The Hateful Memes Dataset

恶意模因数据集
Python数据分析
展示

Niklas Muennighoff，孟念，1800092850

FACEBOOK AI

内容

1. 背景
2. 问题和数据介绍
3. 方案
   a. 数据准备
   b. 模型
4. 结果

背景

- Facebook AI 提出的奖金为100K USD 的比赛
- >3400人参加

Hateful Memes: Phase 2

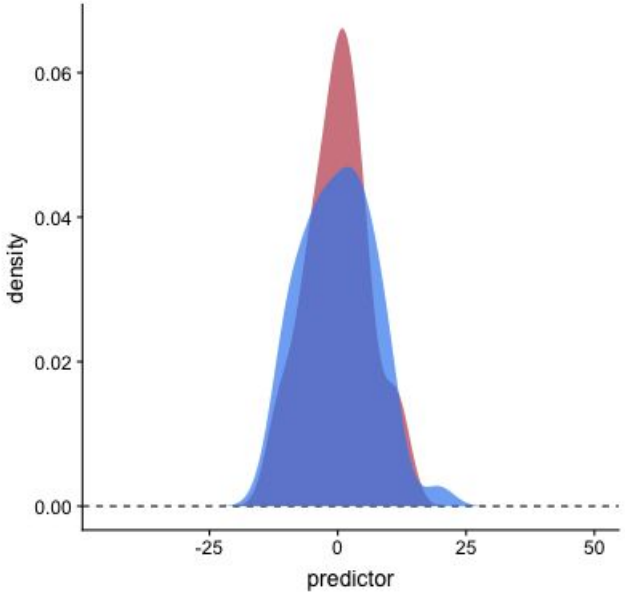
HOSTED BY FACEBOOK

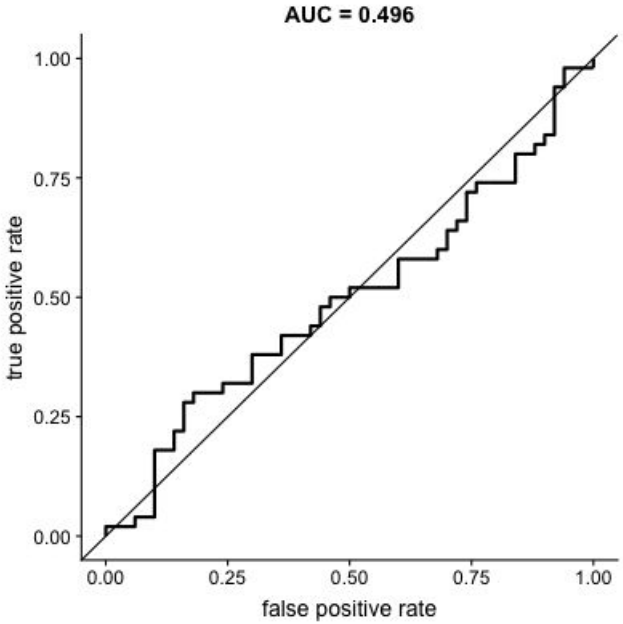
	User or team		Best private AUROC	Accuracy	Timestamp	Trend (last 10)	# Entries
	alfred lab	1	0.8450	0.7320	2020-10-31 19:15:09		3
	Muennighoff	2	0.8310	0.6950	2020-10-31 23:34:40		1
	HateDetectron	3	0.8108	0.7650	2020-10-16 23:02:31		1
	kingsterdam	4	0.8053	0.7385	2020-10-31 23:20:27		3
	burebista	5	0.7943	0.7430	2020-10-30 09:38:08		3
	naoki	6	0.7886	0.7305	2020-10-31 04:43:28		3
	MemeLords	7	0.7884	0.7450	2020-10-31 23:39:13		3
	AiTingting	8	0.7848	0.7295	2020-10-31 12:56:43		3
	mobot	9	0.7832	0.7320	2020-10-28 02:46:48		3
	james005	10	0.7814	0.7280	2020-10-31 20:28:47		3
	hate-alert	11	0.7808	0.7270	2020-10-26 13:13:22		3
	mrsio	12	0.7806	0.7430	2020-10-20 16:30:18		3
	letsgo	13	0.7801	0.7285	2020-10-28 12:51:03		3
	QMUL-NUAA	14	0.7784	0.7300	2020-10-28 05:46:55		3
	xyxyxxxy	15	0.7780	0.7270	2020-10-28 05:17:36		3

问题和数据介绍 - 什么是恶意模因和混杂模因？

恶意　　不恶意　　不恶意

问题和数据介绍 - 评分





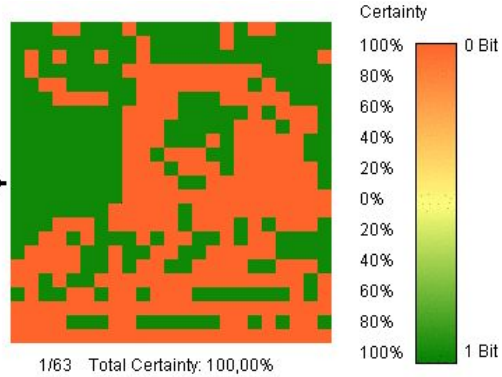
User or team		Best private AUROC	Accuracy	Timestamp
alfred lab	1	0.8450	0.7320	2020-10-31 19:15:09
Muennighoff	2	0.8310	0.6950	2020-10-31 23:34:40
HateDetectron	3	0.8108	0.7650	2020-10-16 23:02:31
kingsterdam	4	0.8053	0.7385	2020-10-31 23:20:27
burebista	5	0.7943	0.7430	2020-10-30 09:38:08
naoki	6	0.7886	0.7305	2020-10-31 04:43:28
MemeLords	7	0.7884	0.7450	2020-10-31 23:39:13
AiTingting	8	0.7848	0.7295	2020-10-31 12:56:43
mobot	9	0.7832	0.7320	2020-10-28 02:46:48
james005	10	0.7814	0.7280	2020-10-31 20:28:47
hate-alert	11	0.7808	0.7270	2020-10-26 13:13:22
mrsio	12	0.7806	0.7430	2020-10-20 16:30:18
letsgo	13	0.7801	0.7285	2020-10-28 12:51:03
QMUL-NUAA	14	0.7784	0.7300	2020-10-28 05:46:55
xyxyxxxy	15	0.7780	0.7270	2020-10-28 05:17:36

方案 - 数据清洗

对比文本：Levenshtein 距离

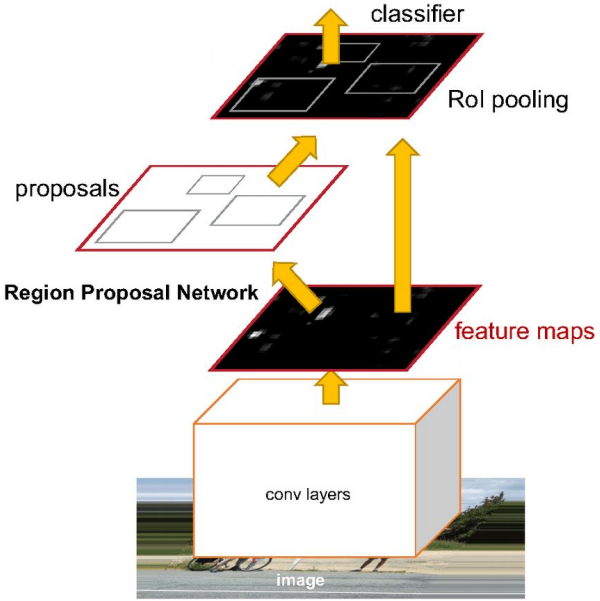
$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

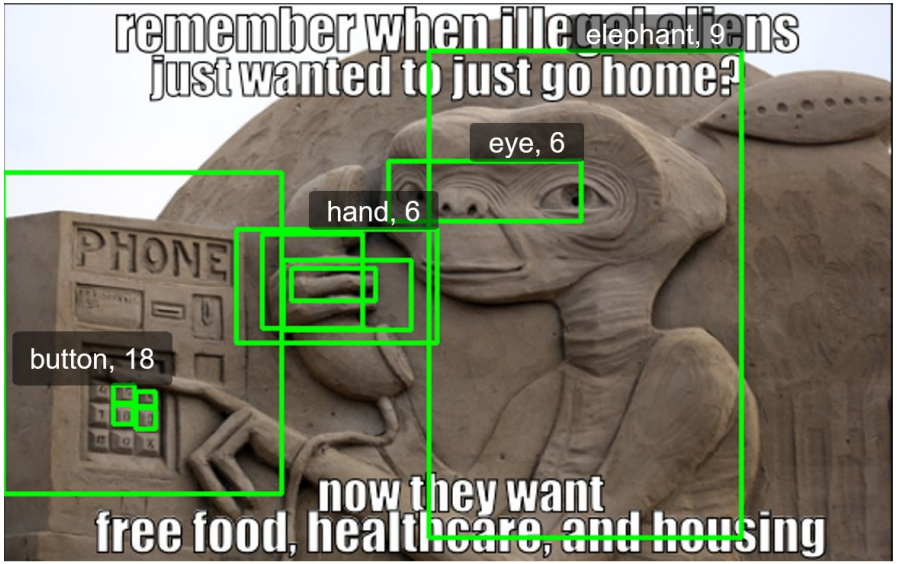
对比图片：感知哈希函数



方案 - 数据准备

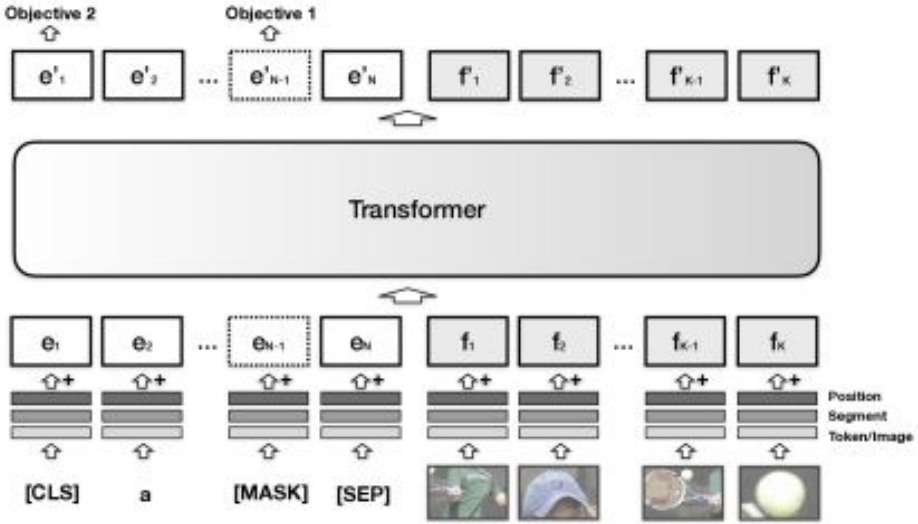
用detectron2的FasterRCNN提取RoI (Regions Of Interest)





方案 - 模型 - VisualBERT

A person hits a ball with a tennis racket



文本 RoI

Liunian Harold Li et al. (北大校友)

结果

来源	模型	验证集 AUC	测试集 AUC
Facebook AI 提供的 Hateful Memes 基线	人类	-	82.65
	ViLBERT	71.13	70.45
	VisualBERT	70.60	71.33
	ViLBERT CC	70.07	70.03
	VisualBERT COCO	73.97	71.41
我的方案	VisualBERT	75.49	75.75
	OSCAR	77.16	77.30
	UNITER	77.75	78.65
	ERNIE-ViL Base	78.18	77.02
	ERNIE-ViL Large	78.76	80.59
	Ensemble	81.56	82.52


THANK YOU
GOOD SIR