
Detecting Adverse Drug Reactions from Twitter through Domain-Specific Preprocessing and BERT Ensembling

— Amy Breden and Lee Moore —

Pharmacovigilance and Adverse Drug Reactions (ADRs)

- Background in healthcare and pharma/biotech
- Pharmacovigilance is critical to ensure the safety of the drugs we use
- Current methods don't make use of social media data
- Idea: use NLP to identify ADRs in Twitter



An Ongoing Competition (and our final score)

Meeting: Social Media Mining for Health Applications (SMM4H) Workshop & Shared Task

- Last Meeting Date: August 2019, Milan

Task: Classify Twitter tweets as including an Adverse Drug Reaction (ADR) mention or not.

Annotated Dataset:

- 25,678 tweets for training
- 4,575 tweets for testing on a competition platform

Performance Metric Refined: **Recall** is most important for our use case

Model	Author	Description	f1 score	precision	recall
Current State-of-the-Art	Chen et al	Retrained BERT on 1.5m scrapped tweets	0.646	0.608	0.689
Our Baseline	Chen at al	BERT fine-tuned	0.618	0.654	0.578
Our model	Amy & Lee	BERT/BioBERT/ClinicalBERT ensemble with preprocessor	0.668	0.590	0.770

Our Models

BERT_{LARGE} - uncased (original)

- 24-layer, 1024-hidden, 16-heads, 340M parameters

BioBERT

- Initialized from BERT_{BASE} (12-layer, 768-hidden, 12-heads, 110M parameters)
- Pre-trained on Biomedical corpora
 - PubMed abstracts (4.5B words)
 - PubMed Central full-text articles (13.5B words)

Bio + clinicalBERT (ClinicalBERT)

- Initialized from BioBERT
- Pre-trained on 2M clinical notes

Domain-Specific Preprocessor

Medical social media preprocessor

- A customizable pipeline for social media text normalization ([Sarker, 2017](#))
- Data-Driven Lexical Normalization for Medical Social Media ([Dirkson et al, 2019](#))

Preprocessor: Our Implementation

Component	Status	Description
Anonymize	Unchanged	Replaces URLs with "-URL-", removes usernames from email addresses while retaining the email domain, and removes copyright and trademark logos
Expand contractions	Removed	Expands contractions (e.g. expand "don't" into "do not")
Expand abbreviations	Removed	Expands domain-specific (patient forum) abbreviations
Sarker normalizer	Removed	Normalizes British English to American English, normalizes slang abbreviations, and removes repeated characters
Spell corrector	Removed	For tokens not in dictionary, updates spelling based on edit distance, similarity, and relative frequency in a biomedical reference dictionary
Replace handles	Modified	Replaces twitter handles with "-TH-" (original code removed handles)
Remove hashtags	Added	If word starts with "#", remove "#" symbol but keep the word
Lowercase	Modified	Converts tokens to lowercase (required for Drug Normalizer step)
Drug normalize	Added	Converts drug brand names to generic names

Preprocessor: Spell Corrector

Original

arthrotec

followfriday

tt

xanaxdreams

toastie

spondyloarthritis

surgeryblue

dysthymia

aderol

injectable

Corrected

arthritis

follow Friday

ass

xanax dreams

chocolate

arthritis

surgery blue

schizophrenia

adderall

injection

Original

bussing

undersleep

doxycycline

sai

macrobid

crazytown

chd

hbp

sleepdisorders

infront

Corrected

missing

understand

ciprofloxacin

said

ciprofloxacin

crazy town

ischemic

metoprolol

sleep disorders

instead



Incorrect translation of a medical term



Incorrect translation of a non-medical term

Preprocessor: Drug Normalizer

Component	Status	Description
Anonymize	Unchanged	Replaces URLs with "-URL-", removes usernames from email addresses while retaining the email domain, and removes copyright and trademark logos
Expand contractions	Removed	Expands contractions (e.g. expand "don't" into "do not")
Expand abbreviations	Removed	Expands domain-specific (patient forum) abbreviations
Sarker normalizer	Removed	Normalizes British English to American English, normalizes slang abbreviations, and removes repeated characters
Spell corrector	Removed	For tokens not in dictionary, updates spelling based on edit distance, similarity, and relative frequency in a biomedical reference dictionary
Replace handles	Modified	Replaces twitter handles with "-TH-" (original code removed handles)
Remove hashtags	Added	If word starts with "#", remove "#" symbol but keep the word
Lowercase	Modified	Converts tokens to lowercase (required for Drug Normalizer step)
Drug normalize	Added	Converts drug brand names to generic names

Preprocessor: Drug Normalizer

Without drug normalizer

- “this se ##ro ##quel got me sleepy”
- “z ##y ##p ##re ##xa is supposed to make you a fat bastard i got off lucky . lit ##hium makes my hand shake . bi ##pol ##ar or s ##chi ##zo drugs .”

With drug normalizer:

- “this que ##tia ##pine got me sleepy”
- “o ##lan ##za ##pine is supposed to make you a fat bastard i got off lucky . lit ##hium makes my hand shake . bi ##pol ##ar or s ##chi ##zo drugs .”

Key Challenge: Model Variability

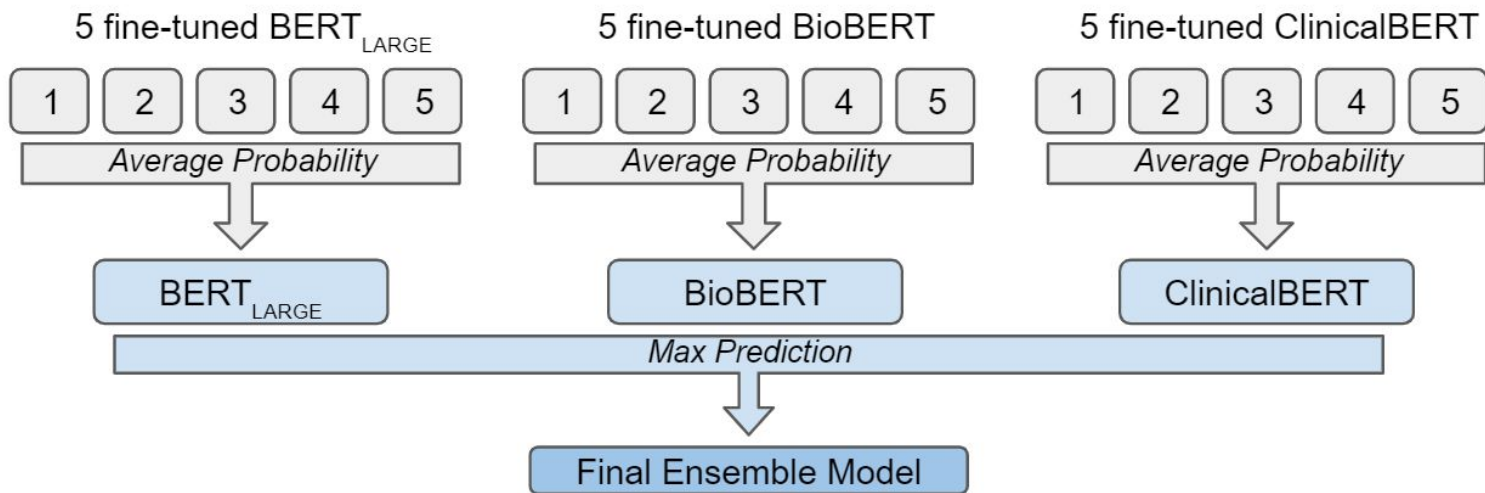
Each fine-tuned model had slightly different results as seen in our baseline model:

	1	2	3	4	5	Avg
F_1	0.59	0.63	0.62	0.63	0.62	0.618
P	0.64	0.66	0.64	0.66	0.67	0.654
R	0.55	0.61	0.61	0.60	0.57	0.587

Our sensitivity analysis did not uncover any consistent opportunity to reduce the variability (measured by standard deviation)

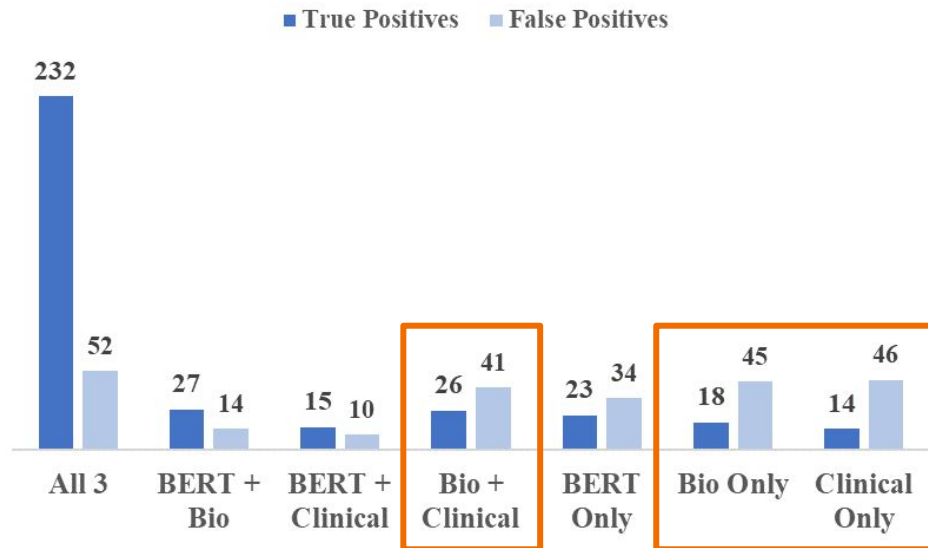
Index	Scenario	BERT _{LARGE} ¹		BioBERT	
		F_1 StDev	Recall StDev	F_1 StDev	Recall StDev
0	Original	0.014	0.017	0.007	0.015
1	Training set with 3x positive examples	0.010	0.024	0.011	0.011
2	Batch size = 128	0.013	0.022	0.004	0.011
3	Loss with 2x positive weights	0.007	0.013	0.011	0.017
4	Dropout = 0.2	0.012	0.022	0.007	0.006
5	1&2 combined	0.007	0.009	0.012	0.013
6	1&2&3&4: All combined	0.008	0.011	0.007	0.020

Max Ensembling Approach



Model Results

<i>No preprocessor</i>	BERT	Bio BERT	Clinical BERT	Max Ensemble
F1-score	0.6446	0.5915	0.5809	0.6378
Precision	0.6695	0.6200	0.6180	0.5663
Recall	0.6214	0.5655	0.5479	0.7300
<i>Preprocessor</i>	BERT	Bio BERT	Clinical BERT	Max Ensemble
F1-score	0.6475	0.6153	0.6212	0.6681
Precision	0.6907	0.6577	0.6360	0.5900
Recall	0.6097	0.5793	0.6076	0.7700



Post-evaluation competition leaderboard

Amy

Lee

Sub-Task 1, ADR classification results							
#	User	Entries	Date of Last Entry	Team Name	F1 ▲	P ▲	R ▲
1	aab213	20	11/23/19		0.67 (1)	0.6081 (6)	0.75 (2)
2	leebean337	10	11/17/19		0.67 (2)	0.6000 (7)	0.76 (1)
3	manrajsingh	11	08/13/19		0.65 (3)	0.6974 (1)	0.60 (7)
4	prady	32	09/05/19		0.64 (4)	0.6605 (3)	0.62 (5)
5	sgondane	37	10/24/19		0.64 (5)	0.6136 (5)	0.66 (4)

<https://competitions.codalab.org/competitions/20798#results>

Questions?

Appendix

False Positive Examples

- @hcorini ive been on actemra w/ mtx for 11 mths which works reasonably(not 100%)before that enbrel and cimzia w/mtx which didn't work at all
- @lulubot75 the shaking is a common side effect of effexor
- @burberrywife lamictal has a black box warning for rash? and they titrate you slowly. no side effects at all for me. bad migraine today.
- paxil why you not let me sleep?!?
- of all things, a summer cold! feeling so weak... tysabri infusion not until tuesday :(
- oddball medication side effects: levofloxacin can cause achilles' tendon rupture.
- the one time i'm glad for seroquel memory loss
- i can only drink when i take adderall and i have no idea why. maybe because it keeps me alert and awake
- seroquel as the sleeping aid. i'm about to knocc out hopefully for the whole night.

ClinicalBERT uniquely identified tweets

- morning fluoxetine burp makes me want to ralph.
- this oxycodone got me feeling goodddd... shit dangerous
- i have a lozenge i cant feel my throat
- @newday 'non hormonal treatment' for hot flashes not new; 10mg/paxil(gsk), an ssri now called 'brisdelle'. ssri withdrawal syndrome alert!
- my ventolin inhaler never works
- @thestomabaglady i still get the pins & needles feeling in my hands since doing humira
- @tjanaepierce lol no. adderall only makes me emotional, no physical pain.
- lamotrigine causing steven johnson syndrome

BioBERT uniquely identified tweets

- @quetiapinefella /swiftly taken off that for vimpat. vimpat made seizures less well controlled but less side effects. on that i had speech/
- i need more medication than this. 🤢 my body is already building a tolerance for the prozac i've been taking.
- has anyone ever had to take avelox? i need some tricks to curb the nausea and help me with the hypotension. #avelox #antibiotic #evil
- @dkegel @ppiaction @msmagazine @itsmepanda1 and "we" expected and stated vioxx was safe and effective. reduced swelling and caused attacks ❤️
- anyone else have worsening #acne problems since starting #enbrel for autoimmune disease?