

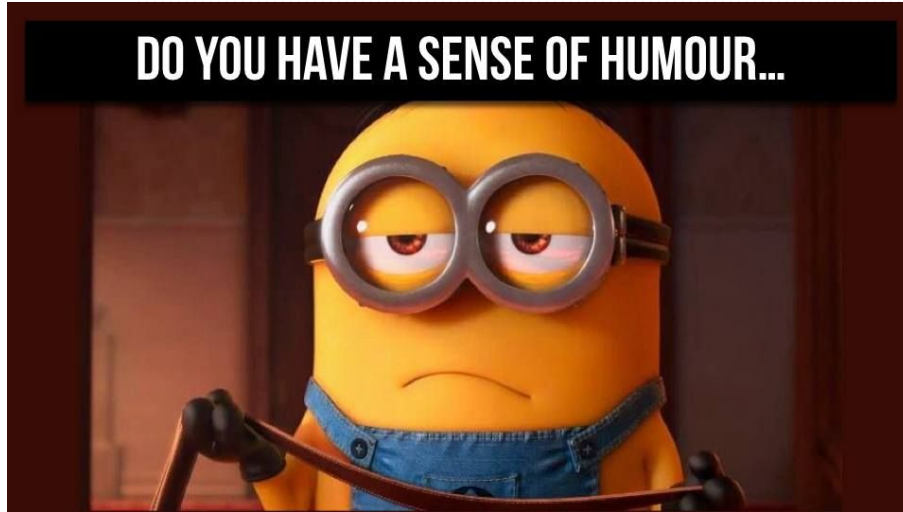
Laugh or Lout (LOL)

Humor and Offense Detection from Sentences

Ming-Hsien Chien, Liuyi Jin, Tian Liu, Yijun Zhang

Apr 26, 2021

Background and Motivation



Data and Tasks

SemEval-2021

The 15th International Workshop on Semantic Evaluation

The Task 7 of SemEval-2021, which was held on the CodaLab

- Task 1a: is_humor (binary classification 0-1)
- Task 1b: humor_rating (regression between 0 to 5)
- Task 1c: humor_controversy (binary classification 0-1)
- Task 2: offense_rating (regression between 0 to 5)
- 8,000 train data
- 1,000 dev
- 1,000 test

id	text	is_humor	humor_rating	humor_controversy	offense_rating
8001	What's the difference	1	2.45	0	1.7
8002	Vodka, whisky, tequil	1	2	0	0
8003	French people don't	1	2.95	0	1.15
8004	A lot of Suicide bom	1	1.38	0	3.75

<https://competitions.codalab.org/competitions/27446>

- This is the first task to combine humor and offensive detection

Challenges in Humor Detection

- Humor is subjective, depending on one's age, education, culture, experiences, etc.
- Various types of humor: wordplay, irony, sarcasm, sexual, etc.
- Some types of humor requires extra knowledge

*"The one who invented the door knocker got a **No Bell prize**"*

*"Your village called. They want their **Idiot** back."*

Did you hear about the guy whose whole left side was cut off? **He's all right now.**

Related Work

- Yang et al. (2015): handcrafted features
 - 4 structures: meaning incongruity, ambiguity, phonetic style, and personal effect
 - Random Forest
- Bertero et al. (2016): Long-Term-Short-Memory (LSTM) model to capture the latent sentence structures in humorous texts
- Chen et al. (2018): Convolutional Neural Network (CNN) with augmented filter sizes and numbers to capture the humor features from texts
- Annamoradnejad (2021): BERT embedding

Our Approach - BERT Sentence Embeddings

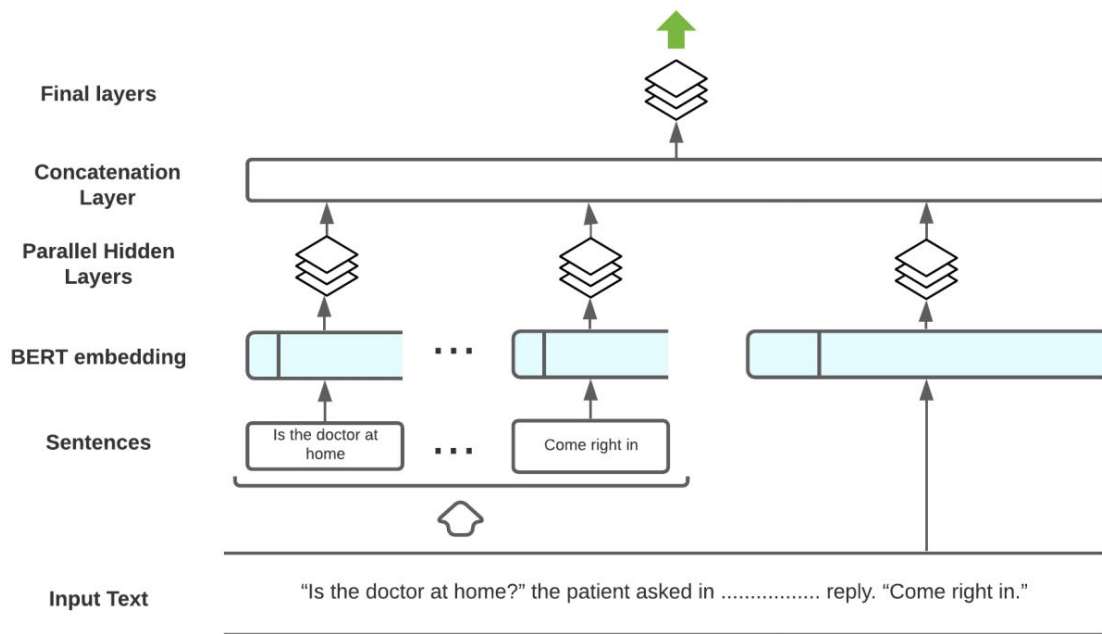
2 types of humor structure to capture:

- **relationships between sentences** (the presence of punchline)
- **word-level connections** (synonyms and antonyms)

Did you hear about the guy whose whole left side was cut off? *He's all right now.*

*"The one who invented the *door knocker* got a *No Bell prize*"*

CoBERT



(Annamoradnejad, 2020)

- **BERT-Base-Uncased:**
 - 12-layer, 768-hidden, 12-heads, 110M params
 - Trained on BooksCorpus w/ 800M words
- **BERT-Large-Uncased:**
 - 24-layer, 1024-hidden, 16-heads, 340M params
 - Trained on English Wikipedia w/ 2,500M words
- **ALBERT:**
 - 24-layer, 1024-hidden, 128 embedding, 18M params, params sharing
- **Baselines: Naive Bayes, Perceptron w/ Bag-of-Words representation**

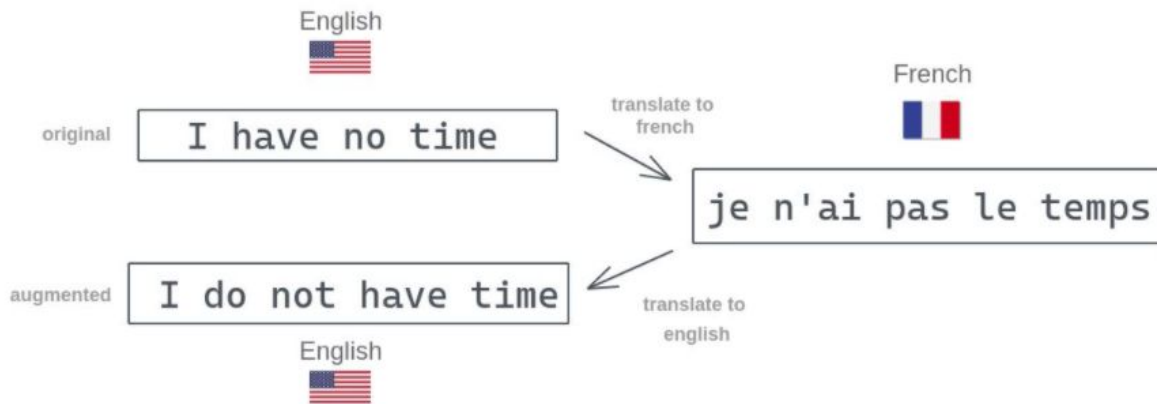
Problems Encountered and Solutions

- **Slow converging:**

- Manually tuned several hyperparameters, including batchsize, training epochs, learning rate

- **Overfitting:**

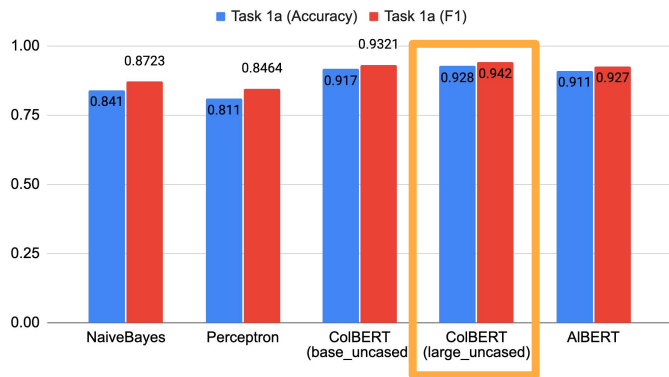
- For BERT-LARGE case, applied text data augmentation using **Back Translation**



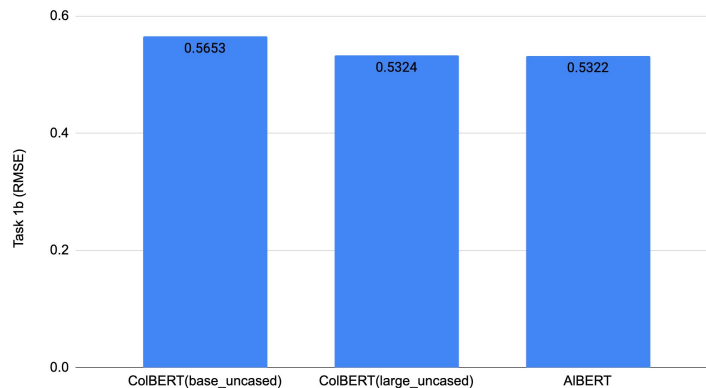
Results and Discussions

Test Data

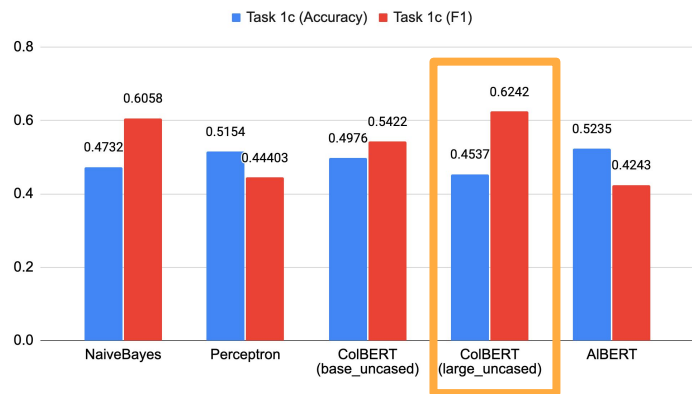
Task 1a (Accuracy and F1)



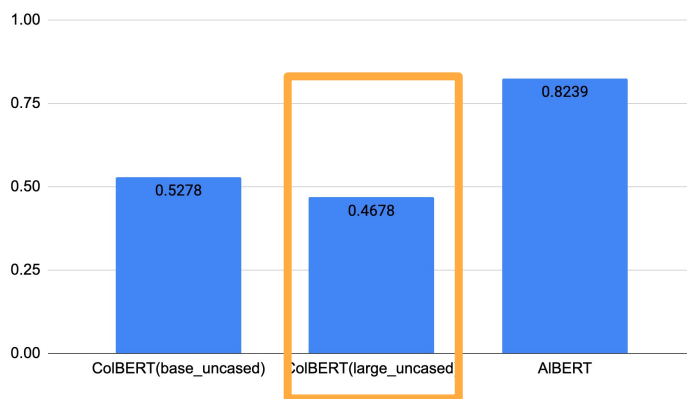
Task 1b (RMSE)



Task 1c (Accuracy and F1)

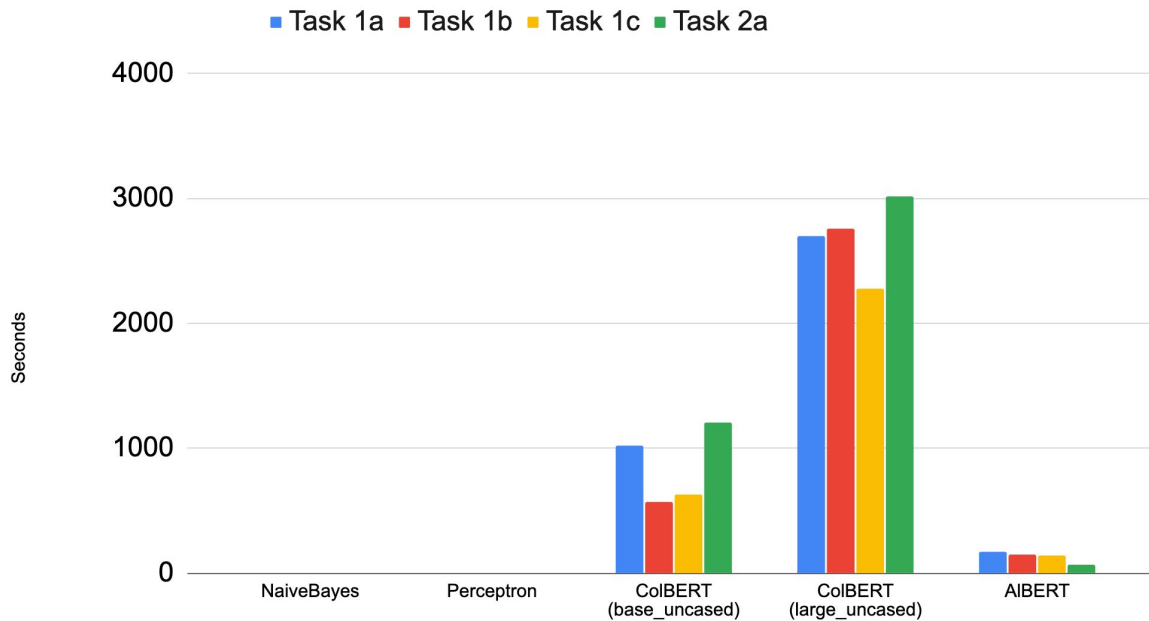


Task 2a (RMSE)



Results and Discussions

Training Time for 8000 training texts



Top Rankings (Total 43 teams globally)

Task 1a Humor Detection

#	User	Entries	Date of Last Entry	Team Name	F-Score ▲	Accuracy ▲
1	DeepBlueAI	6	02/01/21	DeepBlueAI	0.9676 (1)	0.9600 (1)
2	dalya	32	02/02/21	SarcasmDet	0.9675 (2)	0.9600 (1)
3	ThisIsTheEnd	14	02/03/21	EndTimes	0.9655 (3)	0.9570 (2)
24	calamitylink	4	02/01/21	HumorHunter	0.9498 (23)	0.9380 (16)
25	mayukh	10	02/01/21	YoungSheldon	0.9468 (24)	0.9330 (17)
26	csecudsg	46	03/26/21	CSECU-DSG	0.9462 (25)	0.9330 (17)
27	tmarchitan	13	02/07/21	Unibuc_NLP	0.9434 (26)	0.9290 (18)
28	TLIU	7	04/16/21		0.9421 (27)	0.9280 (19)
29	MihaiSamson	8	02/01/21		0.9374 (28)	0.9220 (20)
30	Zehao_Liu	5	02/19/21	UoR	0.9366 (29)	0.9210 (21)
31	zhaoyingjia123	2	02/01/21	ZYJ	0.9348 (30)	0.9210 (21)

Task 1c Humor Controversy

#	User	Entries	Date of Last Entry	Team Name	F-Score ▲	Accuracy ▲
1	CHAOYUDENG	110	04/08/21		0.6299 (1)	0.5089 (13)
2	dalya	32	02/02/21	SarcasmDet	0.6270 (2)	0.4699 (20)
3	ThisIsTheEnd	14	02/03/21	EndTimes	0.6249 (3)	0.4553 (24)
4	mmmm	6	02/01/21		0.6247 (4)	0.4569 (23)
5	TLIU	7	04/16/21		0.6242 (5)	0.4537 (25)
6	abdelkader	22	02/01/21	CS-UM6P	0.6242 (5)	0.4537 (25)
7	fdabek	1	02/01/21	Amobee	0.6233 (6)	0.4537 (25)
8	jam	2	03/29/21	smash	0.6233 (7)	0.4732 (19)

Task 1b Average Humor Score

#	User	Entries	Date of Last Entry	Team Name	RMSE ▲
1	mmmm	6	02/01/21		0.5003 (1)
2	mayukh	10	02/01/21	YoungSheldon	0.5257 (2)
3	fdabek	1	02/01/21	Amobee	0.5271 (3)
4	TLIU	7	04/16/21		0.5324 (4)
5	Amherst685	9	02/01/21		0.5339 (5)
6	tathagataraha	22	02/01/21		0.5441 (6)
7	dalya	32	02/02/21	SarcasmDet	0.5446 (7)
8	bholeswar	12	03/31/21		0.5481 (8)
9	calamitylink	4	02/01/21	HumorHunter	0.5510 (9)
10	abdelkader	22	02/01/21	CS-UM6P	0.5519 (10)

Task 2 Average Offensiveness Score

#	User	Entries	Date of Last Entry	Team Name	RMSE ▲
1	DeepBlueAI	6	02/01/21	DeepBlueAI	0.4120 (1)
2	mmmm	6	02/01/21		0.4197 (2)
3	fdabek	1	02/01/21	Amobee	0.4406 (3)
4	stevenhuahua	2	02/21/21		0.4454 (4)
5	emran	1	02/09/21	ES-JUST	0.4467 (5)
6	dalya	32	02/02/21	SarcasmDet	0.4469 (6)
7	reynier	21	02/01/21	RoMa	0.4532 (7)
8	calamitylink	4	02/01/21	HumorHunter	0.4546 (8)
9	Amherst685	9	02/01/21		0.4564 (9)
10	TLIU	7	04/16/21		0.4678 (10)
11	Sakrah	9	03/11/21	DuluthNLP	0.4690 (11)
12	ThisIsTheEnd	14	02/03/21	EndTimes	0.4691 (12)

<https://competitions.codalab.org/competitions/27446#results>

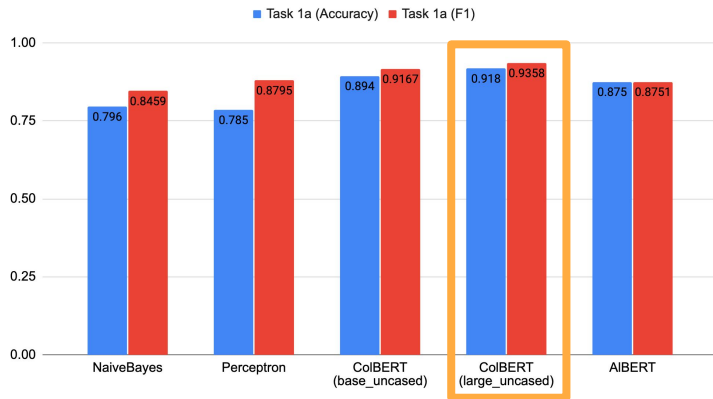
Contribution and Conclusions

- The first task to combine the humor and offensive detection
- Implement four approaches to solve the problem
 - CoBERT-LARGE (performed the best, but requires more training time)
 - CoBERT-BASE
 - ALBERT
 - Naive Bayes and Perceptron
- Achieved top rankings among 43 teams globally
- Humor detection seems easy (task1a), but deciding whether a joke is humorous or offensive is still a very challenging task (task1c)

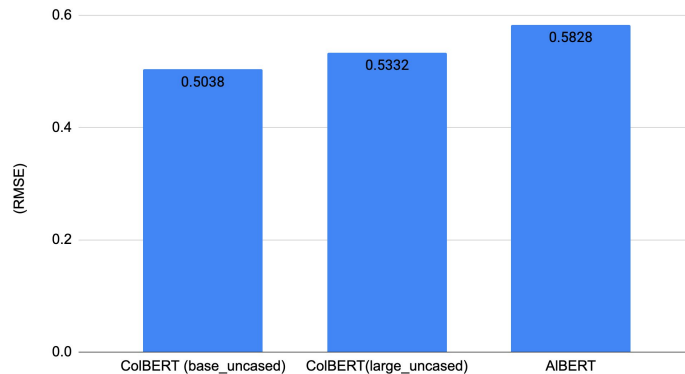
Back up

Backup

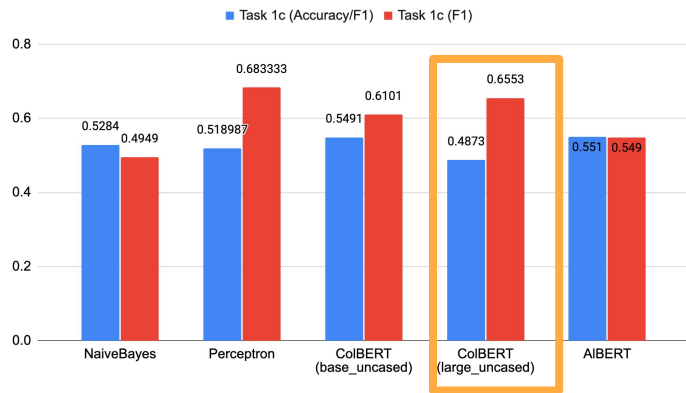
Task 1a (Accuracy and F1)



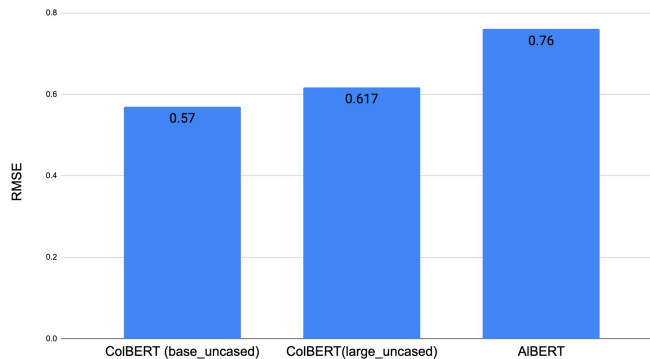
Task 1b (RMSE)



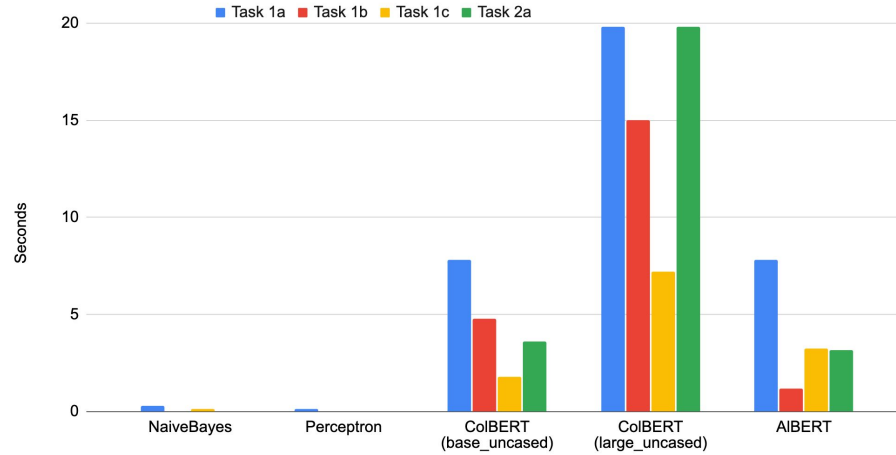
Task 1c (Accuracy and F1)



Task 2a (RMSE)



Predicting Time for 1000 test texts



Examples of Jokes

*“Veni, Vidi, Visa: I came, I saw, I **did a little shopping**” (original poem: I came, I saw, I conquered)*

*My wife tells me I’m a **skeptic**, but I **don’t believe** a word she says.”*

*What is the difference between a nicely dressed man on a tricycle and a poorly dressed man on a bicycle? **A tire.***

Humor Anchor Extraction (Yang et al. 2015)

