TRECVID 2019 INSTANCE RETRIEVAL

INTRODUCTION AND TASK OVERVIEW

Wessel Kraaij
Leiden University; Netherlands
Organisation for Applied Scientific Research (TNO)

George Awad
Georgetown University; National Institute of Standards and Technology

Keith Curtis
National Institute of Standards and Technology

Table of contents

- Task Definition
- Data
- Topics (Queries)
- Participating teams
- Evaluation & results
- General observation

Task

From 2013 – 2015

- The task asked systems to find a specific object, person or location in any context using a small set of image and video examples.

From 2016 - 2018

- A different query type was used: *find a specific person in a specific location.*

In 2019 - 2021

- A new query type is being used: *find a specific person doing a specific action.*

System task:

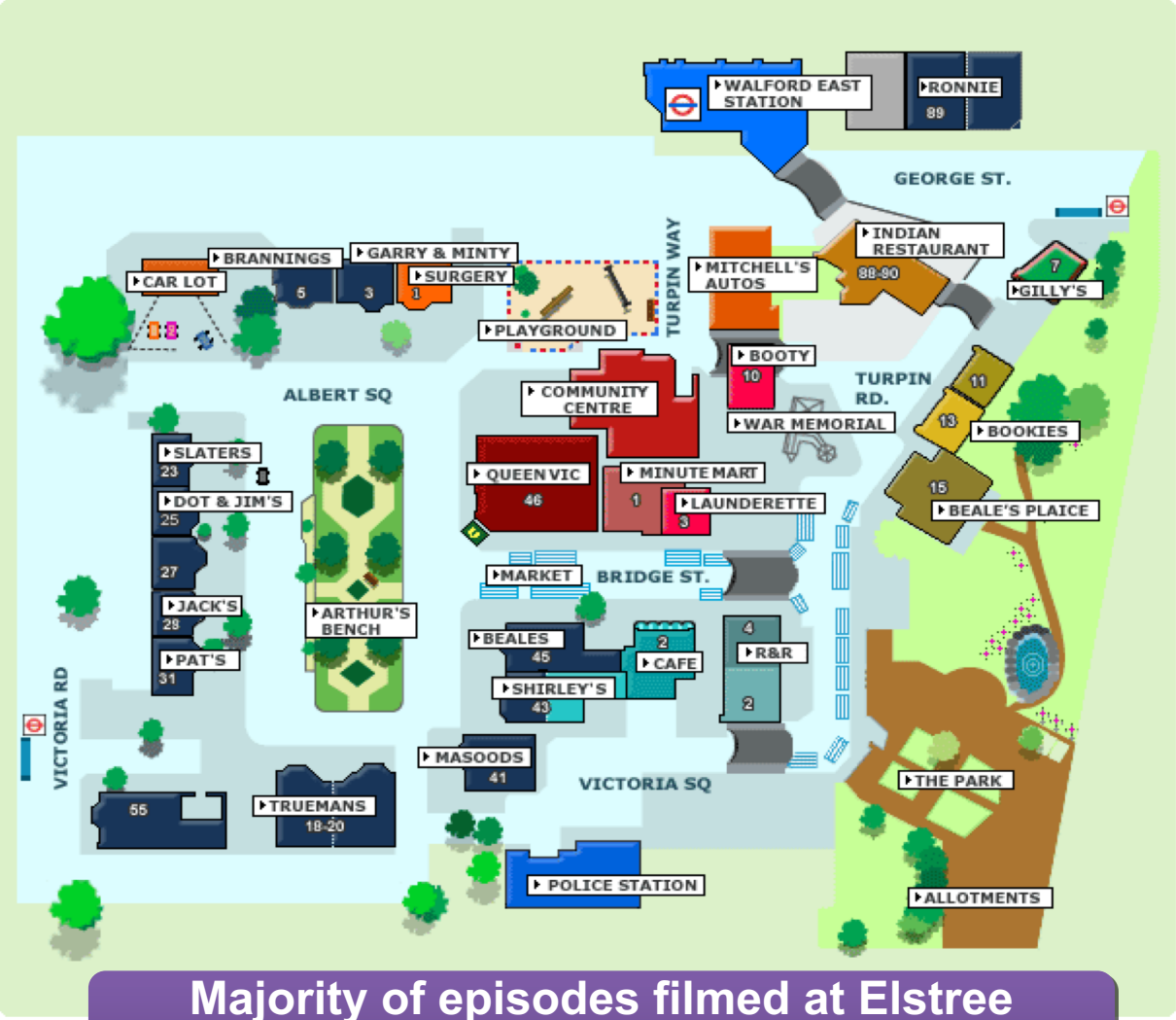
- Given a topic with:
  - 4 example images of the target person
  - 4 Region of Interest (ROI)-masked images of the target person
  - 4 to 6 video examples of a specific action
- Return a list of up to 1000 shots ranked by likelihood that they contain the target person doing the target action
- **Automatic** or **interactive** runs are accepted

Data …

- The British Broadcasting Corporation (BBC) and the Access to Audiovisual Archives (AXES) project made **464 h** of the BBC soap opera EastEnders available for research
  - 244 weekly "omnibus" files (MPEG-4) from 5 years of broadcasts
  - 471527 shots
  - Average shot length: 3.5 seconds
  - Transcripts from BBC
  - Per-file metadata

- Represents a "small world" with a slowly changing set of:
  - People (several dozen)
  - Locales: homes, workplaces, pubs, cafes, open-air market, clubs
  - Objects: clothes, cars, household goods, personal possessions, pets, etc
  - Views: various camera positions, times of year, times of day,
  - Use of fan community metadata allowed, if documented

EastEnders' world



Majority of episodes filmed at Elstree studios. Sometimes filmed on 'location'.

Topic creation procedure @ NIST

- Viewed several videos to develop a list of recurring people, actions and their overlapping.
- Listed in order the most frequent actions and most frequent person's performing them
- Created ≈90 topics targeting recurring specific persons doing specific actions.
- Chose 50 topics as a representative sample, including 30 unique topics for 2019 and 20 common topics for 2019 - 2021. Each topic includes images for target persons and example videos of the specific actions.
- Filtered example shots from the submissions if it satisfies the topic.

Global test condition: type of training data

Effect of examples – 2 conditions:

- A – one or more provided images – no video
- E - video examples (+ optional image examples)

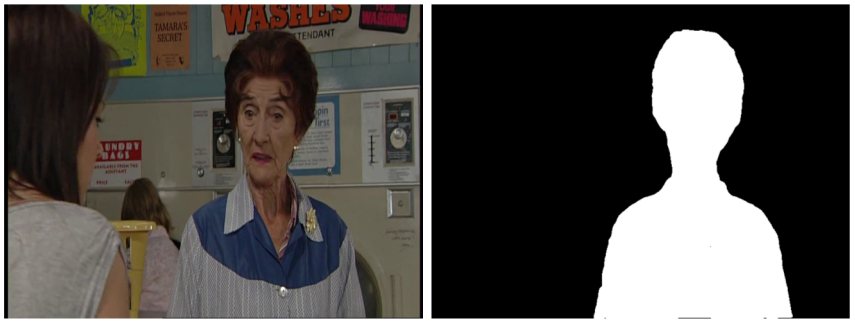
Sources of Training Data:

- A – Only sample video 0
- B - Other external data only
- C – Only provided images/videos in the official query
- D - Sample video 0 AND provided images/videos in the official query (A+C)
- E – External data AND NIST provided data (sample video 0 OR official query images/videos)

Topics – segmented "person" example images

Bradley

Denise



Dot

Heather

Topics – segmented "person" example images

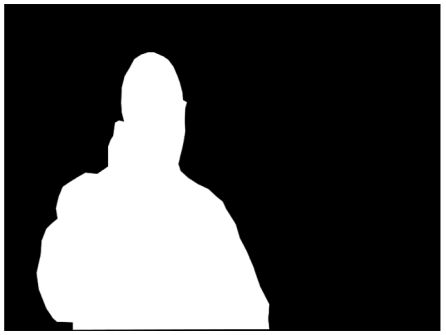

Ian


Jack


Jane


Max

Topics – segmented "person" example images

Phil


Sean


Shirley


Stacey

Sample Actions

Open door & enter

Sit on couch

Sample Actions

Eating

Hugging

30 Unique Queries – 2019

	Max	Pat	Ian	Denise	Phil	Jane	Dot	Bradley	Jack	Stacey
Holding glass	x	x	x	x						
Sit on couch		x		x					x	
Holding phone			x		x	x				
Drinking		x							x	x
Open door & enter			x				x			
Open door & leave							x		x	
Shouting	x				x				x	
Eating			x				x			
Crying	x									x
Laughing						x		x		
Go up / down stairs					x					x
Carrying bag								x		x

30 x unique queries : find {Max, Pat, Ian, Denise, Phil, Jane, Dot, Bradley, Jack, Stacey} doing {Holding glass, Sit on couch, Holding phone, Drinking, Eating, Crying, Laughing, Shouting, Open door & leave, Open door & enter, Go up / down stairs, Carrying bag}

20 Common Queries – 2019-2021

	Sean	Max	Denise	Phil	Dot	Heather	Jack	Shirley	Stacey
Kissing			x				x		
Sit on couch				x		x			
Holding phone						x	x		
Drinking				x				x	
Open door & enter	x			x					
Open door & leave		x							x
Shouting	x							x	
Hugging			x						x
Close door without leaving					x		x		
Stand & talk at door		x			x				

20 x common queries : find {Sean, Max, Denise, Phil, Dot, Heather, Jack, Shirley, Stacey} doing {Kissing, Sit on couch, Holding phone, Drinking, Shouting, Hugging, Open door & leave, Open door & enter, Close door without leaving, Stand & talk at door}

INS 2019: 6 Finishers (out of 12)

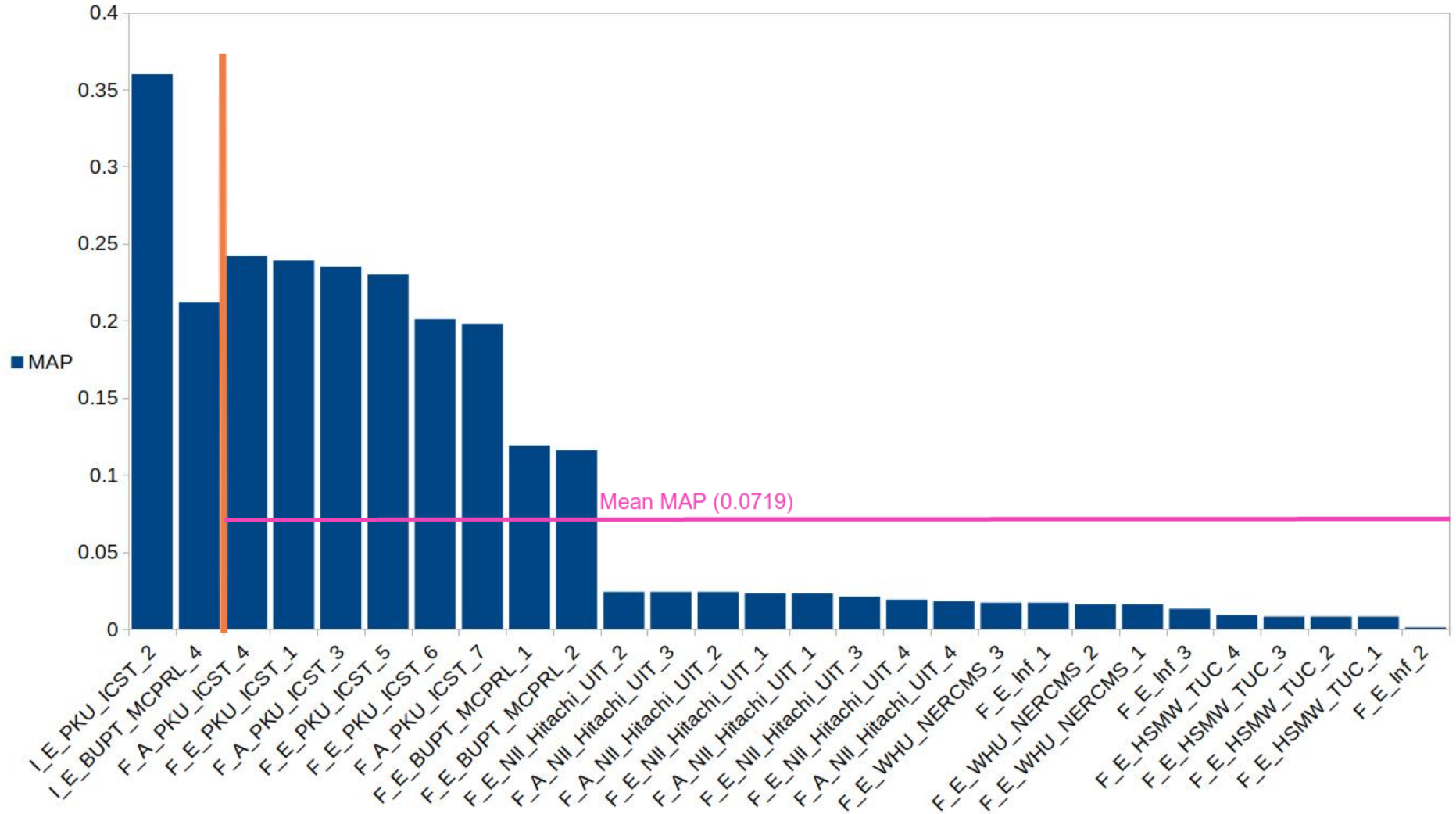
Team	Organization	Run Types Submitted F: automatic, I: Interactive
BUPT_MCPRL	Beijing University of Posts and Telecommunications	F_E (2), I_E (1)
HSMW_TUC	Chemnitz University of Technology, University of Applied Sciences Mittweida	F_E (4)
Inf	Monash University, Renmin University, Shandong University	F_E (3)
WHU_NERCMS	National Engineering Research Center for Multimedia Software, Wuhan University	F_E (3)
NII_Hitachi_UIT	National Institute of Informatics, Japan (NII); Hitachi, Ltd; University of Information Technology, VNU-HCM	F_A (4), F_E(4)
PKU_ICST	Peking University	F_A (3), F_E (3), I_E (1)

Evaluation

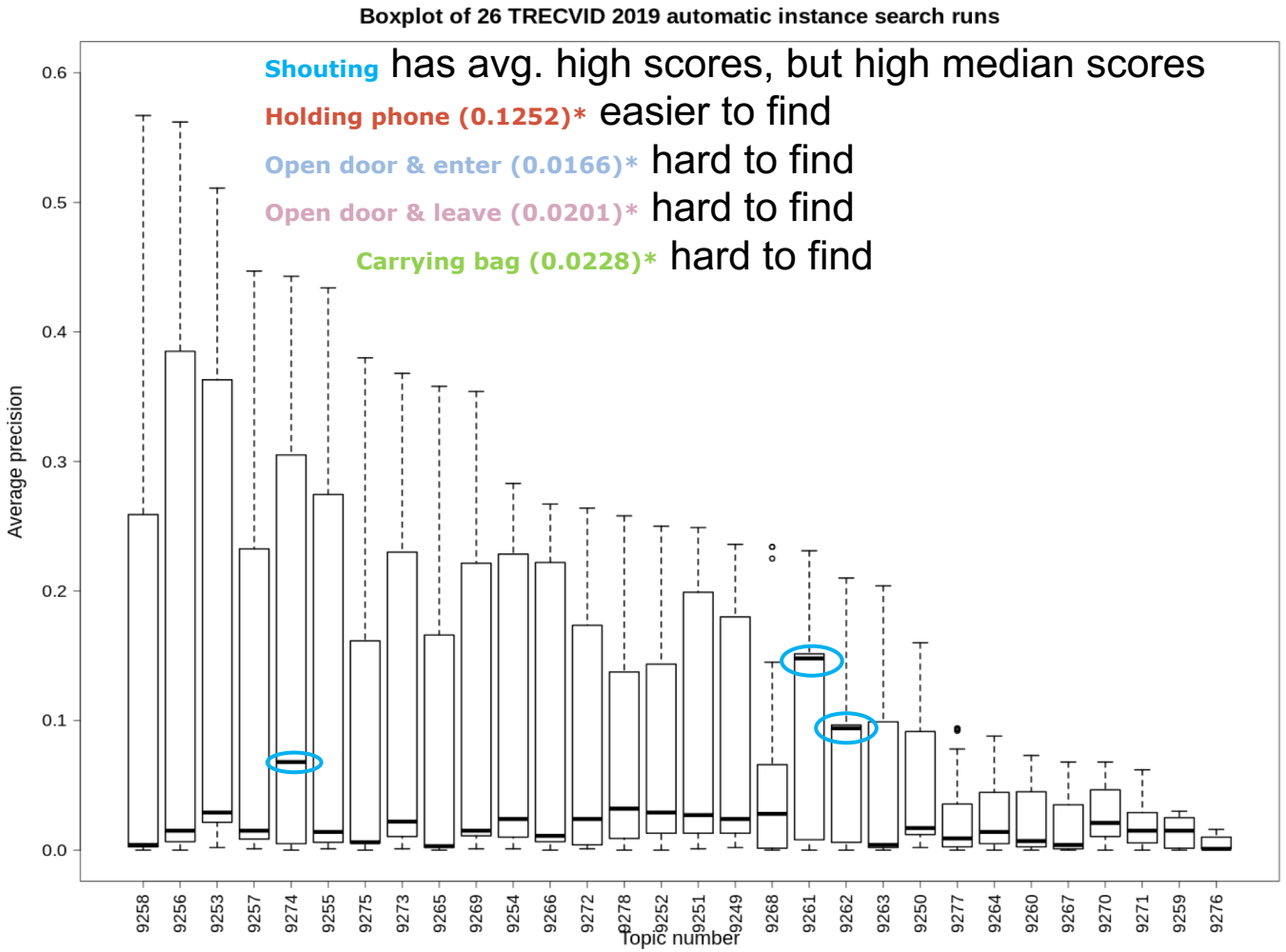
For each topic the submissions were pooled and judged down to max rank 520, resulting in 141 599 judged shots (≈ 473 person-h).

- 10 NIST assessors played the clips and determined if they contained the topic target or not.
- 6 592 clips (avg. 220 / topic) contained the topic target (4.66 %)
- True positives per topic: min 29 med 187 max 575
- The task is treated as a form of ranking and thus the trec_eval_video tool was used to calculate average precision, recall, precision, etc.
- To measure efficiency, speed was also measured.
- In total, 26 automatic and 2 interactive runs were submitted.

Results by team (Automatic)



Results by topics - automatic



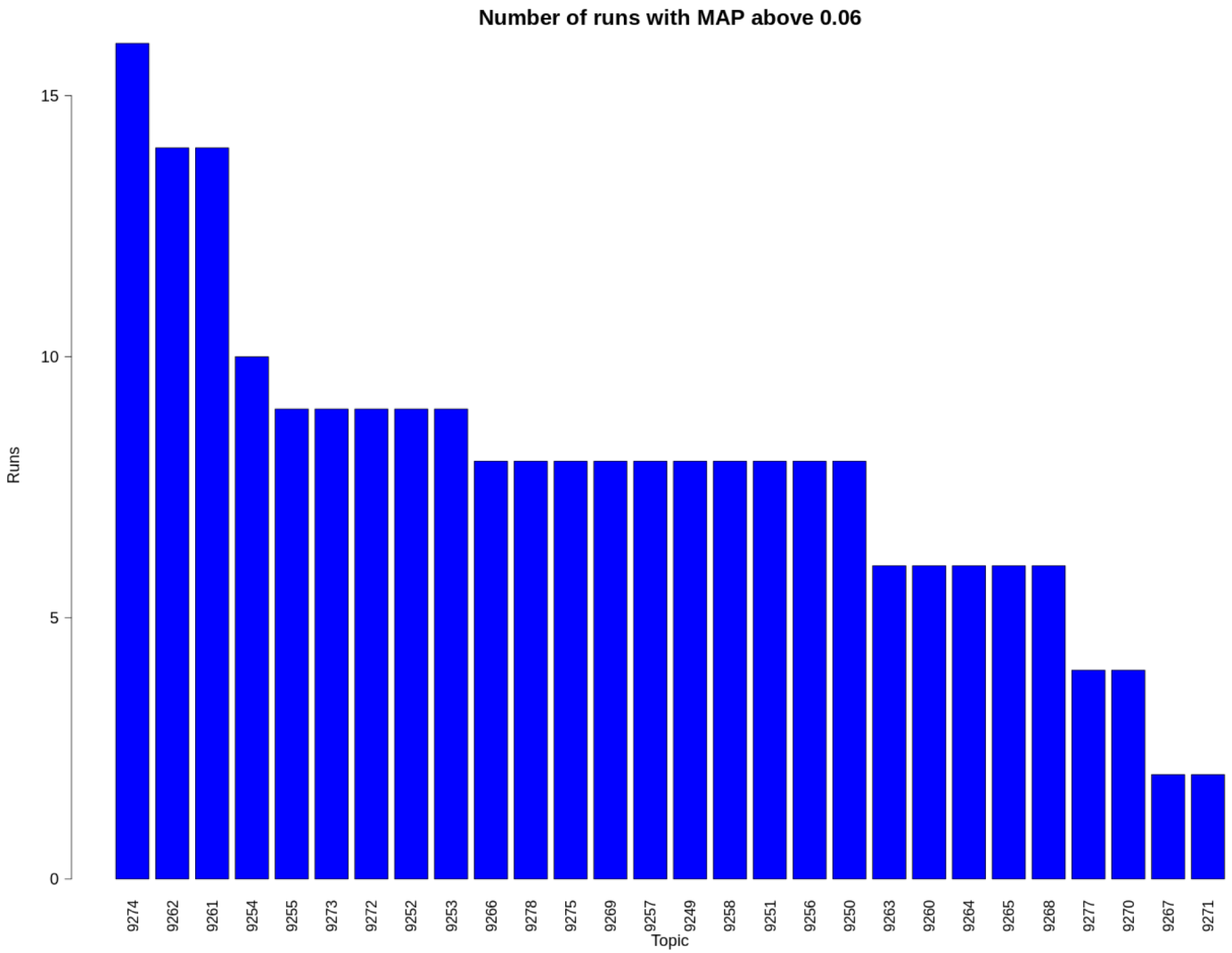
#	Query
9258	Find Pat Drinking
9256	Find Phil Holding phone
9253	Find Pat Sit on couch
9257	Find Jane Holding phone
9274	Find Jack Shouting
9255	Find Ian Holding phone
9275	Find Stacey Crying
9273	Find Jack Drinking
9265	Find Max Crying
9269	Find Jack Sit on couch
9254	Find Denise Sit on couch
9266	Find Jane Laughing
9272	Find Stacey Drinking
9278	Find Stacey Go up/down stairs
9252	Find Denise Holding Cup/Glass
9251	Find Pat Holding Cup/Glass
9249	Find Max Holding Cup/Glass
9268	Find Phil Go up/down stairs
9261	Find Max Shouting
9262	Find Phil Shouting
9263	Find Jane Eating
9250	Find Ian Holding Cup/Glass
9277	Find Jack Open door & leave
9264	Find Dot Eating
9260	Find Dot Open door & enter
9267	Find Dot Open door & leave
9270	Find Stacey Carrying bag
9271	Find Bradley Carrying bag
9259	Find Ian Open door & enter
9276	Find Bradley Laughing

*Mean score of Average Precision per character/action

Some observations..

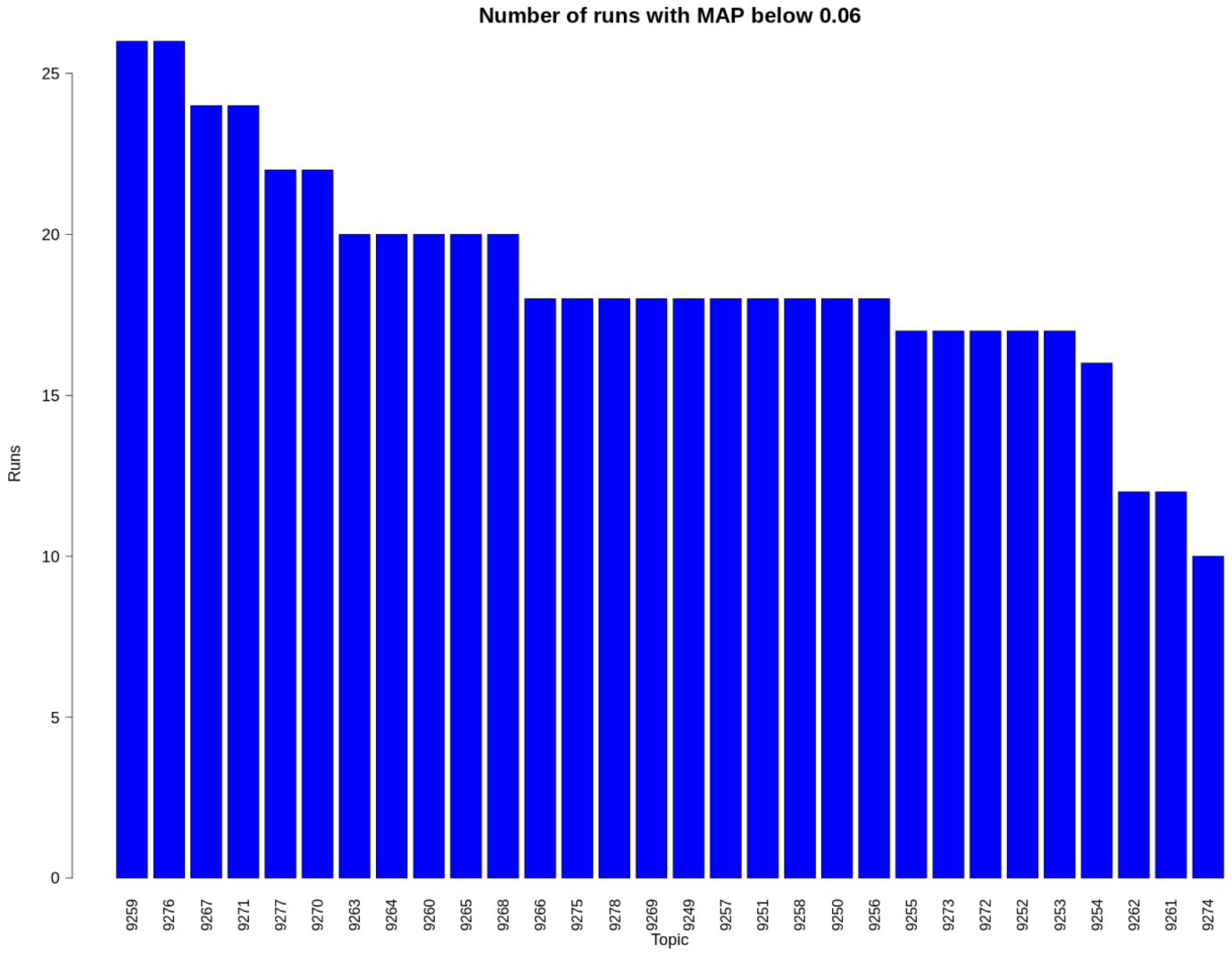
- Poor results for topics involving Dot and Bradley could indicate that they are hard people to find.
- However - previous iterations of the INS task showed them to be among the easiest people to find. What gives?
- Actions involving Dot consistently score poorly, whether it is Dot or another character involved. Seems to be more a case of hard actions to recognise.
- Bradley laughing - very poor results - but looking at frequent false positives on this topic reveal lots of instances of contrived laughter from Bradley. Obvious instances of exaggerated faked / contrived laughter do not count as laughing.

Easier Topics



#	Query
74	Find Jack Shouting
52	Find Phil Shouting
51	Find Max Shouting
54	Find Denise Sit on couch
55	Find Ian Holding phone
73	Find Jack Drinking
72	Find Stacey Drinking
52	Find Denise Holding Cup/Glass
53	Find Pat Sit on couch
56	Find Jane Laughing
78	Find Stacey Go up/down stairs
75	Find Stacey Crying
59	Find Jack Sit on couch
57	Find Jane Holding phone
49	Find Max Holding Cup/Glass
58	Find Pat Drinking
51	Find Pat Holding Cup/Glass
56	Find Phil Holding phone
50	Find Ian Holding Cup/Glass
53	Find Jane Eating
50	Find Dot Open door & enter
54	Find Dot Eating
55	Find Max Crying
58	Find Phil Go up/down stairs
77	Find Jack Open door & leave
70	Find Stacey Carrying bag
57	Find Dot Open door & leave
71	Find Bradley Carrying bag

Hard Topics



#	Query
59	Find Ian Open door & enter
76	Find Bradley Laughing
67	Find Dot Open door & leave
71	Find Bradley Carrying bag
77	Find Jack Open door & leave
70	Find Stacey Carrying bag
63	Find Jane Eating
64	Find Dot Eating
60	Find Dot Open door & enter
65	Find Max Crying
68	Find Phil Go up/down stairs
66	Find Jane Laughing
75	Find Stacey Crying
78	Find Stacey Go up/down stairs
69	Find Jack Sit on couch
49	Find Max Holding Cup/Glass
57	Find Jane Holding phone
51	Find Pat Holding Cup/Glass
58	Find Pat Drinking
50	Find Ian Holding Cup/Glass
56	Find Phil Holding phone
55	Find Ian Holding phone
73	Find Jack Drinking
72	Find Stacey Drinking
52	Find Denise Holding Cup/Glass
53	Find Pat Sit on couch
54	Find Denise Sit on couch
62	Find Phil Shouting
61	Find Max Shouting
74	Find Jack Shouting

Some observations..

- From the previous two bar charts we can safely say that shouting is the easiest topic to find. This was not obvious from the boxplot of results by topics.
- Drinking, sitting on couch, and holding phone are also among the easiest topics to find.
- Open door & leave, open door & enter, and carrying bag are among the hardest topics to find.

Some Frequent False Positives

Jack sit on couch

Jack is sitting on an armchair - a single seating structure. Topic specifies a couch - a comfortable seating structure which seats more than one person.

Bradley carrying bag

Bradley is seated next to Stacey. Dot comes into the picture carrying a bag.

Some Frequent False Positives

Dot open door & leave

Dot opens the door to let people in and then closes the door again. Does not leave the room / house.

Stacey Drinking

In this shot we see Stacey holding a glass. Later in the shot Mo is seen drinking. Topic specifies that the person must be seen moving the glass/cup to their mouth and performing a sipping or drinking action.

Some Frequent False Positives

Stacey crying

Stacey appears to have hurt herself, with blood around her left eye. She is rubbing her eye but does not appear to be crying.

Jack shouting

This appears to be a frank discussion between Jack and another man. The other person appears to be angry and raises his voice at Jack, however Jack does not raise his voice.

Some Frequent False Positives

Phil holding phone

Phil is seen singing into a microphone, along with Garry. Another person is holding a phone recording them.

Pat drinking

Pat is seen clapping hands. Later in the shot she turns her head to look over at someone. Another person moves a glass to their mouth.

Further observations from viewing most frequent false positives of worst performing topics

- Open door & enter - Systems tended to classify any shots with target person and a doorway as a positive detection. More work needed on training systems to classify the action itself.
- Open door & leave - Same as above, systems classifying any shots with target person and a doorway as positive detection. More work needed on training systems to classify the action of opening a door and leaving.
- Carrying bag - Fewer conclusions can be drawn. Many instances where a shot is classified as a positive detection if the target person appears in the shot and a different person is carrying a bag, however, many other shots contain the target person with no bag visible in the shot at all.

Automatic Run results + Randomization testing

Top 10 runs across all teams (automatic)

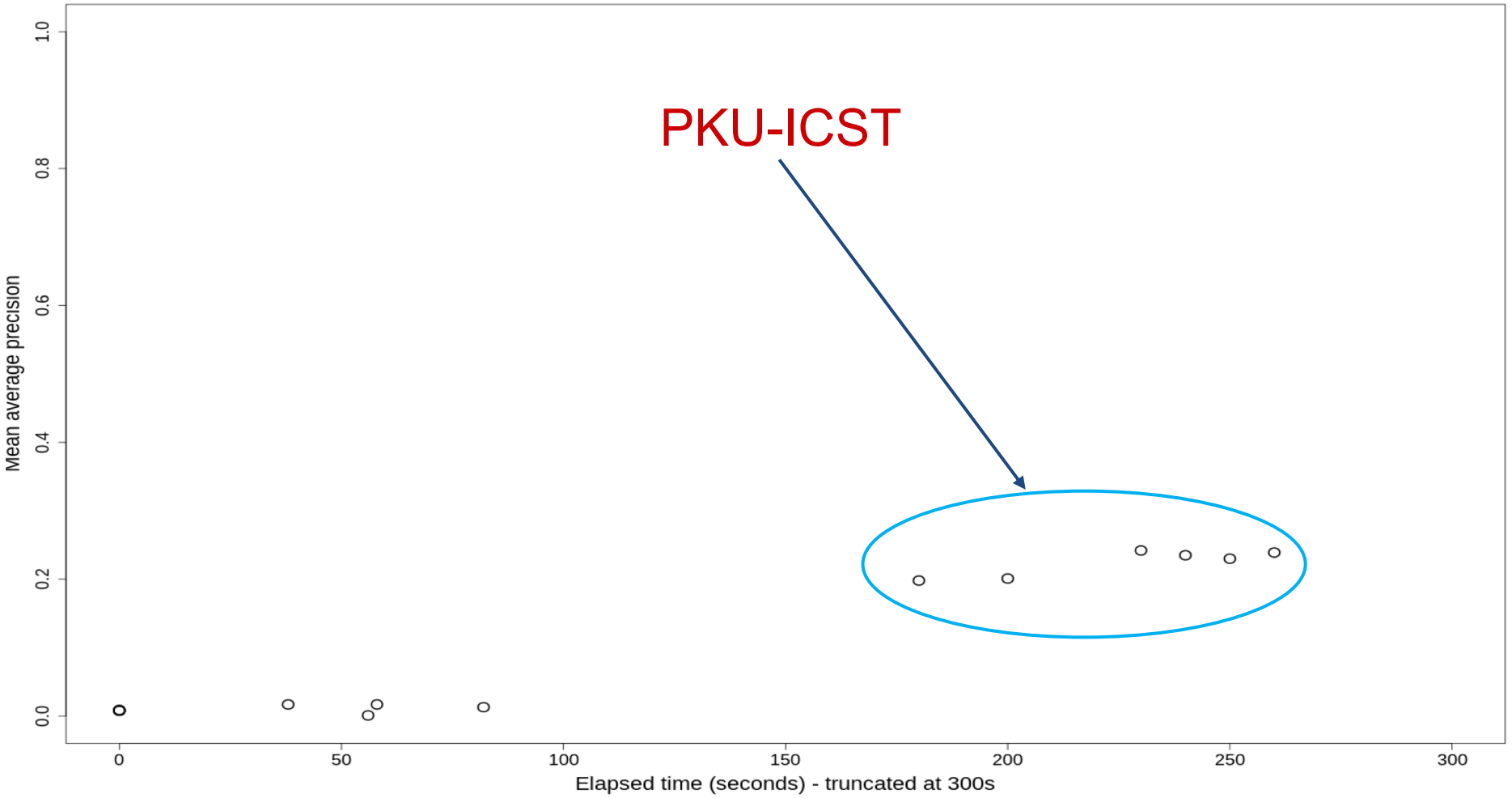
MAP										
0.242	F_A_PKU_ICST_4*^	=			>	>	>	>	>	>	>
0.239	F_E_PKU_ICST_1*↑		=	>		>	>	>	>	>	>
0.235	F_A_PKU_ICST_3^↕			=		>	>	>	>	>	>
0.230	F_E_PKU_ICST_5↑↕				=	>	>	>	>	>	>
0.201	F_E_PKU_ICST_6					=	>	>	>	>	>
0.198	F_A_PKU_ICST_7						=	>	>	>	>
0.119	F_E_BUPT_MCPRL_1							=	>	>	>
0.116	F_E_BUPT_MCPRL_2								=	>	>
0.024	F_E_NII_Hitachi_UIT_2†									=	
0.024	F_A_NII_Hitachi_UIT_3†										=
		1	2	3	4	5	6	7	8	9	10

*^↑↕† = difference not statistically significant

p = probability the row run scored better than the column run due to chance

> p < 0.05

Mean Average Precision vs. per run clock processing time (automatic)



Interactive Run results + Randomization testing

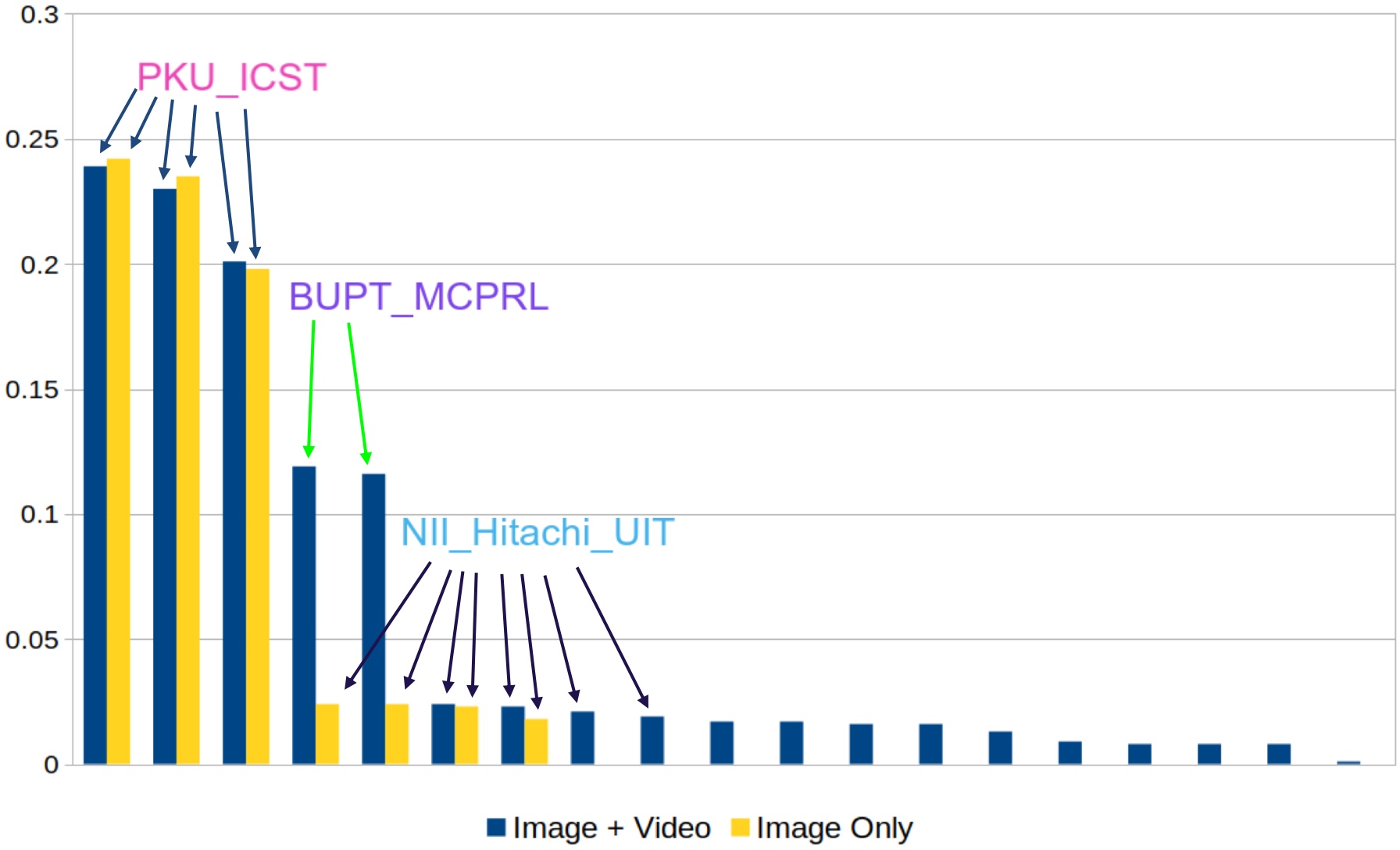
MAP **Runs across all teams (interactive)**

```
0.360 I_E_PKU_ICST_2          =    >
0.212 I_E_BUPT_MCPRL_4             =
                              1    2
```

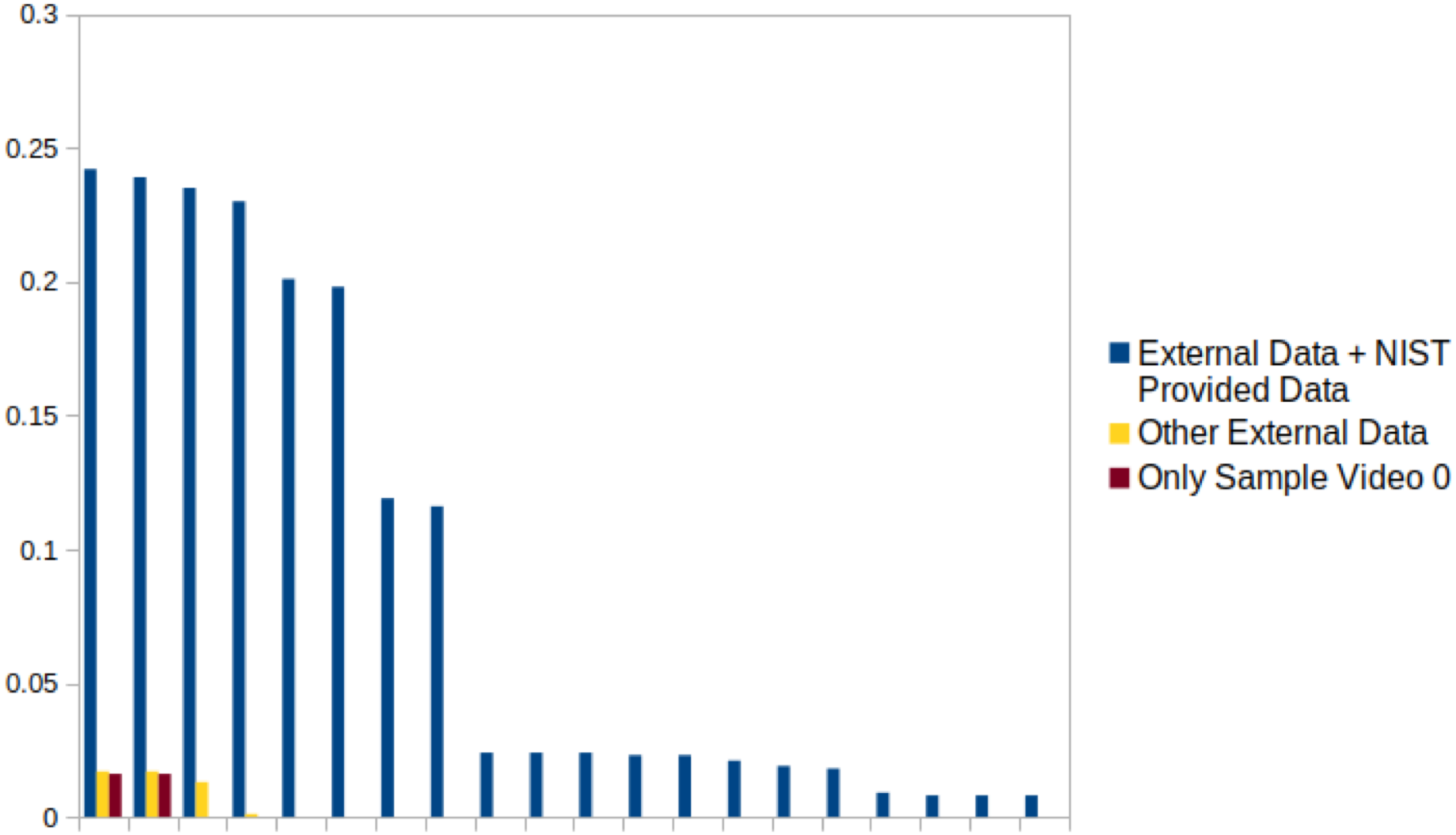
p = probability the row run scored better than the column run due to chance

```
>   p < 0.05
```

Results by example set (A/E) - automatic



Results by Data Source



Some general observations about the task

- Slight decrease in number of participants and finishers, but higher % of participants finished the task.
- Many more teams now using E condition - training with video examples. Perhaps more necessary now with action recognition. But - Results from teams using both show little difference between image & video and image only!
- Interactive search task:
  - Limited participation - only two interactive runs this year.
- First year of updated task - results cannot be compared in any way to previous years - in subsequent years we can compare using the common topics.

Some general observations about the task – Data Source

- Best results by far achieved using external data plus NIST provided data.
- Huge gap to results from systems trained using only external data or using only sample video 0.

Further Conclusions

- Person recognition has been a feature of the INS task since 2013 and is very mature by this stage. Very few frequent false positives misidentify the person.
- Action recognition is a new feature of INS task. The much increased difficulty of the new INS task is due to this. Requires much more work to reach an acceptable level of maturity.

Further Conclusions

- Visual Concepts very important.
- Easier tasks mostly those with obvious visual context (sit on couch, hold phone, hold glass, etc.)
- Harder tasks tend to be more independent from obvious visual context (crying, laughing, eating, different actions involving a doorway hard to isolate from others).

Overview of submissions (1)

- 6 out of 6 teams described INS runs for the TV notebook
- 2 teams will present their INS experiments

2:15 - 2:45, (**BUPT_MCPRL Team– Beijing University of Posts and Telecommunications**)

2:45 - 3:15, (**HSMW_TUC Team– University of Applied Sciences Mittweida**)

3:15 - 3:35, INS Discussion

INS 2019 Discussion

- What do teams think of the new task (query type)?
- Are the selected actions important in real life applications?

- What is the main challenge in the new query type?
  - No enough training data?
  - actions are difficult?
  - fusion of persons + action detection results?

- Is the task still of an ad-hoc nature? Or converting to a supervised learning?

- Do we need additional run categories?

2019 to 2021 Progress Runs

- 20 common topics.
- Evaluate progress of participating teams 2019-2021 using a set of common topics.
- 12 runs submitted by 3 separate teams in 2019. Additional teams can still submit progress runs in 2020 on the 10 topics to be evaluated in 2021.
- 10 common topics will be evaluated in 2020.
- 10 remaining common topics evaluated in 2021.