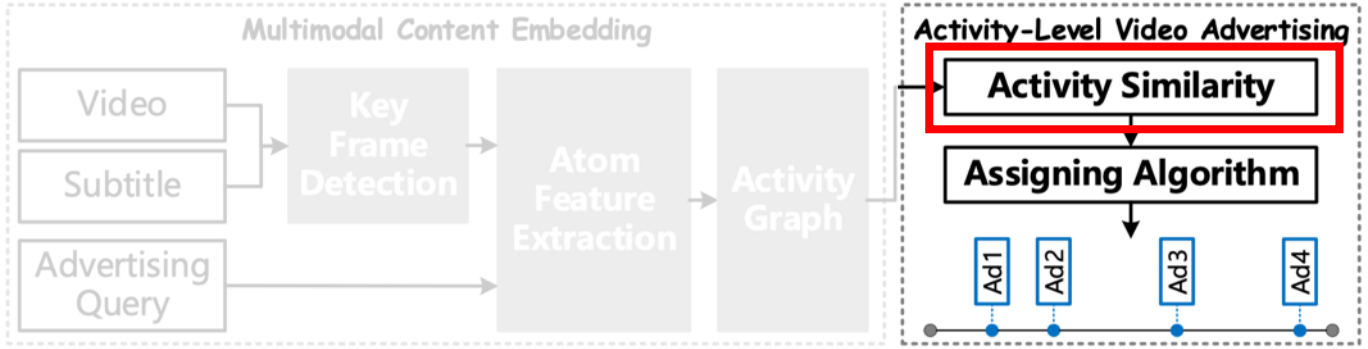
numerical

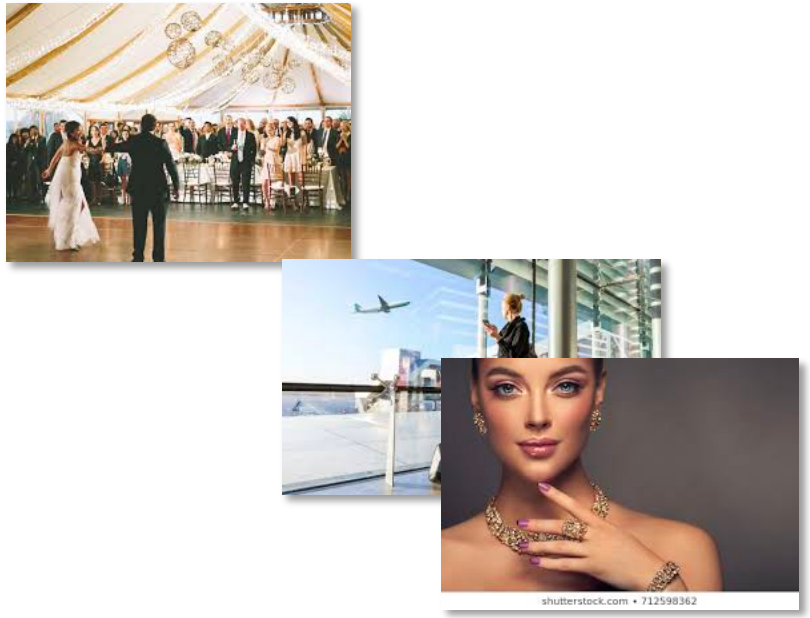
Pose-Pose / Pose-Object / Pose-Face
categorical

$$s_3(e_i, e_j) = \begin{cases} 1 - ||e_i - e_j||_2, & \text{numerical} \\ \rho(e_i, e_j), & \text{categorical} \end{cases}$$

Activity-Level Video Advertising

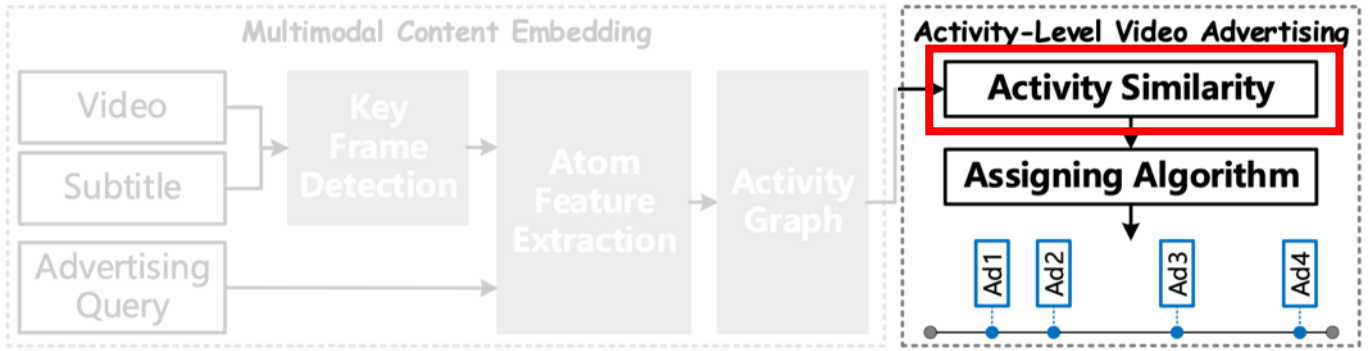
Activity Similarity





Activity-Level Video Advertising

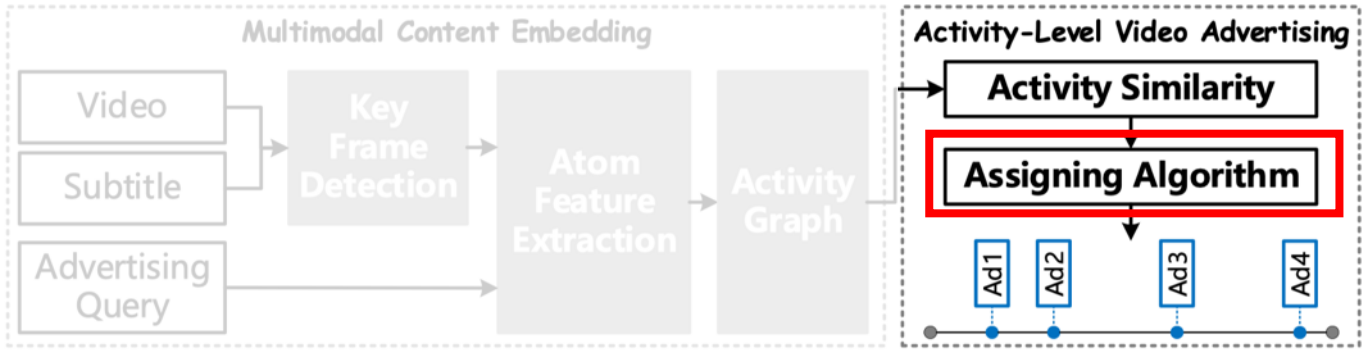
Activity Similarity



Scene: Airport

Activity-Level Video Advertising

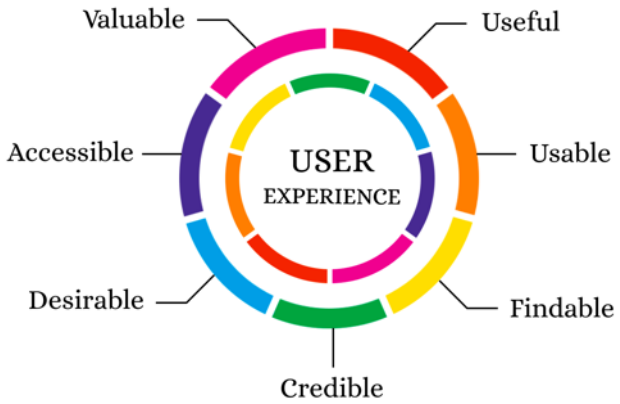
Ads Assignment



Ads Revenue

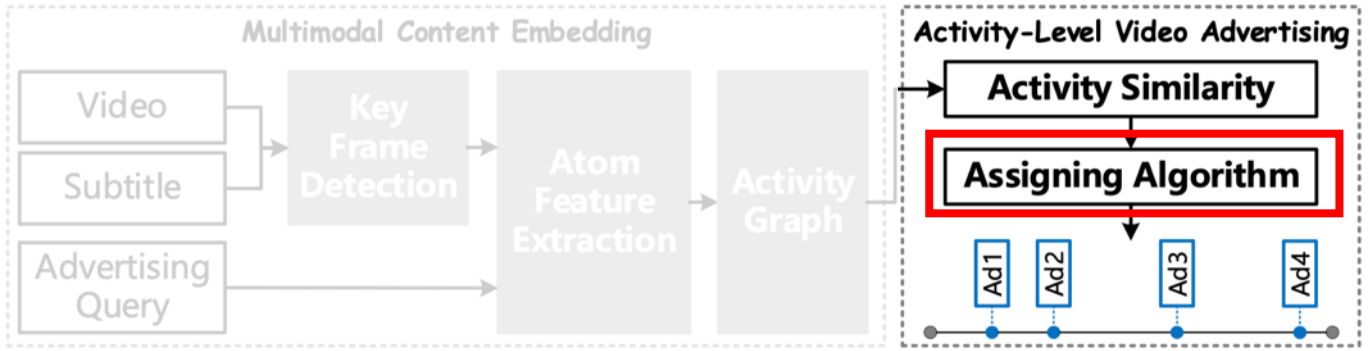
VS.

User Experience



Activity-Level Video Advertising

Ads Assignment: Ads Revenue



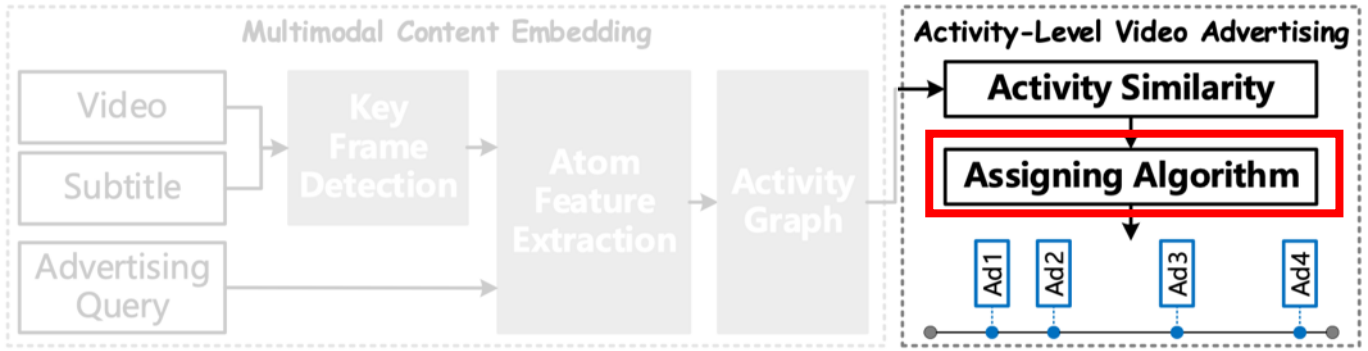
Ads Revenue

$$f(S) = \sum_{(x_i,y_j)\in S} r_j s(x_i, y_j)$$

The subset of selected ads assignments.

Activity-Level Video Advertising

Ads Assignment: Ads Revenue



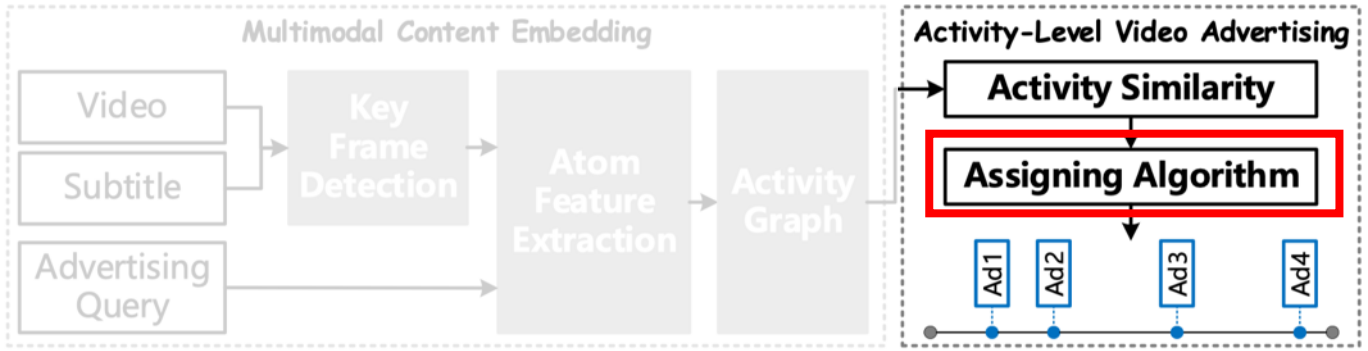
Ads Revenue

$$f(S) = \sum_{(x_i, y_j) \in S} r_j s(x_i, y_j)$$

An assignment is a pair of **key frame** and **ad query**.

Activity-Level Video Advertising

Ads Assignment: Ads Revenue



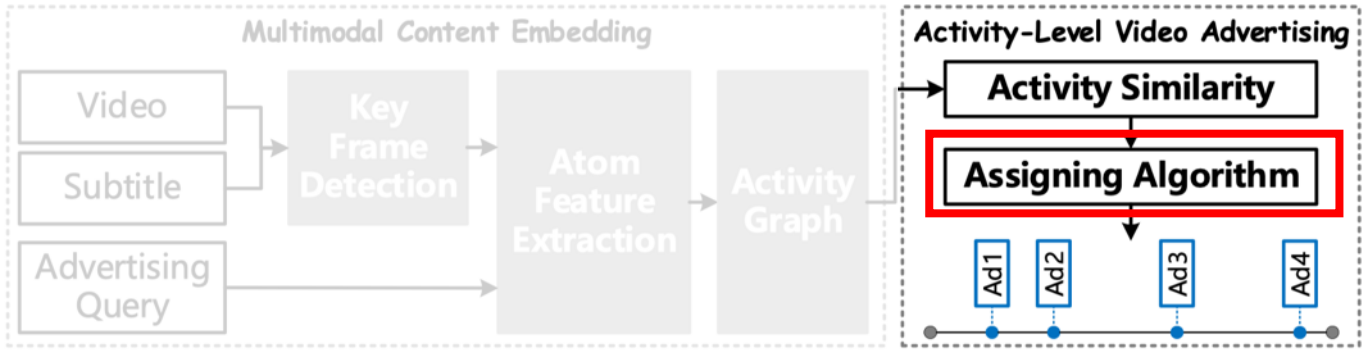
Ads Revenue

$$f(S) = \sum_{(x_i, y_j) \in S} r_j s(x_i, y_j)$$

Activity similarity between the key frame and ad query.

Activity-Level Video Advertising

Ads Assignment: Ads Revenue



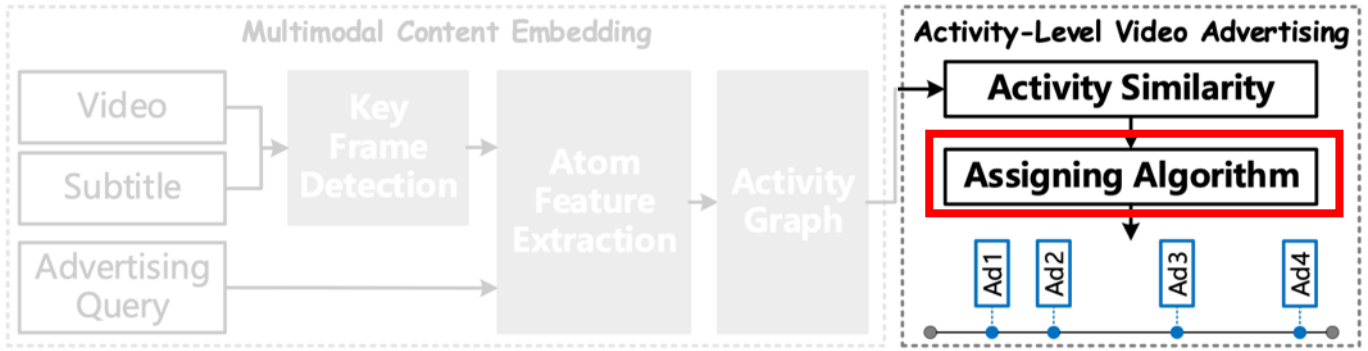
Ads Revenue

$$f(S) = \sum_{(x_i, y_j) \in S} r_j s(x_i, y_j)$$

Query-j revenue.

Activity-Level Video Advertising

Ads Assignment: Intrusiveness Perception

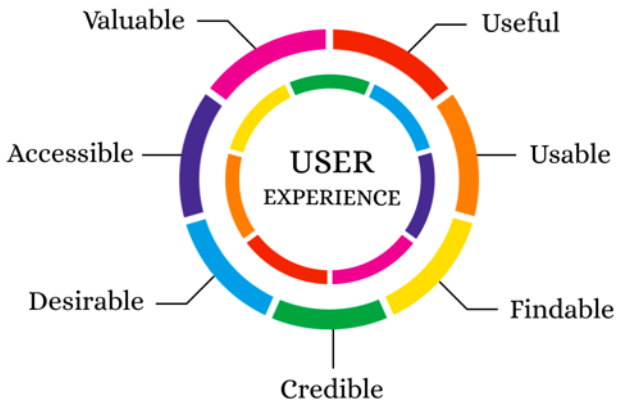


Intrusiveness Distribution Function

$$g(t|S) = 0 \quad ,0 \leq t \leq t_1$$

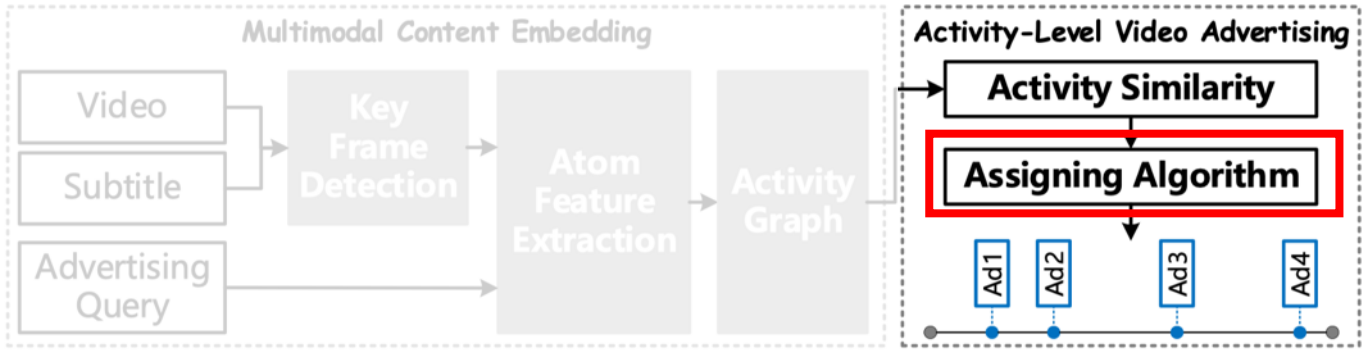
The intrusiveness of ads is the function over **viewing time** given an assigning output.

User Experience



Activity-Level Video Advertising

Ads Assignment: Intrusiveness Perception

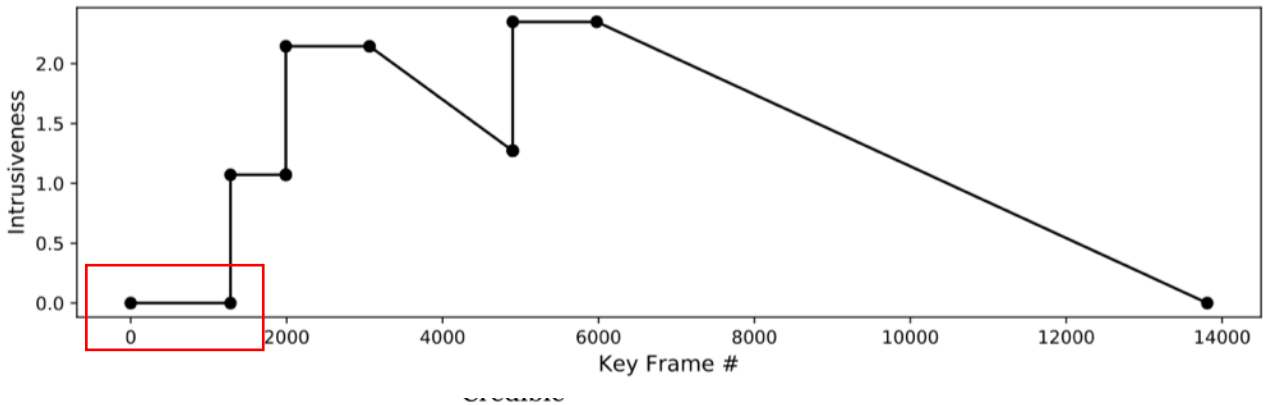


Intrusiveness Distribution Function

$$g(t|S) = 0 \quad ,0 \leq t \leq t_1$$

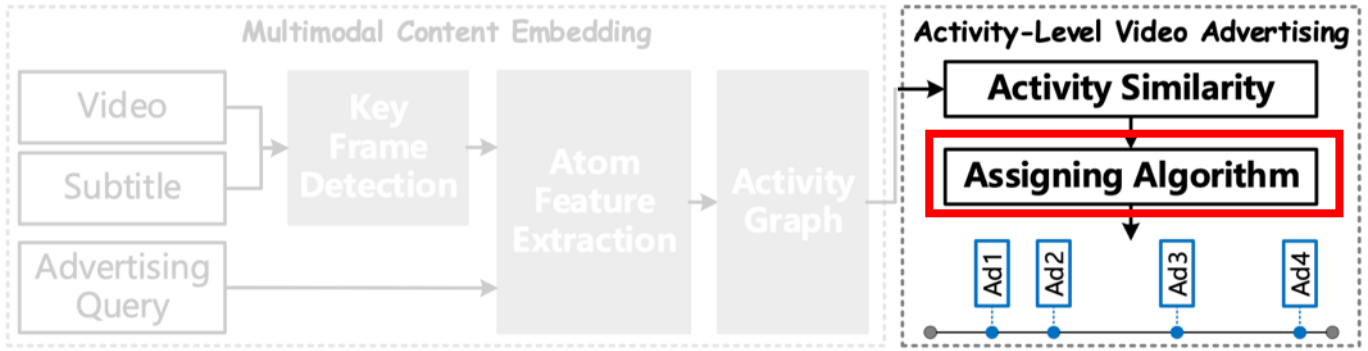
Before the first ad, the intrusiveness is 0.

User Experience



Activity-Level Video Advertising

Ads Assignment: Intrusiveness Perception



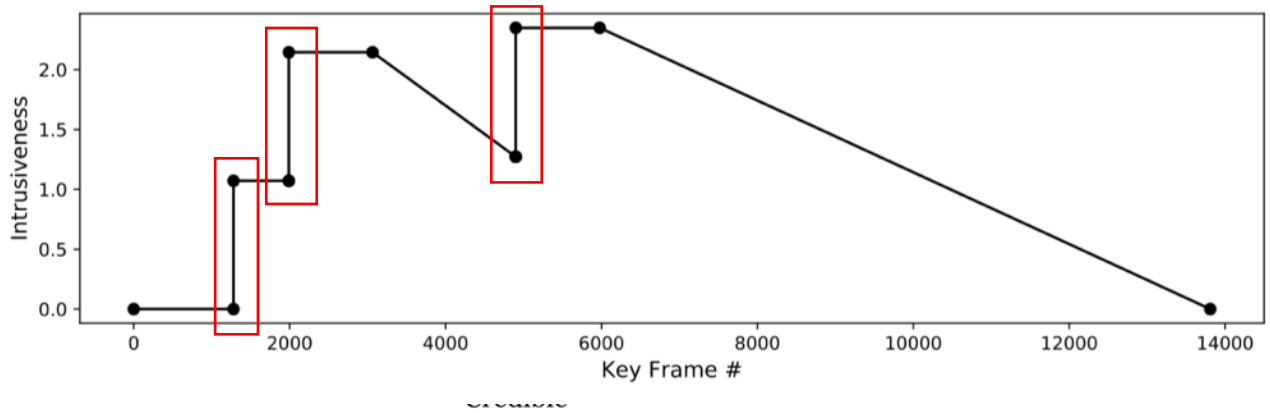
Intrusiveness Distribution Function

$$g(t|S) = 0 \quad , 0 \le t \le t_1$$

$$g(t|S) = g(t_i) + \frac{\alpha}{s(p_i)} \quad , t_i < t \le t_{i+1}$$

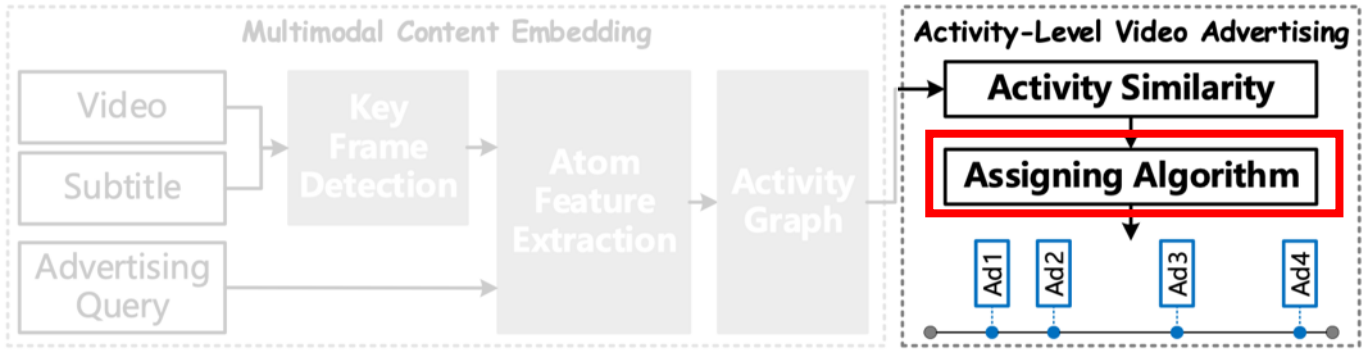
The **incremental intrusiveness** of a new ad is inversely proportional to the content similarity.

User Experience



Activity-Level Video Advertising

Ads Assignment: Intrusiveness Perception



Intrusiveness Distribution Function

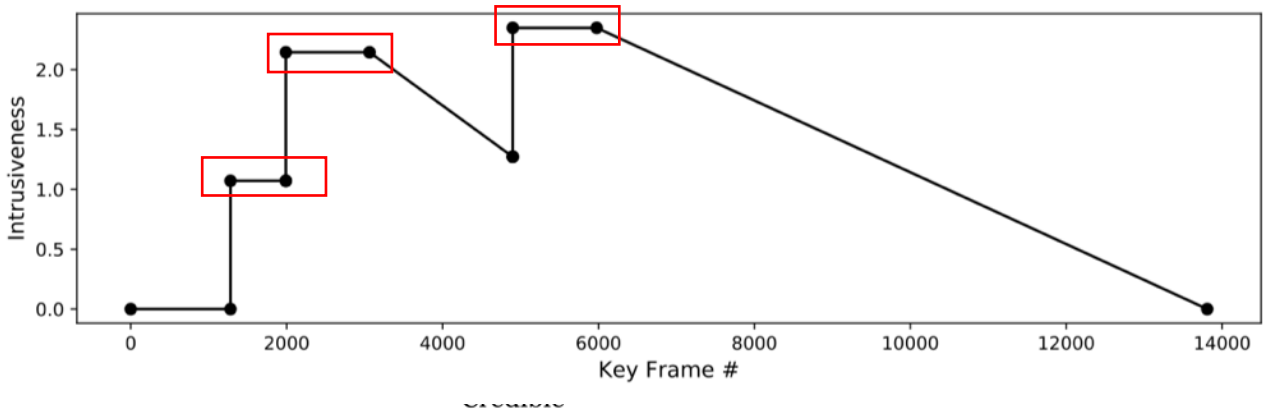
$$g(t|S) = 0 \quad , 0 \le t \le t_1$$

$$g(t|S) = g(t_i) + \frac{\alpha}{s(p_i)} \quad , t_i < t \le t_{i+1}$$

$$g(t|S) = g(t_i) + \frac{\alpha}{s(p_i)} - \gamma(t - t_i - \frac{\beta}{s(p_i)})$$

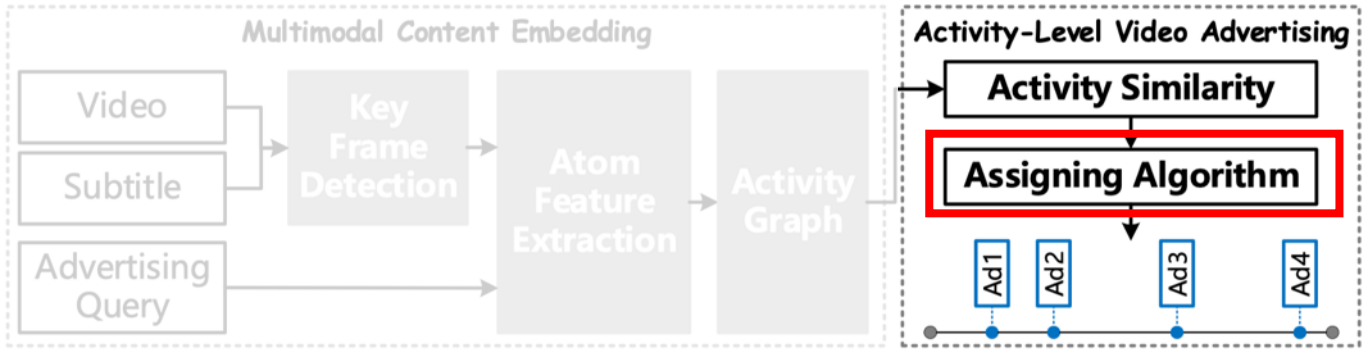
The duration of high intrusive perception is also inversely proportional to the similarity score.

User Experience



Activity-Level Video Advertising

Ads Assignment: Intrusiveness Perception



Intrusiveness Distribution Function

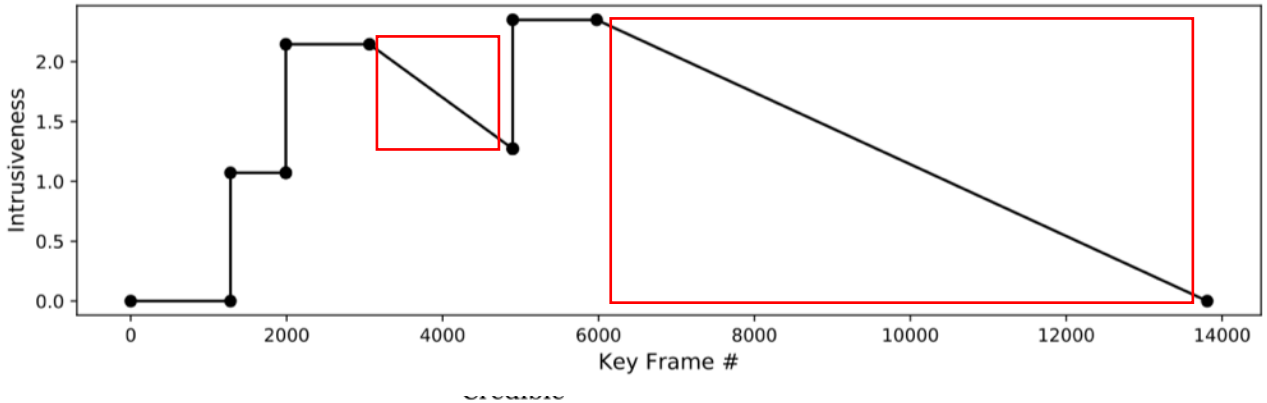
$$g(t|S) = 0 \quad , 0 \le t \le t_1$$

$$g(t|S) = g(t_i) + \frac{\alpha}{s(p_i)} \quad , t_i < t \le t_{i+1}$$

$$g(t|S) = g(t_i) + \frac{\alpha}{s(p_i)} - \gamma(t - t_i - \frac{\beta}{s(p_i)})$$

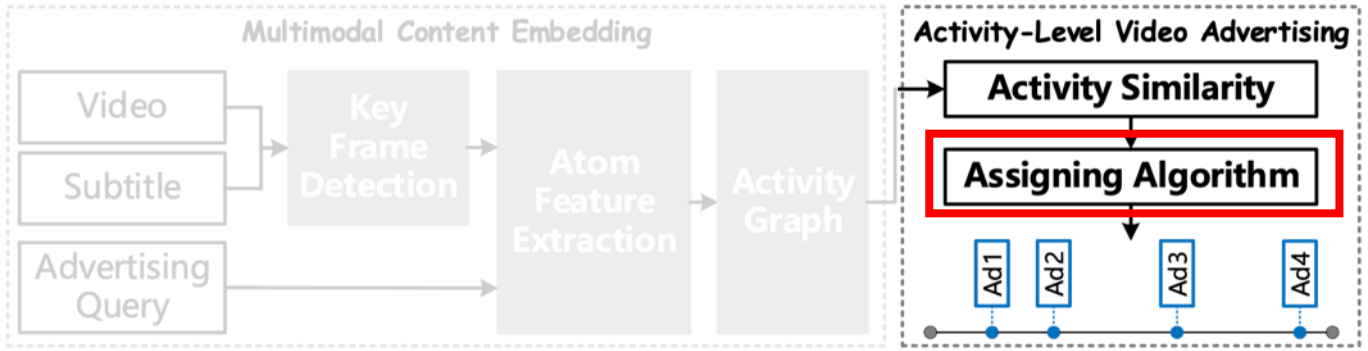
As the time goes on, the intrusiveness will decline and γ controls the descent speed.

User Experience



Activity-Level Video Advertising

Ads Assignment: Trade-Off Model



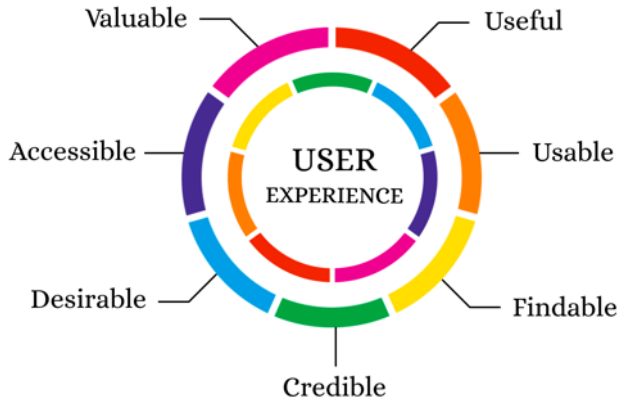
Ads Revenue

Trade-Off Model: Revenue Optimization under Intrusiveness Constraint

$$\max_{S \subseteq P} f(S) + \delta E(S)$$

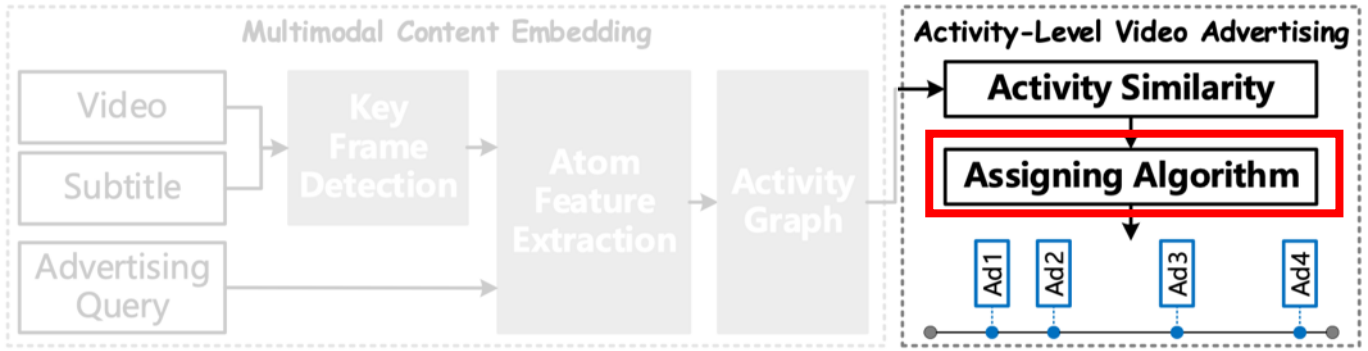
$$\text{s.t.} \quad \int_0^{t_{max}} g(t|S)\mathrm{d}t \leq B$$

User Experience



Activity-Level Video Advertising

Ads Assignment: Trade-Off Model



Ads Revenue

$$E(S) = -\sum_{(x_i,y_j)\in S} p_{y_j} \log_2 \frac{p_{y_j}}{N_y}$$

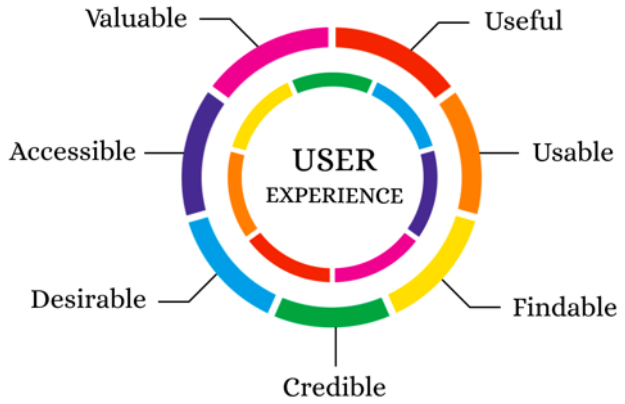
Trade-Off Model: Revenue Optimization under Intrusiveness Constraint

$$\max_{S\subseteq P} f(S) + \delta E(S)$$

$$\text{s.t.} \quad \int_0^{t_{max}} g(t|S)\mathrm{d}t \le B$$

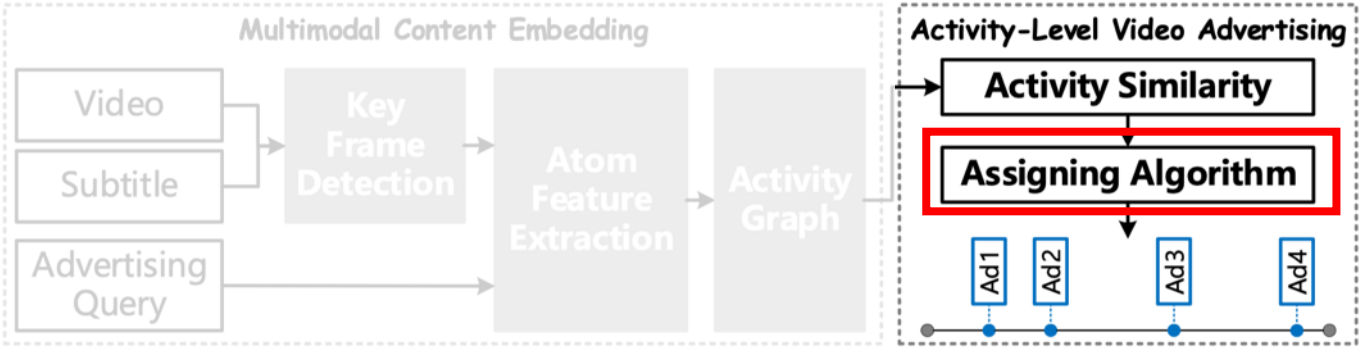
E(S) is an entropy-like function that measures balance of served ad queries distribution.

User Experience



Activity-Level Video Advertising

Ads Assignment: NMS-Greedy Algorithm



Trade-Off Model:
Revenue Optimization under Intrusiveness Constraint

$$\max_{S \subseteq P} f(S) + \delta E(S)$$

$$\text{s.t.} \quad \int_0^{t_{max}} g(t|S)\mathrm{d}t \leq B$$

Algorithm 1 NMS-Greedy algorithm for Equation 9.

Require: key frame set X, ad set Y, intrusiveness budget B
Ensure: assignments S

1: Initialize the value matrix $M_{N_y \times N_x}$, where $m_{ij} = r_i s(x_j, y_i)$.
2: Apply the kernel $K_{11 \times N_x}$ to M.
3: Apply the kernel $K_{2 N_y \times N_k}$ to M.
4: Greedily select assignments with $\arg\max m_{ij}$ into S until $\int_0^{t_{max}} g(t|S)\mathrm{d}t$ exceeds the intrusiveness budget B.
5: **return** S

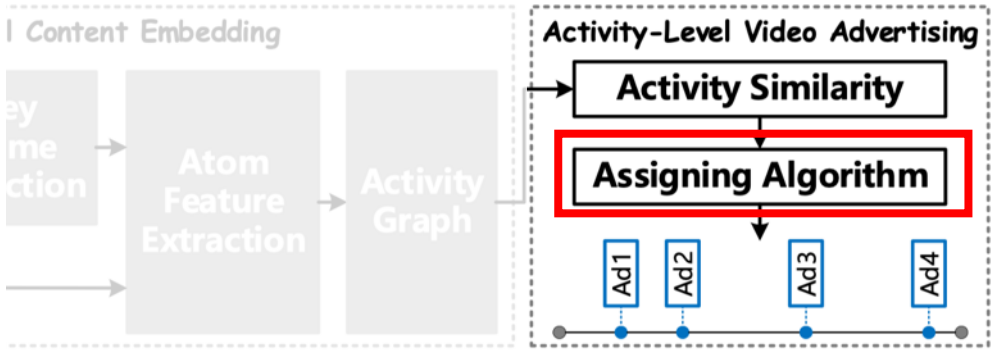
Activity-Level Video Advertising

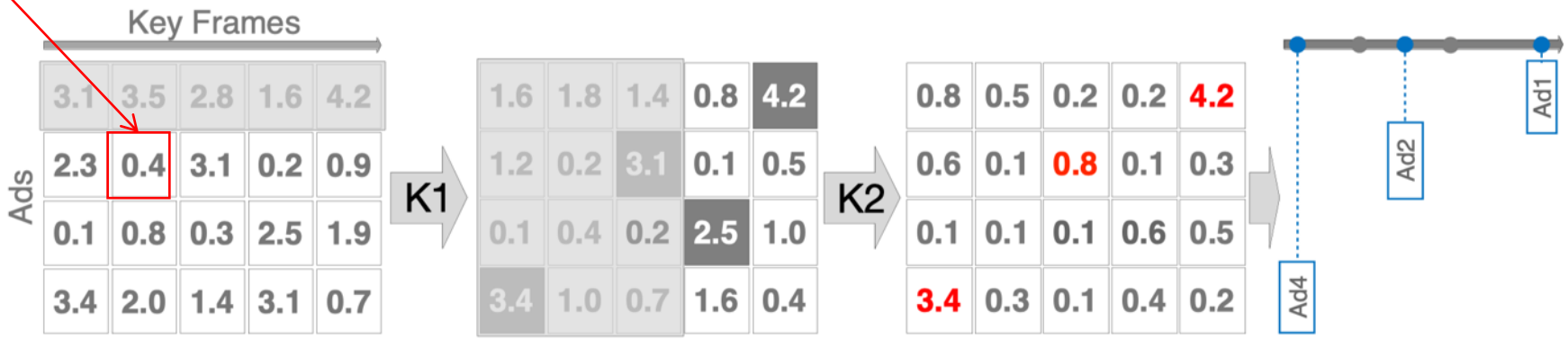
Ads Assignment: NMS-Greedy Algorithm

Algorithm 1 NMS-Greedy algorithm for Equation 9.

Require: key frame set X, ad set Y, intrusiveness budget B
Ensure: assignments S

1: Initialize the value matrix $M_{N_y \times N_x}$, where $m_{ij} = r_i s(x_j, y_i)$.
2: Apply the kernel $K_{11 \times N_x}$ to M.
3: Apply the kernel $K_{2 N_y \times N_k}$ to M.
4: Greedily select assignments with $\arg\max m_{ij}$ into S until $\int_0^{t_{max}} g(t|S)\mathrm{d}t$ exceeds the intrusiveness budget B.
5: **return** S





Calculating the pairwise similarity to construct the value matrix.

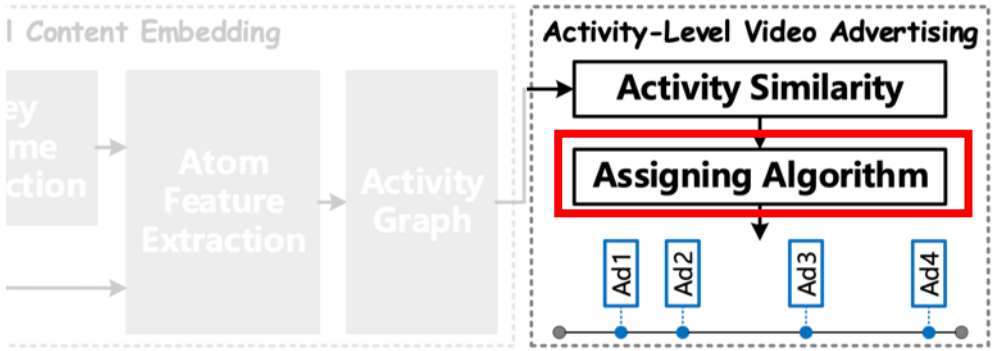
Activity-Level Video Advertising

Ads Assignment: NMS-Greedy Algorithm

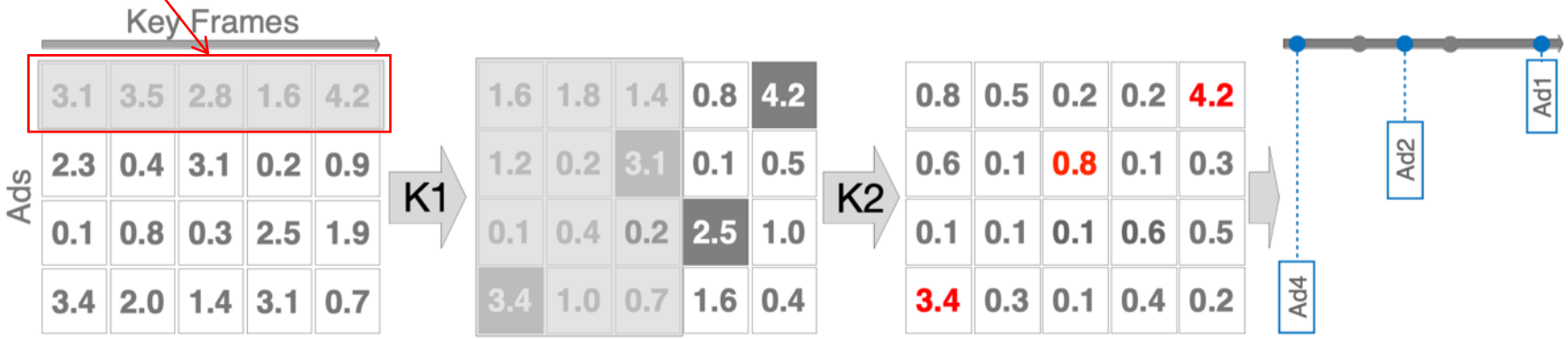
Algorithm 1 NMS-Greedy algorithm for Equation 9.

Require: key frame set X, ad set Y, intrusiveness budget B
Ensure: assignments S

1: Initialize the value matrix $M_{N_y \times N_x}$, where $m_{ij} = r_i s(x_j, y_i)$.
2: Apply the kernel $K_{11 \times N_x}$ to M.
3: Apply the kernel $K_{2N_y \times N_k}$ to M.
4: Greedily select assignments with $\arg\max m_{ij}$ into S until $\int_0^{t_{max}} g(t|S)\mathrm{d}t$ exceeds the intrusiveness budget B.
5: **return** S



K1: Balancing the assignments for each ad.



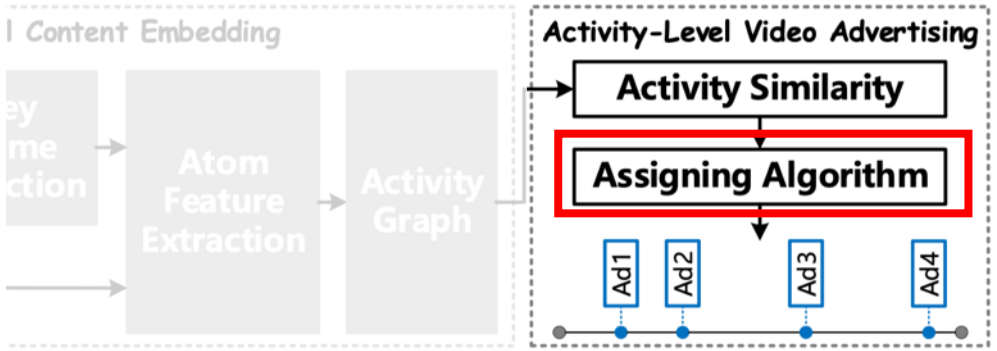
Activity-Level Video Advertising

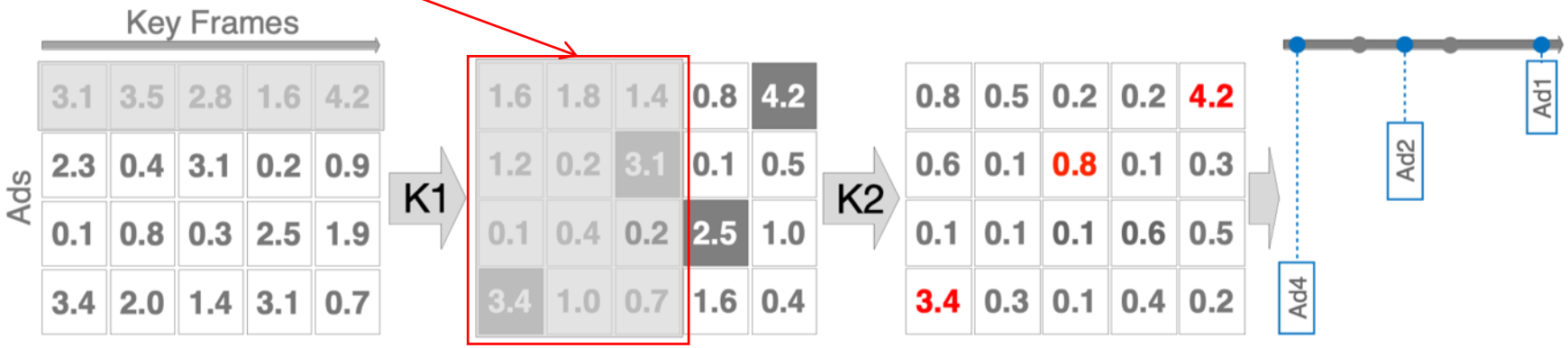
Ads Assignment: NMS-Greedy Algorithm

Algorithm 1 NMS-Greedy algorithm for Equation 9.

Require: key frame set X, ad set Y, intrusiveness budget B
Ensure: assignments S

1: Initialize the value matrix $M_{N_y \times N_x}$, where $m_{ij} = r_i s(x_j, y_i)$.
2: Apply the kernel $K_{11 \times N_x}$ to M.
3: Apply the kernel $K_{2 N_y \times N_k}$ to M.
4: Greedily select assignments with $\arg\max m_{ij}$ into S until $\int_0^{t_{max}} g(t|S)\mathrm{d}t$ exceeds the intrusiveness budget B.
5: **return** S





K2: Avoiding successive ads within short interval.

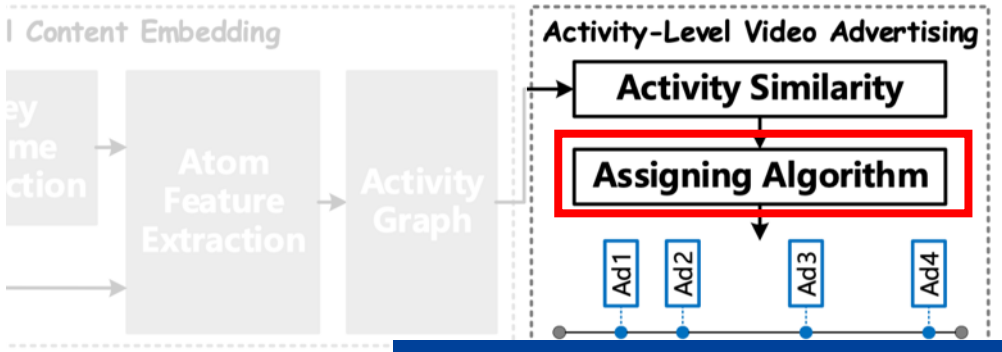
Activity-Level Video Advertising

Ads Assignment: NMS-Greedy Algorithm

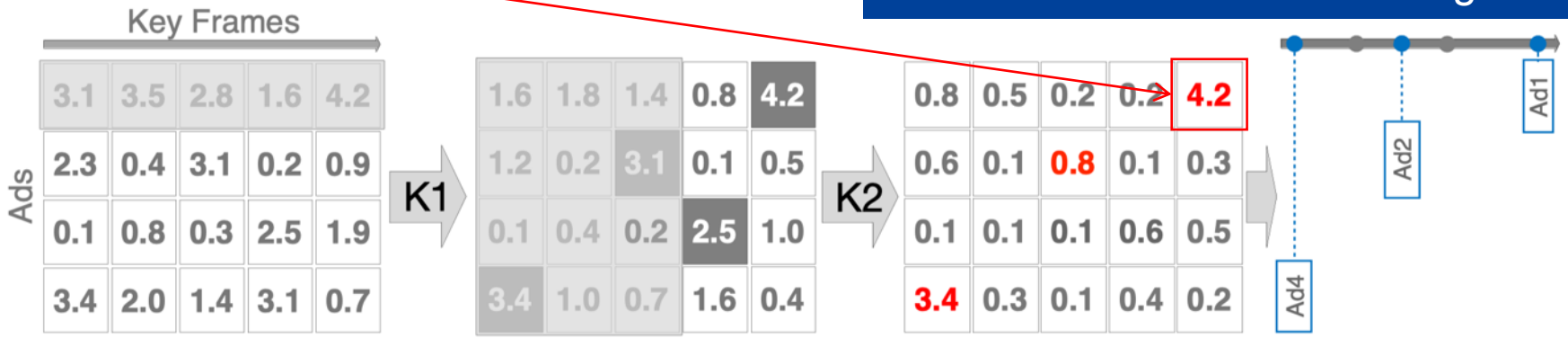
Algorithm 1 NMS-Greedy algorithm for Equation 9.

Require: key frame set X, ad set Y, intrusiveness budget B
Ensure: assignments S

1: Initialize the value matrix $M_{N_y \times N_x}$, where $m_{ij} = r_i s(x_j, y_i)$.
2: Apply the kernel $K_{1 1 \times N_x}$ to M.
3: Apply the kernel $K_{2 N_y \times N_k}$ to M.
4: Greedily select assignments with $\arg\max m_{ij}$ into S until $\int_0^{t_{max}} g(t|S)\mathrm{d}t$ exceeds the intrusiveness budget B.
5: **return** S

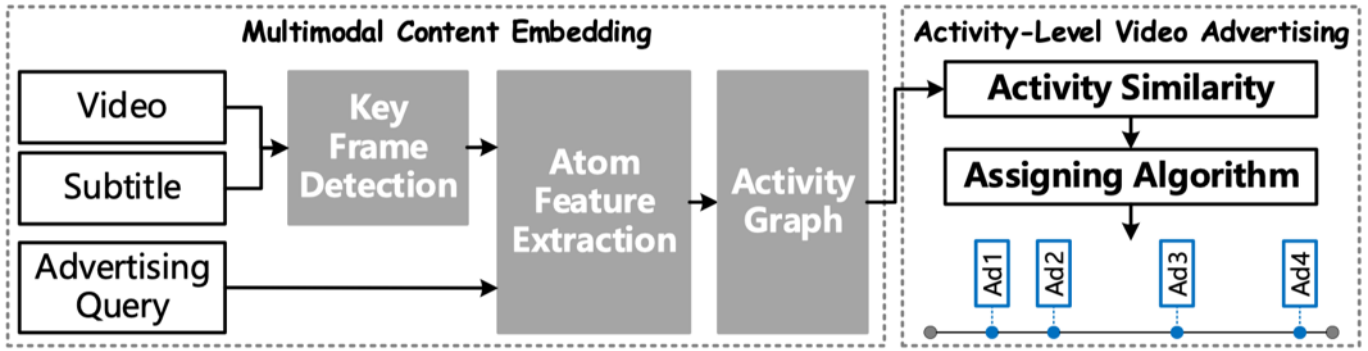


Greedy assigning ads until reaching intrusiveness budget.



Activity-Level Video Advertising

ActVA: Extensibility

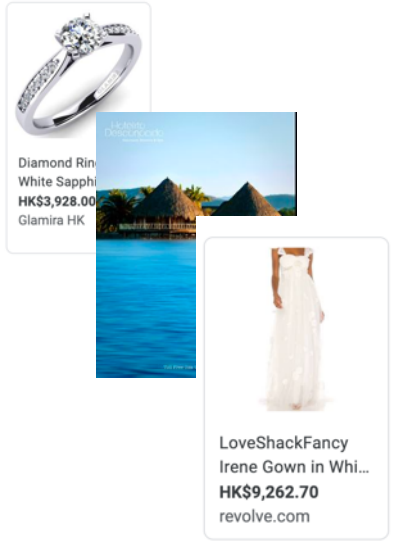


No need for activity-level labelled training data.

Video Data

&

Ads Query



→ **Activity Graph Representation**

Outline

- Introduction
- Multimodal Content Embedding
- Activity-Level Video Advertising
- **Evaluation**
- Conclusion

Evaluation

Experiment Setup

140 Videos

100X

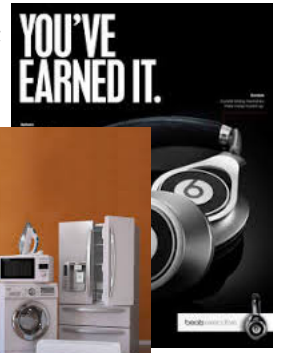
10X

30X

150 Real Ads Needs

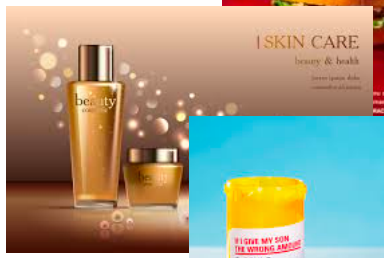
Vehicle

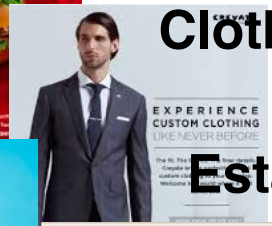
Electronic



Food and Beverage

Cosmetic

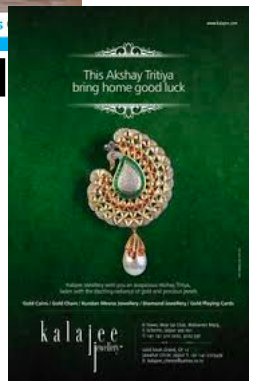




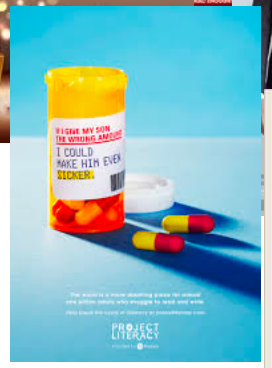
Clothing

Estate

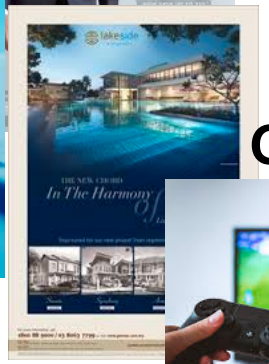
Household



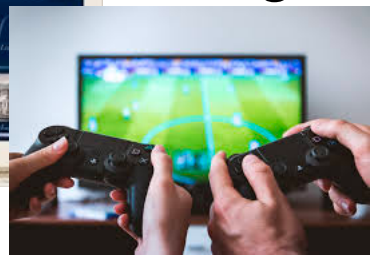
Jewelry



Medicine

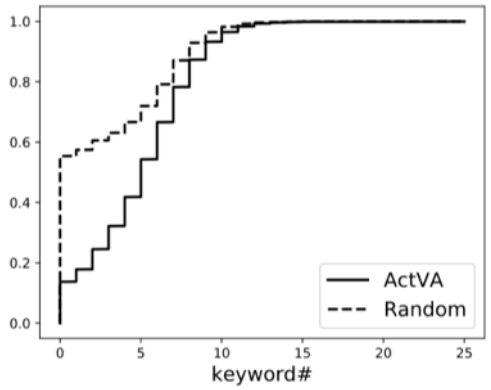


Gaming

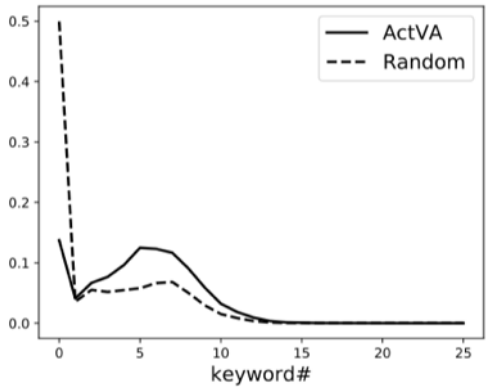


Evaluation

Semantic-Shot Key Frame Detection



(a) Keyword# CDF.



(b) Keyword# distribution.

Richer Textual Information

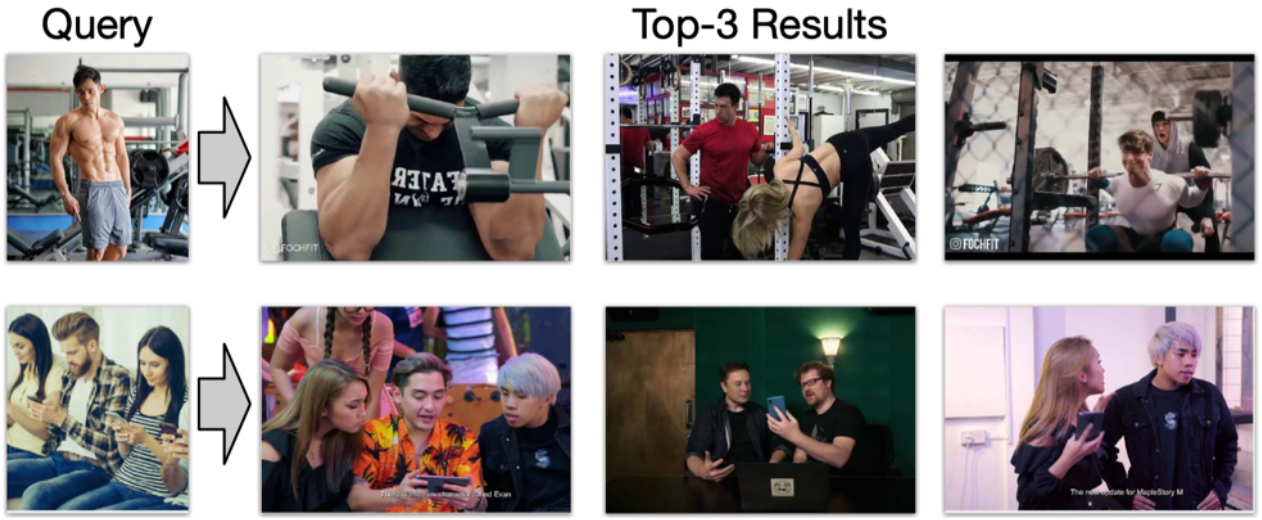
High Accuracy with >10k FPS retrieval speed

Evaluation

Advertising Position Retrieval

(a) Keyword# CDF.

(b) Keyword# distribution.

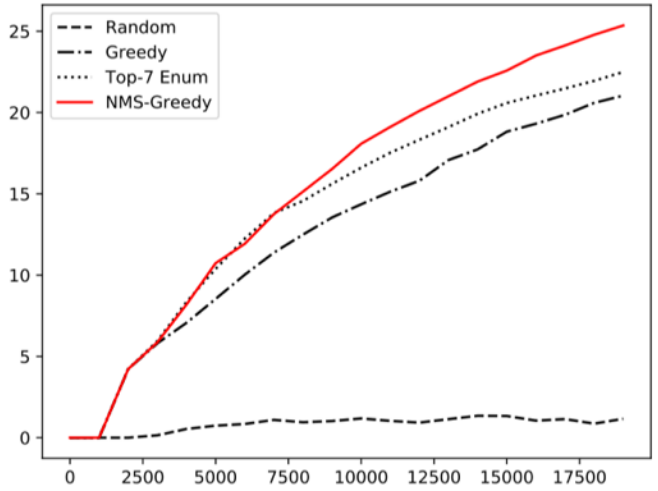


Richer Textual Information

High Accuracy with >10k FPS retrieval speed

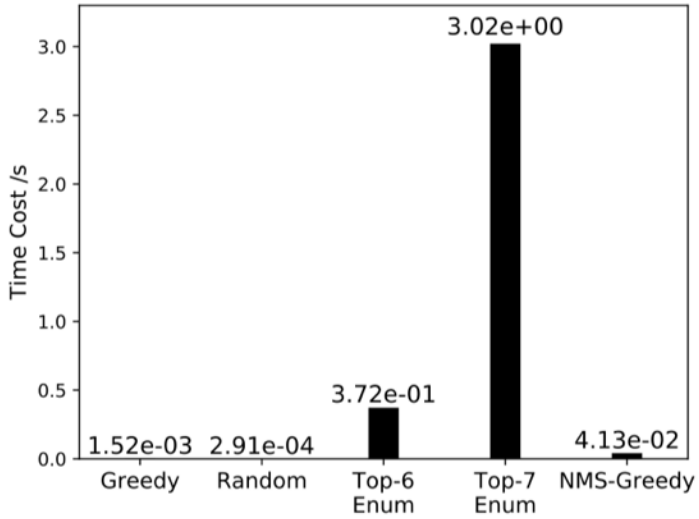
Evaluation

NMS-Greedy Ads Assignment



(a) Average Assignment Value V.S. Intrusiveness Budget

Highest Value



(b) Average Time Cost

Low Cost

(c) Subject Intrusiveness Evaluation

User Friendly

Evaluation

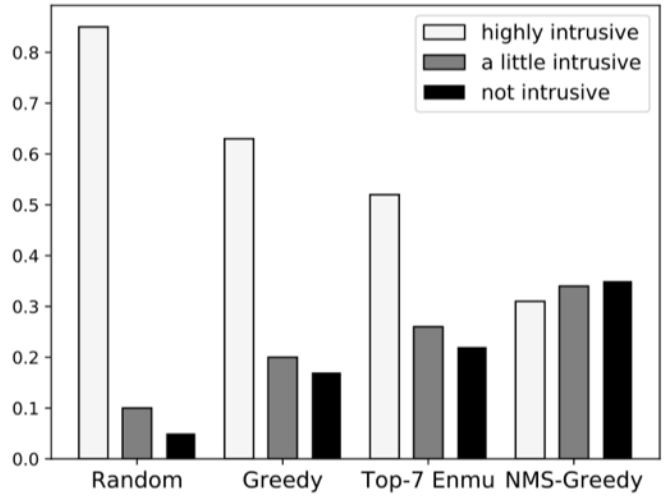
NMS-Greedy Ads Assignment

(a) Average Assignment Value V.S. Intrusiveness Budget

Highest Value

(b) Average Time Cost

Low Cost



(c) Subject Intrusiveness Evaluation

User Friendly

Outline

- Introduction
- Multimodal Content Embedding
- Activity-Level Video Advertising
- Evaluation
- **Conclusion**

Take-Home Message

Based on **atom features** and **activity-graph representation**, we can implement an efficient and scalable activity-level video advertising system.

Lab for Intelligent Networking & Knowledge Engineering

12 Faculty Members, 2 Post-Docs, 3 Secretaries; 7 with PhD from abroad

XiangYang Li

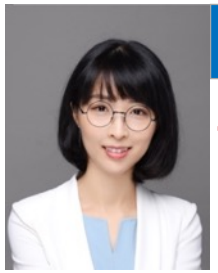
IEEE Fellow
ACM Fellow
ACM China Co-Chair

Panlong Yang

CCF Dist Speaker
Wireless Network
Mobile Computing

Nikolaos M.Freris

USA NYU A.P.
CPS, Algorithms,
Distributed optimization
Machine learning



Lan Zhang

CCF, ACM China Doctor Thesis Award, Youqing Qingcheng Award
Data Understanding/Trading
Privacy Protection

Bei Hua

High-Performance Computing
Edge Computing

Yu Zhang

System Software, Software Optimization/Security,
Quantum software

Hao Zhou

Japan NTII
Wireless Network Resource Management

Yanyong Zhang

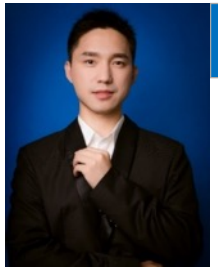
IEEE Fellow
Prof. in Rutgers
NSF Career

Haisheng Tan

HK, Tsinghua Post-Doc
Cloud Computing
Algorithms Analysis

YuBo Yan

Wireless/Passive Network, IntelliSense, IoT, SDR



Xin He

Doc. University of Oulu
Passive Network
Theories of Information and Coding

Xin Guo

Edge Computing
Security of IoT

Xuerong Huang

Master in HKBU
Research Assistant

Ludi Xue

Research Assistant

High-Quality Activity-Level Video Advertising

Mu Yuan

University of Science and Technology of China

School of Computer Science and Technology

ym0813@mail.ustc.edu.cn

Lan Zhang

University of Science and Technology of China

School of Computer Science and Technology

zhanglan@ustc.edu.cn