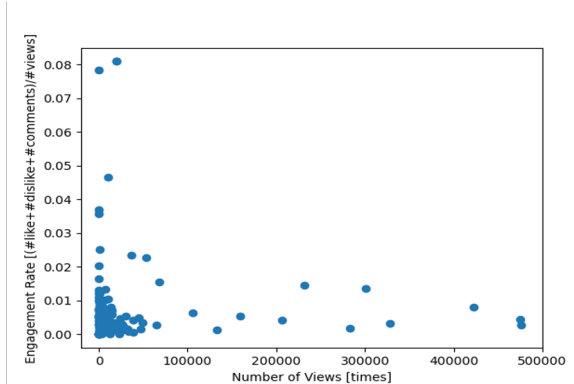


Online Video Engagement Detection

Why Engagement?

- Motivation:
Online video engagement underexplored yet promising (misinformation, Accessibility, UX design...).
- Dataset:
 - based on the YouTube prostate cancer dataset of 250 videos (Hou, Rui, et al. 2019).
 - Engagement Labeling:
 - Average view percentage, Click-through-rate, Click stream...
 - **(#Like + #Dislike + #Comment) / #views**
- Experiments: Formulation

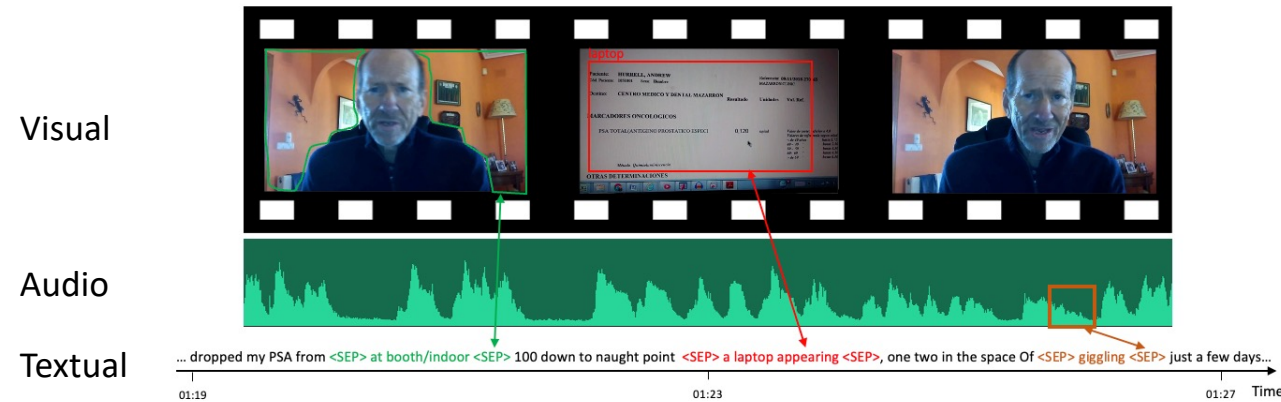
Category	Model	Binary [0,10,+inf]				Multiclass (3 way) [0,10,20,+inf]				Multiclass (4 way) [0,10,20,30,+inf]			
		Accuray	Precision	Recall	F1-score	Accuray	Precision	Recall	F1-score	Accuray	Precision	Recall	F1-score
Baseline	Random	0.5	0.5	0.5	0.5	0.33	0.33	0.33	0.33	0.25	0.25	0.25	0.25
	Majority	0.55	0.275	0.5	0.355	0.55	0.183	0.333	0.236	0.55	0.137	0.25	0.177
Text	Roberta-base	0.624	0.604	0.618	0.592	0.506	0.398	0.422	0.376	0.484	0.300	0.310	0.2680
	Bag of words	0.536	0.552	0.566	0.520	0.478	0.276	0.350	0.276	0.478	0.204	0.262	0.206
Audio	Wav2Vec	0.534	0.534	0.536	0.5	0.5175	0.23	0.28	0.2425	0.414	0.346	0.358	0.328
	Wav2Vec double length	0.566	0.566	0.558	0.536	0.446	0.294	0.298	0.274	0.390	0.324	0.310	0.288
Text+Audio	Roberta-Wav2Vec Concatenation	0.666	0.638	0.644	0.626	0.466	0.248	0.258	0.242	0.470	0.354	0.380	0.336



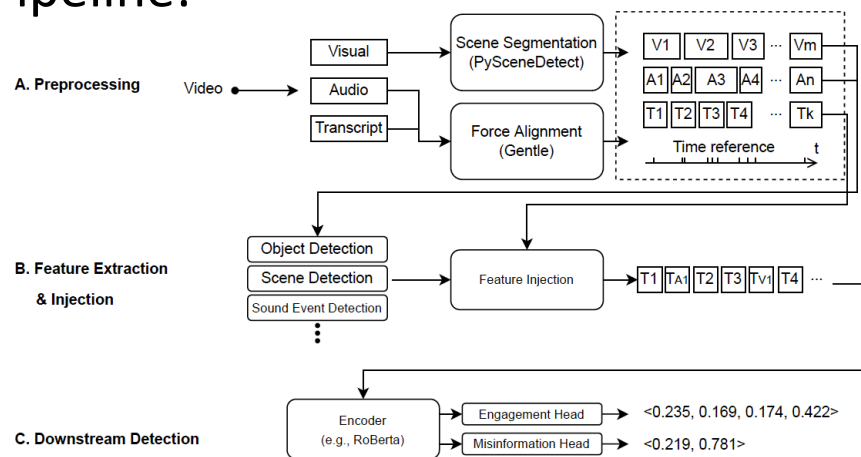
- Finding 1: In-distribution generalization is not significant (Inconsistency labeling?)
- Finding 2: Learning was not distinguishable for audio modality (>>memory)

How to learn engagement?

“Very Early Fusion”: Inject features (visual & audio) to single modality (text)



• Pipeline:



• Experiments

Category	Model	Binary [0,10,+inf]				Multiclass (4 way) [0,10,20,30,+inf]			
		Accuray	Precision	Recall	F1-score	Accuray	Precision	Recall	F1-score
Text	Roberta-base (text)	0.624	0.604	0.618	0.592	0.484	0.300	0.310	0.2680
Text+Audio	Roberta-Wav2Vec concatenation	0.666	0.638	0.644	0.626	0.470	0.354	0.380	0.336
Text+Video	VEF (scene)	0.674	0.624	0.644	0.616	0.548	0.310	0.330	0.300
	VEF (object)	0.668	0.630	0.634	0.624	0.536	0.288	0.338	0.300
	VEF (combined)	0.680	0.652	0.654	0.628	0.514	0.306	0.320	0.300

Engagement vs. Misinformation

- Can engagement be applied to assist misinformation detection?
 - “People may try to grab audience’s eye to better sell their lies.”
- Correlation Analysis:

```
print("Corr of misinformation w/:")
print("\tengagement_rate {:.3f}".format(pearsonr(new_df["Misinformation"],new_df["engagement_rate"]).statistic))
print("\tlike_count {:.3f}".format(pearsonr(new_df["Misinformation"],new_df["like_count"]).statistic))
print("\tdislike_count {:.3f}".format(pearsonr(new_df["Misinformation"],new_df["dislike_count"]).statistic))
print("\tcomment_count {:.3f}".format(pearsonr(new_df["Misinformation"],new_df["comment_count"]).statistic))
print("\tview_count {:.3f}".format(pearsonr(new_df["Misinformation"],new_df["view_count"]).statistic))
```

✓ 0.4s

```
Corr of misinformation w/:
engagement_rate 0.097
like_count 0.299
dislike_count 0.180
comment_count 0.270
view_count 0.170
```

→ Representations of “engagement” matter (context-dependent)!

- Hypothesis: more appropriate at a smaller scale (how engagement rate varies locally)?...